Recursive construction of continuum random trees and some contributions to the theory of tree-valued stochastic processes

Franz Rembart

Department of Statistics

University of Oxford

This dissertation is submitted for the degree of

Doctor of Philosophy

Christ Church Trinity, 2016

“A mathematician is a device for turning coffee into theorems.”

Alfréd Rényi

Declaration

I hereby declare that except where specific reference is made to the work of others, the contents of this dissertation are original. This thesis is submitted to the University of Oxford for the degree Doctor of Philosophy, and has not been submitted in whole or in part for consideration for any other degree or qualification in this, or any other university.

Franz Rembart

Trinity, 2016

Acknowledgements

First and foremost, I would like to thank my supervisor Matthias Winkel for his great support, guidance and patience during my DPhil. I very much appreciate all his efforts to help me grow mathematically and personally, and value all the fruitful discussions, comments and suggestions over the years.

I would also like to thank Christina Goldschmidt and Bénédicte Haas for agreeing to examine this thesis.

Furthermore, I am grateful to all my previous supervisors and mathematical supporters throughout the years, in particular to Nina Gantert, who made and kept me interested in Probability Theory. I would like to thank her and Kavita Ramanan for introducing me to mathematical research and for their great support.

Finally, a special thank you to everyone who supported or challenged me on my way, and to all people who made my time at Oxford a mathematically as well as personally enriching and broadening experience.

Abstract

We introduce a general recursive method to construct continuum random trees (CRTs) from i.i.d. copies of a string of beads, that is, any random interval equipped with a random discrete measure. We prove the existence of these CRTs as a new application of the fixpoint method formalised in high generality by Aldous and Bandyopadhyay. We further apply this recursive construction to "embed" Duquesne and Le Gall's stable tree into a binary compact CRT in a way that solves an open problem posed by Goldschmidt and Haas. Some of these developments are carried out in a space of ∞-marked metric spaces generalising Miermont's notion of a k-marked metric space.

Furthermore, we introduce a family of branch merging operations on continuum trees and show that Ford CRTs are distributionally invariant. This operation is new even in the special case of the Brownian CRT, which we explore in more detail. The operations are based on spinal decompositions and a regenerativity preserving merging procedure of (α, θ)-strings of beads, that is, regenerative strings of beads arising in the limit of ordered (α, θ)-Chinese restaurant processes as introduced by Pitman and Winkel. Indeed, we iterate the branch merging operation recursively and give a new approach to the leaf embedding problem on Ford CRTs related to $(\alpha, 2-\alpha)$-tree growth processes.

Keywords: string of beads, continuum random tree, stable tree, marked metric space, recursive distribution equation, line-breaking construction, Chinese restaurant process, regenerative interval partition, Poisson-Dirichlet, Ford CRT, Brownian CRT, branch merging, tree-valued Markov process

Table of contents

Chapter 1

Introduction

Random trees have ever since played a central role in probability theory and combinatorics due to their wide range of applications in biology, physics, computer science and other areas, see e.g. [10, 21, 25, 26, 30, 36, 51, 63]. The study of large discrete trees has directed the attention to scaling limits of random tree structures. The limiting objects, provided their existence, have been identified as continuum random trees (CRTs) with potentially complex branching structures such as branch points which are dense on the edges, or an uncountable number of leaves, cf. [7, 8, 9, 46, 47]. We will use the framework of an $\mathbb{R}$-*tree*, that is, a separable metric space $(\mathcal{T}, d)$ such that for every $\sigma_1, \sigma_2 \in \mathcal{T}$,

(i) there is an isometric map $h_{\sigma_1,\sigma_2} : [0, d(\sigma_1, \sigma_2)] \to \mathcal{T}$ such that $h_{\sigma_1,\sigma_2}(0) = \sigma_1$ and $h_{\sigma_1,\sigma_2}(d(\sigma_1, \sigma_2)) = \sigma_2$, and

(ii) for every injective path $q : [0, 1] \to \mathcal{T}$ with $q(0) = \sigma_1$ and $q(1) = \sigma_2$ we have $q([0, 1]) = h_{\sigma_1,\sigma_2}([0, d(\sigma_1, \sigma_2)])$.

The $\mathbb{R}$-trees in this thesis are rooted, i.e. they have a distinguished vertex ρ called the *root*. An $\mathbb{R}$-tree can be equipped with a probability measure μ. We call $(\mathcal{T}, d, \mu)$ a *continuum tree* if $(\mathcal{T}, d)$ is an $\mathbb{R}$-tree, and μ is a probability measure which is supported by the set of non-root degree-1 vertices of $\mathcal{T}$ (the *leaves* of $\mathcal{T}$), has no atoms, and assigns positive mass to each subtree $\mathcal{T}_x$ above each point $x \in \mathcal{T}$ that is not a leaf. This thesis is centred around various models of continuum random trees, that is, random variables taking values in the space of continuum trees, see e.g. Section 2.3.1 for a precise introduction of $\mathbb{R}$-trees and continuum (random) trees.

Ford and stable CRTs

David Aldous [7, 8, 9] published his groundbreaking work on the Brownian CRT in a three-paper series in 1991-93. The Brownian CRT can be considered as the prototype of a CRT, and appears naturally as a universal scaling limit of discrete random trees, random maps and other graph structures. For instance, Aldous showed that the Brownian CRT is the scaling limit as $n \to \infty$ of Galton-Watson trees with critical offspring distribution of finite variance conditioned to have total progeny n.

Various equivalent definitions of the Brownian CRT were given in Aldous' initial work. We use the Poissonian line-breaking construction of the Brownian CRT, directly building the Brownian CRT as an increasing limit of discrete trees with edge lengths.

Algorithm (Aldous' line-breaking construction of the Brownian CRT)**.** Let $(S_k, k \geq 0)$ be the times of an inhomogeneous Poisson process of intensity tdt. We build a sequence of trees $(\mathcal{T}_k, k \geq 1)$ with edge lengths as follows.

0. Let $\mathcal{T}_1$ be the tree consisting of one edge of length S_0, connecting the root ρ and a leaf Ω_1.

For $k \geq 1$, to construct $\mathcal{T}_{k+1}$ conditionally given $\mathcal{T}_k$,

1. select J_k uniformly at random according to the length measure on $\mathcal{T}_k$;

2. attach to J_k an edge of length $S_k - S_{k-1}$ leading to a new leaf Ω_{k+1}.

The metric space completion $\mathcal{T} := \overline{\bigcup_{k \geq 1} \mathcal{T}_k}$ is called the *Brownian Continuum Random Tree*, cf. Figure 1.1 for a realisation of the Brownian CRT.

The Brownian CRT is equipped with a natural mass measure that can be defined as the a.s. weak limit as $k \to \infty$ of the empirical mass measure on the leaves of $\mathcal{T}_k$. The Brownian CRT is self-similar in the sense that subtrees above fixed heights (and more generally) are scaled independent Brownian CRTs. In the past 25 years, the literature on random trees has been enriched by a variety of other models. We will mostly be

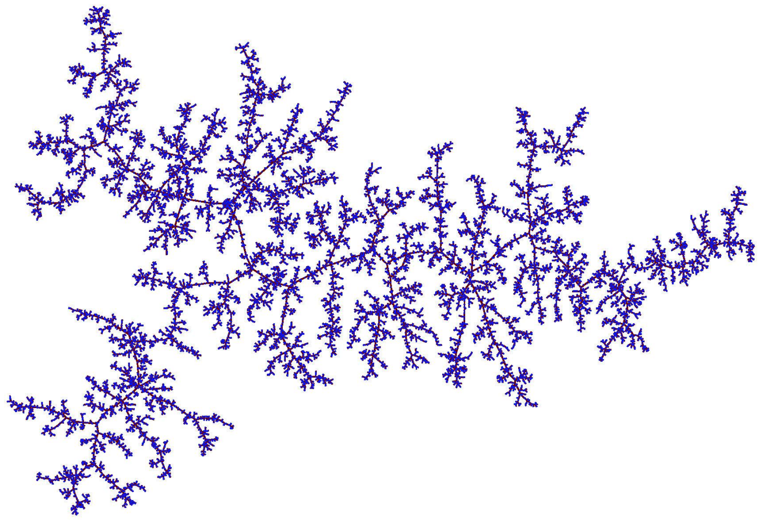

Fig. 1.1 An approximation of a realisation of the Brownian CRT.
Picture by Igor Kortchemski.
.

interested in Ford CRTs and stable CRTs, parametrised by $\alpha \in (0,1)$ and $\beta \in (0,1/2]$, respectively. They both have a self-similarity property, and contain the Brownian CRT as a special case ($\alpha = \beta = 1/2$). Note the non-standard parametrisation for stable trees used in this thesis ($\beta \in (0,1/2]$ rather than $\alpha' = 1/(1-\beta) \in (1,2]$).

Ford CRTs are binary CRTs, and can be defined via a line-breaking construction, generalising Aldous' construction of the Brownian CRT. The cut points are the arrivals of the Mittag-Leffler Markov chain $(S_k, k \geq 0)$ with parameters $\alpha \in (0,1)$ and $\theta = 1-\alpha$. The *Mittag-Leffler Markov chain* $(S_k, k \geq 0)$ of parameters (α, θ) for $\alpha \in (0,1)$ and $\theta > 0$ is the Markov chain starting from $S_0 \sim \mathrm{ML}(\alpha, \theta)$, where $\mathrm{ML}(\alpha,\theta)$ is the Mittag-Leffler distribution with parameters (α, θ), and with transition density

$$f_{S_{k+1}|S_k=z}(y) = f(z,y) = \frac{1-\alpha}{\Gamma(1/\alpha)}(y-z)^{1/\alpha-2}\frac{y g_\alpha(y)}{g_\alpha(z)}, \quad y > z > 0,$$

where $g_\beta(\cdot)$ denotes the density of $\mathrm{ML}(\beta, 0)$, cf. [39, 44] and e.g. Section 2.2.1 for a definition of the Mittag-Leffler distribution. Note that $(S_k, k \geq 0)$ depends on θ only via the initial state $S_0 \sim \mathrm{ML}(\beta, \theta)$.

The attachment procedure in the Ford CRT construction is slightly more enhanced than in the Brownian setting. In particular, at step $k \geq 1$, we select an edge E_k of $\mathcal{T}_k$ proportionally to length. Conditionally given the selected edge E_k, we choose the attachment point J_k uniformly according to the length measure on E_k if E_k is internal, and according to a certain Beta split if E_k is external. The following line-breaking construction of Ford CRTs was given by Haas et al. [42].

Algorithm (Haas et al.'s line-breaking construction of Ford CRTs)**.** Let $\alpha \in (0,1)$, and let $(S_m, m \geq 0)$ be the Mittag-Leffler Markov chain with parameters $(\alpha, 1-\alpha)$. We build a sequence of trees $(\mathcal{F}_m, m \geq 1)$ with edge lengths as follows.

0. Let $\mathcal{F}_1$ be the tree consisting of one edge of length S_0, connecting the root ρ and a leaf Ω_1.

For $m \geq 1$, to construct $\mathcal{F}_{m+1}$ conditionally given $\mathcal{F}_m$,

1. select an edge E_m of $\mathcal{F}_m$ proportionally to length, and select a point $J_m \in E_m$ as follows; if E_m is an internal edge of $\mathcal{F}_m$, sample J_m uniformly from the normalised length measure on E_m; if E_m is external, sample $B_m \sim \text{Beta}(1, 1/\alpha - 1)$ independently, and let J_m be such that E_m is split into length proportions B_m and $1 - B_m$ with the part of proportion B_m being closer to the root;

2. attach to J_m an edge of length $S_m - S_{m-1}$ leading to a new leaf Ω_{m+1}.

The tree $\mathcal{F} := \overline{\bigcup_{m \geq 1} \mathcal{F}_m}$ is called *Ford CRT* of index $\alpha \in (0,1)$.

Note that, for $\alpha = 1/2$, we have $(\alpha, 1-\alpha) = (1/2, 1/2)$ and $B_m \sim U([0,1])$, $m \geq 1$. One can show that, in this case, the Mittag-Leffler Markov chain corresponds to the sequence of times of an inhomogeneous Poisson process of intensity tdt (up to a scaling factor), i.e. we recover Aldous' line-breaking construction of the Brownian CRT.

Le Gall and Le Jan [48] introduced a family of multifurcating CRTs, so-called Lévy trees, which capture all possible scaling limits of Galton-Watson trees, cf. [28]. Stable trees are a special subclass of Lévy trees parametrised by $\beta \in (0, 1/2]$; they were introduced by Duquesne and Le Gall [28, 29]. The stable tree of parameter $\beta \in (0, 1/2]$ is the scaling limit as $n \to \infty$ of Galton-Watson trees with critical offspring distribution in the domain of attraction of a stable distribution of index $1/(1-\beta)$ conditioned to have total progeny n.

Goldschmidt and Haas [39] developed two line-breaking constructions of the stable trees, based on the Mittag-Leffler Markov chain with parameters (β, β), $\beta \in (0, 1/2]$.

The two versions differ with regard to branch point selection in the attachment procedure; one version is based on assigning weights to branch points whereas the second version is based on branch point degrees. We will mostly be interested in the weight-based version of the construction. In contrast to the line-breaking construction of Ford CRTs, only an independent Beta$(1, 1/\beta - 2)$ part of the increment $(S_k - S_{k-1})$ of the Mittag-Leffler Markov chain is used as length of the attached branch, where the remaining part is added to the weight of the branch point of degree ≥ 2 the new edge is attached to. In the following, for any $\mathbb{R}$-tree, $\mathrm{Br}(\mathcal{T})$ denotes the set of branch points of $\mathcal{T}$, and $\mathrm{Leb}(\cdot)$ is the length measure. We write $[n] := \{1, 2, \ldots, n\}$ for $n \in \mathbb{N}$, and $\#S$ for the number of elements of a set S.

Algorithm (Goldschmidt-Haas' line-breaking construction of stable trees)**.** Let $\beta \in (0, 1/2]$. We grow $\mathbb{R}$-trees $\mathcal{T}_k$ with weights $W_k^{(i)}$ in branch points $v_i \in \mathrm{Br}(\mathcal{T}_k)$, $i \in [b_k]$, where $b_k = \#\mathrm{Br}(\mathcal{T}_k)$, $k \geq 0$, as follows.

0. Let $\mathcal{T}_0$ be the tree consisting of one branch of length $S_0 \sim \mathrm{ML}(\beta, \beta)$, let $b_0 = 0$ and $W_0^{(i)} = 0$, $i \geq 1$.

Given $(\mathcal{T}_k, (v_i, i \in [b_k]), (W_k^{(i)}, i \in [b_k]))$ and $S_k = L_k + \sum_{i=1}^{b_k} W_k^{(i)}$, where $L_k = \mathrm{Leb}(\mathcal{T}_k)$,

1. select $I_k = i$ for each branch point $v_i \in \mathrm{Br}(\mathcal{T}_k)$ with probability proportional to $W_k^{(i)}$, $i \in [b_k]$; or select an edge $E_k \subset \mathcal{T}_k$ with probability proportional to its length and let $b_{k+1} = b_k + 1$, $I_k = b_{k+1}$;

2. if an edge E_k is selected, sample $v_{b_{k+1}}$ from the normalised length measure on E_k;

3. sample S_{k+1} with density $f(S_k, \cdot)$ and an independent $B_k \sim \mathrm{Beta}(1, 1/\beta - 2)$; attach to $\mathcal{T}_k$ at $J_k := v_{I_k}$ a new branch of length $(S_{k+1} - S_k)B_k$ to form $\mathcal{T}_{k+1}$; set $W_{k+1}^{(i)} = W_k^{(i)}, i \neq I_k$, and increase the weight of $J_k = v_{I_k}$ to
$$W_{k+1}^{(I_k)} = W_k^{(I_k)} + (S_{k+1} - S_k)(1 - B_k).$$

The tree $\mathcal{T} := \overline{\bigcup_{k \geq 0} \mathcal{T}_k}$ is a *stable tree* of parameter β, cf. Figure 1.2 for a realisation of the stable tree when $\beta = 1/3$.

Note that each branch point has an initial weight of 0 in the case when $\beta = 1/2$, i.e. we obtain a binary tree. One can show that, in this case, we obtain Aldous' line-breaking construction of the Brownian CRT (up to a scaling factor).

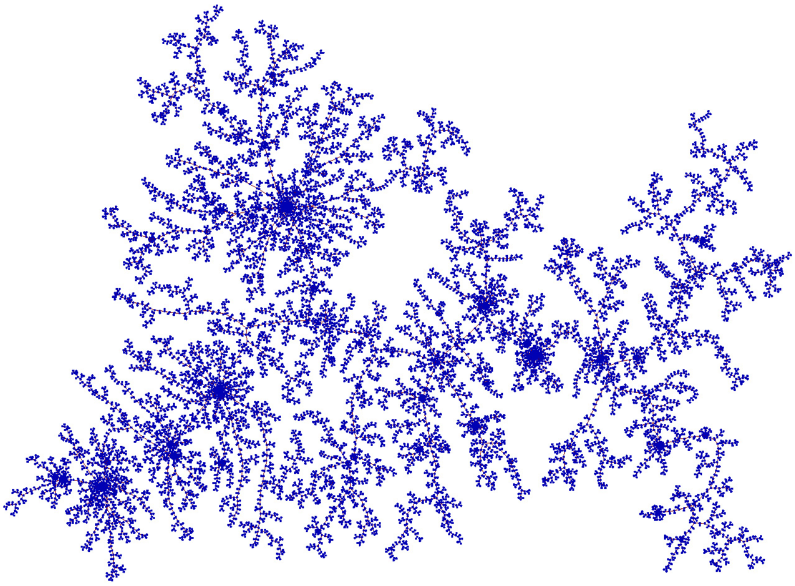

Fig. 1.2 An approximation of a realisation of the stable tree of index $\beta = 1/3$. Picture by Igor Kortchemski.

Regenerative tree growth processes

Ford and stable CRTs, as well as some other types of self-similar CRTs, can be obtained as scaling limits of random tree growth processes, where leaves carry labels in the order of appearance in the growth procedure.

Ford [35] introduced a discrete, binary tree growth model for $\alpha \in (0, 1)$, the so-called Ford alpha-model, that interpolates between several standard models of random tree growth (Yule, uniform and comb models). It was shown by Haas et al. [42] that the distributional scaling limit of the delabelled trees of Ford's alpha-model is a Ford CRT of index $\alpha \in (0, 1)$. The case $\alpha = 1/2$ corresponds to the uniform model whose scaling limit was identified as the Brownian CRT in Aldous [7, 8, 9]. Ford's alpha model is contained as the special case $\theta = 1-\alpha$ in the (α, θ)-tree growth process [60] for $\alpha \in (0, 1)$ and $\theta > 0$ whose delabelled trees have a self-similar CRT as their scaling limit. In a similar manner, Marchal [50] introduced a discrete tree growth process for $\beta \in (0, 1/2]$, whose scaling limit is the stable tree of parameter $\beta \in (0, 1/2]$, see also [23]. We will give precise definitions of these tree growth processes in the relevant chapters.

The mass measures of Ford and stable CRTs, and of the self-similar CRTs related to the (α, θ)-model can, for instance, be obtained via the empirical probability measure on the vertex set in the related discrete tree growth processes. This yields probability measures supported on the set of leaves of these CRTs, and sampling from this measure provides exchangeable leaves.

The leaves with labels in order of appearance in Marchal's model are exchangeable, whereas, in the (α, θ)-tree growth process, leaf labels are exchangeable if and only if $\alpha = \theta = 1/2$, which corresponds to the uniform case related to the Brownian CRT. Since scaling limits are only obtained for the delabelled trees, a fundamental problem related to (α, θ)-tree growth is to embed the leaf labels in the limiting CRT. We refer to this problem as the *leaf embedding problem*.

When leaf labelling is exchangeable this problem can be solved directly by sampling independent leaves from the mass measure on the CRT. In general, the leaf embedding problem was solved in [60] via an (α, θ)-coin tossing construction. Informally, to find the first leaf of the (α, θ)-model in the limiting CRT, we consider a walker climbing up the spine from the root to a(n exchangeable) leaf chosen uniformly at random from the mass measure, tossing a coin for each of the countably infinite number of subtree masses along the spine, where the heads probability depends on the relative remaining mass after each subtree. The walker stops as soon as he sees heads, selects the related subtree, and the same construction is applied recursively in the selected subtree.

Given an embedding of the first leaf, iterating this procedure to embed all subsequent leaves (in particular the second leaf) requires finding the subtree on the spine from the root to leaf 1 which contains leaf 2. This is solved by so-called (α, θ)-*coin tossing sampling* on the spine from the root to the first leaf equipped with projected subtree masses. The notion of a string of beads naturally captures this setting.

Strings of beads

A *string of beads* is an interval of random length equipped with a random discrete measure. Particular emphasis will be on so-called (α, θ)-strings of beads for $\alpha \in (0, 1)$

and $\theta > 0$. They arise in the scaling limit of ordered (α, θ)-Chinese restaurant processes as introduced by Pitman and Winkel [60] in the context of the (α, θ)-model, where they describe the limit of the path from the root to the first leaf equipped with relative projected empirical subtree masses. An (α, θ)-*string of beads* is an interval of length $K \sim \mathrm{ML}(\alpha, \theta)$ equipped with a discrete probability measure whose atom sizes are Poisson-Dirichlet with parameters (α, θ), arranged in a random order that yields a regenerative property.

The recursive embedding of leaf labels in the self-similar CRT obtained from the (α, θ)-model is directly related to (α, θ)-coin tossing sampling on an (α, θ)-string of beads, that allows one to select an atom of an (α, θ)-string of beads such that the two intervals obtained are rescaled independent (α, α)- and (α, θ)-strings of beads (the latter one being further away from the root), which are independent of the Dirichlet$(\alpha, 1-\alpha, \theta)$ mass split between the two intervals and the selected atom (with parameters assigned according to increasing distance from the root).

Outline of this thesis

This thesis consists of two independent, self-contained pieces of research. We give a recursive construction of continuum random trees and various contributions to the theory of tree-valued stochastic processes.

Chapter 2: Recursive construction of binary CRTs and a binary embedding of the stable-line-breaking construction

This chapter constitutes the main part of joint work [65] with my supervisor Dr Matthias Winkel.

We develop a new recursive method to construct continuum random trees based on i.i.d. copies of a string of beads, which is a new application of the fixpoint method developed by Aldous and Bandyopadhyay [13]. This extends the current literature on construction methods for CRTs as studied by Haas, Miermont, Stephenson and others, cf. [14, 40, 61, 66]. Our construction method is based on inductively replacing the atoms of a string of beads by rescaled independent copies of itself. We show that this method builds a compact CRT.

Theorem. *Let $p \geq 1$, $\beta > 1/p$, and let $(\check{\mathcal{T}}_0, \check{\mu}_0)$ be an isometric copy of a string of beads $\xi = ([0,K], \sum_{i\geq 1} P_i \delta_{x_i})$ such that $K = \sup_{i\geq 1}\{x_i : P_i > 0\}$ and $\mathbb{E}[K^p] < \infty$. For $n \geq 0$, to obtain $(\check{\mathcal{T}}_{n+1}, \check{\mu}_{n+1})$, attach to each atom $x \in \check{\mathcal{T}}_n$ of $\check{\mu}_n$ an independent isometric copy of ξ with metric rescaled by $\check{\mu}_n(x)^\beta$, and mass measure rescaled by $\check{\mu}_n(x)$. Then there exists a binary compact CRT $(\check{\mathcal{T}}, \check{\mu})$ such that*

$$\lim_{n\to\infty} \check{\mathcal{T}}_n = \check{\mathcal{T}} \quad a.s.$$

with respect to the Gromov-Hausdorff distance d_{GH}. Furthermore,

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

with respect to the Gromov-Hausdorff Prokhorov distance d_{GHP}.

We use this method to embed the stable tree of index $\beta \in (0, 1/2]$ into a compact binary CRT, which solves an open problem posed by Goldschmidt and Haas [39]: they asked if there is a sensible way to associate a notion of length with the branch point weights arising in their line-breaking construction of the stable trees.

We enhance the line-breaking construction of the stable tree by using the vertex weights to build rescaled Ford CRTs of index $\beta/(1-\beta)$. This construction evolves in a space of ∞-marked metric spaces, generalising Miermont's notion of a k-marked metric space, where the marked components, that is, compact connected non-empty subsets $\mathcal{R}^{(i)} \subset \mathcal{T}, i \geq 1$, of an $\mathbb{R}$-tree $\mathcal{T}$, and the unmarked remainder $\mathcal{T} \setminus \bigcup_{i=1}^\infty \mathcal{R}^{(i)}$ are associated with two different colours. Our two-colour line-breaking construction combines the line-breaking constructions of the stable tree of parameter β and the Ford CRT of index $\beta/(1-\beta)$.

Algorithm (Two-colour line-breaking construction)**.** Let $\beta \in (0, 1/2]$. We grow ∞-marked $\mathbb{R}$-trees $(\mathcal{T}_k^*, (\mathcal{R}_k^{(i)}, i \geq 1))$, $k \geq 0$, as follows.

0. Let $\mathcal{T}_0^*$ be the tree consisting of one branch of length $S_0 \sim \mathrm{ML}(\beta, \beta)$ connecting the root ρ with another degree-1 vertex, let $r_0 = 0$ and $\mathcal{R}_0^{(i)} = \{\rho\}$, $i \geq 1$.

Given $(\mathcal{T}_k^*, (\mathcal{R}_k^{(i)}, i \geq 1))$, let $S_k = \mathrm{Leb}(\mathcal{T}_k^*)$ be the length of $\mathcal{T}_k^*$ and $r_k = \#\{i \geq 1 \colon \mathcal{R}_k^{(i)} \neq \{\rho\}\}$;

1. select an edge $E_k^* \subset \mathcal{T}_k^*$ with probability proportional to its length; if $E_k^* \subset \mathcal{R}_k^{(i)}$ for some $i \in [r_k]$, let $I_k = i$; otherwise, i.e. if $E_k^* \subset \mathcal{T}_k^* \setminus \bigcup_{i\in[r_k]} \mathcal{R}_k^{(i)}$, let $r_{k+1} = r_k + 1$, $I_k = r_{k+1}$;

2. if E_k^* is an external edge of $\mathcal{R}_k^{(i)}$, sample $D_k \sim \mathrm{Beta}(1, 1/\beta - 2)$ and place J_k^* to split E_k^* into length proportions D_k and $1 - D_k$; otherwise, i.e. if $E_k^* \subset \mathcal{T}_k^* \setminus \bigcup_{i\in[r_k]} \mathcal{R}_k^{(i)}$ or if E_k^* is an internal edge of $\mathcal{R}_k^{(i)}$, sample J_k^* from the normalised length measure on E_k^*;

3. sample S_{k+1} with density $f(S_k, \cdot)$ and an independent $B_k \sim \mathrm{Beta}(1, 1/\beta - 2)$; attach to $\mathcal{T}_k^*$ at J_k^* a new branch of length $S_{k+1} - S_k$ to form $\mathcal{T}_{k+1}^*$, and add to $\mathcal{R}_k^{(I_k)}$ the part of length $(S_{k+1} - S_k)(1 - B_k)$ closest to the root to form $\mathcal{R}_{k+1}^{(I_k)}$; set $\mathcal{R}_{k+1}^{(j)} = \mathcal{R}_k^{(j)}$, $j \neq I_k$.

We define $\mathcal{T}^* := \overline{\bigcup_{k\geq 0} \mathcal{T}_k^*}$ and $\mathcal{R}^{(i)} := \overline{\bigcup_{k\geq 0} \mathcal{R}_k^{(i)}}$, $i \geq 1$.

We obtain the correspondence of the branch point weights in the line-breaking construction of the stable tree and the lengths of the marked subtrees in the two-colour line-breaking algorithm. The following theorem will be made precise in Section 2.5.4, Theorem 2.58.

Theorem (Weight-length representation)**.** *Let $(\mathcal{T}_k^*, (\mathcal{R}_k^{(i)}, i \geq 1), k \geq 0)$ be the sequence of ∞-marked $\mathbb{R}$-trees constructed in the two-colour line-breaking construction, and let $\widetilde{W}_k^{(i)} = \mathrm{Leb}(\mathcal{R}_k^{(i)})$ denote the length of $\mathcal{R}_k^{(i)}$, $i \geq 1$, respectively. For $k \geq 1$, contract each component $\mathcal{R}_k^{(i)}$ to a single branch point $\tilde{v}_i$ by using an equivalence relation (cf. Section 2.5.4 for a precise definition), and denote the resulting tree by $\widetilde{\mathcal{T}}_k$. Then*

$$\left(\widetilde{\mathcal{T}}_k, \left(\widetilde{W}_k^{(i)}, i \geq 1\right), k \geq 0\right) \stackrel{d}{=} \left(\mathcal{T}_k, \left(W_k^{(i)}, i \geq 1\right), k \geq 0\right),$$

see Figure 1.3 for an illustration of the weight-length representation.

The sequences $(\mathcal{T}_k, k \geq 1)$ and $(\mathcal{F}_m, m \geq 1)$ converge to CRTs $\mathcal{T}$ and $\mathcal{F}$, respectively. Similarly, we obtain convergence of the sequence $(\mathcal{T}_k^*, (\mathcal{R}_k^{(i)}, i \geq 1), k \geq 0)$ with respect to d_{GH}^{∞}, the analogue of the Gromov-Hausdorff distance on the space of ∞-marked $\mathbb{R}$-trees.

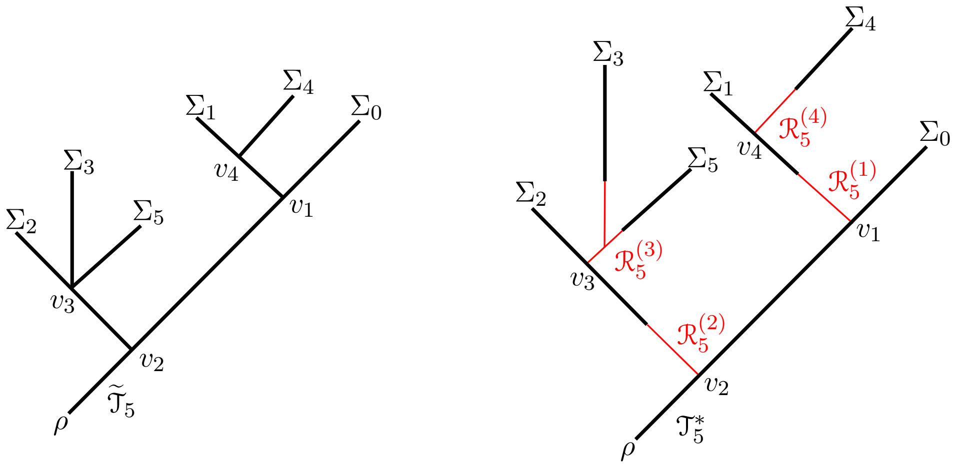

Fig. 1.3 Example of $\widetilde{\mathcal{T}}_5$ with four branch points $v_1, \ldots, v_4$. Branch point weights $W_5^{(1)}, \ldots, W_5^{(4)}$ are represented as lengths of marked subtrees $\mathcal{R}_5^{(1)}, \ldots, \mathcal{R}_5^{(4)}$.

Theorem. *Let the sequence* $(\mathcal{T}_k^*, (\mathcal{R}_k^{(i)}, i \geq 1), k \geq 0)$ *be as above. Then there exists a compact CRT* $\mathcal{T}^*$ *and a sequence of i.i.d. Ford CRTs* $(\mathcal{F}^{(i)}, i \geq 1)$ *of index* $\beta' = \beta/(1-\beta)$ *and scaling factors* $(C_i, i \geq 1)$ *such that*

$$\lim_{k\to\infty} \left(\mathcal{T}_k^*, \left(\mathcal{R}_k^{(i)}, i \geq 1\right)\right) = \left(\mathcal{T}^*, \left(\mathcal{R}^{(i)}, i \geq 1\right)\right) \quad a.s.$$

with respect to d_{GH}^{∞} *where* $\mathcal{R}^{(i)} = C_i \mathcal{F}^{(i)}$.

In fact, the trees $(\mathcal{T}_k^*, (\mathcal{R}_k^{(i)}, i \geq 1))$ are equipped with mass measures μ_k^*, $k \geq 0$, and we will show that

$$\lim_{k\to\infty} \left(\mathcal{T}_k^*, \left(\mathcal{R}_k^{(i)}, i \geq 1\right), \mu_k^*\right) = \left(\mathcal{T}^*, \left(\mathcal{R}^{(i)}, i \geq 1\right), \mu^*\right) \quad \text{a.s.}$$

with respect to $d_{\mathrm{GHP}}^{\infty}$ for some compact CRT $(\mathcal{T}^*, \mu^*)$, where $d_{\mathrm{GHP}}^{\infty}$ is the analogue of the Gromov-Hausdorff-Prokhorov distance on the space of ∞-marked $\mathbb{R}$-trees.

The compactness of the CRT $\mathcal{T}^*$ can be shown by applying the recursive CRT construction method presented in this thesis (cf. Theorem 2.1) to a specific distribution ν of a string of beads, obtained by concatenating an independent $(\beta, 1-2\beta)$- and an independent (β, β)-string of beads rescaled by the two parts of an independent Beta$(1-2\beta, \beta)$ variable.

We obtain a stable tree $\mathcal{T}$ embedded in a binary compact CRT $\mathcal{T}^*$.

Theorem. *The tree $\widetilde{\mathcal{T}}$, obtained from $\mathcal{T}^*$ by contracting each component $\mathcal{R}^{(i)}$ to a single branch point $\tilde{v}_i$, is a stable tree of parameter β.*

Eventually, we develop a two-colour discrete tree growth process, whose scaling limit of the reduced trees spanned by the root and the first $k+1$ leaves, is the tree $(\mathcal{T}_k^*, (\mathcal{R}_k^{(i)}, i \geq 1))$ of the two-colour line-breaking construction for any $k \geq 0$.

Chapter 3: Branch merging on continuum trees with applications to regenerative tree growth

The article version [64] of this chapter appeared in the Latin American Journal of Probability and Mathematical Statistics (ALEA) in July 2016.

The main contribution of this paper is a family of branch merging operations on continuum trees. These are based on merging underlying strings of beads, as the notion of a string of beads naturally captures the branches of reduced trees equipped with projected subtree masses.

We can merge independent (α, θ_i)-strings of beads for $\alpha \in (0,1)$ and $\theta_1, \ldots, \theta_n > 0$ in a way that preserves the regenerative property. Figure 1.4 illustrates the following algorithm in the case when $k = 3$.

Algorithm/Theorem (Merging of (α, θ)-strings of beads)**.** *Let $\alpha \in (0,1)$ and let $\theta_1, \ldots, \theta_k > 0$ for some $k \in \mathbb{N}$. Furthermore, let $E_i = (\rho_i, \rho_i')$, $i \in [k]$, be k disjoint intervals and let μ be a mass measure on $E = \bigcup_{i=1}^n E_i$ such that $(\mu(E_1), \ldots, \mu(E_k)) \sim$* Dirichlet$(\theta_1, \ldots, \theta_k)$*, and $(\mu(E_i)^{-\alpha} E_i, \mu(E_i)^{-1} \mu \restriction_{E_i})$ are independent (α, θ_i)-strings of beads, $i \in [k]$, respectively. Construct $E_{\uplus}$ as follows. Set $E^{(1)} := E$, $E_i^{(1)} := E_i$, $\rho_i^{(1)} := \rho_i$, $i \in [k]$. For $n \geq 1$,*

1. *select an interval $E_i^{(n)}$ proportionally to its mass assigned by μ;*

2. *conditionally given $E_i^{(n)}$, select an atom X_n of $E_i^{(n)}$ according to (α, θ_i)-coin tossing sampling, and let $a_n := \rho_i^{(n)}$ and $b_n := X_n$;*

3. *update $\rho_i^{(n+1)} := X_n$, and $\rho_j^{(n+1)} := \rho_j^{(n)}$ for $j \in [k] \setminus \{i\}$, $E_i^{(n+1)} := (\rho_i^{(n+1)}, \rho_i')$ for $i \in [k]$ and $E^{(n+1)} := \bigcup_{i=1}^n E_i^{(n+1)}$.*

Align the interval components $(a_n, b_n]$, $n \geq 1$, by identifying b_n with a_{n+1}, $n \geq 1$, to obtain an interval $E_{\biguplus}$. Then, $(E_{\biguplus}, \mu)$ is an $(\alpha, \sum_{i=1}^{k} \theta_i)$ string of beads.

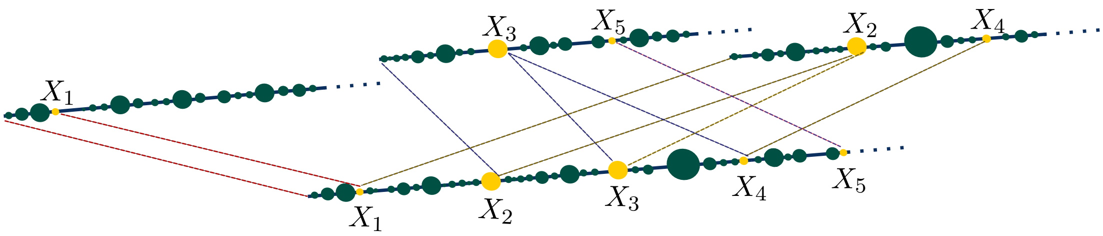

Fig. 1.4 Example of merging of strings of beads for $k = 3$. First five cut points $X_1, \ldots, X_5$ are displayed.

Based on merging of (α, θ)-strings of beads, we define branch merging on Ford CRTs. The branch merging operation on a Ford CRT of index $\alpha \in (0, 1)$ is defined via an embedding of the first two leaves of the related $(\alpha, 1-\alpha)$-tree growth process using the $(\alpha, 1-\alpha)$-coin tossing construction, aligning the three branches of the reduced tree structure and replanting the subtrees in a certain random order. Embedding leaves 1 and 2 of the $(\alpha, 1-\alpha)$-tree growth process (Ford's alpha model) into its scaling limit, a Ford CRT of parameter α, yields two rescaled independent $(\alpha, 1-\alpha)$-strings of beads and one rescaled independent (α, α)-string of beads, subject to a Dirichlet$(1-\alpha, 1-\alpha, \alpha)$ mass split. Merging these strings of beads and replanting the subtrees at the related atoms, yields an $(\alpha, 2-\alpha)$-strings of beads with subtrees attached to its atoms. We obtain the self-similar CRT arising as scaling limit of the $(\alpha, 2-\alpha)$-model with the first leaf embedded. It is well-known that this CRT is as a Ford CRT of index α.

Theorem. *Ford CRTs are distributionally invariant under branch merging.*

Applying the branch merging procedure recursively in subtrees, we obtain an embedding of all leaves of the $(\alpha, 2-\alpha)$-model in the corresponding CRT, given an embedding of the leaves of the $(\alpha, 1-\alpha)$-model in a Ford CRT of parameter $\alpha \in (0, 1)$.

Branch merging is particularly simple in the case when $\alpha = \theta = 1/2$ since $(1/2, 1/2)$-coin tossing sampling corresponds to uniform sampling from the mass measure, i.e. leaves are sampled independently from the mass measure on the CRT and the arising $(1/2, 1/2)$-strings of beads are merged via uniform sampling. Since the $(1/2, 3/2)$-tree growth process also has the Brownian CRT as its scaling limit, we obtain a remarkable simplification of leaf embedding for the $(1/2, 3/2)$-tree growth process.

Theorem. *The Brownian CRT is distributionally invariant under branch merging. Furthermore, the leaf embedding problem related to the* $(1/2, 3/2)$*-tree growth process can be solved by sampling leaves independently from the mass measure of the Brownian CRT and applying branch merging recursively.*

We use this simplified branch merging operation based on uniform sampling from the mass measure as transition rule of a new tree-valued Markov process, the so-called *Branch Merging Markov Chain* (BMMC). Tree-valued Markov processes have attracted particular interest in recent years. Various operations such as cutting, pruning or root-growth with re-grafting have been studied by Abraham, Broutin, Delmas, Evans, He, Pitman, Wang, Winter, and others, cf. [1, 2, 3, 5, 6, 11, 12, 16, 18, 19, 33, 43, 55, 56].

The BMMC can be defined in a discrete as well as a continuous setting. In the continuous case we obtain a continuum-tree valued Markov process, cf. Figure 1.5. The following is a direct consequence of the distributional invariance of the Brownian CRT under branch merging.

Theorem. *The Brownian CRT is a stationary distribution of the BMMC.*

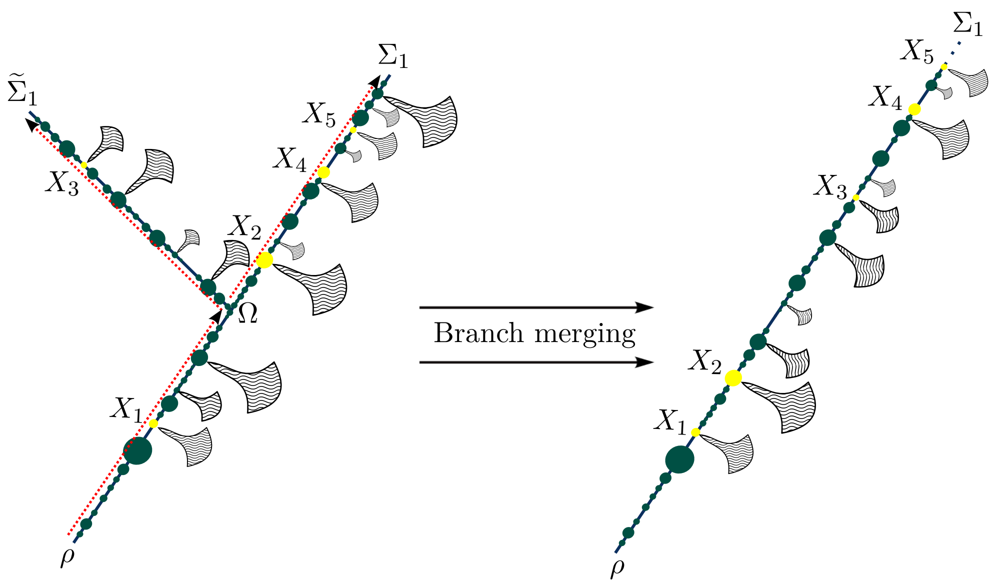

Fig. 1.5 Transition rule of the continuous BMMC. Reduced tree structure related to two independently sampled leaves Σ_1 and $\widetilde{\Sigma}_1$ with subtrees attached in the atoms. Strings of beads are merged and subtrees are carried forward.

In the discrete setting, for any $n \geq 1$, we consider an analogue of the BMMC on the space of discrete (rooted and unlabelled) trees with n leaves and no degree-2 vertices. We show the following.

Theorem. *The discrete BMMC on the space of discrete (rooted and unlabelled) trees with n leaves and no degree-2 vertices has a unique stationary distribution, which is supported on the space of binary discrete (rooted and unlabelled) trees with n leaves and no degree-2 vertices.*

In a nutshell, in the continuous case, we can identify a stationary distribution of the BMMC (the Brownian CRT) but we do not know if it is unique (up to scaling distances by a random factor). In contrast, the discrete BMMC can be shown to have a unique stationary distribution. As there is a direct relation between the uniform model and the Brownian CRT, we conjecture the following.

Conjecture. *The BMMC on the space of (isometry classes of) continuum trees has a unique stationary distribution, which is the law of the Brownian CRT, and the BMMC converges in distribution to the Brownian CRT with respect to the Gromov-Hausdorff-Prokhorov topology (both up to scaling distances by a random factor).*

Chapter 2

Recursive construction of binary CRTs and a binary embedding of the stable line-breaking construction

Abstract

We introduce a general recursive method to construct continuum random trees (CRTs) from i.i.d. copies of a string of beads, that is, any random interval equipped with a random discrete measure. We prove the existence of these CRTs as a new application of the fixpoint method formalised in high generality by Aldous and Bandyopadhyay. We apply this recursive method to "embed" Duquesne and Le Gall's stable tree into a binary compact CRT in a way that solves an open problem posed by Goldschmidt and Haas. Some of these developments are carried out in a space of ∞-marked metric spaces generalising Miermont's notion of a k-marked metric space.

Keywords: string of beads, continuum random tree, stable tree, marked metric space, recursive distribution equation, line-breaking construction

2.1 Introduction

In this chapter, we introduce a new recursive method to construct continuum random trees (CRTs) from i.i.d. copies of a string of beads, that is, any random interval equipped with a random discrete measure. Our construction is based on the concept of a recursive tree framework as introduced by Aldous and Bandyopadhyay [13] to unify various constructions that relate to some *discrete* branching structure of potentially infinite depth, but not a priori to construct CRTs. This allows us to go beyond the class of self-similar CRTs constructed by Haas, Miermont and Stephenson [40, 66], and in particular generalises the bead-splitting processes of [61].

In this setting, we build a random $\mathbb{R}$-tree from a string of beads and rescaled i.i.d. $\mathbb{R}$-trees attached to its atoms, i.e. we view a random tree $\mathcal{T}$ as the output of a map ϕ_β, taking a sequence of i.i.d. $\mathbb{R}$-trees $(\mathcal{T}_i, i \geq 1)$ and attaching these trees to the locations of the atoms of a string of beads ξ with lengths rescaled by the β-power of the respective atom mass for some $\beta \in (0, \infty)$,

$$\mathcal{T} := \phi_\beta(\xi, \mathcal{T}_i, i \geq 1). \tag{2.1}$$

Given some distribution on the space of strings of beads, (2.1) yields a map from $\mathcal{P}(\mathbb{T})$ to $\mathcal{P}(\mathbb{T})$ where $\mathcal{P}(\mathbb{T})$ is the space of probability measures on the space $\mathbb{T}$ of (equivalence classes of) compact $\mathbb{R}$-trees equipped with the Gromov-Hausdorff topology. We can also interpret ϕ_β as a map from the space $\mathbb{T}_{\mathrm{w}}$ of weighted $\mathbb{R}$-trees (that is $\mathbb{R}$-trees equipped with a probability measure) to $\mathbb{T}_{\mathrm{w}}$, equipped with the Gromov-Hausdorff-Prokhorov topology. See Section 2.3.1 for a formal introduction to $\mathbb{R}$-trees and CRTs, and the discussion leading up to Lemma 2.31, which establishes the measurability of ϕ_β on a suitable space of strings of beads.

We show that in our setting the conditions for the existence of a fixpoint distribution developed in [13] are satisfied. Those are mainly based on a contraction property of the Wasserstein metric on a suitable subspace $\mathcal{P} \subset \mathcal{P}(\mathbb{T})$. This establishes the existence of a large family of new CRTs, also including all self-similar CRTs constructed differently by Haas, Miermont and Stephenson [40, 66], which allows us to prove the convergence to such a CRT of various tree growth procedures based on an i.i.d. family of the given random string of beads. The following is a recursive construction.

Theorem 2.1 (Recursive construction of CRTs)**.** *Let $p \geq 1$, $\beta > 1/p$, and let $(\check{\mathcal{T}}_0, \check{\mu}_0)$ be an isometric copy of a string of beads $\xi = ([0,K], \sum_{i\geq 1} P_i \delta_{x_i})$ with some distribution ν such that $K = \sup_{i\geq 1}\{x_i : P_i > 0\}$ and $\mathbb{E}[K^p] < \infty$. For $n \geq 0$, to obtain $(\check{\mathcal{T}}_{n+1}, \check{\mu}_{n+1})$ conditionally given $(\check{\mathcal{T}}_n, \check{\mu}_n)$, attach to each atom $x \in \check{\mathcal{T}}_n$ of $\check{\mu}_n$ an independent isometric copy of ξ with metric rescaled by $\check{\mu}_n(x)^\beta$, and mass measure rescaled by $\check{\mu}_n(x)$. Then there exists a compact CRT $(\check{\mathcal{T}}, \check{\mu})$ such that*

$$\lim_{n\to\infty} \left(\check{\mathcal{T}}_n, \check{\mu}_n\right) = \left(\check{\mathcal{T}}, \check{\mu}\right) \quad \textit{a.s. in the Gromov-Hausdorff-Prokhorov topology.}$$

The attachment procedure for the above construction (and all following algorithms) will be defined precisely in Section 2.4.2.

Corollary 2.2 (Bead splitting processes)**.** *Let $p \geq 1$, $\beta > 1/p$, and let $(\check{\mathcal{T}}_0, \check{\mu}_0)$ be an isometric copy of a string of beads $\xi = ([0,K], \sum_{i\geq 1} P_i \delta_{x_i})$ with some distribution ν such that $K = \sup_{i\geq 1}\{x_i : P_i > 0\}$ and $\mathbb{E}[K^p] < \infty$. For $k \geq 0$, to obtain $(\mathcal{T}_{k+1}, \mu_{k+1})$ conditionally given $(\mathcal{T}_k, \mu_k)$, pick an atom $J_k \in \mathcal{T}_k$ from μ_k and attach at J_k an independent isometric copy of ξ with metric rescaled by $\mu_k(x)^\beta$, and mass measure rescaled by $\mu_k(x)$. Then there exists a compact CRT $(\mathcal{T}, \mu)$ such that*

$$\lim_{k\to\infty} (\mathcal{T}_k, \mu_k) = (\mathcal{T}, \mu) \quad \textit{a.s. in the Gromov-Hausdorff-Prohorov topology.}$$

We will prove this corollary by embedding $(\mathcal{T}_k, \mu_k)$, $k \geq 0$, into $(\check{\mathcal{T}}, \check{\mu})$, and then show $(\mathcal{T}, \mu) = (\check{\mathcal{T}}, \check{\mu})$. These bead splitting processes were introduced in [61] Definition 18. Corollary 2.2 was proved in [61] Theorem 21 in the special case when the string of beads is regenerative in the sense of Gnedin and Pitman [38]. In this special case, it was shown that $(\mathcal{T}, \mu)$ is a self-similar CRT in the sense of Haas and Miermont [40]. Since we show $(\mathcal{T}, \mu) = (\check{\mathcal{T}}, \check{\mu})$, Theorem 2.1 gives an alternative construction of the binary self-similar CRTs of [40].

We will apply Theorem 2.1 to "embed" Duquesne and Le Gall's stable tree [28] into a compact binary CRT in a way that solves an open problem posed by Goldschmidt and Haas [39].

Marchal [50] introduced a Markovian random tree growth model related to the stable tree. Specifically, he built a sequence of discrete trees $(T_n, n \geq 0)$, which we view as rooted $\mathbb{R}$-trees with unit edge lengths where the edges of an $\mathbb{R}$-tree $\mathcal{T}$ are the connected components of $\mathcal{T} \setminus \mathrm{Br}(\mathcal{T})$ and where $\mathrm{Br}(\mathcal{T})$ is the set of branch points of $\mathcal{T}$. We equip T_n with the graph distance, i.e. the distance between two vertices $x, y \in T_n$ is the number of edges between x and y in T_n, $n \geq 0$.

Algorithm 2.3 (Marchal)**.** Let $\beta \in (0, 1/2]$. We grow discrete trees T_n, $n \geq 0$, as follows.

0. Let T_0 consist of two vertices connected by an edge.

Given $T_j, 0 \leq j \leq n$,

1. distribute a total weight of $n + \beta$ by assigning $(d-3)(1-\beta) + 1 - 2\beta$ to each vertex of T_n of degree $d \geq 3$ and β to each edge of T_n; select a vertex or an edge in T_n at random according to these weights;
2. if an edge is selected, insert a new vertex, i.e. replace the selected edge by two edges connecting the new vertex to the vertices of the selected edge; proceed with the new vertex as the selected vertex;
3. in all cases, add a new edge from the selected vertex to a new degree-1 vertex to form T_{n+1}.

It was shown by Curien and Haas [23] that the sequence of trees $(T_n, n \geq 0)$ has a CRT $\mathcal{T}$ as its a.s. scaling limit, i.e.

$$\lim_{n \to \infty} n^{-\beta} T_n = \mathcal{T} \quad \text{a.s. in the Gromov-Hausdorff topology,}$$

where, for any $\alpha > 0$, αT_n denotes the metric space obtained when all distances are multiplied by α. The tree $\mathcal{T}$ was identified as a stable tree of index $1/(1-\beta)$, introduced earlier by Duquesne and Le Gall [28] in the context of continuous-state branching processes, and studied by Miermont and others [40, 42, 43, 52, 53] in the context of fragmentations. Note that the stable trees in this chapter are characterised by a parameter $\beta \in (0, 1/2]$ rather than an index $1/(1-\beta) \in (1, 2]$. A stable tree $\mathcal{T}$ is naturally equipped with a mass measure μ with $\mu(\mathcal{T}) = 1$, supported on the set of its leaves. Given $(\mathcal{T}, \mu)$, sampling leaves $(\Sigma_k, k \geq 0)$ independently from μ yields a sequence of random compact $\mathbb{R}$-trees

$$(\mathcal{T}_k, k \geq 0) \tag{2.2}$$

where $\mathcal{T}_k$ is the reduced subtree of $\mathcal{T}$ spanned by the root and the first $k+1$ leaves $\Sigma_0, \ldots, \Sigma_k$. In the case when $\beta = 1/2$, the stable tree is Aldous' Brownian CRT [7, 8, 9]. Note that, in Algorithm 2.3 for $\beta = 1/2$, each branch point has weight 0 after it is created, and hence no vertices of degree $d \geq 4$ will arise in the sequential construction.

Aldous [7] introduced a line-breaking construction of the trees $(\mathcal{T}_k, k \geq 0)$ in the Brownian case ($\beta = 1/2$), where the branches $\mathcal{T}_0$ and $\mathcal{T}_{k+1} \setminus \mathcal{T}_k$, $k \geq 0$, can be thought of as being broken off a half-infinite line at cut points according to an inhomogeneous Poisson point process. Goldschmidt and Haas [39] generalised Aldous' line-breaking construction of the Brownian CRT to all stable CRTs of parameter $\beta \in (0, 1/2]$, building the sequence of trees $(\mathcal{T}_k, k \geq 0)$ from (2.2) and assigning weights

$$W_k^{(i)}, \quad i \in [b_k], \quad k \geq 1, \tag{2.3}$$

to each branch point $v_i \in \mathrm{Br}(\mathcal{T}_k)$, where $b_k := \#\mathrm{Br}(\mathcal{T}_k)$ and $\#S$ denotes the number of elements of a set S, $[k] := \{1, \ldots, k\}$ for $k \geq 1$, and $(v_i, i \geq 1)$ denotes the sequence of branch points in order of appearance in the construction of $(\mathcal{T}_k, k \geq 0)$. We set $W_k^{(i)} = 0$ for all $i \geq 1$ with $v_i \notin \mathrm{Br}(\mathcal{T}_k)$. The branch point weights (2.3) are derived from the lengths of the edges attached to the respective vertex. For $k \geq 0$, the sum of the branch point weights $(W_k^{(i)}, i \geq 1)$ and the total length $L_k = \mathrm{Leb}(\mathcal{T}_k)$ of $\mathcal{T}_k$ is given by S_k where $(S_k, k \geq 0)$ is the Mittag-Leffler Markov chain [39, 42, 44] with parameters

(β, β), that is, the Markov chain starting from $S_0 \sim \mathrm{ML}(\beta, \beta)$ with transition density

$$f_{S_{k+1}|S_k=z}(y) = f_\beta(z, y) = \frac{1-\beta}{\Gamma(1/\beta)}(y-z)^{1/\beta-2}\frac{y g_\beta(y)}{g_\beta(z)}, \quad y > z > 0,$$

for $k \geq 0$, where $\mathrm{ML}(\alpha, \theta)$ denotes the Mittag-Leffler distribution with parameters $0 < \alpha < 1$ and $\theta > -\alpha$, see Section 2.2.1, and $g_\beta(\cdot)$ is the density of $\mathrm{ML}(\beta, 0)$. Then

$$S_k = \sum_{i=1}^{b_k} W_k^{(i)} + L_k \sim \mathrm{ML}(\beta, \beta + k).$$

In the following, $\mathrm{Leb}(\cdot)$ denotes the length measure.

Algorithm 2.4 (Line-breaking construction of the stable tree, Goldschmidt-Haas [39])**.** Let $\beta \in (0, 1/2]$. We grow $\mathbb{R}$-trees $\mathcal{T}_k$ with weights $W_k^{(i)}$ in branch points $v_i \in \mathrm{Br}(\mathcal{T}_k)$, $i \in [b_k]$, where $b_k = \#\mathrm{Br}(\mathcal{T}_k)$, $k \geq 0$, as follows.

0. Let $(\mathcal{T}_0, \rho)$ be isometric to $([0, S_0], 0)$, where $S_0 \sim \mathrm{ML}(\beta, \beta)$; let $b_0 = 0$ and $W_0^{(i)} = 0$, $i \geq 1$.

Given $(\mathcal{T}_j, (v_i, i \in [b_j]), (W_j^{(i)}, i \in [b_j]))$, $0 \leq j \leq k$ and $S_k = L_k + W_k^{(1)} + \cdots + W_k^{(b_k)}$, where $L_k = \mathrm{Leb}(\mathcal{T}_k)$,

1. select $I_k = i$ for each branch point $v_i \in \mathrm{Br}(\mathcal{T}_k)$ with probability proportional to $W_k^{(i)}$, $i \in [b_k]$; or select an edge $E_k \subset \mathcal{T}_k$ with probability proportional to its length and let $b_{k+1} = b_k + 1$, $I_k = b_{k+1}$;

2. if an edge E_k is selected, sample $v_{b_{k+1}}$ from the normalised length measure on E_k;

3. sample S_{k+1} with density $f_\beta(S_k, \cdot)$ and an independent $B_k \sim \mathrm{Beta}(1, 1/\beta - 2)$; attach to $\mathcal{T}_k$ at $J_k := v_{I_k}$ a new branch of length $(S_{k+1} - S_k)B_k$ to form $\mathcal{T}_{k+1}$; increase the weight of $J_k = v_{I_k}$ to

$$W_{k+1}^{(I_k)} = W_k^{(I_k)} + (S_{k+1} - S_k)(1 - B_k), \qquad \text{and set} \quad W_{k+1}^{(i)} = W_k^{(i)}, \quad i \neq I_k.$$

Note that, when $\beta = 1/2$, $W_k^{(i)} = 0$ for all $i \geq 1$ and $k \geq 0$, i.e. $L_k = S_k$ for all $k \geq 0$. It was shown in [39] that the sequence of trees $(\mathcal{T}_k, k \geq 0)$ of Algorithm 2.4 has the same distribution as the sequence in (2.2), i.e. the metric space completion

$$\mathcal{T} := \overline{\bigcup_{k\geq 0} \mathcal{T}_k}$$

is the stable tree of parameter β; see also [39] for another line-breaking construction building the sequence of trees $(\mathcal{T}_k, k \geq 0)$, where branch point selection is based on vertex degrees and not weights.

Goldschmidt and Haas [39] asked the question if (and if yes how) the branch point weights in Algorithm 2.4 can naturally be expressed in terms of lengths. We answer this question by using the branch point weights to build rescaled Ford trees whose lengths correspond to these weights. Ford trees arise in the scaling limit of Ford's alpha model studied in [35, 42] and also in the context of the alpha-gamma-model [22] for $\gamma = \alpha$, which is also related to the stable tree in the case when $\gamma = 1 - \alpha$. Ford trees are examples of binary self-similar CRTs; they have also been constructed via line-breaking [60]. We call an edge of an $\mathbb{R}$-tree $\mathcal{T}$ internal if it connects two branch points (or the root and a branch point) of $\mathcal{T}$, and external otherwise.

The following algorithm slightly restates results obtained in [42, 60].

Algorithm 2.5 (Line-breaking construction of Ford trees, Haas-Miermont-Pitman-Winkel [42, 60])**.** Let $\beta' \in (0, 1)$. We grow $\mathbb{R}$-trees $\mathcal{F}_m$, $m \geq 1$, as follows.

0. Let $(\mathcal{F}_1, \rho)$ be isometric to $([0, S_1'], 0)$, where $S_1' \sim \mathrm{ML}(\beta', 1 - \beta')$.

Given $\mathcal{F}_j$, $1 \leq j \leq m$, let $S_m' = \mathrm{Leb}(\mathcal{F}_m)$ denote the length of $\mathcal{F}_m$,

1. select an edge $E_m \subset \mathcal{F}_m$ with probability proportional to its length;
2. if E_m is external, sample $D_m \sim \mathrm{Beta}(1, 1/\beta' - 1)$ and place $J_m \in E_m$ to split E_m into length proportions D_m and $1 - D_m$, the former closer to the root ρ; otherwise, sample J_m from the normalised length measure on E_m;
3. sample S_{m+1}' with density $f_{\beta'}(S_m', \cdot)$; attach to $\mathcal{F}_m$ at J_m an edge of length $S_{m+1}' - S_m'$ to form $\mathcal{F}_{m+1}$.

The sequence of trees $(\mathcal{F}_m, m \geq 1)$ converges a.s. with respect to the Gromov-Hausdorff topology to a continuum random tree $\mathcal{F}$ as $k \to \infty$, a so-called *Ford CRT* of index $\beta' \in (0,1)$, see [42, 60]. We refer to the trees $\mathcal{F}_m, m \geq 1$, as *Ford trees.*

We combine the line-breaking constructions given in Algorithms 2.4 and 2.5 in the framework of ∞-marked $\mathbb{R}$-trees, which we introduce in Section 2.3.2 as a natural extension of Miermont's notion of k-marked trees [54]. An ∞-marked $\mathbb{R}$-tree $(\mathcal{T}, (\mathcal{R}^{(i)}, i \geq 1))$ is an $\mathbb{R}$-tree $(\mathcal{T}, d, \rho)$ with compact non-empty connected subsets $\mathcal{R}^{(i)} \subset \mathcal{T}$, $i \geq 1$. We will refer to this setting as a *two-colour* framework, meaning that the marked sets $\bigcup_{i\geq 1} \mathcal{R}^{(i)}$ and the unmarked remainder $\mathcal{T} \setminus \bigcup_{i\geq 1} \mathcal{R}^{(i)}$ are associated with two different colours. The marked components in the line-breaking construction below correspond to rescaled Ford trees with lengths equal to the branch point weights in Algorithm 2.4 and the unmarked remainder gives rise to a stable tree. Selection of a branch point in Algorithm 2.4 corresponds to an insertion into the respective marked component in the enhanced line-breaking construction of Algorithm 2.6.

Algorithm 2.6 (Two-colour line-breaking construction)**.** Let $\beta \in (0, 1/2]$. We grow ∞-marked $\mathbb{R}$-trees $(\mathcal{T}_k^*, (\mathcal{R}_k^{(i)}, i \geq 1))$, $k \geq 0$, as follows.

0. Let $(\mathcal{T}_0^*, \rho)$ be isometric to $([0, S_0], 0)$, where $S_0 \sim \mathrm{ML}(\beta, \beta)$. Let $r_0 = 0$ and $\mathcal{R}_0^{(i)} = \{\rho\}$, $i \geq 1$.

Given $(\mathcal{T}_j^*, (\mathcal{R}_j^{(i)}, i \geq 1))$, $0 \leq j \leq k$, let $S_k = \mathrm{Leb}(\mathcal{T}_k^*)$ be the length of $\mathcal{T}_k^*$ and $r_k = \#\{i \geq 1 \colon \mathcal{R}_k^{(i)} \neq \{\rho\}\}$;

1. select an edge $E_k^* \subset \mathcal{T}_k^*$ with probability proportional to its length; if $E_k^* \subset \mathcal{R}_k^{(i)}$ for some $i \in [r_k]$, let $I_k = i$; otherwise, i.e. if $E_k^* \subset \mathcal{T}_k^* \setminus \bigcup_{i \in [r_k]} \mathcal{R}_k^{(i)}$, let $r_{k+1} = r_k + 1$, $I_k = r_{k+1}$;

2. if E_k^* is an external edge of $\mathcal{R}_k^{(i)}$, sample $D_k \sim \mathrm{Beta}(1, 1/\beta - 2)$ and place J_k^* to split E_k^* into length proportions D_k and $1 - D_k$, the former closer to the root ρ; otherwise, i.e. if $E_k^* \subset \mathcal{T}_k^* \setminus \bigcup_{i \in [r_k]} \mathcal{R}_k^{(i)}$ or if E_k^* is an internal edge of $\mathcal{R}_k^{(i)}$, sample J_k^* from the normalised length measure on E_k^*;

3. sample S_{k+1} with density $f(S_k, \cdot)$ and an independent $B_k \sim \mathrm{Beta}(1, 1/\beta - 2)$; attach to $\mathcal{T}_k^*$ at J_k^* a new branch of length $S_{k+1} - S_k$ to form $\mathcal{T}_{k+1}^*$, and add to

$\mathcal{R}_k^{(I_k)}$ the part of length $(S_{k+1} - S_k)(1 - B_k)$ closest to the root to form $\mathcal{R}_{k+1}^{(I_k)}$; set $\mathcal{R}_{k+1}^{(j)} = \mathcal{R}_k^{(j)}$, $j \neq I_k$.

Indeed, we obtain the correspondence of the branch point weights in Algorithm 2.4 and the lengths of the marked subtrees in Algorithm 2.6, cf. Figure 2.1.

Theorem 2.7 (Weight-length representation)**.** *Let $(\mathcal{T}_k^*, (\mathcal{R}_k^{(i)}, i \geq 1), k \geq 0)$ be the sequence of ∞-marked $\mathbb{R}$-trees constructed in Algorithm 2.6, and let $\widetilde{W}_k^{(i)} = \mathrm{Leb}(\mathcal{R}_k^{(i)})$ denote the length of $\mathcal{R}_k^{(i)}$, $i \geq 1$, respectively. For $k \geq 1$, contract each component $\mathcal{R}_k^{(i)}$ to a single branch point $\tilde{v}_i$ by using an equivalence relation, and denote the resulting tree by $\widetilde{\mathcal{T}}_k$. Then*

$$\left(\widetilde{\mathcal{T}}_k, \left(\widetilde{W}_k^{(i)}, i \geq 1\right), k \geq 0\right) \stackrel{d}{=} \left(\mathcal{T}_k, \left(W_k^{(i)}, i \geq 1\right), k \geq 0\right), \tag{2.4}$$

The sequences $(\mathcal{T}_k, k \geq 1)$ and $(\mathcal{F}_m, m \geq 1)$ converge to CRTs $\mathcal{T}$ and $\mathcal{F}$, respectively. Similarly, we obtain convergence of the sequence $(\mathcal{T}_k^*, (\mathcal{R}_k^{(i)}, i \geq 1), k \geq 0)$ with respect to d_{GH}^{∞}, an analogue of the Gromov-Hausdorff distance on the space of ∞-marked $\mathbb{R}$-trees.

Theorem 2.8 (Convergence of two-colour trees)**.** *Let $(\mathcal{T}_k^*, (\mathcal{R}_k^{(i)}, i \geq 1), k \geq 0)$ be as above. Then*

$$\lim_{k \to \infty} \left(\mathcal{T}_k^*, \left(\mathcal{R}_k^{(i)}, i \geq 1\right)\right) = \left(\mathcal{T}^*, \left(\mathcal{R}^{(i)}, i \geq 1\right)\right) \quad a.s. \tag{2.5}$$

with respect to d_{GH}^{∞} where $(\mathcal{T}^*, (\mathcal{R}^{(i)}, i \geq 1))$ is a compact ∞-marked $\mathbb{R}$-tree.

Furthermore, the tree $\widetilde{\mathcal{T}}$, obtained from $\mathcal{T}^*$ by contracting each component $\mathcal{R}^{(i)}$ to a single branch point $\tilde{v}_i$, is a stable tree of parameter β. The trees $C_i^{-1}\mathcal{R}^{(i)}$, $i \geq 1$, are i.i.d. copies of a Ford CRT $\mathcal{F}$ of index $\beta' = \beta/(1-\beta)$.

The proof of Theorem 2.8, in particular the compactness of $\mathcal{T}^*$, are based on an embedding of $(\mathcal{T}_k^*, k \geq 0)$ into a compact CRT whose existence follows from Theorem 2.1 applied for a specific distribution ν on a certain space of strings of beads. The distribution ν is related to the distribution of an (α, θ)-*string of beads*, arising in the framework of ordered (α, θ)-Chinese restaurant processes for $\alpha > 0$ and $\theta > -\alpha$ as introduced in [60]. An (α, θ)-string of beads is an interval of length $K \sim \mathrm{ML}(\alpha, \theta)$ equipped with a discrete probability measure whose atom sizes are $\mathrm{PD}(\alpha, \theta)$, arranged in a random order that

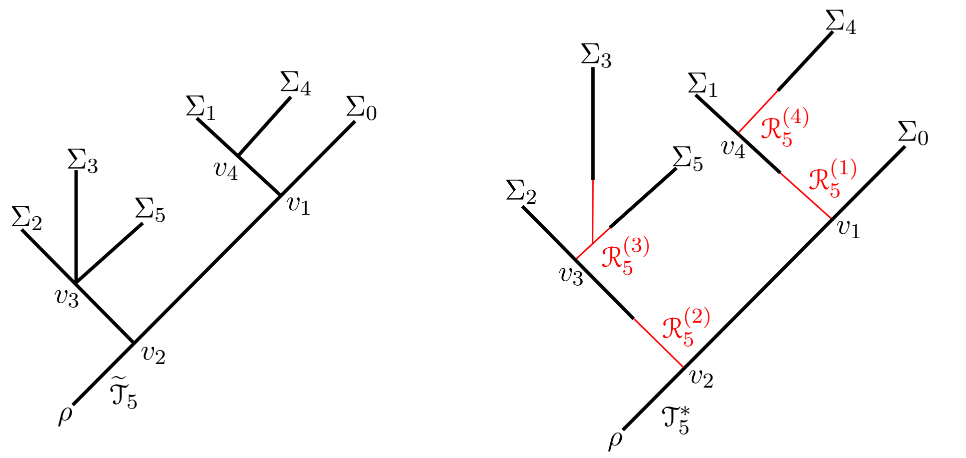

Fig. 2.1 Example of $\widetilde{\mathcal{T}}_5$ with four branch points $v_1, \ldots, v_4$. Branch point weights $W_5^{(1)}, \ldots, W_5^{(4)}$ are represented as lengths of marked subtrees $\mathcal{R}_5^{(1)}, \ldots, \mathcal{R}_5^{(4)}$.

yields a regenerative property (see Section 2.2.2 for a precise definition). Specifically, a string of beads with distribution ν is obtained by concatenating independent $(\beta, 1-2\beta)$- and (β,β)-strings of beads, scaled by the parts of an independent Beta$(1-2\beta,\beta)$ split.

It is crucial for our argument to equip each reduced tree with a mass measure which effectively captures projected subtree masses. This naturally leads to a line-breaking construction of the stable tree where attachment point selection is based on masses rather than lengths.

Algorithm 2.9 (Construction of the stable tree based on masses)**.** Let $\beta \in (0, 1/2]$. We grow weighted $\mathbb{R}$-trees $(\mathcal{T}_k, \mu_k)$, $k \geq 0$, as follows.

0. Let $(\mathcal{T}_0, \mu_0)$ be isometric to a (β,β)-string of beads ξ_0.

Given $(\mathcal{T}_j, \mu_j)$ with $\mu_j = \sum_{x \in \mathcal{T}_j} \mu_j(x)\delta_x$, $0 \leq j \leq k$,

1.-2. sample J_k from μ_k;

3. given $\deg(J_k, \mathcal{T}_k) = d \geq 2$, let $Q_k \sim$ Beta$(\beta, (d-2)(1-\beta)+1-2\beta)$, and let ξ_k be an independent (β,β)-string of beads; to form $(\mathcal{T}_{k+1}, \mu_{k+1})$, remove $Q_k\mu_k(J_k)\delta_{J_k}$ from μ_k and attach to $\mathcal{T}_k$ at J_k an isometric copy of ξ_k with measure rescaled by $Q_k\mu_k(J_k)$ and metric rescaled by $(Q_k\mu_k(J_k))^\beta$;

Theorem 2.10. *In Algorithm 2.9,* $(\mathcal{T}_k, k \geq 0)$ *has the same distribution as the sequence of trees in* (2.2) *(i.e. as in Algorithm 2.4). In particular,* $\lim_{k\to\infty} \mathcal{T}_k = \mathcal{T}$ *a.s. in the*

Gromov-Hausdorff topology for a stable tree $\mathcal{T}$*. Furthermore,* $\lim_{k\to\infty}(\mathcal{T}_k, \mu_k) = (\mathcal{T}, \mu)$ *a.s. in the Gromov-Hausdorff-Prokhorov topology.*

The proof of Theorem 2.10 is based on the following well-known property of the stable tree, phrased in different terminology in [43] Corollary 10(3), and [60] (see the discussion after Corollary 8), where the link between (β, β)-strings of beads and a Bessel bridge of dimension 2β was established.

Proposition 2.11. *Let* $(\mathcal{T}, \mu)$ *be a stable tree of parameter* $\beta \in (0, 1/2]$*, and let* $\Sigma \sim \mu$*. Denote the trivial tree spanned by the root* ρ *and* Σ *by* $\mathcal{T}_0$*, and equip* $\mathcal{T}_0$ *with the mass measure* μ_0*, capturing the masses of the connected components of* $\mathcal{T} \setminus \mathcal{T}_0$ *projected onto* $\mathcal{T}_0$*. Then* $(\mathcal{T}_0, \mu_0)$ *is a* (β, β)*-string of beads.*

This chapter is organised as follows. In Section 2.2 we present some preliminary results on Dirichlet and Mittag-Leffler distributions, focussing on the connection with (ordered) two-parameter Chinese restaurant processes and the Mittag-Leffler Markov chain. We introduce the framework of ∞-marked $\mathbb{R}$-trees in Section 2.3, and develop the recursive construction of CRTs via recursive tree frameworks in Section 2.4, which includes the proof of Theorem 2.1. The two-colour line-breaking construction which allows us to obtain a binary embedding of the stable line-breaking construction is discussed in Section 2.5 (including the proofs of Theorems 2.7, 2.8 and 2.10). In Section 2.6 we present a discrete two-colour tree growth process which converges to the trees in the two-colour line-breaking construction. An appendix, Section 2.7, contains some technical proofs postponed from earlier parts of this chapter.

2.2 Preliminaries

2.2.1 Dirichlet and Mittag-Leffler distributions

In this section, we present the distributional relationships that are key for our constructions. A random variable L follows a (generalised) Mittag-Leffler distribution with parameters (α, θ) for $\alpha > 0$ and $\theta > -\alpha$ if its p-th moment is given by

$$\mathbb{E}[L^p] = \frac{\Gamma(\theta+1)\Gamma(\theta/\alpha+1+p)}{\Gamma(\theta/\alpha+1)\Gamma(\theta+p\alpha+1)}, \quad p \geq 1, \tag{2.6}$$

for short $L \sim \mathrm{ML}(\alpha, \theta)$. The moments (2.6) uniquely characterise $\mathrm{ML}(\alpha, \theta)$, cf. [58]. The following proposition is related to the scaling factors of the Ford CRTs built in our construction.

Proposition 2.12 (Relationship between Mittag-Leffler variables)**.** *Let $\beta \in (0, 1/2]$, and consider independent random variables $C \sim \mathrm{ML}(1-\beta, k(1-\beta)-\beta)$ and $S \sim \mathrm{ML}(\beta/(1-\beta), k-\beta/(1-\beta))$ for some $k \geq 1$. Then*

$$C^{\beta/(1-\beta)} S \stackrel{d}{=} K \tag{2.7}$$

where $K \sim \mathrm{ML}(\beta, k(1-\beta)-\beta)$.

Proof. Recall that the Mittag-Leffler distribution is uniquely characterised by its moments (2.6). It is easily checked that, for each $p \geq 1$, $\mathbb{E}[(C^{\beta/(1-\beta)}S)^p] = \mathbb{E}[K^p]$. □

The Mittag-Leffler distribution naturally appears when we study lengths in the trees considered in this chapter. To analyse mass and length splits across the branches of these trees we have to consider Dirichlet distributions. We will be able to relate mass and length splits on the edges using the following result.

Proposition 2.13 ([39] Proposition 4.2)**.** *Let $\beta \in (0,1)$ and $\theta > 0$. For $n \geq 2$, let $\theta_1, \ldots, \theta_n > 0$ and let $\theta := \sum_{i=1}^n \theta_i$. Consider $S \sim \mathrm{ML}(\beta, \theta)$ and an independent vector $(Y_1, \ldots, Y_n) \sim \mathrm{Dirichlet}(\theta_1/\beta, \ldots, \theta_n/\beta)$. Then,*

$$S(Y_1, \ldots, Y_n) \stackrel{d}{=} \left(X_1^\beta S^{(1)}, \ldots, X_n^\beta S^{(n)}\right) \tag{2.8}$$

for independent $(X_1, \ldots, X_n) \sim \mathrm{Dirichlet}(\theta_1, \ldots, \theta_n)$ and $S^{(i)} \sim \mathrm{ML}(\beta, \theta_i)$, $i \in [n]$.

We will also need some standard properties of the Dirichlet distribution.

Proposition 2.14. *Let $X := (X_1, \ldots, X_n) \sim \mathrm{Dirichlet}(\theta_1, \ldots, \theta_n)$ for some $\theta_1, \ldots, \theta_n > 0$, $n \in \mathbb{N}$.*

(i) *Symmetry. For any permutation $\sigma : [n] \to [n]$,*

$$\left(X_{\sigma(1)}, \ldots, X_{\sigma(n)}\right) \sim \mathrm{Dirichlet}\left(\theta_{\sigma(1)}, \ldots, \theta_{\sigma(n)}\right).$$

(ii) *Aggregation and deletion. Let* $X' := (X_1 + \cdots + X_m, X_{m+1}, \ldots, X_n)$ *for some* $m \in [n]$. *Then*

$$X' \sim \mathrm{Dirichlet}(\theta_1 + \cdots + \theta_m, \theta_{m+1}, \ldots, \theta_n)$$

is independent of

$$(X_1/(X_1 + \cdots + X_m), \ldots, X_m/(X_1 + \cdots + X_m)) \sim \mathrm{Dirichlet}(\theta_1, \ldots, \theta_m).$$

(iii) *Decimation. Let* $i \in [n]$, $m \in \mathbb{N}$, *and let* $\theta_{i,1}, \ldots, \theta_{i,m} > 0$ *be such that* $\theta_{i,1} + \cdots + \theta_{i,m} = \theta_i$. *Consider an independent random vector*

$$(P_1, \ldots, P_m) \sim \mathrm{Dirichlet}(\theta_{i,1}, \ldots, \theta_{i,m}),$$

and define $X'' := (X_1, \ldots, X_{i-1}, P_1 X_i, \ldots, P_m X_i, X_{i+1}, \ldots, X_n)$. *Then*

$$X'' \sim \mathrm{Dirichlet}(\theta_1, \ldots, \theta_{i-1}, \theta_{i,1}, \ldots, \theta_{i,m}, \theta_{i+1}, \ldots, \theta_n).$$

(iv) *Size-bias. Let* $I \in [n]$ *be a random index such that* $\mathbb{P}(I = i | X) = X_i$ *a.s. for* $i \in [n]$. *Then, conditionally given* $I = i$ *for some* $i \in [n]$,

$$X \sim \mathrm{Dirichlet}(\theta_1, \ldots, \theta_{i-1}, \theta_i + 1, \theta_{i+1}, \ldots, \theta_n).$$

Furthermore, we have $\mathbb{P}(I = i) = \theta_i/(\theta_1 + \cdots + \theta_n)$.

Proof. We refer to [67], Propositions 13 and 14, Remark 15, and the Gamma variable representation of the Dirichlet distribution. □

Proposition 2.15. *Let* $S \sim \mathrm{ML}(\beta, \theta + k + 1)$ *and* $A \sim \mathrm{Beta}((\theta + k)/\beta + 1, 1/\beta - 1)$ *for some* $\beta > 0$, $\theta > -\beta$, $k \in \mathbb{N}$, *be independent. Then* $SA \sim \mathrm{ML}(\beta, \theta + k)$.

Proof. We consider the moments of SA. By independence, for $p \geq 1$, we have

$$\mathbb{E}[(SA)^p] = \mathbb{E}[S^p]\mathbb{E}[A^p] = \frac{\Gamma(\theta + k + 2)\Gamma((\theta + k + 1)/\beta + 1 + p)}{\Gamma((\theta + k + 1)/\beta + 1)\Gamma(\theta + k + p\beta + 2)}$$

$$\cdot \frac{\Gamma((\theta+k)/\beta+1+p)\Gamma((\theta+k+1)/\beta)}{\Gamma((\theta+k)/\beta+1)\Gamma((\theta+k+1)/\beta+p)}.$$

We can apply $\Gamma(t+1) = t\Gamma(t)$ for $t \geq 0$ to show that

$$\mathbb{E}[(SA)^p] = \frac{\Gamma(\theta+k+1)\Gamma((\theta+k)/\beta+1+p)}{\Gamma((\theta+k)/\beta+1)\Gamma(\theta+k+p\beta+1)},$$

which is the p-th moment of $\mathrm{ML}(\beta,\theta)$. □

2.2.2 Chinese restaurant processes and strings of beads

We consider (α,θ)-strings of beads for $\alpha \in (0,1), \theta > 0$, arising in the scaling limit of ordered (α,θ)-Chinese restaurant processes, cf. [58, 60].

Consider customers labelled by $[n] = \{1,\ldots,n\}$ sitting at a random number of tables in a Chinese restaurant as follows. Let customer 1 sit at the first table. At step $n+1$, conditionally given that we have k tables with $n_1,\ldots,n_k$ customers at each table, the next customer labelled by $n+1$

- sits at the i-th occupied table with probability $(n_i-\alpha)/(n+\theta)$, $i \in [k]$;
- opens a new table with probability $(k\alpha+\theta)/(n+\theta)$.

The induced sequence of random partitions $(\Pi_n, n \geq 1)$ of $[n]$ is called an (α,θ)-*Chinese restaurant process* (CRP). For each partition π of $[n]$ into k blocks of sizes $n_1,\ldots,n_k$, we have

$$\mathbb{P}(\Pi_n = \pi) = \frac{\prod_{i=1}^{k-1}(\theta+\alpha i)}{[1+\theta]_{n-1}} \prod_{i=1}^{k} [1-\alpha]_{n_i-1}, \tag{2.9}$$

where, for $x \in \mathbb{R}$ and $m \in \mathbb{N}$, we define the m-rising factorial by $[x]_m := x(x+1)(x+2)\cdots(x+m)$. Note that the distribution of Π_n, given by (2.9), only depends on block sizes, i.e. the partitions Π_n are exchangeable. They are also consistent as n varies, i.e. they induce a random partition Π_∞ of $\mathbb{N}$, which yields Π_n when restricted to $[n]$.

Pitman and Winkel [60] enhanced the (α,θ)-CRP with a *table order*, i.e. a random total order $<$ on the tables, independently of the actual seating process. The tables are ordered from left to right as follows. The second table is put to the right of the first table with probability $\theta/(\alpha+\theta)$, and to the left with probability $\alpha/(\alpha+\theta)$. For $k \geq 1$, conditionally given any of the $k!$ possible orderings of the first k tables, the $(k+1)$-st table is put

- to the left of the first table, or between any two tables with probability $\alpha/(k\alpha+\theta)$ each;
- to the right of the last table with probability $\theta/(k\alpha+\theta)$.

This induces the *ordered* (α,θ)-CRP $(\widetilde{\Pi}_n, n \geq 1)$, that is, the (α,θ)-CRP $(\Pi_n, n \geq 1)$ with tables ordered according to $<$. Note that $(\Pi_n, n \geq 1)$ can be recovered from $(\widetilde{\Pi}_n, n \geq 1)$ by ordering the blocks in *order of appearance*, i.e. in order of least labelled elements. For $n \in \mathbb{N}$, we write

$$\Pi_n := (\Pi_{n,1}, \ldots, \Pi_{n,K_n}), \qquad \widetilde{\Pi}_n := \left(\widetilde{\Pi}_{n,1}, \ldots, \widetilde{\Pi}_{n,K_n}\right)$$

for the blocks of the two partitions Π_n, $\widetilde{\Pi}_n$ of $[n]$, where K_n denotes the number of tables at step n. The sizes of these blocks at step n form random *compositions* of n, $n \geq 1$, that is, sequences of positive integers $(n_1, \ldots, n_k)$ with sum $n = \sum_{j=1}^{k} n_j$. The composition related to $\widetilde{\Pi}_n$, $n \geq 1$, can be shown to be regenerative in the sense of Gnedin and Pitman [38], where a sequence of random compositions $\mathcal{C}_n$ of n, $n \geq 1$, is called *regenerative* if for all $1 \leq m < n$, conditionally given that the first part of $\mathcal{C}_n$ is m, the remaining composition of $n-m$ is distributed as $\mathcal{C}_{n-m}$.

The number of tables K_n at step n rescaled by n^α converges a.s. as $n \to \infty$, i.e. there is $L_{\alpha,\theta} > 0$ a.s. such that

$$L_{\alpha,\theta} = \lim_{n\to\infty} n^{-\alpha} K_n \quad \text{a.s.}. \tag{2.10}$$

The distribution of $L_{\alpha,\theta}$ can be identified as $\mathrm{ML}(\alpha,\theta)$. Furthermore, we obtain limiting proportions $(P_1, P_2, \ldots)$ of the relative table sizes $n^{-1}\#\Pi_{n,i}$, $i \in [K_n]$, as $n \to \infty$ in the order of appearance, i.e.

$$\lim_{n\to\infty} \left(n^{-1}\#\Pi_{n,1}, \ldots, n^{-1}\#\Pi_{n,K_n}\right) = (P_1, P_2, P_3, \ldots) = \left(V_1, \overline{V}_1 V_2, \overline{V}_1 \overline{V}_2 V_3, \ldots\right) \tag{2.11}$$

a.s. where $(V_i, i \geq 1)$ are independent with $V_i \sim \text{Beta}(1-\alpha, \theta + i\alpha)$, and $\overline{V}_i := 1 - V_i$. If we rank the components of the vector $(P_i, i \geq 1)$ in decreasing order we obtain a Poisson-Dirichlet distribution with parameters (α, θ), i.e. $(P_i^{\downarrow}, i \geq 1) := (P_i, i \geq 1)^{\downarrow} \sim \text{PD}(\alpha, \theta)$. On the other hand, each of the atom masses $P_i, i \geq 1$, is associated with a position on the limiting interval $[0, L_{\alpha,\theta}]$ induced by the table order.

Lemma 2.16 ([60] Proposition 6)**.** *Consider an ordered (α, θ)-CRP*

$$\left(\widetilde{\Pi}_n = \left(\widetilde{\Pi}_{n,1}, \ldots, \widetilde{\Pi}_{n,K_n}\right), n \geq 1\right)$$

for $\alpha \in (0,1)$, $\theta > 0$. Let $N_{n,j} := \sum_{i=1}^{j} \#\widetilde{\Pi}_{n,i}$, $j \in [n]$, be the number of customers at the first j tables from the left. Then,

$$\lim_{n \to \infty} \left\{ n^{-1} N_{n,j}, j \geq 0 \right\} = \mathcal{N}_{\alpha,\theta} := \left\{ 1 - e^{-G_t}, t \geq 0 \right\}^{\text{cl}} \quad a.s. \tag{2.12}$$

with respect to the Hausdorff metric on closed subsets of $[0,1]$, where $^{\text{cl}}$ denotes the closure in $[0,1]$, and $(G_t, t \geq 0)$ is a subordinator with Laplace exponent $\Phi_{\alpha,\theta}(s) = s\Gamma(s+\theta)\Gamma(1-\alpha)/\Gamma(s+\theta+1-\alpha)$. Furthermore, there is a continuous local time process $\mathcal{L} = (\mathcal{L}(u), u \in [0,1])$ for $\mathcal{L}_n(u) := \#\{j \in [K_n] : n^{-1}N_{n,j} \leq u\}$, $u \in [0,1]$, such that

$$\lim_{n \to \infty} \sup_{u \in [0,1]} \left| n^{-\alpha} \mathcal{L}_n(u) - \mathcal{L}(u) \right| = 0 \qquad a.s.$$

where $\mathcal{N}_{\alpha,\theta}$ is the set of points at which $\mathcal{L}$ increases a.s..

We refer to the collection of open intervals in $[0,1] \setminus \mathcal{N}_{\alpha,\theta}$ as the *(α,θ)-regenerative interval partition* associated with the local time process $\mathcal{L}$, where $\mathcal{L}(1) = L_{\alpha,\theta}$ a.s.. Note that the joint law of ranked lengths of components of this interval partition is $\text{PD}(\alpha, \theta)$, and that $(\mathcal{L}(u), 0 \leq u \leq 1)$ can be characterised in terms of $(G_t, t \geq 0)$ via

$$\mathcal{L}\left(1 - e^{-G_t}\right) = \Gamma(1-\alpha) \int_0^t e^{-\alpha G_s} ds, \quad t \geq 0.$$

The inverse local time $\mathcal{L}^{-1}$ defined by

$$\mathcal{L}^{-1} \colon [0, L_{\alpha,\theta}] \to [0,1], \qquad \mathcal{L}^{-1}(x) := \inf\{u \in [0,1] : \mathcal{L}(u) > x\}, \tag{2.13}$$

is increasing and right-continuous. We equip the random interval $[0, L_{\alpha,\theta}]$ with the Stieltjes measure $d\mathcal{L}^{-1}$, and refer to the pair $([0, L_{\alpha,\theta}], d\mathcal{L}^{-1})$ as an (α, θ)-string of beads in the following sense.

Definition 2.17 (String of beads)**.** A *string of beads* (I, λ) is an interval I equipped with a discrete mass measure λ. An (α, θ)*-string of beads* is the weighted random interval $([0, L_{\alpha,\theta}], d\mathcal{L}^{-1})$ associated with an (α, θ)-regenerative interval partition $[0, 1] \setminus \mathcal{N}_{\alpha,\theta}$ for $\alpha \in (0, 1), \theta > 0$.

For any string of beads $\xi = (I, \lambda)$, we will write $x \in \xi$ to mean $x \in I$. We would like to consider random variables taking values in a set of strings of beads. To this end, we equip the space of strings of beads with the Borel σ-algebra associated with the topology of weak convergence of probability measures on $[0, \infty)$, where the length of a string of beads (I, λ) can be obtained as the supremum of the support of the measure λ.

We also refer to measure-preserving isometric copies of the weighted interval $([0, L_{\alpha,\theta}], d\mathcal{L}^{-1})$ as an (α, θ)-string of beads. Since the lengths of the interval components of an (α, θ)-regenerative interval partition $[0, 1] \setminus \mathcal{N}_{\alpha,\theta}$ are the masses of the atoms of the associated (α, θ)-string of beads, we conclude that the joint law of the masses $(P_i^{\downarrow}, i \geq 1)$ of the atoms of an (α, θ)-string of beads ranked in decreasing order is $\mathrm{PD}(\alpha, \theta)$. It is well-known that the length $L_{\alpha,\theta} \sim \mathrm{ML}(\alpha, \theta)$ of an (α, θ)-string of beads can be recovered from the ranked atom masses $(P_i^{\downarrow}, i \geq 1)$ or the vector $(P_i, i \geq 1)$ of the stick-breaking representation (2.11) via

$$L_{\alpha,\theta} = \lim_{i\to\infty} i\Gamma(1 - \alpha)(P_i^{\downarrow})^{\alpha} = \lim_{k\to\infty} \left(1 - \sum_{i=1}^{k} P_i\right)^{\alpha} \alpha^{-\alpha} k^{1-\alpha}, \tag{2.14}$$

which is the so-called α*-diversity* of $(P_i^{\downarrow}, i \geq 1) \sim \mathrm{PD}(\alpha, \theta)$, cf. [58].

One of the key properties of (α, θ)-strings of beads is the regenerative nature inherited from the underlying regenerative interval partition, cf. [58]. Pitman and Winkel [60] developed a method to sample an atom of an (α, θ)-string of beads such that the two strings of beads obtained in this way are rescaled independent (α, α)- and (α, θ)-strings of beads (the first one being the one closer to the origin). The mass split between the two induced interval components and the selected atom is Dirichlet$(\alpha, 1-\alpha, \theta)$, with parameters assigned in order of appearance on the interval $[0, L_{\alpha,\theta}]$.

Proposition 2.18 ([60] Proposition 10/14(b), Corollary 15)**.** *Consider an (α,θ)-string of beads $(I,\lambda) := ([0, L_{\alpha,\theta}], d\mathcal{L}^{-1})$ for some $\alpha \in (0,1), \theta > 0$ with associated (α,θ)-regenerative interval partition $[0,1]\setminus\mathcal{N}_{\alpha,\theta}$ as in* (2.12)*. Consider the switching probability function*

$$p\colon [0,1] \to [0,1], \quad u \mapsto \frac{(1-u)\theta}{(1-u)\theta + u\alpha}.$$

For $t \geq 0$, select a block $(1-e^{-G_{t-}}, 1-e^{-G_t})$ of $[0,1]\setminus\mathcal{N}_{\alpha,\theta}$ with probability

$$p\Big(e^{-\Delta G_t}\Big) \prod_{s<t} \Big(1 - p\Big(e^{-\Delta G_s}\Big)\Big)$$

where $\Delta G_s := G_s - G_{s-} > 0$, and consider the local time $J := \mathcal{L}(1-e^{-G_t})$ related to the selected block $(1-e^{-G_{t-}}, 1-e^{-G_t})$. Then the following random variables are independent.

- *The mass split $(\lambda([0,J)), \lambda(J), \lambda((J, L_{\alpha,\theta}])) \sim \mathrm{Dirichlet}(\alpha, 1-\alpha, \theta)$;*
- *the (α,α)-string of beads $(\lambda([0,J))^{-\alpha}[0,J), \lambda([0,J))^{-1}\lambda\restriction_{[0,J)})$;*
- *the (α,θ)-string of beads $(\lambda((J,L_{\alpha,\theta}])^{-\alpha}(J, L_{\alpha,\theta}], \lambda((J,L_{\alpha,\theta}])^{-1}\lambda\restriction_{(J,L_{\alpha,\theta}]})$.*

We refer to the sampling procedure of Proposition 2.18 as (α,θ)-*coin tossing sampling* in the (α,θ)-string of beads $([0, L_{\alpha,\theta}], d\mathcal{L}^{-1})$. When $\alpha = \theta$, this is just uniform sampling from the mass measure λ.

There is an informal description of coin tossing sampling: consider a walker going along the string of beads $([0, L_{\alpha,\theta}], d\mathcal{L}^{-1})$ from 0, tossing a coin for each of the (countably infinite number of atom) masses aligned along the interval; the heads probability is given by $p(\exp(-\Delta G_t))$, i.e. it depends on the relative remaining mass after an atom; the walker stops when he sees heads for the first time, and we select the associated atom.

In Section 2.5 we will formulate the algorithms of the Introduction based on masses rather than lengths. In particular, the attachment points in the update step will be sampled from the mass and not the length measure. The following statement will be needed to conclude that that the algorithms based on masses imply the length versions.

Lemma 2.19. *Let* $(X_1,\ldots,X_n)\sim\mathrm{Dirichlet}(\theta_1,\ldots,\theta_n)$ *for some* $\theta_1,\ldots,\theta_n>0$ *and* $n\in\mathbb{N}$*, and let* $([0,L_i],\lambda_i)$ *be independent* (α,θ_i)*-strings of beads, respectively.*

- *Select* $I'=j\in[n]$ *with probability* X_j *and, conditionally given* $I'=j$*, select* $L'\in[0,L_j]$ *via* (α,θ_j)*-coin tossing sampling on* $([0,L_j],\lambda_j)$*.*
- *Select* $I''=j\in[n]$ *with probability proportional to* $X_j^\alpha L_j$ *and, conditionally given* $I''=j$*, select* $L''=BL_j$ *where* $B\sim\mathrm{Beta}(1,\theta_j/\alpha)$ *is independent.*

Then

$$\begin{aligned}&(I',L_1,\ldots,L_{I'-1},L',L_{I'}-L',L_{I'+1},\ldots L_n)\\&\qquad\qquad\overset{d}{=}(I'',L_1,\ldots,L_{I''-1},L'',L_{I''}-L'',L_{I''+1},\ldots L_n).\end{aligned}$$

Proof. We show that, for any bounded continuous function $f\colon\mathbb{R}^{n+2}\to\mathbb{R}$

$$\begin{aligned}&\mathbb{E}[f(I',L_1,\ldots,L_{I'-1},L',L_{I'}-L',L_{I'+1},\ldots,L_n)]\\&\qquad=\mathbb{E}[f(I'',L_1,\ldots,L_{I''-1},L'',L_{I''}-L'',L_{I''+1},\ldots L_n)].\end{aligned}\tag{2.15}$$

Conditioning on $I'=j$, and using Proposition 2.14(iv), the LHS of (2.15) is

$$\sum_{j=1}^n\mathbb{E}[f(I',L_1,\ldots,L_{I'-1},L',L_{I'}-L',L_{I'+1},\ldots L_n)\mid I'=j]\left(\theta_j\Big/\sum_{i=1}^n\theta_i\right).$$

Conditionally given $I'=j$, we select an atom of the relevant (α,θ_j)-string of beads via (α,θ_j)-coin tossing sampling. We conclude by Proposition 2.18 and Proposition 2.14(ii) that the mass split $(\lambda_j([0,L')),\lambda_j((L',L_j]))\sim\mathrm{Dirichlet}(\alpha,\theta_j)$, the (α,α)-string of beads $(\lambda_j([0,L'))^{-\alpha}[0,L'),\lambda_j([0,L'))^{-1}\lambda_j\restriction_{[0,L')})$ and the (α,θ)-string of beads $(\lambda_j((L',L_j])^{-\alpha}(L',L_j],\lambda_j((L',L_j])^{-1}\lambda_j\restriction_{(L',L_j]})$ are independent. By Proposition 2.13, we conclude that the relative length split on $[0,L_j]$ is $L'/L_j\sim\mathrm{Beta}(1,\theta_j/\alpha)$. To conclude (2.15), proceed in the same way with the RHS of (2.15), using that, by Proposition 2.13, $(L_1,\ldots,L_n)\sim\mathrm{Dirichlet}(\theta_1/\alpha,\ldots,\theta_n/\alpha)$. More precisely, note that

$\mathbb{P}(I'' = j) = (\theta_j/\alpha)/(\sum_{i=1}^n \theta_i/\alpha) = \theta_j/\sum_{i=1}^n \theta_i$, and that, conditionally given $I'' = j$, we have $L''/L_j \sim \mathrm{Beta}(1, \theta_j/\alpha)$, as before. □

We will also need the following statement about sampling from Poisson-Dirichlet distributions.

Proposition 2.20 (Sampling from $\mathrm{PD}(\alpha,\theta)$, [62] Proposition 34)**.** *Let* $(P_i, i \geq 1) \sim \mathrm{PD}(\alpha, \theta)$ *for some* $0 \leq \alpha < 1$ *and* $\theta > -\alpha$*, and let* N *be an index such that*

$$\mathbb{P}(N = i \mid (P_i, i \geq 1)) = P_i \quad a.s., \qquad i \geq 1.$$

Let $(P_i', i \geq 1)$ *be obtained from* P *by deleting* P_N*, and set* $P_i'' := P_i'/(1 - P_N)$ *for* $i \geq 1$*. Then,* $P_N \sim \mathrm{Beta}(1 - \alpha, \alpha + \theta)$*, and* $(P_i'', i \geq 1) \sim \mathrm{PD}(\alpha, \alpha + \theta)$ *is independent of* P_N*.*

2.2.3 The Mittag-Leffler Markov Chain

The Mittag-Leffler Markov chain $(S_k, k \geq 0)$ is one of the key ingredients in the line-breaking construction of the stable tree, Algorithm 2.4. In the $(k+1)$-st update step a new branch whose length is a Beta proportion B_k of the increment $(S_{k+1} - S_k)$ is attached to the tree constructed in the first k steps, where the remaining proportion $(1 - B_k)$ of $(S_{k+1} - S_k)$ is added to the respective branch point weight. The following results are mainly taken from [39] but note that we consider a more general parameter regime. The proofs can be directly adapted from [39].

The Markov chain $(S_k, k \geq 0)$ can be introduced via a coupling of underlying Chinese restaurant processes as follows. Let $\beta \in (0, 1)$ and $\theta > -\beta$. Consider infinitely many Chinese restaurant processes $(\Pi_n^{(k)}, n \geq k)$, $k \geq 0$, where $\Pi^{(k)}$ is a $(\beta, \theta + k)$-CRP. We couple these CRPs as follows. Inductively, for $k \geq 1$, to start the CRP $\Pi^{(k)}$, ignore the first k customers in the CRP $\Pi^{(k-1)}$ and add their total weight of k to the weight for opening a new table. Let S_k be the scaling limit of the table number in the k-th CRP, i.e. $S_k \sim \mathrm{ML}(\beta, \theta + k)$. One can show that a $\mathrm{Beta}((\theta + k)/\beta + 1, 1/\beta - 1)$ proportion of customers opening a new table in $\Pi^{(k+1)}$ will also open a new table in $\Pi^{(k)}$. The following result is a direct generalisation of [39] Lemma 1.1, where the special case $\theta = 1 - \beta$ was studied. See also [42] Proposition 18(c) for the special case $\theta = \beta$.

Proposition 2.21 (Mittag-Leffler Markov chain)**.** *The process* $(S_k, k \geq 0)$ *satisfies* $S_k = A_k S_{k+1}, k \geq 0$*, where* $A_k \sim \mathrm{Beta}((\theta+k)/\beta+1, 1/\beta-1)$ *and* S_{k+1} *is independent of* $(A_0, \ldots, A_k)$*. Furthermore,* $(S_k, k \geq 0)$ *is a time-homogeneous Markov chain with transition density*

$$f_{S_{k+1}|S_k=z}(y) = f(z,y) = \frac{1-\beta}{\Gamma(1/\beta)}(y-z)^{1/\beta-2}\frac{y g_\beta(y)}{g_\beta(z)}, \quad y > z > 0,$$

where $g_\beta \colon (0,\infty) \to (0,\infty)$ *denotes the density of* $\mathrm{ML}(\beta, 0)$*.*

We showed that $S_k \sim \mathrm{ML}(\beta, \theta+k)$ in Proposition 2.15. An expression for g_β in terms of densities of stable subordinators can be found in [58]. Recall that S_k is the sum of the branch point weights and the length of the tree at step k in Algorithm 2.4. In the two-colour line-breaking construction, Algorithm 2.6, S_k is the length of the tree at step k, as needed for the weight-length representation.

We also refer to James [44] who obtained the Mittag-Leffler Markov chain via limits in certain preferential attachment models.

2.3 Marked metric spaces

2.3.1 $\mathbb{R}$-trees and the Gromov-Hausdorff topology

We use the notion of an $\mathbb{R}$-tree, that is, a separable metric space $(\mathcal{T}, d)$ such that the following two properties hold for every $\sigma_1, \sigma_2 \in \mathcal{T}$.

(i) There is an isometric map $h_{\sigma_1,\sigma_2} \colon [0, d(\sigma_1,\sigma_2)] \to \mathcal{T}$ such that

$$h_{\sigma_1,\sigma_2}(0) = \sigma_1 \text{ and } h_{\sigma_1,\sigma_2}(d(\sigma_1,\sigma_2)) = \sigma_2.$$

(ii) For every injective path $q \colon [0,1] \to \mathcal{T}$ with $q(0) = \sigma_1$ and $q(1) = \sigma_2$ we have

$$q([0,1]) = h_{\sigma_1,\sigma_2}([0, d(\sigma_1,\sigma_2)]).$$

We write $[[\sigma_1,\sigma_2]] := h_{\sigma_1,\sigma_2}([0, d(\sigma_1,\sigma_2)])$ for the range of h_{σ_1,σ_2}, and refer to the combinatorial graph structure associated with $(\mathcal{T}, d)$, that is, the finite connected and labelled graph with no cycles and a-priori no degree-2 vertices, as the *shape* of $\mathcal{T}$, i.e. the shape of $\mathcal{T}$ is the graph whose vertices are the root, the leaves and the branch points of $\mathcal{T}$.

The trees considered in this chapter are usually compact, but we also allow non-compact $\mathbb{R}$-trees. A *rooted* $\mathbb{R}$-tree $(\mathcal{T}, d, \rho)$ is an $\mathbb{R}$-tree $(\mathcal{T}, d)$ with a distinguished element $\rho = \rho(\mathcal{T})$, the *root*. We only consider rooted $\mathbb{R}$-trees, and will often refer to $\mathcal{T}$ as an $\mathbb{R}$-tree without mentioning the distance d and the root ρ explicitly. For any $\alpha > 0$ and any metric space $(\mathcal{T}, d)$, we write $\alpha\mathcal{T}$ for $(\mathcal{T}, \alpha d)$, the metric space obtained when all distances are multiplied by α.

We are only interested in equivalence (or isometry) classes of rooted $\mathbb{R}$-trees. Two rooted $\mathbb{R}$-trees $(\mathcal{T}, d, \rho)$ and $(\mathcal{T}', d', \rho')$ are *equivalent* if there exists an isometry from $\mathcal{T}$ onto $\mathcal{T}'$ such that ρ is mapped onto ρ'. The set of equivalence classes of compact rooted $\mathbb{R}$-trees is denoted by $\mathbb{T}$.

We follow [32] and equip $\mathbb{T}$ with the pointed Gromov-Hausdorff distance d_{GH} to obtain the Polish space $(\mathbb{T}, d_{\mathrm{GH}})$. The *pointed Gromov-Hausdorff distance* between rooted $\mathbb{R}$-trees $(\mathcal{T}, d, \rho)$, $(\mathcal{T}', d', \rho')$ is

$$d_{\mathrm{GH}}\big((\mathcal{T}, d, \rho), (\mathcal{T}', d', \rho')\big) := \inf_{\varphi, \varphi'} \{\max\{\delta(\varphi(\rho), \varphi'(\rho')), \delta_{\mathrm{H}}(\varphi(\mathcal{T}), \varphi'(\mathcal{T}'))\}\}, \tag{2.16}$$

where δ_{H} is the Hausdorff distance between compact subsets of $(\mathcal{M}, \delta)$, and the infimum is taken over all metric spaces $(\mathcal{M}, \delta)$ and all isometric embeddings $\varphi \colon \mathcal{T} \to \mathcal{M}$, $\varphi' \colon \mathcal{T}' \to \mathcal{M}$ into $(\mathcal{M}, \delta)$. It can be shown that the Gromov-Hausdorff distance only depends on equivalence classes of rooted $\mathbb{R}$-trees. We equip $\mathbb{T}$ with its Borel σ-algebra $\mathcal{B}(\mathbb{T})$ induced by d_{GH}. We often consider random $\mathbb{R}$-trees whose equivalence class in $\mathbb{T}$ has the same distribution as a Ford CRT or a stable tree of parameter β. We also refer to these trees as Ford CRT or stable tree, respectively.

A *weighted* $\mathbb{R}$-tree $(\mathcal{T}, d, \rho, \mu)$ is a rooted $\mathbb{R}$-tree $(\mathcal{T}, d, \rho)$ equipped with a probability measure μ on the Borel sets $\mathcal{B}(\mathcal{T})$. Two weighted $\mathbb{R}$-trees $(\mathcal{T}, d, \rho, \mu)$ and $(\mathcal{T}', d', \rho', \mu')$ are *equivalent* if there is an isometry from $(\mathcal{T}, d, \rho)$ onto $(\mathcal{T}', d', \rho')$ such that μ' is the push-forward of μ under this isometry. The set of equivalence (or isometry) classes of weighted compact $\mathbb{R}$-trees is denoted by $\mathbb{T}_{\mathrm{w}}$.

The Gromov-Hausdorff distance can be extended to a metric on weighted $\mathbb{R}$-trees, the Gromov-Hausdorff-Prokhorov metric, which only depends on equivalence classes of weighted $\mathbb{R}$-trees, see e.g. [54] Section 6. The *pointed Gromov-Hausdorff-Prokhorov distance* $d_{\mathrm{GHP}}((\mathcal{T}, d, \rho, \mu), (\mathcal{T}', d', \rho', \mu'))$ between two weighted $\mathbb{R}$-trees $(\mathcal{T}, d, \rho, \mu)$ and $(\mathcal{T}', d', \rho', \mu')$ is

$$\inf_{\varphi,\varphi'}\{\max\{\delta(\varphi(\rho), \varphi'(\rho')), \delta_{\mathrm{H}}(\varphi(\mathcal{T}), \varphi'(\mathcal{T}')), \delta_{\mathrm{P}}(\varphi_*\mu, \varphi'_*\mu')\}\}, \tag{2.17}$$

where $\varphi, \varphi', \delta_{\mathrm{H}}$ are as in (2.16), $\varphi_*\mu$, $\varphi'_*\mu$ are the push-forwards of μ, μ' via φ, φ', respectively, and δ_{P} is the Prokhorov distance on the space of Borel probability measures on $(\mathcal{M}, \delta)$ given by

$$\delta_{\mathrm{P}}(\mu, \mu') = \inf\{\epsilon > 0 : \mu(D) \leq \mu'(D^\epsilon) + \epsilon \text{ and } \mu'(D) \leq \mu(D^\epsilon) + \epsilon \quad \forall D \subset \mathcal{M} \text{ closed}\}$$

where $D^\epsilon := \{x \in \mathcal{M} : \inf_{y\in D} \delta(x, y) < \epsilon\}$ is the *ϵ-thickening* of C.

Proposition 2.22 ([54] Proposition 8(i))**.** *The space* $(\mathbb{T}_{\mathrm{w}}, d_{\mathrm{GHP}})$ *is separable and complete.*

We will also need some terminology to describe an $\mathbb{R}$-tree $(\mathcal{T}, d, \rho)$. For any $x \in \mathcal{T}$, we call $d(\rho, x)$ the *height* of x, $\mathrm{ht}(\mathcal{T}) := \sup_{x\in\mathcal{T}} d(\rho, x)$ the *height* of $\mathcal{T}$, and $\mathrm{diam}(\mathcal{T}) := \sup_{x,y\in\mathcal{T}} d(x, y)$ the *diameter* of $\mathcal{T}$. A *leaf* is an element $x \in \mathcal{T} \setminus \{\rho\}$ such that $\mathcal{T} \setminus \{x\}$ is connected. We denote the set of all leaves of $\mathcal{T}$ by $\mathrm{Lf}(\mathcal{T})$, and call $\#\mathrm{Lf}(\mathcal{T})$ the *size* of $\mathcal{T}$. An element $x \in \mathcal{T} \setminus \{\rho\}$ is a *branch point* if $\mathcal{T} \setminus \{x\}$ has at least three connected components. The *degree* $\deg(x, \mathcal{T})$ of a vertex $x \in \mathcal{T}$ is the number of connected components of $\mathcal{T} \setminus \{x\}$. The connected components of $\mathcal{T}$ obtained when we remove all leaves and all branch points of $\mathcal{T}$ are called *edges*.

A weighted $\mathbb{R}$-tree $(\mathcal{T}, d, \rho, \mu)$ is a *continuum tree* if the probability measure μ on $\mathcal{T}$ satisfies the following three properties.

(i) μ is supported by $\mathrm{Lf}(\mathcal{T})$, the set of leaves of $\mathcal{T}$.

(ii) μ has no atom, i.e. for any singleton $x \in \mathrm{Lf}(\mathcal{T})$ we have $\mu(x) = 0$.

(iii) For every $x \in \mathcal{T} \setminus \mathrm{Lf}(\mathcal{T})$, $\mu(\mathcal{T}_x) > 0$, where $\mathcal{T}_x := \{\sigma \in \mathcal{T} : x \in [[\rho, \sigma]]\}$.

Note that these properties imply that $\mathrm{Lf}(\mathcal{T})$ is uncountable and has no isolated points. We will also need the notion of a *reduced* tree: for an $\mathbb{R}$-tree $\mathcal{T}$ and any $x_1, x_2, \ldots, x_n \in \mathrm{Lf}(\mathcal{T})$ let

$$\mathcal{R}(\mathcal{T}, x_1, \ldots, x_n) := \bigcup_{i=1}^{n} [[\rho, x_i]] \tag{2.18}$$

be the *reduced* subtree associated with $\mathcal{T}, x_1, \ldots, x_n$, that is, the tree spanned by the root ρ and $x_1, \ldots, x_n$.

$\mathcal{R}(\mathcal{T}, x_1, \ldots, x_n)$ is an $\mathbb{R}$-tree with root ρ and leaves $x_1, \ldots, x_n$. We consider the projection map

$$\pi^{\mathcal{T}_k} : \mathcal{T} \to \mathcal{T}_k, \quad u \mapsto h_{\rho,u}(\sup\{t \geq 0 : h_{\rho,u}(t) \in \mathcal{T}_k\}) \tag{2.19}$$

where $\mathcal{T}_k = \mathcal{R}(\mathcal{T}, x_1, \ldots, x_k)$ for some $x_1, \ldots, x_k \in \mathrm{Lf}(\mathcal{T})$, and $h_{\rho,u} \colon [0, d(\rho, u)] \to \mathcal{T}$ is the unique isometry with $h_{\rho,u}(0) = \rho$ and $h_{\rho,u}(d(\rho, u)) = u$ from the definition of an $\mathbb{R}$-tree. For any probability measure μ on $\mathcal{T}$, we denote the push-forward of μ via this projection map by $\pi_*^{\mathcal{R}_k}\mu$, i.e.

$$\pi_*^{\mathcal{R}_k}\mu(D) = \mu\Big(\Big(\pi^{\mathcal{R}_k}\Big)^{-1}(D)\Big), \quad D \in \mathcal{B}(\mathcal{T}_k) := \{D' \subset \mathcal{T}_k : D' \text{ Borel measurable}\}. \tag{2.20}$$

We refer to [20, 31, 47] for a rigorous study and details on the topic of $\mathbb{R}$-trees.

2.3.2 ∞-marked $\mathbb{R}$-trees

We introduce ∞-marked $\mathbb{R}$-trees to capture the framework of an $\mathbb{R}$-tree with infinitely many marked components. This is a generalisation of Miermont's concept of a k-marked metric space, [54] Section 6.4. In the two-colour line-breaking construction, the marked components correspond to the rescaled Ford CRTs by which we replace the branch points in the stable tree. Each Ford CRT, i.e. each connected red component, is related to a new marked subset of the ∞-marked metric space.

A *k-marked* $\mathbb{R}$-tree $(\mathcal{T}, d, \rho, (\mathcal{R}^{(1)}, \ldots, \mathcal{R}^{(k)}))$, $k \geq 1$, is a rooted $\mathbb{R}$-tree $(\mathcal{T}, d, \rho)$ with compact non-empty connected subsets $\mathcal{R}^{(1)}, \ldots, \mathcal{R}^{(k)} \subset \mathcal{T}$. As in the unmarked case, we do not always mention the distance d and the root ρ explicitly, and rather write $(\mathcal{T}, (\mathcal{R}^{(1)}, \dots, \mathcal{R}^{(k)}))$. We call two k-marked $\mathbb{R}$-trees $(\mathcal{T}, d, \rho, (\mathcal{R}^{(1)}, \dots, \mathcal{R}^{(k)}))$ and $(\mathcal{T}', d', \rho, (\mathcal{R}'^{(1)}, \dots, \mathcal{R}'^{(k)}))$ *equivalent* if there exists an isometry from $\mathcal{T}$ to $\mathcal{T}'$ such that each $\mathcal{R}^{(i)}$ is mapped onto $\mathcal{R}'^{(i)}$, $i \in [k]$, respectively. If $\mathcal{T}$ and $\mathcal{T}'$ are equipped with mass measures μ and μ', we speak of *weighted* k-marked $\mathbb{R}$-trees, and we call them *equivalent* if there is an isometry from $\mathcal{T}$ to $\mathcal{T}'$ such that each $\mathcal{R}^{(i)}$ is mapped onto $\mathcal{R}'^{(i)}$, $i \in [k]$, respectively, and μ' is the push-forward of μ under this isometry. The set of all equivalence classes of k-marked compact $\mathbb{R}$-trees is denoted by $\mathbb{T}^{[k]}$, and $\mathbb{T}^{[k]}_{\mathrm{w}}$ is the set of equivalence classes of weighted k-marked compact $\mathbb{R}$-trees.

We equip $\mathbb{T}^{[k]}$ with $d_{\mathrm{GH}}^{[k]}$, a generalisation of the Gromov-Hausdorff distance on the set of compact metric spaces. For $(\mathcal{T}, (\mathcal{R}^{(1)}, \dots, \mathcal{R}^{(k)}))$, $(\mathcal{T}', (\mathcal{R}'^{(1)}, \dots, \mathcal{R}'^{(k)})) \in \mathbb{T}^{[k]}$ with roots ρ, ρ', respectively, define

$$
\begin{aligned}
&d_{\mathrm{GH}}^{[k]}\Big(\Big(\mathcal{T}, \Big(\mathcal{R}^{(1)}, \dots, \mathcal{R}^{(k)}\Big), \rho\Big), \Big(\mathcal{T}', \Big(\mathcal{R}'^{(1)}, \dots, \mathcal{R}'^{(k)}\Big), \rho'\Big)\Big) \\
&\quad := \inf_{\varphi, \varphi'} \Big\{ \max\Big\{ \delta_{\mathrm{H}}(\varphi(\mathcal{T}), \varphi'(\mathcal{T}')), \max_{1 \le i \le k} \delta_{\mathrm{H}}\Big(\varphi\Big(\mathcal{R}^{(i)}\Big), \varphi'\Big(\mathcal{R}'^{(i)}\Big)\Big), \delta(\varphi(\rho), \varphi(\rho')) \Big\}\Big\}
\end{aligned}
\tag{2.21}
$$

where the infimum is taken over all isometric embeddings φ, φ' of $\mathcal{T}$, $\mathcal{T}'$ into a common metric space $(\mathcal{M}, \delta)$, and δ_{H} is the Hausdorff distance on $(\mathcal{M}, \delta)$. It was shown in [54] that $d_{\mathrm{GH}}^{[k]}$ is a metric on $\mathbb{T}^{[k]}$.

Lemma 2.23 ([54] Proposition 9(ii))**.** *The space* $(\mathbb{T}^{[k]}, d_{\mathrm{GH}}^{[k]})$ *is separable and complete.*

We extend the definition of a k-marked $\mathbb{R}$-trees to the notion of an ∞-*marked* $\mathbb{R}$-tree $(\mathcal{T}, d, \rho, (\mathcal{R}^{(i)}, i \ge 1))$, that is a rooted $\mathbb{R}$-tree $(\mathcal{T}, d, \rho)$ with compact non-empty subsets $\mathcal{R}^{(i)} \subset \mathcal{T}$, $i \ge 1$. As in the k-marked case, ∞-*marked* $\mathbb{R}$-trees $(\mathcal{T}, d, \rho, (\mathcal{R}^{(i)}, i \ge 1))$, $(\mathcal{T}', d', \rho', (\mathcal{R}'^{(i)}, i \ge 1))$ are *equivalent* if there is an isometry from $\mathcal{T}$ to $\mathcal{T}'$ such that $\mathcal{T}$ is mapped onto $\mathcal{T}'$, and each $\mathcal{R}^{(i)}$ is mapped onto $\mathcal{R}'^{(i)}$, $i \ge 1$, respectively. We write $\mathbb{T}^{\infty}$ for the set of equivalence classes of ∞-marked compact $\mathbb{R}$-trees, and equip it with the metric $d_{\mathrm{GH}}^{\infty} := \sum_{k \ge 1} 2^{-k} d_{\mathrm{GH}}^{[k]}$, i.e. for $(\mathcal{T}, (\mathcal{R}^{(i)}, i \ge 1))$, $(\mathcal{T}', (\mathcal{R}'^{(i)}, i \ge 1)) \in \mathbb{T}^{\infty}$ with roots ρ, ρ', respectively,

$$
d_{\mathrm{GH}}^{\infty}\Big(\Big(\mathcal{T}, \Big(\mathcal{R}^{(i)}, i \ge 1\Big), \rho\Big), \Big(\mathcal{T}', \Big(\mathcal{R}'^{(i)}, i \ge 1\Big), \rho'\Big)\Big)
$$

$$:= \sum_{k\geq 1} 2^{-k} d_{\mathrm{GH}}^{[k]}\Big(\Big(\mathcal{T}, \Big(\mathcal{R}^{(i)}, \ldots, \mathcal{R}^{(k)}\Big), \rho\Big), \Big(\mathcal{T}', \Big(\mathcal{R}'^{(1)}, \ldots, \mathcal{R}'^{(k)}\Big), \rho', \Big)\Big). \quad (2.22)$$

Corollary 2.24. *The space* $(\mathbb{T}^{\infty}, d_{\mathrm{GH}}^{\infty})$ *is separable and complete.*

Proof. By Lemma 2.23, for any $k \geq 1$, there exists a countable dense subset $\mathbb{C}_k \subset \mathbb{T}_k$ such that for any $\epsilon > 0$ and any $(\mathcal{T}, (\mathcal{R}^{(1)}, \ldots, \mathcal{R}^{(k)})) \in \mathbb{T}^{[k]}$, there is $(\mathcal{T}', (\mathcal{R}'^{(1)}, \ldots, \mathcal{R}'^{(k)})) \in \mathbb{C}_k$ with

$$d_{\mathrm{GH}}^{[k]}\Big(\Big(\mathcal{T}, \Big(\mathcal{R}^{(1)}, \ldots, \mathcal{R}^{(k)}\Big)\Big), \Big(\mathcal{T}', \Big(\mathcal{R}'^{(1)}, \ldots, \mathcal{R}'^{(k)}\Big)\Big)\Big) < \epsilon. \quad (2.23)$$

Let ρ' denote the root of $\mathcal{T}'$. For any $k \geq 1$, the set

$$\mathbb{C}_k' := \Big\{\Big(\mathcal{T}', \Big(\mathcal{R}^{(1')}, \ldots, \mathcal{R}^{(k')}, \{\rho'\}, \{\rho'\}, \ldots\Big)\Big) : \Big(\mathcal{T}', \Big(\mathcal{R}'^{(1)}, \ldots, \mathcal{R}'^{(k)}\Big)\Big) \in \mathbb{C}_k\Big\}$$

is clearly countable. For any $(\mathcal{T}, (\mathcal{R}^{(1)}, \mathcal{R}^{(2)}, \ldots)) \in \mathbb{T}^{\infty}$ there is an element

$$\Big(\mathcal{T}', \Big(\mathcal{R}'^{(1)}, \ldots, \mathcal{R}'^{(k)}, \{\rho'\}, \{\rho'\}, \ldots\Big)\Big) \in \mathbb{C}_k'$$

such that (2.23) holds when we restrict to the first k marked components, i.e.

$$d_{\mathrm{GH}}^{\infty}\Big(\Big(\mathcal{T}, \Big(\mathcal{R}^{(i)}, i \geq 1\Big)\Big), \Big(\mathcal{T}', \Big(\mathcal{R}'^{(i)}, i \geq 1\Big)\Big)\Big) \leq 2^{-k}\epsilon \sum_{i=0}^{k-1} 2^i + \sum_{n\geq k+1} 2^{-n}\mathrm{diam}(\mathcal{T}) < \epsilon$$

for k large enough. Therefore, $\mathbb{C}_\infty := \bigcup_{k\geq 1} \mathbb{C}_k'$ is countable and dense in $\mathbb{T}^{\infty}$, i.e. $(\mathbb{T}^{\infty}, d_{\mathrm{GH}}^{\infty})$ is separable.

To see completeness, consider a Cauchy sequence $(\mathcal{T}_n, (\mathcal{R}_n^{(i)}, i \geq 1), n \geq 1)$ in $\mathbb{T}^{\infty}$. By definition of d_{GH}^{∞}, for any $k \geq 1$, the sequence $(\mathcal{T}_n, (\mathcal{R}_n^{(1)}, \ldots, \mathcal{R}_n^{(k)}), n \geq 1)$ is clearly Cauchy in $\mathbb{T}^{[k]}$. By Lemma 2.23, $(\mathbb{T}^{[k]}, d_{\mathrm{GH}}^{[k]})$ is complete, and we conclude that there is $(\mathcal{T}, (\mathcal{R}^{(1)}, \ldots, \mathcal{R}^{(k)})) \in \mathbb{T}^{[k]}$ such that

$$\lim_{n\to\infty}\Big(\mathcal{T}, \Big(\mathcal{R}_n^{(1)}, \ldots, \mathcal{R}_n^{(k)}\Big)\Big) = \Big(\mathcal{T}, \Big(\mathcal{R}^{(1)}, \ldots, \mathcal{R}^{(k)}\Big)\Big). \quad (2.24)$$

Furthermore, the limits (2.24) are easily seen to be consistent as k varies, i.e. given (2.24) for some $k \geq 2$,

$$\lim_{n\to\infty}\Big(\mathcal{T},\Big(\mathcal{R}_n^{(1)},\ldots,\mathcal{R}_n^{(k-1)}\Big)\Big)=\Big(\mathcal{T},\Big(\mathcal{R}^{(1)},\ldots,\mathcal{R}^{(k-1)}\Big)\Big) \tag{2.25}$$

in $(\mathbb{T}^{[k-1]}, d_{\mathrm{GH}}^{[k-1]})$. We conclude that there is $(\mathcal{T},(\mathcal{R}^{(1)},\mathcal{R}^{(2)},\ldots))$ such that

$$\lim_{n\to\infty} d_{\mathrm{GH}}^{\infty}\Big(\Big(\mathcal{T}_n,\Big(\mathcal{R}_n^{(i)},i\geq 1\Big)\Big),\Big(\mathcal{T},\Big(\mathcal{R}^{(i)},i\geq 1\Big)\Big)\Big)=0. \qquad \square$$

We can extend $d_{\mathrm{GH}}^{[k]}$ to a metric on $\mathbb{T}_{\mathrm{w}}^{[k]}$ by adding a Prokhorov component to $d_{\mathrm{GH}}^{[k]}$. For any $k \in \{1,2,\ldots\}$ and $(\mathcal{T},(\mathcal{R}^{(1)},\ldots,\mathcal{R}^{(k)}),\mu)$, $(\mathcal{T}',(\mathcal{R}'^{(1)},\ldots,\mathcal{R}'^{(k)}),\mu') \in \mathbb{T}_{\mathrm{w}}^{[k]}$, we define $d_{\mathrm{GHP}}^{[k]}((\mathcal{T},(\mathcal{R}^{(1)},\ldots,\mathcal{R}^{(k)}),\mu),(\mathcal{T}',(\mathcal{R}'^{(1)},\ldots,\mathcal{R}'^{(k)}),\mu'))$ as the infimum over all φ,φ' of

$$\max\Big\{\delta_{\mathrm{H}}(\varphi(\mathcal{T}),\varphi'(\mathcal{T}')),\max_{1\leq i\leq k}\delta_{\mathrm{H}}\Big(\varphi\Big(\mathcal{R}^{(i)}\Big),\varphi'\Big(\mathcal{R}'^{(i)}\Big)\Big),\delta(\varphi(\rho),\varphi'(\rho')),\delta_{\mathrm{P}}(\varphi_*\mu,\varphi'_*\mu')\Big\}$$

where φ,φ' and $\varphi_*\mu$, $\varphi'_*\mu'$ are as in (2.17) and (2.21). In the spirit of (2.22), we define

$$\begin{aligned} &d_{\mathrm{GHP}}^{\infty}\Big(\Big(\mathcal{T},\Big(\mathcal{R}^{(i)},i\geq 1\Big),\mu\Big),\Big(\mathcal{T}',\Big(\mathcal{R}'^{(i)},i\geq 1\Big),\mu'\Big)\Big)\\ &\qquad=\sum_{k\geq 1}2^{-k}d_{\mathrm{GHP}}^{[k]}\Big(\Big(\mathcal{T},\Big(\mathcal{R}^{(1)},\ldots,\mathcal{R}^{(k)}\Big),\mu\Big),\Big(\mathcal{T}',\Big(\mathcal{R}'^{(1)},\ldots,\mathcal{R}'^{(k)}\Big),\mu'\Big)\Big) \end{aligned} \tag{2.26}$$

for two weighted ∞-marked $\mathbb{R}$-trees $(\mathcal{T},(\mathcal{R}^{(i)},i\geq 1),\mu)$ and $(\mathcal{T}',(\mathcal{R}'^{(i)},i\geq 1),\mu')$.

Lemma 2.25. *The function $d_{\mathrm{GHP}}^{[k]}$ defines a distance on $\mathbb{T}_{\mathrm{w}}^{[k]}$, and the space $(\mathbb{T}_{\mathrm{w}}^{[k]}, d_{\mathrm{GHP}}^{[k]})$ is separable and complete, for any $k \in \{1,2,\ldots;\infty\}$.*

Proof. For $k \geq 1$, the proof is a direct generalisation of the proof of Lemma 2.23. In particular, it is straightforward to generalise the results about d_{GHP} in [54] Section 6.2 and Section 6.3 to $d_{\mathrm{GHP}}^{[k]}$. For $k = \infty$, the claim can then be deduced as in the proof of Corollary 2.24. $\square$

As a natural special case of weighted k-marked $\mathbb{R}$-trees we can introduce *k-marked strings of beads* of the form

$$(I = [0, K], I_1 = [K_1, K_1'], \ldots, I_k = [K_k, K_k'], \lambda),$$

i.e. intervals $I = [0, K]$ with marked subsets $I_i = [K_i, K_i'] \subset [0, K]$, $1 \leq i \leq k$, equipped with a discrete mass measure λ. We can view a k-marked string of beads as a weighted k-marked $\mathbb{R}$-tree. It will be sufficient to restrict our considerations to equivalence classes of 1-marked strings of beads, where the marked component $[K_1, K_1']$ is of the form $[0, K_1']$, i.e. $K_1 = 0$.

Remark 2.26*.* We view the points of colour changes in k- and ∞-marked $\mathbb{R}$-trees as branch points of degree 2, i.e. an interval $[0, K]$ with a marked component $[0, K_1]$ as in a 1-marked string of beads is related to two edges in the tree.

Remark 2.27*.* Miermont [54] introduced the more general concept of a k-marked metric space, and studied the space $\mathbb{M}^{[k]}$ of equivalence classes of k-marked metric spaces. $\mathbb{T}^{[k]}$ is a closed subset of $\mathbb{M}^{[k]}$ ([32] Lemma 2.1), i.e. the results on $(\mathbb{T}^{[k]}, d_{\mathrm{GH}}^{[k]})$ presented here are a direct consequence of his studies of $(\mathbb{M}^{[k]}, d_{\mathrm{GH}}^{[k]})$.

2.4 Construction of CRTs using recursive tree frameworks

2.4.1 Recursive tree frameworks

We briefly recap the concept of a *recursive tree framework* (RTF) introduced by Aldous and Bandyopadhyay [13], where general recursive distributional equations were studied with regard to the existence of fixpoints. Such recursive relationships arise in a variety of contexts, e.g. in algorithmic structures, Galton-Watson branching processes and random trees. We will use the recursive distributional relations underlying an RTF to give a recursive construction of CRTs based on a string of beads.

Consider two measurable spaces $(\mathbb{T}, \mathcal{A}_{\mathbb{T}})$ and $(\Xi, \mathcal{A}_{\Xi})$, and construct

$$\Xi^* := \Xi \times \bigcup_{0 \leq m \leq \infty} \mathbb{T}^m, \tag{2.27}$$

where $\mathbb{T}^m$ denotes the space of $\mathbb{T}$-valued sequences of length m, $0 \leq m \leq \infty$, and where we denote the empty sequence by Υ, i.e. $\mathbb{T}^0 = \{\Upsilon\}$. Furthermore, consider a measurable map

$$\phi : \Xi^* \to \mathbb{T}, \tag{2.28}$$

and random variables $(\xi, N) \in \Xi \times \overline{\mathbb{N}} := \Xi \times \{1, 2, \ldots; \infty\}$, and $(\tau_i, i \geq 1) \in \mathbb{T}^\infty$ as follows.

(i) The pair (ξ, N) has some distribution ν, i.e. $(\xi, N) \sim \nu$.

(ii) The sequence $(\tau_i, i \geq 1)$ is i.i.d. with some distribution η, i.e. $\tau_1 \sim \eta$.

(iii) The random variables in (i) and (ii) are independent.

We denote by $\mathcal{P}(\mathbb{T})$ the space of probability measures on the space $\mathbb{T}$. For any given distribution on $\Xi \times \overline{\mathbb{N}}$, we then obtain a map

$$\Phi : \mathcal{P}(\mathbb{T}) \to \mathcal{P}(\mathbb{T}), \quad \eta \mapsto \Phi(\eta) \tag{2.29}$$

as an analogue of (2.28) on $\mathcal{P}(\mathbb{T})$, i.e. $\Phi(\eta)$ is defined as the distribution of

$$\tau := \phi(\xi, \tau_i, 1 \leq i \leq N). \tag{2.30}$$

We call τ the *parent* value of $(\tau_i, 1 \leq i \leq N)$, and refer to $(\tau_i, 1 \leq i \leq N)$, as the values of the *children* of τ. We now proceed recursively, and view the random variables τ_i as the parent values of random variables $\tau_{ij}, 1 \leq j \leq N_i$, associated with ξ_i, i.e.

$$\tau_i = \phi(\xi_i, \tau_{ij}, 1 \leq j \leq N_i), \quad 1 \leq i \leq N.$$

This setting can be extended recursively to each of the following generations, i.e. each child is considered as a parent itself. We refer to this setting as a *recursive tree framework*. More precisely, we define

$$\mathcal{N} := \bigcup_{n \geq 0} \mathbb{N}^n$$

and use the Ulam-Harris notation to describe the set of all possible descendants $\mathbf{i}$, where $\mathbf{i} = i_1 i_2 \cdots i_n$ denotes an individual at generation $n \geq 1$, which is the i_n-th child of the parent $i_1 i_2 \cdots i_{n-1}$. If we let the empty vector $\varnothing$ (the only element of $\mathbb{N}^0$) be the root of $\mathcal{N}$ and link parents and children via edges, the space $\mathcal{N}$ can be viewed as a (discrete) tree.

Definition 2.28 (Recursive tree framework)**.** A *recursive tree framework* (RTF) is a pair $\{(\xi_{\mathbf{i}}, N_{\mathbf{i}}, \mathbf{i} \in \mathcal{N}); \phi\}$ where $(\xi_{\mathbf{i}}, N_{\mathbf{i}}, \mathbf{i} \in \mathcal{N})$ is a sequence of $\Xi \times \overline{\mathbb{N}}$-valued i.i.d. random variables $(\xi_{\mathbf{i}}, N_{\mathbf{i}}) \sim \nu$, $\mathbf{i} \in \mathcal{N}$, and $\phi : \Xi^* \to \mathbb{T}$ is a measurable map.

Suppose there are random variables $\tau_{\mathbf{i}}, \mathbf{i} \in \mathcal{N}$, possibly on an extended probability space, as follows.

(ii) For all $\mathbf{i} \in \mathcal{N}$,

$$\tau_{\mathbf{i}} = \phi(\xi_{\mathbf{i}}, \tau_{\mathbf{i}j}, 1 \le j \le N_{\mathbf{i}}) \quad \text{a.s..} \tag{2.31}$$

(ii) The random variables

$$\{\tau_{\mathbf{i}} : \mathbf{i} \text{ in generation } n\} \tag{2.32}$$

are i.i.d. with some distribution η_n.

(iii) The random variables $\{\tau_{\mathbf{i}} : \mathbf{i} \text{ in generation } n\}$ are independent of the random variables

$$\{(\xi_{\mathbf{i}}, N_{\mathbf{i}}) : \mathbf{i} \text{ in first } n-1 \text{ generations}\}. \tag{2.33}$$

.

Note that $\{\tau_{\mathbf{i}} : \mathbf{i} \text{ in generation } n\} = \{\tau_{\mathbf{i}} : \mathbf{i} \in \mathbb{N}^n\}$. A *recursive tree process* (RTP) is a recursive tree framework with random variables $\tau_{\mathbf{i}}, \mathbf{i} \in \mathcal{N}$, as in (2.31)-(2.32), i.e. an RTP is an RTF enriched by the random variables $\tau_{\mathbf{i}}, \mathbf{i} \in \mathcal{N}$. An RTP with random variables $\tau_{\mathbf{i}}$ only defined up to generation n, i.e. only for $\mathbf{i} \in \bigcup_{m=0}^{n} \mathbb{N}^m$, is called an RTP *of depth n*. While RTPs of depth n can always be defined using (2.32) for any distribution η_n and (2.31) for generations $n-1, \ldots, 0$, RTPs of infinite depth do not exist in general. We refer to [13] Section 2.3 for more details on RTFs and RTPs, in particular with regard to connections to Markov chains and Markov transition kernels.

In what follows, consider a fixed recursive tree framework, i.e. let $(\xi_{\mathbf{i}}, N_{\mathbf{i}})$, $\mathbf{i} \in \mathcal{N}$, be i.i.d. with distribution ν, and let $\phi : \Xi^* \to \mathbb{T}$ be a measurable map. Given $n \ge 1$ and an arbitrary distribution η_n on $\mathbb{T}$, we consider a recursive tree process where the values of n-th generation individuals are i.i.d. with distribution η_n. The distributions of the values of j-th generation individuals are then given by

$$\tau_{\mathbf{i}} \sim \eta_j := \Phi^{n-j}(\eta_n), \quad \mathbf{i} \in \mathbb{N}^j, \tag{2.34}$$

for $1 \leq j \leq n-1$. Aldous and Bandyopadhyay [13] studied fixpoints of Φ. Note that the existence of a fixpoint η^* of Φ ensures the existence of a stationary RTP, i.e. an RTP with $\eta_n = \eta$, $n \geq 0$, by Kolomogorov's consistency theorem.

Lemma 2.29 (The contraction method, [13] Lemma 5)**.** *Let $\mathcal{P} \subset \mathcal{P}(\mathbb{T})$ such that $\Phi(\mathcal{P}) \subset \mathcal{P}$, i.e. consider $\Phi\colon \mathcal{P} \to \mathcal{P}$ as in* (2.29)-(2.30) *related to a recursive tree process as above. Furthermore, let $d_{\mathcal{P}}$ be a complete metric on $\mathcal{P}$ such that the contraction property holds, i.e.*

$$\sup_{\eta,\eta' \in \mathcal{P}, \eta \neq \eta'} \frac{d_{\mathcal{P}}(\Phi(\eta), \Phi(\eta'))}{d_{\mathcal{P}}(\eta, \eta')} < 1. \tag{2.35}$$

Then the map Φ has a unique fixpoint $\eta^* \in \mathcal{P}$, and the domain of attraction of η^* is the whole set $\mathcal{P}$, i.e. for all $\eta \in \mathcal{P}$, $\lim_{j\to\infty} \Phi^j(\eta) = \eta^*$ in the sense of weak convergence.

2.4.2 The setting for recursive tree processes in the context of CRTs

We now use a specific recursive tree framework to construct a continuum random tree. We will apply this construction to build a compact CRT which is related to the tree constructed in the two-colour line-breaking construction (without colour marks).

Let $\mathbb{T}$ be the space of all equivalence classes of (rooted) compact $\mathbb{R}$-trees, as in Section 2.3.1, and let $\widetilde{\Xi}$ be the space of strings of beads defined by

$$\widetilde{\Xi} = \left\{ \xi : \xi = \left(I = [0, K], \lambda = \sum_{i \geq 1} P_i \delta_{x_i} \right) \text{ such that (i)-(iii) hold} \right\} \tag{2.36}$$

where the properties (i)-(iii) are given by

(i) $K \geq 0$ and $0 \leq x_i \leq K$ for all $i \geq 1$;

(ii) $0 < \sum_{i \geq 1} P_i \leq 1$ and $0 \leq P_i < 1$ for all $i \geq 1$;

(iii) $x_i \neq x_j$ for all $i, j \geq 1$ with $i \neq j$, where the x_i's are enumerated in decreasing order of their mass P_i; indices are assigned according to increasing distance to 0 if atom masses have the same size.

Each string of beads $\xi \in \widetilde{\Xi}$ can be characterised via a constant $K > 0$ and a sequence of atoms $(x_i)_{i\geq 1}$ with respective masses $(P_i)_{i\geq 1}$ in decreasing order. This gives a natural bijection between $\widetilde{\Xi}$ and the space

$$\Xi := \left\{\left(K, (x_i)_{i\geq 1}, (P_i)_{i\geq 1}\right) : \text{(i)-(iii) hold}\right\} \subset [0, \infty) \times [0, \infty)^{\mathbb{N}} \times l^1(\mathbb{N}).$$

We will obtain a construction of a random compact weighted $\mathbb{R}$-tree $(\check{\mathfrak{T}}, \check{\mu})$ based on any string of beads ξ satisfying (i)-(iii) and an additional moment condition on its length K. $(\check{\mathfrak{T}}, \check{\mu})$ will be a CRT in the strict sense if and only if ξ satisfies the additional properties $K = \sup_{i\geq 1}\{x_i : P_i > 0\}$ and $\sum_{i\geq 1} P_i = 1$. When $\sum_{i\geq 1} P_i < 1$, we will still obtain convergence of our contruction to a random $\mathbb{R}$-tree in the Gromov-Hausdorff sense, but we do not obtain a CRT (as the total mass of the tree is strictly less than 1).

Henceforth, we work with the space Ξ, and define $\Xi^* := \Xi \times \mathbb{T}^\infty$ where $\mathbb{T}^\infty$ is the set of infinite sequences in $\mathbb{T}$. From now on, let $\beta \in (0, \infty)$ be fixed. We equip Ξ^* with the metric d_β where $d_\beta(\vartheta, \vartheta')$ is defined as

$$|K - K'| \vee \sup_{i\geq 1}\left(|x_i - x_i'| \vee \left|(P_i)^\beta - (P_i')^\beta\right| \vee d_{\mathrm{GH}}(\tau_i, \tau_i') \vee d_{\mathrm{GH}}\left((P_i)^\beta \tau_i, (P_i')^\beta \tau_i'\right)\right)$$

for $\vartheta = (K, (x_i)_{i\geq 1}, (P_i)_{i\geq 1}, (\tau_i)_{i\geq 1})$, $\vartheta' = (K', (x_i')_{i\geq 1}, (P_i')_{i\geq 1}, (\tau_i')_{i\geq 1}) \in \Xi^*$, where we recall that $P_i^\beta \tau$ denotes the $\mathbb{R}$-tree τ_i with distances scaled by P_i^β.

Proposition 2.30. *The space (Ξ^*, d_β) is separable.*

Proof. We first show that the space $([0, 1] \times \mathbb{T}, \tilde{d}_\beta)$ is separable, where

$$\tilde{d}_\beta((P, \tau), (P', \tau')) := \left|P^\beta - P'^\beta\right| \vee d_{\mathrm{GH}}(\tau, \tau') \vee d_{\mathrm{GH}}\left(P^\beta \tau, (P')^\beta \tau'\right). \qquad (2.37)$$

Recall that $\mathbb{T}$ equipped with the Gromov-Hausdorff distance d_{GH} and the space $[0, 1]$ equipped with the Euclidian distance are separable, i.e. there exist countable dense subsets $\mathbb{T}' \subset \mathbb{T}$, $\mathbb{Q} \cap [0, 1] \subset [0, 1]$ such that for any $\epsilon > 0$ and any $\tau \in \mathbb{T}$, $P \in [0, 1]$ there are $\tau' \in \mathbb{T}'$, $P' \in \mathbb{Q} \cap [0,1]$ such that

$$d_{\mathrm{GH}}(\tau, \tau') < \epsilon/2, \quad \left|P^\beta - (P')^\beta\right| < \epsilon/(2\mathrm{ht}(\tau)),$$

see e.g. [31] Theorem 4.23 for the separability of $(\mathbb{T}, d_{\mathrm{GH}})$. Clearly, the space $D := (\mathbb{Q} \cap [0,1]) \times \mathbb{T}'$ is countable. To prove that D is dense in $[0,1] \times \mathbb{T}$, it remains to show that $d_{\mathrm{GH}}(P^\beta \tau, (P')^\beta \tau') < \epsilon$. Indeed,

$$\begin{aligned} d_{\mathrm{GH}}\Big(P^\beta \tau, (P')^\beta \tau'\Big) &\leq d_{\mathrm{GH}}\Big(P^\beta \tau, (P')^\beta \tau\Big) + d_{\mathrm{GH}}\Big((P')^\beta \tau, (P')^\beta \tau'\Big) \\ &\leq \left|P^\beta - (P')^\beta\right| \mathrm{ht}(\tau) + (P')^\beta d_{\mathrm{GH}}(\tau, \tau') \\ &< (\epsilon/(2\mathrm{ht}(\tau)))\mathrm{ht}(\tau) + \epsilon/2 = \epsilon, \end{aligned}$$

where we applied the triangle inequality twice. Hence, $([0,1] \times \mathbb{T}, \tilde{d}_\beta)$ is separable. Note that $[0, \infty)$ equipped with the Euclidian distance is separable. Proposition 2.30 now follows from the fact that if $(A_1, d_1), (A_2, d_2), \ldots$ are separable metric spaces, then the space $((A_1 \times A_2 \times \ldots), d_\infty)$ is separable, where d_∞ is defined by $d_\infty((x_1, x_2, \ldots), (x_1', x_2', \ldots)) := \sup_{j \geq 1} d_j(x_j, x_j')$. □

Next, we consider

$$(K, (x_i)_{i \geq 1}, (P_i)_{i \geq 1}, (\tau_i, d_i, \rho_i)_{i \geq 1})$$

where $(K, (x_i)_{i \geq 1}, (P_i)_{i \geq 1}) \in \Xi$ and $(\tau_i, d_i, \rho_i)_{i \geq 1}$ is a sequence of $\mathbb{R}$-trees, that we do not consider as elements of $\mathbb{T}$.

For any $i \geq 1$, rescale the tree (τ_i, d_i, ρ_i) by P_i^β, i.e. consider the tree $P_i^\beta \tau_i := (\tau_i, P_i^\beta d_i, \rho_i)$, and attach $P_i^\beta \tau_i$ to the point x_i, replacing the related atom $P_i \delta_{x_i}$ of the string of beads ξ related to $(K, (x_i)_{i \geq 1}, (P_i)_{i \geq 1})$, i.e. identify the root ρ_i with the point x_i of mass P_i. More formally, define

$$(\tau', d', \rho') \tag{2.38}$$

by $\tau' := [0,K] \sqcup \bigsqcup_{i\geq 1} \tau_i \setminus \{\rho_i\}$ and the metric d' on τ' via

$$d'(x,y) := \begin{cases} |x-y| & \text{if } x,y \in [0,K], \\ P_j^\beta d_j(x,y) & \text{if } x,y \in \tau_j, j \geq 1, \\ |x-x_j| + P_j^\beta d_j(\rho_j, y) & \text{if } x \in [0,K], y \in \tau_j, j \geq 1, \\ P_{j_1}^\beta d_{j_1}(x,\rho_{j_1}) + |x_{j_1} - x_{j_2}| + P_{j_2}^\beta d_{j_2}(\rho_{j_2}, y) & \text{if } x \in \tau_{j_1}, y \in \tau_{j_2}, j_1 \neq j_2, \end{cases} \tag{2.39}$$

where we define the root by $\rho' := 0$. We only consider (τ', d', ρ') when compact. It is easy to see that the equivalence class of (τ', d', ρ') only depends on the equivalence classes of (τ_i, d_i, ρ_i), $i \geq 1$, hence we can define the subset $C_{\phi_\beta} \subset \Xi^*$ as the set of all elements $(K, (x_i)_{i\geq 1}, (P_i)_{i\geq 1}, (\tau_i)_{i\geq 1}) \in \Xi^*$ such that (τ', d', ρ') is compact for any representative associated with $(K, (x_i)_{i\geq 1}, (P_i)_{i\geq 1}, (\tau_i)_{i\geq 1})$, and the map $\phi_\beta \colon \Xi^* \to \mathbb{T}$,

$$\vartheta := (K, (x_i)_{i\geq 1}, (P_i)_{i\geq 1}, (\tau_i)_{i\geq 1}) \mapsto \phi_\beta(K, (x_i)_{i\geq 1}, (P_i)_{i\geq 1}, (\tau_i)_{i\geq 1})$$

mapping ϑ onto the equivalence class associated with (τ', d', ρ') if $\vartheta \in C_{\phi_\beta}$, and onto $\{\rho\}$ otherwise.

As we have a natural bijection between $\widetilde{\Xi}$ and Ξ, we will also write $\phi_\beta(\xi, (\tau_i, i \geq 1))$ for any $\xi = ([0,K], \sum_{i\geq 1} P_i \delta_{x_i}) \in \widetilde{\Xi}$ to mean $\phi_\beta(K, (x_i)_{i\geq 1}, (P_i)_{i\geq 1}, (\tau_i, i \geq 1))$.

Lemma 2.31. *The map $\phi_\beta : \Xi^* \to \mathbb{T}$ is measurable.*

Proof. We first show that $\phi_\beta \restriction_{C_{\phi_\beta}}$ is continuous.

Let $(\vartheta^{(n)})_{n\geq 1}$ be a sequence in C_{ϕ_β} such that $\lim_{n\to\infty} d_\beta(\vartheta^{(n)}, \vartheta) = 0$ for some $\vartheta \in C_{\phi_\beta}$. Then, using the notation

$$\vartheta^{(n)} = \left(K^{(n)}, \left(x_i^{(n)}\right)_{i\geq 1}, \left(P_i^{(n)}\right)_{i\geq 1}, \left(\tau_i^{(n)}\right)_{i\geq 1}\right), \quad n \geq 1,$$

and $\vartheta = (K, (x_i)_{i\geq 1}, (P_i)_{i\geq 1}, (\tau_i)_{i\geq 1})$, we obtain

$$\begin{aligned} d_{\mathrm{GH}}\left(\phi_\beta\left(\vartheta^{(n)}\right), \phi_\beta(\vartheta)\right) &\leq \left|K^{(n)} - K\right| + \sup_{i\geq 1}\left|x_i^{(n)} - x_i\right| + \sup_{i\geq 1} d_{\mathrm{GH}}\left(\left(P_i^{(n)}\right)^\beta \tau_i^{(n)}, (P_i)^\beta \tau_i\right) \\ &\leq 3 d_\beta\left(\left(\xi^{(n)}, \tau^{(n)}\right), (\xi, \tau)\right), \end{aligned}$$

i.e. $\lim_{n\to\infty} d_{\mathrm{GH}}(\phi_\beta(\vartheta^{(n)}), \phi_\beta(\vartheta)) = 0$, and hence, $\phi_\beta \restriction_{C_{\phi_\beta}}$ is continuous.

The measurability of ϕ_β will follow from Lemma 2.33 and Lemma 2.34. $\square$

The proof of Lemma 2.33 is based on the following proposition, whose proof we provide for completeness.

Proposition 2.32. *Let $(A, \mathcal{A})$ and $(B, \mathcal{B})$ be two separable metrisable Borel spaces, and let $f : A \to B$ be continuous. Then f is measurable.*

Proof. Consider the collection of sets

$$\mathcal{G} := \Big\{ B_0 \in \mathcal{B} : f^{-1}(B_0) \in \mathcal{A} \Big\}.$$

By continuity, the pre-image $f^{-1}(B_0)$ of any open set $B_0 \in \mathcal{B}$ is open, and hence measurable, i.e. $B_0 \in \mathcal{G}$ for all open sets $B_0 \in \mathcal{B}$. Furthermore, by properties of the pre-image function f^{-1}, the collection of sets $\mathcal{G}$ is a σ-algebra. Since the Borel σ-algebra $\mathcal{B}$ on B is generated by all open sets, we conclude that $\mathcal{B} = \mathcal{G}$, i.e. f is measurable. $\square$

Lemma 2.33. *Let $(A, \mathcal{A})$ and $(B, \mathcal{B})$ be two separable metrisable Borel spaces, and let $A_0 \in \mathcal{A}$. Then any function $f : A \to B$ with $f(a) \equiv b_0$ for all $a \in A \setminus A_0$ and $f \restriction_{A_0}$ continuous is measurable.*

Proof. For every $A_0 \in \mathcal{A}$, the subset topology on A_0 is separable and induced by the metric restricted to A_0. By Proposition 2.32, $f_0 = f \restriction_{A_0}$ is measurable. We need to show that $f^{-1}(B_0) \in \mathcal{A}$ for any $B_0 \in \mathcal{B}$. We have that

$$f^{-1}(B_0) = \begin{cases} \Big(f^{-1}(b_0) \setminus f_0^{-1}(b_0)\Big) \cup f_0^{-1}(B_0) & \text{if } b_0 \in B_0, \\ f_0^{-1}(B_0) & \text{if } b_0 \notin B_0. \end{cases}$$

Since f_0 is measurable, we obtain $f_0^{-1}(B_0) \in \mathcal{A}$. Furthermore, $f^{-1}(b_0) \setminus f_0^{-1}(b_0) = A \setminus A_0 \in \mathcal{A}$ since $A, A_0 \in \mathcal{A}$. Thus, $f^{-1}(B_0) \in \mathcal{A}$ for all $B_0 \in \mathcal{B}$, i.e. f is measurable. $\square$

Now take $A = \Xi^*$ equipped with its Borel σ-algebra $\mathcal{A} = \mathcal{B}(\Xi^*)$, and similarly $B = \mathbb{T}$, $\mathcal{B} = \mathcal{B}(\mathbb{T})$. Furthermore, consider $A_0 = C_{\phi_\beta}$, $f = \phi_\beta$, $b_0 = \{\rho\}$. We already know that $\phi_\beta \restriction_{C_{\phi_\beta}}$ is continuous. It remains to show the following.

Lemma 2.34. *The set $C_{\phi_\beta} \subset \Xi^*$ is measurable, i.e. $C_{\phi_\beta} \in \mathcal{B}(\Xi^*)$.*

Proof. For any $(K,(x_i)_{i\geq 1},(P_i)_{i\geq 1},(\tau_i)_{i\geq 1}) \in \Xi^*$, $\phi_\beta(K,(x_i)_{i\geq 1},(P_i)_{i\geq 1},(\tau_i)_{i\geq 1})$ is compact if and only if for all $N \geq 1$, we can find $I \geq 1$ such that for all $i \geq I$, $\mathrm{ht}(P_i^\beta \tau_i) < 1/N$, i.e.

$$C_{\tilde{\phi}_\beta} = \bigcap_{N\geq 1}\bigcup_{I\geq 1}\bigcap_{i\geq I}\{(K,(x_i)_{i\geq 1},(P_i)_{i\geq 1},(\tau_i)_{i\geq 1}) \in \Xi^* : P_i^\beta \mathrm{ht}(\tau_i) < 1/N\}.$$

Since the height function $\mathrm{ht}\colon \mathbb{T} \to [0,\infty)$ is measurable, C_{ϕ_β} is measurable. □

Consider the space of probability measures $\mathcal{P}(\mathbb{T})$ on the space $\mathbb{T}$ of equivalence classes of compact $\mathbb{R}$-trees, and the subsets $\mathcal{P}_p \subset \mathcal{P}(\mathbb{T})$, $p > 0$, given by

$$\mathcal{P}_p := \{\eta : \eta \text{ is a probability measure on } \mathbb{T} \text{ and } \mathbb{E}[\mathrm{ht}(\tau)^p] < \infty \text{ for } \tau \sim \eta\}. \tag{2.40}$$

We follow [13], and equip $\mathcal{P}_p$ with the *Wasserstein metric* of order $p \geq 1$, which is defined by

$$W_p(\eta,\eta') := \left(\inf \mathbb{E}\left[|d_{\mathrm{GH}}(\tau,\tau')|^p\right]\right)^{1/p}, \qquad \eta,\eta' \in \mathcal{P}(\mathbb{T}), \quad \tau \sim \eta, \tau' \sim \eta', \tag{2.41}$$

provided that $E[|d_{\mathrm{GH}}(\tau,\tau')|^p] < \infty$, where the infimum is taken over all couplings of random variables τ and τ' with marginal distributions η and η', respectively. The space $(\mathcal{P}_p, W_p)$ is complete since d_{GH} is a complete metric on $\mathbb{T}$, see e.g. [17, 32].

2.4.3 Existence of a fixpoint distribution

We consider the recursive tree framework induced by some string of beads $\xi \in \tilde{\Xi}$ with distribution ν and an i.i.d. sequence of compact $\mathbb{R}$-trees $(\tau_i)_{i\geq 1} = (\tau_i, d_i, \rho_i)_{i\geq 1}$ with some distribution $\eta \in \mathcal{P}_p$. We have $N = \infty$ throughout this section, and, with a slight abuse of notation, we write $(\xi, N) \sim \nu$.

Lemma 2.35 (Contraction)**.** *Let $\beta > 0$, $p > 1/\beta$ and $\xi = ([0,K], \sum_{i\geq 1} P_i \delta_{x_i}) \in \tilde{\Xi}$ with some distribution ν be such that $\mathbb{E}[K^p] < \infty$. The map $\Phi_\beta\colon \mathcal{P}_p \to \mathcal{P}_p$ related to ξ and $\phi_\beta : \Xi^* \to \mathbb{T}$ is a strict contraction with respect to the p-th Wasserstein distance, i.e.*

$$\sup_{\eta,\eta' \in \mathcal{P}_p, \eta \neq \eta'} \frac{W_p(\Phi_\beta(\eta), \Phi_\beta(\eta'))}{W_p(\eta,\eta')} < 1. \tag{2.42}$$

Proof. Let $\tau \sim \eta$ and $\tau' \sim \eta'$ be two elements of $\mathbb{T}$. Furthermore, let $(\tau_i)_{i\geq 1}$ and $(\tau_i')_{i\geq 1}$ be two i.i.d. sequences with $\tau_1 \sim \eta$ and $\tau_1' \sim \eta'$, independent of ξ. We first check that $\Phi_\beta(\eta) \in \mathcal{P}_p$. Indeed,

$$\begin{aligned}\mathbb{E}[\mathrm{ht}(\phi_\beta(\xi,(\tau_i,i\geq 1)))^p] &\leq \mathbb{E}\left[\left(K + \sup_{i\geq 1} P_i^\beta \mathrm{ht}(\tau_i)\right)^p\right] \\ &\leq 2^p\mathbb{E}[K^p] + 2^p\mathbb{E}\left[\sup_{i\geq 1} P_i^{p\beta}\mathrm{ht}(\tau_i)^p\right] < \infty\end{aligned}$$

since the first term is finite by assumption, and the second term can further be estimated above using

$$\mathbb{E}\left[\sup_{i\geq 1} P_i^{p\beta}\mathrm{ht}(\tau_i)^p\right] \leq \mathbb{E}\left[\sum_{i\geq 1} P_i^{p\beta}\mathrm{ht}(\tau_i)^p\right] = \left(\sum_{i\geq 1}\mathbb{E}\Big[P_i^{p\beta}\Big]\right)\mathbb{E}[\mathrm{ht}(\tau_1)^p]$$

which is finite since $\eta \in \mathcal{P}_p$, and since $p\beta > 1$, $\sum_{i\geq 1} P_i \leq 1$, and $P_i \in [0,1)$ for all $i \geq 1$ by the assumption that $\xi \in \widetilde{\Xi}$. Then

$$\begin{aligned}W_p^p(\Phi_\beta(\eta),\Phi_\beta(\eta')) &\leq \inf \mathbb{E}\big[(d_{\mathrm{GH}}(\phi_\beta(\xi,\tau_i,i\geq 1),\phi_\beta(\xi,\tau_i',i\geq 1)))^p\big] \\ &\leq \inf \mathbb{E}\left[\sup_{i\geq 1}\Big(d_{\mathrm{GH}}\Big(P_i^\beta\tau_i, P_i^\beta\tau_i'\Big)\Big)^p\right] \\ &\leq \inf \mathbb{E}\left[\sum_{i=1}^\infty d_{\mathrm{GH}}\Big(P_i^\beta\tau_i, P_i^\beta\tau_i'\Big)^p\right] \\ &= \inf \mathbb{E}\left[\sum_{i=1}^\infty (P_i)^{p\beta} d_{\mathrm{GH}}(\tau_i,\tau_i')^p\right] \\ &= \inf \sum_{i=1}^\infty \mathbb{E}\Big[(P_i)^{p\beta}\Big]\mathbb{E}\big[d_{\mathrm{GH}}(\tau_i,\tau_i')^p\big] \\ &= \left(\sum_{i=1}^\infty \mathbb{E}\Big[(P_i)^{p\beta}\Big]\right)\big(\inf \mathbb{E}\big[d_{\mathrm{GH}}(\tau_i,\tau_i')^p\big]\big) \\ &= \left(\sum_{i=1}^\infty \mathbb{E}\Big[(P_i)^{p\beta}\Big]\right)W_p^p(\eta,\eta') \end{aligned} \tag{2.43}$$

where the infimum is taken over all couplings of $(\tau_i)_{i\geq 1}$ and $(\tau_i')_{i\geq 1}$.

Dividing by $W_p^p(\eta,\eta')$, taking the $1/p$-th power and the supremum over all $\eta,\eta' \in \mathcal{P}_p$, $\eta \neq \eta'$, we obtain, for any $p \geq 1$,

$$\sup_{\eta,\eta'\in\mathcal{P}_p,\eta\neq\eta'} \frac{W_p(\Phi(\eta),\Phi(\eta'))}{W_p(\eta,\eta')} \leq \left(\sum_{i=1}^\infty \mathbb{E}\Big[(P_i)^{p\beta}\Big]\right)^{1/p}. \tag{2.44}$$

By assumption, $p\beta > 1$. Then, since $0 < \sum_{i\geq 1} P_i \leq 1$ and $0 \leq P_i < 1$ for all $i \geq 1$, we have

$$\sum_{i=1}^{\infty} \mathbb{E}\Big[(P_i)^{p\beta}\Big] < \sum_{i=1}^{\infty} \mathbb{E}[P_i] = \mathbb{E}\left[\sum_{i=1}^{\infty} P_i\right] \leq 1. \tag{2.45}$$

□

Corollary 2.36 (Fixpoint)**.** *In the setting of Lemma 2.35, the map $\Phi_\beta : \mathcal{P}_p \to \mathcal{P}_p$ has a unique fixpoint $\eta^* \in \mathcal{P}_p$, i.e. $\Phi_\beta(\eta^*) = \eta^*$ and $\Phi_\beta^n(\eta) \to \eta^*$ as $n \to \infty$ for all $\eta \in \mathcal{P}_p$ in the sense of weak convergence.*

Proof. Recall that W_p is a complete metric on $\mathcal{P}_p$ for any $1 \leq p < \infty$. Furthermore, Φ is a strict contraction by Lemma 2.35. Hence, we conclude by Lemma 2.29 that $\Phi_\beta\colon \mathcal{P}_p \to \mathcal{P}_p$ has a unique fixpoint $\eta^* \in \mathcal{P}_p$ such that $\Phi_\beta^n(\eta) \to \eta^*$ as $n \to \infty$ for all $\eta \in \mathcal{P}_p$. □

This yields a distribution η^* on $\mathbb{T}$, and a recursive construction of an $\mathbb{R}$-tree $\check{\mathcal{T}} \sim \eta^*$ based on a string of beads $\xi \in \widetilde{\Xi}$ with distribution ν, cf. [13].

Corollary 2.37. *In the setting of Lemma 2.35 and Corollary 2.36 the height of $\check{\mathcal{T}}$ has finite moments $\mathbb{E}[\mathrm{ht}(\check{\mathcal{T}})^p] < \infty$ for all $1 \leq p < p^* := \sup\{p \geq 1 : \mathbb{E}[K^p] < \infty\}$.*

Proof. Note that, as a direct consequence of Lemma 2.35,

$$\eta^* \in \bigcap_{1/\beta < p < p^*} \mathcal{P}_p = \bigcap_{1 \leq p < p^*} \mathcal{P}_p.$$

□

We write

$$\xi_{\mathbf{i}} = \left(E_{\mathbf{i}}, \sum_{j\geq 1} P_{\mathbf{i}j}\delta_{x_{\mathbf{i}j}}\right), \quad \mathbf{i} \in \mathcal{N}, \tag{2.46}$$

for isometric copies of i.i.d. strings of beads with distribution ν, where $E_{\mathbf{i}}, \mathbf{i} \in \mathcal{N}$, are disjoint intervals.

We will slightly abuse notation by writing $\xi_{\mathbf{i}} \subset \mathcal{T}$ to mean that $E_{\mathbf{i}} \subset \mathcal{T}$ for some set $\mathcal{T}$. In Corollary 2.38, we construct a sequence of trees $(\check{\mathcal{T}}_n, n \geq 0)$ by successively replacing the atoms of the mass measure $\check{\mu}_n$ on $\check{\mathcal{T}}_n$ by rescaled independent copies of the initial string of beads $\xi_\varnothing = (\check{\mathcal{T}}_0, \check{\mu}_0)$ (simultaneously for all atoms at the same depth $n \geq 0$). The almost-sure limit $\check{\mathfrak{T}}$ of this sequence is a tree with the fixpoint distribution determined by the underlying recursive tree framework $(\xi_{\mathbf{i}}, \mathbf{i} \in \mathcal{N})$. Note that, by construction, $\check{\mathfrak{T}} = \overline{\bigcup_{n\geq 0} \check{\mathfrak{T}}_n} = \overline{\bigcup_{\mathbf{i}\in\mathcal{N}} E_{\mathbf{i}}}$.

The following result is a restatement of Theorem 2.1 from the introduction.

Corollary 2.38 (Recursive construction of CRTs)**.** *We construct an increasing sequence of weighted $\mathbb{R}$-trees $(\check{\mathfrak{T}}_n, \check{\mu}_n, n \geq 0)$ as follows.*

Let $\xi_{\mathbf{i}}, \mathbf{i} \in \mathcal{N}$, be isometric copies of i.i.d. strings of beads with distribution ν, and set $(\check{\mathfrak{T}}_0, \check{\mu}_0) = \xi_\varnothing$. For $n \geq 1$, set $\check{\mathfrak{T}}_n := \phi_\beta(\xi_\varnothing, (\tau_i^{(n)}, i \geq 1))$ where $(\tau_{\mathbf{i}}^{(n)}, \mathbf{i} \in \bigcup_{k=1}^n \mathbb{N}^k)$ is a sequence of trees defined recursively via $\tau_{\mathbf{i}}^{(1)} = \xi_{\mathbf{i}}$, $\mathbf{i} \in \mathbb{N}$, and, for $n \geq 2$,

$$\tau_{\mathbf{i}}^{(n)} := \begin{cases} [0, K_i], & \mathbf{i} \in \mathbb{N}^n, \\ \phi\Big(\xi_{\mathbf{i}}, \Big(\tau_{\mathbf{i}j}^{(n)}, j \geq 1\Big)\Big) & \mathbf{i} \in \mathbb{N}^m, 1 \leq m \leq n-1. \end{cases}$$

We equip $\check{\mathfrak{T}}_n$ with the mass measure $\check{\mu}_n$ given by

$$\check{\mu}_n = \sum_{\mathbf{i}=i_1 i_2 \ldots i_{n+1} \in \mathbb{N}^{n+1}} P_{i_1} P_{i_1 i_2} \cdots P_{i_1 i_2 \ldots i_{n+1}} \delta_{x_{\mathbf{i}}} =: \sum_{\mathbf{i}=i_1 i_2 \ldots i_{n+1} \in \mathbb{N}^{n+1}} \check{P}_{\mathbf{i}} \delta_{x_{\mathbf{i}}}, \quad n \geq 0. \tag{2.47}$$

Then we have the convergence

$$\lim_{n\to\infty} d_{\mathrm{GH}}\Big(\check{\mathfrak{T}}_n, \check{\mathfrak{T}}\Big) = 0 \qquad a.s. \tag{2.48}$$

where $\check{\mathfrak{T}} := \overline{\bigcup_{n\geq 0} \check{\mathfrak{T}}_n}$ has the fixpoint distribution η^* related to the RTF $(\xi_{\mathbf{i}}, \mathbf{i} \in \mathcal{N})$.

Proof. The sequence $(\check{\mathfrak{T}}_n, n \geq 0)$ and the tree $\check{\mathfrak{T}}$ are naturally coupled via the underlying recursive tree framework $(\xi_{\mathbf{i}}, \mathbf{i} \in \mathcal{N})$. More precisely, $\check{\mathfrak{T}}_n \subset \check{\mathfrak{T}}_{n+1} \subset \check{\mathfrak{T}}$ for all $n \geq 0$.

The tree $\check{\mathfrak{T}}$ has the fixpoint distribution η^* given in Corollary 2.36, i.e. $\check{\mathfrak{T}}$ admits the representation

$$\check{\mathfrak{T}} = \phi_\beta(\xi_\varnothing, (\mathfrak{T}_i, i \geq 1)) \tag{2.49}$$

where $\mathcal{T}_{\mathbf{i}}, \mathbf{i} \in \mathcal{N}$, are i.i.d. copies of $\check{\mathcal{T}}$, and $\mathcal{T}_{\mathbf{i}} = \phi_\beta(\xi_{\mathbf{i}}, (\mathcal{T}_{\mathbf{i}j}, j \geq 1)), \mathbf{i} \in \mathcal{N}$. For any $p \geq 1$, we have

$$\mathbb{E}\Big[d_{\mathrm{GH}}\Big(\check{\mathcal{T}}_n, \check{\mathcal{T}}\Big)^p\Big] \leq \mathbb{E}\Bigg[\Bigg(\sup_{\mathbf{i}\in\mathbb{N}^n, j\geq 1} \check{P}_{\mathbf{i}j}^{\beta}\mathrm{ht}(\mathcal{T}_{\mathbf{i}j})\Bigg)^p\Bigg] \leq \sum_{\mathbf{i}\in\mathbb{N}^n} \mathbb{E}\Bigg[\sum_{j\geq 1} \check{P}_{\mathbf{i}j}^{p\beta}\Bigg]\mathbb{E}\Big[\mathrm{ht}\Big(\check{\mathcal{T}}\Big)^p\Big].$$

Note that, for each $\mathbf{i} \in \mathcal{N}$, the sequence $(\check{P}_{\mathbf{i}j}, j \geq 1)$ can be represented as

$$\Big(\check{P}_{\mathbf{i}j}, j \geq 1\Big) = \Big(\check{P}_{\mathbf{i}}(P_{\mathbf{i}j}, j \geq 1)\Big) \tag{2.50}$$

where $(P_{\mathbf{i}j}, j \geq 1)$ is independent of $\check{P}_{\mathbf{i}}$, $\mathbf{i} \in \mathbb{N}^n$, and has the same distribution as the ranked atom masses $(P_j, j \geq 1)$ of the initial string of beads $\xi_\varnothing$. Therefore,

$$\mathbb{E}\Big[d_{\mathrm{GH}}\Big(\check{\mathcal{T}}_n, \check{\mathcal{T}}\Big)^p\Big] \leq \mathbb{E}\Big[\mathrm{ht}(\check{\mathcal{T}})^p\Big]\mathbb{E}\Bigg[\sum_{j\geq 1} P_j^{p\beta}\Bigg]\mathbb{E}\Bigg[\sum_{\mathbf{i}\in\mathbb{N}^n} \check{P}_{\mathbf{i}}^{p\beta}\Bigg].$$

Applying (2.50) recursively, we obtain

$$\mathbb{E}\Big[d_{\mathrm{GH}}\Big(\check{\mathcal{T}}_n, \check{\mathcal{T}}\Big)^p\Big] \leq \mathbb{E}\Big[\mathrm{ht}\Big(\check{\mathcal{T}}\Big)^p\Big]\Bigg(\mathbb{E}\Bigg[\sum_{j\geq 1} P_j^{p\beta}\Bigg]\Bigg)^n.$$

By the Markov inequality, we have

$$\mathbb{P}\Big(d_{\mathrm{GH}}\Big(\check{\mathcal{T}}_n, \check{\mathcal{T}}\Big) > \epsilon\Big) \leq \epsilon^{-p}\mathbb{E}\Big[d_{\mathrm{GH}}\Big(\check{\mathcal{T}}_n, \check{\mathcal{T}}\Big)^p\Big] \leq \mathbb{E}\Big[\mathrm{ht}\Big(\check{\mathcal{T}}\Big)^p\Big]\epsilon^{-p}\Bigg(\mathbb{E}\Bigg[\sum_{j\geq 1} P_j^{p\beta}\Bigg]\Bigg)^n. \tag{2.51}$$

We choose p such that $p\beta > 1$ to obtain $\mathbb{E}\Big[\sum_{j\geq 1} P_j^{p\beta}\Big] < \mathbb{E}\Big[\sum_{j\geq 1} P_j\Big] \leq 1$ as in (2.45). By Corollary 2.37 and the first Borel-Cantelli Lemma, we conclude

$$\mathbb{P}\Big(d_{\mathrm{GH}}\Big(\check{\mathcal{T}}_n, \check{\mathcal{T}}\Big) > \epsilon \text{ for infinitely many } n\Big) = 0, \tag{2.52}$$

and hence $\lim_{n\to\infty} \check{\mathcal{T}}_n = \check{\mathcal{T}}$ a.s. in the Gromov-Hausdorff topology. $\square$

If the initial string of beads satisfies some extra conditions, we obtain a mass measure $\check{\mu}$ as the a.s. limit of $\check{\mu}_n$ as $n \to \infty$ such that $(\check{\mathcal{T}}, \check{\mu})$ is a CRT in the strict sense.

Proposition 2.39 (Mass measure on $\check{\mathcal{T}}$)**.** *Consider the setting of Corollary 2.38. If* $\sum_{j\geq 1} P_j = 1$ *and* $K = \sup_{i\geq 1}\{x_i : P_i > 0\}$ *a.s., there exists a continuum random tree* $(\check{\mathcal{T}}, \check{\mu})$ *such that*

$$\lim_{n\to\infty} d_{\mathrm{GHP}}\Big(\Big(\check{\mathcal{T}}_n, \check{\mu}_n\Big), \Big(\check{\mathcal{T}}, \check{\mu}\Big)\Big) = 0 \quad a.s.. \tag{2.53}$$

Proof. By Corollary 2.38, it remains to show that the mass measures $\check{\mu}_n$ converge a.s.. We follow Aldous' formalism of trees embedded in $l^1(\mathbb{N})$, cf. [7, 8, 9]. For any $x \in \check{\mathcal{T}}_n$ there exists a unique index $\mathbf{i} \in \mathcal{N}$ such that $x \in \xi_{\mathbf{i}}$ (excluding left endpoints of the strings of beads $\xi_{\mathbf{i}}$, $\mathbf{i} \in \mathcal{N} \setminus \{\varnothing\}$). $\mathcal{N}$ is countable, i.e. there is a bijection $\sigma : \mathcal{N} \to \mathbb{N}$, and hence we can embed the branches of the trees $\check{\mathcal{T}}_n$ into $l^1(\mathbb{N})$ (instead of embedding them into $l^1(\mathcal{N})$). This embedding is consistent as n varies, as, for each $n \geq 0$, $\check{\mathcal{T}}_{n+1} \setminus \check{\mathcal{T}}_n = \bigcup_{\mathbf{i}\in\mathbb{N}^n} \xi_{\mathbf{i}}$. The mass measure $\check{\mu}_n$ can hence be viewed as a measure on $l^1(\mathbb{N})$, and we need to show that the $\check{\mu}_n$ converge a.s., as random probability measures on $l^1(\mathbb{N})$. This is easier than Aldous' approximation of the measure representation of the Brownian CRT ([7], Theorem 3) but the formalism and the key steps are the same.

Firstly, mass is preserved from $\check{\mu}_n$ to $\check{\mu}_{n+1}$, and, for $n \geq k \geq 0$, the projection of $\check{\mu}_n$ onto $\check{\mathcal{T}}_k$ is μ_k and does not depend on $n \geq k$. Hence, we obtain a.s. convergence to a consistent system of finite-dimensional marginals of a random probability measure $\check{\mu}$ on $[0,\infty)^{\mathbb{N}}$. Secondly, the sequence $(\check{\mu}_n)$ is (a.s.) tight as a family of (random) probability measures on $l^1(\mathbb{N})$ since $\mathcal{T}$ (represented in $l^1(\mathbb{N})$) is compact a.s.

Hence $\check{\mu} := \lim_{n\to\infty} \check{\mu}$ is a probability measure on $l^1(\mathbb{N})$. It remains to check that $(\check{\mathcal{T}}, \check{\mu})$ is a CRT, i.e. that $\check{\mu}$ a.s. satisfies the properties of a continuum tree from Section 2.3.1. Note that $\max_{x\in\check{\mathcal{T}}_n} \check{\mu}_n(x) = \max_{\mathbf{i}=i_1\cdots i_{n+1}\in\mathbb{N}^{n+1}} \check{P}_{\mathbf{i}}$, and that for $\epsilon > 0$ and $p > 1$,

$$\begin{aligned}\sum_{n\geq 0} \mathbb{P}\left(\max_{\mathbf{i}=i_1\cdots i_{n+1}\in\mathbb{N}^{n+1}} \check{P}_{\mathbf{i}} > \epsilon\right) &\leq \sum_{n\geq 0} \epsilon^{-p}\mathbb{E}\left[\left(\max_{\mathbf{i}=i_1\cdots i_{n+1}\in\mathbb{N}^{n+1}} \check{P}_{\mathbf{i}}\right)^p\right] \\ &\leq \sum_{n\geq 0} \epsilon^{-p}\mathbb{E}\left[\sum_{\mathbf{i}=i_1\cdots i_{n+1}\in\mathbb{N}^{n+1}} \check{P}_{\mathbf{i}}^p\right] \\ &\leq \sum_{n\geq 0} \epsilon^{-p}\left(\mathbb{E}\left[\sum_{j\geq 1} P_j^p\right]\right)^{n+1} < \infty,\end{aligned}$$

where the second last inequality can be derived as in (2.50), and the last inequality follows from the fact that $\mathbb{E}[\sum_{j\geq 1} P_j^p] < 1$ for $p > 1$. Hence, $\lim_{n\to\infty} \max_{\mathbf{i}\in\mathbb{N}^n} \check{P}_{\mathbf{i}} = 0$

a.s.. Note that $\check{\mu}_n(E_{\mathbf{i}}) = 0$ for any $\mathbf{i} \in \bigcup_{1\le m\le n-1} \mathbb{N}^m$, i.e. for any $x \in \check{\mathcal{T}}$,

$$\check{\mu}([[\rho, x]]) = \lim_{n\to\infty} \check{\mu}_n([[\rho, x]]) \le \lim_{n\to\infty} \max_{\mathbf{i}\in\mathbb{N}^n} \check{P}_{\mathbf{i}} = 0 \quad \text{a.s..} \tag{2.54}$$

Properties (i) and (ii) of a continuum tree now follow from (2.54). To see (iii), note that, for $x \in \check{\mathcal{T}} \setminus \mathrm{Lf}(\check{\mathcal{T}})$, there is $\mathbf{i} = i_1 \cdots i_n \in \mathcal{N}$ for some $n \ge 0$ with $x_{\mathbf{i}} \in \mathcal{T}_x$, i.e. $\xi_{\mathbf{i}} \subset \mathcal{T}_x$. Hence, $\check{\mu}(\mathcal{T}_x) \ge \check{\mu}_n(\xi_i) = \check{P}_{\mathbf{i}} > 0$. □

We now prove Corollary 2.2, using an embedding of $(\mathcal{T}_k, \mu_k)$, $k \ge 0$, into $(\check{\mathcal{T}}, \check{\mu})$.

Proof of Corollary 2.2. Consider $(\check{\mathcal{T}}_n, \mu_n)$, $n \ge 0$, built from $\xi_{\mathbf{i}}$, $i \in \mathcal{N}$, as in Corollary 2.38, with limit $\check{\mathcal{T}}$, and mass measure $\check{\mu}$ as in Proposition 2.39. Now let $(\overline{\mathcal{T}}_0, \overline{\mu}_0) = \xi_\varnothing$ and given $(\overline{\mathcal{T}}_j, \overline{\mu}_j)$, $0 \le j \le k$, with $\overline{\mathcal{T}}_k \subseteq \check{\mathcal{T}}$ and $\overline{\mu}_k$ the push-forward of $\check{\mu}_k$ under the natural projection from $\check{\mathcal{T}}$ to $\overline{\mathcal{T}}_k$, pick $\overline{J}_k$ from $\overline{\mu}_k$. If $\overline{J}_k = x_{\mathbf{i}j}$, remove $\overline{\mu}_k(\overline{J}_k)\delta_{\overline{J}_k}$ from $\overline{\mu}_k$ and add to $\overline{\mathcal{T}}_k$ the rescaled string of beads $\xi_{\mathbf{i}j} \subset \check{\mathcal{T}}$ to form $(\overline{\mathcal{T}}_{k+1}, \overline{\mu}_{k+1})$.

Then $(\overline{\mathcal{T}}_k, \overline{\mu}_k, k \ge 0)$ has the same distribution as $(\mathcal{T}_k, \mu_k, k \ge 0)$. Since $\overline{\bigcup_{k\ge 0} \overline{\mathcal{T}}_k} \subseteq \check{\mathcal{T}}$ is compact, it remains to show that the inclusion is actually an equality. To see this, we employ a simpler variant of an argument of [60] Proposition 22. Let $\epsilon > 0$ and consider the connected components of

$$\left\{\sigma \in \check{\mathcal{T}} \colon \mathrm{ht}\left(\check{\mathcal{T}}_\sigma\right) \le \epsilon\right\}$$

where $\check{\mathcal{T}}_\sigma = \{x \in \check{\mathcal{T}} : \sigma \in [[\rho, x]]\}$. Since $\check{\mathcal{T}}$ is compact, only finitely many attain height ϵ. Each of these intersects $\xi_{\mathbf{i}}$ for some $\mathbf{i} = i_1 i_2 \cdots i_n \in \mathcal{N}$, and each $x_{i_1\cdots i_j}$, $1 \le j \le n$, is picked as some $\overline{J}_{k_j}$ where $k_j := \inf\{k \ge 1 : \overline{J}_k = x_{i_1\cdots i_j}\}$ after a geometric number of steps with parameter $\overline{\mu}_{k_{j-1}}(x_{i_1\cdots i_j})$,$1 \le j \le n$, and the tree $\overline{\mathcal{T}}_{k_d+1}$ will intersect the component. When all components of height ϵ have been intersected the Gromov-Hausdorff-Prohorov distance from $(\check{\mathcal{T}}, \check{\mu})$ is below ε. This completes the proof. □

2.5 Binary embedding of the stable line-breaking construction

In the previous section, we obtained a tree $\check{\mathfrak{T}}$ via a recursive construction using isometric copies of i.i.d. strings of beads with some distribution ν on the space $\widetilde{\Xi}$. We are now interested in the construction induced by the (1-marked) string of beads (I, I_1, λ) defined as follows. Let $\beta \in (0, 1/2]$, and consider a function

$$\Psi_\beta : \widetilde{\Xi} \times \widetilde{\Xi} \times [0,1] \to \widetilde{\Xi}^{[1]}, \tag{2.55}$$

where $\widetilde{\Xi}^{[1]}$ is the space of 1-marked strings of beads given by

$$\widetilde{\Xi}^{[1]} = \Big\{ (I = [a,b], I' = [a',b'], \lambda) : (I, \lambda) \in \widetilde{\Xi}, I' \subset I \Big\}. \tag{2.56}$$

For $(I_1 = [0, K_1], \lambda_1), (I_2 = [0, K_2], \lambda_2) \in \widetilde{\Xi}$ and $B \in [0,1]$, define

$$(I = [0,K], I' = [0,K'], \lambda) := \Psi_\beta((I_1, \lambda_1), (I_2, \lambda_2), B) \tag{2.57}$$

via

$$K := B^\beta K_1 + (1-B)^\beta K_2, \quad K' := B^\beta K_1$$

and the mass measure λ on $[0,K]$ given by

$$\lambda([0,x]) = \begin{cases} B\lambda_1\Big(\Big[0, B^{-\beta} x\Big]\Big), & \text{if } x \in [0, K'], \\ B + (1-B)\lambda_2\Big(\Big[0, (1-B)^{-\beta}(x - K')\Big]\Big), & \text{if } x \in [K', K]. \end{cases}$$

The string of beads obtained via (2.57) when (I_1, λ_1) and (I_2, λ_2) are independent $(\beta, 1-2\beta)$- and (β,β)-strings of beads, respectively, and $B \sim \mathrm{Beta}(1-2\beta, \beta)$ is independent is called a *β-mixed* string of beads. Recall from Section 2.2.2 that for any (α, θ)-string of beads $([0,K], \lambda)$, $\alpha \in (0,1)$ and $\theta > -\alpha$, we have that $K \sim \mathrm{ML}(\alpha, \theta)$, and

$$(\lambda(x) : x \in [0,K], \lambda(x) > 0)^\downarrow \sim \mathrm{PD}(\alpha, \theta). \tag{2.58}$$

Hence, any (α, θ)-string of beads a.s. satisfies properties (i)-(iii) of $\widetilde{\Xi}$ as in (2.36). Since each β-mixed string of beads is composed of a rescaled $(\beta, 1-2\beta)$- and a rescaled (β, β)-string of beads, we conclude that a β-mixed string of beads is an element of $\widetilde{\Xi}$ a.s.. We denote the distributions of (I, λ) and (I, I', λ) on $\widetilde{\Xi}$ and $\widetilde{\Xi}^{[1]}$ by ν_β and $\nu_\beta^{[1]}$, respectively, i.e.

$$\Psi_\beta((I_1, \lambda_1), (I_2, \lambda_2), B) \sim \nu_\beta^{[1]}. \tag{2.59}$$

Remark 2.40. The total length of a β-mixed string of beads is given by $K = B^\beta K_1 + (1-B)^\beta K_2$ where $B \sim \text{Beta}(1-2\beta, \beta)$ and $K_1 \sim \text{ML}(\beta, 1-2\beta)$, $K_2 \sim \text{ML}(\beta, \beta)$ are independent. Therefore, by Proposition 2.13 with $\theta_1 = 1-2\beta, \theta_2 = \beta$, noting that $(B, 1-B) \sim \text{Dirichlet}(1-2\beta, \beta)$, we have

$$\left(B^\beta K_1, (1-B)^\beta K_2\right) \overset{d}{=} K(B', 1-B') \tag{2.60}$$

where $B' \sim \text{Beta}(1/\beta - 2, 1)$ is independent of K, and $K \sim \text{ML}(\beta, 1-\beta)$, but note that we cannot expect that $([0, K], \lambda)$ is a $(\beta, 1-\beta)$-string of beads when $\beta \in (0, 1/2)$. In particular, at the junction point in a $(\beta, 1-\beta)$-string of beads, we would expect a $\text{Beta}(\beta, 1-2\beta)$ mass split into a rescaled (β, β)- and a rescaled $(\beta, 1-2\beta)$-string of beads in this order (and not vice versa).

We now apply Corollary 2.38 and Proposition 2.39 with $\nu = \nu_\beta$ to obtain a CRT $(\check{\mathcal{T}}, \check{\mu})$. We will embed the stable line-breaking construction of Algorithm 2.4 into the tree $\mathcal{T}^*$ defined via

$$\mathcal{T}^* := \phi_\beta(\xi_{\beta,\beta}, (\mathcal{T}_i^*, i \geq 1)) \tag{2.61}$$

where $(\mathcal{T}_i^*, \mu_i^*, i \geq 1)$ is an i.i.d. sequence of weighted $\mathbb{R}$-trees with the same distribution as $(\check{\mathcal{T}}, \check{\mu})$, and $\xi_{\beta,\beta}$ is an independent (β, β)-string of beads with ranked atom masses $(D_i, i \geq 1) \sim \text{PD}(\beta, \beta)$.

The tree $\mathcal{T}^*$ is constructed by replacing the atoms of a (β, β)-string of beads by independent copies of $\check{\mathcal{T}}$ with metric rescaled by D_i^β, which yields a natural mass measure μ^* on $\mathcal{T}^*$ via

$$\mu^* = \sum_{i \geq 1} D_i \mu_i^*(\mathcal{T}_i^* \cap \cdot).$$

Lemma 2.41. *The tree $\mathcal{T}^*$ is compact a.s..*

Proof. The trees attached to the atoms of $\xi_{\beta,\beta}$ have heights $D_i^\beta \mathrm{ht}(\mathcal{T}_i^*), i \geq 1$. We have, for any $\epsilon > 0$,

$$\sum_{i=1}^\infty \mathbb{P}\Big(D_i^\beta \mathrm{ht}(\mathcal{T}_i^*) > \epsilon\Big) \leq \epsilon^{-p} \sum_{i=1}^\infty \mathbb{E}\Big[D_i^{p\beta}\Big] \mathbb{E}[\mathrm{ht}(\mathcal{T}_i^*)^p] \leq \epsilon^{-p} \mathbb{E}\Big[\mathrm{ht}(\check{\mathcal{T}})^p\Big] < \infty,$$

where we choose p such that $p\beta \geq 1$ and $\mathbb{E}[\mathrm{ht}(\check{\mathcal{T}})^p] < \infty$, cf. (2.45) and Corollary 2.37. By the first Borel-Cantelli lemma and the compactness of $\mathcal{T}_i^*$, $i \geq 1$, and $\xi_{\beta,\beta}$, we conclude that $\mathcal{T}^*$ is compact. $\square$

We would like to exploit the recursive structure of $\mathcal{T}^*$, so we need to introduce some notation first. Let

$$\hat{\xi}_{\mathbf{i}} = \left(\hat{E}_{\mathbf{i}}, \sum_{j\geq 1} \hat{P}_{\mathbf{i}j} \delta_{\hat{x}_{\mathbf{i}j}}\right), \quad \mathbf{i} \in \mathcal{N}, \tag{2.62}$$

be such that

$$\mathcal{T}^* = \phi_\beta\Big(\hat{\xi}_\varnothing, \Big(\hat{\mathcal{T}}_j, j \geq 1\Big)\Big), \tag{2.63}$$

i.e. $\hat{\xi}_\varnothing = \xi_{\beta,\beta}$ and $\widehat{\mathcal{T}}_j, j \geq 1$, are i.i.d. with the same distribution as $\check{\mathcal{T}}$. Hence, we may assume that for all $\mathbf{i} \in \mathcal{N}$,

$$\hat{\mathcal{T}}_{\mathbf{i}} = \phi_\beta\Big(\hat{\xi}_{\mathbf{i}}, \Big(\hat{\mathcal{T}}_{\mathbf{i}j}, j \geq 1\Big)\Big), \quad \mathbf{i} \in \mathcal{N}, \tag{2.64}$$

where $\hat{\mathcal{T}}_{\mathbf{i}j}, j \geq 1$, are i.i.d. with the same distribution as $\check{\mathcal{T}}$. We can now describe the mass measure μ^* by considering the masses of the subtrees $\mathcal{T}_x = \{\sigma \in \mathcal{T}^* : x \in [[\rho, \sigma]]\}$, and obtain, $\mu^*(\mathcal{T}_x) = \sum_{j\geq 1 : \hat{x}_j \geq x} \hat{P}_j$ if $x \in \hat{E}_\varnothing$, and

$$\mu^*(\mathcal{T}_x) = \sum_{j \geq 1 : x_{\mathbf{i}j} \geq x} \hat{P}_{i_1} \hat{P}_{i_1 i_2} \dots \hat{P}_{i_1 i_2 \cdots i_n} \hat{P}_{i_1 i_2 \cdots i_n j} \tag{2.65}$$

if $x \in \hat{E}_{\mathbf{i}}$ for some $\mathbf{i} = i_1 \cdots i_n$ with $n \geq 1$. Note that, for each $\mathbf{i} \in \mathcal{N} \setminus \{\varnothing\}$, $(\hat{\xi}_{\mathbf{i}j}, j \geq 1)$ is a sequence of i.i.d. copies of the string of beads $\xi \sim \nu_\beta$ as in Corollary 2.38.

2.5.1 The binary two-colour line-breaking construction

The aim of this section is to show that the tree $\mathcal{T}^*$ can be obtained as the limit of the sequence of trees constructed in the two-colour line-breaking construction given by Algorithm 2.6 in the Introduction. Note that the tree $\mathcal{T}^*$ is an $\mathbb{R}$-tree without any marked components, whereas the two-colour line-breaking construction evolves in the space of ∞-marked $\mathbb{R}$-trees. We present an enhanced version of Algorithm 2.6, which is based on sampling from the mass measure, and we show that we can embed this sequence of weighted $\mathbb{R}$-trees into $(\mathcal{T}^*, \mu^*)$.

Let us first make the attachment procedure precise. For any $\mathbb{R}$-tree $\mathcal{T}$ with metric d and discrete mass measure μ, a parameter $\beta \in (0, 1/2]$, an element $J \in \mathcal{T}$ and a string of beads $(I^+ = [0, K^+], \mu^+)$ with metric d^+, the tree $(\mathcal{T}', d', \mu')$ created from $(\mathcal{T}, d, \mu)$ by *attaching* to J the string of beads (I^+, μ^+) with mass measure μ^+ rescaled by $\mu(J)$ and metric d^+ rescaled by $\mu(J)^\beta$ is defined as follows. Set

$$\mathcal{T}' := \mathcal{T} \setminus \{J\} \sqcup I^+, \tag{2.66}$$

and equip $\mathcal{T}'$ with the metric d' defined by

$$d'(x,y) := \begin{cases} d(x,y) & \text{if } x, y \in \mathcal{T}, \\ d(x,J) + (\mu(J))^\beta d^+(0,y) & \text{if } x \in \mathcal{T}, y \in I, \\ (\mu(J))^\beta d^+(x,y) & \text{if } x, y \in I, \end{cases} \tag{2.67}$$

and the mass measure μ' given by

$$\mu' = \sum_{x \in \mathcal{T}'} \mu'(x)\delta_x := \sum_{x \in \mathcal{T} \setminus \{J\}} \mu(x)\delta_x + \mu(J) \sum_{x \in I^+} \mu^+(x)\delta_x. \tag{2.68}$$

We now consider ∞-marked $\mathbb{R}$-trees with a discrete branching structure. Note that the marked components $\mathcal{R}^{(i)}, i \geq 1$, of an ∞-marked $\mathbb{R}$-tree $(\mathcal{T}, (\mathcal{R}^{(i)}, i \geq 1))$ are compact $\mathbb{R}$-trees when equipped with the metric restricted to $\mathcal{R}^{(i)}, i \geq 1$, respectively. We call an element of $\mathcal{R}^{(i)}$ a *branch point* of $\mathcal{R}^{(i)}$ if its removal disconnects $\mathcal{R}^{(i)}$ into three or more components; it is a *leaf* of $\mathcal{R}^{(i)}$ if its removal does not disconnect $\mathcal{R}^{(i)}$. $\mathrm{Lf}(\mathcal{R}^{(i)})$ and $\mathrm{Br}(\mathcal{R}^{(i)})$ are the set of leaves and the set of branch points of $\mathcal{R}^{(i)}$, respectively.

Recall that we distinguish between *internal* and *external* edges of $\mathcal{R}^{(i)}$ where the *edges* of $\mathcal{R}^{(i)}$ are the connected components of $\mathcal{R}^{(i)} \setminus (\mathrm{Br}(\mathcal{R}^{(i)}) \cup \mathrm{Lf}(\mathcal{R}^{(i)}))$. External edges of $\mathcal{R}^{(i)}$ are edges connecting a branch point and a leaf of $\mathcal{R}^{(i)}$, internal edges connect two branch points or the root and a branch point. The same definition is applied to distinguish between internal and external edges of $\mathcal{T}$. In the following, we will slightly abuse the definition of an edge; we will also refer to the set containing the endpoints or only one endpoint as an edge.

Algorithm 2.42 (Two-colour line-breaking construction)**.** Let $\beta \in (0, 1/2]$. We grow weighted ∞-marked $\mathbb{R}$-trees $(\mathcal{T}_k^*, (\mathcal{R}_k^{(i)}, i \geq 1), \mu_k^*)$, $k \geq 0$, as follows.

0. Let $(\mathcal{T}_0^*, \mu_0^*)$ be isometric to a (β, β)-string of beads ξ_0; let $r_0 = 0$ and $\mathcal{R}_0^{(i)} = \{\rho\}$, $i \geq 1$.

Given $(\mathcal{T}_j^*, (\mathcal{R}_j^{(i)}, i \geq 1), \mu_j^*)$ with $\mu_j^* = \sum_{x \in \mathcal{T}_j^*} \mu_j^*(x)\delta_x$, $0 \leq j \leq k$, let $r_k = \#\{i \geq 1 \colon \mathcal{R}_k^{(i)} \neq \{\rho\}\}$;

1. select an edge $E_k^* \subset \mathcal{T}_k^*$ with probability proportional to its mass $\mu_k^*(E_k^*)$; if $E_k^* \subset \mathcal{R}_k^{(i)}$ for some $i \in [r_k]$, let $I_k = i$; otherwise, i.e. if $E_k^* \subset \mathcal{T}_k^* \setminus \bigcup_{i \in [r_k]} \mathcal{R}_k^{(i)}$, let $r_{k+1} = r_k + 1$, $I_k = r_{k+1}$;
2. if E_k^* is an external edge of $\mathcal{R}_k^{(i)}$, perform $(\beta, 1-2\beta)$-coin tossing sampling on E_k^* to determine $J_k^* \in E_k^*$ (cf. Proposition 2.18); otherwise, i.e. if $E_k^* \subset \mathcal{T}_k^* \setminus \bigcup_{i \in [r_k]} \mathcal{R}_k^{(i)}$ or if E_k^* is an internal edge of $\mathcal{R}_k^{(i)}$, sample J_k^* from the normalised mass measure on E_k^*;
3. let (E_k^+, R_k^+, μ_k^+) be an independent β-mixed string of beads; to form $(\mathcal{T}_{k+1}^*, \mu_{k+1}^*)$, remove $\mu_k^*(J_k^*)\delta_{J_k^*}$ from μ_k^* and attach to $\mathcal{T}_k^*$ at J_k^* an isometric copy of (E_k^+, μ_k^+) with measure rescaled by $\mu_k^*(J_k^*)$ and metric rescaled by $(\mu_k^*(J_k^*))^\beta$; add to $\mathcal{R}_k^{(I_k)}$ the (image under the isometry of) R_k^+ to form $\mathcal{R}_{k+1}^{(I_k)}$; set $\mathcal{R}_{k+1}^{(i)} = \mathcal{R}_k^{(i)}$, $i \neq I_k$.

To analyse Algorithm 2.42, we will need some more notation, in particular with regard to the marked subtree growth processes $(\mathcal{R}_k^{(i)}, k \geq 0)$, $i \geq 1$. Define the random subsequences $(k_m^{(i)}, m \geq 1)$, $i \geq 1$, by

$$k_1^{(i)} := \inf\Big\{n \geq 1 : \mathcal{R}_n^{(i)} \neq \mathcal{R}_0^{(i)}\Big\} = \inf\Big\{n \geq 1 : \mathcal{R}_n^{(i)} \neq \{\rho\}\Big\}, \tag{2.69}$$

and, for $m \geq 1$,

$$k_{m+1}^{(i)} := \inf\Big\{n \geq k_m^{(i)} : \mathcal{R}_n^{(i)} \neq \mathcal{R}_{k_m^{(i)}}^{(i)}\Big\}, \tag{2.70}$$

i.e. there is a change in $(\mathcal{R}_k^{(i)}, k \geq 1)$ when $k = k_m^{(i)}$ for some $m \geq 1$. Note that $\bigcup_{i\geq 1}\{k_m^{(i)}, m \geq 1\} = \{1, 2, \ldots\}$ is a disjoint union, and that, for any $i \geq 1$, $\mathcal{R}_k^{(i)}$ is a binary tree, rooted at $J^*_{k_1^{(i)}-1}$ for any $k \geq 1$, where we also use the convention that $\rho \notin \mathcal{R}_k^{(i)}$ for $k \geq k_1^{(i)}$. For $k = k_m^{(i)} - 1$, we write

$$R_k^+ = [[J_k^*, \Omega_m^{(i)}]] \subset E_k^+ = [[J_k^*, \Sigma_{k+1}]], \qquad \text{i.e.} \quad]]J_k^*, \Omega_m^{(i)}]] = \mathcal{R}_{k+1}^{(i)} \setminus \mathcal{R}_k^{(i)}.$$

In other words, at step $k = k_m^{(i)} - 1$, $\Omega_m^{(i)}$ and Σ_{k+1} denote the leaves added to $\mathcal{R}_k^{(i)}$ and $\mathcal{T}_k^*$, respectively.

We write $\xi_k^{(1)}, \xi_k^{(2)}$ and γ_k for the random variables inducing the β-mixed string of beads (E_k^+, R_k^+, μ_k^+) as in (2.57), i.e.

$$\Big(E_k^+, R_k^+, \mu_k^+\Big) := \Psi_\beta\Big(\xi_k^{(1)}, \xi_k^{(2)}, \gamma_k\Big).$$

We now establish a result on the total mass split in the trees $(\mathcal{T}_k^*, k \geq 0)$, where, for any $\mathbb{R}$-tree $(\mathcal{T}, \mu)$ with n edges $E_1, \ldots, E_n$, a vector $(X_1, \ldots, X_n)$ with $X_i := \mu(E_i)/\mu(\mathcal{T})$, $i \in [n]$, is called *mass split* in $\mathcal{T}$ if $\sum_{i=1}^n X_i = 1$. We will also consider mass splits in subtrees of a tree $\mathcal{R} \subset \mathcal{T}$, that is, the mass split in $(\mathcal{R}, \mu(\mathcal{R}^{-1})\mu\restriction_{\mathcal{R}})$. To distinguish between the mass split in the "big" tree $\mathcal{T}$, and the mass splits in "small" subtrees $\mathcal{R}$ we will speak of the *total* and *internal* (*relative*) mass splits, respectively.

Proposition 2.43 (Mass split in $\mathcal{T}_k^*$). *Let the sequence* $(\mathcal{T}_k^*, (\mathcal{R}_k^{(i)}, i \geq 1), \mu_k^*, k \geq 0)$ *be as in Algorithm 2.42 for some* $\beta \in (0, 1/2]$. *The total mass split between the edges of* $\mathcal{T}_k^*$ *has a*

$$\text{Dirichlet}(\beta, \ldots, \beta, 1 - 2\beta, \ldots, 1 - 2\beta) \tag{2.71}$$

distribution, with parameter β for each internal marked and each unmarked edge, and parameter $1-2\beta$ for each external marked edge. Furthermore, for each edge E of $\mathcal{T}_k^*$,

$$\left(\mu_k^*(E)^{-\beta}E, \mu_k^*(E)^{-1}\mu_k^* \restriction_E\right) \tag{2.72}$$

is a (β,θ)-string of beads, where $\theta = 1-2\beta$ if E is an external marked edge of $\mathcal{R}_k^{(i)}$ for some $i \in [r_k]$, and $\theta = \beta$ otherwise. The strings of beads (2.72) *are independent of each other, and of the mass split* (2.71)*.*

Proof. This proof is an application of the properties of the Dirichlet distribution, Proposition 2.14, and of coin tossing sampling, Proposition 2.18. We give a brief sketch of the proof, using an induction on k.

For $k = 0$, $(\mathcal{T}_0^*, \mu_0^*)$ is a (β,β)-string of beads by definition. Assume that the claim holds up to k. In the update step from $\mathcal{T}_k^*$ to $\mathcal{T}_{k+1}^*$, we first select an edge of $\mathcal{T}_k^*$ proportionally to mass. By Proposition 2.14(iv), the parameter for this edge in the Dirichlet split (2.71), conditionally given that it has been selected, is then increased by 1. We select an atom J_k^* on this edge via (β,θ)-coin tossing, where $\theta = 1-2\beta$ for external marked edges, and $\theta = \beta$ otherwise, and, by Proposition 2.18, the selected edge is split by J_k^* into a rescaled independent (β,β)- and a rescaled independent (β,θ)-string of beads where the relative mass split on this edge is Dirichlet$(\beta, 1-\beta, \theta)$, which is conditionally independent of the total mass split. Furthermore, the mass $\mu_k^*(J_k^*)$ is split by the independent random variable $\gamma_k \sim \text{Beta}(1-2\beta,\beta)$ into a marked $(\beta, 1-2\beta)$-string of beads, and an unmarked (β,β)-string of beads, which are independent, i.e., by Proposition 2.14(iii), the claim follows. □

Remark 2.44. By Proposition 2.43 and Lemma 2.19 we conclude that Algorithm 2.42 implies Algorithm 2.6.

We will show that $\mathcal{T}^* = \overline{\bigcup_{k\geq 0}\mathcal{T}_k^*}$, equipped with the metric d^* which is the natural extension of the metrics d_k^* on $\mathcal{T}_k^*$, $k \geq 0$, and obtain an ∞-marked metric space via

$$\left(\mathcal{T}^*, \left(\mathcal{R}^{(i)}, i \geq 1\right)\right) \quad \text{where } \mathcal{R}^{(i)} := \overline{\bigcup_{k\geq 0}\mathcal{R}_k^{(i)}}, \quad i \geq 1. \tag{2.73}$$

Given $\mathcal{T}^*$ with mass measure μ^*, and the recursive description (2.62) – (2.65) via the strings of beads $\hat{\xi}_{\mathbf{i}}, \mathbf{i} \in \mathcal{N}$, we will obtain the sequence of weighted ∞-marked $\mathbb{R}$-trees of Algorithm 2.42 as an increasing sequence of trees with mass measures embedded in $\mathcal{T}^*$. Note that the strings of beads $\hat{\xi}_{\mathbf{i}}$, $\mathbf{i} \in \mathcal{N} \setminus \{\varnothing\}$, are obtained as the β-mixed strings of beads used in Algorithm 2.42 but are not elements of the space of (equivalence classes of) weighted 1-marked $\mathbb{R}$-trees $\mathbb{T}_{\mathrm{w}}^{[1]}$, as there is no marked component. Therefore, given some $\hat{\xi} = (I = [0, K], \lambda) \sim \nu_\beta$, we need to determine $I_1 = [0, K_1] \subset I = [0, K]$ such that $(I, I_1, \lambda) \sim \nu_\beta^{[1]}$, where ν_β and $\nu_\beta^{[1]}$ were introduced in (2.59). To this end, we need the conditional distribution of the point of the colour change K_1 given $\hat{\xi}$, which exists due to the following proposition.

Proposition 2.45. *Let $\mathbb{T}$ be the space of equivalence classes of compact $\mathbb{R}$-trees equipped with its Borel σ-algebra, and let $\hat{\xi} \sim \nu_\beta$. Then there exists a unique probability kernel κ from $\mathbb{T}$ to $\mathbb{R}$ such that*

$$\mathbb{P}\Big(K_1 \in \cdot \Big| \hat{\xi}\Big) = \kappa\Big(\hat{\xi}, \cdot\Big) \quad a.s.. \tag{2.74}$$

Proof. This is a special case of Theorem 6.3 in [45], since $\mathbb{R}$ is a Borel space. □

Given the weighted $\mathbb{R}$-tree $(\mathcal{T}^*, \mu^*)$, we will obtain a sequence of weighted ∞-marked $\mathbb{R}$-trees

$$\Big(\overline{\mathcal{T}}_k^*, \Big(\overline{\mathcal{R}}_k^{(i)}, i \geq 1\Big), \overline{\mu}_k^*, k \geq 0\Big)$$

with the same distribution as $(\mathcal{T}_k^*, (\mathcal{R}_k^{(i)}, i \geq 1), \mu_k^*, k \geq 0)$ as an increasing sequence of subsets $\overline{\mathcal{T}}_k^* \subset \mathcal{T}^*$, $k \geq 0$, where the mass measure $\overline{\mu}_k^*$ captures the masses of the connected components of $\mathcal{T}^* \setminus \overline{\mathcal{T}}_k^*$ projected onto $\overline{\mathcal{T}}_k^*$, $k \geq 0$. The recursive structure (2.64) provides the i.i.d. strings of beads needed in Algorithm 2.42, which the colour change kernel (2.74) turns into i.i.d. 1-marked strings of beads.

Algorithm 2.46 (Two-colour embedding)**.** Let $\beta \in (0, 1/2]$. We embed into the tree $(\mathcal{T}^*, \mu^*)$ constructed in (2.61)-(2.65) weighted ∞-marked $\mathbb{R}$-trees $(\overline{\mathcal{T}}_k^*, (\overline{\mathcal{R}}_k^{(i)}, i \geq 1), \overline{\mu}_k^*)$, $k \geq 0$, as follows.

0. Let $(\overline{\mathcal{T}}_0^*, \overline{\mu}_0^*) = \hat{\xi}_\varnothing$ be the initial (β, β)-string of beads ξ_0; let $\overline{r}_0 = 0$ and $\overline{\mathcal{R}}_0^{(i)} = \{\rho\}$, $i \geq 1$.

Given $(\overline{\mathcal{T}}_j^*, (\overline{\mathcal{R}}_j^{(i)}, i \geq 1), \overline{\mu}_j^*)$ with $\overline{\mu}_j^* = \sum_{x \in \mathcal{T}_j^*} \overline{\mu}_j^*(x)\delta_x$, $0 \leq j \leq k$, let $\overline{r}_k = \#\{i \geq 1 \colon \overline{\mathcal{R}}_k^{(i)} \neq \{\rho\}\}$;

1. select an edge $\overline{E}_k^* \subset \overline{\mathcal{T}}_k^*$ with probability proportional to its mass $\overline{\mu}_k^*(\overline{E}_k^*)$; if $\overline{E}_k^* \subset \overline{\mathcal{R}}_k^{(i)}$ for some $i \in [\overline{r}_k]$, let $\overline{I}_k = i$; otherwise, i.e. if $\overline{E}_k^* \subset \overline{\mathcal{T}}_k^* \setminus \bigcup_{i \in [\overline{r}_k]} \overline{\mathcal{R}}_k^{(i)}$, let $\overline{r}_{k+1} = \overline{r}_k + 1$, $\overline{I}_k = \overline{r}_{k+1}$;
2. if $\overline{E}_k^*$ is an external edge of $\overline{\mathcal{R}}_k^{(i)}$, perform $(\beta, 1-2\beta)$-coin tossing sampling on $\overline{E}_k^*$ to determine $\overline{J}_k^* \in \overline{E}_k^*$; otherwise, i.e. if $\overline{E}_k^* \subset \overline{\mathcal{T}}_k^* \setminus \bigcup_{i \in [\overline{r}_k]} \overline{\mathcal{R}}_k^{(i)}$ or if $\overline{E}_k^*$ is an internal edge of $\overline{\mathcal{R}}_k^{(i)}$, sample $\overline{J}_k^*$ from the normalised mass measure on $\overline{E}_k^*$;
3. let $\mathbf{i} = i_1 i_2 \cdots i_n \in \mathcal{N}$ and $j \geq 1$ be such that $\overline{J}_k^* = \hat{x}_{\mathbf{i}j} \in \hat{\xi}_{\mathbf{i}}$ and $\overline{\mu}_k^*(\overline{J}_k^*) = \hat{P}_{i_1} \hat{P}_{i_1 i_2} \cdots \hat{P}_{i_1 i_2 \cdots i_n} \hat{P}_{\mathbf{i}j}$, see (2.65); sample a point $\overline{\Omega}_k$ from $\kappa(\hat{\xi}_{\mathbf{i}j}, \cdot)$; to form $(\overline{\mathcal{T}}_{k+1}^*, \overline{\mu}_{k+1}^*)$, remove $\overline{\mu}_k^*(\overline{J}_k^*)\delta_{\overline{J}_k^*}$ from $\overline{\mu}_k^*$ and add to $\overline{\mathcal{T}}_k^*$ the string of beads $\hat{\xi}_{\mathbf{i}j} \subset \mathcal{T}^*$; set $\overline{\mathcal{R}}_{k+1}^{(\overline{I}_k)} = \overline{\mathcal{R}}_k^{(\overline{I}_k)} \cup [[\overline{J}_k^*, \overline{\Omega}_k]]$ and $\overline{\mathcal{R}}_{k+1}^{(i)} = \overline{\mathcal{R}}_k^{(i)}$, $i \neq \overline{I}_k$.

The proof of the following statement can be found in the appendix.

Proposition 2.47. *The sequences of trees constructed in Algorithm 2.42 and Algorithm 2.46 have the same distribution, i.e.*

$$\left(\mathcal{T}_k^*, \left(\mathcal{R}_k^{(i)}, i \geq 1\right), \mu_k^*, k \geq 0\right) \stackrel{d}{=} \left(\overline{\mathcal{T}}_k^*, \left(\overline{\mathcal{R}}_k^{(i)}, i \geq 1\right), \overline{\mu}_k^*, k \geq 0\right).$$

Before we identify a tree growth process related to the stable tree embedded in the two-colour line-breaking construction, we need to recall and establish some preliminary results on stable trees.

2.5.2 An introduction to stable trees

The stable trees are a family of compact random $\mathbb{R}$-trees, parametrised by $\beta \in (0, 1/2]$. They describe the genealogical structure of continuous-state branching processes with branching mechanism $\lambda \mapsto \lambda^{1/(1-\beta)}$, and are related to various fragmentation and coalescence models, see [28, 29, 40, 43, 48, 52, 53]. We define the stable tree based on Goldschmidt and Haas' line-breaking construction.

Definition 2.48 (Stable tree)**.** Let $(\mathcal{T}_k, k \geq 0)$ be the sequence of discrete trees constructed in Algorithm 2.4 for some $\beta \in (0, 1/2]$. The tree

$$\mathcal{T} := \overline{\bigcup_{k \geq 0} \mathcal{T}_k} \tag{2.75}$$

is called *stable tree* of parameter β.

Many discrete random tree growth models such as the one described by Marchal, Algorithm 2.3, have a stable tree as their scaling limit (in the Gromov-Hausdorff sense). Critical Galton-Watson trees whose offspring distribution lies in the domain of attraction of a stable distribution of index $(1-\beta)^{-1}$ conditioned on having n vertices at step n converge in distribution to the stable tree of parameter β, see [9, 27]. Considering the limit of the empirical probability measure on the vertices of these Galton-Watson trees naturally leads to a uniform probability measure supported by the leaves of the stable tree.

Goldschmidt and Haas [39] gave two line-breaking constructions of the stable trees, where branch point selection can either be based on weights, see Algorithm 2.4, or vertex degrees, see [39] for more details. In the following, we consider the tree growth process $(\mathcal{T}_k, k \geq 0)$ obtained via Algorithm 2.4.

Leaves and branch points have a natural order induced by time of appearance, i.e. we can write $(v_i, i \geq 1)$ for the branch points, and $W_k^{(i)}$ for the branch point weight of v_i in $\mathcal{T}_k$ (if $v_i \notin \mathrm{Br}(\mathcal{T}_k)$ or $i > b_k$, set $W_k^{(i)} = 0$). We will also need an order of the edges and the edge lengths of $(\mathcal{T}_k, k \geq 0)$ in Algorithm 2.4. It was suggested in [39] to proceed inductively as follows.

Set the length $L_0^{(1)}$ of the first edge equal to S_0, and let $|\mathcal{T}_k|$ denote the number of edges of $\mathcal{T}_k$ for $k \geq 0$. Conditionally given a vector

$$L_k = \left(L_k^{(1)}, \ldots, L_k^{(|\mathcal{T}_k|)}\right) \tag{2.76}$$

of edge lengths of $\mathcal{T}_k$, let $L_{k+1}^{(|\mathcal{T}_k|+1)}$ denote the length of the edge added at step $k+1$, and recall that $J_k \in \mathcal{T}_k$ is the point to which this edge is attached to. We have two possibilities for the location of J_k.

- If J_k is a branch point of $\mathcal{T}_k$, i.e. $J_k \in \mathrm{Br}(\mathcal{T}_k)$, then $|\mathcal{T}_{k+1}| = |\mathcal{T}_k| + 1$ and $L_{k+1}^{(|\mathcal{T}_k|+1)} = (S_{k+1} - S_k)B_k$. We set $L_{k+1}^{(i)} = L_k^{(i)}, i \in [|\mathcal{T}_k|]$.
- If J_k is a point along an edge of $\mathcal{T}_k$, i.e. $J_k \in \mathcal{T}_k \setminus \mathrm{Br}(\mathcal{T}_k)$, then $|\mathcal{T}_{k+1}| = |\mathcal{T}_k| + 2$, and we add one edge of length $L_{k+1}^{(|\mathcal{T}_k|+1)}$, and split an edge of length $L_k^{(j)}$ for some $j \in [|\mathcal{T}_k|]$, say, into two edges whose lengths we denote by $L_{k+1}^{(j)}$ and $L_{k+1}^{(|\mathcal{T}_k|+2)}$, using the convention that $L_{k+1}^{(j)}$ denotes the length of the edge closest to the root. We set $L_{k+1}^{(i)} = L_k^{(i)}, i \in [|\mathcal{T}_k|] \setminus \{j\}$.

This description yields an order of the edges of $\mathcal{T}_k, k \geq 0$, and their lengths, respectively denoted by

$$E_k = \left(E_k^{(1)}, \ldots, E_k^{(|\mathcal{T}_k|)}\right), \quad L_k = \left(L_k^{(1)}, \ldots, L_k^{(|\mathcal{T}_k|)}\right). \tag{2.77}$$

The following result is directly taken from [39].

Lemma 2.49 ([39] Proposition 3.2)**.** *For $k \geq 1$, given the shapes of $\mathcal{T}_0, \ldots, \mathcal{T}_k$, and $|\mathcal{T}_k| - (k+1) = l$, i.e. conditionally given that the tree $\mathcal{T}_k$ has l branch points $(v_i, i \in [l])$ and l internal edges, we have*

$$\left(L_k^{(1)}, \ldots, L_k^{(k+1+l)}, W_k^{(1)}, \ldots, W_k^{(l)}\right) = S_k\left(Z_k^{(1)}, \ldots, Z_k^{(k+l+1)}, Z_k^{(k+l+2)}, \ldots Z_k^{(k+1+2l)}\right) \tag{2.78}$$

where $S_k \sim \mathrm{ML}(\beta, \beta + k)$ and $Z_k := (Z_k^{(1)}, \ldots, Z_k^{(k+l+1)}, Z_k^{(k+l+2)}, \ldots Z_k^{(k+1+2l)}) \sim \mathrm{Dirichlet}(1, \ldots, 1, w(d_1)/\beta, \ldots, w(d_l)/\beta)$ are independent, $w(d_i) = (d_i - 3)(1-\beta) + 1 - 2\beta$ and $d_i = \deg(v_i, \mathcal{T}_k)$ is the degree of v_i.

The following corollary is key for our weight-length representation.

Corollary 2.50 (Masses as lengths)**.** *For $k \geq 1$, given the shapes of $\mathcal{T}_0, \ldots, \mathcal{T}_k$, and $|\mathcal{T}_k| - (k+1) = l$,*

$$\left(L_k^{(1)}, \ldots, L_k^{(k+1+l)}, W_k^{(1)}, \ldots, W_k^{(l)}\right) = \left(X_1^\beta M_k^{(1)}, \ldots, X_{k+1+2l}^\beta M_k^{(k+1+2l)}\right) \tag{2.79}$$

where the random variables

$$X = (X_1, \ldots, X_{k+1+l}, X_{k+2+l}, \ldots X_{k+1+2l}) \sim \mathrm{Dirichlet}(\beta, \ldots, \beta, w(d_1), \ldots, w(d_l))$$

$$M_k^{(i)} \sim \mathrm{ML}(\beta,\beta), \quad i \in [k+1+l], \qquad M_k^{(k+1+l+i)} \sim \mathrm{ML}(\beta, w(d_i)), \quad i \in [l],$$

are independent, and $w(d_i) = (d_i - 3)(1-\beta) + 1 - 2\beta$ *with* $d_i = \deg(v_i, \mathcal{T}_k)$*.*

Proof. We apply Lemma 2.49, and Proposition 2.13 with $n = k+1+2l$, $\theta_i = \beta, i \in [k+1+l]$, and $\theta_{k+1+l+i} = w(d_i)$, $i \in [l]$. It remains to check that $\theta = \sum_{i=1}^n \theta_i = \beta + k - 1$, i.e. that

$$(\beta + k - 1)/\beta = \left(\sum_{i=1}^{l} (d_i - 3)(1/\beta - 1) + (1/\beta - 2)\right) + (k+l).$$

This follows from the fact that the sum of the vertex degrees in a tree with m edges is $2m$, i.e. $\sum_{i=1}^{l} d_i = 2(k+1+l) - (k+1) - 1$, since $\mathcal{T}_k$ has $k+1+l$ edges and $(k+1)+1$ degree-1 vertices. □

Haas et al. [43] analysed the stable tree as an example of a self-similar CRT. Let $(\mathcal{T}, d, \rho)$ with mass measure μ be the stable tree of parameter $\beta \in (0, 1/2]$, and let $\Sigma \sim \mu$ be a leaf sampled from μ. Consider the *spine*, that is, the path $[[\rho, \Sigma]]$ from the root to this leaf, and remove all vertices of degree one or two from this path. This yields a sequence of connected components $(\mathcal{S}_i, i \in I)$ rooted at $\rho_i \in [[\rho, \Sigma]]$, $i \in I$, respectively, where $\rho_i, i \in I$, has infinite degree a.s..

- The *coarse spinal mass partition* is $(P^{(i)\downarrow}, i \geq 1) := (P^{(i)}, i \in I)^{\downarrow} = (\mu(\mathcal{S}_i), i \in I)^{\downarrow}$, i.e. the sequence of masses $(\mu(\mathcal{S}_i), i \in I)$ ranked in decreasing order.
- The *fine spinal mass partition* is the sequence $(P_j^{(i)}, j \in I_i, i \in I)^{\downarrow} := (\mu(\mathcal{S}_j^{(i)}), j \in I_i, i \in I)^{\downarrow}$ where $(\mathcal{S}_j^{(i)}, j \in I_i)$ are the connected components of $\mathcal{S}_i \setminus \{\rho_i\}$, $i \in I$. In other words, the fine spinal mass partition is the sequence of masses ranked in decreasing order of the connected components that we obtain when we remove the whole spine.

Theorem 2.51 (Mass partition in the stable tree, [43] Corollary 10)**.** *Let* $\beta \in (0, 1/2]$*, and let* $\mathcal{T}$ *be the stable tree of parameter* β*. Then the following statements hold.*

(i) *The coarse spinal mass partition has a Poisson-Dirichlet distribution with parameters* (β, β)*, i.e.* $(P^{(i)\downarrow}, i \geq 1) = (\mu(\mathcal{S}_i), i \in I)^{\downarrow} \sim \mathrm{PD}(\beta, \beta)$*.*

(ii) *Fragmenting each block $\mu(\mathcal{S}_i), i \in I$, of the coarse spinal mass partition by an independent random vector with distribution* $\mathrm{PD}(1-\beta,-\beta)$ *and re-ranking in decreasing order yields the fine spinal mass partition, i.e.*

$$\left(\mu\left(\mathcal{S}_j^{(i)}\right), j \in I_i, i \in I\right)^{\downarrow} = \left(F^{(i)} \cdot \mu(\mathcal{S}_i), i \in I\right)^{\downarrow} \quad a.s.$$

for i.i.d. random variables $F^{(i)} := (F_j^{(i)}, j \geq 1) \sim \mathrm{PD}(1-\beta,-\beta)$, $i \in I$.

(iii) *Conditionally given the fine spinal mass partition $(\mu(S_j^{(i)}), j \in I_i, i \in I)^{\downarrow}$ of $\mathcal{T}$, the rescaled trees equipped with restricted mass measures*

$$\left(\mu\left(\mathcal{S}_j^{(i)}\right)^{-\beta}\mathcal{S}_j^{(i)}, \mu(\mathcal{S}_j^{(i)})^{-1}\mu\restriction_{\mathcal{S}_j^{(i)}}\right), \quad j \in I_i, i \in I, \tag{2.80}$$

are i.i.d. copies of $(\mathcal{T},\mu)$.

The α-diversities of these $\mathrm{PD}(\alpha,\theta)$ partitions can naturally be interpreted as lengths in the tree $\mathcal{T}^*$. In particular the β-diversity of the coarse spinal mass partition has distribution $S_0 \sim \mathrm{ML}(\beta,\beta)$ which is the starting point of Goldschmidt-Haas' line-breaking constructions. Similarly, the $(1-\beta)$-diversities of the fragmenting $\mathrm{PD}(1-\beta,-\beta)$ random partitions for each block of the coarse spinal mass partition in (ii) are related to the lengths of the Ford CRTs by which we replace the branch points in the stable tree. The independence of these $\mathrm{PD}(1-\beta,-\beta)$ vectors reflects the independence of the related Ford trees.

We now consider a third construction of the stable tree, Algorithm 2.9, which is based on masses, and yields a tree growth process $(\mathcal{T}_k, k \geq 0)$ as in (2.2) and Algorithm 2.4. We refer to (2.66)-(2.68) for a precise definition of the attachment procedure.

Proposition 2.52. *Let $(\mathcal{T},\mu)$ be the stable tree of parameter β, and consider the reduced tree $\hat{\mathcal{T}}_k = \mathcal{R}(\mathcal{T},\Sigma_0,\ldots,\Sigma_k)$ associated with $\mathcal{T}$ and independently sampled leaves $\Sigma_0,\ldots,\Sigma_k \sim \mu, k \geq 0$. The sequence of weighted $\mathbb{R}$-trees $(\mathcal{T}_k,\mu_k,k\geq 0)$ from Algorithm 2.9 has the same distribution as the sequence of trees $(\hat{\mathcal{T}}_k,\hat{\mu}_k)$ as in* (2.2) *where, for any $k \geq 1$, the mass measure $\hat{\mu}_k = \pi_*^{\hat{\mathcal{T}}_k}\mu$ captures the masses of the connected components of $\mathcal{T}\setminus\hat{\mathcal{T}}_k$ projected onto $\hat{\mathcal{T}}_k$ as in* (2.19)-(2.20).

Furthermore, conditionally given $|\mathcal{T}_k| = k+1+l$ *and the shape of* $\mathcal{T}_k$*, the edges of* $\mathcal{T}_k$ *equipped with the mass measure* μ_k *restricted to each edge, are rescaled independent* (β,β)*-strings of beads given via*

$$\left(\mu_k\left(E_k^{(i)}\right)^{-\beta} E_k^{(i)}, \mu_k\left(E_k^{(i)}\right)^{-1} \mu_k \restriction_{E_k^{(i)}}\right), \quad i \in [k+1+l], \tag{2.81}$$

and the total mass distribution

$$\left(\mu_k\left(E_k^{(1)}\right), \ldots, \mu_k\left(E_k^{(k+1+l)}\right), \mu_k(v_1), \ldots, \mu_k(v_l)\right)$$

which is independent and has a $\mathrm{Dirichlet}(\beta, \ldots, \beta, w(d_1), \ldots, w(d_l))$ *distribution where* $v_i, i \in [l]$*, are the branch points of* $\mathcal{T}_k$ *of degrees* $d_i = \deg(v_i, \mathcal{T}_k)$*,* $i \in [l]$*, respectively, and* $w(d_i) = (d_i - 3)(1-\beta) + (1-2\beta)$*,* $i \in [l]$*, and where we number the edges* $E_k^{(i)}, i \in [k+1+l]$*, of* $\mathcal{T}_k$ *as in* (2.77)*.*

The proof of Proposition 2.52 is part of the appendix.

2.5.3 The marked subtree growth processes

We first consider the evolution of the marked subtrees $\mathcal{R}_k^{(i)}$, $k \geq 1$, for $i \geq 1$. Recall the notation used in Algorithm 2.42. Following (2.77), given that $\mathcal{R}_k^{(i)}$ has m leaves, i.e. $k_m^{(i)} \leq k \leq k_{m+1}^{(i)} - 1$, we denote the edges and the edge lengths of $\mathcal{R}_k^{(i)}$ by

$$E_{m,i} = \left(E_{m,i}^{(1)}, \ldots, E_{m,i}^{(2m-1)}\right), \quad L_{m,i} = \left(L_{m,i}^{(1)}, \ldots, L_{m,i}^{(2m-1)}\right), \tag{2.82}$$

respectively, where we note that $\mathcal{R}_k^{(i)}$ is a binary tree, i.e. $\mathcal{R}_k^{(i)}$ has $2m-1$ edges for $k_m^{(i)} \leq k \leq k_{m+1}^{(i)} - 1$. Recall that $E_{m,i}^{(j)}$ is an internal edge of $\mathcal{R}_k^{(i)}$ if $1 \leq j \leq m-1$, and an external edge if $m \leq j \leq 2m-1$.

We will need the following characterisation of the distribution of Ford trees.

Lemma 2.53 ([42] Proposition 18)**.** *Consider the process* $(\mathcal{F}_m, m \geq 1)$ *of Algorithm 2.5 for some* $\beta' \in (0,1)$*. The distribution of* $\mathcal{F}_m$ *is characterised by the following three independent random variables. Its shape* F_m*, the total length* $S'_m \sim \mathrm{ML}(\beta', m-\beta')$ *and a* $\mathrm{Dirichlet}(1, \ldots, 1, (1-\beta')/\beta', \ldots, (1-\beta')/\beta')$ *length split between the edges of* $\mathcal{F}_m$ *with* *a parameter of* 1 *for each of the* $m-1$ *internal edges, and a parameter of* $(1-\beta')/\beta'$ *for each of the* m *external edges of* $\mathcal{F}_m$.

Proposition 2.54 (Marked subtrees)**.** *Let* $(\mathcal{T}_k^*, (\mathcal{R}_k^{(i)}, i \geq 1), \mu_k^*, k \geq 0)$ *be as in Algorithm 2.42, and fix some* $i \geq 1$. *Then, for* $m \geq 1$, *the following random variables are independent.*

- *The initial mass* $\mu_{k_1^{(i)}}(\mathcal{R}^*_{k_1^{(i)}})$;
- *the shape* $F_m^{(i)}$ *of* $\mathcal{R}^{(i)}_{k_m^{(i)}}$ *which has the same distribution as the shape* F_m *of a Ford tree* $\mathcal{F}_m$ *of index* $\beta' = \beta/(1-\beta)$;
- *the relative mass split* $D_m^{(i)}$ *in* $\mathcal{R}^{(i)}_{k_m^{(i)}}$ *given by*

$$\mu^*_{k_m^{(i)}}\left(\mathcal{R}^{(i)}_{k_m^{(i)}}\right)^{-1}\left(\mu^*_{k_m^{(i)}}\left(E_{m,i}^{(1)}\right),\ldots,\mu^*_{k_m^{(i)}}\left(E_{m,i}^{(m-1)}\right),\mu^*_{k_m^{(i)}}\left(E_{m,i}^{(m)}\right),\ldots,\mu^*_{k_m^{(i)}}\left(E_{m,i}^{(2m-1)}\right)\right) \tag{2.83}$$

  which has a $\mathrm{Dirichlet}(\beta,\ldots,\beta,1-2\beta,\ldots,1-2\beta)$ *distribution;*
- *the independent* (β,θ)*-strings of beads*

$$\xi_{m,i}^{(j)} := \left(\mu^*_{k_m^{(i)}}\left(E_{m,i}^{(j)}\right)^{-\beta} E_{m,i}^{(j)}, \mu^*_{k_m^{(i)}}\left(E_{m,i}^{(j)}\right)^{-1}\mu^*_{k_m^{(i)}}\restriction_{E_{m,i}^{(j)}}\right), \quad 1 \leq j \leq 2m-1, \tag{2.84}$$

  where $\theta = \beta$ *for* $1 \leq j \leq m-1$ *and* $\theta = 1-2\beta$ *for* $m \leq j \leq 2m-1$;
- *the independent random variables* $Q_1^{(i)},\ldots,Q_{m-1}^{(i)}$ *where*

$$\mu^*_{k_{n+1}^{(i)}}\left(\mathcal{R}^{(i)}_{k_{n+1}^{(i)}}\right) = \left(1-Q_n^{(i)}\right)\mu^*_{k_n^{(i)}}\left(\mathcal{R}^{(i)}_{k_n^{(i)}}\right) \tag{2.85}$$

  and $Q_n^{(i)} \sim \mathrm{Beta}(\beta, n(1-\beta)+1-2\beta)$, $1 \leq n \leq m-1$.

The proof of Proposition 2.54 will use the following standard lemma.

Lemma 2.55. *Let* I *be a random variable taking values in a countable set* $\mathbb{I}$. *Futhermore, consider a random variable* Z *taking values in some measurable space* $(\mathbb{V},\mathcal{Z})$ *whose distribution conditional on* $I=j$ *does not depend on* $j \in \mathbb{I}$. *Then,* Z *is independent of* I, *and for any other random variable* Z' *taking values in some measurable space* $(\mathbb{V}',\mathcal{Z}')$ *that is conditionally independent of* Z *given* I, *the pair* (Z',I) *is independent of* Z.

Proof of Proposition 2.54. We prove Proposition 2.54 by induction on $m \geq 1$. The case $m = 1$ is trivial since the shape of $\mathcal{R}_{k_1}^{(i)}$ is deterministic, and the initial $(\beta, 1-2\beta)$-string of beads is independent of the scaling factor $\mu^*_{k_1^{(i)}}(\mathcal{R}^{(i)}_{k_1^{(i)}})$, cf. Algorithm 2.42.

Assume that the claim holds for some $m \geq 1$, i.e. that $\mu^*_{k_1^{(i)}}(\mathcal{R}^{(i)}_{k_1^{(i)}})$, the shape $F_m^{(i)}$ of $\mathcal{R}^{(i)}_{k_m^{(i)}}$, the internal mass split $D_m^{(i)}$ in $\mathcal{R}^{(i)}_{k_m^{(i)}}$, the strings of beads $\xi^{(j)}_{m,i}$ on the edges $E^{(j)}_{m,i}$, $j \in [2m-1]$, and $Q_1^{(i)}, \ldots, Q_{m-1}^{(i)}$ are independent, and have the claimed distributions.

Consider the construction of $\mathcal{R}^{(i)}_{k_{m+1}^{(i)}}$ conditionally given $\mathcal{R}^{(i)}_{k_m^{(i)}}$.

- First, we select an edge E of $\mathcal{R}_{k_m^{(i)}}$ according to the internal mass split in $\mathcal{R}_{k_m^{(i)}}$, that is, conditionally given the Dirichlet$(\beta, \ldots, \beta, 1-2\beta, \ldots, 1-2\beta)$ mass split (2.83), the probability of selecting $E^{(j)}_{m,i}$ is $\mu^*_{k_m^{(i)}}(E^{(j)}_{m,i})$ a.s.. By Proposition 2.14(iv), we have, unconditionally,

$$\mathbb{P}\Big(E = E^{(j)}_{m,i}\Big) = \frac{\theta}{(m-1)(1-\beta)+1-2\beta} = \frac{\theta/(1-\beta)}{(m-1)+(1-2\beta)/(1-\beta)}$$

  where $\theta = \beta$ if $j \in [m-1]$ and $\theta = 1-2\beta$ if $j \in [2m-1] \setminus [m-1]$, as required in a Ford tree of index β'. Note that, by the induction hypothesis and the fact that the edge selection is independent of the shape $F_m^{(i)}$ as it only depends on the internal mass split $D_m^{(i)}$, we conclude that the distribution of the shape $F_{m+1}^{(i)}$ of $\mathcal{R}^{(i)}_{k_{m+1}}$ is as the shape F_m of a Ford tree $\mathcal{F}_{m+1}$ of index β'.

- Second, conditionally given the selected edge E we perform (β, θ)-coin tossing sampling on the independent (β, θ)-string of beads related to E, where $\theta = \beta$ if E is internal and $\theta = 1-2\beta$ if E is external. (β, θ)-coin tossing sampling, cf. Proposition 2.18, yields a rescaled independent (β, β)-string of beads, an atom, and a rescaled independent (β, θ)-string of beads subject to an independent Dirichlet$(\beta, 1-\beta, \theta)$ mass split. By the two-colour line-breaking construction, the selected atom is split according to an independent Beta$(\beta, 1-2\beta)$ variable into a new marked independent $(\beta, 1-2\beta)$-string of beads attached to this atom, and an umarked independent (β, β)-string of beads. Therefore, the mass $\mu^*_{k_m^{(i)}}(\mathcal{R}^{(i)}_{k_m^{(i)}})$ is split according to Dirichlet$(\beta, \ldots, \beta, 1-2\beta, \ldots, 1-2\beta, \beta)$ with a parameter of β

for the new unmarked (β,β)-string of beads and each internal edge, and weight $1-2\beta$ for each external edge.

As the mass of the new unmarked rescaled (β,β)-string of beads is the only component of $\mu^*_{k_m^{(i)}}(\mathcal{R}^{(i)}_{k_m^{(i)}})$ which is not part of $\mu^*_{k_{m+1}^{(i)}}(\mathcal{R}^{(i)}_{k_{m+1}^{(i)}})$, we conclude by Proposition 2.14(i) that

$$\mu^*_{k_{m+1}^{(i)}}\left(\mathcal{R}^{(i)}_{k_{m+1}^{(i)}}\right) = \left(1-Q_m^{(i)}\right)\mu^*_{k_m^{(i)}}\left(\mathcal{R}^{(i)}_{k_m^{(i)}}\right)$$

where $Q_m^{(i)} \sim \mathrm{Beta}(\beta,(m-1)(1-\beta)+1-2\beta)$ is independent of the mass split $D_{m+1}^{(i)}$ in $\mathcal{R}_{k_{m+1}^{(i)}}$ which is Dirichlet$(\beta,\ldots,\beta,1-2\beta,\ldots,1-2\beta)$.

By the induction hypothesis, $Q_1^{(i)},\ldots,Q_{m-1}^{(i)}$ are independent of the internal mass split $D_m^{(i)}$ in $\mathcal{R}_{k_m^{(i)}}$ which determines $Q_m^{(i)}$ via edge selection based on the internal mass split, coin tossing on the independent string of beads related to the selected edge, and the independent Beta variable splitting the selected atom. Hence, $Q_1^{(i)},\ldots,Q_m^{(i)}$ are independent.

Similarly, conditionally given the shape $F_{m+1}^{(i)}$ of $\mathcal{R}^{(i)}_{k_{m+1}^{(i)}}$, by the induction hypothesis and arguments as above, $\mu^*_{k_1^{(i)}}(\mathcal{R}^{(i)}_{k_1^{(i)}})$, the mass split $D_{m+1}^{(i)}$, the strings of beads $\xi_{m+1,i}^{(j)}$, $j\in[2m+1]$, and $Q_1^{(i)},\ldots,Q_m^{(i)}$ are independent and have the required distributions.

- Finally, note that the distributions of all of these random variables, conditionally given $F_{m+1}^{(i)}$, do not depend on $F_{m+1}^{(i)}$ where we recall that $F_{m+1}^{(i)}$ is fully determined by the selected edge E and the shape $F_m^{(i)}$ of $\mathcal{R}^{(i)}_{k_m^{(i)}}$. By Lemma 2.55, we conclude the independence between these variables and the shape $F_{m+1}^{(i)}$ of $\mathcal{R}^{(i)}_{k_{m+1}^{(i)}}$. $\square$

Lemma 2.56 (Length split in marked subtrees)**.** *In the setting of Proposition 2.54, let* $\tilde{S}_{m,i}=\sum_{j=1}^{2m-1}L_{m,i}^{(j)}$ *denote the total length of* $\mathcal{R}^{(i)}_{k_m^{(i)}}$, $m\geq 1$. *Then,*

$$\left(L_{m,i}^{(1)},\ldots,L_{m,i}^{(m-1)},L_{m,i}^{(m)},\ldots,L_{m,i}^{(2m-1)}\right)=\tilde{S}_{m,i}\left(Z_{m,i}^{(1)},\ldots,Z_{m,i}^{(m-1)},Z_{m,i}^{(m)},\ldots,Z_{m,i}^{(2m-1)}\right) \tag{2.86}$$

where $\tilde{S}_{m,i} = \mu^*_{k_m^{(i)}}(\mathcal{R}^{(i)}_{k_m^{(i)}})^\beta S_{m,i}$*, and where* $\mu^*_{k_m^{(i)}}(\mathcal{R}^{(i)}_{k_m^{(i)}})^\beta$*,* $S_{m,i} \sim \mathrm{ML}(\beta,(m-1)(1-\beta)+1-2\beta)$ *and*

$$\left(Z^{(1)}_{m,i},\ldots,Z^{(m-1)}_{m,i},Z^{(m)}_{m,i},\ldots,Z^{(2m-1)}_{m,i}\right) \sim \mathrm{Dirichlet}(1,\ldots,1,1/\beta-2,\ldots,1/\beta-2)$$

are independent. Furthermore, for $m \geq 1$*,*

$$\tilde{S}_{m,i} = B_{m,i}\tilde{S}_{m+1,i} \tag{2.87}$$

where $B_{m,i} \sim \mathrm{Beta}(m(1/\beta-1),1/\beta-2)$ *and* $\tilde{S}_{m+1,i}$ *are independent.*

Proof. Fix $i \geq 1$, and set

$$X_j = \mu^*_{k_m^{(i)}}\left(E^{(j)}_{m,i}\right), \quad j \in [2m-1], \qquad \text{so that } \sum_{j=1}^{2m-1} X_j = \mu^*_{k_m^{(i)}}\left(\mathcal{R}^{(i)}_{k_m^{(i)}}\right).$$

By Proposition 2.54, the edge lengths $L^{(j)}_{m,i}$, $j \in [2m-1]$, are given by $L^{(j)}_{m,i} = X_j^\beta M_m^{(j)}$ where $M_m^{(j)} \sim \mathrm{ML}(\beta,\beta)$ for $j \in [m-1]$, $M_m^{(j)} \sim \mathrm{ML}(\beta,1-2\beta)$ for $j \in [2m-1]\setminus[m-1]$, $\sum_{j=1}^{2m-1} X_j$ and

$$\left(\sum_{j=1}^{2m-1} X_j\right)^{-1}(X_1,\ldots,X_{m-1},X_m,\ldots X_{2m-1}) \sim \mathrm{Dirichlet}(\beta,\ldots,\beta,1-2\beta,\ldots,1-2\beta)$$

are independent. We apply Proposition 2.13 with $n = 2m-1$, $\theta_j = \beta$ for $j \in [m-1]$ and $\theta_j = 1-2\beta$ for $j \in [2m-1]\setminus[m-1]$ to the vector

$$\left(\sum_{j=1}^{2m-1} X_j\right)^\beta \left(\left(\frac{X_1}{\sum_{j=1}^{2m-1} X_j}\right)^\beta M_m^{(1)},\ldots,\left(\frac{X_{2m-1}}{\sum_{j=1}^{2m-1} X_j}\right)^\beta M_m^{(2m-1)}\right). \tag{2.88}$$

Then $\theta = \sum_{j=1}^{2m-1}\theta_j = (m-1)(1-\beta)+1-2\beta$, and hence (2.86) follows.

To see (2.87), recall that $E^{(2m)}_{m+1,i} = \mathcal{R}^{(i)}_{k_{m+1}^{(i)}} \setminus \mathcal{R}^{(i)}_{k_m^{(i)}}$. By (2.88) for $m+1$, and Proposition 2.14(i)-(ii),

$$\mu^*_{k_m^{(i)}}\left(\mathcal{R}^{(i)}_{k_m^{(i)}}\right)^\beta S_{m,i} = B_{m,i}\mu^*_{k_{m+1}^{(i)}}\left(\mathcal{R}^{(i)}_{k_{m+1}^{(i)}}\right)^\beta S_{m+1,i}$$

where $S_{m+1,i} \sim \mathrm{ML}(\beta,m(1-\beta)+1-2\beta)$, $B_{m,i} \sim \mathrm{Beta}(m(1/\beta-1),1/\beta-2)$ and $\mu^*_{k_{m+1}^{(i)}}(\mathcal{R}^{(i)}_{k_{m+1}^{(i)}})$ are independent, i.e. $\tilde{S}_{m,i} = B_{m,i}\tilde{S}_{m+1,i}$. □

We are now ready to present our main result on the subtree growth processes $(\mathcal{R}^{(i)}_{k^{(i)}_m}, m \geq 1), i \geq 1$.

Theorem 2.57 (Embedded Ford trees)**.** *Let $(\mathcal{T}^*_k, (\mathcal{R}^{(i)}_k, i \geq 1), \mu^*_k, k \geq 0)$ be as in Algorithm 2.42.*

(i) *The weighted tree growth processes $(\mathcal{G}^{(i)}_m, \mu^{(i)}_m, m \geq 1)$, $i \geq 1$, defined by*

$$\left(\mathcal{G}^{(i)}_m, \mu^{(i)}_m, m \geq 1\right) := \left(\mu^*_{k^{(i)}_1}\left(\mathcal{R}^{(i)}_{k^{(i)}_1}\right)^{-\beta}\mathcal{R}^{(i)}_{k^{(i)}_m}, \mu^*_{k^{(i)}_1}\left(\mathcal{R}^{(i)}_{k^{(i)}_1}\right)^{-1}\mu^*_k\restriction_{\mathcal{R}^{(i)}_k}, m \geq 1\right), \tag{2.89}$$

are independent and identically distributed.

(ii) *The tree growth processes*

$$\left(\mu^*_{k^{(i)}_1}\left(\mathcal{R}^{(i)}_{k^{(i)}_1}\right)^{-\beta}\mathcal{R}^{(i)}_{k^{(i)}_m}, m \geq 1\right), \quad i \geq 1,$$

are i.i.d. tree growth processes starting from the length $\mathrm{ML}(\beta, 1-2\beta)$ *of a $(\beta, 1-2\beta)$-string of beads, and the shapes $(F^{(i)}_m, m \geq 1)$ of $(\mathcal{R}^{(i)}_{k_m}, m \geq 1)$ evolve as the shapes $(F_m, m \geq 1)$ of a Ford tree tree growth process $(\mathcal{F}_m, m \geq 1)$ of index $\beta' = \beta/(1-\beta)$. Furthermore, for each $i \geq 1$, there is $\mathcal{R}^{(i)}$ such that* $\lim_{m\to\infty}\mathcal{R}^{(i)}_{k^{(i)}_m} = \mathcal{R}^{(i)}$ *a.s. in the Gromov-Hausdorff topology.*

(iii) *For $i \geq 1$, define $C_i(m) := (1-\beta)^\beta m^{-\beta^2/(1-\beta)}\mu^*_{k^{(i)}_m}(\mathcal{R}^{(i)}_{k^{(i)}_m})^{-\beta}$. The processes $(C_i(m)\mathcal{R}^{(i)}_{k^{(i)}_m}, m \geq 1)$, $i \geq 1$, are i.i.d.,*

$$\lim_{m\to\infty} C_i(m) = C_i^{-\beta/(1-\beta)}\mu^*_{k^{(i)}_1}\left(\mathcal{R}^{(i)}_{k^{(i)}_1}\right)^{-\beta} \quad a.s.$$

where $C_i \sim \mathrm{ML}(1-\beta, 1-2\beta)$, and there are i.i.d. Ford CRTs $(\mathcal{F}^{(i)}, i \geq 1)$ of index $\beta' = \beta/1-\beta$ such that

$$\lim_{m\to\infty} C_i(m)\mathcal{R}^{(i)}_{k^{(i)}_m} = \mathcal{F}^{(i)} \quad a.s. \tag{2.90}$$

in the Gromov-Hausdorff topology.

Proof. (i) Informally, our argument is as follows. The initial states are independent and identically distributed as they correspond to independent $(\beta, 1-2\beta)$-strings of beads related to the independent β-mixed strings of beads used in the two-colour line-breaking construction. The transition steps, whenever they happen, are carried out irrespective of other components already created.

A formal and technical proof can be found in the appendix.

The convergence of $\mathcal{R}_k^{(i)}$ to some CRT $\mathcal{R}^{(i)}$ follows from the fact that, by Proposition 2.47, $(\mathcal{R}_k^{(i)}, k \geq 1)$ can be embedded as an increasing sequence into a compact CRT, and hence converges a.s..

(ii) Note that, by Proposition 2.54 (and Proposition 2.43), for each i and $k = k_m^{(i)} - 1$ for some $m \geq 1$, we are in the situation of Lemma 2.19 with $n = 2m - 1$, $\theta_1 = \cdots = \theta_{m-1} = \beta$, $\theta_m = \cdots = \theta_{2m-1} = 1 - 2\beta$, $\alpha = \beta$. We recover the shape evolution of Algorithm 2.5 with index β', cf. Proposition 2.54.

(iii) First note that, by Lemma 2.56, the lengths of the trees $C_i(m)\mathcal{R}_{k_m^{(i)}}^{(i)}$ do not depend on $\mu^*_{k_m^{(i)}}(\mathcal{R}_{k_m^{(i)}}^{(i)})$. Fix some $i \geq 1$ and recall from Proposition 2.54 that there are independent random variables $Q_m^{(i)} \sim \mathrm{Beta}(\beta, m(1-\beta) + 1 - 2\beta)$ such that

$$\mu^*_{k_{m+1}^{(i)}}\left(\mathcal{R}_{k_{m+1}^{(i)}}^{(i)}\right) = \left(1 - Q_m^{(i)}\right)\mu^*_{k_m^{(i)}}\left(\mathcal{R}_{k_m^{(i)}}^{(i)}\right), \quad m \geq 1.$$

Define $P_1^{(i)} := Q_1^{(i)} \sim \mathrm{Beta}(\beta, 2 - 3\beta)$, and, for $m \geq 1$,

$$P_m^{(i)} := \overline{Q}_1^{(i)}\overline{Q}_2^{(i)} \cdots \overline{Q}_{m-1}^{(i)} Q_m^{(i)}$$

where $\overline{Q} = 1 - Q$ for any random variable Q. Note that $P_m^{(i)}$ is the proportion of the mass of $\mu^*_{k_1^{(i)}}(\mathcal{R}_{k_1^{(i)}}^{(i)})$ attached to the $(m+1)$-st leaf of the i-th marked component.

We recognise the stick-breaking construction (2.11) of a $\mathrm{PD}(1-\beta, 1-2\beta)$ vector $(P_m^{(i)}, m \geq 1)^\downarrow$, and obtain the corresponding $(1-\beta)$-diversity C_i by

$$C_i = \lim_{m \to \infty} \left(1 - \sum_{j=1}^{m} P_j^{(i)}\right)^{1-\beta} (1-\beta)^{-(1-\beta)} m^\beta \sim \mathrm{ML}(1-\beta, 1-2\beta), \qquad (2.91)$$

as in (2.14). Now fix some $m_0 \geq 1$ and let $k \geq k_{m_0}^{(i)}$. We consider the reduced tree

$$\mathcal{R}\Big(C_i(m)\mathcal{R}_{k_m^{(i)}}^{(i)}, \Omega_1^{(i)}, \ldots, \Omega_{m_0}^{(i)}\Big) \tag{2.92}$$

spanned by the root v_i and the leaves $\Omega_1^{(i)}, \ldots, \Omega_{m_0}^{(i)}$ of $\mathcal{R}_k^{(i)}$. Recall that the shape and the Dirichlet$(1, \ldots, 1, 1/\beta - 2, \ldots, 1/\beta - 2)$ length split between the edges $E_{m_0,i}^{(1)}, \ldots, E_{m_0,i}^{(2m_0-1)}$ of $\mathcal{R}_{k_{m_0}^{(i)}}^{(i)}$ are as required for the reduced tree associated with a Ford CRT of index β', cf. Proposition 2.54 and Lemma 2.56.

Scaling by $C_i(m)$ is independent of the Dirichlet length split and the shape of $\mathcal{R}_{k_{m_0}^{(i)}}^{(i)}$ and only affects the total length of the reduced tree (2.92), cf. Proposition 2.54 and Lemma 2.56. We will show that the total length of (2.92) scaled by $C_i(m)$ converges a.s. to some $S'_{m_0,i} \sim \mathrm{ML}(\beta', m_0 - \beta')$, which is the total length of the reduced tree spanned by the root and the first m_0 leaves of a Ford CRT of index β', i.e. that

$$\lim_{m\to\infty} \mathrm{Leb}\Big(\mathcal{R}\Big(C_i(m)\mathcal{R}_{k_m^{(i)}}^{(i)}, \Omega_1^{(i)}, \ldots, \Omega_{m_0}^{(i)}\Big)\Big) = S'_{m_0,i} \sim \mathrm{ML}(\beta', m_0 - \beta')$$

where we will use that

$$\begin{aligned} C_i(m) :&= (1-\beta)^\beta m^{-\beta^2/(1-\beta)} \mu^*_{k_m^{(i)}}\left(\mathcal{R}_{k_m^{(i)}}^{(i)}\right)^{-\beta} \\ &= \left(1 - \sum_{j=1}^m P_j^{(i)}\right)^{-\beta} (1-\beta)^\beta m^{-\beta^2/(1-\beta)} \mu^*_{k_1^{(i)}}\left(\mathcal{R}_{k_1^{(i)}}^{(i)}\right)^{-\beta}, \end{aligned}$$

since $1 - \sum_{j=1}^m P_j^{(i)} = \mu^*_{k_m^{(i)}}(\mathcal{R}_{k_m^{(i)}}^{(i)})\mu^*_{k_1^{(i)}}(\mathcal{R}_{k_1^{(i)}}^{(i)})^{-1}$. Hence

$$\lim_{m\to\infty} C_i(m) = C_i^{-\beta/(1-\beta)} \mu^*_{k_1^{(i)}}(\mathcal{R}_{k_1^{(i)}}^{(i)})^{-\beta} \quad \text{a.s.}.$$

Note that C_i is independent of $\mu^*_{k_1^{(i)}}(\mathcal{R}_{k_1^{(i)}}^{(i)})$ as it only depends on the sequence of independent random variables $(Q_j^{(i)}, j \geq 1)$ which is independent of $\mu^*_{k_1^{(i)}}(\mathcal{R}_{k_1^{(i)}}^{(i)})$. The shape of $\mathcal{R}_{k_m^{(i)}}^{(i)}$ has the same distribution as the shape of $\mathcal{F}_m$ where $(\mathcal{F}_m, m \geq 1)$ is a Ford tree growth process of index β'. In particular, by Proposition 2.54 and (ii), we already know that the number of edges $T_{m_0+m} + (2m_0 - 1), m \geq 0$, of the reduced trees (2.92) as a subset of $\mathcal{R}^{(i)}_{k^{(i)}_m}$ behaves like the number of tables in a $(\beta', m_0-\beta')$-CRP, started at time m_0, that is, by (2.10),

$$\lim_{m\to\infty} m^{-\beta/(1-\beta)} T_{m_0+m} = \lim_{m\to\infty} m^{-\beta/(1-\beta)} T_m = S'_{m_0} \quad \text{a.s.} \tag{2.93}$$

where $S'_{m_0} \sim \mathrm{ML}(\beta', m_0-\beta')$, cf. [42]. By Proposition 2.54, we conclude that, in the limit, the length of $\mathcal{R}^{(i)}_{k^{(i)}_{m_0}}$ is given by

$$\lim_{m\to\infty} \mu^*_{k^{(i)}_m}\left(\mathcal{R}^{(i)}_{k^{(i)}_m}\right)^\beta \sum_{j=1}^{T_m}\left(X'_j\right)^\beta M^{(j)}_m = \mu^*_{k^{(i)}_{m_0}}\left(\mathcal{R}^{(i)}_{k^{(i)}_{m_0}}\right)^\beta S_{m_0,i} = \tilde{S}_{m_0,i} \tag{2.94}$$

where, the mass split $X' := (X'_1,\dots,X'_{m-1},X'_m,\dots,X'_{2m-1})$ of $\mathcal{R}^{(i)}_{k^{(i)}_m}$ with

$$X' \sim \mathrm{Dirichlet}(\beta,\dots,\beta,1-2\beta,\dots,1-2\beta),$$

$\mu^*_{k^{(i)}_m}(\mathcal{R}^{(i)}_{k^{(i)}_m})$, $(T_m, m\geq 1)$ and the i.i.d. random variables $M^{(j)}_m \sim \mathrm{ML}(\beta,\beta), j\geq 1$, are independent. Note that we do not consider the lengths of the m_0 external edges leading to the leaves of the reduced tree (2.92) and the initial m_0-1 internal edges, which do not affect the asymptotics. We will use the representation of a Dirichlet vector $X' \sim \mathrm{Dirichlet}(\beta,\dots,\beta,1-2\beta,\dots,1-2\beta)$ in terms of independent Gamma variables, that is,

$$X' \overset{d}{=} Y^{-1}(Y_1,\dots,Y_{m-1},Y'_1,\dots,Y'_m)$$

where the i.i.d. sequences $(Y_j, j\geq 1)$ and $(Y'_j, j\geq 1)$ are independent with $Y_1 \sim \mathrm{Gamma}(\beta,1)$, $Y'_1 \sim \mathrm{Gamma}(1-2\beta,1)$, and $Y = \sum_{j=1}^{m-1} Y_j + \sum_{j=1}^m Y'_j \sim \mathrm{Gamma}((m-1)(1-\beta)+1-2\beta,1)$. By (2.94),

$$C_i(m)\tilde{S}_{m_0,i} = C_i(m)\mu^*_{k^{(i)}_m}\left(\mathcal{R}^{(i)}_{k^{(i)}_m}\right)^\beta \left(\sum_{j=1}^{T_m+(m_0-1)}\left(X'_j\right)^\beta M^{(j)}_m + \sum_{j=0}^{m_0-1}\left(X'_{j+m}\right)^\beta \overline{M}^{(j)}_m\right)$$

where $\overline{M}_m^{(j)}$, $j \geq 1$, are i.i.d. with $\overline{M}_m^{(1)} \sim \mathrm{ML}(\beta, 1-2\beta)$ and independent of X' and T_m, $m \geq 1$, and hence $C_i(m)\tilde{S}_{m_0,i}$ has the same distribution as

$$T_m(1-\beta)^\beta m^{-\beta/(1-\beta)} \left(m^{-1} \left(\sum_{j=1}^{m-1} Y_j + \sum_{j=1}^{m} Y'_j \right) \right)^{-\beta} \left(T_m^{-1} \left(\sum_{j=1}^{T_m+(m_0-1)} Y_j^\beta M^{(j)} + \sum_{j=1}^{m_0} Y_j'^\beta \overline{M}^{(j)} \right) \right) \tag{2.95}$$

where the i.i.d. sequences $(Y_j, j \geq 1)$, $(Y'_j, j \geq 1)$, $(M^{(j)}, j \geq 1)$ and $(\overline{M}^{(j)}, j \geq 1)$, and the sequence $(T_m, m \geq 1)$ are independent with $M^{(1)} \sim \mathrm{ML}(\beta, \beta)$ and $\overline{M}^{(1)} \sim \mathrm{ML}(\beta, 1-2\beta)$.

By the strong law of large numbers, we have $\lim_{m\to\infty} T_m^{-1} \sum_{j=1}^{T_m} Y_j^\beta M^{(j)} = \mathbb{E}[Y_1^\beta M^{(j)}] = 1$ a.s. since $T_m \to \infty$ a.s., $\mathbb{E}[Y_1^\beta] = \Gamma(2\beta)/\Gamma(\beta)$, and where we use the first moment of the Mittag-Leffler distribution (2.6). Furthermore, note that $Y''_j := Y_j + Y'_j \sim \mathrm{Gamma}(1-\beta, 1)$, $j \in [m-1]$, are i.i.d., and hence $m^{-1}(\sum_{j=1}^{m-1} Y_j + \sum_{j=1}^{m} Y'_j) \to \mathbb{E}[Y''_1] = (1-\beta)$ a.s.. By (2.93), we conclude that the expression in (2.95) converges to S'_{m_0} a.s. where $S'_{m_0} \sim \mathrm{ML}(\beta', m_0 - \beta')$.

We already know that $\mathcal{R}_{k_m^{(i)}}^{(i)}$ and the scaling factor $C_i(m)$ converge almost-surely, and hence, by Lemma 2.53,

$$\lim_{m\to\infty} C_i(m) \mathcal{R}_{k_{m_0}^{(i)}}^{(i)} = \mathcal{F}_{m_0}^{(i)} \quad \text{a.s.}$$

where $\mathcal{F}_{m_0}^{(i)}$ is a Ford tree of index β'. Let $m_0 \to \infty$ to conclude (iii). □

2.5.4 The stable tree growth process

We use an equivalence relation $\sim$ on $(\mathcal{T}_k^*, (\mathcal{R}_k^{(i)}, i \geq 1))$ to contract each marked component $\mathcal{R}_k^{(i)}$, $i \geq 1$, of $\mathcal{T}_k^*$ to a single point, i.e.

$$x \sim y :\Leftrightarrow x, y \in \mathcal{R}_k^{(i)} \text{ for some } i \geq 1. \tag{2.96}$$

Note that, for all i with $\mathcal{R}_k^{(i)} \neq \{\rho\}$, $x, y \in \mathcal{R}_k^{(i)}$ implies $x, y \in \mathcal{R}_{k'}^{(i)}$ for all $k' \geq k$, and hence the equivalence relation $\sim$ is consistent as k varies. Denote the equivalence class related to $\mathcal{R}_k^{(i)}$ by $\tilde{v}_i := [\mathcal{R}_k^{(i)}]_\sim$, and let

$$\widetilde{\mathcal{T}}_k := \mathcal{T}_k^* / \sim \tag{2.97}$$

denote the set of equivalence classes of $\mathcal{T}_k^*, k \geq 0$. Furthermore, for $k \geq 0$, let $\tilde{\mu}_k$ be the push-forward of μ_k^* under the projection map from $\mathcal{T}_k^*$ onto $\widetilde{\mathcal{T}}_k$. This yields another formulation of the atom selection procedure on $\mathcal{T}_k^*$ in Algorithm 2.42.

Given $(\mathcal{T}_j^*, (\mathcal{R}_j^{(i)}, i \geq 1), \mu_j^*)$ and $r_k = \#\{i \geq 1 : \mathcal{R}_k^{(i)} \neq \{\rho\}\} = \#\{i \geq 1 : \tilde{v}_i \neq \{\rho\}\}$,

1. select $\tilde{J}_k$ from $\tilde{\mu}_k$;

2. if $\tilde{J}_k \neq \tilde{v}_i$ for all $i \in [r_k]$, set $J_k^* = \tilde{J}_k$; otherwise, if $\tilde{J}_k = \tilde{v}_i$ for some $i \in [r_k]$, sample an edge E_k^* of $\mathcal{R}_k^{(i)}$ proportionally to its mass $\mu_k^*(E_k^*)$; if E_k^* is an internal edge of $\mathcal{R}_k^{(i)}$, sample J_k^* from the normalised mass measure on E_k^*; if E_k^* is an external edge of $\mathcal{R}_k^{(i)}$, perform $(\beta, 1-2\beta)$-coin tossing sampling on E_k^* to determine $J_k^* \in E_k^*$..

It is this view on Algorithm 2.42 that we pursue further now. The following theorem contains the desired weight-length transformation, i.e. the branch point weights in Goldschmidt-Haas' stable line-breaking construction (Algorithm 2.4) are indeed as the lengths of the marked subtrees in the two-colour line-breaking construction (Algorithm 2.42). Its proof is given in the appendix.

Theorem 2.58. *Let the sequence* $(\mathcal{T}_k^*, (\mathcal{R}_k^{(i)}, i \geq 1), \mu_k^*, k \geq 0)$ *be as in Algorithm 2.42, and let*

$$\left(\widetilde{\mathcal{T}}_k, \left(\tilde{v}_i = [\mathcal{R}_k^{(i)}]_\sim, i \geq 1\right), \tilde{\mu}_k, k \geq 0\right)$$

be as in (2.97)*. Then the following hold.*

(i) The sequence of trees with mass measures from Algorithm 2.42 and (2.97) *has the same distribution as the sequence in Algorithm 2.9, i.e.*

$$\left(\widetilde{\mathcal{T}}_k, \tilde{\mu}_k, k \geq 0\right) \stackrel{d}{=} (\mathcal{T}_k, \mu_k, k \geq 0). \tag{2.98}$$

(ii) The sequence of trees with marked component lengths from Algorithm 2.42 and (2.97) *has the same distribution as the sequence of trees with weights from Algorithm*

2.4, i.e.

$$\left(\widetilde{\mathcal{T}}_k, \left(\widetilde{W}_k^{(i)}, i \geq 1\right), k \geq 0\right) \stackrel{d}{=} \left(\mathcal{T}_k, \left(W_k^{(i)}, i \geq 1\right), k \geq 0\right), \tag{2.99}$$

where $\widetilde{W}_k^{(i)} = \mathrm{Leb}(\mathcal{R}_k^{(i)})$ *is the length of* $\mathcal{R}_k^{(i)}, i \geq 1$, *respectively. In particular, the sequence* $(S_k^*, k \geq 0)$, *where* $S_k^* = \mathrm{Leb}(\mathcal{T}_k^*)$ *is the length of* $\mathcal{T}_k^*$, *evolves like the Mittag-Leffler Markov chain* $(S_k, k \geq 0)$ *with* $\theta = \beta$ *(cf. Section 2.2.3), i.e.*

$$(S_k^*, k \geq 0) \stackrel{d}{=} (S_k, k \geq 0).$$

Remark 2.59. We do not claim independence of the tree growth processes $(\mathcal{G}_m^{(i)}, \mu_m^{(i)}, m \geq 1), i \geq 1$, as in Theorem 2.57(i), and $(\widetilde{\mathcal{T}}_k, \tilde{\mu}_k, k \geq 0)$. Informally, if we see a long red component (as recorded in $\mathcal{G}_m^{(i)}$) of small mass (as recorded in $\widetilde{\mu}_{k_m^{(i)}}(\{\tilde{v}_i\})$), the (implicit) mass distribution will not be Poisson-Dirichlet, as is the case when averaging out according to the appropriate length distribution, but will be a Poisson-Dirichlet variable conditioned on having as its β-diversity the length of $\mathcal{G}_m^{(i)}$. In particular, a (sampling plus coin-tossing) pick from the conditioned Poisson-Dirichlet variable will not give the appropriately Beta distribution. But this is needed to obtain an independent factor to further scale the unmarked (β, β)-string of beads (which is also scaled by the second factor $\widetilde{\mu}_{k_m^{(i)}}(\{v_i\})$, which is independent of $\mathcal{G}_m^{(i)}$). However, we conjecture that $(\mathcal{G}_m^{(i)}, \mu_m^{(i)}, m \geq 1), i \geq 1$, are conditionally independent given the stable tree $(\widetilde{\mathcal{T}}, \widetilde{\mu})$. We believe that his can be proven using coagulation-fragmentation dualities of [58].

2.5.5 Convergence of two-colour trees

Theorems 2.57 and 2.58 demonstrate that the two-colour line-breaking construction naturally combines the stable tree growth process, and infinitely many rescaled subtree growth processes that build rescaled independent Ford CRTs. By Proposition 2.47, we can show that the tree growth process $(\mathcal{T}_k^*, k \geq 0)$ converges to a compact CRT $\mathcal{T}^*$, using the embedding of Algorithm 2.46.

Proposition 2.60 (Convergence of $(\mathcal{T}_k^*, \mu_k^*, k \geq 0)$)**.** *Let* $(\mathcal{T}_k^*, \mu_k^*, k \geq 0)$ *be the sequence of weighted* $\mathbb{R}$*-trees from Algorithm 2.46. Then, there is a compact CRT* $(\mathcal{T}^*, \mu^*)$ *such* *that*

$$\lim_{k\to\infty} d_{\mathrm{GH}}(\mathcal{T}_k^*, \mathcal{T}^*) = 0 \quad a.s.. \tag{2.100}$$

Furthermore, we have

$$\lim_{k\to\infty} d_{\mathrm{GHP}}((\mathcal{T}_k^*, \mu_k^*), (\mathcal{T}^*, \mu^*)) = 0 \quad a.s.. \tag{2.101}$$

Proof. We prove the claim for the sequence of weighted $\mathbb{R}$-trees $(\overline{\mathcal{T}}_k^*, \overline{\mu}_k^*, k \geq 0)$ embedded in a given $(\mathcal{T}^*, \mu^*)$ obtained in Lemma 2.41. Then (2.100) and (2.101) will follow from Proposition 2.47.

We follow the lines of the proof of [60] Proposition 22. Conditionally given $\mathcal{T}^*$, we show that for any $\epsilon > 0$ there is $k_1 \geq 0$ such that

- all edges of $\overline{\mathcal{T}}_{k_1}$ have length less than ϵ, and
- all connected components of $\mathcal{T}^* \setminus \overline{\mathcal{T}}_{k_1}$ have height less than ϵ.

Consider the set of connected components $\Upsilon_\epsilon := \{\overline{\mathcal{T}}^{(i)}, i \in I\}$ of

$$\{\sigma \in \mathcal{T}^* : \{\sigma' \in \mathcal{T}_\sigma^* : d(\sigma, \sigma') \geq \epsilon\} = \emptyset\}, \tag{2.102}$$

completed by their root vertices on $\mathcal{T}^*$, where

$$\mathcal{T}_\sigma^* := \{\sigma' \in \mathcal{T}^* : \sigma \in [[\rho, \sigma']]\}.$$

By Lemma 2.41 the tree $\mathcal{T}^*$ is compact a.s., and hence Υ_ϵ has finitely many elements. Fix some $\overline{\mathcal{T}}^{(i)} \in \Upsilon_\epsilon$, and recall the recursive structure (2.63)-(2.65) of $\mathcal{T}^*$ given via the strings of beads $\hat{\xi}_{\mathbf{i}}, \mathbf{i} \in \mathbb{N}$.

Clearly, $\hat{\xi}_{\mathbf{i}} \cap \overline{\mathcal{T}}^{(i)} \neq \emptyset$ for some $\mathbf{i} = i_1 i_2 \cdots i_d \in \mathbb{N}$. We show that $\overline{\mathcal{T}}^{(i)} \cap \overline{\mathcal{T}}_k^* = \emptyset$ for only finitely many k. To this end, we exploit the recursive structure of $\mathcal{T}^*$ and apply a Borel-Cantelli argument inductively.

Note that there exists $N_1 < \infty$ a.s. such that, for all $k \geq N_1$, $\hat{\xi}_{i_1} \subset \overline{\mathcal{T}}_k^*$. Indeed, we have $\mu^*(\hat{\xi}_{i_1}) = \overline{\mu}_0^*(x_{i_1}) = P_{i_1} \in (0,1)$ and

$$p_{1,k} := \mathbb{P}\Big(\hat{\xi}_{i_1} \cap \overline{\mathcal{T}}_k^* = \emptyset \;\Big|\; \big(\overline{\mathcal{T}}_0^*, \overline{\mu}_0^*\big)\Big) = (1 - P_{i_1})^k,$$

i.e. $\sum_{k=1}^{\infty} p_{1,k} < \infty$, and the first Borel-Cantelli Lemma yields the claim. In the following, we work conditionally given N_1.

- First, we show that there is $N_1 \le N_2 < \infty$ a.s. such that the atom $x_{i_1 i_2}$ of $\hat{\xi}_{i_1}$ is on an internal marked edge or an unmarked edge of $\overline{\mathcal{T}}_k^*$ for all $k \ge N_2$. For $k \ge N_1$, let E_k denote the edge of $\overline{\mathcal{T}}_k$ such that $x_{i_1 i_2} \in E_k$. Recall that the probability of selecting E_k is $\overline{\mu}_k^*(E_k)$, and that, conditionally given that E_k is an external marked edge,
$$\left(\overline{\mu}_k^*(E_k)^{-\beta} E_k, \overline{\mu}_k^*(E_k)^{-1} \overline{\mu}_k^* \restriction_{E_k}\right)$$
is a $(\beta, 1-2\beta)$-string of beads on which Algorithm 2.46 selects an atom via $(\beta, 1-2\beta)$-coin tossing sampling (Proposition 2.18), and that the switching probability function $p\colon [0,1] \to [0,1]$ is given by
$$p(u) = \frac{(1-2\beta)(1-u)}{(1-2\beta)(1-u) + \beta u}.$$
Note that $p(\cdot)$ is decreasing.

  The probability that E_{k+1} is an internal marked edge, given that E_k is an external marked edge, is bounded below by
$$\begin{aligned} &\overline{\mu}_k^*(E_k) \prod_{j \ge 1\colon\, x_{i_1 j} \le x_{i_1 i_2}, x_{i_1 j} \in E_k} \left(1 - p\left(\frac{P_{i_1 j}}{\overline{\mu}_k^*(E_k)}\right)\right) \\ &\quad \ge P_{i_1 i_2} \prod_{j \ge 1\colon\, x_{i_1 j} \le x_{i_1 i_2}, x_{i_1 j} \in E_{N_1}} \left(1 - p\left(\frac{P_{i_1 j}}{\overline{\mu}_{N_1}^*(E_{N_1})}\right)\right) > 0 \end{aligned} \tag{2.103}$$
as $E_k \subset E_{N_1}$, i.e. $\overline{\mu}_k^*(E_k) \le \overline{\mu}_{N_1}^*(E_{N_1})$, and note that (2.103) does not depend on k. The random time N_2 determining the first time when E_k is an internal edge of $\overline{\mathcal{T}}_k^*$ is hence bounded above by a geometric random variable, in particular it is finite a.s..

- Second, conditionally given N_1 and N_2, we show that there is $N_2 \le N_3 < \infty$ a.s. such that, for all $k \ge N_3$, $\hat{\xi}_{i_1 i_2} \subset \overline{\mathcal{T}}_k^*$. Since $\overline{\mu}_{N_2}^*(x_{i_1 i_2}) = P_{i_1} P_{i_1 i_2} \in (0,1)$, we have, for $k \geq N_2$,

$$p_{3,k} := \mathbb{P}\Big(\hat{\xi}_{i_1 i_2} \cap \overline{\mathcal{T}}_k^* = \emptyset \;\Big|\; \Big(\overline{\mathcal{T}}_{N_2}^*, \overline{\mu}_{N_2}^*\Big)\Big) \leq (1 - P_{i_1} P_{i_1 i_2})^{k-N_2},$$

and we apply Borel-Cantelli again.

We proceed inductively up to depth d, and conclude that there is $N_{2d+1} < \infty$ a.s. such that $\xi_{\mathbf{i}} \subset \overline{\mathcal{T}}_k^*$ for all $k \geq N_{2d+1}$, and therefore, $\overline{\mathcal{T}}^{(i)} \cap \overline{\mathcal{T}}_k^* \neq \emptyset$ for all $k \geq N_{2d+1}$. Since Υ_ϵ only contains finitely many components attaining height ϵ, we can find k_0 such that all connected components of $\mathcal{T}^* \setminus \overline{\mathcal{T}}_{k_0}$ have height less than ϵ, i.e. no subtrees of height ϵ can persist outside $\overline{\mathcal{T}}_k$ forever.

Now consider the tree $\overline{\mathcal{T}}_{k_0}$ with its marked components $\overline{\mathcal{R}}_{k_0}^{(i)}, i \geq 1$. This tree has at most $2k_0 - 1$ marked edges and at most $2k_0 + 1$ unmarked edges. Hence, at most $4k_0$ edges have length $\geq \epsilon$. Unmarked and internal marked edges are related to rescaled (β, β)-strings of beads with atoms being dense on these edges. Atoms of these (β, β)-strings of beads are sampled from the mass measure, and hence, for k large enough, all of these unmarked and internal marked edges will be split up into edges of lengths $< \epsilon$.

External marked edges are selected proportionally to mass, and are related to rescaled $(\beta, 1-2\beta)$-strings of beads on which atoms are selected via $(\beta, 1-2\beta)$-coin tossing sampling, cf. Proposition 2.18. We can use the same argument as for subtrees on external marked edges (i.e. we can show that for any atom x on an external marked edge of $\overline{\mathcal{T}}_{k_0}^*$, we can find $N < \infty$ a.s. such that x is an element of an internal edge of $\overline{\mathcal{T}}_k^*$ for all $k \geq N$), and conclude that there is $k_1 \geq k_0$ such that all edges of $\overline{\mathcal{T}}_{k_1}^*$ have length $< \epsilon$.

Note that each edge added in the growth procedure after step k_1 has length $< \epsilon$, since it is part of a subtree of $\mathcal{T}^* \setminus \overline{\mathcal{T}}_{k_1}^*$. Therefore, we conclude that for all $k \geq k_1$,

$$d_{\mathrm{GH}}\Big(\overline{\mathcal{T}}_k^*, \mathcal{T}^*\Big) < \epsilon, \tag{2.104}$$

and hence $\lim_{k\to\infty} d_{GH}(\overline{\mathcal{T}}_k^*, \mathcal{T}^*) = 0$ a.s., i.e. (2.100) follows.

The convergence in the Gromov-Hausdorff-Prokhorov sense follows directly from (2.104) since the mass measure $\overline{\mu}_k^*$ is the projection of μ^* onto $\overline{\mathcal{T}}_k^*$. We refer to the proof of [60] Corollary 23, for details of this argument. $\square$

Theorem 2.61 (Convergence of two-colour trees)**.** *Let* $(\mathcal{T}_k^*, (\mathcal{R}_k^{(i)}, i \geq 1), \mu_k^*, k \geq 0)$ *be the two-colour tree growth process from Algorithm 2.42 for some* $\beta \in (0, 1/2]$*. Then there exist a compact CRT* $(\mathcal{T}^*, \mu^*)$*, an i.i.d. sequence* $(\mathcal{F}^{(i)}, i \geq 1)$ *of Ford CRTs of index* $\beta' = \beta/(1-\beta)$ *and scaling factors* $(C_i, i \geq 1)$ *as in Theorem 2.57 such that*

$$\lim_{k\to\infty} d_{\mathrm{GH}}^{\infty}\Big(\Big(\mathcal{T}_k^*, \Big(\mathcal{R}_k^{(i)}, i \geq 1\Big)\Big), \Big(\mathcal{T}^*, \Big(C_i \mathcal{F}^{(i)}, i \geq 1\Big)\Big)\Big) = 0 \quad a.s..$$

Furthermore,

$$\lim_{k\to\infty} d_{\mathrm{GHP}}^{\infty}\Big(\Big(\mathcal{T}_k^*, \Big(\mathcal{R}_k^{(i)}, i \geq 1\Big), \mu_k^*\Big), \Big(\mathcal{T}^*, \Big(C_i \mathcal{F}^{(i)}, i \geq 1\Big)\Big), \mu^*\Big) = 0 \quad a.s..$$

Proof. This is a direct consequence of Proposition 2.60 and Theorem 2.57(iii). $\square$

2.6 Discrete two-colour tree growth processes

We discussed Marchal's discrete tree growth model yielding the stable tree in Algorithm 2.3. In a similar manner, the trees $(\mathcal{F}_m, m \geq 1)$ of a Ford tree growth process can be obtained as a.s. scaling limits of reduced trees of a (discrete) tree growth process, the so-called Ford alpha-model.

Both Marchal's model related to the stable tree and Ford's alpha-model that gives rise to Ford CRTs are contained as special cases in the alpha-gamma-model studied in [22]. We view discrete trees as $\mathbb{R}$-trees with unit edge lengths.

Definition 2.62 (The alpha-gamma-model)**.** Let $\alpha \in [0, 1]$ and $\gamma \in (0, \alpha]$. We grow discrete trees $T_n, n \geq 1$, as follows.

0. Let T_1 consist of two vertices connected by an edge.

Given $T_j, 0 \leq j \leq n$,

1. distribute a total weight of $n-\alpha$ by assigning $(d-2)\alpha-\gamma$ to each vertex of T_n of degree $d\geq 2$, $1-\alpha$ to each external edge of T_n, and γ to each internal edge of T_n; select a vertex or an edge in T_n at random according to these weights;

2. if an edge is selected, insert a new vertex, i.e. replace the selected edge by two edges connecting the new vertex to the vertices of the selected edge; proceed with the new vertex as the selected vertex;

3. in all cases, add a new edge from the selected vertex to a new degree-1 vertex to form T_{n+1}.

Note that the case $\gamma = 1-\alpha = \beta$ gives Marchal's model, Algorithm 2.3, where the case $\gamma=\alpha=\beta'$ is Ford's alpha-model. In the latter branch points get assigned weight 0 after their creation, i.e. the trees constructed in Ford's alpha model are binary.

Lemma 2.63 (Convergence of reduced trees, [22] Theorem 5)**.** *Let $(T_n, n\geq 1)$ be an alpha-gamma tree growth process for some $\alpha\in(0,1)$ and $\gamma\in(0,\alpha]$. For $k\geq 1$, consider the reduced tree $\mathcal{R}(T_n,\Sigma_1,\ldots,\Sigma_k)$ spanned by the root and the first k leaves $\Sigma_1,\ldots,\Sigma_k$, equipped with the graph distance on T_n, that is, for any edge $a\to b$ in $\mathcal{R}(T_n,\Sigma_1,\ldots,\Sigma_k)$ the number of edges between a and b in T_n. Then there exists an $\mathbb{R}$-tree $\mathcal{T}_k$ such that*

$$\lim_{n\to\infty} n^{-\gamma}\mathcal{R}(T_n,\Sigma_0,\ldots,\Sigma_k)=\mathcal{T}_k \quad a.s.$$

in the Gromov-Hausdorff topology. Furthermore, conditionally given that T_k has a total of $k+l$ edges, i.e. that T_k has l branch points, the edge lengths of $\mathcal{T}_k$ are given by $L_kV_k^{\gamma}D_k$ where $D_k\sim \mathrm{Dirichlet}((1-\alpha)/\gamma,\ldots,(1-\alpha)/\gamma,1,\ldots,1)$ with a weight of $(1-\alpha)/\gamma$ for each external edge, and weight 1 for each internal edge, and $L_k\sim \mathrm{ML}(\gamma, l\gamma+k(1-\alpha)), V_k\sim \mathrm{Beta}(k(1-\alpha)+l\gamma,(k-1)\alpha-l\gamma)$ are conditionally independent.

Note that in the stable case, $V_k\sim\mathrm{Beta}((k+l)(1-\alpha),(k-1-l)\alpha-l)$, and the total length is a $V_k^{1-\alpha}$ proportion of $L_k\sim\mathrm{ML}(1-\alpha,(k+l)(1-\alpha))$, which is uniformly distributed amongst the $k+l$ edges. In Ford's model, we have $l=k-1$, and we distribute the "full" length $L_k\sim\mathrm{ML}(\alpha,k-\alpha)$ according to a Dirichlet variable D_k with a parameter of $1/\alpha-1$ for each external edge and parameter 1 for each internal edge.

In a similar manner, we can obtain the two-colour trees $(\mathcal{T}_k^*, (\mathcal{R}_k^{(i)}, i \geq 1))$, $k \geq 0$, as a.s. scaling limits of the following discrete tree growth process in the space of ∞-marked $\mathbb{R}$-trees with unit edge lengths.

Definition 2.64 (The discrete two-colour model)**.** Let $\beta \in (0, 1/2]$. We grow discrete two-colour trees $(T_n^*, (R_n^{(i)}, i \geq 1))$, $n \geq 0$, as follows.

0. Let T_0 consist of the root ρ and another degree-1 vertex connected by an edge, let $R_0^{(i)} = \{\rho\}$, $i \geq 1$, and $r_0 = 0$.

Given $(T_j^*, (R_j^{(i)}, i \geq 1))$, $0 \leq j \leq n$, and $r_k = \#\{i \geq 1 : R_n^{(i)} \neq \{\rho\}\}$,

1. distribute a total weight of $n + \beta$ by assigning β to each unmarked and each internal marked edge of T_n, and $1 - 2\beta$ to each external marked edge of T_n; select an edge in T_n at random according to these weights;

2. if the selected edge is unmarked, replace it by two unmarked edges connecting the new vertex to the vertices of the selected edge; if the selected edge is a marked edge of $R_n^{(i)}$ for some $i \geq 1$, replace it by two marked edges; proceed with the new vertex as the selected vertex;

3. add a new degree-2 vertex and connect it to the selected vertex by a marked edge, and to a new degree-1 vertex by an unmarked edge; add the marked edge to $R_n^{(i)}$ if the selected vertex is an element of $R_n^{(i)}$, otherwise add the edge to $R_n^{(r_k+1)}$ and remove ρ from $R_n^{(r_k+1)}$.

Proposition 2.65 (Convergence of the discrete two-colour model)**.** *Consider the discrete two-colour tree growth process $(T_n^*, (R_n^{(i)}, i \geq 1), n \geq 0)$ from Definition 2.64, which we view as a sequence of ∞-marked $\mathbb{R}$-trees with unit edge lengths. For $k \geq 0$, let*

$$\mathcal{R}\Big(T_n^*, \Big(R_n^{(i)}, i \geq 1\Big), \Sigma_0, \ldots, \Sigma_k\Big)$$

denote the reduced tree spanned by the root ρ and the first $k+1$ leaves $\Sigma_0, \ldots, \Sigma_k$. Then

$$\lim_{n \to \infty} n^{-\beta} R\Big(T_n^*, \Big(R_n^{(i)}, i \geq 1\Big), \Sigma_0, \ldots, \Sigma_k\Big) = \Big(\mathcal{T}_k^*, \Big(\mathcal{R}_k^{(i)}, i \geq 1\Big)\Big) \quad a.s.$$

with respect to the distance $d_{\rm GH}^{\infty}$ *defined in* (2.21)*.*

Conditionally given that T_k^* *has* r_k *marked components* $R_k^{(i)} \neq \{\rho\}$ *with* $d_1 - 2, \ldots, d_{r_k} - 2$ *leaves, the distribution of the edge lengths of* $(\mathcal{T}_k^*, (\mathcal{R}_k^{(i)}, i \geq 1))$ *is given by* $S_k^* D_k$ *where* $S_k^* \sim \mathrm{ML}(\beta, \beta + k)$ *and* $D_k \sim \mathrm{Dirichlet}(1, \ldots, 1, 1/\beta - 2, \ldots, 1/\beta - 2)$ *with a parameter of* 1 *for each unmarked edge and each internal marked edge, and a parameter of* $1/\beta - 2$ *for each external marked edge, are conditionally independent.*

The proof of Proposition 2.65 is based on exactly the same techniques as the proof of the corresponding result for the alpha-gamma model, cf. [22] Proposition 21 and Proposition 22, and the result for (α, θ)-tree growth processes, cf. [60] Proposition 14.

Remark 2.66. One can obtain the mass measures μ_k^*, $k \geq 0$, as the scaling limits of the empirical probability measures on the leaves of T_n^*, projected onto the reduced trees, using the same methods as in [60]. In particular, each edge equipped with limiting relative projected subtree masses is a rescaled (β, θ)-string of beads where $\theta = \beta$ for internal marked and unmarked edges, and $\theta = 1 - 2\beta$ for external marked edges. It can be shown that these (β, θ)-string of beads are independent of each other and of the mass split on $\mathcal{T}_k^*$ which has distribution $\mathrm{Dirichlet}(\beta, \ldots, \beta, 1 - 2\beta, \ldots, 1 - 2\beta)$, with parameter β for each internal marked and each unmarked edge, and parameter $1 - 2\beta$ for each external marked edge of $\mathcal{T}_k^*$.

2.7 Appendix

We present the proofs postponed from earlier parts of this chapter.

Proof of Theorem 2.57(i). Consider a space $\mathbb{T}_{[m]}$ of weighted discrete $\mathbb{R}$-trees $(\mathcal{T}, \mu)$ with m leaves labelled by $[m]$ and mass measure μ of total mass $\mu(\mathcal{T}) \in (0, 1]$, $m \geq 1$, see e.g. [60] Section 3.3. for a formal introduction. We consider the transition kernels $\kappa_m \colon \mathbb{T}_{[m]} \to \mathbb{T}_{[m+1]}$, $m \geq 1$, defined as follows. Given any $(\mathcal{T}, \mu) \in \mathbb{T}_{[m]}$,

- select an edge E of $\mathcal{T}$ according to the normalised mass measure $\mu(\mathcal{T})^{-1}\mu$; given E, select an atom J of $\mu \restriction_E$ according to (β, θ)-coin tossing sampling where $\theta = \beta$ if E is internal, and $\theta = 1 - 2\beta$ if E is external; this determines a selection probability $p_m(x)$ for each atom $x \in \mathcal{T}$;

- given J, let $\gamma \sim \text{Beta}(1-2\beta, \beta)$ be independent, and attach to J an independent $(\beta, 1-2\beta)$-string of beads with mass measure rescaled by $\gamma\mu(J)$ and metric rescaled by $(\gamma\mu(J))^{\beta}$, and label the new leaf by $m+1$.

We use the convention that if no atom is selected, we apply a scaling factor of 0. Note that, in our setting with (β,β)-strings of beads on internal edges and $(\beta, 1-2\beta)$-strings of beads on external edges, this does not happen almost surely.

Denote by $\kappa_m((\mathcal{T},\mu),\cdot)$ the distribution of the resulting tree. We further consider the kernel $\kappa_0(\cdot) = \kappa_0((\{\rho\},\delta_\rho),\cdot)$ taking the singleton tree $\{\rho\}$ of mass 1, and associating a $(\beta, 1-2\beta)$-string of beads with $\{\rho\}$. We will show that each process in (2.89) evolves according to the transition kernels $\kappa_m, m \geq 1$, starting from an independent $(\beta, 1-2\beta)$-string of beads whose distribution is given by $\kappa_0(v_i)$.

More formally, for $l \geq 1$ and some $m_i \geq 1$, $i \in [l]$, we will show that

$$\begin{aligned}&\mathbb{E}\left[\prod_{i\in[l]} f_i\Big(\Big(\mathcal{G}_m^{(i)}, \mu_m^{(i)}\Big), m \in [m_i]\Big)\right]\\ &= \prod_{i\in[l]} \int\int\cdots\int f_i(R_1,\ldots,R_{m_i})\kappa_{m_i-1}(R_{m_i-1}, dR_{m_i})\cdots\kappa_1(R_1,dR_2)\kappa_0(dR_1) \quad (2.105)\end{aligned}$$

for any bounded continuous functions $f_i\colon \mathbb{T}_{[1]}\times\cdots\times\mathbb{T}_{[m_i]} \to \mathbb{R}$, $i \in [l]$.

We first show the equation (2.105) for $l=1$. For notational convenience, we write $(\mathcal{G}_m, \mu_m) = (\mathcal{G}_m^{(1)}, \mu_m^{(1)})$ and $f = f_1$. We further use the notation $\xi_{\beta,\beta}$ and $\xi_{\beta,1-2\beta}$ for (β,β)- and $(\beta, 1-2\beta)$-strings of beads, respectively, and recall that we denote by $p_m(x)$ the selection probability of $x \in \mathcal{T}$ for $\mathcal{T} \in \mathbb{T}_{[m]}$ using the edge selection rule in combination with coin tossing sampling, as described above. $B_{\beta,1-2\beta}(\cdot)$ denotes the density of $\text{Beta}(\beta, 1-2\beta)$. We obtain,

$$\begin{aligned}\mathbb{E}[f(\mathcal{G}_1,\ldots,\mathcal{G}_{m_1})] = &\sum_{k_1^{(1)}=1,k_2^{(1)},\ldots,k_{m_1}^{(1)}} \int_{\xi_0} \sum_{v\in\xi_0} \mu_0(v) \int_{x_1} B_{\beta,1-2\beta}(x_1) \int_{\xi_1}\\ &(1-\mu_0(v)(1-\overline{x}_1))^{k_2^{(1)}-k_1^{(1)}-1}\mu_0(v)(1-\overline{x}_1) \sum_{w_1\in R_1} p_1(w_1) \int_{x_2} B_{\beta,1-2\beta}(x_2)\int_{\xi_2}\cdots\\ &\left(1-\mu_0(v)\prod_{i\in[m_1-1]}(1-\overline{x}_i)\right)^{k_{m_1}^{(1)}-k_{m_1-1}^{(1)}-1} \mu_0(v)\prod_{i\in[m_1-1]}(1-\overline{x}_i)\end{aligned}$$

$$\times \sum_{w_{m_1-1}\in R_{m_1-1}} p_{m_1-1}(w_{m_1-1}) \int_{x_{m_1}} B_{\beta,1-2\beta}(x_{m_1}) \int_{\xi_{m_1}} f(R_1,\ldots,R_{m_1})$$

$$\times \mathbb{P}(\xi_{\beta,1-2\beta}\in d\xi_{m_1})dx_{m_1}\cdots\mathbb{P}(\xi_{\beta,1-2\beta}\in d\xi_2)dx_2\mathbb{P}(\xi_{\beta,1-2\beta}\in d\xi_1)dx_1$$

$$\times \mathbb{P}(\xi_{\beta,\beta}\in d\xi_0)$$

where

- μ_0 is the mass measure of ξ_0;
- $R_1=\xi_1$ with mass measure $\mu_1^{(1)}$ is the initial string of beads, and, for $m\geq 2$, R_m with mass measure $\mu_m^{(1)}$ is created by attaching to $w_{m-1}\in R_{m-1}$ the string of beads ξ_m rescaled by the proportion x_{m-1} of the mass of w_{m-1};
- the sequence $(\overline{x}_i, i\geq 1)$ is defined by

$$\overline{x}_1=x_1,\qquad \overline{x}_i=1-\frac{\mu_{i-1}^{(1)}(w_{i-1})}{\mu_{i-1}^{(1)}(R_{i-1})}(1-x_i),\quad i=2,\ldots,m_1;$$

- the integrals are taken over the whole ranges of $x_i\in[0,1]$ and the subspaces of (β,β)- and $(\beta,1-2\beta)$-strings of beads $\xi_i\in\mathbb{T}_{[1]}$.

Note that

$$\mu_0(v)\prod_{i\in m-1}(1-\overline{x}_i)$$

is the relative remaining mass of the first marked component after m transition steps have been carried out in this component.

We can move the sum over $k_1^{(1)},\ldots,k_{m_1}^{(1)}$ inside the integrals, and note that there is only one term which depends on $k_{m_1}^{(1)}$. Moving the sum over $k_{m_1}^{(1)}$ in front of this factor, we obtain

$$\sum_{k_{m_1}^{(1)}\geq k_{m_1-1}^{(1)}+1}\left(1-\mu_0(v)\prod_{i\in[m_1-1]}(1-\overline{x}_i)\right)^{k_{m_1}^{(1)}-k_{m_1-1}^{(1)}-1}\mu_0(v)\prod_{i\in[m_1-1]}(1-\overline{x}_i)=1$$

as this is the sum over the probability mass function of a geometric random variable (there are infinitely many insertions into the first marked component almost surely). We can proceed inductively and sum the corresponding geometric probabilities over

$k_1^{(1)}, \ldots, k_{m_1-1}^{(1)}$ to obtain

$$\mathbb{E}[f(\mathcal{G}_1, \ldots, \mathcal{G}_{m_1})] = \int_{\xi_0} \sum_{v \in \xi_0} \mu_0(v) \int_{x_1} B_{\beta,1-2\beta}(x_1) \int_{\xi_1} \sum_{w_1 \in R_1} p_1(w_1)$$
$$\int_{x_2} B_{\beta,1-2\beta}(x_2) \int_{\xi_2} \cdots \sum_{w_{m_1-1} \in R_{m_1-1}} p_{m_1-1}(w_{m_1-1}) \int_{x_{m_1}} B_{\beta,1-2\beta}(x_{m_1}) \int_{\xi_{m_1}}$$
$$f(R_1, \ldots, R_{m_1}) \mathbb{P}(\xi_{\beta,1-2\beta} \in d\xi_{m_1}) dx_{m_1} \cdots \mathbb{P}(\xi_{\beta,1-2\beta} \in d\xi_2) dx_2$$
$$\mathbb{P}(\xi_{\beta,1-2\beta} \in d\xi_1) dx_1 \mathbb{P}(\xi_{\beta,\beta} \in d\xi_0).$$

We can now take the sum $\sum_{v \in \xi_0} \mu_0(v) = 1$ and the outer integral, as the inner terms are independent of $\mu_0(v)$ and ξ_0. This results in

$$\mathbb{E}[f(\mathcal{G}_1, \ldots, \mathcal{G}_{m_1})] = \int_{x_1} B_{\beta,1-2\beta}(x_1) \int_{\xi_1} \sum_{w_1 \in R_1} p_1(w_1) \int_{x_2} B_{\beta,1-2\beta}(x_2) \int_{\xi_2}$$
$$\cdots \sum_{w_{m_1-1} \in R_{m_1-1}} p_{m_1-1}(w_{m_1-1}) \int_{x_{m_1}} B_{\beta,1-2\beta}(x_{m_1}) \int_{\xi_{m_1}} f(R_1, \ldots, R_{m_1})$$
$$\mathbb{P}(\xi_{\beta,1-2\beta} \in d\xi_{m_1}) dx_{m_1} \cdots \mathbb{P}(\xi_{\beta,1-2\beta} \in d\xi_2) dx_2 \mathbb{P}(\xi_{\beta,1-2\beta} \in d\xi_1) dx_1.$$

We recognise the definition of the transition kernels $\kappa_m, m \geq 1$, and rewrite this integral in the form

$$\mathbb{E}[f(\mathcal{G}_1, \ldots, \mathcal{G}_{m_1})]$$
$$= \int \int \cdots \int f(R_1, \ldots, R_m) \kappa_{m-1}(R_{m-1}, dR_m) \cdots \kappa_1(R_1, dR_2) \kappa_0(dR_1)$$

To see (2.105) for $l \geq 2$, we express the left-hand side in terms of the distribution of $(\mathcal{T}_0^*, \mu_0^*)$ and the two-colour transition kernels, which can be described via Algorithm 2.42, as a sum over $k_j^{(i)}, j \in [m_i], i \in [l]$. Then we can proceed as follows.

- First integrate out irrelevant transitions which affect components $i \geq l+1$ and parts of earlier transitions such as unmarked strings of beads after the creation of the l-th component. These transitions do not affect the marked components $i \in [l]$.

- Move the sums over $k^{(l)}_{m_l},\dots,k^{(l)}_2$ inside the integrals. Notice that there is only one term depending on $k^{(l)}_{m_l}$, i.e. we obtain the sum over $k^{(l)}_{m_l} \geq k^{(l)}_{m_l-1}$, $k^{(l)}_{m_l} \neq k^{(j)}_i, j \in [m_i], i \in [l-1]$ of the probabilities of selecting the l-th marked component at step $k^{(l)}_{m_l}$, provided that no insertion into other marked components $i \in [l-1]$ happens at this time, that is,

$$\sum_{k^{(l)}_{m_l} \geq k^{(l)}_{m_l-1}+1, k^{(l)}_{m_l} \neq k^{(j)}_i, j\in[m_i], i\in[l-1]} \left(1-\mu_{k^{(l)}_1-1}(v_l)\prod_{i\in[m_l-1]}\left(1-\overline{x}^{(l)}_i\right)\right)^{k(m,l)} \mu_{k^{(l)}_1-1}(v_l)\prod_{i\in[m_l-1]}\left(1-\overline{x}^{(l)}_i\right)$$

where

$$k(m,l) := k^{(l)}_{m_l} - k^{(l)}_{m_l-1} - \#\left\{k^{(l)}_{m_l-1} < k < k^{(l)}_{m_l} : k = k^{(j)}_i, j \in [m_i], i \in [l-1]\right\},$$

and where the sequences $(x^{(l)}_i, i \geq 1)$ and $(\overline{x}^{(l)}_i, i \geq 1)$ are defined as $(x_i, i \geq 1)$ and $(\overline{x}_i, i \geq 1)$, respectively. Note that

$$\mu_{k^{(l)}_1-1}(v_l)\prod_{i\in[m-1]}\left(1-\overline{x}^{(l)}_i\right)$$

is the mass of the l-th marked component after m transition steps have been carried out in this component.

As we have a sum over the probability mass function of a geometric random variable, no matter when insertions into components $i \in [l-1]$ happen, this sum is 1. We can proceed inductively down to $k^{(l)}_2$.

- The sum over the insertion point v_l is just a sum over the bead selection probabilities $\mu_{k^{(l)}_1-1}(v_l)$, $k^{(l)}_1 \geq k^{(l-1)}_1 + 1$, which sum to the probability of creating the l-th component (no matter what the sizes of the other components are at this step). The sum over $k^{(l)}_1$ is not geometric but it is a sum over the probabilities of success in a Bernoulli sequence with increasing success probability. This sum is again 1 (as we will open the l-th marked component with probability one).

- We can put the integrals over the ingredients for the l-th subtree growth process in front of the other integrals, as they do not depend on anything else.

- Inductively, for $j = l-1, \dots, 1$, repeat these steps to lose all sums over insertion times $k_i^{(j)}$ and first insertion points v_i, $i \in [l]$.

- Finally, the integrand of the outer integral over the distribution of ξ_0 is constant, so the integral can be dropped. We obtain precisely the product form of the right-hand side (2.105). □

The proofs of Theorem 2.58(i), Proposition 2.47 and Proposition 2.52 are based on coupling arguments and are quite similar to one another. We present the proof of Theorem 2.58(i) first.

Proof of Theorem 2.58(i). Recall the constructions of $(\mathcal{T}_k^*, \mu_k^*)$ in Algorithm 2.42 and $(\widetilde{\mathcal{T}}_k, \tilde{\mu}_k)$ in (2.97). We use the following coupling to construct $(\mathcal{T}_k, \mu_k, k \geq 0)$ from $(\mathcal{T}_k^*, \mu_k^*, k \geq 0)$ with the same distribution as in Algorithm 2.9.

- Let the trees with mass measures coincide initially, i.e. consider the same isometric copy of a (β, β)-string of beads given by

$$(\mathcal{T}_0, \mu_0) = \left(\widetilde{\mathcal{T}}_0, \tilde{\mu}_0\right) = (\mathcal{T}_0^*, \mu_0^*).$$

- Supposing that $(\mathcal{T}_k, \mu_k) = (\widetilde{\mathcal{T}}_k, \tilde{\mu}_k)$ for some $k \geq 0$, set
  - $J_k := \tilde{J}_k = [J_k^*]_\sim$;
  - $\xi_k = \xi_k^{(2)}$ (i.e. consider the same (β, β)-strings of beads); and
  - $Q_k := (1-\gamma_k)\mu_k^*(J_k^*)/\tilde{\mu}_k(\tilde{J}_k)$, where we recall that $(1-\gamma_k) \sim \mathrm{Beta}(\beta, 1-2\beta)$ is the independent scaling factor for $\xi_k^{(2)}$ in the construction of the independent β-mixed string of beads $\Psi(\xi_k^{(1)}, \xi_k^{(2)}, \gamma_k) \sim \nu_\beta$ in Algorithm 2.42. Note that, if the selected atom J_k^* is an element of a marked component, Q_k is the proportion of the mass of J_k^* which is added to this marked component in form of a rescaled independent $(\beta, 1-2\beta)$-string of beads, i.e. a proportion of $1 - Q_k$ is split into an unmarked rescaled (β, β)-string of beads.

Since $\tilde{J}_k$ was sampled from $\tilde{\mu}_k$, J_k is sampled from μ_k, as required for Algorithm 2.9. It remains to check that the scaling factor $Q_k\tilde{\mu}_k(\tilde{J}_k)$ induced by Algorithm 2.42 for the (β,β)-string of beads $\xi_k = \xi_k^{(2)}$ used in the attachment procedure is as needed for Algorithm 2.9. We work conditionally given that $\widetilde{\mathcal{T}}_k$ has l branch points $\tilde{v}_i$ of sizes $d_i = \deg(\tilde{v}_i, \widetilde{\mathcal{T}}_k)$, $i \in [l]$, respectively.

- If $\tilde{J}_k \neq \tilde{v}_i$ for $i \in [l]$, then $J_k = \tilde{J}_k = J_k^*$, and a new branch point $\tilde{J}_k$ of degree $\deg(\tilde{J}_k, \widetilde{\mathcal{T}}_{k+1}) = 3 = 1 + \deg(\tilde{J}_k + 1, \widetilde{\mathcal{T}}_k)$ is created. The mass $\mu_k^*(J_k^*) = \tilde{\mu}_k(\tilde{J}_k)$ is split by the independent random variable $\gamma_k \sim \mathrm{Beta}(1-2\beta, \beta)$ into a branch point weight $\tilde{\mu}_{k+1}(\tilde{J}_k) = \gamma_k\tilde{\mu}_k(\tilde{J}_k)$ and the rescaled isometric copy of the (β,β)-string of beads $\xi_k^{(2)} = \xi_k$, i.e. $\tilde{\mu}_k(\tilde{J}_k)(1-\gamma_k) = \tilde{\mu}_k(\tilde{J}_k)Q_k$ where $Q_k \sim \mathrm{Beta}(\beta, 1-2\beta)$ is conditionally independent of ξ_k and $(\mathcal{T}_k, \mu_k, J_k)$ given $\deg(J_k, \mathcal{T}_k) = 2$, as required.
- If $\tilde{J}_k = \tilde{v}_i$ of degree $\deg(\tilde{v}_i, \widetilde{\mathcal{T}}_k) = d_i$ for some $i \in [l]$, we first select an edge E_k^* of $\mathcal{R}_k^{(i)}$ from μ_k^* restricted to $\mathcal{R}_k^{(i)}$. Conditionally given that E_k^* has been selected, we choose $J_k^* \in E_k^*$ according to (β,θ)-coin tossing sampling, where $\theta = \beta$ if E_k^* is an internal edge of $\mathcal{R}_k^{(i)}$, and $\theta = 1-2\beta$ otherwise. Hence, by Proposition 2.18 and Proposition 2.14(iii)-(iv), conditionally given $J_k^* \in E_k^*$, the relative mass split in $\mathcal{R}_k^{(i)}$ is

  $$\mathrm{Dirichlet}(\beta, \ldots, \beta, 1-2\beta, \ldots, 1-2\beta, \beta, 1-\beta, \theta)$$

  with parameter β for each non-selected internal edge of $\mathcal{R}_k^{(i)}$, $1-2\beta$ for each non-selected external edge of $\mathcal{R}_k^{(i)}$, β for the part of E_k^* closer to the root, θ for the other part of E_k^*, and $1-\beta$ for the atom J_k^*. In any case (i.e. no matter if E_k^* is internal or external), we get by Proposition 2.14(i)-(ii) that, conditionally given $\tilde{J}_k = \tilde{v}_i$,

  $$\mu_k^*(J_k^*)/\tilde{\mu}_k(\tilde{J}_k) \sim \mathrm{Beta}(1-\beta, (d_i-2)(1-\beta))$$

  is independent of $\tilde{\mu}_k(\tilde{J}_k)$, as the internal relative mass split in $\mathcal{R}_k^{(i)}$ is independent of its total mass, see Proposition 2.54. Overall, still conditionally given $\tilde{J}_k = \tilde{v}_i$, we have that

  $$\mu_k^*(J_k^*)(1-\gamma_k) = (1-\gamma_k)\Big(\mu_k^*(J_k^*)\tilde{\mu}_k(\tilde{J}_k)^{-1}\Big)\tilde{\mu}_k(\tilde{J}_k) = Q_k\tilde{\mu}_k(\tilde{J}_k)$$

where $Q_k \sim \text{Beta}(\beta, d_i(1-\beta)-1)$, as is easily checked using Proposition 2.14(i)-(iii). Note that Q_k is independent of $\tilde{\mu}_k(\tilde{J}_k)$. This is due to the fact that the mass split within $\mathcal{R}_k^{(i)}$, and the mass split between the edges of $\widetilde{\mathcal{T}}_k$ and its branch points are conditionally independent given l branch points $\tilde{v}_i$ with $\deg(\tilde{v}_i, \widetilde{\mathcal{T}}_k) = d_i$, $i \in [l]$, i.e. $\gamma_k, \mu_k^*(J_k^*)/\tilde{\mu}_k(\tilde{J}_k)$ and $\tilde{\mu}_k(\tilde{J}_k)$ are independent. $\square$

Proof of Proposition 2.47. Similarly to the proof of Theorem 2.58(i), let the initial weighted ∞-marked $\mathbb{R}$-trees coincide, i.e. let $(\mathcal{T}_0^*, (\mathcal{R}_0^{(i)}, i \geq 1), \mu_0^*) = (\overline{\mathcal{T}}_0^*, (\overline{\mathcal{R}}_0^{(i)}, i \geq 1), \overline{\mu}_0^*)$. Then, $(\mathcal{T}_0^*, \mu_0^*)$ is a (β,β)-string of beads, and $\mathcal{R}_0^{(i)} = \{\rho\}$ for all $i \geq 1$, as required in Algorithm 2.42.

Supposing that $(\mathcal{T}_k^*, (\mathcal{R}_k^{(i)}, i \geq 1), \mu_k^*) = (\overline{\mathcal{T}}_k^*, (\overline{\mathcal{R}}_k^{(i)}, i \geq 1), \overline{\mu}_k^*)$ for some $k \geq 0$, set $J_k^* := \overline{J}_k^*$, $I_k := \overline{I}_k$, and, conditionally given $\overline{I}_k = x_{\mathbf{i}j}$ and $\overline{\Omega}_k$, set $(E_k^+, \mu_k^+) = \xi_{\mathbf{i}j}$ and $R_k^+ = [[\overline{J}_k^*, \overline{\Omega}_k]]$. We need to check that the induced update step from $(\mathcal{T}_k^*, (\mathcal{R}_k^{(i)}, i \geq 1), \mu_k^*)$ to $(\mathcal{T}_{k+1}^*, (\mathcal{R}_{k+1}^{(i)}, i \geq 1), \mu_{k+1}^*)$ is as required in Algorithm 2.42.

To select $\overline{J}_k^*$ in Algorithm 2.46, we first select an edge $\overline{E}_k^*$ of $\mathcal{T}_k^*$ proportionally to $\overline{\mu}_k^*(\overline{E}_k^*)$, and perform $(\beta, 1-2\beta)$-coin tossing if $\overline{E}_k^*$ is an external marked edge, and uniform sampling from $\overline{\mu}_k \restriction_{\overline{E}_k^*}$ otherwise, i.e. J_k is sampled from μ_k as required in Algorithm 2.42, and $\mu_k^*(J_k^*) = \overline{\mu}_k^*(\overline{J}_k^*)$. In particular, $\overline{I}_k = I_k$.

Furthermore, (E_k^+, R_k^+, μ_k^+) is an independent β-mixed string of beads, which is a direct consequence of the definition of $\xi_{\mathbf{i}j}$ from the recursive construction of $\mathcal{T}^*$ in (2.62)-(2.65), and the transition kernel $\kappa(\xi_{\mathbf{i}j}, \cdot)$. Therefore,

$$\left(\left(\mathcal{T}_k^*, \left(\mathcal{R}_k^{(i)}, i \geq 1\right), \mu_k^*\right), \left(\mathcal{T}_{k+1}^*, \left(\mathcal{R}_{k+1}^{(i)}, i \geq 1\right), \mu_{k+1}^*\right)\right)$$

has the same distribution as

$$\left(\left(\mathcal{T}_k^*, \left(\mathcal{R}_k^{(i)}, i \geq 1\right), \mu_k^*\right), \left(\overline{\mathcal{T}}_{k+1}^*, \left(\overline{\mathcal{R}}_{k+1}^{(i)}, i \geq 1\right), \overline{\mu}_{k+1}^*\right)\right),$$

which proves Proposition 2.47, as Algorithm 2.42 and Algorithm 2.46 specify Markov chains. $\square$

Proof of Proposition 2.52. We denote the sequence of trees of (2.2) by $(\widehat{\mathcal{T}}, \hat{\mu}_k, k \geq 0)$, which is obtained by sampling leaves $\Sigma_k, k \geq 0$, independently from the mass measure μ

of a stable tree $(\mathcal{T}, d, \rho)$ where $\hat{\mu}_k := \pi_*^{\widehat{\mathcal{T}}_k}\mu$ is the projected mass measure on $\widehat{\mathcal{T}}_k$, $k \geq 0$. $(\mathcal{T}_k, \mu_k, k \geq 0)$ denotes the sequence of Algorithm 2.9, where $(J_k, k \geq 0)$ is the sequence of attachment points, and $(Q_k, k \geq 0)$ is the sequence of update random variables.

The coupling is as follows. Set $(\mathcal{T}_0, \mu_0) = (\widehat{\mathcal{T}}_0, \hat{\mu}_0)$, and, given $(\mathcal{T}_k, \mu_k) = (\widehat{\mathcal{T}}_k, \hat{\mu}_k)$ for some $k \geq 0$, set $J_k := \hat{J}_k$ where

$$\hat{J}_k := \arg\inf\Big\{d(\rho, x) : x \in \widehat{\mathcal{T}}_{k+1} \setminus \widehat{\mathcal{T}}_k\Big\},$$

let $Q_k = 1 - \hat{\mu}_{k+1}(\hat{J}_k)/\hat{\mu}_k(\hat{J}_k)$, and

$$\xi_k := \left(\hat{\mu}_{k+1}\Big(\widehat{\mathcal{T}}_{k+1} \setminus \widehat{\mathcal{T}}_k\Big)^{-\beta}\widehat{\mathcal{T}}_{k+1} \setminus \widehat{\mathcal{T}}_k, \hat{\mu}_{k+1}\Big(\widehat{\mathcal{T}}_{k+1} \setminus \widehat{\mathcal{T}}_k\Big)^{-1}\hat{\mu}_{k+1} \upharpoonright_{\widehat{\mathcal{T}}_{k+1}\setminus\widehat{\mathcal{T}}_k}\right).$$

By Proposition 2.11, $(\hat{\mathcal{T}}_0, \hat{\mu}_0)$ is a (β, β)-string of beads, as required in Algorithm 2.9. Now assume that $(\mathcal{T}_k, \mu_k) = (\widehat{\mathcal{T}}_k, \hat{\mu}_k)$ for some $k \geq 0$ with the distribution claimed in Proposition 2.52. Denote the connected components of $\mathcal{T} \setminus \widehat{\mathcal{T}}_k$ by $\mathcal{S}_i, i \in I$, completed by their root vertices $\rho_i \in \widehat{\mathcal{T}}_k$, $i \in I$, respectively. Note that $\hat{\mu}_k(\rho_i) = \sum_{j:\rho_j=\rho_i} \mu(\mathcal{S}_j)$.

Since we sample Σ_{k+1} from the mass measure μ on $\mathcal{T}$, the conditional probability that $\Sigma_{k+1} \in \mathcal{S}_i$, given $(\mathcal{T}, \mu)$, $(\widehat{\mathcal{T}}_k, \hat{\mu}_k)$ and $(\mathcal{S}_i, i \in I)$, is $\mu(\mathcal{S}_i) = \hat{\mu}_k(\rho_i)(\mu(\mathcal{S}_i)/\hat{\mu}_k(\rho_i))$, i.e. we can sample $\hat{J}_k$ in two steps: first select one of the atoms ρ_i of $\widehat{\mathcal{T}}_k$ proportionally to their mass $\hat{\mu}_k(\rho_i)$, and then we select one of the components $\mathcal{S}_j$ with root $\rho_j = \rho_i$ proportionally to relative mass $\mu(\mathcal{S}_j)/\hat{\mu}_k(\rho_i)$, where we note that by Theorem 2.51(ii), $(\mu(\mathcal{S}_j)/\hat{\mu}_k(\rho_i), j : \rho_j = \rho_i)^{\downarrow} \sim \mathrm{PD}(1-\beta, -\beta)$ is independent of $\hat{\mu}_k(\rho_i)$.

We have $\hat{J}_k = \rho_i$ with probability $\hat{\mu}_k(\rho_i)$, and hence J_k is sampled from μ_k, as required in Algorithm 2.9. By Theorem 2.51(iii), the weighted $\mathbb{R}$-trees $(\mu(\mathcal{S}_i)^{-\beta}\mathcal{S}_i, \mu(\mathcal{S}_i)^{-1}\mu \upharpoonright_{\mathcal{S}_i})$ are independent copies of $(\mathcal{T}, \mu)$, i.e. conditionally given $\Sigma_{k+1} \in \mathcal{S}_i$, the sampling procedure of $\Sigma_{k+1} \in \mathcal{S}_i$ from $\mu(\mathcal{S}_i)^{-1}\mu \upharpoonright_{\mathcal{S}_i}$ is like sampling $\Sigma_0 \in \mathcal{T}$ from μ. Hence ξ_k is an independent (β, β)-string of beads, as required in Algorithm 2.9.

Let us consider the distribution of Q_k. Conditionally given $\deg(\hat{J}_k, \widehat{\mathcal{T}}_k) = 2$, Σ_{k+1} is a leaf of a connected component $\mathcal{S}_j$ of $\mathcal{T} \setminus \widehat{\mathcal{T}}_k$ with root $\rho_j = \rho_i = \hat{J}_k$, which is chosen independently and proportionally to relative mass $\mu(\mathcal{S}_j)/\hat{\mu}_k(\rho_i)$. By Theorem

2.51, the relative mass partition above $\hat{J}_k$ is PD$(1-\beta,-\beta)$, i.e. by Proposition 2.20, $Q_k \sim \text{Beta}(\beta, 1-2\beta)$, as required in Algorithm 2.9.

Conditionally given $\deg(\hat{J}_k, \widehat{\mathcal{T}}_k) = d$ for some $d \geq 3$, Σ_{k+1} is a leaf of a connected component $\mathcal{S}_j$ of $\mathcal{T} \setminus \widehat{\mathcal{T}}_k$ with root $\rho_j = \rho_i = \hat{J}_k$. By Theorem 2.51 and Proposition 2.20, the relative mass partition of the connected components $\mathcal{T} \setminus \widehat{\mathcal{T}}_k$ with root ρ_i is PD$(1-\beta, (d-3)(1-\beta)+1-2\beta)$ where we note that $\hat{J}_k$ must have been selected $d-2$ times up step k in order to obtain $\deg(\hat{J}_k, \widehat{\mathcal{T}}_k) = d$. Therefore, by Proposition 2.20, conditionally given $\deg(\hat{J}_k, \widehat{\mathcal{T}}_k) = d$, $Q_k \sim \text{Beta}(\beta, (d-3)(1-\beta)+1-2\beta)$, as required in Algorithm 2.9. We also note that, by Proposition 2.20, Q_k is conditionally independent of $\hat{\mu}_k(\hat{J}_k)$ given $\deg(\hat{J}_k, \widehat{\mathcal{T}}_k) = d$.

The mass distribution in $(\widehat{\mathcal{T}}_{k+1}, \hat{\mu}_{k+1})$ can easily be found from Proposition 2.14, as in Proposition 2.43. $\square$

Proof of Theorem 2.58(ii). Recall that the ingredients in Algorithm 2.4 to construct the sequence on the RHS of (2.99) are the Mittag-Leffer Markov chain $(S_k, k \geq 0)$, attachment points $(J_k, k \geq 0)$, and an independent sequence of i.i.d. random variables $(B_k, k \geq 0)$ with $B_1 \sim \text{Beta}(1, 1/\beta - 2)$. We recover these ingredients from the random variables incorporated in the construction of the LHS of (2.99) via the following coupling.

- Set $S_0 = S_0^*$, i.e. $S_0 \sim \text{ML}(\beta, \beta)$ is the length of the initial (β,β)-string of beads $(\mathcal{T}_0^*, \mu_0^*) = (\widetilde{\mathcal{T}}_0, \tilde{\mu}_0)$. For $k \geq 0$, set S_k equal to the total length of $\mathcal{T}_k^*$, i.e. $S_k = S_k^*$.
- Set $(J_k, k \geq 0) = (\tilde{J}_k, k \geq 0)$.
- Set $(B_k, k \geq 0) = (B_k^*, k \geq 0)$, where B_k^* denotes the length split between the unmarked and the marked part of the independent β-mixed string of beads $(E_k^+, R_k^+, \mu_k^+) = \Psi_\beta(\xi_k^{(1)}, \xi_k^{(2)}, \gamma_k)$. By Remark 2.40, $(B_k, k \geq 0)$ is an i.i.d. sequence with $B_1 \sim \text{Beta}(1, 1/\beta - 2)$, as required.

Since Algorithm 2.4 and Algorithm 2.42 specify Markov chains, we have to check

$$\begin{aligned}&\left(\left(\widetilde{\mathcal{T}}_k, \left(\widetilde{W}_k^{(i)}, i \geq 1\right)\right), \left(\widetilde{\mathcal{T}}_{k+1}, \left(\widetilde{W}_{k+1}^{(i)}, i \geq 1\right)\right)\right) \\ &\qquad \overset{d}{=} \left(\left(\mathcal{T}_k, \left(W_k^{(i)}, i \geq 1\right)\right), \left(\mathcal{T}_{k+1}, \left(W_{k+1}^{(i)}, i \geq 1\right)\right)\right) \qquad (2.106)\end{aligned}$$

for all $k \geq 0$. We prove (2.106) by induction on k.

The shape of the tree $\widetilde{\mathcal{T}}_1$ is clearly the same as the shape of $\mathcal{T}_1$. In particular, $\widetilde{\mathcal{T}}_1$ and $\mathcal{T}_1$ have three edges, connected via one branch point. For $k = 0$ the distribution of the LHS of (2.106) is hence fully characterised by the joint distribution of the lengths $(L_1^{(1)}, L_1^{(2)}, L_1^{(3)}, L_{1,1}^{(1)})$ of the four edges $(E_1^{(1)}, E_1^{(2)}, E_1^{(3)}, E_{1,1}^{(1)})$ of $\mathcal{T}_1^*$, where we recall that $E_{1,1}^{(1)}$ is the only marked edge of $\mathcal{T}_1^*$, and $\mathcal{T}_0^* = E_1^{(1)} \cup E_1^{(3)}$.

By Theorem 2.58(i), Proposition 2.52 and Proposition 2.54, the total mass split in $\mathcal{T}_1^*$ is given by

$$\left(\mu_1^*\left(E_1^{(1)}\right), \mu_1^*\left(E_1^{(2)}\right), \mu_1^*\left(E_1^{(3)}\right), \mu_1^*\left(E_{1,1}^{(1)}\right)\right) \sim \mathrm{Dirichlet}(\beta, \beta, \beta, 1-2\beta),$$

and the edges of $\mathcal{T}_1^*$ equipped with restricted mass measures are rescaled independent (β, θ)-strings of beads where $\theta = \beta$ for the unmarked edges $E_1^{(1)}, E_1^{(2)}, E_1^{(3)}$, and $\theta = 1-2\beta$ for the marked edge $E_{1,1}^{(1)}$. Therefore, by Proposition 2.13, the total length S_1^* of $\mathcal{T}_1^*$ is given by

$$S_1^* \sim \mathrm{ML}(\beta, 3\beta + (1-2\beta)) = \mathrm{ML}(\beta, 1+\beta),$$

and the distribution of the length split in $\mathcal{T}_1^*$ is obtained via

$$(S_1^*)^{-1}\left(L_1^{(1)}, L_1^{(2)}, L_1^{(3)}, L_{1,1}^{(1)}\right) \sim \mathrm{Dirichlet}(1, 1, 1, 1/\beta - 2) \tag{2.107}$$

where $(S_1^*)^{-1}(L_1^{(1)}, L_1^{(2)}, L_1^{(3)}, L_{1,1}^{(1)})$ is independent of S_1^*. By Proposition 2.14(i)-(ii), we conclude that

$$S_0^* = L_1^{(1)} + L_1^{(3)} =: A_1^* S_1^*$$

where $A_1^* \sim \mathrm{Beta}(2, 1/\beta - 1)$ is independent of S_1^*, i.e. (S_0^*, S_1^*) has the same distribution as (S_0, S_1), see Proposition 2.21. Furthermore, the attachment point J_0^* has an independently and uniformly chosen location on $\mathcal{T}_0^*$ since

$$\left(L_1^{(1)} + L_1^{(3)}\right)^{-1} L_1^{(1)} \sim \mathrm{Beta}(1, 1) = U([0, 1])$$

is independent of S_0^*. Similarly, it follows that the weight $\widetilde{W}_1^{(1)} = L_{1,1}^{(1)}$, is given by

$$\widetilde{W}_1^{(1)} = L_{1,1}^{(1)} = B_0^*\Big(L_1^{(2)} + L_{1,1}^{(1)}\Big) = B_0^*(S_1^* - S_0^*)$$

where $B_0^* \sim \mathrm{Beta}(1/\beta - 2, 1)$ is independent of $S_1^* - S_0^*$, $(\mathcal{T}_k^*, \mu_k^*)$, $\xi_k^{(1)}$ and $\xi_k^{(2)}$. By the coupling, we obtain (2.106) for $k = 0$.

Now assume that (2.106) holds up to k, in particular that

$$\Big(\widetilde{\mathcal{T}}_k, \Big(\widetilde{W}_k^{(i)}, i \geq 1\Big)\Big) \stackrel{d}{=} \Big(\mathcal{T}_k, \Big(W_k^{(i)}, i \geq 1\Big)\Big). \tag{2.108}$$

In the tree growth process $(\widetilde{\mathcal{T}}_k, k \geq 0)$ edge and branch point selection is based on masses, whereas in $(\mathcal{T}_k, k \geq 0)$ edges are selected based on length and branch points based on weights. We first prove the correspondence of the selection rules, where we work conditionally given the shape of the tree $\widetilde{\mathcal{T}}_k = \mathcal{T}_k$, in particular conditionally given that $\mathcal{T}_k^*$ has l marked components $\mathcal{R}_k^{(i)} \neq \{\rho\}$, $i \in [l]$, of sizes $d_i - 2$, $i \in [l]$, respectively, or, in other words, that $\widetilde{\mathcal{T}}_k$ has l branch points $\tilde{v}_i, i \in [l]$, of degrees $d_i, i \in [l]$, respectively, and a total of $k + l + 1$ edges. By (i) and Proposition 2.52, the total mass split in $\widetilde{\mathcal{T}}_k$ is

$$\begin{aligned}\Big(\tilde{\mu}_k\Big(E_k^{(1)}\Big), \ldots, \tilde{\mu}_k\Big(E_k^{(k+l+1)}\Big), \tilde{\mu}_k(\tilde{v}_1), \ldots, \tilde{\mu}_k(\tilde{v}_l)\Big)& \\ \sim \mathrm{Dirichlet}(\beta, \ldots, \beta, w(d_1), \ldots, w(d_l))&\end{aligned} \tag{2.109}$$

where $w(d_i) = (d_i - 3)(1 - \beta) + 1 - 2\beta$ for $i \in [l]$. We denote the edge lengths and the branch point weights in $\widetilde{\mathcal{T}}_k$ by

$$\widetilde{L}_k = \Big(\widetilde{L}_k^{(1)}, \ldots, \widetilde{L}_k^{(k+l+1)}\Big), \quad \widetilde{W}_k = \Big(\widetilde{W}_k^{(1)}, \ldots, \widetilde{W}_k^{(l)}\Big), \tag{2.110}$$

respectively, and use corresponding notation in $\mathcal{T}_k$. We first show that the joint distributions of edge lengths, weights and the selected attachment points $\tilde{J}_k$ and J_k in $\widetilde{\mathcal{T}}_k$ and $\mathcal{T}_k$, respectively, are the same in Algorithm 2.42 and Algorithm 2.4, i.e. for any $k \geq 0$, and any continuous and bounded function $f : \mathbb{R}^{k+2l+1} \to \mathbb{R}$,

$$\mathbb{E}\Big[f\Big(\widetilde{L}_k, \widetilde{W}_k\Big)\mathbb{1}_{\{\tilde{J}_k \in E_k^{(j)}\}}\Big] = \mathbb{E}\Big[f(L_k, W_k)\mathbb{1}_{\{J_k \in E_k^{(j)}\}}\Big] \tag{2.111}$$

for any $j \in [k+l+1]$, and

$$\mathbb{E}\Big[f\Big(\widetilde{L}_k, \widetilde{W}_k\Big)\mathbb{1}_{\{\tilde{J}_k=v_j\}}\Big] = \mathbb{E}\Big[f(L_k, W_k)\mathbb{1}_{\{J_k=v_j\}}\Big] \tag{2.112}$$

for any $j \in [l]$.

- *Proof of* (2.111). Fix some $j \in [k+l+1]$, and consider the LHS of (2.111) first. Conditioning on $\widetilde{J}_k \in E_k^{(j)}$, and using the mass split (2.109) and Proposition 2.14(iv), we obtain

  $$\mathbb{E}\Big[f\Big(\widetilde{L}_k, \widetilde{W}_k\Big)\mathbb{1}_{\{\tilde{J}_k \in E_k^{(j)}\}}\Big] = \frac{\beta}{k+\beta}\mathbb{E}\Big[f\Big(\widetilde{L}_k, \widetilde{W}_k\Big)\Big|\tilde{J}_k \in E_k^{(j)}\Big].$$

  By Proposition 2.14(iv) and (2.109), conditionally given $\tilde{J}_k \in E_k^{(j)}$, the distribution of the mass split

  $$\Big(X_k^{(1)}, \dots, X_k^{(j-1)}, X_k^{(j)}, X_k^{(j+1)}, \dots, X_k^{(k+l+1)}, X_k^{(k+l+2)}, \dots, X_k^{(k+2l+1)}\Big) \tag{2.113}$$

  with $X_k^{(i)} = \tilde{\mu}_k(E_k^{(i)})$ for $i \in [k+l+1]$ and $X_k^{(i)} = \tilde{\mu}_k(\tilde{v}_{i-(k+l+1)})$ for $i \in [k+2l+1] \setminus [k+l+1]$ is

  $$\text{Dirichlet}(\beta, \dots, \beta, 1+\beta, \beta, \dots, \beta, w(d_1), \dots, w(d_l)). \tag{2.114}$$

  Furthermore, still conditionally given $\tilde{J}_k \in E_k^{(j)}$, $\tilde{J}_k$ is an atom of mass $\tilde{\mu}_k(\tilde{J}_k) =: U_k^{(j)}X_k^{(j)}$ sampled from the rescaled independent (β,β)-string of beads related to $E_k^{(j)}$, splitting $E_k^{(j)}$ into two edges $E_{k+1}^{(j)}$ and $E_{k+1}^{(k+l+3)}$ of masses $\tilde{\mu}_k(E_{k+1}^{(j)}) =: U_k^{(j-)}X_k^{(j)}$ and $\tilde{\mu}_k(E_{k+1}^{(k+l+3)}) =: U_k^{(j+)}X_k^{(j)}$, respectively. By Proposition 2.18, the relative mass split on $E_k^{(j)}$ is given by

  $$\Big(U_k^{(j-)}, U_k^{(j)}, U_k^{(j+)}\Big) \sim \text{Dirichlet}(\beta, 1-\beta, \beta),$$

and is independent of $X_k^{(j)} = \tilde{\mu}_k(E_k^{(j)})$, since, by (i) and Proposition 2.52, the (β,β)-string of beads

$$\left(\left(X_k^{(j)}\right)^{-\beta} E_k^{(j)}, \left(X_k^{(j)}\right)^{-1} \tilde{\mu}_k \restriction_{E_k^{(j)}}\right)$$

is independent of the scaling factor $X_k^{(j)}$. We obtain the refined mass split

$$\left(\overline{X}_k^{(1)}, \ldots, \overline{X}_k^{(j-1)}, \overline{X}_k^{(j-)}, \overline{X}_k^{(j)}, \overline{X}_k^{(j+)}, \overline{X}_k^{(j+1)}, \ldots, \ldots, \overline{X}_k^{(k+2l+1)}\right) \tag{2.115}$$

where $\overline{X}_k^{(i)} = X_k^{(i)}$, $i \in [k+2l+1] \setminus \{j, j+, j-\}$ and $\overline{X}_k^{(j-)} = U_k^{(j-)} X_k^{(j)}$, $\overline{X}_k^{(j)} = U_k^{(j)} X_k^{(j)}$ and $\overline{X}_k^{(j+)} = U_k^{(j+)} X_k^{(j)}$.

By Proposition 2.14(iii), the distribution of (2.115) is

$$\text{Dirichlet}(\beta, \ldots, \beta, \beta, 1-\beta, \beta, \beta, \ldots, \beta, w(d_1), \ldots, w(d_l)).$$

Furthermore, the atom $\tilde{J}_k$ induces the two rescaled independent (β,β)-strings of beads

$$\left(\left(X_k^{(j-)}\right)^{-\beta} E_{k+1}^{(j)}, \left(X_k^{(j-)}\right)^{-1} \tilde{\mu}_k \restriction_{E_{k+1}^{(j)}}\right)$$

and

$$\left(\left(X_k^{(j+)}\right)^{-\beta} E_{k+1}^{(k+l+3)}, \left(X_k^{(j+)}\right)^{-1} \tilde{\mu}_k \restriction_{E_{k+1}^{(k+l+3)}}\right)$$

where $X_k^{(i)} = U_k^{(i)} X_k^{(j)}, i \in \{j-, j+\}$, i.e. the lengths of the edges $E_{k+1}^{(j)}$ and $E_{k+1}^{(k+l+3)}$ are given by

$$\widetilde{L}_{k+1}^{(j)} = \left(U_k^{(j-)} X_k^{(j)}\right)^{\beta} M_k^{(j-)}, \quad \widetilde{L}_{k+1}^{(k+l+3)} = \left(U_k^{(j+)} X_k^{(j)}\right)^{\beta} M_k^{(j+)},$$

respectively, where $M_k^{(i)} \sim \text{ML}(\beta,\beta)$, $i \in \{j-, j+\}$, are independent, see Proposition 2.18. Conditionally given $\tilde{J}_k \in E_k^{(j)}$, by (2.108) and Corollary 2.50, the weights $\widetilde{W}_k^{(i)}$ of $\widetilde{\mathfrak{T}}_k$ are therefore

$$\widetilde{W}_k^{(i-(k+l+1))} = \left(X_k^{(i)}\right)^{\beta} M_k^{(i)}, \quad i \in [k+2l+1] \setminus [k+l+1],$$

where the lengths $\widetilde{L}_k^{(i)}$ are

$$\widetilde{L}_k^{(i)} = \begin{cases} \left(X_k^{(i)}\right)^\beta M_k^{(i)}, & i \in [k+l+1] \setminus [j] \\ \left(U_k^{(j-)} X_k^{(j)}\right)^\beta M_k^{(j-)} + \left(U_k^{(j+)} X_k^{(j)}\right)^\beta M_k^{(j+)}, & i = j, \end{cases}$$

for independent random variables

$$M_k^{(i)} \sim \begin{cases} \mathrm{ML}(\beta, \beta), & i \in [k+l+1] \setminus \{j\} \cup \{j-, j+\}, \\ \mathrm{ML}(\beta, w(d_{i-(k+l+1)})), & i \in [k+2l+1] \setminus [k+l+1]. \end{cases}$$

Also note that, by the definition of $(S_k^*, k \geq 0)$ and the attachment procedure,

$$S_{k+1}^* - S_k^* = \tilde{\mu}_k\left(\tilde{J}_k\right)^\beta M_k^* = \left(U_k^{(j)} X_k^{(j)}\right)^\beta M_k^*$$

where $M_k^* \sim \mathrm{ML}(\beta, 1-\beta)$ is the length of the attached, independent β-mixed string of beads. We conclude by Proposition 2.13 and Proposition 2.14(i)-(ii) that

$$S_k^* = A_k^* S_{k+1}^* \sim \mathrm{ML}(\beta, k+\beta)$$

where $S_{k+1}^* \sim \mathrm{ML}(\beta, k+1+\beta)$ and $A_k^* \sim \mathrm{Beta}(k/\beta+2, 1/\beta-1)$ are independent, and that, conditionally given $\tilde{J}_k \in E_k^{(j)}$, we have $(\widetilde{L}_k, \widetilde{W}_k) = S_{k+1}^* A_k^* \widetilde{Z}_k$ where

$$\widetilde{Z}_k = \left(\widetilde{Z}_k^{(1)}, \ldots, \widetilde{Z}_k^{(j-1)}, \widetilde{Z}_k^{(j)}, \widetilde{Z}_k^{(j+1)}, \ldots, \widetilde{Z}_k^{(k+l+1)}, \widetilde{Z}_k^{(k+l+2)}, \ldots, \widetilde{Z}_k^{(k+2l+1)}\right)$$

is independent of S_{k+1}^* and A_k^*, and has a

$$\mathrm{Dirichlet}(1, \ldots, 1, 2, 1, \ldots, 1, w(d_1)/\beta, \ldots, w(d_l)/\beta)$$

distribution. Hence,

$$\mathbb{E}\left[f\left(\widetilde{L}_k, \widetilde{W}_k\right) \mathbb{1}_{\{\tilde{J}_k \in E_k^{(j)}\}}\right] = \frac{\beta}{k+\beta} \mathbb{E}\left[f\left(S_k^* \widetilde{Z}_k\right) \middle| \tilde{J}_k \in E_k^{(j)}\right].$$

We now consider the RHS of (2.111). We condition on $J_k \in E_k^{(j)}$, and apply Lemma 2.49 and Proposition 2.14(iv) to obtain

$$\mathbb{E}\Big[f\Big(L_k^{(1)},\dots,L_k^{(k+l+1)},W_k^{(1)},\dots,W_k^{(l)}\Big)\mathbb{1}_{\{J_k\in E_k^{(j)}\}}\Big]=\frac{\beta}{k+\beta}\mathbb{E}\Big[f(S_kZ_k)\Big|J_k\in E_k^{(j)}\Big]$$

where $S_k\sim\mathrm{ML}(\beta,k+\beta)$ is independent of

$$Z_k=\Big(Z_k^{(1)},\dots,Z_k^{(j-1)},Z_k^{(j)},Z_k^{(j+1)},\dots,Z_k^{(k+l+1)},Z_k^{(k+l+2)},\dots,Z_k^{(k+2l+1)}\Big)$$

and $Z_k\sim\mathrm{Dirichlet}(1,\dots,1,2,1,\dots,1,w(d_1)/\beta,\dots,w(d_l)/\beta)$.

Hence, we conclude (2.111).

- *Proof of* (2.112). Consider now the LHS of (2.112). We follow the lines of the proof of (2.111). Conditionally given $\widetilde{J}_k=\tilde{v}_j$, the mass split (2.113) has distribution

$$\mathrm{Dirichlet}(\beta,\dots,\beta,w(d_1),\dots,w(d_{j-1}),1+w(d_j),w(d_j+1),\dots,w(d_l)) \tag{2.116}$$

where we recall that $w(d_i)=(d_i-3)(1-\beta)+1-2\beta$ for $i\in[l]$. By (i), the mass $\tilde{\mu}_k(\tilde{v}_j)$ is split by an independent random variable $(Q_k,1-Q_k)\sim\mathrm{Dirichlet}(\beta,w(d_j)+(1-\beta))$ into a new unmarked (β,β)-string of beads

$$\Big(\tilde{\mu}_{k+1}\Big(E_{k+1}^{(k+l+2)}\Big)^{-\beta}E_{k+1}^{(k+l+2)},\tilde{\mu}_{k+1}\Big(E_{k+1}^{(k+l+2)}\Big)^{-1}\tilde{\mu}_{k+1}\restriction_{E_{k+1}^{(k+l+2)}}\Big)$$

attached to $\tilde{v}_j$ and the mass $\tilde{\mu}_{k+1}(\tilde{v}_j)$, i.e.

$$\tilde{\mu}_{k+1}\Big(E_{k+1}^{(k+l+2)}\Big)=Q_k\tilde{\mu}_k(\tilde{v}_j),\quad \tilde{\mu}_{k+1}(\tilde{v}_j)=(1-Q_k)\tilde{\mu}_k(\tilde{v}_j).$$

By (i), Lemma 2.56 and the induction hypothesis, we have $\widetilde{W}_{k+1}^{(i-(k+l+2))}=\tilde{\mu}_{k+1}(\tilde{v}_{i-(k+l+2)})^{\beta}M_{k+1}^{(i)}$ for $i\in[k+2l+2]\setminus[k+l+2],i\neq j+(k+l+2)$ and $\tilde{L}_{k+1}^{(i)}=\tilde{\mu}_{k+1}\Big(E_{k+1}^{(i)}\Big)^{\beta}M_{k+1}^{(i)}$ for $i\in[k+l+2]$ where the random variables $M_{k+1}^{(i)}\sim\mathrm{ML}(\beta,\beta)$ for $i\in[k+l+2]$, $M_{k+1}^{(i)}\sim\mathrm{ML}(\beta,w(d_{i-(k+l+2)}))$ for $i\in[k+2l+2]\setminus[k+l+2],i\neq j+(k+l+2)$ and $M_{k+1}^{(i)}\sim\mathrm{ML}(\beta,w(d_{i-(k+l+2)})+(1-\beta))$

for $i = j + (k + l + 2)$, are independent of the mass split (2.116). Note that, by Lemma 2.56 and recalling the definition of $\widetilde{W}_k^{(i)}, i \in [l]$,

$$\widetilde{W}_k^{(j)} = B_k^* \widetilde{W}_{k+1}^{(j)}$$

where $B_k^* \sim \mathrm{Beta}((d_j - 2)(1/\beta - 1), 1/\beta - 2)$ is independent of $\widetilde{W}_{k+1}^{(j)}$. Note that $\widetilde{W}_{k+1}^{(i)} = \widetilde{W}_k^{(i)}$ for $i \neq j$, $\tilde{L}_{k+1}^{(i)} = \tilde{L}_k^{(i)}$ for $i \in [k + l + 1]$, i.e.

$$S_{k+1}^* - S_k^* = \tilde{L}_{k+1}^{(k+l+2)} + \left(\widetilde{W}_{k+1}^{(j)} - \widetilde{W}_k^{(j)}\right).$$

By Proposition 2.13, $S_k^* = A_k^* S_{k+1}^*$ where $S_{k+1}^* \sim \mathrm{ML}(\beta, k + 1 + \beta)$ and $A_k^* \sim \mathrm{Beta}(k/\beta + 2, 1/\beta - 1)$ are independent. The rest of the proof of (2.112) is analogous to the proof of (2.111).

The statement in (ii) now directly follows from (2.111)-(2.112) and the coupling. $\square$

Chapter 3

Branch merging on continuum trees with applications to regenerative tree growth

Abstract

We introduce a family of branch merging operations on continuum trees and show that Ford CRTs are distributionally invariant. This operation is new even in the special case of the Brownian CRT, which we explore in more detail. The operations are based on spinal decompositions and a regenerativity preserving merging procedure of (α, θ)-strings of beads, that is, random intervals $[0, L_{\alpha,\theta}]$ equipped with a random discrete measure dL^{-1} arising in the limit of ordered (α, θ)-Chinese restaurant processes as introduced by Pitman and Winkel. Indeed, we iterate the branch merging operation recursively and give a new approach to the leaf embedding problem on Ford CRTs related to $(\alpha, 2 - \alpha)$-tree growth processes.

Keywords: string of beads, Chinese restaurant process, regenerative interval partition, Poisson-Dirichlet, Ford CRT, Brownian CRT, branch merging, tree-valued Markov process

3.1 Introduction

Tree-valued Markov processes and operations on continuum random trees (CRTs) such as pruning have recently attracted particular interest in probability theory. Evans and Winter, for instance, consider regrafting in combination with subtree pruning and show in [33] that Aldous' Brownian CRT arises as the stationary distribution of a certain reversible $\mathbb{R}$-tree valued Markov process. Evans et al [32] present a similar study involving root growth with regrafting. Aldous [11, 12] and Pal [55, 56] consider a Markov chain operating on the space of binary rooted $\mathbb{R}$-trees randomly removing and reinserting leaves, related to a diffusion limit on the space of continuum trees. See also [1, 2, 3, 5, 16, 18, 19, 43] for related work.

We study branch merging as a new operation on (continuum) trees, which leaves Ford CRTs, and in particular the Brownian CRT, distributionally invariant. Our branch merging operation is based on a leaf sampling procedure, and on the study of the resulting reduced subtrees equipped with projected subtree masses. The notion of a string of beads naturally captures this projected mass on a branch.

Following [60], we consider (α, θ)-strings of beads for $\alpha \in (0,1)$ and $\theta > 0$. An (α, θ)-string of beads is a random interval $[0, L_{\alpha,\theta}]$ equipped with a random discrete measure dL^{-1}, arising in the framework of the two-parameter family of (α, θ)-Chinese restaurant processes (CRP) when equipped with a regenerative table order. An (α, θ)-Chinese restaurant process as introduced by Dubins and Pitman (see e.g. [58]) is a sequence of exchangeable random partitions $(\Pi_n, n \geq 1)$ of $[n] := \{1, \ldots, n\}$ defined by customers labelled by $[n]$ sitting at a random number of tables. Customer 1 sits at the first table, and, at step $n+1$, conditionally given k tables in the restaurant with $n_1, \ldots, n_k$ customers at each table, customer $n+1$

- sits at the i-th occupied table with probability $(n_i - \alpha)/(n + \theta)$, $i \in [k]$;
- opens a new table with probability $(k\alpha + \theta)/(n + \theta)$.

The joint distribution of the asymptotic relative table sizes arranged in decreasing order is known to be Poisson-Dirichlet with parameters (α, θ), for short $\mathrm{PD}(\alpha, \theta)$. $\mathrm{PD}(\alpha, \theta)$ vectors have been widely studied in the literature: see e.g. [34, 62] for constructions and properties, [24] and [58] for coagulation-fragmentation-dualities, and [60] for Poisson-Dirichlet compositions arising in the framework of continuum random trees.

Pitman and Winkel [60] equipped the (α, θ)-CRP with a table order which is independent of the actual seating process. These structures arise naturally in the study of the (α, θ)-tree growth process $(T_n^{\alpha,\theta}, n \geq 1)$, that is, a rooted binary regenerative tree growth process with n leaves at step n labelled by $[n]$ such that the partition structure induced by the subtrees (tables) encountered on the unique path from the root to the first leaf defines an ordered (α, θ)-CRP, see Section 3.4.1.2.

It was shown that the table sizes in the ordered (α, θ)-CRP induce a regenerative composition structure of $[n]$ in the sense of Gnedin and Pitman [38], and yield a so-called (α, θ)-regenerative interval partition of $[0, 1]$ in the scaling limit. The lengths of the components of this interval partition are the limiting proportions of table sizes in the CRP arranged in regenerative order, and correspond to the masses of the atoms of the discrete measure dL^{-1} associated with an (α, θ)-string of beads. (α, θ)-strings of beads in this sense align the limiting table proportions of an (α, θ)-CRP along an interval of length $L_{\alpha,\theta}$ in regenerative order where $L_{\alpha,\theta}$ is the almost-sure limit as $n \to \infty$ of the number of tables in the CRP at step n scaled by $n^{-\alpha}$ (see Section 3.2.1).

In this chapter, we study a merging operation for (α, θ_i)-strings of beads, $i \in [k]$ for some $k \in \mathbb{N}$ fixed, subject to a Dirichlet$(\theta_1, \ldots, \theta_k)$ mass split with $\alpha \in (0, 1)$, $\theta_1, \ldots, \theta_k > 0$. This situation arises naturally in the framework of ordered CRPs. Consider k ordered CRPs with parameters (α, θ_i), $i \in [k]$. At each step, let one of these restaurants be selected proportionally to the total weight in each CRP with initial weights $\theta_i, i \in [k]$. Conditionally given that the j-th restaurant is selected with $m_j(n)$ customers present, the $(n+1)$-st customer is customer $m_j(n)+1$ in the (α, θ_j)-CRP, and chooses a table according to the (α, θ_j)-selection rule, independent of the restaurant selection process. Then, by a simple urn model, we obtain a Dirichlet$(\theta_1, \ldots, \theta_k)$ mass split between these restaurants in the limit, and each CRP is associated with an (α, θ_i)-string of beads, when we rescale distances and masses by appropriate scaling factors depending on the Dirichlet vector. Remarkably, the identified (α, θ_i)-strings of beads are independent of each other and of the Dirichlet$(\theta_1, \ldots, \theta_k)$ mass split. We ask the following questions.

- (How) can we merge these ordered (α, θ_i)-CRPs, $i \in [k]$, by completing the partial table orders to obtain the seating rule of an ordered (α, θ)-Chinese restaurant process for some $\theta > 0$?
- How does the parameter θ depend on α and θ_i, $i \in [k]$?

Our merging algorithm allows to merge (α, θ_i)-strings of beads preserving the regenerative property. It becomes particularly simple when $\alpha = \theta_i = 1/2$ for all $i \in [k]$, which is the case relevant for branch merging on the Brownian CRT (see Figure 3.1 for an example with $k = 3$).

Algorithm/Theorem 3.1 (Merging $(1/2, 1/2)$-strings of beads). *Let $E_i = (\rho_i, \rho_i')$, $i \in [k]$, be k disjoint intervals and μ a mass measure on $E := \bigcup_{i=1}^k E_i$ such that $(\mu(E_1), \ldots, \mu(E_k)) \sim$ Dirichlet$(1/2, \ldots, 1/2)$. Suppose that the rescaled pairs*

$$\left(\mu(E_i)^{-1/2} E_i, \mu(E_i)^{-1} \mu \restriction_{E_i}\right)$$

are $(1/2, 1/2)$-strings of beads, $i \in [k]$, independent of each other and of the Dirichlet mass split $(\mu(E_1), \ldots, \mu(E_k))$.

Define the metric space $(E_{\biguplus}, d_{\biguplus})$ as follows.

- *Set $E^{(1)} := E$, $E_i^{(1)} := E_i$, $\rho_i^{(1)} := \rho_i$, $i \in [k]$. For $n \geq 1$, do the following.*
  - (i) *Select an atom X_n from $\mu \restriction_{E^{(n)}}$ with probability proportional to mass.*
  - (ii) *Let $a_n := \rho_{I_n}^{(n)}$ and $b_n := X_n$ where $I_n := i$ if $X_n \in E_i$, $i \in [k]$.*
  - (iii) *Update $\rho_{I_n}^{(n+1)} := X_n$, $\rho_j^{(n+1)} := \rho_j^{(n)}$ for any $j \in [k] \setminus \{I_n\}$, $E_i^{(n+1)} := (\rho_i^{(n+1)}, \rho_i')$ for any $i \in [k]$ and $E^{(n+1)} := \bigcup_{i=1}^k E_i^{(n+1)}$.*
- *Align the interval components $(a_n, b_n]$ in order by identifying a_{n+1} with b_n, $n \geq 1$, to obtain the set $E_{\biguplus}$ equipped with a metric $d_{\biguplus}$ induced by this operation.*

Then the pair $(E_{\biguplus}, \mu)$ equipped with the metric $d_{\biguplus}$ is an $(1/2, k/2)$-string of beads.

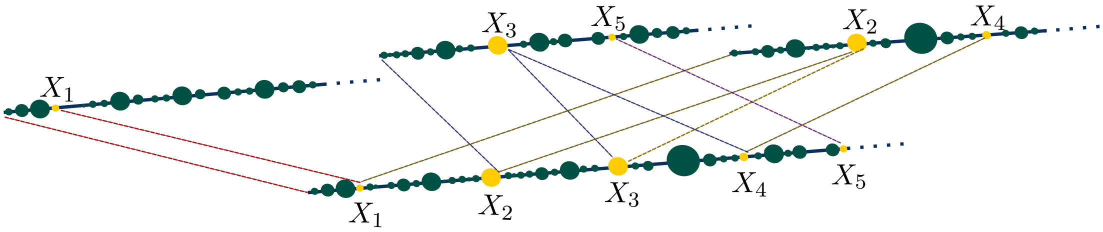

Fig. 3.1 Example of merging of strings of beads for $k = 3$.
First five cut points $X_1, \dots, X_5$ are displayed.

This algorithm will be made mathematically precise, and will be presented for general (α, θ_i)-strings of beads, $\alpha \in (0, 1)$ and $\theta_i > 0$, $i \in [k]$, in Section 3.3, see Algorithm 3.14 and Theorem 3.15. In general, our merging procedure yields an $(\alpha, \theta := \sum_{i=1}^{k} \theta_i)$-string of beads. As a corollary of our result, we recover the following property of $\mathrm{PD}(\alpha, \theta)$ vectors, disregarding the regenerative order incorporated by (α, θ)-strings of beads.

Corollary 3.2. *Let $\alpha \in (0, 1)$ and $\theta_i > 0$, $i \in [k]$. Let $P_i, i \in [k]$, be a sequence of independent random vectors such that $P_i = (P_{i,1}, P_{i,2}, \dots) \sim \mathrm{PD}(\alpha, \theta_i), i \in [k]$, and let $Y := (Y_1, \dots, Y_k)$ be an independent k-dimensional vector Y such that $Y \sim$ $\mathrm{Dirichlet}(\theta_1, \dots, \theta_k)$. Then*

$$P := (Y_1 P_{1,1}, Y_2 P_{2,1}, \dots, Y_k P_{k,1}, Y_1 P_{1,2}, \dots, Y_k P_{k,2}, Y_1 P_{1,3}, \dots)^{\downarrow} \sim \mathrm{PD}(\alpha, \theta)$$

where $\theta := \sum_{i=1}^{k} \theta_i$ and $\downarrow$ denotes the decreasing rearrangement of components.

The proof of our main result, Theorem 3.15, goes back to decomposition results for (α, θ)-strings of beads [60] and a stick-breaking construction of (α, θ)-regenerative interval partitions [38].

The (α, θ)-tree growth model [60] related to ordered (α, θ)-CRPs generalises Ford's model (i.e. $\theta = 1 - \alpha$) [35] whose asymptotics were studied earlier by Haas et al. [42]. See also [22] for a multifurcating tree growth process related to Ford's model, and [59] for structural results on regenerative tree growth.

It was proved in [41] that, for $0 < \alpha < 1$ and $\theta \geq 0$, the delabelled trees in the (α, θ)-model have a binary fragmentation continuum random tree (CRT) $\mathcal{T}^{\alpha,\theta}$ as distributional scaling limit, where a two-stage limit was provided earlier in [60].

The problem of appropriately embedding leaf labels of $(T_n^{\alpha,\theta}, n \geq 1)$ and the reduced trees with edge lengths $\mathcal{R}_k^{\alpha,\theta}$, arising as scaling limit of the growth process $(T_n^{\alpha,\theta}, n \geq 1)$ reduced to the first k leaves, into $\mathcal{T}^{\alpha,\theta}$ was solved in [60] by introducing the so-called (α,θ)-*coin tossing construction*.

Based on our merging operation for strings of beads, Algorithm 3.14, we develop a branch merging operation (Algorithm 3.44) on Ford CRTs, the class of CRTs $\mathcal{T}^{\alpha,1-\alpha}$ arising for $\theta = 1-\alpha$, $\alpha \in (0,1)$. We can couple $\mathcal{T}^{\alpha,1-\alpha}$ and $\mathcal{T}^{\alpha,2-\alpha}$ in the sense that we can embed $\mathcal{R}_k^{\alpha,2-\alpha}$ into $\mathcal{T}^{\alpha,2-\alpha}$ given an embedding of $\mathcal{R}_k^{\alpha,1-\alpha}$ into $\mathcal{T}^{\alpha,1-\alpha}$, $k \geq 1$. In particular, we construct $\mathcal{T}^{\alpha,2-\alpha}$ with leaves embedded from the limiting tree $\mathcal{T}^{\alpha,1-\alpha}$ of Ford's model.

In the case when $\alpha = 1/2$, and $\theta = 1/2$ or $\theta = 3/2$, Aldous' Brownian Continuum Random Tree [7, 8, 9] arises as distributional scaling limit of the delabelled (α,θ)-tree growth process. Leaf labelling is exchangeable in the case when $\alpha = \theta = 1/2$, and uniform sampling from the natural mass measure on the Brownian CRT allows to embed leaf labels in this case, using a simplified version of the branch merging algorithm. For the case when $\alpha = 1/2$ and $\theta = 3/2$ we obtain embedded leaves applying our branch merging operation yielding a much simpler approach than the coin tossing construction from [60].

As a by-product of these developments, we obtain the distributional invariance under branch merging of Ford trees $\mathcal{T}^{\alpha,1-\alpha}$, $\alpha \in (0,1)$, and in particular of the Brownian CRT. Detached from the aim of leaf identification in CRTs, we define the *Branch Merging Markov Chain* (BMMC) using a simplified branch merging operation as the transition rule. The BMMC operates on continuum trees. We prove that, for any $n \in \mathbb{N}$, a discrete analogue of the BMMC on the space $\mathbb{T}_n^o$ of rooted unlabelled trees with n leaves and no degree-2 vertices (except the root) has a unique stationary distribution supported on the space $\mathbb{T}_n^{b,o}$ of binary rooted unlabelled trees with n leaves (and no degree-2 vertices), to which it converges as time goes to infinity. The Brownian CRT is a stationary distribution of the BMMC, and we conjecture that, for any continuum tree as initial state, the BMMC converges in distribution to the Brownian CRT.

This chapter is organized as follows. In Section 3.2 we recall some background on regenerative composition structures, and give a proper definition of ordered Chinese restaurant processes and (α,θ)-strings of beads. In Section 3.3, we introduce the merging algorithm for strings of beads and state our main result, which is proved in Section 3.3.3. Branch merging is explained in Section 3.4, where we also recap some background on $\mathbb{R}$-trees and explain the (α,θ)-tree growth process including its convergence properties. We use a simplified version of the branch merging algorithm to define the Branch Merging Markov Chain including a discrete analogue, which is shown to have a unique stationary distribution. As an application of general branch merging, we describe leaf embedding for $(\alpha, 2-\alpha)$-tree growth processes in general, and the $(1/2, 3/2)$-tree growth process related to the Brownian CRT in particular.

3.2 Preliminaries

3.2.1 (α,θ)-Chinese restaurant processes and (α,θ)-strings of beads.

We present some properties of the two-parameter Chinese restaurant process, and refer to [58] for further details. Recall the definition of an (α,θ)-Chinese restaurant process (CRP) $(\Pi_n, n \geq 1)$ from the Introduction. The state of the system after n customers have been seated is a random partition Π_n of $[n]$. It is easy to see that, for each particular partition π of $[n]$ into k classes of sizes $n_1, \ldots, n_k$, we have

$$\mathbb{P}(\Pi_n = \pi) = \frac{\prod_{i=1}^{k-1}(\theta+\alpha i)}{[1+\theta]_{n-1}} \prod_{i=1}^{k} [1-\alpha]_{n_i-1}, \tag{3.1}$$

where, for $x \in \mathbb{R}$ and $m \in \mathbb{N}$, we define $[x]_m := x(x+1)(x+2)\cdots(x+m)$. It follows immediately from the fact that the distribution of Π_n, given by (3.1), only depends on block sizes, that the partitions Π_n are exchangeable. They are consistent as n varies, and hence induce a random partition Π_∞ of $\mathbb{N}$ whose restriction to $[n]$ is Π_n.

We equip the CRP $(\Pi_n, n \geq 1)$ with a random total order $<$ on the tables, which we call *table order*, see [60]. Independently of the process of seating of customers at tables, we order the tables from left to right according to the following scheme. The second table is put to the right of the first table with probability $\theta/(\alpha+\theta)$, and to the left with probability $\alpha/(\alpha+\theta)$. Conditionally given any of the $k!$ possible orderings of the first k tables, $k \geq 1$, the $(k+1)$-st table is put

- to the left of the first table, or between any two tables with probability $\alpha/(k\alpha+\theta)$ each;
- to the right of the last table with probability $\theta/(k\alpha+\theta)$.

We refer to the CRP with tables ordered according to $<$ as *ordered* CRP, and write $(\widetilde{\Pi}_n, n \geq 1)$ for the process of random partitions of $[n]$ with blocks ordered according to $<$, where we use the convention that $(\Pi_n, n \geq 1)$ orders the blocks according to least labels (*birth order*). For $n \in \mathbb{N}$, we write

$$\Pi_n := (\Pi_{n,1}, \ldots, \Pi_{n,K_n}), \qquad \widetilde{\Pi}_n := \left(\widetilde{\Pi}_{n,1}, \ldots, \widetilde{\Pi}_{n,K_n}\right)$$

for the blocks of the two partitions Π_n, $\widetilde{\Pi}_n$ of $[n]$, where K_n denotes the number of tables at step n. The sizes of these blocks at step n form two *compositions* of n, $n \geq 1$, that is, sequences of positive integers $(n_1, \ldots, n_k)$ with sum $n = \sum_{j=1}^{k} n_j$. The theory of CRPs gives us an almost-sure limit for the number of tables K_n, i.e.

$$L_{\alpha,\theta} = \lim_{n\to\infty} \frac{K_n}{n^\alpha}, \qquad L_{\alpha,\theta} > 0 \quad \text{a.s.},$$

as well as limiting proportions $(P_1, P_2, \ldots)$ of customers at each table in *birth* order, represented as $(P_1, P_2, P_3, \ldots) = (M_1, \overline{M}_1 M_2, \overline{M}_1 \overline{M}_2 M_3, \ldots)$ where the random variables $(M_i, i \geq 1)$ are independent, $M_i \sim \text{Beta}(1-\alpha, \theta + i\alpha)$, and $\overline{M}_i := 1 - M_i$. The distribution of ranked limiting proportions is $\text{PD}(\alpha, \theta)$. The table sizes in the ordered (α, θ)-CRP $(\widetilde{\Pi}_n, n \geq 1)$ induce a *regenerative composition structure* in the sense of Gnedin and Pitman [38], see also Section 3.2.2 here.

Definition 3.3 (Regenerative composition structure)**.** A *composition structure* $(\mathcal{C}_n, n \geq 1)$ is a Markovian sequence of random compositions of n, $n = 1, 2, \ldots$, where the transition probabilities satisfy the property of *sampling consistency*, defined via the following description. Let n identical balls be distributed into an ordered series of boxes according to $\mathcal{C}_n$, $n \geq 1$, pick a ball uniformly at random and remove it (and delete an empty box if one is created). Then the composition of the remaining $n-1$ balls has the same distribution as $\mathcal{C}_{n-1}$.

We call a composition structure $(\mathcal{C}_n, n \geq 1)$ *regenerative* if for all $n > n_1 > 1$, given that the first part of $\mathcal{C}_n$ is n_1, the remaining composition of $n - n_1$ has the same distribution as $\mathcal{C}_{n-n_1}$.

Lemma 3.4 ([60] Proposition 6)**.** *Let $\alpha \in (0,1)$ and $\theta > 0$, and consider an ordered CRP $(\widetilde{\Pi}_n = (\widetilde{\Pi}_{n,1}, \ldots, \widetilde{\Pi}_{n,K_n}), n \geq 1)$. For $j \in [n]$, define $S_{n,j} := \sum_{i=1}^{j} |\widetilde{\Pi}_{n,i}|$, i.e. $S_{n,j}$ is the number of customers seated at the first j tables from the left. Then,*

$$\left\{ \frac{S_{n,j}}{n}, j \geq 0 \right\} \to \mathcal{Z}_{\alpha,\theta} := \left\{ 1 - e^{-\xi_t}, t \geq 0 \right\}^{\mathrm{cl}} \qquad a.s.\ as\ n \to \infty,$$

with respect to the Hausdorff metric on closed subsets of $[0,1]$, where $^{\mathrm{cl}}$ denotes the closure in $[0,1]$, and $(\xi_t, t \geq 0)$ is a subordinator with Laplace exponent

$$\Phi_{\alpha,\theta}(s) = \frac{s\Gamma(s+\theta)\Gamma(1-\alpha)}{\Gamma(s+\theta+1-\alpha)}.$$

Furthermore, define $L_n(u) := \#\{j \in [K_n] : S_{n,j}/n \leq u\}$ for $u \in [0,1]$, where for any set S, $\#S$ denotes the number of elements in S. Then

$$\lim_{n\to\infty} \sup_{u \in [0,1]} |n^{-\alpha} L_n(u) - L(u)| = 0 \qquad a.s.,$$

where $L := (L(u), u \in [0,1])$ is a continuous local time process for $\mathcal{Z}_{\alpha,\theta}$ which means that the random set of points of increase of L is $\mathcal{Z}_{\alpha,\theta}$ a.s..

We refer to the collection of open intervals in $[0,1] \setminus \mathcal{Z}_{\alpha,\theta}$ as the *(α,θ)-regenerative interval partition* associated with $(\mathcal{C}_n, n \geq 1)$ and the local time process L, where $L(1) = L_{\alpha,\theta}$ a.s.. Note that the joint law of ranked lengths of components of this interval partition is $\mathrm{PD}(\alpha,\theta)$. The local time process $(L(u), 0 \leq u \leq 1)$ can be characterised via ξ, i.e. we have

$$L\Big(1 - e^{-\xi_t}\Big) = \Gamma(1-\alpha) \int_0^t e^{-\alpha \xi_s} ds, \quad t \geq 0.$$

We consider the inverse local time L^{-1} given by

$$L^{-1} \colon [0, L_{\alpha,\theta}] \to [0,1], \qquad L^{-1}(x) := \inf\{u \in [0,1] : L(u) > x\}. \tag{3.2}$$

Note that L^{-1} is increasing and right-continuous, and hence we can equip the random interval $[0, L_{\alpha,\theta}]$ with the Stieltjes measure dL^{-1}. We refer to the pair $([0, L_{\alpha,\theta}], dL^{-1})$ as an (α, θ)-string of beads in the following way.

Definition 3.5 (String of beads). A *string of beads* (I, μ) is an interval I equipped with a discrete mass measure μ. An (α, θ)*-string of beads* is the weighted random interval $([0, L_{\alpha,\theta}], dL^{-1})$ associated with an (α, θ)-regenerative interval partition.

We equip the space of strings of beads with the Borel σ-algebra associated with the topology of weak convergence of probability measures on $[0, \infty)$, where the length of a string of beads (I, μ) is determined by the supremum of the support of the measure μ.

We also use the term (α, θ)-string of beads for measure-preserving isometric copies of the weighted interval $([0, L_{\alpha,\theta}], dL^{-1})$. Since the lengths of the interval components of an (α, θ)-regenerative interval partition are the masses of the atoms of the associated (α, θ)-string of beads, we know that the joint law of ranked masses of the atoms of an (α, θ)-string of beads is $\mathrm{PD}(\alpha, \theta)$.

3.2.2 Regenerative composition structures

We recap some well-known results for regenerative composition structures from [38], some of which are based on [49]. For a (random) closed subset $\mathcal{M}$ of $[0, \infty]$ let

$$G(\mathcal{M}, t) := \sup \mathcal{M} \cap [0, t], \qquad D(\mathcal{M}, t) := \inf \mathcal{M} \cap (t, \infty].$$

A random closed subset $\mathcal{M} \subset [0, \infty]$ is *regenerative* if for each $t \in [0, \infty)$, conditionally given $\{D(\mathcal{M}, t) < \infty\}$, the random set $(\mathcal{M} - D(\mathcal{M}, t)) \cap [0, \infty]$ has the same distribution as $\mathcal{M}$ and is independent of $[0, D(\mathcal{M}, t)] \cap \mathcal{M}$.

Theorem 3.6 ([38] Theorem 5.1). *The closed range $\{S_t, t \geq 0\}^{cl}$ of a subordinator $(S_t, t \geq 0)$ is a regenerative random subset of $[0, \infty]$. Furthermore, for every regenerative random subset $\mathcal{M}$ of $[0, \infty]$ there is a subordinator $(S_t, t \geq 0)$ such that $\mathcal{M} \stackrel{d}{=} \{S_t, t \geq 0\}^{\mathrm{cl}}$.*

We call a process $(\widetilde{S}_t, t \geq 0)$ a *multiplicative subordinator* if for $t' > t$, the ratio $(1 - \widetilde{S}_{t'})/(1 - \widetilde{S}_t)$ has the same distribution as $(1 - \widetilde{S}_{t'-t})$ and is independent of $(\widetilde{S}_u, 0 \leq u \leq t)$. Furthermore, let $\widetilde{\mathcal{M}}$ denote the closed range of a multiplicative subordinator

$(\widetilde{S}_t, t \geq 0)$, i.e. $\widetilde{\mathcal{M}} := \{\widetilde{S}_t, t \geq 0\}^{cl}$. A multiplicative subordinator $(\widetilde{S}_t, t \geq 0)$ can be obtained from a subordinator $(S_t, t \geq 0)$ via the the mapping from $[0, \infty]$ onto $[0, 1]$ defined by $z \mapsto 1 - \exp(-z)$, i.e. $\widetilde{S}_t := 1 - \exp(-S_t)$.

For any closed subset M of $[0, 1]$ and any $z \in [0, 1]$ such that $M \cap (z, 1) \neq \emptyset$, we define $M(z)$ as the part of M strictly right to $D(M, z)$, scaled back to $[0, 1]$, i.e.

$$M(z) := \left\{ \frac{y - D(M, z)}{1 - D(M, z)} : y \in M \cap [D(M, z), 1] \right\}.$$

A random closed subset $\widetilde{\mathcal{M}} \subset [0, 1]$ is called *multiplicatively regenerative* if, for each $z \in [0, 1)$, conditionally given $\{D(\widetilde{\mathcal{M}}, z) < 1\}$ the random set $\widetilde{\mathcal{M}}(z)$ has the same distribution as $\widetilde{\mathcal{M}}$ and is independent of $[0, D(\widetilde{\mathcal{M}}, z)] \cap \widetilde{\mathcal{M}}$.

Proposition 3.7 ([38] Proposition 6.2). *For two random closed subsets $\widetilde{\mathcal{M}} \subset [0, 1]$ and $\mathcal{M} \subset [0, \infty]$ related by $\widetilde{\mathcal{M}} = 1 - \exp\{-\mathcal{M}\}$, the random set $\mathcal{M}$ is regenerative if and only if $\widetilde{\mathcal{M}}$ is multiplicatively regenerative.*

We associate with each composition $(n_1, \ldots, n_k)$ of n the finite closed set whose points are partial sums of the parts $n_1, \ldots, n_k$ divided by n, i.e.

$$(n_1, \ldots, n_k) \mapsto \{0, n_1/n, (n_1 + n_2)/n, \ldots, (n - n_k)/n, 1\} =: \widetilde{\mathcal{M}}_n. \tag{3.3}$$

Every regenerative composition structure $(\mathcal{C}_n, n \geq 1)$ is hence associated with a sequence of random sets $(\widetilde{\mathcal{M}}_n, n \geq 1)$. Note that each $\widetilde{\mathcal{M}}_n$ defines an interval partition of $[0, 1]$, namely $[0, 1] \setminus \widetilde{\mathcal{M}}_n$, $n \geq 1$.

Lemma 3.8 ([38] Lemma 6.3). *Let $(\mathcal{C}_n, n \geq 1)$ be a composition structure and let $(\widetilde{\mathcal{M}}_n, n \geq 1)$ be the associated sequence of random sets as in (3.3). Then $\widetilde{\mathcal{M}}_n \to \widetilde{\mathcal{M}}$ a.s. as $n \to \infty$ in the Hausdorff metric for some random closed subset $\widetilde{\mathcal{M}}$.*

Lemma 3.9 ([38] Corollary 6.4). *In the setting of Lemma 3.8, a composition structure $(\mathcal{C}_n, n \geq 1)$ is regenerative if and only if $\widetilde{\mathcal{M}}$ is multiplicatively regenerative.*

For $\alpha \in (0, 1)$ we can construct an $(\alpha, 0)$-regenerative interval partition, arising in the limit of the regenerative composition structure induced by an ordered $(\alpha, 0)$-CRP, as the restriction to $[0, 1]$ of the range of a stable subordinator of index α.

Definition 3.10 (Stable subordinator)**.** A subordinator $(S_t, t \geq 0)$ is called *stable* with index $\alpha \in (0,1)$ if it has the *self-similarity* property, that is, for all $t > 0$, $t^{-1/\alpha} S_t \stackrel{d}{=} S_1$.

Proposition 3.11 ([38] Section 8.3)**.** *Let $\alpha \in (0,1)$, and let $\mathcal{M}_\alpha$ be the range of a stable subordinator of index $\alpha > 0$. Then, the interval partition of $[0,1]$ generated by $\mathcal{M}_\alpha \cap [0,1]$ is an $(\alpha, 0)$-regenerative interval partition.*

3.3 Merging strings of beads

3.3.1 The merging algorithm

We first explain the basic structure of our merging operation for strings of beads, given a sequence of cut points $(x_n)_{n\geq 1}$, that is, a sequence of atoms on the strings of beads. We use $\mathbb{R}$ equipped with the usual distance function as the underlying metric space, and refer to isometric copies of the related intervals when we present applications to continuum trees in Section 3.4.

Let $k \in \mathbb{N}$, and consider k disjoint strings of beads (E_i, μ_i), $i \in [k]$, given by

$$E_i := (\rho_i, \rho_i'), \qquad i \in [k],$$

where $E_i \cap E_j = \emptyset$ for all $i, j \in [k], i \neq j$. Furthermore, suppose that, for every $i \in [k]$, we have a sequence $(\rho_{i,j})_{j\geq 1}$ such that

$$E_i = \bigcup_{j=0}^{\infty} (\rho_{i,j}, \rho_{i,j+1}] \quad \text{and } d(\rho_i, \rho_{i,j}) < d(\rho_i, \rho_{i,j+1}) \text{ for all } j \geq 1, \tag{3.4}$$

where $\rho_{i,0} := \rho_i$, $i \in [k]$. We refer to the sequences $(\rho_{i,j})_{j\geq 1}$, $i \in [k]$, as *cut points* for the string of beads (E_i, μ_i), $i \in [k]$. Define the *set of cut points* Υ by

$$\Upsilon := \{\rho_{i,j}, j \geq 1, i \in [k]\}, \tag{3.5}$$

and assume that $\sigma_\Upsilon : \Upsilon \to \mathbb{N}$ gives a total order on Υ (i.e. σ_Υ is bijective) which is consistent with the natural order given by $\mathbb{N}$ when restricted to $\{\rho_{i,j}, j \geq 1\}$, $i \in [k]$, in the sense that $\sigma_\Upsilon(\rho_{i,j}) < \sigma_\Upsilon(\rho_{i,j+1})$ for all $j \geq 1$. For $n \in \mathbb{N}$, we write $x_n := \sigma_\Upsilon^{-1}(n)$, i.e. $\Upsilon = \{x_n, n \geq 1\}$.

We define the *merged* string of beads $(E_{\uplus}, \mu)$ equipped with the metric $d_{\uplus}$ by

$$E_{\uplus} := \biguplus_{i=1}^{k} E_i = \biguplus_{i=1}^{k} (\rho_i, \rho_i') \tag{3.6}$$

where the operator $\uplus$ on the intervals E_i, $i \in [k]$, the metric $d_{\uplus} : E_{\uplus} \times E_{\uplus} \to \mathbb{R}_0^+$ and the mass measure μ on $E_{\uplus}$ are defined as follows.

- For any $n \geq 1$, let $(a_n, b_n]$ be defined by

$$(a_n, b_n] := (\rho_{i,j-1}, \rho_{i,j}] \tag{3.7}$$

  where $\rho_{i,j}$ denotes the unique element of Υ with $\sigma_\Upsilon(\rho_{i,j}) = n$, i.e. $x_n = \rho_{i,j}$.

- Define the set $E_{\uplus}$ by

$$E_{\uplus} := \bigcup_{n=1}^{\infty} (a_n, b_n] = \bigcup_{i=1}^{k} E_i, \tag{3.8}$$

  and equip it with the metric $d_{\uplus}$ carried forward via these operations, i.e.

$$d_{\uplus}(x, y) := \begin{cases} d(x, b_n) + \sum\limits_{k=n+1}^{m-1} d(a_k, b_k) + d(a_m, y), & x \in (a_n, b_n], y \in (a_m, b_m], \\ d(x, y), & x, y \in (a_n, b_n], n \geq 1, \end{cases} \tag{3.9}$$

  where $n < m$ in the first case. We use the subscript $\uplus$ to underline that the set $E_{\uplus} = \bigcup_{i=1}^{k} E_i$ is equipped with the metric $d_{\uplus}$.

- Considering the natural mass measure μ on the Borel sets $\mathcal{B}(E_{\uplus})$ given by

$$\mu(A) = \sum_{i=1}^{k} \mu_i(A \cap E_i), \quad A \in \mathcal{B}\Big(E_{\uplus}\Big), \tag{3.10}$$

  yields a metric space $(E_{\uplus}, d_{\uplus})$ endowed with a mass measure μ.

Note that $(E_{\uplus}, d_{\uplus})$ is isometric to an open interval. Considering its completion $[\rho, \rho']$, we can write $E_{\uplus} = (\rho, \rho')$ where $\rho := a_0$ and ρ' is the unique element in the completion of $E_{\uplus}$ such that $d_{\uplus}(b_n, \rho') \to 0$ as $n \to \infty$. Hence, $(E_{\uplus}, \mu)$ is a string of beads equipped with the metric $d_{\uplus}$. Furthermore, note that $d_{\uplus}(\rho, \rho') = \sum_{i=1}^{k} d(\rho_i, \rho_i')$.

Remark 3.12. Note that we do not necessarily need *open* intervals (ρ_i, ρ_i'), $i \in [k]$, in the merging algorithm. In the case when $E_l = [\rho_l, \rho_l')$ for some $l \in [k]$, we interpret our algorithm in the following way. Consider the first cut point $\rho_{l,1}$ in E_l, i.e. $\rho_{l,1} = x_n$ such that $x_n \in E_l$ and $x_m \notin E_l$ for all $1 \leq m \leq n-1$.

- If $\rho_{l,1} \neq x_1$, then there is $n \in \mathbb{N}$, such that $b_n = \rho_{l,1}$ and $b_{n-1} = \rho_{i,j}$ for some $j \geq 1$ and $i \in [k] \backslash \{l\}$. In this case, we consider the interval component $(a_n, b_n] = (\rho_l, \rho_{l,1}]$ in our algorithm, and set $\mu(b_{n-1}) := \mu_i(\rho_{i,j}) + \mu_l(\rho_l)$.
- If $\rho_{l,1} = x_1$, then $[a_1, b_1] = [\rho_l, x_1]$ and we obtain an interval $[\rho = \rho_l, \rho')$ as a result of the merging procedure with $\mu(a_1) := \mu_l(\rho_l)$.

If $E_l = (\rho_l, \rho_l']$, we might have that $x_n = \rho_l'$ for some $n \in \mathbb{N}$. Then the interval E_l is split into only finitely many components. If $E_l = [\rho_l, \rho_l']$ is a closed interval, we naturally combine the described conventions.

3.3.2 Cut point sampling and main result

We now assume that our strings of beads are rescaled (α, θ_i)-string of beads for $\alpha \in (0,1)$ and $\theta_i > 0$, $i \in [k]$. More precisely, let $E_i = (\rho_i, \rho_i')$, $i \in [k]$, and let μ be a mass measure on $E = \bigcup_{i=1}^k E_i$ such that

$$(\mu(E_1), \ldots, \mu(E_k)) \sim \text{Dirichlet}(\theta_1, \ldots, \theta_k). \tag{3.11}$$

Furthermore, suppose that

$$\Big(\mu(E_i)^{-\alpha} E_i, \mu(E_i)^{-1} \mu \restriction_{E_i}\Big), \qquad i \in [k],$$

are (α, θ_i)-strings of beads, respectively, independent of each other and of (3.11). We construct an $(\alpha, \theta = \sum_{i=1}^k \theta_i)$-string of beads from $(E_i, \mu \restriction_{E_i}), i \in [k]$, using the merging algorithm presented in Section 3.3.1. The procedure is based on sampling an atom of an (α, θ)-string of beads such that the induced mass split has a Dirichlet$(\alpha, 1-\alpha, \theta)$ distribution. Pitman and Winkel [60] describe a sampling procedure for such an atom, and show that the atom splits the (α, θ)-string of beads into a rescaled independent (α, α)- and a rescaled independent (α, θ)-string of beads.

Proposition 3.13 ([60] Proposition 10/14(b), Corollary 15)**.** *Consider an (α, θ)-string of beads $(I, \mu) := ([0, L_{\alpha,\theta}], dL^{-1})$ for some $\alpha \in (0, 1)$ and $\theta > 0$ with associated (α, θ)-regenerative interval partition $[0, 1] \setminus \mathcal{Z}_{\alpha,\theta}$, where $\mathcal{Z}_{\alpha,\theta} = \{1 - \exp(-\xi_t), t \geq 0\}^{cl}$, cf. Lemma 3.4. Define a switching probability function by*

$$p\colon [0,1] \to [0,1], \quad u \mapsto \frac{(1-u)\theta}{(1-u)\theta + u\alpha}.$$

For $t \geq 0$, select a block $(1 - e^{-\xi_{t-}}, 1 - e^{-\xi_t})$ of $[0, 1] \setminus \mathcal{Z}_{\alpha,\theta}$ with $\Delta\xi_t := \xi_t - \xi_{t-} > 0$ with probability

$$p\Big(e^{-\Delta\xi_t}\Big) \prod_{s<t} \Big(1 - p\Big(e^{-\Delta\xi_s}\Big)\Big)$$

and let X be the local time at the selected block $(1 - e^{-\xi_{t-}}, 1 - e^{-\xi_t})$, which we call switching time, i.e. $X := L(1 - e^{-\xi_t})$. The following random variables are independent.

- *The mass split $(\mu([0, X)), \mu(X), \mu((X, L_{\alpha,\theta}])) \sim \text{Dirichlet}(\alpha, 1 - \alpha, \theta)$;*
- *the (α, α)-string of beads $(\mu([0, X))^{-\alpha}[0, X), \mu([0, X))^{-1}\mu \restriction_{[0,X)})$;*
- *the (α, θ)-string of beads $(\mu((X, L_{\alpha,\theta}])^{-\alpha}(X, L_{\alpha,\theta}], \mu((X, L_{\alpha,\theta}])^{-1}\mu \restriction_{(X,L_{\alpha,\theta}]})$.*

We refer to the sampling procedure described in Proposition 3.13 as (α, θ)-*coin tossing sampling* in the string of beads $([0, L_{\alpha,\theta}], dL^{-1})$. Informally speaking, we can visualise this by a walker starting in 0, walking up the string of beads $([0, L_{\alpha,\theta}], dL^{-1})$, tossing a coin for each of the (infinite number of atom) masses where the probability for heads is given by $p(\exp(-\Delta\xi_t))$, and hence depends on the relative remaining mass after an atom; let the walker stop the first time the tossed coin shows heads and select the related atom.

Algorithm 3.14 (Construction of an (α, θ)-string of beads)**.** Let $\alpha \in (0, 1)$ and $\theta_1, \ldots, \theta_k > 0$.

- *Input.* A mass measure μ on $E := \bigcup_{i=1}^k E_i$, $E_i = (\rho_i, \rho_i')$, $i \in [k]$, such that

$$(\mu(E_1), \ldots, \mu(E_k)) \sim \text{Dirichlet}(\theta_1, \ldots, \theta_k) \tag{3.12}$$

where the pairs $(\mu(E_i)^{-\alpha}E_i, \mu(E_i)^{-1}\mu\restriction_{E_i})$, $i \in [k]$, are (α, θ_i)-strings of beads, respectively, independent of each other and of the mass split (3.12).

- *Output.* A sequence of cut points $(X_n)_{n\geq 1}$ and a sequence of interval components $\{(a_n, b_n]\}_{n\geq 1}$ of $E_i, i \in [k]$.

Initialisation. $n = 1$, $E_i^{(1)} := E_i$, $i \in [k]$, and $E^{(1)} := \bigcup_{i\in[k]} E_i$, $\rho_i^{(0)} := \rho_i$, $i \in [k]$.

For $n \geq 1$, given (E, μ) and the previous steps of the algorithm, do the following.

(i) Select one of the intervals $E_i^{(n)}$, $i \in [k]$, proportionally to mass, i.e. let I_n be a random variable taking values in $\{1, \ldots, k\}$ such that, for $i \in [k]$, $I_n = i$ with probability $\mu(E_i^{(n)})/\mu(E^{(n)})$.

Conditionally given I_n, pick an atom X_n of the interval $E_{I_n}^{(n)}$ according to (α, θ_{I_n})-coin tossing sampling in the string of beads

$$\left(\mu\left(E_{I_n}^{(n)}\right)^{-\alpha} E_{I_n}^{(n)}, \mu\left(E_{I_n}^{(n)}\right)^{-1} \mu\restriction_{E_{I_n}^{(n)}}\right).$$

(ii) Define the interval $(a_n, b_n]$ by

$$(a_n, b_n] := (\rho_{I_n}^{(n-1)}, X_n]. \tag{3.13}$$

(iii) Update $\rho_{I_n}^{(n)} := X_n$ and $\rho_i^{(n)} := \rho_i^{(n-1)}$ for $i \in [k] \setminus \{I_n\}$, i.e.

$$E_{I_n}^{(n+1)} := (X_n, \rho'_{I_n}), \quad E_i^{(n+1)} := E_i^{(n)} \text{ for } i \in [k] \setminus \{I_n\}, \tag{3.14}$$

and set $E^{(n)} = \bigcup_{i=1}^k E_i^{(n)}$.

We can define $E_{\uplus}$ by aligning the intervals $(a_n, b_n]$ in order, i.e. by identifying a_{n+1} with b_n, $n \geq 1$. More precisely, we consider $\Upsilon := \{X_n, n \geq 1\}$, $\sigma_\Upsilon : \Upsilon \to \mathbb{N}$, $\sigma(X_n) := n$, and define

$$E_{\uplus} := \biguplus_{n\geq 1} (a_n, b_n],$$

equipped with the metric $d_{\uplus}$, where the operator $\uplus$ was defined in (3.6)-(3.9). We can write $E_{\uplus} = (\rho, \rho')$. Our main result here is the following.

Theorem 3.15 (Merging of (α,θ)-strings of beads)**.** *Let $\alpha \in (0,1)$ and $\theta_1,\ldots,\theta_k > 0$, $k \in \mathbb{N}$. Let $E_1,\ldots,E_k$ be k disjoint intervals and μ a mass measure on $E := \bigcup_{i=1}^k E_i$ such that $(\mu(E_1),\ldots,\mu(E_k)) \sim \mathrm{Dirichlet}(\theta_1,\ldots,\theta_k)$. Furthermore, suppose that the rescaled pairs $(\mu(E_i)^{-\alpha}E_i, \mu(E_i)^{-1}\mu\restriction_{E_i})$ are (α,θ_i)-string of beads, $i \in [k]$, respectively, independent of each other and of the Dirichlet mass split $(\mu(E_1),\ldots,\mu(E_k))$. Let $E_{\biguplus} := \biguplus_{i=1}^k E_i$ be constructed as in Algorithm 3.14 with associated distance $d_{\biguplus}$ and mass measure μ. Then the pair $(E_{\biguplus},\mu)$ is an (α,θ)-string of beads with $\theta := \sum_{i=1}^k \theta_i$.*

3.3.3 Proof of the main result

We first recall and establish some preliminary results, including decomposition and construction rules for (α,θ)-regenerative interval partitions and (α,θ)-strings of beads.

3.3.3.1 Some preliminary results

(α,θ)-regenerative interval partition can be constructed as follows.

Proposition 3.16 ([60] Corollary 8)**.** *Let $\alpha \in (0,1)$ and $\theta \geq 0$, and consider a random vector $(G, D-G, 1-D) \sim \mathrm{Dirichlet}(\alpha, 1-\alpha, \theta)$. Conditionally given (G,D), construct an interval partition of $[0,1]$ as follows. Let (G,D) be one component, and construct the interval components in $[0,G]$ and in $[D,1]$ by linear scaling of an independent (α,α)- and an independent (α,θ)-regenerative interval partition, respectively. The interval partition obtained in this way is an (α,θ)-regenerative interval partition.*

Remark 3.17 (The special case $\theta = 0$)*.* The case $\theta = 0$ is considered in [57] Proposition 15. In this case, $G \sim \mathrm{Beta}(\alpha, 1-\alpha)$ and $(G,1)$ is the last interval component of $[0,1] \setminus \mathcal{Z}_{\alpha,0}$, and the restriction of $\mathcal{Z}_{\alpha,0}$ to $[0,G]$ is a rescaled copy of $\mathcal{Z}_{\alpha,\alpha}$, where $\mathcal{Z}_{\alpha,\theta}$, $\alpha \in (0,1)$, $\theta \geq 0$, was defined in Lemma 3.4.

We can formulate an analogous version of Proposition 3.16 for (α,θ)-strings of beads, $\alpha \in (0,1)$ and $\theta \geq 0$, which is based upon [60] Proposition 14(b).

Proposition 3.18 ([60] Proposition 14(b))**.** *Let $\alpha \in (0,1)$ and $\theta \geq 0$, and consider a random vector $(G, D-G, 1-D) \sim \mathrm{Dirichlet}(\alpha, 1-\alpha, \theta)$. Conditionally given (G,D), let $(I_1 = [0, L_{\alpha,\alpha}], \mu_1)$ be an independent (α,α)-string of beads and $(I_2 = [0, L_{\alpha,\theta}], \mu_2)$*

an independent (α,θ)-string of beads. Define

$$L'_{\alpha,\theta} := G^\alpha L_{\alpha,\alpha} + (1-D)^\alpha L_{\alpha,\theta},$$

the interval $I := [0, L'_{\alpha,\theta}]$ and the mass measure μ on I by

$$\mu(y) = \begin{cases} G \cdot \mu_1(G^{-\alpha} y) & \text{if } y \in [0, G^\alpha L_{\alpha,\alpha}), \\ D - G & \text{if } y = G^\alpha L_{\alpha,\alpha}, \\ (1-D)\cdot \mu_2((1-D)^{-\alpha}(y - G^\alpha L_{\alpha,\alpha})) & \text{if } y \in (G^\alpha L_{\alpha,\alpha}, L'_{\alpha,\theta}]. \end{cases}$$

Then (I,μ) is an (α,θ)-string of beads.

The following theorem shows how an (α,θ)-regenerative interval partition, $\alpha \in (0,1)$ and $\theta > 0$, can be constructed via a stick-breaking scheme, and a sequence of i.i.d. stable subordinators of index α.

Proposition 3.19 ([38] Theorem 8.3)**.** *Let $(Y_i)_{i\geq 1}$ be i.i.d. with $B_1 \sim \mathrm{Beta}(1,\theta)$ for some $\theta > 0$, and define the sequence $(V_n)_{n\geq 1}$ by*

$$V_n := 1 - \prod_{i=1}^{n} (1 - Y_i), \qquad n \geq 1. \tag{3.15}$$

For $\alpha \in (0,1)$, let $\mathcal{M}_\alpha(n)$, $n \geq 1$, be independent copies of the range $\mathcal{M}_\alpha$ of a stable subordinator of index α, and define a random closed subset $\widetilde{\mathcal{M}}(\alpha,\theta) \subset [0,1]$ by

$$\widetilde{\mathcal{M}}(\alpha,\theta) := \{1\} \cup \bigcup_{n=1}^{\infty} ([V_{n-1}, V_n] \cap (V_{n-1} + \mathcal{M}_\alpha(n))) \tag{3.16}$$

where $V_0 := 0$. Then, $\widetilde{\mathcal{M}}(\alpha,\theta)$ is a multiplicatively regenerative random subset of $[0,1]$ which can be represented as $\widetilde{\mathcal{M}}(\alpha,\theta) = 1 - \exp(-\mathcal{M}(\alpha,\theta))$, where $\mathcal{M}(\alpha,\theta)$ is the range of a subordinator with Laplace exponent

$$\Phi_{\alpha,\theta}(s) = \frac{s\Gamma(s+\theta)\Gamma(1-\alpha)}{\Gamma(s+\theta+1-\alpha)}.$$

Remark 3.20 (Sliced splitting). Note that $[0,1]\setminus\widetilde{\mathcal{M}}(\alpha,\theta)$ is an (α,θ)-regenerative interval partition. Gnedin [37] refers to the method presented in Proposition 3.19 as *sliced* *splitting*: first split the interval $[0,1]$ according to the stick-breaking scheme with $\mathrm{Beta}(1,\theta)$ variables, $\theta>0$, to obtain the stick-breaking points $(V_n)_{n\geq 1}$. Then, for each $n\geq 1$, use an independent copy of a regenerative set derived from a stable subordinator with index $\alpha\in(0,1)$ to split the interval (V_{n-1},V_n), $n\geq 1$, $V_0:=0$. In other words, we first split the interval $[0,1]$ according to a $(0,\theta)$-regenerative interval partition with $\theta>0$, and then shatter each part according to an $(\alpha,0)$-regenerative interval partition, $\alpha\in(0,1)$.

As a consequence of Proposition 3.19, an (α,θ)-regenerative interval partition can be constructed via independent $(\alpha,0)$-regenerative interval partitions.

Corollary 3.21. *Let $(Y_i)_{i\geq 1}$ be i.i.d. with $Y_1\sim\mathrm{Beta}(1,\theta)$ for some $\theta>0$, and let $(V_n)_{n\geq 1}$ be defined as in (3.15). For $\alpha\in(0,1)$, let $\mathcal{M}_\alpha^*(n)$, $n\geq 1$, be independent copies of the random closed subset $\mathcal{M}_\alpha^*$ of $[0,1]$ associated with an $(\alpha,0)$-regenerative interval partition of $[0,1]$ given by $[0,1]\setminus\mathcal{M}_\alpha^*$. Define the random set $\widetilde{\mathcal{M}}^*(\alpha,\theta)$ by*

$$\widetilde{\mathcal{M}}^*(\alpha,\theta):=\{1\}\cup\bigcup_{n=1}^{\infty}(V_{n-1}+(V_n-V_{n-1})\mathcal{M}_\alpha^*(n)). \tag{3.17}$$

Then $\widetilde{\mathcal{M}}^*(\alpha,\theta)$ is a multiplicatively regenerative random subset of $[0,1]$, and hence defines an (α,θ)-regenerative interval partition of $[0,1]$ via $[0,1]\setminus\widetilde{\mathcal{M}}^*(\alpha,\theta)$.

Proof. We apply Proposition 3.19, i.e. we show that the sets $V_{n-1}+(V_n-V_{n-1})\mathcal{M}_\alpha^*(i)$ are of the form $[V_{n-1},V_n]\cap(V_{n-1}+\mathcal{M}_\alpha(n))$ as in (3.16). Consider the closure $\mathcal{M}_\alpha$ of the range of a stable subordinator S_α of index $\alpha\in(0,1)$ as defined in Proposition 3.19, i.e. $\mathcal{M}_\alpha:=\{S_\alpha(t),t\geq 0\}^{\mathrm{cl}}$. Since S_α is stable, we obtain, for any $c>0$,

$$\{S_\alpha(t),t\geq 0\}^{\mathrm{cl}}=\{S_\alpha(tc),t\geq 0\}^{\mathrm{cl}}\overset{d}{=}\{c^{1/\alpha}S_\alpha(t),t\geq 0\}^{\mathrm{cl}}=c^{1/\alpha}\{S_\alpha(t),t\geq 0\}^{\mathrm{cl}}.$$

Therefore, for any $r>0$, $r\mathcal{M}_\alpha\overset{d}{=}\mathcal{M}_\alpha$, and hence, for any $r>0$, we have

$$[0,r]\cap\mathcal{M}_\alpha\overset{d}{=}r\cdot([0,1]\cap\mathcal{M}_\alpha).$$

Furthermore, by stationary and independent increments of subordinators, and the independence of $(V_n)_{n\geq 1}$ and $\mathcal{M}_\alpha(n)$, $n \geq 1$, in Proposition 3.19, we obtain

$$[V_{n-1}, V_n] \cap (V_{n-1} + \mathcal{M}_\alpha(n)) \stackrel{d}{=} V_{n-1} + [0, V_n - V_{n-1}] \cap \mathcal{M}_\alpha(n).$$

Hence, by Proposition 3.11, the sets $[0, V_n - V_{n-1}] \cap \mathcal{M}_\alpha(n)$, $n \geq 1$, define independent $(\alpha, 0)$-regenerative interval partitions of $[0, 1]$ rescaled by the factor $(V_n - V_{n-1})$. More precisely, we use that if $\mathcal{V}$ is a random interval partition or closed subset of $[0, \infty]$ such that, for any $r > 0$, $r\mathcal{V} \stackrel{d}{=} \mathcal{V}$ and W is an independent random variable such that $0 < W < \infty$ a.s., then $W\mathcal{V} \stackrel{d}{=} \mathcal{V}$.

Consequently, the sets $[0, V_n - V_{n-1}] \cap \mathcal{M}_\alpha(n)$, $n \geq 1$, are independent $(\alpha, 0)$-regenerative interval partitions rescaled by $(V_n - V_{n-1})$, and hence each set $[0, V_n - V_{n-1}] \cap \mathcal{M}_\alpha(n)$, $n \geq 1$, has the same distribution as $(V_n - V_{n-1})\mathcal{M}^*_\alpha(n)$. By Proposition 3.19, we conclude that $\widetilde{\mathcal{M}}^*(\alpha, \theta)$ defines an (α, θ)-regenerative interval partition. □

3.3.3.2 Strings of beads: splitting and merging

The proof of Theorem 3.15 uses an induction on the splitting steps $n \geq 1$. In Lemma 3.23 we describe the initial step. Its proof requires some basic properties of the Dirichlet distribution.

Proposition 3.22. *Let* $X := (X_1, \dots, X_k) \sim \mathrm{Dirichlet}(\theta_1, \dots, \theta_k)$ *for some* $\theta_i > 0$, $i \in [k]$, *and* $k \in \mathbb{N}$.

(i) *Aggregation. For* $i, j \in [k]$ *with* $i < j$ *define*

$$X' := (X_1, \dots, X_{i-1}, X_i + X_j, X_{i+1}, \dots, X_{j-1}, X_{j+1}, \dots, X_k).$$

Then, $X' \sim \mathrm{Dirichlet}(\theta_1, \dots, \theta_{i-1}, \theta_i + \theta_j, \theta_{i+1}, \dots, \theta_{j-1}, \theta_{j+1}, \dots, \theta_k)$.

(ii) *Decimation. Let* $\alpha_1, \alpha_2, \alpha_3 \in (0, 1)$ *be such that* $\alpha_1 + \alpha_2 + \alpha_3 = 1$. *Fix* $i \in [k]$, *and consider a random vector* $(P_1, P_2, P_3) \sim \mathrm{Dirichlet}(\alpha_1\theta_i, \alpha_2\theta_i, \alpha_3\theta_i)$ *which is independent of* X. *Then*

$$X'' := (X_1, \dots, X_{i-1}, P_1X_i, P_2X_i, P_3X_i, X_{i+1}, \dots, X_k).$$

has a $\mathrm{Dirichlet}(\theta_1, \ldots, \theta_{i-1}, \alpha_1\theta_i, \alpha_2\theta_i, \alpha_3\theta_i, \theta_{i+1}, \ldots, \theta_k)$ *distribution.*

(iii) *Size-bias. Let I be an index chosen such that $\mathbb{P}(I = i | (X_1, \ldots, X_k)) = X_i$ a.s. for $i \in [k]$. Then, for any $i \in [k]$, conditionally given $I = i$,*

$$X = (X_1, \ldots, X_k) \sim \mathrm{Dirichlet}(\theta_1, \ldots, \theta_{i-1}, \theta_i + 1, \theta_{i+1}, \ldots, \theta_k).$$

(iv) *Two-dimensional marginals. For any $i, j \in [k]$ with $i \neq j$,*

$$X_i/(X_i + X_j) \sim \mathrm{Beta}(\theta_i, \theta_j).$$

(v) *Deletion. For any $i \in [k]$, the vector*

$$X^* := (X_1/(1 - X_i), \ldots, X_{i-1}/(1 - X_i), X_{i+1}/(1 - X_i), \ldots, X_k/(1 - X_i))$$

has a $\mathrm{Dirichlet}(\theta_1, \ldots, \theta_{i-1}, \theta_{i+1}, \ldots, \theta_k)$ *distribution.*

Proof. (i) and (ii) can be found in [67] as Proposition 13 and Proposition 14/Remark 15, for instance. (iii) is Lemma 17 in [4]. (iv) and (v) follow directly from the representation of the Dirichlet distribution via independent Gamma variables. □

Lemma 3.23. *In Algorithm 3.14, the following random variables are independent.*

- *The mass* $\mu((\rho_{I_1}, X_1]) \sim \mathrm{Beta}(1, \theta)$;
- *the $(\alpha, 0)$-string of beads* $(\mu((\rho_{I_1}, X_1])^{-\alpha}(\rho_{I_1}, X_1], \mu((\rho_{I_1}, X_1])^{-1}\mu \restriction_{(\rho_{I_1}, X_1]})$;
- *the (α, θ_i)-strings of beads* $(\mu(E_i^{(2)})^{-\alpha}E_i^{(2)}, \mu(E_i^{(2)})^{-1}\mu \restriction_{E_i^{(2)}}), i \in [k]$;
- *the mass split* $\mu(E^{(2)})^{-1}(\mu(E_1^{(2)}), \ldots, \mu(E_k^{(2)})) \sim \mathrm{Dirichlet}(\theta_1, \ldots, \theta_k)$.

The proof of Lemma 3.23 will use the following basic proposition, whose proof is a simple exercise.

Proposition 3.24. *Let I be a random variable taking values in a countable set $\mathbb{I}$. Futhermore, consider a random variable Z taking values in some measurable space $(\mathbb{V}, \mathcal{Z})$ whose distribution conditional on $I = i$ does not depend on $i \in \mathbb{I}$. Then, Z is independent* *of I, and for any other random variable Z' taking values in some measurable space $(\mathbb{V}', \mathcal{Z}')$ that is conditionally independent of Z given I, (Z', I) is independent of Z.*

Proof of Lemma 3.23. Conditionally given I_1, i.e. given that the interval E_{I_1} is selected in the first step of Algorithm 3.14, by Proposition 3.22(iii), we have

$$(\mu(E_1), \ldots, \mu(E_{I_1-1}), \mu(E_{I_1}), \mu(E_{I_1+1}), \ldots, \mu(E_k))$$
$$\sim \text{Dirichlet}(\theta_1, \ldots, \theta_{I_1-1}, \theta_{I_1}+1, \theta_{I_1+1}, \ldots, \theta_{I_k}).$$

Note that, conditionally given I_1, X_1 is an element of E_{I_1} which is independent of $E_i, i \in [k] \setminus \{I_1\}$. We apply (α, θ_{I_1})-coin tossing sampling in the (α, θ_{I_1})-string of beads $(\mu(E_{I_1})^{-\alpha} E_{I_1}, \mu(E_{I_1})^{-1} \mu \restriction_{E_{I_1}})$, which gives an atom X_1 splitting E_{I_1} according to Dirichlet$(\alpha, 1-\alpha, \theta_{I_1})$, i.e.

$$\mu(E_{I_1})^{-1}(\mu((\rho_{I_1}, X_1)), \mu(X_1), \mu((X_1, \rho'_{I_1}))) \sim \text{Dirichlet}(\alpha, 1-\alpha, \theta_{I_1}), \tag{3.18}$$

see Proposition 3.13. Proposition 3.22(ii), with $i = I_1$, $P_1 = \mu((\rho_{I_1}, X_1))/\mu(E_{I_1})$, $P_2 = \mu(X_1)/\mu(E_{I_1})$, $P_3 = \mu((X_1, \rho'_{I_1}))/\mu(E_{I_1})$, and $\alpha_1 = \alpha/(\theta_{I_1}+1)$, $\alpha_2 = (1-\alpha)/(\theta_{I_1}+1)$, $\alpha_3 = \theta_{I_1}/(\theta_{I_1}+1)$, yields that the distribution of

$$(\mu((\rho_{I_1}, X_1)), \mu(X_1), \mu(E_1), \ldots, \mu(E_{I_1-1}), \mu((X_1, \rho'_{I_1})), \mu(E_{I_1+1}), \ldots, \mu(E_k)) \tag{3.19}$$

is Dirichlet$(\alpha, 1-\alpha, \theta_1, \ldots, \theta_{I_1-1}, \theta_{I_1}, \theta_{I_1+1}, \ldots, \theta_k)$. By Proposition 3.13, the pairs

$$\left(\mu((\rho_{I_1}, X_1))^{-\alpha}(\rho_{I_1}, X_1), \mu((\rho_{I_1}, X_1))^{-1}\mu \restriction_{(\rho_{I_1}, X_1)}\right) \tag{3.20}$$

and

$$\left(\mu((X_1, \rho'_{I_1}))^{-\alpha}(X_1, \rho'_{I_1}), \mu((X_1, \rho'_{I_1}))^{-1}\mu \restriction_{(X_1, \rho'_{I_1})}\right) \tag{3.21}$$

are (α, α)- and (α, θ_{I_1})-strings of beads, respectively, independent of each other and of the mass split (3.18).

By Proposition 3.22(iv) and (3.19) we have $\mu((\rho_{I_1}, X_1))/\mu((\rho_{I_1}, X_1]) \sim \text{Beta}(\alpha, 1-\alpha)$. Furthermore, by Proposition 3.22 (i), $\mu((\rho_{I_1}, X_1]) \sim \text{Beta}(1, \theta)$ with $\theta = \sum_{i=1}^k \theta_i$. By

Proposition 3.16, Remark 3.17 and Proposition 3.18, we conclude that

$$\left(\mu((\rho_{I_1}, X_1])^{-\alpha}(\rho_{I_1}, X_1], \mu((\rho_{I_1}, X_1])^{-1}\mu\restriction_{(\rho_{I_1}, X_1]}\right)$$

is an $(\alpha, 0)$-string of beads. The distribution of $\mu(E^{(2)})^{-1}(\mu(E_1^{(2)}), \ldots, \mu(E_k^{(2)}))$ follows directly from Proposition 3.22(v). The strings of beads

$$\left(\mu(E_i^{(2)})^{-\alpha}E_i^{(2)}, \mu(E_i^{(2)})^{-1}\mu\restriction_{E_i^{(2)}}\right), \quad i \in [k],$$

are independent (α, θ_i)-strings of beads, as, for $i \in [k] \setminus \{I_1\}$, these are not affected by the algorithm, conditionally given I_1, and $E_{I_1}^{(2)} = (X_1, \rho'_{I_1})$.

To deduce the claimed independence, note that the distributions of the four random variables mentioned in Lemma 3.23 do not depend on I_1, i.e. we can apply Proposition 3.24 to these variables inductively with $I = I_1$. □

We are now ready to prove our main result, Theorem 3.15.

Proof of Theorem 3.15. We use Corollary 3.21 to show the claim. Let $(X_n)_{n\geq 1}$ be a sequence of random atoms of E, as sampled in Step (i) of Algorithm 3.14.

By Lemma 3.23, the distribution of the mass split obtained via X_1

$$(\mu((\rho_{I_1}, X_1)), \mu(X_1), \mu(E_1), \ldots, \mu(E_{I_1-1}), \mu((X_1, \rho'_{I_1})), \mu(E_{I_1+1}), \ldots, \mu(E_k))$$

is Dirichlet$(\alpha, 1-\alpha, \theta_1, \ldots, \theta_{I_1-1}, \theta_{I_1}, \theta_{I_1+1}, \ldots, \theta_k)$. Again by Lemma 3.23,

$$Y_1 := \mu((\rho_{I_1}, X_1]) \sim \text{Beta}(1, \theta),$$

where Y_1 is independent of the (α, θ_i)-strings of beads

$$\left(\mu(E_i^{(2)})^{-\alpha}E_i^{(2)}, \mu(E_i^{(2)})^{-1}\mu\restriction_{E_i^{(2)}}\right), \quad i \in [k], \tag{3.22}$$

which are also independent of each other. Furthermore, Y_1 and the k strings of beads in (3.22) are also independent of the $(\alpha, 0)$-string of beads

$$\Big(\mu((\rho_{I_1}, X_1])^{-\alpha}(\rho_{I_1}, X_1], \mu((\rho_{I_1}, X_1])^{-1}\mu\restriction_{(\rho_{I_1}, X_1]}\Big).$$

X_2 is now a random atom picked from $E^{(2)} = E^{(1)} \setminus (\rho_{I_1}, X_1]$. By Lemma 3.23,

$$\left(\frac{\mu((\rho_1, \rho_1'))}{\mu(E^{(2)})}, \ldots, \frac{\mu((\rho_{I_1-1}, \rho_{I_1-1}'))}{\mu(E^{(2)})}, \frac{\mu((X_1, \rho_{I_1}'))}{\mu(E^{(2)})}, \frac{\mu((\rho_{I_1+1}, \rho_{I_1+1}'))}{\mu(E^{(2)})}, \ldots, \frac{\mu((\rho_k, \rho_k'))}{\mu(E^{(2)})}\right)$$

has a Dirichlet$(\theta_1, \ldots, \theta_{I_1-1}, \theta_{I_1}, \theta_{I_1+1}, \ldots, \theta_k)$ distribution, and hence X_2 splits the set $E^{(2)}$ according to Dirichlet$(\alpha, 1-\alpha, \theta_1, \ldots, \theta_k)$. Therefore, inductively, for $n \geq 1$, we obtain the representation

$$V_n := \mu\left(\bigcup_{m=1}^{n} (a_m, b_m]\right) = 1 - \prod_{m=1}^{n} (1 - Y_m),$$

where $(Y_m)_{m\geq 1}$ is a sequence of i.i.d. random variables with $Y_m \sim$ Beta$(1, \theta)$, using notation from (3.13). Consider now the sets of interval components defined via $\{(a_n, b_n], n \geq 1\}$. Applying Lemma 3.23 inductively yields that the sequence $(Y_m)_{m\geq 1}$ is i.i.d. and independent of the $(\alpha, 0)$-strings of beads

$$\Big(\mu((a_n, b_n])^{-\alpha}(a_n, b_n], \mu((a_n, b_n])^{-1}\mu\restriction_{(a_n, b_n]}\Big), \quad n \geq 1.$$

We are now in the situation of Corollary 3.21, and conclude that $(E_{\uplus} = (\rho, \rho'), \mu)$ is indeed associated with an (α, θ)-regenerative interval partition of $[0, 1]$. By Proposition 3.18, $(E_{\uplus} = (\rho, \rho'), \mu)$ equipped with $d_{\uplus}$ defines an (α, θ)-string of beads. $\square$

3.4 Branch merging on (continuum) trees

We introduce a branch merging operation on Ford CRTs, i.e. on binary fragmentation continuum random trees of the form $\mathcal{T}^{\alpha,1-\alpha}$ arising in the scaling limit of the (α, θ)-tree growth process for $\theta = 1 - \alpha$. We construct a sequence of reduced trees $(\mathcal{R}_k^{\alpha,2-\alpha}, k \geq 1)$ associated with $\mathcal{T}^{\alpha,2-\alpha}$, the scaling limit of the $(\alpha, 2-\alpha)$-tree growth process, based on branch merging and the $(\alpha, 1-\alpha)$-coin tossing construction of embedded leaves according to the $(\alpha, 1-\alpha)$-tree growth model.

The $(\alpha, 1-\alpha)$-coin tossing construction reduces to uniform sampling from the mass measure in the case of the Brownian CRT, i.e. when $\alpha = \theta = 1/2$, which yields a simplified branch merging operation, that we use to construct a new tree-valued Markov process on a space of continuum trees.

3.4.1 Preliminaries

3.4.1.1 $\mathbb{R}$-trees, self-similar CRTs and the Brownian CRT

We recap some background on $\mathbb{R}$-trees. As we will only work with compact $\mathbb{R}$-trees, we include compactness in the definition. An $\mathbb{R}$-*tree* is a compact metric space $(\mathcal{T}, d)$ such that the following two properties hold for every $\sigma_1, \sigma_2 \in \mathcal{T}$.

(i) There is an isometric map $\Phi_{\sigma_1,\sigma_2}\colon [0, d(\sigma_1, \sigma_2)] \to \mathcal{T}$ such that

$$\Phi_{\sigma_1,\sigma_2}(0) = \sigma_1 \text{ and } \Phi_{\sigma_1,\sigma_2}(d(\sigma_1, \sigma_2)) = \sigma_2.$$

(ii) For every injective path $q\colon [0, 1] \to \mathcal{T}$ with $q(0) = \sigma_1$ and $q(1) = \sigma_2$,

$$q([0, 1]) = \Phi_{\sigma_1,\sigma_2}([0, d(\sigma_1, \sigma_2)]).$$

We write $[[\sigma_1, \sigma_2]] := \Phi_{\sigma_1,\sigma_2}([0, d(\sigma_1, \sigma_2)])$ for the range of Φ_{σ_1,σ_2}.

A *rooted* $\mathbb{R}$-tree $(\mathcal{T}, d, \rho)$ is an $\mathbb{R}$-tree $(\mathcal{T}, d)$ with a distinguished element ρ, the *root*. In what follows, we only work with rooted $\mathbb{R}$-trees. Sometimes we refer to $\mathcal{T}$ as an $\mathbb{R}$-tree, the distance d and the root ρ being implicit. For any $\alpha > 0$ and any metric space $(\mathcal{T}, d)$, in particular any $\mathbb{R}$-tree, we write $\alpha\mathcal{T}$ for $(\mathcal{T}, \alpha d)$.

We only consider equivalence classes of rooted $\mathbb{R}$-trees. Two rooted $\mathbb{R}$-trees $(\mathcal{T}, d, \rho)$ and $(\mathcal{T}', d', \rho')$ are *equivalent* if there exists an isometry from $\mathcal{T}$ onto $\mathcal{T}'$ such that ρ is mapped onto ρ'. We denote by $\mathbb{T}$ the set of all equivalence classes of rooted $\mathbb{R}$-trees. As shown in [32], the space $\mathbb{T}$ is a Polish space when endowed with the pointed Gromov-Hausdorff distance d_{GH}. The *pointed Gromov-Hausdorff distance* between two rooted

$\mathbb{R}$-trees $(\mathcal{T}, d, \rho)$, $(\mathcal{T}', d', \rho')$ is defined as

$$d_{\mathrm{GH}}((\mathcal{T}, d, \rho), (\mathcal{T}', d', \rho')) := \inf\{\max\{\delta(\phi(\rho), \phi'(\rho')), \delta_{\mathrm{H}}(\phi(\mathcal{T}), \phi'(\mathcal{T}'))\}\},$$

where the infimum is taken over all metric spaces $(\mathcal{M}, \delta)$ and all isometric embeddings $\phi\colon \mathcal{T} \to \mathcal{M}$, $\phi'\colon \mathcal{T}' \to \mathcal{M}$ into $(\mathcal{M}, \delta)$, and δ_{H} is the Hausdorff distance between compact subsets of $(\mathcal{M}, \delta)$. In fact, the Gromov-Hausdorff distance only depends on equivalence classes of rooted $\mathbb{R}$-trees and so induces a metric on $\mathbb{T}$. We equip $\mathbb{T}$ with its Borel σ-algebra. See [32] for a formal construction of $\mathbb{T}$.

A *weighted* $\mathbb{R}$-*tree* $(\mathcal{T}, d, \rho, \mu)$ is a rooted $\mathbb{R}$-tree $(\mathcal{T}, d, \rho)$ equipped with a probability measure μ on the Borel sets $\mathcal{B}(\mathcal{T})$. Two weighted $\mathbb{R}$-trees $(\mathcal{T}, d, \rho, \mu)$ and $(\mathcal{T}', d', \rho', \mu')$ are called *equivalent* if there exists an isometry from $(\mathcal{T}, d, \rho)$ onto $(\mathcal{T}', d', \rho')$ such that μ' is the push-forward of μ. The set of equivalence classes of weighted $\mathbb{R}$-trees is denoted by $\mathbb{T}_{\mathrm{w}}$. The Gromov-Hausdorff distance can be extended to a metric on $\mathbb{T}_{\mathrm{w}}$, the Gromov-Hausdorff-Prokhorov metric, see e.g. [54]. For any rooted $\mathbb{R}$-tree $(\mathcal{T}, d, \rho)$ and $x \in \mathcal{T}$, we call $d(\rho, x)$ the *height* of x, and $\sup_{x\in\mathcal{T}} d(\rho, x)$ the *height* of $\mathcal{T}$. A *leaf* is an element $x \in \mathcal{T}$ with $x \neq \rho$ whose removal does not disconnect $\mathcal{T}$. We denote the set of all leaves of $\mathcal{T}$ by $\mathrm{Lf}(\mathcal{T})$. An element $x \in \mathcal{T}$, $x \neq \rho$, is a *branch point* if its removal disconnects the $\mathbb{R}$-tree into three or more components. The *degree* of a vertex $x \in \mathcal{T}$ is the number of connected components of $\mathcal{T} \setminus \{x\}$; we call an $\mathbb{R}$-tree *binary* if the highest branch point degree is 3.

A pair $(\mathcal{T}, \mu)$ is a *continuum tree* if $\mathcal{T}$ is an $\mathbb{R}$- tree, and μ is a probability measure on $\mathcal{T}$ satisfying the following three properties.

(i) μ is supported by $\mathrm{Lf}(\mathcal{T})$, the set of leaves of $\mathcal{T}$;

(ii) μ has no atom, i.e. for any singleton $x \in \mathrm{Lf}(\mathcal{T})$ we have $\mu(\{x\}) = 0$;

(iii) for every $x \in \mathcal{T} \setminus \mathrm{Lf}(\mathcal{T})$, $\mu(\mathcal{T}_x) > 0$ where $\mathcal{T}_x := \{\sigma \in \mathcal{T} : x \in [[\rho, \sigma]]\}$.

By definition of a continuum tree, $\mathrm{Lf}(\mathcal{T})$ is uncountable and has no isolated points.

It will be useful to consider *reduced* trees: for any rooted $\mathbb{R}$-tree $\mathcal{T}$ and any $x_1, x_2, \ldots, x_n \in \mathrm{Lf}(\mathcal{T})$ let

$$\mathcal{R}(\mathcal{T}, x_1, \ldots, x_n) := \bigcup_{i=1}^{n} [[\rho, x_i]] \tag{3.23}$$

be the *reduced* subtree associated with $\mathcal{T}, x_1, \ldots, x_n$. $\mathcal{R}(\mathcal{T}, x_1, \ldots, x_n)$ is an $\mathbb{R}$-tree, whose root is ρ and whose set of leaves is $\{x_1, \ldots, x_n\}$.

For any weighted $\mathbb{R}$-tree $(\mathcal{T}, d, \rho, \mu)$ and any $\mathcal{R} \subset \mathcal{T}$, we define the projection

$$\pi^{\mathcal{R}} \colon \mathcal{T} \to \mathcal{R}, \qquad \sigma \mapsto \Phi_{\rho,\sigma}(\sup\{t \geq 0 : \Phi_{\rho,\sigma}(t) \in \mathcal{R}\}), \tag{3.24}$$

where the function $\Phi_{\rho,\sigma} \colon [0, d(\rho, \sigma)] \to \mathcal{T}$ is the isometry with $\Phi_{\rho,\sigma}(0) = \rho$ and $\Phi_{\rho,\sigma}(d(\rho,\sigma)) = \sigma$. We write

$$\mu_{\mathcal{R}}(C) := \mu\Big(\big(\pi^{\mathcal{R}}\big)^{-1}(C)\Big), \quad C \in \mathcal{B}(\mathcal{R}),$$

for the push-forward of μ via $\pi^{\mathcal{R}}$, where $\mathcal{B}(\mathcal{R})$ is the Borel σ-algebra on $\mathcal{R}$.

Weighted $\mathbb{R}$-trees can be obtained from *height functions*, that is, continuous functions $h \colon [0,1] \to [0,\infty)$ with $h(0) = h(1) = 0$. We define a distance by

$$d_h(x,y) := h(x) + h(y) - 2 \inf_{\min\{x,y\} \leq z \leq \max\{x,y\}} h(z), \quad x, y \in [0,1]. \tag{3.25}$$

Let $y \sim y'$ if $d_h(y,y') = 0$, and take the quotient $\mathcal{T}_h = [0,1)/\sim$. Then $(\mathcal{T}_h, d_h, \rho)$ is a rooted $\mathbb{R}$-tree coded by the function h, where the root ρ is the equivalence class of 0. When the height function $h \colon [0,1] \to [0,\infty)$ is random we obtain a *random* $\mathbb{R}$-tree $(\mathcal{T}_h, d_h, \rho)$. Note that (random) $\mathbb{R}$-trees can be equipped with a natural mass measure μ_h induced by the Lebesgue measure on $[0,1]$.

We study a certain type of random $\mathbb{R}$-trees, namely binary fragmentation continuum random trees, arising in the scaling limit of (α, θ)-tree growth processes.

Definition 3.25 (Binary fragmentation continuum random tree)**.** A random weighted rooted binary $\mathbb{R}$–tree $(\mathcal{T}, d, \rho, \mu)$ is called a *binary fragmentation continuum random tree* (CRT) of index $\gamma > 0$, if

(i) the measure μ has no atoms and $\mu(\mathcal{T}_x) > 0$ a.s. for all $x \in \mathcal{T} \setminus \mathrm{Lf}(\mathcal{T})$, where $\mathcal{T}_x = \{\sigma \in \mathcal{T} : x \in [[\rho, \sigma]]\}$, and $\mu([[\rho, x]]) = 0$ for all $x \in \mathcal{T}$;

(ii) for each $t \geq 0$, given the masses of the connected components $(\mathcal{T}_i^t, i \geq 1)$ of $\{\sigma \in \mathcal{T} : d(\rho, \sigma) > t\}$, indexed in decreasing order of mass and completed by a root vertex ρ_i, i.e. given $(\mu(\mathcal{T}_i^t), i \geq 1) = (m_i, i \geq 1)$ for some $m_1 \geq m_2 \geq \cdots \geq 0$, the rescaled trees

$$\left(\mathcal{T}_i^t, m_i^{-\gamma} d \restriction_{\mathcal{T}_i^t}, \rho_i, m_i^{-1} \mu \restriction_{\mathcal{T}_i^t}\right)$$

are independent identically distributed isometric copies of $(\mathcal{T}, d, \rho, \mu)$.

Binary fragmentation CRTs are characterised by a *self-similarity parameter* $\gamma > 0$ and a so-called *dislocation measure* $\nu(du)$, that is, a σ-finite measure on $[1/2, 1)$ satisfying $\int_{[1/2,1)} (1-u)\nu(du) < \infty$. They can be decomposed into isometric i.i.d. copies of the (in a probabilistic sense) "same" tree, when split into subtrees along the spine from the root to a leaf sampled from the mass measure.

Theorem 3.26 (Spinal decomposition, [60] Proposition 18)**.** *Let $(\mathcal{T}, d, \rho, \mu)$ be a binary fragmentation CRT with self-similarity parameter $\gamma > 0$. Select a leaf $\Sigma^* \sim \mu$ in $\mathcal{T}$, consider the spine $[[\rho, \Sigma^*]]$ and the connected components $(\mathcal{T}_i, i \in I)$ of $\mathcal{T} \setminus [[\rho, \Sigma^*]]$, each completed by a root vertex ρ_i. Furthermore, assign mass $m_i := \mu(\mathcal{T}_i)$ to each point $\rho_i \in [[\rho, \Sigma^*]]$ and denote the resulting distribution on $[[\rho, \Sigma^*]]$ by μ^*. Then, given the string of beads $([[\rho, \Sigma^*]], \mu^*)$, the rescaled trees*

$$\left(\mathcal{T}_i, m_i^{-\gamma} d \restriction_{\mathcal{T}_i}, \rho_i, m_i^{-1} \mu \restriction_{\mathcal{T}_i}\right), \qquad i \in I,$$

are independent identically distributed isometric copies of $(\mathcal{T}, d, \rho, \mu)$.

For any $\Sigma \in \mathrm{Lf}(\mathcal{T})$, we say that the *spinal decomposition theorem holds for the spine* $[[\rho, \Sigma]]$ to mean that Theorem 3.26 holds when we replace the leaf Σ^* by Σ.

Examples for binary fragmentation CRTs include Ford CRTs, i.e. the trees of the form $\mathcal{T}^{\alpha, 1-\alpha}$, $\alpha \in (0, 1)$, with self-similarity index $\gamma = \alpha$, see Section 3.4.1.2. When $\gamma = \alpha = 1/2$, the tree $\mathcal{T}^{1/2, 1/2}$ is the Brownian CRT, which can be defined in terms of Brownian excursion, see e.g. [15]. Specifically, let $W = (W(t), 0 \leq t \leq 1)$ be a standard

Brownian excursion. The tree $(\mathcal{T}_{2W}, d_{2W}, \rho)$ defined via $2W$ as height function, with mass measure μ_{2W} induced by the Lebesgue measure on $[0,1]$, is called the *Brownian continuum random tree* (Brownian CRT), see [7].

We often need a random $\mathbb{R}$-tree whose equivalence class has the same distribution as a Ford CRT on $\mathbb{T}$ (or $\mathbb{T}_{\mathrm{w}}$). We also refer to such $\mathbb{R}$-trees as Ford CRTs.

3.4.1.2 Regenerative tree growth: the (α,θ)-model

Ordered (α,θ)-CRPs and related (α,θ)-strings of beads naturally appear in the study of the (α,θ)-tree growth process $(T_n^{\alpha,\theta}, n \geq 1)$, as studied in [60].

Consider the set $\mathbb{T}_n^b$ of rooted binary trees with n leaves labelled by $[n] = \{1,\ldots,n\}$. We study the growth model given by the following growth procedure. Let T_1 be the tree consisting of one single edge, joining the root ρ and the leaf with label 1. At step 2 we select this edge and split it into a Y-shaped tree T_2, which has one edge that connects the root vertex and a binary branch point; there are two leaves labelled by 1 and 2 which are linked to the single branch point by one edge each. Inductively, to construct T_{n+1} conditionally given T_n, select an edge of T_n according to some weights (which we specify later). We select the edge $\Omega_n \to \Omega_n''$, say, directed away from the root. We replace this edge by three edges $\Omega_n \to \Omega_n'$, $\Omega_n' \to n+1$ and $\Omega_n' \to \Omega_n''$, meaning that there is one edge connecting Ω_n to a new branch point Ω_n' which is linked to the existing vertex Ω_n'' and a new leaf $n+1$. Clearly, $T_n \in \mathbb{T}_n^b$ for all $n \geq 2$.

The stochastic process $(T_n, n \geq 1)$ of random trees created in the described way, where each edge is selected randomly according to some selection rule, is called a *binary tree growth process*. A range of selection rules, which are partly related to each other, was studied in the literature. We focus on the (α,θ)-rule.

Definition 3.27 ((α,θ)-tree growth process)**.** For $0 \leq \alpha < 1$ and $\theta \geq 0$, the (α,θ)-selection rule is defined as follows.

(i) For any $n \geq 2$, consider the branch point of the tree T_n adjacent to the root, and the two subtrees $T_{n,0}$ and $T_{n,1}$ above this point, where $T_{n,1}$ contains the smallest label in T_n . Denote their sizes, i.e. their number of leaves, by m and $n-m$, respectively. Assign weight α to the edge connecting the root and the adjacent branch point, and weights $m-\alpha$, $n-m-1+\theta$ to the subtrees $T_{n,0}$, $T_{n,1}$, respectively.

(ii) Choose one of the two subtrees or the edge adjacent to the root proportionally to these weights (note that the total weight is $n-1+\theta$). If a subtree with two or more leaves was selected, recursively apply the weighting procedure and the random selection until an edge or a subtree with a single leaf is chosen. If a subtree with a single leaf was selected, select the unique edge of this subtree.

A binary tree growth process $(T_n^{\alpha,\theta}, n \geq 1)$ grown via the (α,θ)-rule for some $0 \leq \alpha < 1$ and $\theta \geq 0$ is called an (α,θ)-*tree growth process*, see Figure 3.2 for an example showing the growth procedure.

We write $(T_n^{\alpha,\theta,o}, n \geq 1)$ for the delabelled tree growth process. The trees $T_n^{\alpha,\theta}$ and $T_n^{\alpha,\theta,o}, n \geq 1$, are considered as $\mathbb{R}$–trees with unit edge lengths.

Remark 3.28 (Ford's rule). The selection rule for $\theta = 1-\alpha$, $0 \leq \alpha < 1$, is called *Ford's rule*, see [35].

Consider the spine $[[\rho, 1]]$, i.e. the unique path in $T_n^{\alpha,\theta}$ connecting the root and leaf 1. Referring to the subtrees along the spine as *tables* and the leaves within these subtrees as *customers* we obtain an ordered (α,θ)-CRP with label set $\{2, 3, \ldots\}$. Whenever an edge on the spine is selected, the height of leaf 1, that is, the graph distance between the root ρ and leaf 1, i.e. the number of tables K_n in the CRP, increases by 1. Note that the j-th customer in the restaurant carries label $j+1$ as a leaf in the tree, since leaf 1 is not contained in a subtree off the spine. Each subtree is uniquely characterised by its leaf labels which is a natural consequence of the growth procedure of $T_n^{\alpha,\theta}$, $n \geq 1$. Moreover, the order of tables on the spine (or *spinal* order) is consistent across different values of n.

(α,θ)-tree growth processes are *regenerative* in the following sense.

Definition 3.29 (Regenerative tree growth process)**.** A binary tree growth process $(T_n, n \geq 1)$ is called *regenerative* if for each $n \geq 2$, conditionally given that the first split of T_n is into two subtrees $T_{n,0}$, $T_{n,1}$ with label sets B_0, B_1, respectively, the relabelled trees $\widetilde{T}_{n,0}$ and $\widetilde{T}_{n,1}$ are independent copies of $T_{\#B_i}$, $i = 0, 1$, respectively where $\widetilde{T}_{n,i}$ is the tree $T_{n,i}$ with leaves relabelled by the increasing bijection $B_i \to [\#B_i]$, $i = 0, 1$.

The trees grown via the (α, θ)-selection rule have leaves which can be identified by successively assigned labels from $\{1, 2, \ldots\}$. Exchangeability of leaf labelling was studied in [60].

Lemma 3.30 ([60] Proposition 1)**.** *Let $(T_n^{\alpha,\theta}, n \geq 1)$ be an (α, θ)-tree growth process for some $\alpha \in (0,1)$ and $\theta \geq 0$. The leaf labels of $T_n^{\alpha,\theta}$ are exchangeable for all $n \geq 1$ if and only if $\alpha = \theta = 1/2$.*

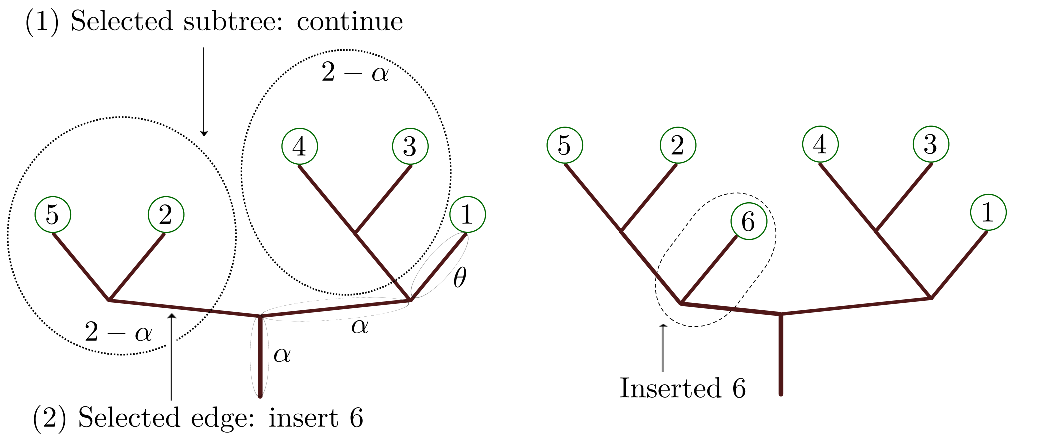

Fig. 3.2 Example of two-step recursion in the (α, θ)-model. First subtree is selected with probability $(2-\alpha)/(4+\theta)$; within this subtree leaf 6 is inserted at the root edge with probability $\alpha/(1+\theta)$.

Limits of the delabelled tree growth process $(T_n^{\alpha,\theta,o}, n \geq 1)$ were established in [60] by considering the reduced trees $\mathcal{R}(T_n^{\alpha,\theta}, [k])$, $k \geq 1$, where each edge between two vertices v_1 and v_2 is equipped with the graph distance in $T_n^{\alpha,\theta}$, i.e. with the number of edges between v_1 and v_2 in $T_n^{\alpha,\theta}$.

Proposition 3.31 ([60] Proposition 2)**.** *Let $(T_n^{\alpha,\theta}, n \geq 1)$ be an (α, θ)-tree growth process for some $\alpha \in (0,1)$ and $\theta \geq 0$. Then, for all $k \geq 1$,*

$$\lim_{n\to\infty} n^{-\alpha}\mathcal{R}(T_n^{\alpha,\theta}, [k]) = \mathcal{R}_k^{\alpha,\theta} \qquad a.s., \tag{3.26}$$

i.e. the edge lengths of $\mathcal{R}(T_n^{\alpha,\theta}, [k])$ scaled by $n^{-\alpha}$ converge jointly a.s. as $n \to \infty$ to the edge lengths of a tree $\mathcal{R}_k^{\alpha,\theta}$ with the same shape as $\mathcal{R}(T_n^{\alpha,\theta}, [k])$.

Recall the notation $\mathcal{R}_k^{\alpha,\theta,o}$ for the delabelled tree associated with $\mathcal{R}_k^{\alpha,\theta}$.

Theorem 3.32 ([60] Theorem 3)**.** *Consider the situation of Proposition 3.31, and the delabelled trees* $\mathcal{R}_k^{\alpha,\theta,o}$ *associated with* $\mathcal{R}_k^{\alpha,\theta}$ *for* $k \geq 1$*. Then there exists a CRT* $\mathcal{T}^{\alpha,\theta}$ *on the same probability space such that*

$$\lim_{k\to\infty} \mathcal{R}_k^{\alpha,\theta,o} = \mathcal{T}^{\alpha,\theta} \qquad \textit{a.s. in the Gromov-Hausdorff topology.} \tag{3.27}$$

In general, the trees $\mathcal{T}^{\alpha,\theta}$ are binary fragmentation CRTs, characterised by the self-similarity parameter α, and the absolutely continuous dislocation measure $\nu_{\alpha,\theta}(du) = f^o_{\alpha,\theta}(u)du$ for $u \in (1/2, 1)$, where the *dislocation density* is given via

$$\begin{aligned}\Gamma(1-\alpha)f^o_{\alpha,\theta}(u) = \alpha\Big(u^\theta(1-u)^{-\alpha-1} + u^{-\alpha-1}(1-u)^\theta\Big)& \\ + \theta\Big(u^{\theta-1}(1-u)^{-\alpha} + u^{-\alpha}(1-u)^{\theta-1}\Big),& \quad 1/2 < u < 1.\end{aligned} \tag{3.28}$$

Corollary 3.33. *For any* $\alpha \in (0,1)$*,* $\mathcal{T}^{\alpha,1-\alpha} \stackrel{d}{=} \mathcal{T}^{\alpha,2-\alpha}$*.*

The family of trees $\{\mathcal{T}^{\alpha,1-\alpha}, \alpha \in (0,1)\}$ is called *Ford CRTs*.

Proof. This was shown in [60].

- The marginal distributions of delabelled trees coincide for the $(\alpha, 1-\alpha)$- and the $(\alpha, 2-\alpha)$-tree growth process (see Lemma 12 in [60]).
- The $(\alpha, 1-\alpha)$- and the $(\alpha, 2-\alpha)$-tree growth process have the trees $\mathcal{T}^{\alpha,1-\alpha}$ and $\mathcal{T}^{\alpha,2-\alpha}$ as their distributional scaling limits (Proposition 3.31). □

Corollary 3.34. *For* $\alpha = 1/2$ *and* $\theta \in \{1/2, 3/2\}$*,* $\mathcal{T}^{\alpha,\theta}$ *is the Brownian CRT.*

Proof. It was shown in [42] that the delabelled $(1/2, 1/2)$-tree growth process has the Brownian CRT as its distributional scaling limit. By Corollary 3.33, the Brownian CRT is also the limit of the delabelled $(1/2, 3/2)$-tree growth process. □

Pitman and Winkel [60] used an embedding of the reduced trees $\mathcal{R}_k^{\alpha,\theta}$, $k \geq 1$, into the CRT $\mathcal{T}^{\alpha,\theta}$ to prove convergence results for $(T_n^{\alpha,\theta}, n \geq 1)$.

Theorem 3.35. *Let* $(\mathcal{T}^{\alpha,\theta}, d, \rho, \mu)$ *be a CRT as in Theorem 3.32. Then there exists a sequence of leaves* $\Sigma_k, k \geq 1$*, on a suitably enlarged probability space such that*

$$(\mathcal{R}(\mathcal{T}^{\alpha,\theta}, \Sigma_1, \ldots, \Sigma_k), k \geq 1) \stackrel{d}{=} (\mathcal{R}_k^{\alpha,\theta}, k \geq 1). \tag{3.29}$$

The problem of finding the leaves $\Sigma_k, k \geq 1$, as in Theorem 3.35 inside a CRT was solved in [60] (Lemma 19, Proposition 20 and Corollary 21). We refer to this problem as the *leaf embedding problem.* In particular, given a CRT $\mathcal{T} = \mathcal{T}^{\alpha,\theta}$ for some $\alpha \in (0,1), \theta > 0$, with mass measure μ, it involves the following steps.

(i) Find a leaf $\Sigma_1 \in \mathcal{T}$ such that $([[\rho, \Sigma_1]], \mu_{[[\rho,\Sigma_1]]})$ is an (α,θ)-string of beads, i.e. embed $(\mathcal{R}_1, \mu_1)$ into $(\mathcal{T}, \mu)$ via $(\mathcal{R}_1, \mu_1) = ([[\rho, \Sigma_1]], \mu_{[[\rho,\Sigma_1]]})$.

(ii) Given $(\mathcal{R}_1, \mu_1)$, find the point on $\mathcal{R}_1$ which identifies the root J_1 of the component of $\mathcal{T} \setminus \mathcal{R}_1$ which contains Σ_2, so that $\mathcal{R}_2 = \mathcal{R}_1 \cup]]J_1, \Sigma_2]]$.

(iii) Iterate the procedure to embed $(\mathcal{R}_k, \mu_k)$ for general $k \geq 2$ into $\mathcal{T}$.

When $(\alpha, \theta) = (1/2, 1/2)$ the leaf embedding problem can be solved by sampling leaves $\Sigma_k, k \geq 1$, independently from the mass measure μ. The general technique to solve the leaf embedding problem is more involved, but not needed to understand the Branch Merging Markov Chain, Section 3.4.2, and the new and simple leaf embedding approach for $(1/2, 3/2)$-tree growth processes in Section 3.4.3.2.

We recap the procedure to solve (i) for $\alpha \in (0,1)$, $\theta > 0$, from [60], Section 4.3.

- Let $(\mathcal{T}^{(1)}, d^{(1)}, \rho^{(1)}, \mu^{(1)}) = (\mathcal{T}, d, \rho, \mu)$ and let $\Sigma_1^{*(1)}$ be a random leaf sampled from $\mu^{(1)}$ in $\mathcal{T}^{(1)}$. Consider the string of beads $([[\rho, \Sigma_1^{*(1)}]], \mu_{[[\rho,\Sigma_1^{*(1)}]]})$ obtained by considering the spine from the root to this leaf, equipped with projected subtree masses. $([[\rho, \Sigma_1^{*(1)}]], \mu_{[[\rho,\Sigma_1^{*(1)}]]})$ can be represented via a local time process $(L^{*(1)}(u), 0 \leq u \leq 1)$, i.e. $d(\rho, \Sigma_1^{*(1)}) = L^{*(1)}(1)$ and
$$\mu_{[[\rho,\Sigma_1^{*(1)}]]}\left(\left\{\Phi_{\rho,\Sigma_1^{*(1)}}\left(L^{*(1)}\left(1 - e^{-\xi_t^{*(1)}}\right)\right)\right\}\right) = e^{-\xi_{t-}^{*(1)}} - e^{-\xi_t^{*(1)}} = -\Delta e^{-\xi_t^{*(1)}},$$
for a subordinator $(\xi_t^{*(1)}, t \geq 0)$, the so-called *spinal subordinator*, whose distribution was characterised in [60], Section 4.1. Note that the definition of $\mu_{[[\rho,\Sigma_1^{*(1)}]]}$ above means that the $\mu_{[[\rho,\Sigma_1^{*(1)}]]}$-measure of the point $L^{*(1)}\left(1 - e^{-\xi_t^{*(1)}}\right)$ along the path from the root to $\Sigma_1^{*(1}$ is $-\Delta e^{-\xi_t^{*(1)}}$.

Consider the switching time $\tau^{(1)}$ associated with the switching probability function $(\hat{p}(u), 0 \leq u \leq 1)$ defined by

$$\hat{p}(u) := \frac{(1-u) f_{\alpha,\theta}(1-u)}{f^*_{\alpha,\theta}(u)}, \quad 0 < u < 1,$$

as in Proposition 3.13 where, for $0 < u < 1$, $f_{\alpha,\theta}(u)$ and $f^*_{\alpha,\theta}(u)$ are given by

$$\begin{aligned} f_{\alpha,\theta} &= \Gamma(1-\alpha)^{-1}\Big(\alpha(1-u)^{-\alpha-1}u^{\theta-1} + \theta u^{\theta-2}(1-u)^{-\alpha}\Big), \\ f^*_{\alpha,\theta}(u) &= u f_{\alpha,\theta}(u) + (1-u) f_{\alpha,\theta}(1-u). \end{aligned}$$

Note that $f^o_{\alpha,\theta}(u) = u f_{\alpha,\theta}(u) + (1-u) f_{\alpha,\theta}(1-u), 1/2 < u < 1$, where $f^o_{\alpha,\theta}$ was given in (3.28). We refer to [61] (19) for a derivation of $f^*_{\alpha,\theta}$. Define

$$F_t := \exp\Big(-\Delta\xi^{*(1)}_t\Big), \quad 0 \leq t < \tau^{(1)}, \qquad F_{\tau^{(1)}} = 1 - \exp\Big(-\Delta\xi^{*(1)}_{\tau^{(1)}}\Big).$$

- For $i \geq 1$, consider the interval partition $\{1 - \exp(-\xi^{*(i)}_t), t \geq 0\}^{\mathrm{cl}}$, with associated local time process $(L^{*(i)}(u), 0 \leq u \leq 1)$, and the junction point

$$\rho^{(i+1)} = \Phi_{\rho^{(i)}, \Sigma^{*(i)}_1}\Big(1 - L^{*(i)}\Big(\exp\Big(-\xi^{*(i)}_{\tau^{(i)}}\Big)\Big)\Big),$$

that is, the point on the spine $[[\rho, \Sigma^{*(i)}_1]]$ related to the switching time $\tau^{(i)}$.

Furthermore, define

$$\begin{aligned} \mathcal{T}^{(i+1)} &= \Big\{\sigma \in \mathcal{T}^{(i)} : [[\rho^{(i)}, \sigma]] \cap [[\rho^{(i)}, \Sigma^{*(i)}_1]] = [[\rho^{(i)}, \rho^{(i+1)}]]\Big\}, \\ d^{(i+1)} &= \Big(1 - \exp\Big(-\xi^{*(i)}_{\tau^{(i)}-\tau^{(i-1)}}\Big)\Big)^{-\alpha} d^{(i)} \restriction_{\mathcal{T}^{(i+1)}}, \\ \mu^{(i+1)} &= \Big(1 - \exp\Big(-\xi^{*(i)}_{\tau^{(i)}-\tau^{(i-1)}}\Big)\Big)^{-1} \mu^{(i)} \restriction_{\mathcal{T}^{(i+1)}} . \end{aligned}$$

Note that $(\mathcal{T}^{(i+1)}, d^{(i+1)}, \mu^{(i+1)})$ is the subtree rooted at $\rho^{(i+1)}$ on $[[\rho^{(i)}, \Sigma^{*(i)}_1]]$ with distance and mass measure rescaled by the mass of the selected atom $\rho^{(i+1)}$ to the power $-\alpha$ and -1, respectively.

Sample a leaf $\Sigma_1^{*(i+1)}$ from $\mu^{(i+1)}$ in $\mathcal{T}^{(i+1)}$, and proceed as before to determine the spinal subordinator $\xi^{*(i+1)}$ and the switching time $\tau^{(i+1)}$. Set

$$\begin{aligned} F_{\tau^{(i)}+t} &= \exp\Big(-\Delta\xi_t^{*(i+1)}\Big), \quad 0 \le t < \tau^{(i+1)} - \tau^{(i)}, \\ F_{\tau^{(i+1)}} &= 1 - \exp\Big(-\Delta\xi^{*(i+1)}_{\tau^{(i+1)}-\tau^{(i)}}\Big). \end{aligned}$$

Lemma 3.36 ([60] Proposition 20)**.** *The process $(F_t, t \ge 0)$ is a Poisson point process in $(0,1)$ with intensity measure $uf_{\alpha,\theta}(u)$ and cemetery state 1. Furthermore, consider the space*

$$[[\rho, \Sigma_1[[:= \bigcup_{i\ge 1} [[\rho, \rho^{(i)}]].$$

Then $\Sigma_1 \in \mathcal{T}$ is a leaf a.s., the string of beads $([[\rho,\Sigma_1]], \mu_{[[\rho,\Sigma_1]]})$ is a weight-preserving isometric copy of $(\mathcal{R}_1, \mu_1)$, and the spinal decomposition theorem holds for the spine $[[\rho, \Sigma_1]]$ with connected components $(\mathcal{T}_i, i \in I)$ of $\mathcal{T} \setminus [[\rho, \Sigma_1]]$.

We refer to this procedure to find the leaf Σ_1 as the (α,θ)-*coin tossing construction* (or simply the *coin tossing construction*) in the tree $(\mathcal{T}, \mu)$.

In order to find the branch point where the spine to leaf Σ_1 leaves the spine to leaf Σ_2, we apply (α,θ)-coin tossing sampling (cf. Proposition 3.13) in the (α,θ)-string of beads $(\mathcal{R}_1, \mu_1)$ which we have already embedded into $(\mathcal{T}, \mu)$. By the spinal decomposition theorem, we conclude that the subtrees associated with that atom as branch point base point is a rescaled independent copy of $(\mathcal{T}, \mu)$, and hence we can apply the (α,θ)-coin tossing construction in this subtree to identifiy Σ_2. Therefore, given the embedding of $(\mathcal{R}_1, \mu_1)$ into $(\mathcal{T}, \mu)$, we can embedd $(\mathcal{R}_2, \mu_2)$ by identifying leaf Σ_2, and considering μ_2 as the projection of the measure μ onto $\mathcal{R}_2$. It was shown in [60], Corollary 21, that this method can be used to embed $(\mathcal{R}_k, \mu_k)$ into $(\mathcal{T}, \mu)$ for all $k \ge 1$.

Lemma 3.37 (Embedding of $(\mathcal{R}_k, \mu_k)$, [60] Corollary 21)**.** *For $k \ge 1$, given an embedding of $(\mathcal{R}_k, \mu_k)$ into $(\mathcal{T}, \mu)$, perform the following steps to embed $(\mathcal{R}_{k+1}, \mu_{k+1})$.*

- *Select an edge E of $\mathcal{R}_k$ proportionally to the mass assigned to E by μ_k.*

- *If E is an internal edge of $\mathcal{R}_k$, pick J_k from the mass measure μ_k restricted to E; otherwise perform (α,θ)-coin tossing sampling (see Proposition 3.13) in the string of beads $(\mu_k(E)^{-\alpha}E, \mu_k(E)^{-1}\mu_k\restriction_E)$ to identify J_k.*
- *Define the pair $(\mathcal{R}_{k+1},\mu_{k+1})$ by $\mathcal{R}_{k+1} := \mathcal{R}_k \cup [[J_k,\Sigma_{k+1}]], \mu_{k+1} := \mu_{\mathcal{R}_{k+1}}$.*

3.4.2 The Branch Merging Markov Chain

We present a branch merging operation on continuum trees. Given a continuum tree, we pick two leaves independently from the mass measure on the tree, merge the three branches of the arising reduced Y-shaped tree, and replant the subtrees in a certain random order on the created longer branch. This is in principle a simplified version of the more elaborate and general branch merging operation that we present in Section 3.4.3, and, in contrast to Section 3.4.3, branch merging here is separate from leaf embedding. Recall that $\mathbb{T}_{\mathrm{w}}$ denotes the space of equivalence classes of weighted $\mathbb{R}$-trees.

Definition 3.38 (Branch Merging Markov Chain)**.** The *Branch Merging Markov Chain* (BMMC) is the time-homogeneous Markov Chain $(\mathcal{B}_k, k \geq 0)$ on $\mathbb{T}$ which is defined via the following transition rule. Given a continuum tree $\mathcal{B}_0 := (\mathcal{T}, d, \rho, \mu)$, perform the following steps to obtain a tree $\mathcal{B}_1 := (\widetilde{\mathcal{T}}, \tilde{d}, \rho, \tilde{\mu})$.

(i)$^{\mathrm{MC}}$ *Start configuration.* Pick two leaves Σ_1 and $\widetilde{\Sigma}_1$ independently by uniform sampling from the mass measure μ on $\mathcal{T}$, and denote by Ω the branch point of the spines from the root ρ to Σ_1 and $\widetilde{\Sigma}_1$. We write $\mathcal{R}_2 = \mathcal{R}(\mathcal{T}, \Sigma_1, \widetilde{\Sigma}_1)$, and obtain the three branches $E_1 := [[\rho, \Omega[[, E_2 :=]]\Omega, \Sigma_1[[, E_3 := [[\Omega, \widetilde{\Sigma}_1[[$.

(ii)$^{\mathrm{MC}}$ *Partition sampling.* Consider a sequence of i.i.d. random variables $(X_k)_{k\geq 2} \sim \mu_{\mathcal{R}_2}$, and define $\Pi_x := \{k \geq 2 : X_k = x\}, x \in \mathcal{R}_2$.

(iii)$^{\mathrm{MC}}$ *Tree pruning and spine merging.* Define $\overline{\Pi}_i := \{\Pi_x, x \in E_i\}$, $i \in [3]$, and the collection of blocks

$$\mathcal{C} := \bigcup_{i\in[3]} \{\Pi_x \in \overline{\Pi}_i : \Pi_x \prec \Pi_y \forall y \in \mathcal{R}_2 \text{ with } \Pi_y \in \overline{\Pi}_i, d_{\mathcal{R}_2}(\rho, y) > d_{\mathcal{R}_2}(\rho, x)\},$$

where, for $A, B \subset \mathbb{N}$, $A \cap B \neq \emptyset$, $A \prec B :\Leftrightarrow \min A < \min B$.

Let $\Upsilon := \{x : \Pi_x \in \mathcal{C}\}$ be the associated set of cut points on the branches E_1, E_2, E_3, and write $\Upsilon = \{X_k, k \geq 1\}$ where, for $k \geq 1$,

$$X_k := x \in \mathcal{R}_2 \quad \text{such that} \quad |\{y \in \mathcal{R}_2 | \Pi_y \in \mathcal{C} \text{ and } \Pi_x \prec \Pi_y\}| = k - 1.$$

Note that the set of cut points Υ contains all points $x \in \mathcal{R}_2$ such that the smallest label in the label set Π_x is smaller than the smallest label in all the other label sets Π_y for points $y \in \mathcal{R}_2$ which are on the same branch as x but further away from the root than x, see Figure 3.3.

Define the merged branch $\widetilde{\mathcal{R}}_1$ via

$$\widetilde{\mathcal{R}}_1 := \{\rho\} \cup \left(\biguplus_{i \in [3]} E_i \right) \cup \{\Sigma_1\},$$

where the operator $\biguplus$ was defined in (3.4), Section 3.3.1. We obtain the metric $d_{\widetilde{\mathcal{R}}_1} : \widetilde{\mathcal{R}}_1 \times \widetilde{\mathcal{R}}_1 \to \mathbb{R}_0^+$, $d_{\widetilde{\mathcal{R}}_1}(x, y) := |d_{\widetilde{\mathcal{R}}_1}(\rho, x) - d_{\widetilde{\mathcal{R}}_1}(\rho, y)|$ where

$$d_{\widetilde{\mathcal{R}}_1}(\rho, x) := d_{\biguplus}(\rho, x) \text{ for } x \in \widetilde{\mathcal{R}}_1 \setminus \{\Sigma_1\}, \quad d_{\widetilde{\mathcal{R}}_1}(\rho, \Sigma_1) = d(\rho, \Sigma_1) + d(\Omega, \widetilde{\Sigma}_1).$$

(iv)$^{\mathrm{MC}}$ *Update rule.* Let $\mathcal{S}_x$ be the subtree of $\mathcal{T}$ rooted at $x \in \mathcal{R}_2$, respectively, and define the metric space $(\widetilde{\mathcal{T}}_1, \tilde{d}_1)$ by $\widetilde{\mathcal{T}}_1 := \widetilde{\mathcal{R}}_1 \cup \bigcup_{x \in \widetilde{\mathcal{R}}_1} \mathcal{S}_x$, equipped with the metric $\tilde{d}_1 : \widetilde{\mathcal{T}}_1 \times \widetilde{\mathcal{T}}_1 \to \mathbb{R}_0^+$,

$$\tilde{d}_1(x, y) := \begin{cases} d_{\widetilde{\mathcal{R}}_1}(x, y) & \text{if } x, y \in \widetilde{\mathcal{R}}_1, \\ d_{\widetilde{\mathcal{R}}_1}(x', y') + d(x, x') + d(y, y') & \text{if } x \in \mathcal{S}_{x'}, y \in \mathcal{S}_{y'}, x' \neq y' \in \widetilde{\mathcal{R}}_1, \\ d_{\widetilde{\mathcal{R}}_1}(x, y') + d(y, y') & \text{if } x \in \widetilde{\mathcal{R}}_1, y \in \mathcal{S}_{y'}, y' \in \widetilde{\mathcal{R}}_1, \\ d(x, y) & \text{if } x, y \in \mathcal{S}_{x'}, x' \in \widetilde{\mathcal{R}}_1. \end{cases} \tag{3.30}$$

Furthermore, define the mass measure $\tilde{\mu}_1$ on $(\widetilde{\mathcal{T}}_1, \tilde{d}_1)$ by

$$\tilde{\mu}_1(\overline{\mathcal{T}}_x) := \begin{cases} \sum_{y \in \overline{\mathcal{T}}_x \cap \widetilde{\mathcal{R}}_1} \mu(\mathcal{S}_y) & \text{if } x \in \widetilde{\mathcal{R}}_1, \\ \mu(\overline{\mathcal{T}}_x \cap \mathcal{S}_{x'}) & \text{if } x \in \mathcal{S}_{x'}, x' \in \widetilde{\mathcal{R}}_1, \end{cases} \tag{3.31}$$

where $\overline{\mathcal{T}}_x := \{y \in \widetilde{\mathcal{T}}_1 : x \in [[\rho, y]]\}$, which uniquely identifies $\widetilde{\mu}_1$.

Finally, define the output tree $\mathcal{B}_1 := (\widetilde{\mathcal{T}}_1, \tilde{d}_1, \rho, \tilde{\mu}_1)$.

Remark 3.39. In fact, the merging operation $\uplus$ was only defined for intervals. However, each $E_i = [[\rho_i, \rho_i']]$, $i \in [3]$, is isometric to the interval $[0, d(\rho_i, \rho_i')]$, i.e. we write $\biguplus_{i\in[k]} E_i$ to mean that we apply the merging operation to the strings of beads that we obtain when we consider the intervals $[0, d(\rho_i, \rho_i')]$ with mass measure pushed-forward via the choice of class representatives.

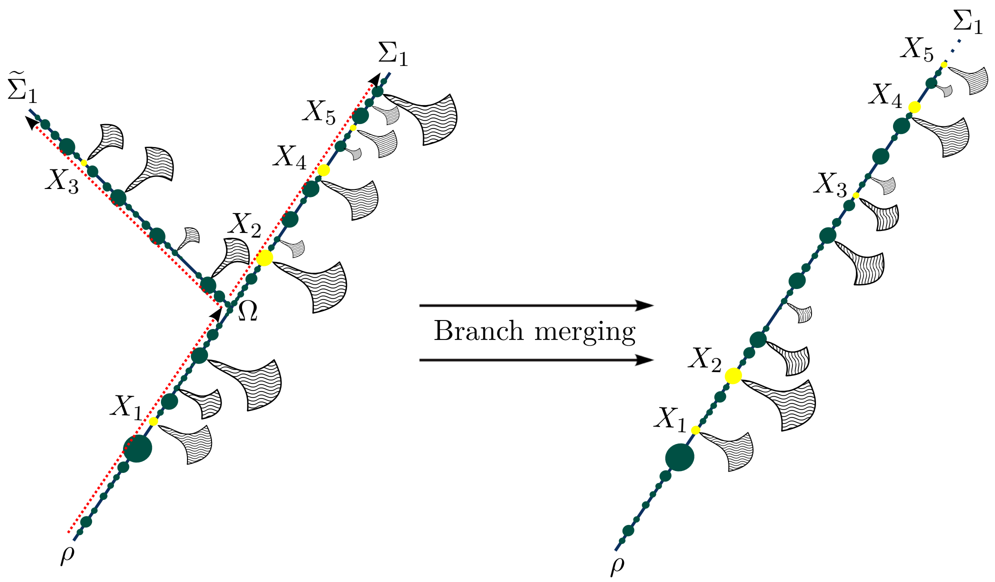

Fig. 3.3 Transition rule of the BMMC. Reduced tree structure with subtrees attached in the atoms. Strings of beads are merged and subtrees are carried forward.

Lemma 3.40. *The distribution of the Brownian CRT on the space* $\mathbb{T}_{\mathrm{w}}$ *is a stationary distribution of the Branch Merging Markov Chain* $(\mathcal{B}_k, k \geq 0)$.

Proof. Recall that, by Lemma 3.30, leaf labelling in the $(1/2, 1/2)$-model is exchangeable, i.e. we can embed $\mathcal{R}_2^{1/2,1/2}$ by sampling two leaves independently from the mass measure. The distribution of the mass split on the three branches of the Y-shaped tree $\mathcal{R}_2^{1/2,1/2}$ is Dirichlet$(1/2, 1/2, 1/2)$, and the three branches are rescaled independent $(1/2, 1/2)$-strings of beads when equipped with the restricted mass measure (Proposition 3.13).

We merge the three $(1/2, 1/2)$-strings of beads to obtain a $(1/2, 3/2)$-string of beads (see Theorem 3.15). By the spinal decomposition theorem, the subtrees rooted on $\mathcal{R}_2^{1/2,1/2}$ which we carry forward, are rescaled independent Brownian CRTs, i.e. by Theorem 3.35 and Corollary 3.34, the merged tree is a Brownian CRT. $\square$

We now study a discrete analogue of the BMMC, and use the framework of $\mathbb{R}$-trees with unit edge lengths. For $n \in \mathbb{N}$, let $\mathbb{T}_n^o$ be the set of rooted unlabelled trees with n leaves and no degree-2 vertices (except the root), and let $\mathbb{T}_n^{b,o}$ be the space of binary rooted unlabelled trees with n leaves and no degree-2 vertices, i.e. $\mathbb{T}_n^{b,o} \subset \mathbb{T}_n^o$. For $T \in \mathbb{T}_n^o$, we consider the uniform probability measure on $\mathrm{Lf}(\mathcal{T})$ by assigning mass $1/n$ to any of the leaves. Let $(\tau_k, k \geq 0)$ be a time-homogeneous Markov chain operating on $\mathbb{T}_n^o$ with the following transition rule. Conditionally given $\tau_0 = T$ with root ρ, perform the following steps to obtain $\tau_1 = \widetilde{T}$.

(i)$^{\mathrm{d}}$ Select one of the edges of T uniformly at random, and insert one leaf at this edge to create the tree $T_{+1} \in \mathbb{T}_{n+1}^o$ according to the growth rule for uniform binary tree growth processes (i.e. as in the $(1/2, 1/2)$-tree growth process), see Section 3.4.1.2 and Figure 3.2.

(ii)$^{\mathrm{d}}$ Sample two distinct leaves Σ_1 and $\widetilde{\Sigma}_1$ from the empirical mass measure on the set of leaves $\mathrm{Lf}(T_{+1})$ of T_{+1}, as in (i)$^{\mathrm{MC}}$, and let Ω be the unique branch point of the reduced tree $\mathcal{R}(T_{+1}, \Sigma_1, \widetilde{\Sigma}_1)$.

(iii)$^{\mathrm{d}}$ Sample leaf labels $2, \ldots, n$ according to a uniform allocation onto the $n-1$ unlabelled leaves (in the sense that leaves are sampled uniformly without replacement). Define the set of cut points as in (iii)$^{\mathrm{MC}}$, and refer to the labelled tree as T_{+1}^l.

(iv)$^{\mathrm{d}}$ Consider $\mathcal{R}_2 = \mathcal{R}(T_{+1}^l, \Sigma_1, \widetilde{\Sigma}_1)$ and the pairs $(E_i, \mu_{\mathcal{R}_2} \restriction_{E_i})$, $i \in [3]$, where

$$E_1 = [[\rho, \Omega[[, \quad E_2 =]]\Omega, \Sigma_1[[, \quad E_3 = [[\Omega, \widetilde{\Sigma}_1[[.$$

Define the merged spine $\widetilde{\mathcal{R}}_1$ and the tree $\widetilde{\mathcal{T}}_1$ as in (iii)$^{\mathrm{MC}}$ and (iv)$^{\mathrm{MC}}$ (and forget the labels) to obtain the output tree $\widetilde{T} = \widetilde{\mathcal{T}}_1$.

Note that $\widetilde{T} \in \mathbb{T}_n^o$ as we lose the leaf $\widetilde{\Sigma}_1$ under branch merging.

Further note that both in the BMMC and in the discrete version, we accept $\mu_{\mathcal{R}_2}(\Omega) > 0$. In this case, when the first part of E_2 is not the first part of the merged string of beads, in step (iv)$^{\mathrm{d}}$, we join the two atoms Ω and x for some $x \in \mathcal{R}_2$, and form a larger atom with mass $\mu_{\mathcal{R}_2}(\Omega) + \mu_{\mathcal{R}_2}(x)$, see the convention from Remark 3.12. Furthermore, when subtrees are carried forward, we join the trees $\mathcal{S}_\Omega = (\pi^{\mathcal{R}_2})^{-1}(\Omega)$ and $\mathcal{S}_x = (\pi^{\mathcal{R}_2})^{-1}(x)$ at their root vertices. We treat the case $\mu_{\mathcal{R}_2}(\rho) > 0$ analogously.

Theorem 3.41. *Let $n \geq 2$, and consider the Markov chain $(\tau_k, k \geq 0)$ operating on $\mathbb{T}_n^o$, with transition rule given by* $(\mathrm{i})^d - (\mathrm{iv})^d$. *The space $\mathbb{T}_n^{b,o}$ is the only closed communicating class of $(\tau_k, k \geq 0)$. Furthermore, $(\tau_k, k \geq 0)$ has a unique stationary distribution π which is supported by $\mathbb{T}_n^{b,o}$, and the distribution of τ_k converges to π as $k \to \infty$.*

Proof. Consider two trees $T, \widetilde{T} \in \mathbb{T}_n^o$. We have to show the following.

- If $T \in \mathbb{T}_n^{b,o}$, then
$$\mathbb{P}\Big(\tau_k \in \mathbb{T}_n^{b,o} \text{ for all } k \geq 1 \Big| \tau_0 = T\Big) = 1. \tag{3.32}$$

- If $T, \widetilde{T} \in \mathbb{T}_n^{b,o}$, then
$$\mathbb{P}\Big(\tau_k = \widetilde{T} \text{ for some } k \geq 1 \Big| \tau_0 = T\Big) > 0. \tag{3.33}$$

- If $T \in \mathbb{T}_n^o \setminus \mathbb{T}_n^{b,o}$, then
$$\mathbb{P}\Big(\tau_k \in \mathbb{T}_n^{b,o} \text{ for some } k \geq 2 \Big| \tau_0 = T\Big) > 0. \tag{3.34}$$

We use an induction on the number of leaves n. For $n = 2$, we have two possible trees, a Y-shaped tree, say T_1, and a tree with two leaves which are directly connected to the root, say T_2. Obviously, given $\tau_0 = T_1$, $\tau_1 = T_1$ with probability 1. On the other hand, to obtain T_1 from T_2, after the leaf insertion, assign label $\widetilde{\Sigma}_1$ to the leaf directly connected to the root or, indeed, to the inserted leaf. This leaf gets lost in the branch merging operation, and we obtain T_1.

Now, assume that (3.32)-(3.34) hold for any $1 \leq m \leq n-1$ where $n \geq 2$.

- *Proof of* (3.32). (3.32) is clear due to the nature of the leaf insertion procedure: only edges can be selected (but no branch points), and, in the binary case, there are no atoms in branch points. Hence, degrees of branch points cannnot increase, and (3.32) follows.

- *Proof of* (3.33). Conditionally given $T \in \mathbb{T}_n^{b,o}$, we consider the two subtrees T_1 and T_2 into which T splits at the first branch point. Let n_1 be the number of leaves in T_1 (i.e. T_2 has $n - n_1$ leaves). By the induction hypothesis, for any $T_1' \in \mathbb{T}_{n_1}^{b,o}$ and any $T_2' \in \mathbb{T}_{n-n_1}^{b,o}$ it holds that

$$\mathbb{P}(\tau_k = T_i' \text{ for some } k \geq 1 | \tau_0 = T_i) > 0, \quad i = 1, 2.$$

  Assigning label 2 to a leaf in T_2, and restricting the remaining leaf insertion and sampling procedure for the leaves $\Sigma_1, \widetilde{\Sigma}_1$ and the labels $3, \ldots, n_1 + 1$ to the tree T_1 (and swapping the roles of T_1 and T_2 afterwards), we obtain positive probability to attain any tree $T' \in \mathbb{T}_n^{b,o}$ which splits at the first branch point into two subtrees T_1' and T_2' of sizes n_1 and $n - n_1$, respectively.

  It remains to show that, for any $1 \leq n_1 \leq n - 1$, we have positive probability of attaining a tree $T' \in \mathbb{T}_n^{b,o}$ which splits into two subtress of sizes $n_1 + 1$ and $n - n_1 - 1$ at the first branch point. This is easy to see, e.g. consider the case when the leaf insertion takes place in the subtree T_1, label 2 is assigned to a leaf in T_1, and Σ_1, $\widetilde{\Sigma}_1$ as well as labels $3, \ldots, n - n_1$ are all in T_2. This shows (3.33).

- *Proof of* (3.34). For a tree $T \in \mathbb{T}_n^o \setminus \mathbb{T}_n^{b,o}$, conditionally given that the degree of the root is k, consider the subtrees $S_1, \ldots, S_k$ of sizes $n_1, \ldots, n_k$ in which T splits at the root (taking one edge each as a root edge for S_i, $1 \leq i \leq k$). First, we want to show that we have positive probability to decrease the degree of the root vertex by one. Therefore, apply the following procedure. In (i)$^{\mathrm{d}}$ insert the leaf in S_k to increase its size by one, assign label 2 to one of the subtrees $S_2, \ldots, S_k$, and consider the subtree S_1 in the following.

  - At the first branch point Ω_1 of S_1, the tree splits into k_1 subtrees, say, $S_{1,1}, \ldots, S_{1,k_1}$. Assign labels $\Sigma_1, \widetilde{\Sigma}_1$ to leaves in $S_{1,1}$, S_{1,k_1}, respectively.
  - Assign label 3 to a leaf in one of the subtrees $S_{1,2}, \ldots, S_{1,k_1-1}$.
  - Perform branch merging to obtain a tree S_1' from S_1 whose size has decreased by one, as leaf $\widetilde{\Sigma}_1$ gets lost under branch merging.

After branch merging, the tree T becomes a tree $T' \in \mathbb{T}_n^o$, which splits at the root into subtrees $S_1', \dots, S_k'$ of sizes $n_1 - 1, n_2, \dots, n_{k-1}, n_k + 1$, respectively. We perform this procedure $n_1 - 1$ times, and obtain a tree T'', which splits into subtrees $S_1'', \dots, S_k''$ of sizes $1, n_2, \dots, n_{k-1}, n_k + n_1 - 1$. Note that S_1'' consists of one single edge and one leaf. Now, perform the leaf insertion in S_k'', assign label Σ_1 to a leaf in S_k'', label $\widetilde{\Sigma}_1$ to the single leaf in S_1'', and label 2 to a leaf in one of the subtrees $S_2'', \dots, S_{k-1}''$. After branch merging, we only have a split into $k-1$ subtrees at the root edge. Clearly, we can perform this procedure another $k-2$ times, and, eventually, with positive probability, we obtain a tree with a root of degree 1, which we denote by $\overline{T}$. The tree $\overline{T}$ has one edge connecting the root and a branch point at which the tree splits into m, say, subtrees $\overline{S}_1, \dots, \overline{S}_m$.

Next, we want to show that there is positive probability to obtain a tree from $\overline{T}$ with root degree one and first branch point degree two. Therefore, repeat the procedure as above $m-2$ times with $\overline{S}_1, \dots, \overline{S}_m$. Eventually, with positive probability, we obtain a tree with a binary branch point, and denote the two subtrees of sizes n_1 and $n - n_1$ by $\overline{T}_1$ and $\overline{T}_2$, respectively.

We now use the induction hypothesis related to (3.34). Assigning label 2 to a leaf in $\overline{T}_2$, and restricting the remaining leaf insertion and sampling procedure for the leaves $\Sigma_1, \widetilde{\Sigma}_1$ and the labels $3, \dots, n_1 + 1$ to $\overline{T}_1$ (and swapping the roles of $\overline{T}_1$ and $\overline{T}_2$ afterwards), by the induction hypothesis, we obtain positive probability to turn $\overline{T}_1$ and $\overline{T}_2$ into binary trees, and hence, positive probability to turn $\overline{T}$ into a binary tree, i.e. (3.34) follows.

Hence, for any $n \in \mathbb{N}$, the Markov chain $(\tau_k, k \geq 0)$ on $\mathbb{T}_n^o$ has exactly one closed communicating class, namely $\mathbb{T}_n^{b,o}$. For $T \in \mathbb{T}_n^0$, $\mathbb{P}(\tau_1 = T \mid \tau_0 = T) > 0$, i.e. the chain is aperiodic. We conclude that $(\tau_k, k \geq 0)$ has a unique stationary distribution π, supported on $\mathbb{T}_n^{b,o}$, and that the distribution of τ_k converges to π as $k \to \infty$. □

Theorem 3.41 gives evidence that the Brownian CRT should be the *unique* stationary distribution of the Branch Merging Markov Chain.

Conjecture 3.42. *The Markov chain* $(\mathcal{B}_k, k \geq 0)$ *operating on* $\mathbb{T}_{\mathrm{w}}$ *has a unique stationary distribution given by the Brownian CRT, and* $\mathcal{B}_k$ *converges in distribution to* *the Brownian CRT as $k \to \infty$ with respect to the Gromov-Hausdorff topology (both up to scaling distances by a random factor).*

Remark 3.43. Note that the $\mathbb{R}$-trees we consider are compact, and hence, continuum trees are of finite height. The convergence to the Brownian CRT in Conjecture 3.42 is with respect to the Gromov-Hausdorff topology. Fix $\epsilon > 0$. For any continuum tree $\mathcal{B}_0 := \mathcal{T}$ and any branch point, we consider the subtrees attached of height $\geq \epsilon$. The probability that the two leaves Σ_1 and $\widetilde{\Sigma}_1$ in the leaf sampling procedure are in two distinct subtrees of this branch point is positive, and the number of subtrees with height $\geq \epsilon$ is reduced by one after the branch merging transition. Eventually, the BMMC will turn this branch point into a binary branch point (ignoring subtrees of size $< \epsilon$). On the other hand, subtrees of size $< \epsilon$ do not matter for convergence in the Gromov-Hausdorff sense, and hence we conjecture the convergence in distribution of the BMMC $(\mathcal{B}_k, k \geq 0)$ to the Brownian CRT.

3.4.3 General branch merging and the leaf embedding problem

We now turn to a more general setting where leaf sampling is not based on uniform sampling from the mass measure as in the BMMC. As a tree analogue of merging of (α, θ_i)-strings of beads, $i \in [k]$, in the special case when $k = 3$ and $\theta_1 = \theta_2 = 1 - \alpha$, $\theta_3 = \alpha$, we introduce branch merging on Ford CRTs, and give two applications.

This procedure will be more involved than the branch merging algorithm of the BMMC. In order to define branch merging on general Ford CRTs, using our merging algorithm for strings of beads, we need a Dirichlet$(\alpha, 1-\alpha, 1-\alpha)$ mass split between the three branches of the reduced tree structure. In general, this cannot be achieved via uniform sampling of leaves from the mass measure, see e.g. [60], Section 4.2. Therefore, we have to apply the $(\alpha, 1-\alpha)$-coin tossing construction in $\mathcal{T}^{\alpha, 1-\alpha}$ (which corresponds to uniform sampling when $\alpha = 1/2$). A similar generalisation is needed for the partition sampling procedure in (ii)$^{\mathrm{MC}}$.

It is a straightforward generalisation of Lemma 3.40 to conclude that Ford CRTs are distributionally invariant under this general procedure. However, our goal is to establish a coupling between the sequences of reduced trees related to the $(\alpha, 1-\alpha)$-

and the $(\alpha, 2-\alpha)$-model. The simple procedure involving two sampled leaves is enough to relate $\mathcal{R}_1^{\alpha,1-\alpha}$ and $\mathcal{R}_1^{\alpha,2-\alpha}$ but not to apply the procedure recursively, and to couple the sequences $(\mathcal{R}_k^{\alpha,1-\alpha}, k \geq 1)$ and $(\mathcal{R}_k^{\alpha,2-\alpha}, k \geq 1)$: it is necessary to find the bead of the second leaf on $\mathcal{R}_1^{\alpha,2-\alpha}$ to obtain a Dirichlet$(\alpha, 1-\alpha, 2-\alpha)$ mass split at the atom of label 2 in $\widetilde{\mathcal{R}}_1$ which identifies the subtree we have to select for the recursive application of the merging procedure. Therefore, it is crucial for the recursive leaf embedding procedure to obtain successively finer allocations of the non-embedded leaf labels, that is, after step $k \geq 1$, the labels $\{k+1, k+2, \ldots\}$, onto the atoms of the mass measure $\mu_{\widetilde{\mathcal{R}}_k}$ on the reduced tree $\widetilde{\mathcal{R}}_k$.

In the case $\alpha = 1/2$ and $\theta = 3/2$, in which the Brownian CRT arises as scaling limit, we do not only obtain a coupling of the reduced trees; in fact, we provide a significant simplification of the solution to the leaf embedding problem for $(1/2, 3/2)$-tree growth process. We construct a Brownian CRT with embedded leaf labels according to the $(1/2, 3/2)$-rule from a Brownian CRT equipped with the natural mass measure. In this sense, uniform sampling from the mass measure on the Brownian CRT in combination with branch merging allows to embed leaf labels in the $(1/2, 3/2)$ case by using a much simpler approach than the $(1/2, 3/2)$-coin tossing construction of Lemma 3.36.

3.4.3.1 Branch merging on Ford CRTs

We now present our branch merging algorithm on Ford CRTs $\mathcal{T}^{\alpha,1-\alpha}$, $\alpha \in (0,1)$.

Algorithm 3.44 (Branch merging)**.** Let $(\mathcal{T}^{\alpha,1-\alpha}, d, \rho, \mu)$ be a Ford CRT with parameter $\alpha \in (0,1)$, root ρ, mass measure μ and metric $d : \mathcal{T}^{\alpha,1-\alpha} \times \mathcal{T}^{\alpha,1-\alpha} \to \mathbb{R}_0^+$.

(i) *Start configuration.* Embed three leaves $\Omega_0, \Omega_1, \Omega_2$, i.e. the tree $\mathcal{R}_3^{\alpha,1-\alpha}$, into $\mathcal{T}^{\alpha,1-\alpha}$ by using the $(\alpha, 1-\alpha)$-coin tossing construction, and denote by Ω the first branch point of the spines from the root ρ to $\Omega_0, \Omega_1, \Omega_2$. We obtain the three branches $\widetilde{E}_0 :=]]\Omega, \Omega_0]], \widetilde{E}_1 :=]]\Omega, \Omega_1]], \widetilde{E}_2 :=]]\Omega, \Omega_2]]$. Define the start configuration $(\Sigma_1, \widetilde{\Sigma}_1, \Theta)$ by

$$\left(\Sigma_1, \widetilde{\Sigma}_1, \Theta\right) := \begin{cases} (\Omega_2, \Omega_0, \Omega_1) & \text{if } \widetilde{E}_0 \cap \widetilde{E}_1 \neq \emptyset, \\ (\Omega_1, \Omega_0, \Omega_2) & \text{if } \widetilde{E}_0 \cap \widetilde{E}_2 \neq \emptyset, \\ (\Omega_0, \Omega_1, \Omega_2) & \text{if } \widetilde{E}_1 \cap \widetilde{E}_2 \neq \emptyset. \end{cases}$$

Write $\mathcal{R}_2 = \mathcal{R}(\mathcal{T}^{\alpha.1-\alpha}, \Sigma_1, \widetilde{\Sigma}_1)$ for the reduced tree spanned by ρ, Σ_1, $\widetilde{\Sigma}_1$, $\rho_\Theta := \pi^{\mathcal{R}_2}(\Theta)$ and $\mathcal{S}_x := (\pi^{\mathcal{R}_2})^{-1}(x)$ for any $x \in \mathcal{R}_2$. As illustrated in Figure 3.4, we obtain a split of $\mathcal{R}_2$ into the 4 branches $E_0 :=]]\rho, \Omega[[, E_1 :=]]\Omega, \Sigma_1[[, E_2 :=]]\Omega, \rho_\Theta[[, E_3 :=]]\rho_\Theta, \widetilde{\Sigma}_1[[$.

(ii) *Partition sampling.* Consider i.i.d. random variables $(Z_n)_{n\geq 3} \sim \mu_{\mathcal{R}_2}$. Let $Y_2 := \rho_\Theta$, and, given Y_k, $2 \leq k \leq n$, define Y_{n+1} as follows.

- If $Z_{n+1} \in \bigcup_{k=2}^n [[\rho, Y_k]]$, set $Y_{n+1} = Z_{n+1}$.
- Otherwise, perform $(\alpha, 1-\alpha)$-coin tossing sampling on the connected component of Z_{n+1} in $\mathcal{R}_2 \setminus \bigcup_{k=2}^n [[\rho, Y_k]]$ to identify the atom Y_{n+1}, i.e. on the string of beads related to the branch $[[Y_n^*, \Sigma_n^*]]$ in $\mathcal{R}_2$ where

$$\Sigma_n^* := \begin{cases} \Sigma_1 & \text{if } Z_{n+1} \in [[\rho, \Sigma_1]], \\ \widetilde{\Sigma}_1 & \text{if } Z_{n+1} \in [[\rho, \widetilde{\Sigma}_1]], \end{cases}$$

  and $Y_n^* := \arg\max\{d(\rho, Y) | Y \in \{\Omega, Y_2 \ldots, Y_n\} \text{ and } Y \in [[\rho, \Sigma_n^*]]\}$.

Define the label sets $\Pi_x := \{k \geq 2 : Y_k = x\}, x \in \mathcal{R}_2$.

(iii) *Tree pruning and spine merging.* Define $\overline{\Pi}_i := \{\Pi_x, x \in E_i\}$, $i \in [3]$, and the collection of blocks

$$\mathcal{C} := \bigcup_{i\in[3]} \{\Pi_x \in \overline{\Pi}_i : \Pi_x \prec \Pi_y \forall y \in \mathcal{R}_2 \text{ with } \Pi_y \in \overline{\Pi}_i, d_{\mathcal{R}_2}(\rho_\Theta, y) > d_{\mathcal{R}_2}(\rho_\Theta, x)\},$$

where, for $A, B \subset \mathbb{N}$, $A \cap B \neq \emptyset$, $A \prec B :\Leftrightarrow \min A < \min B$.

Let $\Upsilon := \{x : \Pi_x \in \mathcal{C}\}$ be the associated set of cut points on the branches E_1, E_2, E_3 and write $\Upsilon = \{X_k, k \geq 1\}$ where, for $k \geq 1$,

$$X_k := x \in \mathcal{R}_2 \quad \text{such that} \quad |\{y \in \mathcal{R}_2 | \Pi_y \in \mathcal{C} \text{ and } \Pi_x \prec \Pi_y\}| = k - 1.$$

Define the merged branch $\widetilde{\mathcal{R}}_1$ via

$$\widetilde{\mathcal{R}}_1 := [[\rho, \Omega[[\cup \{\rho_\Theta\} \cup \left(\biguplus_{i \in [3]} E_i \right) \cup \{\Sigma_1\}, \tag{3.35}$$

where the operator $\biguplus$ was defined in (3.4), Section 3.3.1. We obtain the metric $d_{\widetilde{\mathcal{R}}_1} : \widetilde{\mathcal{R}}_1 \times \widetilde{\mathcal{R}}_1 \to \mathbb{R}_0^+$, $d_{\widetilde{\mathcal{R}}_1}(x, y) := |d_{\widetilde{\mathcal{R}}_1}(\rho, x) - d_{\widetilde{\mathcal{R}}_1}(\rho, y)|$ where

$$d_{\widetilde{\mathcal{R}}_1}(\rho, x) := \begin{cases} d(\rho, x) & \text{if } x \in [[\rho, \Omega[[, \\ d(\rho, \Omega) & \text{if } x = \rho_\Theta, \\ d(\rho, \Omega) + d_{\biguplus}(\rho_\Theta, x) & \text{if } x \in \bigcup_{i=1}^3 E_i, \\ d(\rho, \Omega) + d(\rho_\Theta, \Omega) + d(\rho_\Theta, \widetilde{\Sigma}_1) + d(\Omega, \Sigma_1) & \text{if } x = \Sigma_1. \end{cases}$$

(iv) *Update rule.* Define the metric space $(\widetilde{\mathcal{T}}_1, \tilde{d}_1)$ by $\widetilde{\mathcal{T}}_1 := \widetilde{\mathcal{R}}_1 \cup \bigcup_{x \in \widetilde{\mathcal{R}}_1} \mathcal{S}_x$, equipped with the metric $\tilde{d}_1 : \widetilde{\mathcal{T}}_1 \times \widetilde{\mathcal{T}}_1 \to \mathbb{R}_0^+$,

$$\tilde{d}_1(x, y) := \begin{cases} d_{\widetilde{\mathcal{R}}_1}(x, y) & \text{if } x, y \in \widetilde{\mathcal{R}}_1, \\ d_{\widetilde{\mathcal{R}}_1}(x', y') + d(x, x') + d(y, y') & \text{if } x \in \mathcal{S}_{x'}, y \in \mathcal{S}_{y'}, x' \neq y' \in \widetilde{\mathcal{R}}_1, \\ d_{\widetilde{\mathcal{R}}_1}(x, y') + d(y, y') & \text{if } x \in \widetilde{\mathcal{R}}_1, y \in \mathcal{S}_{y'}, y' \in \widetilde{\mathcal{R}}_1, \\ d(x, y) & \text{if } x, y \in \mathcal{S}_{x'}, x' \in \widetilde{\mathcal{R}}_1. \end{cases} \tag{3.36}$$

Furthermore, define the mass measure $\tilde{\mu}_1$ on $(\widetilde{\mathcal{T}}_1, \tilde{d}_1)$ by

$$\tilde{\mu}_1(\overline{\mathcal{T}}_x) := \begin{cases} \sum_{y \in \overline{\mathcal{T}}_x \cap \widetilde{\mathcal{R}}_1} \mu(\mathcal{S}_y) & \text{if } x \in \widetilde{\mathcal{R}}_1, \\ \mu(\overline{\mathcal{T}}_x \cap \mathcal{S}_{x'}) & \text{if } x \in \mathcal{S}_{x'}, x' \in \widetilde{\mathcal{R}}_1, \end{cases} \tag{3.37}$$

where $\overline{\mathcal{T}}_x := \{y \in \widetilde{\mathcal{T}}_1 : x \in [[\rho, y]]\}$, which uniquely identifies $\tilde{\mu}_1$.

Note that the only step which was directly carried forward from the transition rule of the BMMC (i.e. without any generalisation) is step (iv)$^{\mathrm{MC}}$=(iv), and that we keep the information from the label sets.

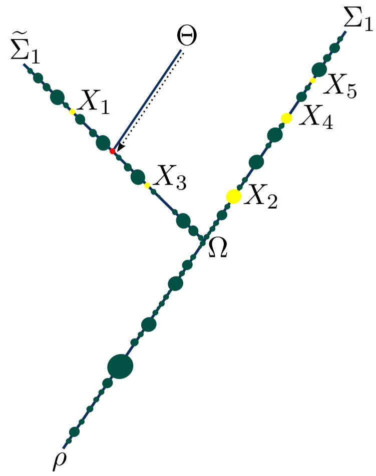

Fig. 3.4 The reduced tree split into four branches with subtree masses as atoms. First five cut points $X_1, \ldots, X_5$ are displayed.

Theorem 3.45 (The merged tree)**.** *The quadruple $(\widetilde{\mathcal{T}}_1, \tilde{d}_1, \rho, \tilde{\mu}_1)$ constructed in Algorithm 3.44, see Figure 3.5, is a random $\mathbb{R}$-tree with the following properties.*

(i) *Let $\mathcal{S}_x$, $x \in \widetilde{\mathcal{R}}_1$, denote the connected components of $\widetilde{\mathcal{T}}_1 \setminus \widetilde{\mathcal{R}}_1$, where the branch point base point of $\mathcal{S}_x$ in $\widetilde{\mathcal{T}}_1$ is x, and denote the discrete mass measure on $\widetilde{\mathcal{R}}_1$ obtained by assigning mass $\mu(\mathcal{S}_x)$ to $x \in \widetilde{\mathcal{R}}_1$ by $\mu_{\widetilde{\mathcal{R}}_1}$. Then $(\widetilde{\mathcal{R}}_1 = [[\rho, \Sigma_1]], \mu_{\widetilde{\mathcal{R}}_1})$ is an $(\alpha, 2-\alpha)$-string of beads.*

(ii) *Given $(\widetilde{\mathcal{R}}_1, \mu_{\widetilde{\mathcal{R}}_1})$, the trees*

$$\left(\mathcal{S}_x, \tilde{\mu}_1(\mathcal{S}_x)^{-\alpha}\tilde{d}_1 \restriction_{\mathcal{S}_x}, \rho_{\mathcal{S}_x} := x, \tilde{\mu}_1(\mathcal{S}_x)^{-1}\tilde{\mu}_1 \restriction_{\mathcal{S}_x}\right), \qquad x \in \widetilde{\mathcal{R}}_1,$$

are isometric to independent copies of $(\mathcal{T}^{\alpha,1-\alpha}, d, \rho, \mu)$.

(iii) *The tree $(\widetilde{\mathcal{T}}_1, \tilde{d}_1, \rho, \tilde{\mu}_1)$ is a Ford CRT with parameter $\alpha \in (0,1)$, i.e.*

$$(\widetilde{\mathcal{T}}_1, \tilde{d}_1, \rho, \tilde{\mu}_1) \stackrel{d}{=} (\mathcal{T}^{\alpha,1-\alpha}, d, \rho, \mu) \stackrel{d}{=} (\mathcal{T}^{\alpha,2-\alpha}, d, \rho, \mu).$$

We prove Theorem 3.45 in a series of propositions.

Proposition 3.46. *The leaf embedding procedure in step* (i) *of Algorithm 3.44 induces the mass split*

$$(\mu_{\mathcal{R}_2}(E_0), \mu_{\mathcal{R}_2}(E_1), \mu_{\mathcal{R}_2}(E_2), \mu_{\mathcal{R}_2}(E_3), \mu_{\mathcal{R}_2}(\rho_\Theta))$$

which has a Dirichlet$(\alpha, 1-\alpha, \alpha, 1-\alpha, 1-\alpha)$ *distribution. Moreover, the aggregated mass split*

$$(\mu_{\mathcal{R}_2}(E_0), \mu_{\mathcal{R}_2}(\rho_\Theta), \mu_{\mathcal{R}_2}(E_1) + \mu_{\mathcal{R}_2}(E_2) + \mu_{\mathcal{R}_2}(E_3))$$

has a Dirichlet$(\alpha, 1-\alpha, 2-\alpha)$ *distribution.*

Proof. We refer to [60], Section 3.5. It was shown that the leaf embedding procedure on $\mathcal{T}^{\alpha,1-\alpha}$ yields a reduced tree $\mathcal{R}(\mathcal{T}^{\alpha,1-\alpha}, \Sigma_1, \widetilde{\Sigma}_1, \Theta)$ subject to a Dirichlet mass split with parameter α for each internal branch, and $(1-\alpha)$ for each external branch. To see the first claim note that the mass of the branch $]]\rho_\Theta, \Theta[[$ of $\mathcal{R}(\mathcal{T}^{\alpha,1-\alpha}, \Sigma_1, \widetilde{\Sigma}_1, \Theta)$ corresponds to the mass of the atom ρ_Θ on $\mathcal{R}_2$. The second claim follows from the aggregation property of the Dirichlet distribution, Proposition 3.22(i). $\square$

Proposition 3.47. *Let* Π_x*,* $x \in \mathcal{R}_2$*, be the label sets of step* (ii) *of Algorithm 3.44. Furthermore, let* $(\Sigma_k, k \geq 1)$ *be a sequence of leaves as in Theorem 3.35. Then, for any* $x \in \mathcal{R}_2$ *and* $\mathcal{S}_x = (\pi^{\mathcal{R}_2})^{-1}(x)$*, we have* $\Pi_x \stackrel{d}{=} \{k \geq 2 : \Sigma_k \in \mathcal{S}_x\}$*.*

Proof. This follows from Lemma 3.37. Given $\mathcal{R}(\mathcal{T}^{\alpha,1-\alpha}, \Sigma_1, \ldots, \Sigma_k)$ the selection rule for the atom x which Σ_{k+1} is attached to, is such that an edge of the reduced tree is selected proportionally to mass. If an external edge is selected, the $(\alpha, 1-\alpha)$-coin tossing construction of Lemma 3.36 is applied. If an internal edge is selected, an atom is chosen proportionally to weight. Projecting the labels of $\Sigma_k, k \geq 2$, onto $\mathcal{R}_2$ yields the claimed equality in distribution of the label sets, where we set $\Sigma_2 = \Theta$. $\square$

Proposition 3.48. *Let* $\mathcal{A}(\mu_{\mathcal{R}_2}) := \{x \in \mathcal{R}_2 : \mu_{\mathcal{R}_2}(x) > 0\}$ *be the set of atoms of* $\mu_{\mathcal{R}_2}$*. Conditionally given* $\mathcal{A}(\mu_{\mathcal{R}_2})$*, the quadruples*

$$\left(\mathcal{S}_x, \mu(\mathcal{S}_x)^{-\alpha} d\restriction_{\mathcal{S}_x}, \rho_{\mathcal{S}_x} := x, \mu(\mathcal{S}_x)^{-1} \mu\restriction_{\mathcal{S}_x}\right), \quad x \in \mathcal{A}(\mu_{\mathcal{R}_2}),$$

are isometric to independent copies of $(\mathcal{T}^{\alpha,1-\alpha}, d, \rho, \mu)$*.*

Proof. This follows directly from Theorem 3.35 and Lemma 3.36. $\square$

Proposition 3.49. *The atom ρ_Θ splits $\widetilde{\mathcal{R}}_1$ into $]]\rho, \rho_\Theta[[$, ρ_Θ, $]]\rho_\Theta, \Sigma_1[[$ such that*

$$\left(\mu_{\widetilde{\mathcal{R}}_1}(]]\rho, \rho_\Theta[[), \mu_{\widetilde{\mathcal{R}}_1}(\rho_\Theta), \mu_{\widetilde{\mathcal{R}}_1}(]]\rho_\Theta, \Sigma_1[[)\right) \sim \text{Dirichlet}(\alpha, 1-\alpha, 2-\alpha). \tag{3.38}$$

Furthermore, the pair

$$\left(\mu_{\widetilde{\mathcal{R}}_1}(]]\rho_\Theta, \Sigma_1[[)^{-\alpha}]]\rho_\Theta, \Sigma_1[[, \mu_{\widetilde{\mathcal{R}}_1}(]]\rho_\Theta, \Sigma_1[[)^{-1}\mu_{\widetilde{\mathcal{R}}_1} \restriction_{]]\rho_\Theta,\Sigma_1[[}\right) \tag{3.39}$$

is an $(\alpha, 2-\alpha)$-string of beads, and the pair

$$\left(\mu_{\widetilde{\mathcal{R}}_1}(]]\rho, \rho_\Theta[[)^{-\alpha}]]\rho, \rho_\Theta[[, \mu_{\widetilde{\mathcal{R}}_1}(]]\rho, \rho_\Theta[[)^{-1}\mu_{\widetilde{\mathcal{R}}_1} \restriction_{]]\rho,\rho_\Theta[[}\right) \tag{3.40}$$

is an (α, α)-string of beads. The strings of beads (3.39) *and* (3.40) *are independent of each other, and of the Dirichlet mass split* (3.38)*.*

Proof. First of all, recall the definition of E_i, $i \in [3]$, i.e.

$$E_1 =]]\Omega, \Sigma_1[[, \quad E_2 =]]\Omega, \rho_\Theta[[, \quad E_3 =]]\rho_\Theta, \widetilde{\Sigma}_1[[,$$

and note that, as a consequence of the embedding procedure, $\mu_{\widetilde{\mathcal{R}}_1}(E_1)^{-\alpha}E_i$, $i \in [3]$, equipped with $\mu_{\widetilde{\mathcal{R}}_1}(E_i)^{-1}\mu_{\widetilde{\mathcal{R}}_1} \restriction_{E_i}$, $i \in [3]$, are independent (α, θ_i)-strings of beads, respectively, where $\theta_1 = \theta_3 = 1-\alpha$, $\theta_2 = \alpha$, see [60], Lemma 3.13. Hence, by construction of $]]\rho_\Theta, \Sigma_1[[\subset \widetilde{\mathcal{R}}_1$ using the operator $\uplus$ from (3.4) and Theorem 3.15,

$$\left(\mu_{\widetilde{\mathcal{R}}_1}(]]\rho_\Theta, \Sigma_1[[)^{-\alpha}]]\rho_\Theta, \Sigma_1[[, \mu_{\widetilde{\mathcal{R}}_1}(]]\rho_\Theta, \Sigma_1[[)^{-1}\mu_{\widetilde{\mathcal{R}}_1} \restriction_{]]\rho_\Theta,\Sigma_1[[}\right)$$

is an $(\alpha, 2-\alpha)$-string of beads. On the other hand, the interval $E_0 =]]\rho, \rho_\Theta[[$ with distance $d \restriction_{]]\rho,\rho_\Theta[[}$ is not affected by Algorithm 3.44, in particular $d \restriction_{]]\rho,\rho_\Theta[[} = d_{\widetilde{\mathcal{R}}_1} \restriction_{]]\rho,\rho_\Theta[[}$. Therefore, $(\mu_{\widetilde{\mathcal{R}}_1}(]]\rho, \rho_\Theta[[)^{-\alpha}]]\rho, \rho_\Theta[[, \mu_{\widetilde{\mathcal{R}}_1}(]]\rho, \rho_\Theta[[)^{-1}\mu_{\widetilde{\mathcal{R}}_1} \restriction_{]]\rho,\rho_\Theta[[})$ is an (α, α)-string of beads, cf. Theorem 3.35. By Proposition 3.13, the strings of beads (3.39) and (3.40) are independent of each other, and of the mass split (3.38). (3.38) is a direct consequence of the branch merging operation and Proposition 3.46. □

Lemma 3.50. *The pair $(\widetilde{\mathcal{R}}_1, \mu_{\widetilde{\mathcal{R}}_1})$ is an $(\alpha, 2-\alpha)$-string of beads.*

Proof. By Proposition 3.49 we have a Dirichlet$(\alpha, 1-\alpha, 2-\alpha)$ mass split, an independent (α, α)- and an independent $(\alpha, 2-\alpha)$-string of beads as in Proposition 3.18, and we conclude that $(\widetilde{\mathcal{R}}_1, \mu_{\widetilde{\mathcal{R}}_1})$ is an $(\alpha, 2-\alpha)$-string of beads. □

Proof of Theorem 3.45. Lemma 3.50 proves Theorem 3.45(i). For (ii), note that, as an immediate consequence of the nature of the branch merging algorithm, the rescaled subtrees $\tilde{\mu}_1(\mathcal{S}_x)^{-\alpha}\mathcal{S}_x = \mu_{\widetilde{\mathcal{R}}_1}(x)^{-\alpha}\mathcal{S}_x, x \in \widetilde{\mathcal{R}}_1$, are i.i.d. copies of $\mathcal{T}^{\alpha,1-\alpha}$, see Proposition 3.48. (iii) is a direct consequence of the spinal decomposition theorem for $\theta = 2-\alpha$, cf. Lemma 3.36, and Corollary 3.33. □

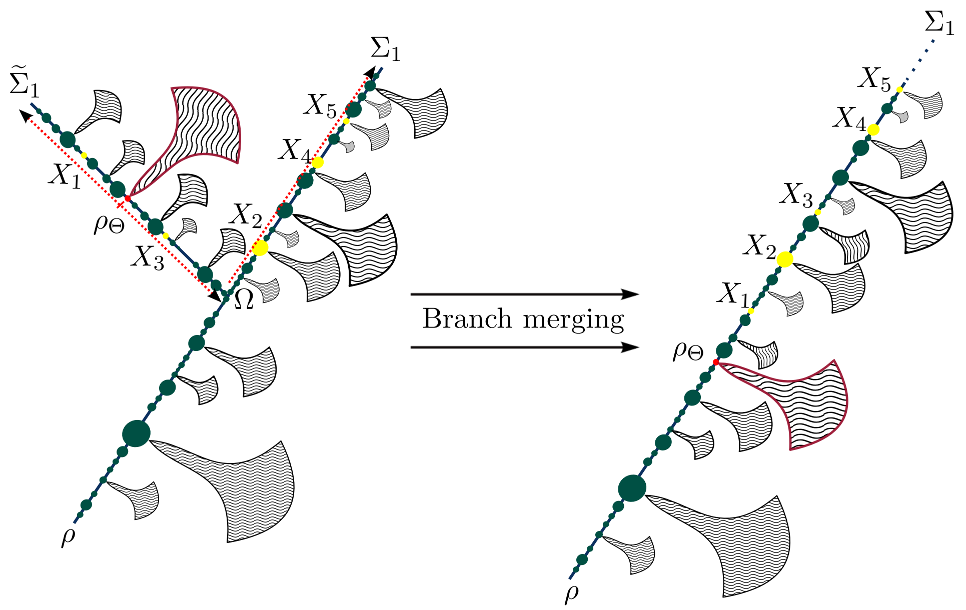

Fig. 3.5 Branch merging on Ford CRTs. Reduced tree structure with subtrees attached in the atoms. Strings of beads are merged and subtrees are carried forward.

3.4.3.2 Application: the leaf embedding problem

In order to apply the branch merging procedure to the leaf embedding problem we first describe how to apply the branch merging procedure recursively on Ford CRTs.

Recursive branch merging. Let $\mathcal{T}^{\alpha,1-\alpha}$ be a Ford CRT for some $\alpha \in (0,1)$ as in Section 3.4.3.1. Consider $(\widetilde{\mathcal{T}}_k, \tilde{d}_k, \rho, \tilde{\mu}_k)$, $k \in \mathbb{N}$, with leaves $\Sigma_1, \ldots, \Sigma_k$ and root ρ, denoting the random $\mathbb{R}$-tree obtained by performing the branch merging operation given by Algorithm 3.44 in $(\mathcal{T}^{\alpha,1-\alpha}, d, \rho, \mu)$, and in $k-1$ subsequently selected subtrees according to the following scheme, which explains how to obtain $(\widetilde{\mathcal{T}}_{k+1}, \tilde{d}_{k+1}, \rho, \tilde{\mu}_{k+1})$ from $(\widetilde{\mathcal{T}}_k, \tilde{d}_k, \rho, \tilde{\mu}_k)$, $k \geq 1$.

Consider the subtree $\widetilde{\mathcal{R}}_k := \mathcal{R}(\widetilde{\mathcal{T}}_k, \Sigma_1, .., \Sigma_k)$ spanned by the leaves $\Sigma_1, \ldots, \Sigma_k$ (as embedded in the previous steps of the algorithm), which is equipped with a random point measure $\mathcal{P}_{\widetilde{\mathcal{R}}_k}$ given by

$$\mathcal{P}_{\widetilde{\mathcal{R}}_k} := \sum_{x \in \mathcal{A}(\mu_{\widetilde{\mathcal{R}}_k})} \delta_{(x, \mathcal{S}_x, \Pi_x)}$$

where the mass measure $\mu_{\widetilde{\mathcal{R}}_k}$ is the push-forward of $\tilde{\mu}_k$ via the projection

$$\pi^{\widetilde{\mathcal{R}}_k} : \widetilde{\mathcal{T}}_k \to \widetilde{\mathcal{R}}_k, \qquad \sigma \to \tilde{\Phi}^{(k)}_{\rho,\sigma}\Big(\sup\Big\{t \geq 0 : \tilde{\Phi}^{(k)}_{\rho,\sigma}(t) \in \widetilde{\mathcal{R}}_k\Big\}\Big)$$

mapping the connected components of $\widetilde{\mathcal{T}}_k \setminus \widetilde{\mathcal{R}}_k$ onto the closest point of the reduced tree $\widetilde{\mathcal{R}}_k$, the set $\mathcal{A}(\mu_{\widetilde{\mathcal{R}}_k}) = \{x \in \widetilde{\mathcal{R}}_k : \mu_{\widetilde{\mathcal{R}}_k}(x) > 0\}$ is the set of atoms of $\mu_{\widetilde{\mathcal{R}}_k}$, and $\tilde{\Phi}^{(k)}_{\rho,\sigma} : [0, \tilde{d}_k(\rho, \sigma)] \to \widetilde{\mathcal{T}}_k$, $\sigma \in \widetilde{\mathcal{T}}_k$, denote the unique geodesics associated with $\widetilde{\mathcal{T}}_k$. Suppose that the union of the sets Π_x, $x \in \mathcal{A}(\mu_{\widetilde{\mathcal{R}}_k})$, is $\mathbb{N}_{\geq k+1}$, i.e.

$$\bigcup_{x \in \mathcal{A}(\mu_{\widetilde{\mathcal{R}}_k})} \Pi_x = \mathbb{N}_{\geq k+1}.$$

Recursive application of the spinal decomposition theorem for binary fragmentation CRTs $\mathcal{T}^{\alpha,1-\alpha}$ (Lemma 3.36) yields that the point process $\mathcal{P}_{\widetilde{\mathcal{R}}_k}$ defines a labelled bead space in the following sense (see also [60, 61] for subtree decompositions).

Definition 3.51 (Labelled bead space)**.** Let $(\mathcal{T}^{\alpha,1-\alpha}, d, \rho, \mu)$ be a Ford CRT for some $\alpha \in (0, 1)$.

(i) A *labelled bead* is a pair $(\mathcal{S}, \Pi_\mathcal{S})$ where $(\mathcal{S}, d_\mathcal{S})$ is a pointed metric space with a distinguished vertex $\rho_\mathcal{S}$ (the "root" of $\mathcal{S}$) and equipped with a mass measure $\mu_\mathcal{S}$, $\Pi_\mathcal{S}$ is an infinite label set, and, for $\Delta_\mathcal{S} := \mu_\mathcal{S}(\mathcal{S})$,

$$\Big(\mathcal{S}, \Delta_\mathcal{S}^{-\alpha} d_\mathcal{S}, \rho_\mathcal{S}, \Delta_\mathcal{S}^{-1}\mu_\mathcal{S}\Big) \overset{d}{=} \Big(\mathcal{T}^{\alpha,1-\alpha}, d, \rho, \mu\Big).$$

(ii) A *labelled bead space* is a pair $(\mathcal{R}, \mathcal{P}_{\mathcal{R}})$ where $(\mathcal{R}, d_{\mathcal{R}})$ is a metric space with a distinguished vertex $\rho_{\mathcal{R}}$ (the "root" of $\mathcal{R}$) and equipped with a discrete mass measure $\mu_{\mathcal{R}}$, and $\mathcal{P}_{\mathcal{R}}$ is a point process of the form

$$\mathcal{P}_{\mathcal{R}} := \sum_{x \in \mathcal{A}(\mu_{\mathcal{R}})} \delta_{(x, \mathcal{S}_x, \Pi_x)}$$

where $\mathcal{A}(\mu_{\mathcal{R}})$ is the set of atoms of the measure $\mu_{\mathcal{R}}$ on $\mathcal{R}$, i.e. $\mathcal{A}(\mu_{\mathcal{R}}) := \{x \in \mathcal{R} : \mu_{\mathcal{R}}(x) > 0\}$, such that $(\mathcal{S}_x, \Pi_x)$ is a labelled bead for every $x \in \mathcal{A}(\mu_{\mathcal{R}})$ with root $\rho_{\mathcal{S}_x} := x$, distance $d_{\mathcal{S}_x}$ and mass measure $\mu_{\mathcal{S}_x}$.

In order to obtain the branch linking $\widetilde{\mathcal{R}}_k$ and Σ_{k+1} choose the unique labelled bead $(\mathcal{S}_x, \Pi_x)$ whose label set contains $k+1$, say $(\mathcal{S}_k, \Pi_k)$. Note that $d \restriction_{\mathcal{S}_k} = \tilde{d}_k \restriction_{\mathcal{S}_k}$ and $\mu \restriction_{\mathcal{S}_k} = \tilde{\mu}_k \restriction_{\mathcal{S}_k}$, since each subtree related to a labelled bead on $\widetilde{\mathcal{R}}_k$ was present as a subtree of $(\widetilde{\mathcal{T}}_k, \tilde{d}_k, \rho)$ as well as in the initial CRT $(\mathcal{T}^{\alpha,1-\alpha}, d, \rho)$.

We now consider the rescaled tree

$$\left(\mathcal{S}_k, \tilde{d}_{\mathcal{S}_k}, \rho_k, \tilde{\mu}_{\mathcal{S}_k}\right) := \left(\mathcal{S}_k, \mu(\mathcal{S}_k)^{-\alpha} d \restriction_{\mathcal{S}_k}, \rho_k, \mu(\mathcal{S}_k)^{-1} \mu \restriction_{\mathcal{S}_k}\right) \tag{3.41}$$

completed by the root $\rho_k := \pi^{\widetilde{\mathcal{R}}_k}(\mathcal{S}_k)$ and equipped with the label set Π_k. Since

$$\left(\mathcal{S}_k, \tilde{d}_{\mathcal{S}_k}, \rho_k, \tilde{\mu}_{\mathcal{S}_k}\right) \overset{d}{=} \left(T^{\alpha,1-\alpha}, d, \rho, \mu\right)$$

we can apply the procedure of Algorithm 3.44 when interpreted appropriately.

(i) Choose the start configuration as in Algorithm 3.44(i) but label the three selected leaves by $(\Sigma_{k+1}, \widetilde{\Sigma}_{k+1}, \Theta_k)$ instead of $(\Sigma_1, \widetilde{\Sigma}_1, \Theta)$, and the branch point by Ω_k instead of Ω. Furthermore, consider the label set $\Pi_k := \{n_i, i \geq 1\}$ where subscripts are assigned according to the increasing bijection, i.e. $n_1 = k+1$ and $n_2 = k'$, where $k' := \min\{n \in \mathbb{N} : n \in \Pi_k \setminus \{k+1\}\}$.

Let $(Z_{n_i})_{i \geq 3}$, and let $Y_{n_2} = Y_{k'} := \rho_{\Theta_k}$ be a sequence of i.i.d. random variables where ρ_{Θ_k} denotes the projection of Θ_k onto the reduced tree $\mathcal{R}(\mathcal{S}_k, \Sigma_{k+1}, \widetilde{\Sigma}_{k+1})$. Given $Y_{n_i}, 1 \leq i \leq l$, define $Y_{n_{l+1}}$ in the following way.

- If $Z_{n_{l+1}} \in \bigcup_{i=2}^{l} [[\rho_k, Y_{n_i}]]$, set $Y_{n_{l+1}} = Z_{n_{l+1}}$.

- Otherwise, perform $(\alpha, 1-\alpha)$-coin tossing sampling on the connected component of $Z_{n_{l+1}}$ in $\mathcal{R}(\mathcal{S}_k, \Sigma_{k+1}, \widetilde{\Sigma}_{k+1}) \setminus \bigcup_{i=2}^{l}[[\rho_k, Y_{n_i}]]$ to identify the atom $Y_{n_{l+1}}$, i.e. on the string of beads related to $[[Y^*_{n_l}, \Sigma^*_{n_l}]]$ in $\mathcal{R}(\mathcal{S}_k, \Sigma_{k+1}, \widetilde{\Sigma}_{k+1})$ where

$$\Sigma^*_{n_l} := \begin{cases} \Sigma_{k+1} & \text{if } Z_{n_{l+1}} \in [[\rho_k, \Sigma_{k+1}]], \\ \widetilde{\Sigma}_{k+1} & \text{if } Z_{n_{l+1}} \in [[\rho_k, \widetilde{\Sigma}_{k+1}]], \end{cases}$$

$$Y^*_{n_l} := \arg\max\{d(\rho_k, Y) | Y \in \{\Omega_k, Y_{n_2} \dots, Y_{n_l}\} \text{ and } Y \in [[\rho_k, \Sigma^*_{n_l}]]\}.$$

Algorithm 3.44 applied to $(\mathcal{S}_k, \tilde{d}_{\mathcal{S}_k}, \rho_k, \tilde{\mu}_{\mathcal{S}_k})$ yields a CRT incorporating a spine transformation on the spine from ρ_k to the leaf Σ_{k+1}. We obtain a labelled bead space $(]]\rho_k, \Sigma_{k+1}]], \mathcal{P}_{]]\rho_k, \Sigma_{k+1}]]})$, i.e.

$$\mathcal{P}_{]]\rho_k, \Sigma_{k+1}]]} := \sum_{x \in \mathcal{A}(\tilde{\mu}_{]]\rho_k, \Sigma_{k+1}]]})} \delta_{(x, \mathcal{S}_x, \Pi_x)},$$

where $\tilde{\mu}_{]]\rho_k, \Sigma_{k+1}]]}$ denotes the mass measure on the merged spine $]]\rho_k, \Sigma_{k+1}]]$ whose atoms are the connected components of $\mathcal{S}_k \backslash]]\rho_k, \Sigma_{k+1}]]$ projected onto $]]\rho_k, \Sigma_{k+1}]]$, and

$$\mathcal{A}(\tilde{\mu}_{]]\rho_k, \Sigma_{k+1}]]}) = \left\{x \in]]\rho_k, \Sigma_{k+1}]] : \tilde{\mu}_{]]\rho_k, \Sigma_{k+1}]]}(x) > 0\right\}$$

is the associated set of atoms. Note that the union of the label sets Π_x, $x \in]]\rho_k, \Sigma_{k+1}]]$, of the labelled beads aligned along $]]\rho_k, \Sigma_{k+1}]]$ is $\Pi_k \setminus \{k+1\}$.

(ii) Denote the transformed, scaled back tree by $(\mathcal{S}'_k, \tilde{d}'_{\mathcal{S}'_k}, \rho_k, \tilde{\mu}'_{\mathcal{S}'_k})$. The reduced tree spanned by the root ρ and leaves $\Sigma_1, \dots, \Sigma_{k+1}$ is obtained via

$$\widetilde{\mathcal{R}}_{k+1} := \widetilde{\mathcal{R}}_k \cup]]\rho_k, \Sigma_{k+1}]].$$

We replace the point $(\rho_k, \mathcal{S}_k, \Pi_k)$ by a new series of labelled beads aligned along $]]\rho_k, \Sigma_{k+1}]]$ to update the point process $\mathcal{P}_{\widetilde{\mathcal{R}}_k}$, and get

$$\mathcal{P}_{\widetilde{\mathcal{R}}_{k+1}} := \sum_{x \in \mathcal{A}(\mu_{\widetilde{\mathcal{R}}_{k+1}})} \delta_{(x, \mathcal{S}_x, \Pi_x)},$$

where $\mathcal{A}(\mu_{\widetilde{\mathcal{R}}_{k+1}}) = \{x \in \widetilde{\mathcal{R}}_{k+1} : \mu_{\widetilde{\mathcal{R}}_{k+1}}(x) > 0\}$, and $\mu_{\widetilde{\mathcal{R}}_{k+1}}$ is uniquely determined via

$$\mu_{\widetilde{\mathcal{R}}_{k+1}} \restriction_{]]\rho_k,\Sigma_{k+1}]]} = \tilde{\mu}_{]]\rho_k,\Sigma_{k+1}]]}, \quad \mu_{\widetilde{\mathcal{R}}_{k+1}} \restriction_{\widetilde{\mathcal{R}}_k \setminus \{\rho_k\}} = \mu_{\widetilde{\mathcal{R}}_k}, \quad \mu_{\widetilde{\mathcal{R}}_{k+1}}(\rho_k) = 0.$$

More precisely, we have $\mathcal{P}_{\widetilde{\mathcal{R}}_{k+1}} := \mathcal{P}_{\widetilde{\mathcal{R}}_k} - \delta_{(\rho_{\mathcal{S}_k}, \mathcal{S}_k, \Pi_k)} + \mathcal{P}_{]]\rho_k,\Sigma_{k+1}]]}$, i.e.

$$\mathcal{P}_{\widetilde{\mathcal{R}}_{k+1}} = \sum_{x \in \mathcal{A}(\mu_{\widetilde{\mathcal{R}}_k}), x \neq \rho_{\mathcal{S}_k}} \delta_{(x, \mathcal{S}_x, \Pi_x)} + \sum_{x \in \mathcal{A}(\tilde{\mu}_{]]\rho_k,\Sigma_{k+1}]]})} \delta_{(x, \mathcal{S}_x, \Pi_x)}.$$

Informally speaking, we replace $(\mathcal{S}_k, d \restriction_{\mathcal{S}_k}, \rho_k, \mu \restriction_{\mathcal{S}_k})$ by attaching $(\mathcal{S}'_k, \tilde{d}'_{\mathcal{S}'_k}, \rho_k, \tilde{\mu}'_{\mathcal{S}'_k})$ to ρ_k, the base point of the subtree $(\mathcal{S}_k, d \restriction_{\mathcal{S}_k}, \mu \restriction_{\mathcal{S}_k}, \Pi_k)$ on $\widetilde{\mathcal{T}}_k$, where the tree $(\mathcal{S}'_k, \tilde{d}'_{\mathcal{S}'_k}, \rho_k, \tilde{\mu}'_{\mathcal{S}'_k})$ is obtained from the initial subtree $(\mathcal{S}_k, d \restriction_{\mathcal{S}_k}, \mu \restriction_{\mathcal{S}_k}, \Pi_k)$ by incorporating the distance changes related to the spine transformation and removing the two points Ω_k and $\widetilde{\Sigma}_{k+1}$ lost in the branch merging operation.

(iii) We define $(\widetilde{\mathcal{T}}_{k+1}, \tilde{d}_{k+1}, \rho, \tilde{\mu}_{k+1})$ by

$$\widetilde{\mathcal{T}}_{k+1} := \widetilde{\mathcal{R}}_{k+1} \cup \bigcup_{x \in \widetilde{\mathcal{R}}_{k+1}} \mathcal{S}_x$$

where the distance $\tilde{d}_{k+1}$ is the push-forward of $\tilde{d}_k$ via the analogue of (3.36), and the mass measure $\tilde{\mu}_{k+1}$ is induced by $\tilde{\mu}_k$ as in (3.37), see Algorithm 3.44 for the precise definition of $\tilde{d}_{k+1}$ and $\tilde{\mu}_{k+1}$ in the case $k = 0$, which can be easily adapted to the general case $k \geq 0$.

Corollary 3.52. *Let $(\mathcal{T}^{\alpha,1-\alpha}, d, \rho, \mu)$ be a Ford CRT for some $\alpha \in (0,1)$ as in Algorithm 3.44. The sequence of trees $((\widetilde{\mathcal{T}}_k, \tilde{d}_k, \rho, \tilde{\mu}_k), k \geq 1)$ constructed recursively according to Algorithm 3.44 and the recursive leaf embedding procedure, and the sequence of leaves $(\Sigma_k, k \geq 1)$ are such that*

$$(\mathcal{R}(\widetilde{\mathcal{T}}_k, \Sigma_1, \ldots, \Sigma_k), k \geq 1) \stackrel{d}{=} (\mathcal{R}_k^{\alpha,2-\alpha}, k \geq 1),$$

where $\mathcal{R}_k^{\alpha,2-\alpha}$ is the almost-sure limit as $n \to \infty$ of the trees $\mathcal{R}(T_n^{\alpha,2-\alpha}, [k])$ scaled by $n^{-\alpha}$ associated with the $(\alpha, 2-\alpha)$-tree growth process as in Proposition 3.31.

Proof. We prove Corollary 3.52 by induction. By Theorem 3.45, $\mathcal{R}(\widetilde{\mathcal{T}}_1, \Sigma_1)$ defines an $(\alpha, 2-\alpha)$-string of beads and is therefore distributed as $\mathcal{R}_1^{\alpha,2-\alpha}$. Now, let $k \in \mathbb{N}$, and assume that $\mathcal{R}(\widetilde{\mathcal{T}}_k, \Sigma_1, \ldots, \Sigma_k)$ is distributed as $\mathcal{R}_k^{\alpha,2-\alpha}$. Using the labelled bead space property, we conclude that the rescaled, transformed subtree

$$\left(\mathcal{S}'_k, \mu(\mathcal{S}'_k)^{-\alpha}\tilde{d}'_{\mathcal{S}'_k}, \rho_k, \mu(\mathcal{S}_k)^{-1}\tilde{\mu}'_{\mathcal{S}'_k}\right)$$

has the same distribution as $(\widetilde{\mathcal{T}}_1, \tilde{d}_1, \rho, \tilde{\mu}_1)$, and that Σ_{k+1} is such that

$$\mathcal{R}(\mu(S_k)^{-\alpha}S'_k, \Sigma_{k+1}) \overset{d}{=} \mathcal{R}(\widetilde{\mathcal{T}}_1, \Sigma_1).$$

Therefore, the subtree $(\mathcal{S}_k, \tilde{d}_k \restriction_{\mathcal{S}_k}, \rho_k, \tilde{\mu}_k \restriction_{\mathcal{S}_k})$ in $\widetilde{\mathcal{T}}_k$ is replaced by $(\mathcal{S}'_k, \tilde{d}'_{\mathcal{S}'_k}, \rho_k, \tilde{\mu}'_{\mathcal{S}'_k})$ in order to obtain $(\widetilde{\mathcal{T}}_{k+1}, \tilde{d}_{k+1}, \rho, \tilde{\mu}_{k+1})$, and

$$\left(\mu(\mathcal{S}_k)^{-\alpha}\mathcal{S}'_k, \tilde{d}'_{\mathcal{S}'_k}, \rho_k, \mu(\mathcal{S}_k)^{-1}\tilde{\mu}'_{\mathcal{S}'_k}\right) \overset{d}{=} \left(\widetilde{\mathcal{T}}_1, \tilde{d}_1, \rho, \tilde{\mu}_1\right).$$

Since the tree $(\widetilde{\mathcal{T}}_k, \tilde{d}_k, \rho, \tilde{\mu}_k)$ is such that $\mathcal{R}(\widetilde{\mathcal{T}}_k, \Sigma_1, \ldots, \Sigma_k)$ has the same distribution as $\mathcal{R}_k^{\alpha,2-\alpha}$ by the induction hypothesis, and since, by construction, the projection of $\mathcal{R}(\widetilde{\mathcal{T}}_{k+1}, \Sigma_1, \ldots, \Sigma_{k+1})$ onto $\mathcal{R}(\widetilde{\mathcal{T}}_{k+1}, \Sigma_1, \ldots, \Sigma_k)$ is $\mathcal{R}(\widetilde{\mathcal{T}}_k, \Sigma_1, \ldots, \Sigma_k)$, the claim follows. □

Remark 3.53. Note that by the refining nature of the branch merging algorithm,

$$d_{\mathrm{GH}}\left(\widetilde{\mathcal{T}}_{k+1}, \mathcal{R}\left(\widetilde{\mathcal{T}}_{k+1}, \Sigma_1, \ldots, \Sigma_{k+1}\right)\right) \leq d_{\mathrm{GH}}\left(\widetilde{\mathcal{T}}_k, \mathcal{R}\left(\widetilde{\mathcal{T}}_k, \Sigma_1, \ldots, \Sigma_k\right)\right)$$

for all $k \geq 1$. By Corollary 3.52, the increasing limit

$$\widetilde{\mathcal{T}}_\infty = \overline{\bigcup_{k\geq 1} \mathcal{R}\left(\widetilde{\mathcal{T}}_k, \Sigma_1, \ldots, \Sigma_k\right)}$$

is a random compact $\mathbb{R}$-tree distributed as a Ford CRT (since this holds for the increasing sequence $(\mathcal{R}_k^{\alpha,2-\alpha}, k \geq 1)$). By construction,

$$\mathcal{R}\left(\widetilde{\mathcal{T}}_\infty, \Sigma_1, \ldots, \Sigma_k\right) = \mathcal{R}\left(\widetilde{\mathcal{T}}_k, \Sigma_1, \ldots, \Sigma_k\right). \tag{3.42}$$

Hence, $\widetilde{\mathcal{T}}_\infty$ is a Ford CRT with all leaves $\Sigma_k, k \geq 1$, embedded. By the nature of the algorithm, (3.42) can be enriched by the corresponding point processes, i.e.

$$\left(\mathcal{R}\left(\widetilde{\mathcal{T}}_\infty, \Sigma_1, \ldots, \Sigma_k\right), \mathcal{P}_{\mathcal{R}\left(\widetilde{\mathcal{T}}_\infty, \Sigma_1, \ldots, \Sigma_k\right)}\right) \stackrel{d}{=} \left(\mathcal{R}\left(\widetilde{\mathcal{T}}_k, \Sigma_1, \ldots, \Sigma_k\right), \mathcal{P}_{\mathcal{R}\left(\widetilde{\mathcal{T}}_k, \Sigma_1, \ldots, \Sigma_k\right)}\right).$$

Based on a generalisation of Miermont's notion of k-marked trees [54] to ∞-marked trees, it can be shown that $\widetilde{\mathcal{T}}_k \to \widetilde{\mathcal{T}}_\infty$ a.s. as $k \to \infty$ with all leaves $\Sigma_k, k \geq 1$, embedded in $\widetilde{\mathcal{T}}_\infty$, see Chapter 2.

The Brownian case: leaf embedding for $\alpha = 1/2$, $\theta = 3/2$. In the following, $(\mathcal{T}, d, \rho, \mu)$ denotes the Brownian CRT. By Corollary 3.34, for $\alpha = 1/2$, i.e. $1-\alpha = 1/2$, $2-\alpha = 3/2$, we have $\mathcal{T} \stackrel{d}{=} \mathcal{T}^{\alpha, 1-\alpha} \stackrel{d}{=} \mathcal{T}^{\alpha, 2-\alpha}$.

In the Brownian case, by Lemma 3.30, leaf labelling in the $(1/2, 1/2)$-tree growth model is exchangeable, and hence, leaf embedding in Algorithm 3.44 simplifies to uniform sampling from the mass measure. For any continuum tree $(\mathcal{T}, d, \rho, \mu)$ (and in particular the Brownian CRT), let $(\mathrm{i})^{\mathrm{B}}$ and $(\mathrm{ii})^{\mathrm{B}}$ be defined as follows.

$(\mathrm{i})^{\mathrm{B}}$ *Start configuration.* Pick three leaves $\Omega_0, \Omega_1, \Omega_2$ independently by uniform sampling from the mass measure μ on $\mathcal{T}$, and denote by Ω the first branch point of the spines from the root ρ to $\Omega_0, \Omega_1, \Omega_2$. We obtain the three branches $\widetilde{E}_0 :=]]\Omega, \Omega_0]], \widetilde{E}_1 :=]]\Omega, \Omega_1]], \widetilde{E}_2 :=]]\Omega, \Omega_2]]$. Define the start configuration $(\Sigma_1, \widetilde{\Sigma}_1, \Theta)$ by

$$\left(\Sigma_1, \widetilde{\Sigma}_1, \Theta\right) := \begin{cases} (\Omega_2, \Omega_0, \Omega_1) & \text{if } \widetilde{E}_0 \cap \widetilde{E}_1 \neq \emptyset, \\ (\Omega_1, \Omega_0, \Omega_2) & \text{if } \widetilde{E}_0 \cap \widetilde{E}_2 \neq \emptyset, \\ (\Omega_0, \Omega_1, \Omega_2) & \text{if } \widetilde{E}_1 \cap \widetilde{E}_2 \neq \emptyset. \end{cases}$$

Write $\mathcal{R}_2 = \mathcal{R}(\mathcal{T}, \Sigma_1, \widetilde{\Sigma}_1)$, $\rho_\Theta := \pi^{\mathcal{R}_2}(\Theta)$ and $\mathcal{S}_x := (\pi^{\mathcal{R}_2})^{-1}(x)$ for any $x \in \mathcal{R}_2$. We obtain a split of $\mathcal{R}_2$ into the 4 branches

$$E_0 :=]]\rho, \Omega[[, \quad E_1 :=]]\Omega, \Sigma_1[[, \quad E_2 :=]]\Omega, \rho_\Theta[[, \quad E_3 :=]]\rho_\Theta, \widetilde{\Sigma}_1[[.$$

(ii)$^{\mathrm{B}}$ *Partition sampling.* Consider a sequence of i.i.d. random variables $(Y_k)_{k\geq 3} \sim \mu_{\mathcal{R}_2}$, and define

$$\Pi_{\rho_\Theta} := \{k \geq 3 : Y_k = \rho_\Theta\} \cup \{2\}, \qquad \Pi_x := \{k \geq 3 : Y_k = x\}, x \in \mathcal{R}_2 \setminus \{\rho_\Theta\}.$$

Based on the properties of the Brownian CRT as described above, we conclude that (i) and (ii) in Algorithm 3.14 simplify to (i)$^{\mathrm{B}}$ and (ii)$^{\mathrm{B}}$ in the Brownian case. Algorithm 3.44 gives a way to construct a Brownian CRT with an $(1/2, 3/2)$-string of beads, and in particular leaf 1 of the $(1/2, 3/2)$-tree growth process, embedded. By Corollary 3.52 and Corollary 3.34 the Brownian CRT is distributionally invariant under the simplified branch merging algorithm with (i)$^{\mathrm{B}}$ and (ii)$^{\mathrm{B}}$.

Corollary 3.54. *Let $(\mathcal{T}, d, \rho, \mu)$ be the Brownian CRT. The sequence of trees*

$$\left(\left(\widetilde{\mathcal{T}}_k, \tilde{d}_k, \rho, \tilde{\mu}_k\right), k \geq 1\right)$$

constructed recursively according to Algorithm 3.44 with steps $(\mathrm{i})^{\mathrm{B}}, (\mathrm{ii})^{\mathrm{B}}, (\mathrm{iii}), (\mathrm{iv})$ *and the procedure described in Section 3.4.3.2, and the sequence of leaves $(\Sigma_k, k \geq 1)$ are such that*

$$\left(\mathcal{R}\left(\widetilde{\mathcal{T}}_k, \Sigma_1, \ldots, \Sigma_k\right), k \geq 1\right) \stackrel{d}{=} \left(\mathcal{R}_k^{1/2,3/2}, k \geq 1\right),$$

where $\mathcal{R}_k^{1/2,3/2}$ is the almost-sure limit as $n \to \infty$ of the trees $\mathcal{R}(T_n^{1/2,3/2}, [k])$ associated with the $(1/2, 3/2)$-tree growth process $(T_n^{1/2,3/2}, n \geq 1)$ as in Proposition 3.31. Furthermore, for any $k \geq 1$, $(\widetilde{\mathcal{T}}_k, \tilde{d}_k, \rho, \tilde{\mu}_k) \stackrel{d}{=} (\mathcal{T}, d, \rho, \mu)$, i.e. the Brownian CRT is distributionally invariant under the simplified branch merging algorithm.

Bibliography

[1] Abraham, R., and Delmas, J.-F. A continuum tree-valued Markov process. *The Annals of Probability 40*, 3 (2012), 1167–1211.

[2] Abraham, R., Delmas, J.-F., and He, H. Pruning Galton–Watson trees and tree-valued Markov processes. *Annales de l'Institut Henri Poincaré, Probabilités et Statistiques 48*, 3 (2012), 688–705.

[3] Abraham, R., Delmas, J.-F., and He, H. Pruning of CRT sub-trees. *Stochastic Processes and their Applications 125*, 4 (2015), 1569–1604.

[4] Addario-Berry, L., Broutin, N., and Goldschmidt, C. Critical random graphs: Limiting constructions and distributional properties. *Electronic Journal of Probability 15*, 25 (2010), 741–775.

[5] Addario-Berry, L., Broutin, N., and Holmgren, C. Cutting down trees with a Markov chainsaw. *The Annals of Applied Probability 24*, 6 (2014), 2297–2339.

[6] Addario-Berry, L., Dieuleveut, D., and Goldschmidt, C. Inverting the cut-tree transform. http://arxiv.org/abs/1606.04825, 2016.

[7] Aldous, D. The Continuum Random Tree, I. *The Annals of Probability 19* (1991), 1–28.

[8] Aldous, D. The Continuum Random Tree, II. An overview. In *Stochastic Analysis.* Cambridge University Press, 1991.

[9] Aldous, D. The Continuum Random Tree, III. *The Annals of Probability 21* (1993), 248–289.

[10] Aldous, D. Tree-based models for random distribution of mass. *Journal of Statistical Physics 73*, 3-4 (1993), 625–641.

[11] Aldous, D. Wright-Fisher diffusion with negative mutation rate. https://www.stat.berkeley.edu/~aldous/Research/OP/fw.html, 1999.

[12] Aldous, D. Mixing time for a Markov chain on cladograms. *Combinatorics, Probability and Computing 9* (2000), 191–204.

[13] Aldous, D., and Bandyopadhyay, A. A survey of max-type recursive distributional equations. *Annals of Applied Probability 15*, 2 (2005), 1047–1110.

[14] Amini, O., Devroye, L., Griffiths, S., and Olver, N. Explosion and linear transit times in infinite trees. *Probability Theory and Related Fields* (2015), 1–23.

[15] Bertoin, J. Self-similar fragmentations. *Annales de l'Institut Henri Poincaré - Probabilités et Statistiques 38* (2002), 319–340.

[16] Bertoin, J., and Miermont, G. The cut-tree of large Galton-Watson trees and the Brownian CRT. *The Annals of Applied Probability 23*, 4 (2013), 1469–1493.

[17] Bolley, F. Separability and completeness for the Wasserstein distance. *Séminaire de probabilités XLI, Lecture Notes in Mathematics 1934* (2008), 371–377.

[18] Broutin, N., and Wang, M. Cutting down p-trees and inhomogeneous continuum random trees. http://arxiv.org/abs/1408.0144, 2014.

[19] Broutin, N., and Wang, M. Reversing the cut tree of the Brownian continuum random tree. http://arxiv.org/abs/1408.2924, 2014.

[20] Burago, D., Burago, Y., and Ivanov, S. *A course in metric geometry.* American Mathematical Society, Providence, RI, 2001.

[21] Chauvin, B., Flajolet, P., Gardy, D., and Mokkadem, A., Eds. *Mathematics and Computer Science II: Algorithms, Trees, Combinatorics and Probabilities*, 1st ed. Trends in Mathematics. Birkhäuser, 2002.

[22] Chen, B., Ford, D., and Winkel, M. A new family of Markov branching trees: The alpha-gamma model. *Electronic Journal of Probability 14*, 15 (2009), 400–430.

[23] Curien, N., and Haas, B. The stable trees are nested. *Probability Theory and Related Fields 157* (2013), 847–883.

[24] Dong, R., Goldschmidt, C., and Martin, J. Coagulation-fragmentation duality, Poisson-Dirichlet distributions and random recursive trees. *The Annals of Applied Probability 16*, 4 (2006), 1773–1750.

[25] Drmota, M. *Random trees: An interplay between Combinatorics and Probability*, 1st ed. Springer, 2009.

[26] Drmota, M., Flajolet, P., Gardy, D., and Gittenberger, B., Eds. *Mathematics and Computer Science III: Algorithms, Trees, Combinatorics and Probabilities*, 1st ed. Trends in Mathematics. Birkhäuser, 2004.

[27] Duquesne, T. A limit theorem for the contour process of conditioned Galton-Watson trees. *The Annals of Probability 31*, 2 (2003), 996–1027.

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

[29] Duquesne, T., and Le Gall, J.-F. Probabilistic and fractal aspects of Lévy trees. *Probability Theory and Related Fields 131*, 4 (2005).

[30] Etheridge, A. *Some mathematical models from population genetics.* École d'Été de Probabilités de Saint-Flour XXXIX-2009, Lecture Notes in Mathematics. Springer, 2012.

[31] Evans, S. *Probability and real trees.* Ecole d'Été de Probabilités de Saint-Flour XXXV-2005. Lecture Notes in Mathematics. Springer, 2005.

[32] Evans, S., Pitman, J., and Winter, A. Rayleigh Processes, Real Trees, and Root Growth with Re-grafting. *Probability Theory and Related Fields 134* (2006), 81–126.

[33] Evans, S. N., and Winter, A. Subtree prune and regraft: A reversible real tree-valued Markov process. *The Annals of Probability 34* (2006), 918–961.

[34] Feng, S. *The Poisson-Dirichlet Distribution and Related Topics*, 1st ed. Springer, 2010.

[35] Ford, D. *Probabilities on cladograms: Introduction to the alpha model.* Stanford University, 2005.

[36] Gardy, D., and Mokkadem, A., Eds. *Mathematics and Computer Science: Algorithms, Trees, Combinatorics and Probabilities*, 1st ed. Trends in Mathematics. Birkhäuser, 2000.

[37] Gnedin, A. Regeneration in random combinatorial structures. *Probability Surveys 7* (2010), 105–156.

[38] Gnedin, A., and Pitman, J. Regenerative composition structures. *The Annals of Probability 33*, 2 (2005), 445–479.

[39] Goldschmidt, C., and Haas, B. A line-breaking construction of the stable trees. *Electronic Journal of Probability 20*, 16 (2015), 1–24.

[40] Haas, B., and Miermont, G. The genealogy of self-similar fragmentations with negative index as a continuum random tree. *Electronic Journal of Probability 9* (2004), 57–97.

[41] Haas, B., and Miermont, G. Scaling limits of Markov branching trees with applications to Galton-Watson and random unordered trees. *The Annals of Probability 40*, 6 (2012), 2589–2666.

[42] Haas, B., Miermont, G., Pitman, J., and Winkel, M. Continuum tree asymptotics of discrete fragmentations and applications to phylogenetic models. *The Annals of Probability 36*, 5 (2008), 1790–1837.

[43] Haas, B., Pitman, J., and Winkel, M. Spinal partitions and invariance under re-rooting of continuum random trees. *The Annals of Probability 37*, 4 (2009), 1381–1411.

[44] James, L. F. Generalized Mittag Leffler distributions arising as limits in preferential attachment models. http://arxiv.org/pdf/1509.07150.pdf, 2015.

[45] Kallenberg, O. *Foundations of Modern Probability.* Springer, 2001.

[46] Le Gall, J.-F. Random trees and applications. *Probability Surveys 2* (2005), 245–311.

[47] Le Gall, J.-F. Random real trees. *Annales de la faculté des sciences de Toulouse Mathématiques 41*, 1 (2006), 35–62.

[48] Le Gall, J.-F., and Le Jan, Y. Branching processes in Lévy processes: the exploration process. *The Annals of Probability 26*, 1 (1998), 213–252.

[49] Maisonneuve, B. Ensembles régénératifs de la droite. *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete 63* (1983), 501–510.

[50] Marchal, P. A note on the fragmentation of a stable tree. In *Fifth Colloquium on Mathematics and Computer Science*, vol. AI. Discrete Mathematics and Theoretical Computer Science, 2008, pp. 489–500.

[51] McColm, G. L. An introduction to random trees. *Research on Language and Communication 1* (2003), 203–226.

[52] Miermont, G. Self-similar fragmentations derived from the stable tree I: splitting at heights. *Probability Theory and Related Fields 127*, 3 (2003), 423–454.

[53] Miermont, G. Self-similar fragmentations derived from the stable tree II: splitting at nodes. *Probability Theory and Related Fields 131*, 3 (2005), 341–375.

[54] Miermont, G. Tessellations of Random Maps of Arbitrary Genus. *Annales Sciéntifique de l'École Normale Supérieure 42*, 5 (2009), 725–781.

[55] Pal, S. On the Aldous diffusion on continuum trees. I. http://arxiv.org/pdf/1104.4186v1.pdf, 2011.

[56] Pal, S. Wright-Fisher diffusion with negative mutation rates. *The Annals of Probability 41*, 2 (2013), 503–526.

[57] Pitman, J. Partition structures derived from Brownian motion and stable subordinators. *Bernoulli 3* (1997), 79–96.

[58] Pitman, J. Combinatorial stochastic processes: École d'Été de Probabilités de Saint-Flour XXXII-2002. In *Lecture Notes in Mathematics*, vol. 1875. Springer, 2006.

[59] Pitman, J., Rizzolo, D., and Winkel, M. Regenerative tree growth: Structural results and convergence. *The Annals of Probability* (2008).

[60] Pitman, J., and Winkel, M. Regenerative tree growth: Binary self-similar continuum random trees and Poisson-Dirichlet compositions. *The Annals of Probability 37*, 5 (2009), 1999–2041.

[61] Pitman, J., and Winkel, M. Regenerative tree growth: Markovian embedding of fragmenters, bifurcators, and bead splitting processes. *The Annals of Probability 43*, 5 (2015), 2611–2646.

[62] Pitman, J., and Yor, M. The two-parameter Poisson-Dirichlet distribution derived from a stable subordinator. *The Annals of Probability 25*, 2 (1997), 855–900.

[63] Rannala, B., and Yang, Z. Molecular phylogenetics: principles and practice. *Nature Reviews Genetics 13* (2012), 303–314.

[64] Rembart, F. Branch merging on continuum trees with applications to regenerative tree growth. *Latin American Journal of Probability and Mathematical Statistics XIII*, 563-603 (2016).

[65] Rembart, F., and Winkel, M. Recursive construction of CRTs and a binary embedding of the stable line-breaking construction. *Work in progress*, 2016.

[66] Stephenson, R. General fragmentation trees. *Electronic Journal of Probability 18*, 101 (2013), 1–45.

[67] Zhang, X. A very gentle note on the construction of Dirichlet process. http://webdocs.cs.ualberta.ca/~xinhua2/pubDoc/notes/dirichlet_process.pdf, 2008.