Available online at www.sciencedirect.com
ScienceDirect
Stochastic Processes and their Applications 163 (2023) 424–472
stochastic processes and their applications
www.elsevier.com/locate/spa

A binary embedding of the stable line-breaking construction

Franz Rembart[a], Matthias Winkel[b,*]

[a] *Department of Statistics, Columbia University, 1255 Amsterdam Avenue, New York, NY 10027, USA*
[b] *Department of Statistics, University of Oxford, 24–29 St Giles, Oxford OX1 3LB, United Kingdom*

Received 6 July 2022; received in revised form 22 January 2023; accepted 12 June 2023
Available online 16 June 2023

Abstract

We embed Duquesne and Le Gall's stable tree into a binary compact continuum random tree (CRT) in a way that solves an open problem posed by Goldschmidt and Haas. This CRT can be obtained by applying a recursive construction method of compact CRTs as presented in earlier work to a specific distribution of a random string of beads, i.e. a random interval equipped with a random discrete measure. We also express this CRT as a tree built by replacing all branch points of a stable tree by i.i.d. copies of a Ford CRT, each rescaled by a factor intrinsic to the stable CRT.

Keywords: Stable tree; Line-breaking construction; String of beads; Continuum random tree; Marked metric space; Recursive distribution equation

1. Introduction

1.1. Motivation and main results

Stable trees were introduced by Duquesne and Le Gall [14] as a family of continuum random trees (CRTs) parametrised by a self-similarity parameter $\alpha \in (1, 2]$ to describe the genealogical structure of continuous-state branching processes with branching mechanism $\lambda \mapsto \lambda^\alpha$. As such they form a subclass of Lévy trees [32] and contain Aldous's Brownian CRT [1–3] as a special case ($\alpha = 2$). They were studied by Miermont and others [14,15,22,24,25,32,34,35] in the

* Corresponding author.
E-mail addresses: fhr2111@columbia.edu (F. Rembart), winkel@stats.ox.ac.uk (M. Winkel).

https://doi.org/10.1016/j.spa.2023.06.007

context of self-similar fragmentations and by several authors to establish invariance principles [6,11,13,23,29] and other properties [8,10]. Furthermore, they have deeper connections to random maps and Liouville quantum gravity [12,31,37].

We represent trees as $\mathbb{R}$-*trees*, i.e. compact metric spaces $(\mathcal{T}, d)$ such that any two points $x, y \in \mathcal{T}$ are connected by a unique path $[[x, y]]$ in $\mathcal{T}$, which is furthermore required to have length $d(x, y)$. All our $\mathbb{R}$-trees are *rooted* at a distinguished $\rho \in \mathcal{T}$. We refer to a rooted $\mathbb{R}$-tree $(\mathcal{T}, d, \rho)$ equipped with a probability measure μ as a *weighted* $\mathbb{R}$-tree $(\mathcal{T}, d, \rho, \mu)$, and equip sets of isometry classes of $\mathbb{R}$-trees and weighted $\mathbb{R}$-trees with the Gromov–Hausdorff and the Gromov–Hausdorff–Prokhorov topology, respectively.

Ever since Aldous [3], such trees have been built sequentially from a single branch $[[\rho, \Sigma_0]]$, grafting further branches (line segments) $]]J_{k-1}, \Sigma_k]]$ to build trees $\mathcal{T}_k$ spanned by a growing number of points $\rho, \Sigma_0, \ldots, \Sigma_k$, $k \geq 1$, finally passing to the closure/completion $\mathcal{T}$ of $\bigcup_{k\geq 0} \mathcal{T}_k$. In a given weighted $\mathbb{R}$-tree $(\mathcal{T}, d, \rho, \mu)$, a natural sequence $(\Sigma_k, k \geq 0)$ may be obtained as an independent sample from μ. For the Brownian CRT, Aldous [3] gave an autonomous description of the resulting tree-growth process $(\mathcal{T}_k, k \geq 0)$ by breaking the half-line $[0, \infty)$ at the points $(S_k, k \geq 0)$ of an inhomogeneous Poisson process with linearly growing intensity $t dt$ on $[0, \infty)$, each segment $]S_k, S_{k+1}]$ grafted in a point J_k chosen uniformly from the length measure on the structure $\mathcal{T}_k$ already built, with $\mathcal{T}_0 = [0, S_0]$.

In Aldous's construction, the *branch points* J_k, $k \geq 0$, are distinct, the trees *binary*. This construction reveals some of the local complexity of the limiting tree, since elementary thinning of Poisson processes shows that every branch receives a dense set of branch points. Goldschmidt and Haas [20] generalised this line-breaking construction to all stable trees $(\mathcal{T}, d, \rho, \mu)$, which are not binary for $\alpha \in (1, 2)$. They describe

$$\mathcal{T}_k = \bigcup_{i=0}^{k} [[\rho, \Sigma_i]], \quad k \geq 0, \qquad \text{for a sample } \Sigma_i \sim \mu, i \geq 0, \tag{1.1}$$

not quite autonomously, as Aldous does in the special case $\alpha = 2$, but by assigning weights

$$W_k^{(i)}, \quad i \in [b_k], \quad k \geq 0, \tag{1.2}$$

to each branch point v_i of $\mathcal{T}_k$, where $(v_i, i \geq 1)$ is the sequence of *distinct branch points* in their order of appearance in $(\mathcal{T}_k, k \geq 0)$, and $b_k \geq 0$ is the number of branch points of $\mathcal{T}_k$. Here, $[b] := \{1, \ldots, b\}$ with the convention that $[0] = \emptyset$.

Specifically, it will be convenient to change the usual parametrisation of the stable trees from a parameter $\alpha \in (1, 2]$ to an *index* $\beta = 1 - 1/\alpha \in (0, 1/2]$. For $k \geq 0$, the sum of the branch point weights $(W_k^{(i)}, i \in [b_k])$ 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* [20,24,27] with parameter β, starting from $S_0 \sim \mathrm{ML}(\beta, \beta)$ with transition density

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

Here, $\mathrm{ML}(\alpha, \theta)$ denotes the Mittag-Leffler distribution with parameters $0 < \alpha < 1$ and $\theta > -\alpha$ (cf. Section 2.3), and $g_\beta(\cdot)$ is the density of $\mathrm{ML}(\beta, 0)$. Then $S_k = W_k^{(1)} + \cdots + W_k^{(b_k)} + L_k \sim \mathrm{ML}(\beta, \beta + k)$.

Algorithm 1.1 (*Goldschmidt–Haas [20], Line-Breaking Construction II*)**.** Let $\beta \in (0, 1/2]$. We grow discrete random $\mathbb{R}$-trees $\mathcal{T}_k$ with weights $W_k^{(i)}$ in the branch points v_i, $i \in [b_k]$, of $\mathcal{T}_k$, and edge lengths between vertices, 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 of $\mathcal{T}_k$ 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 $\mathcal{T}_{k+1}$; increase the weight of $J_k = v_{I_k}$ to $W_{k+1}^{(I_k)} = W_k^{(I_k)} + (S_{k+1} - S_k)(1 - B_k)$, and set $W_{k+1}^{(j)} = W_k^{(j)}$, $j \neq I_k$.

When $\beta = 1/2$, we understand $B_k = 1$, so $W_k^{(i)} = 0$ for all $i \geq 1$, $k \geq 0$, and $L_k = S_k$ for all $k \geq 0$. We obtain a sequence of compact binary $\mathbb{R}$-trees whose evolution is determined by attachment points chosen uniformly at random according to the length measure, and the total length given by the Mittag-Leffler Markov chain of parameter $\beta = 1/2$, which can be seen to correspond to an inhomogeneous Poisson process of rate $\frac{1}{2}t\,dt$. Hence, this reduces to Aldous's line-breaking construction of the Brownian CRT [3].

It was shown in [20] that the sequence of trees $(\mathcal{T}_k, k \geq 0)$ in Algorithm 1.1 has the same distribution as the sequence of trees from (1.1), i.e. we can formally define *the stable tree of index* $\beta \in (0, 1/2]$ as the (Gromov–Hausdorff) limit $\mathcal{T}$ of $\mathcal{T}_k$, as $k \to \infty$. See also [20] for an alternative line-breaking construction of the sequence $(\mathcal{T}_k, k \geq 0)$, where branch point selection is based on vertex degrees instead of weights.

Goldschmidt and Haas [20] asked if there was a sensible way to associate a notion of length with the branch point weights in Algorithm 1.1. 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 [18,24] and in the context of the alpha–gamma model [7] 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 and have also been constructed via line-breaking:

Algorithm 1.2 (*Haas–Miermont–Pitman–Winkel [24,39]*)**.** Let $\beta' \in (0, 1)$. We grow random $\mathbb{R}$-trees $\mathcal{F}_m$, $m \geq 1$:

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 $1 - D_m$; otherwise, sample J_m from the normalised length measure on E_m;
3. sample S_{m+1}' with density $f(S_m', \cdot)$; 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 $(\mathcal{F}_m, m \geq 1)$ has as its (Gromov–Hausdorff) limit a CRT $\mathcal{F}$ as $k \to \infty$, a so-called *Ford CRT* of index $\beta' \in (0, 1)$, see [24,39]. We refer to the trees $\mathcal{F}_m$, $m \geq 1$, as *Ford trees*. In the case when $\beta' = 1/2$, Algorithm 1.2 corresponds to Aldous's construction of the Brownian CRT.

We combine the line-breaking constructions of Algorithm 1.1 and Algorithm 1.2 in the framework of ∞-marked $\mathbb{R}$-trees, which we introduce in Section 2.2 as a natural extension of Miermont's notion of k-marked trees [36]. An ∞-marked $\mathbb{R}$-tree $(\mathcal{T}, (\mathcal{R}^{(i)}, i \geq 1))$ is an $\mathbb{R}$-tree $(\mathcal{T}, d, \rho)$ with non-empty closed 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* set $\bigcup_{i\geq 1} \mathcal{R}^{(i)}$ 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 1.1 and the unmarked remainder gives rise to a stable tree. Selection of a branch point in Algorithm 1.1 corresponds to an insertion into the respective marked component in the enhanced line-breaking construction given by Algorithm 1.3.

Algorithm 1.3 (*Two-Colour Line-Breaking Construction*)**.** Let $\beta \in (0, 1/2]$. We grow random ∞-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 $r_k = \#\{i \geq 1 : \mathcal{R}_k^{(i)} \neq \{\rho\}\}$;

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 $1 - D_k$, with proportion D_k closer to the root; otherwise, i.e. if $E_k^* \subset \mathcal{T}_k^* \setminus \bigcup_{j\in[r_k]} \mathcal{R}_k^{(j)}$ 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 1.1 and the lengths of the marked subtrees in Algorithm 1.3, as well as marked subtrees as in Algorithm 1.2, up to scaling:

Theorem 1.4 (*Weight-Length Representation*)**.** *Let $(\mathcal{T}_k, (W_k^{(i)}, i \geq 1), k \geq 0)$ be as in Algorithm 1.1. Consider a sequence $(\mathcal{T}_k^*, (\mathcal{R}_k^{(i)}, i \geq 1), k \geq 0)$ of ∞-marked $\mathbb{R}$-trees constructed as in Algorithm 1.3, 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 $\widetilde{v}_i$ by using an equivalence relation, and denote the resulting quotient space by $\widetilde{\mathcal{T}}_k$. Then*

$$\left(\widetilde{\mathcal{T}}_k, \left(\widetilde{W}_k^{(i)}, i \geq 1\right), k \geq 0\right) \stackrel{d}{=} \left(\mathcal{T}_k, \left(W_k^{(i)}, i \geq 1\right), k \geq 0\right). \tag{1.3}$$

See Fig. 1. Moreover, there exist positive random variables $C^{(i)}$ and subsequences $(k_m^{(i)}, m \geq 1)$, $i \geq 1$, such that the rescaled marked subtrees grow like Ford trees of index $\beta' = \beta/(1 - \beta)$, i.e.

$$\left(C^{(i)}\mathcal{R}^{(i)}_{k_m^{(i)}}, m \geq 1\right) \stackrel{d}{=} (\mathcal{F}_m, m \geq 1), \tag{1.4}$$

for all $i \geq 1$ where $(C^{(i)}\mathcal{R}^{(i)}_{k_m^{(i)}}, m \geq 1)$, $i \geq 1$, are independent of each other.

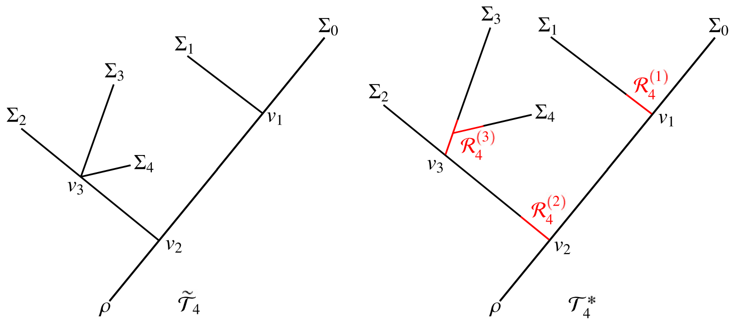

Fig. 1. Example of $\widetilde{\mathcal{T}}_4$ with three branch points v_1, v_2, v_3; branch point weights $W_4^{(1)}, W_4^{(2)}, W_4^{(3)}$ are represented as lengths $\widetilde{W}_4^{(1)}, \widetilde{W}_4^{(2)}, \widetilde{W}_4^{(3)}$ of marked subtrees $\mathcal{R}_4^{(1)}, \mathcal{R}_4^{(2)}, \mathcal{R}_4^{(3)}$ in $\mathcal{T}_4^*$, respectively.

To obtain limiting ∞-marked CRTs, we build on [36] to define a suitable metric d_{GH}^{∞} in Section 2.2.

Theorem 1.5 (*Convergence of Two-Colour Trees*). *Let $(\mathcal{T}_k^*, (\mathcal{R}_k^{(i)}, i \geq 1), k \geq 0)$ be as above. Then*

$$\lim_{k\to\infty} \left(\mathcal{T}_k^*, \left(\mathcal{R}_k^{(i)}, i \geq 1\right)\right) = \left(\mathcal{T}^*, \left(\mathcal{R}^{(i)}, i \geq 1\right)\right) \quad a.s. \tag{1.5}$$

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 $\widetilde{v}_i$, is a stable tree of parameter β;*
- *there exist $\widetilde{\mathcal{T}}$-measurable scaling factors $(C^{(i)}, i \geq 1)$ such that $C^{(i)}\mathcal{R}^{(i)}$, $i \geq 1$, are i.i.d. copies of a Ford CRT $\mathcal{F}$ of index $\beta' = \beta/(1-\beta)$, and the trees $C^{(i)}\mathcal{R}^{(i)}$, $i \geq 1$, are independent of $\widetilde{\mathcal{T}}$.*

The scaling factors $C^{(i)}$ can be given explicitly in terms of the masses of the subtrees of the stable tree $\widetilde{\mathcal{T}}$ above the branch point $\widetilde{v}_i$. We can in fact use this, with the ingredients listed in Theorem 1.5, to construct the two-colour tree $(\mathcal{T}^*, (\mathcal{R}^{(i)}, i \geq 1))$ from a stable tree $(\mathcal{T}, \mu)$ by replacing each branch point by a rescaled independent copy of a Ford CRT:

Theorem 1.6 (*Branch Point Replacement in a Stable Tree*). *Let $(\mathcal{T}, d, \rho, \mu)$ be a stable tree of index $\beta \in (0, 1/2]$ equipped with an i.i.d. sequence of labelled leaves $(\Sigma_k, k \geq 0)$ sampled from μ. Consider the reduced trees $(\mathcal{T}_k, k \geq 0)$ as in (1.1) with branch points $(v_i, i \geq 1)$ in order of appearance. For each $i \geq 1$, consider the path from the root to the leaf with the smallest label above v_i and the following variables:*

- *the total mass $P^{(i)} = \sum_{j\geq 1} P_j^{(i)}$ of the subtrees rooted at v_i on this path with masses $(P_j^{(i)}, j \geq 1)$, in the order of their smallest labels;*
- *the random variable $D^{(i)} = \lim_{n\to\infty}\left(1 - \sum_{j\in[n]} P_j^{(i)}/P^{(i)}\right)^{1-\beta}(1-\beta)^{\beta-1}n^{\beta}$ derived from $(P_j^{(i)}, j \geq 1)$.*

For $i \geq 1$, replace v_i by an independent Ford tree $\mathcal{F}^{(i)}$ of index $\beta' = \beta/(1-\beta)$ with distances rescaled by $\left(C^{(i)}\right)^{-1} = \left(P^{(i)}\right)^{\beta}\left(D^{(i)}\right)^{\beta/(1-\beta)} = \lim_{n\to\infty}\left(P^{(i)} - \sum_{j\in[n]} P_j^{(i)}\right)^{\beta}(1-\beta)^{-\beta}n^{\beta^2/(1-\beta)}$.

Specifically, the root of $\mathcal{F}^{(i)}$ is identified with v_i and the subtrees rooted at v_i are attached to leaves of $\mathcal{F}^{(i)}$ in the order of their appearance in Algorithm 1.2. Then the tree $\mathcal{T}^*$ obtained here in the limit after all replacements has the same distribution as the tree $\mathcal{T}^*$ in Theorem 1.5.

We will formalise this construction in Section 5.3. The random variable $D^{(i)}$ is the so-called $(1-\beta)$-*diversity* of the mass partition $(P_j^{(i)}/P^{(i)}, j \geq 1) \sim \mathrm{GEM}(1-\beta, -\beta)$, where $\mathrm{GEM}(\alpha, \theta)$ denotes the Griffiths–Engen–McCloskey distribution with parameters $\alpha \in [0, 1)$, $\theta > -\alpha$, whose ranked version is the Poisson–Dirichlet distribution $\mathrm{PD}(\alpha, \theta)$. Note that, when $\beta = 1/3$, we have $\beta' = 1/2$, which means that we replace the branch points of the stable tree by rescaled i.i.d. Brownian CRTs. This should be compared with Le Gall [31], who effectively contracts subtrees in the middle of a Brownian CRT to obtain a stable tree of parameter 3/2. Neither his subtrees nor our $\mathcal{T}^*$ appear to be rescaled Brownian CRTs.

1.2. Main ideas for the proofs and further results

The proofs of our main results, Theorems 1.5 and 1.6, 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 earlier work [42] where we constructed CRTs via i.i.d. copies of a random string of beads, see Section 3.3 for details. Specifically, in [42], we introduced a recursive method to construct a CRT via *bead splitting*, i.e., from independent copies of random strings of beads, which are intervals equipped with discrete mass measures. The algorithm starts with a single string of beads. In the first step, each bead is replaced by a rescaled independent copy of a string of beads, whose distribution is fixed throughout the algorithm. The method is applied recursively, i.e., the beads of the tree obtained after n steps are replaced by rescaled independent copies of the string of beads. We have shown the convergence of the resulting trees to a (compact) CRT. To apply this setup in the context of this paper, we need to find the 'right' string of beads. More precisely, the string of beads will have two components: one scaled marked string of beads that leads to rescaled Ford trees, and one unmarked component that leads to a stable tree after contracting all marked components to single branch points.

Mathematically, the distribution ν of the string of beads we use to obtain this CRT combines two (β, θ)-strings of beads (for $\theta = \beta$ and $\theta = 1 - 2\beta$), which arise in the framework of ordered (β, θ)-Chinese restaurant processes as introduced in [39]. A (β, θ)-*string of beads* is an interval of length $K \sim \mathrm{ML}(\beta, \theta)$ equipped with a discrete probability measure whose atom sizes are $\mathrm{PD}(\beta, \theta)$, arranged in a random order that yields a regenerative property. It is crucial for our argument to equip each reduced tree with a measure, which effectively captures projected subtree masses.

This naturally leads to a new description of the sequence of weighted trees as a bead splitting process, in the terminology of [39], which converges to the stable tree. This only uses (β, β)-strings of beads $([0, K], \mu)$, in which $\mu = \sum_{i \geq 1} P_i \delta_{KU_i}$ for $(P_i, i \geq 1) \sim \mathrm{PD}(\beta, \beta)$ and $U_i \sim \mathrm{Unif}(0, 1)$, $i \geq 1$, independent. In this description, the selection of the attachment point J_k is based on bead masses rather than branch lengths, and only a proportion of the mass in J_k is spread over the new branch, depending on the degree $\deg(J_k, \mathcal{T}_k)$ of J_k in $\mathcal{T}_k$.

Algorithm 1.7 (*Bead-Splitting Construction of the Stable Tree*)**.** Let $\beta \in (0, 1/2]$. We grow weighted $\mathbb{R}$-trees $(\mathcal{T}_k, \mu_k)$, $k \geq 0$, as follows (cf. Fig. 2).

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

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

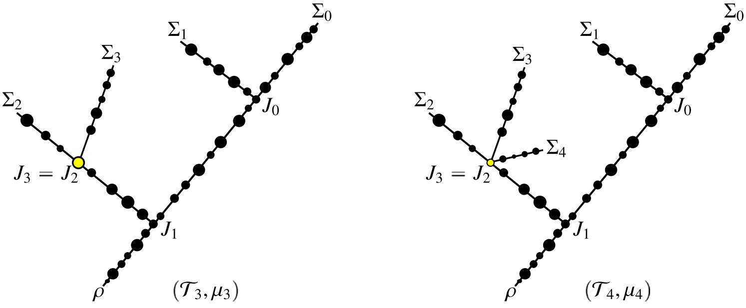

Fig. 2. Example of $(\mathcal{T}_3, \mu_3)$ and $(\mathcal{T}_4, \mu_4)$ constructed via bead splitting. To build $(\mathcal{T}_4, \mu_4)$ from $(\mathcal{T}_3, \mu_3)$, the attachment point J_3 is selected via sampling from μ_3, and a Q_3 proportion of the mass of the selected bead is split up into a rescaled independent copy of the initial string of beads $(\mathcal{T}_0, \mu_0)$, which is attached to J_3.

1.–2. sample J_k from μ_k;

3. given $\deg(J_k, \mathcal{T}_k) = d \geq 2$, let $Q_k \sim \text{Beta}(\beta, (d-2)(1-\beta) + 1 - 2\beta)$, 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 1.8. *In Algorithm 1.7, we have* $\lim_{k\to\infty}(\mathcal{T}_k, \mu_k) = (\mathcal{T}, \mu)$ *a.s. in the Gromov–Hausdorff–Prokhorov topology, where* $(\mathcal{T}, \mu)$ *is a stable tree of index* β.

In fact, part of our argument is to show that the sequence of unweighted trees $(\mathcal{T}_k, k \geq 0)$ has the same distribution as the sequence in (1.1) (and as in Algorithm 1.1), and that the projected mass measures are as required. This claim for $k = 0$ is well-known in different terminology in [25, Corollary 10(3)]. See also [39, discussion after Corollary 8].

We conclude this section by complementing the growth of CRTs by discrete two-colour tree growth. Marchal [33] introduced a 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 $\mathbb{R}$-trees with unit edge lengths equipped 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.

Algorithm 1.9 (*Marchal's Algorithm*)**.** Let $\beta \in (0, 1/2]$. We grow discrete trees T_n, $n \geq 0$, as follows.

0. Let T_0 consist of a root ρ and a leaf Σ_0, connected by an edge.

Given T_n, with leaves $\Sigma_0, \ldots, \Sigma_n$,

1. distribute a total weight of $n + \beta$ by assigning $(d-2)(1-\beta) - \beta$ to each vertex 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 leaf Σ_{n+1} to form T_{n+1}.

Strengthening a result by Marchal [33], Curien and Haas [8] showed that the sequence of trees $(T_n, n \geq 0)$ has the stable tree $\mathcal{T}$ of index β as its a.s. scaling limit, in the following strong sense:

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

In a manner similar to the algorithms for the alpha–gamma model of [7], 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 1.10 (*The Discrete Two-Colour Model*)**.** Let $\beta \in (0, 1/2]$. We grow random discrete two-colour trees $(T_n^*, (R_n^{(i)}, i \geq 1))$, $n \geq 0$, as follows.

0. Let T_0^* consist of a root ρ and a leaf Σ_0 connected by an edge, let $R_0^{(i)} = \{\rho\}$, $i \geq 1$, and $r_0 = 0$.

Given $(T_n^*, (R_n^{(i)}, i \geq 1))$, with leaves $\Sigma_0, \ldots, \Sigma_n$ and $r_n = \#\{i \geq 1 : R_n^{(i)} \neq \{\rho\}\}$,

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 and set $I_n = r_n + 1$; if the selected edge is a marked edge of $R_n^{(i)}$ for some $i \geq 1$, replace it by two marked edges and set $I_n = i$; proceed with the new vertex as the selected vertex;
3. add a new degree-2 vertex, connect it to the selected vertex by a marked edge, and to a new leaf Σ_{n+1} by an unmarked edge; add the marked edge to $R_n^{(I_n)}$ to form $R_{n+1}^{(I_n)}$; set $R_{n+1}^{(i)} = R_n^{(i)}$ for $i \neq I_n$.

Proposition 1.11 (*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 1.10, which we view as a sequence of ∞-marked $\mathbb{R}$-trees with unit edge lengths. For all $k \geq 0$, denote the reduced tree spanned by the root ρ and the leaves $\Sigma_0, \ldots, \Sigma_k$ by $\mathcal{R}(T_n^*, (R_n^{(i)}, i \geq 1), \Sigma_0, \ldots, \Sigma_k)$. Then*

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

with respect to the distance d_{GH}^{∞}, where $(\mathcal{T}_k^*, (\mathcal{R}_k^{(i)}, i \geq 1), k \geq 0)$ is as in Algorithm 1.3.

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 weight 1 for each unmarked edge and each internal marked edge, and weight $1/\beta - 2$ for each external marked edge, are conditionally independent.

This proposition can be proved using exactly the same techniques as the proof of the corresponding result for the alpha–gamma model, cf. [7, Propositions 21 and 22], and the result for (α, θ)-tree growth processes, cf. [39, Proposition 14]. We omit the details.

Remark 1.12. One can obtain mass measures μ_k^* on $\mathcal{T}_k^*$, $k \geq 0$, as scaling limits of the empirical measures on the leaves of T_n^*, projected onto the reduced trees, using the same methods as in [39]. 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 directly that these strings of beads are independent of each other and of the mass split on $\mathcal{T}_k^*$, which has distribution Dirichlet$(\beta, \ldots, \beta, 1 - 2\beta, \ldots, 1 - 2\beta)$, with parameter β for each internal marked and unmarked edge, and parameter $1-2\beta$ for each external marked edge of $\mathcal{T}_k^*$, cf. Proposition 3.3.

There has been a recent sequence of papers by Sénizergues that relates to our work in several ways. Building on a generalised line-breaking construction constructed by Curien and Haas [9], Sénizergues [43,44] develops more general algorithms that grow random graphs or more general metric spaces by gluing substructures. When using preferential attachment, this includes Marchal's algorithm and related line-breaking constructions. See also further joint work [21] for related constructions of stable graphs and [45] for branch point replacement constructions.

This paper is structured as follows. We introduce the framework of ∞-marked $\mathbb{R}$-trees in Section 2 and also collect some preliminary results on strings of beads and stable trees. Section 3 introduces the limiting tree of the two-colour line-breaking construction and states the auxiliary bead-splitting constructions, which are the keys to analysing the two-colour line-breaking construction in Section 4. Section 5 completes the proofs of our main results. An appendix includes the technical proof of a result postponed from an earlier section.

2. Preliminaries on marked $\mathbb{R}$-trees, strings of beads and stable trees

2.1. $\mathbb{R}$-Trees and the Gromov–Hausdorff topology

A compact metric space $(\mathcal{T}, d)$ is called an $\mathbb{R}$-*tree* [16,30] if for each $x, y \in \mathcal{T}$ the following holds.

(i) There is an isometry $f_{x,y} : [0, d(x, y)] \to \mathcal{T}$ such that $f_{x,y}(0) = x$ and $f_{x,y}(d(x, y)) = y$.
(ii) For all injective paths $g : [0, 1] \to \mathcal{T}$ with $g(0) = x$ and $g(1) = y$, we have $g([0, 1]) = f_{x,y}([0, d(x, y)])$.

We denote the range of $f_{x,y}$ by $[[x, y]] := f_{x,y}\,([0, d(x, y)])$. All our $\mathbb{R}$-trees will be *rooted* at a distinguished element ρ, the *root* of $\mathcal{T}$. We call two $\mathbb{R}$-trees $(\mathcal{T}, d, \rho)$ and $(\mathcal{T}', d', \rho')$ *equivalent* if there is an isometry from $\mathcal{T}$ to $\mathcal{T}'$ that maps ρ onto ρ'. We denote by $\mathbb{T}$ the set of equivalence classes of rooted $\mathbb{R}$-trees, which we equip with the Gromov–Hausdorff distance d_{GH} [17] to obtain the Polish space $(\mathbb{T}, d_{\mathrm{GH}})$. The *Gromov–Hausdorff distance* between two $\mathbb{R}$-trees $(\mathcal{T}, d, \rho)$ and $(\mathcal{T}', d', \rho')$ is defined as

$$d_{\mathrm{GH}}\left((\mathcal{T}, d, \rho), (\mathcal{T}', d', \rho')\right) := \inf_{\varphi, \varphi'} \left\{\max\left\{\delta\left(\varphi(\rho), \varphi'(\rho')\right), \delta_{\mathrm{H}}\left(\varphi(\mathcal{T}), \varphi'(\mathcal{T}')\right)\right\}\right\}, \tag{2.1}$$

where the infimum is taken over all metric spaces $(\mathcal{M}, \delta)$ and all isometric embeddings $\varphi : \mathcal{T} \to \mathcal{M}$, $\varphi' : \mathcal{T}' \to \mathcal{M}$ into the common metric space $(\mathcal{M}, \delta)$, and δ_{H} is the Hausdorff distance between compact subsets of $(\mathcal{M}, \delta)$. It is well-known that the Gromov–Hausdorff distance only depends on equivalence classes of rooted $\mathbb{R}$-trees, and we equip $\mathbb{T}$ with the Borel σ-algebra $\mathcal{B}(\mathbb{T})$ induced by d_{GH}.

We can enhance a rooted $\mathbb{R}$-tree by considering a probability measure μ on its Borel sets $\mathcal{B}(\mathcal{T})$, and call $(\mathcal{T}, d, \rho, \mu)$ a *weighted* $\mathbb{R}$-tree. We call $(\mathcal{T}, d, \rho, \mu)$ and $(\mathcal{T}', d', \rho', \mu')$

equivalent if there is an isometry from $\mathcal{T}$ to $\mathcal{T}'$ such that ρ is mapped onto ρ' and μ' is the push-forward of μ under this isometry. We let $\mathbb{T}_{\rm w}$ denote the set of equivalence classes of weighted $\mathbb{R}$-trees. Then $\mathbb{T}_{\rm w}$ is Polish when equipped with the *Gromov–Hausdorff–Prokhorov distance* $d_{\rm GHP}$ induced by

$$\begin{aligned} &d_{\rm GHP}\left((\mathcal{T},d,\rho,\mu),\left(\mathcal{T}',d',\rho',\mu'\right)\right)\\ &:=\inf_{\varphi,\varphi'}\left\{\max\left\{\delta\left(\varphi\left(\rho\right),\varphi'\left(\rho'\right)\right),\delta_{\rm H}\left(\varphi\left(\mathcal{T}\right),\varphi'\left(\mathcal{T}'\right)\right),\delta_{\rm P}\left(\varphi_*\mu,\varphi'_*\mu'\right)\right\}\right\} \end{aligned} \tag{2.2}$$

for weighted $\mathbb{R}$-trees $(\mathcal{T},d,\rho,\mu)$, $(\mathcal{T}',d',\rho',\mu')$, where $\varphi,\varphi',\delta_{\rm H}$ are as in (2.1), $\varphi_*\mu$, $\varphi'_*\mu$ are the push-forwards of μ, μ' via φ,φ', respectively, and $\delta_{\rm P}$ is the Prokhorov distance on the space of Borel probability measures on $(\mathcal{M},\delta)$ given by

$$\delta_{\rm P}\left(\mu,\mu'\right)=\inf\left\{\epsilon>0\colon\mu(D)\le\mu'(D^\epsilon)+\epsilon\text{ for all }D\subset\mathcal{M}\text{ closed}\right\},$$

where $D^\epsilon=\{x\in\mathcal{M}\colon\inf_{y\in D}\delta(x,y)\le\epsilon\}$ denotes the *ϵ-thickening* of D.

While some of our developments are more easily stated in $(\mathbb{T},d_{\rm GH})$ or $(\mathbb{T}_{\rm w},d_{\rm GHP})$, others benefit from more explicit embeddings into a particular metric space $(\mathcal{M},\delta)$, which we will mostly choose as

$$\mathcal{M}=l^1(\mathbb{N}_0^2):=\left\{(s_{i,j})_{i,j\in\mathbb{N}_0}\in[0,\infty)^{\mathbb{N}_0^2}\colon\sum_{i,j\in\mathbb{N}_0}s_{i,j}<\infty\right\}$$

with the metric induced by the l^1-norm. This is a variation of Aldous's [1–3] choice $\mathcal{M}=l^1(\mathbb{N})$. We denote by $\mathbb{T}^{\rm emb}$ the space of all compact $\mathbb{R}$-trees $\mathcal{T}\subset l^1(\mathbb{N}_0^2)$ with root $0\in\mathcal{T}$, which we equip with the Hausdorff metric $\delta_{\rm H}$, and by $\mathbb{T}^{\rm emb}_{\rm w}$ the space of all weighted compact $\mathbb{R}$-trees $(\mathcal{T},\mu)$ with $\mathcal{T}\in\mathbb{T}^{\rm emb}$, which we equip with the metric $\delta_{\rm HP}((\mathcal{T},\mu),(\mathcal{T}',\mu'))=\max\{\delta_{\rm H}(\mathcal{T},\mathcal{T}'),\delta_{\rm P}(\mu,\mu')\}$.

For $\mathcal{T}\in\mathbb{T}^{\rm emb}$ and $c>0$, we define $c\mathcal{T}:=\{cx\colon x\in\mathcal{T}\}$. More generally for any $\mathbb{R}$-tree $(\mathcal{T},d)$, we slightly abuse notation and denote by $c\mathcal{T}$ the metric space $(\mathcal{T},cd)$ obtained when all distances are multiplied by c. We consider random $\mathbb{R}$-trees whose equivalence class in $\mathbb{T}$ has the distribution of a stable or Ford tree, and also refer to these trees as stable or Ford trees, and to the associated law on $\mathbb{T}$ as their *distribution*.

If $x\in\mathcal{T}\setminus\{\rho\}$ is such that $\mathcal{T}\setminus\{x\}$ is connected, we call x a *leaf* of $\mathcal{T}$. A *branch point* is an element $x\in\mathcal{T}$ such that $\mathcal{T}\setminus\{x\}$ has at least three connected components. We refer to the number of these components as the *degree* $\deg(x,\mathcal{T})$ of x. We denote the sets of all leaves and branch points by $\mathrm{Lf}(\mathcal{T})$ and $\mathrm{Br}(\mathcal{T})$. If $\mathcal{T}\setminus\mathrm{Br}(\mathcal{T})$ has only finitely many connected components, we call $\mathcal{T}$ a *discrete* $\mathbb{R}$-tree and these components (with or without one or both endpoints) *edges*. We denote the set of edges by $\mathrm{Edg}(\mathcal{T})$, and call $\#\mathrm{Lf}(\mathcal{T})$ the *size* of $\mathcal{T}$. Also, $|\mathcal{T}|:=\#\mathrm{Edg}(\mathcal{T})$. We call the discrete graph with the edge set $\mathrm{Edg}(\mathcal{T})$ the *shape* of $\mathcal{T}$.

In the case of discrete weighted $\mathbb{R}$-trees it will often be of interest how the total mass of 1 is distributed between the edges, with possibly some mass in branch points, which for convenience we will also write in the form $E=\{v\}$. For any weighted $\mathbb{R}$-tree $(\mathcal{T},\mu)$ with n edges/branch points $E_1,\ldots,E_n$, the vector $(X_1,\ldots,X_n)$ with $X_i:=\mu(E_i)$, $i\in[n]$, is called the *mass split* in $\mathcal{T}$. We will also consider mass splits in subtrees $\mathcal{R}\subset\mathcal{T}$, i.e. mass splits in $(\mathcal{R},\mu(\mathcal{R})^{-1}\mu\restriction_{\mathcal{R}})$, where for any Borel set $A\in\mathcal{B}(\mathcal{T})$, we define $\mu\restriction_{\mathcal{R}}(A)=\mu(A\cap\mathcal{R})$. To distinguish mass splits in the "big" tree $\mathcal{T}$ and in "small" subtrees, we will speak of the *total* and *internal* (or *relative*) mass splits, respectively.

The limiting trees of the weighted $\mathbb{R}$-trees in our constructions will be *continuum trees*, i.e. weighted $\mathbb{R}$-trees $(\mathcal{T},d,\mu)$ such that the probability measure μ on $\mathcal{T}$ satisfies the following three properties: (i) μ is supported by the set $\mathrm{Lf}(\mathcal{T})$ of leaves of $\mathcal{T}$; (ii) μ is non-atomic, i.e. for any $x \in \mathrm{Lf}(\mathcal{T})$, $\mu(x) = 0$; (iii) For any $x \in \mathcal{T} \setminus \mathrm{Lf}(\mathcal{T})$ and $\mathcal{T}_x := \{\sigma \in \mathcal{T} : x \in [[\rho, \sigma]]\}$, we have $\mu(\mathcal{T}_x) > 0$.

It is an immediate consequence of (i)–(iii) that, for any continuum tree $(\mathcal{T}, d)$, the set of leaves $\mathrm{Lf}(\mathcal{T})$ is uncountable and that it has no isolated points. Finally, we introduce the notion of a *reduced* subtree

$$\mathcal{R}(\mathcal{T}, x_1, \ldots, x_n) := \bigcup_{i\in[n]} [[\rho, x_i]] \tag{2.3}$$

of an $\mathbb{R}$-tree $\mathcal{T}$ spanned by the root and $x_1, x_2, \ldots, x_n \in \mathrm{Lf}(\mathcal{T})$. Note that $\mathcal{R}(\mathcal{T}, x_1, \ldots, x_n)$ is a discrete $\mathbb{R}$-tree with root ρ and leaves $x_1, \ldots, x_n$. We further consider the projection map that projects the tree onto the closest part of the reduced tree

$$\pi_k : \mathcal{T} \to \mathcal{R}(\mathcal{T}, x_1, \ldots, x_k), \quad y \mapsto f_{\rho,y}\left(\sup\{t \geq 0 : f_{\rho,y}(t) \in \mathcal{R}(\mathcal{T}, x_1, \ldots, x_k)\}\right), \tag{2.4}$$

where $f_{\rho,y} : [0, d(\rho, y)] \to \mathcal{T}$ is the unique isometry with $f_{\rho,y}(0) = \rho$ and $f_{\rho,y}(d(\rho, y)) = y$ from the definition of an $\mathbb{R}$-tree. The push-forward of a probability measure μ on $\mathcal{T}$ via this projection map is denoted by $(\pi_k)_*\mu$, i.e.

$$(\pi_k)_*\mu\,(D) = \mu\left(\pi_k^{-1}(D)\right), \qquad D \subset \mathcal{R}(\mathcal{T}, x_1, \ldots, x_k) \text{ Borel measurable.} \tag{2.5}$$

More details on $\mathbb{R}$-trees and proofs for the statements made in this section can be found in [5,16,30].

2.2. ∞-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, [36, Section 6.4]. In the context of the two-colour line-breaking construction, the marked components correspond to the rescaled Ford trees by which we replace the branch points in the stable line-breaking construction. Each Ford tree, i.e. each connected red component, is related to a new marked subset of the ∞-marked $\mathbb{R}$-tree.

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 non-empty closed connected subsets $\mathcal{R}^{(1)}, \ldots, \mathcal{R}^{(k)} \subset \mathcal{T}$. We call two k-marked $\mathbb{R}$-trees $(\mathcal{T}, d, \rho, (\mathcal{R}^{(1)}, \ldots, \mathcal{R}^{(k)}))$ and $(\mathcal{T}', d', \rho, (\mathcal{R}'^{(1)}, \ldots, \mathcal{R}'^{(k)}))$ *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, and ρ is mapped onto ρ'. 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 $\mathcal{R}'^{(i)}$, $i \in [k]$, ρ is mapped to ρ' and μ' is the push-forward of μ under this isometry. The set of equivalence classes of k-marked $\mathbb{R}$-trees is denoted by $\mathbb{T}^{[k]}$, and $\mathbb{T}^{[k]}_{\mathrm{w}}$ is the set of equivalence classes of weighted k-marked $\mathbb{R}$-trees.

For k-marked $\mathbb{R}$-trees $(\mathcal{T}, d, \rho, (\mathcal{R}^{(1)}, \ldots, \mathcal{R}^{(k)}))$, $(\mathcal{T}', d', \rho', (\mathcal{R}'^{(1)}, \ldots, \mathcal{R}'^{(k)})) \in \mathbb{T}^{[k]}$, let

$$\begin{aligned} &d_{\mathrm{GH}}^{[k]}\left(\left(\mathcal{T}, d, \rho, \left(\mathcal{R}^{(1)}, \ldots, \mathcal{R}^{(k)}\right)\right), \left(\mathcal{T}', d', \rho', \left(\mathcal{R}'^{(1)}, \ldots, \mathcal{R}'^{(k)}\right)\right)\right) \\ &\quad := \inf_{\varphi,\varphi'} \left\{\max\left\{\delta_{\mathrm{H}}\left(\varphi(\mathcal{T}), \varphi'(\mathcal{T}')\right), \max_{1\leq i\leq k} \delta_{\mathrm{H}}\left(\varphi\left(\mathcal{R}^{(i)}\right), \varphi'\left(\mathcal{R}'^{(i)}\right)\right), \delta\left(\varphi(\rho), \varphi(\rho')\right)\right\}\right\} \end{aligned} \tag{2.6}$$

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 [36] that $d_{\mathrm{GH}}^{[k]}$ is a metric on $\mathbb{T}^{[k]}$.

Lemma 2.1 (*[36, Proposition 9(ii)]*). *The space* $(\mathbb{T}^{[k]}, d_{\mathrm{GH}}^{[k]})$ *is separable and complete.*

We extend the notion of a k-marked $\mathbb{R}$-tree to an ∞-*marked* $\mathbb{R}$-tree $(\mathcal{T}, d, \rho, (\mathcal{R}^{(i)}, i \geq 1))$. The marked components $\mathcal{R}^{(i)}, i \geq 1$, of an ∞-marked $\mathbb{R}$-tree $(\mathcal{T}, (\mathcal{R}^{(i)}, i \geq 1))$ are themselves $\mathbb{R}$-trees when equipped with the metric to $\mathcal{R}^{(i)}$, and rooted at the point of $\mathcal{R}^{(i)}$ closest to the root of $\mathcal{T}$, $i \geq 1$. We will consider ∞-marked $\mathbb{R}$-trees $(\mathcal{T}, d, \rho, (\mathcal{R}^{(i)}, i \geq 1))$ with a discrete branching structure, and distinguish between *internal* and *external* edges of $\mathcal{R}^{(i)}$. External edges of $\mathcal{R}^{(i)}$ are edges connecting a branch point/root and a leaf of $\mathcal{R}^{(i)}$, internal edges connect two branch points or the root and a branch point.

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

$$d_{\mathrm{GH}}^{\infty}\left(\left(\mathcal{T}, d, \rho, \left(\mathcal{R}^{(i)}, i \geq 1\right)\right), \left(\mathcal{T}', d', \rho', \left(\mathcal{R}'^{(i)}, i \geq 1\right)\right)\right)$$
$$:= \sum_{k \geq 1} 2^{-k} d_{\mathrm{GH}}^{[k]}\left(\left(\mathcal{T}, \left(\mathcal{R}^{(i)}, \ldots, \mathcal{R}^{(k)}\right)\right), \left(\mathcal{T}', \left(\mathcal{R}'^{(1)}, \ldots, \mathcal{R}'^{(k)}\right)\right)\right). \tag{2.7}$$

Corollary 2.2. *The space* $(\mathbb{T}^{\infty}, d_{\mathrm{GH}}^{\infty})$ *is separable and complete.*

Proof. This can be deduced from Lemma 2.1. We leave the details to the reader. □

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 $(\mathcal{T}, (\mathcal{R}^{(1)}, \ldots, \mathcal{R}^{(k)}), \mu), (\mathcal{T}', (\mathcal{R}'^{(1)}, \ldots, \mathcal{R}'^{(k)}), \mu') \in \mathbb{T}^{[k]}$, we define

$$d_{\mathrm{GHP}}^{[k]}\left(\left(\mathcal{T}, \left(\mathcal{R}^{(1)}, \ldots, \mathcal{R}^{(k)}\right), \mu\right), \left(\mathcal{T}', \left(\mathcal{R}'^{(1)}, \ldots, \mathcal{R}'^{(k)}\right), \mu'\right)\right)$$
$$:= \inf_{\varphi, \varphi'}\left\{\max\left\{\delta_{\mathrm{H}}\left(\varphi\left(\mathcal{T}\right), \varphi'\left(\mathcal{T}'\right)\right), \max_{1 \leq i \leq k} \delta_{\mathrm{H}}\left(\varphi\left(\mathcal{R}^{(i)}\right), \varphi'\left(\mathcal{R}'^{(i)}\right)\right), \delta\left(\varphi\left(\rho\right), \varphi'\left(\rho'\right)\right), \delta_{\mathrm{P}}\left(\varphi_*\mu, \varphi'_*\mu'\right)\right\}\right\}$$

where φ, φ' and $\varphi_*\mu, \varphi'_*\mu'$ are as in (2.2) and (2.6). In the spirit of (2.7), we define

$$d_{\mathrm{GHP}}^{\infty}\left(\left(\mathcal{T}, \left(\mathcal{R}^{(i)}, i \geq 1\right), \mu\right), \left(\mathcal{T}', \left(\mathcal{R}'^{(i)}, i \geq 1\right), \mu'\right)\right)$$
$$= \sum_{k \geq 1} 2^{-k} d_{\mathrm{GHP}}^{[k]}\left(\left(\mathcal{T}, \left(\mathcal{R}^{(1)}, \ldots, \mathcal{R}^{(k)}\right), \mu\right), \left(\mathcal{T}', \left(\mathcal{R}'^{(1)}, \ldots, \mathcal{R}'^{(k)}\right), \mu'\right)\right) \tag{2.8}$$

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 2.3. *The function* $d_{\mathrm{GHP}}^{[k]}$ *defines a metric on* $\mathbb{T}_{\mathrm{w}}^{[k]}$*, and the space* $(\mathbb{T}_{\mathrm{w}}^{[k]}, d_{\mathrm{GHP}}^{[k]})$ *is separable and complete, for any* $k \in \{0, 1, 2, \ldots; \infty\}$.

Proof. For $k \in \{0, 1, 2, \ldots\}$, the proof is a direct generalisation of the proof of Lemma 2.1. In particular, it is straightforward to generalise the results about d_{GHP} in [36, Section 6.2/6.3] to $d_{\mathrm{GHP}}^{[k]}$. For $k = \infty$, the claim can then be deduced from the finite-k case, as for Corollary 2.2. □

Remark 2.4. Miermont [36] 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 $\mathbb{M}^{[k+1]}$ ([17, Lemma 2.1]), i.e. the results on $(\mathbb{T}^{[k]}, d_{\mathrm{GH}}^{[k]})$ presented here follow from his study of $(\mathbb{M}^{[k]}, d_{\mathrm{GH}}^{[k]})$, $k \geq 0$.

2.3. Dirichlet and Mittag-Leffler distributions

In this section, we recall well-known 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 pth moment is given by

$$\mathbb{E}\left[L^p\right] = \frac{\Gamma(\theta+1)\Gamma(\theta/\alpha+1+p)}{\Gamma(\theta/\alpha+1)\Gamma(\theta+p\alpha+1)}, \quad p \geq 1, \tag{2.9}$$

for short $L \sim \mathrm{ML}(\alpha, \theta)$. The moments (2.9) uniquely characterise $\mathrm{ML}(\alpha, \theta)$, cf. [38].

The Mittag-Leffler distribution naturally appears when we study lengths in the trees considered in this paper. 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 2.5 (*[20] Proposition 4.2*)**.** *Let $\beta \in (0, 1)$. For $n \geq 2$, let $\theta_1, \ldots, \theta_n > 0$ and let $\theta := \sum_{i\in[n]} \theta_i$. 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,*

$$S \cdot (Y_1, \ldots, Y_n) \stackrel{d}{=} \left(X_1^\beta S^{(1)}, \ldots, X_n^\beta S^{(n)}\right) \tag{2.10}$$

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

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

Proposition 2.6. *Let $n \in \mathbb{N}$, $\theta_1, \ldots, \theta_n > 0$ and $X := (X_1, \ldots, X_n) \sim \mathrm{Dirichlet}(\theta_1, \ldots, \theta_n)$.*

(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 $m \in [n-1]$. Then $X' := (\sum_{i\in[m]} X_i, X_{m+1}, \ldots, X_n) \sim \mathrm{Dirichlet}\left(\sum_{i\in[m]} \theta_i, \theta_{m+1}, \ldots, \theta_n\right)$ and $X^* := (X_1/\sum_{i\in[m]} X_i, \ldots, X_m/\sum_{i\in[m]} X_i) \sim \mathrm{Dirichlet}\,(\theta_1, \ldots, \theta_m)$ are independent.*

(iii) *Decimation. Let $i \in [n]$, $m \in \mathbb{N}$, and let $\theta_{i,1}, \ldots, \theta_{i,m} > 0$ be such that $\sum_{j\in[m]} \theta_{i,j} = \theta_i$. Consider an independent random vector $(P_1, \ldots, P_m) \sim \mathrm{Dirichlet}\left(\theta_{i,1}, \ldots, \theta_{i,m}\right)$. Then*

$$\begin{aligned} X'' &:= (X_1, \ldots, X_{i-1}, P_1 X_i, \ldots, P_m X_i, X_{i+1}, \ldots, X_n) \\ &\sim \mathrm{Dirichlet}\left(\theta_1, \ldots, \theta_{i-1}, \theta_{i,1}, \ldots, \theta_{i,m}, \theta_{i+1}, \ldots, \theta_n\right). \end{aligned}$$

(iv) *Size-bias. Let $I \in [n]$ be a random index such that $\mathbb{P}(I = i | X_1, \ldots, X_n) = X_i$ a.s. for $i \in [n]$. Then for any $i \in [n]$, conditionally given $I = i$, we have $X \sim \mathrm{Dirichlet}$ $(\theta_1, \ldots, \theta_{i-1}, \theta_i + 1, \theta_{i+1}, \ldots, \theta_n)$. Furthermore, we have $\mathbb{P}(I = i) = \theta_i \big/ \sum_{j\in[n]} \theta_j$.*

Proof. We refer to [46, Propositions 13–14, Remark 15], and the Gamma variable representation for the Dirichlet distribution. □

2.4. 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 (CRPs), cf. [26,38,39]. Consider customers labelled by $[n] := \{1, \ldots, n\}$ sitting at a random number of tables 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, the next customer labelled by $n+1$

- sits at the ith occupied table with probability $(n_i - \alpha)/(n+\theta)$, $i \in [k]$;
- opens a new table to the left of the first table, or between any two tables with probability $\alpha/(n+\theta)$;
- opens a new table to the right of the last table with probability $\theta/(n+\theta)$.

This induces the *ordered* (α, θ)-CRP $(\widetilde{\Pi}_n, n \geq 1)$. The classical *unordered* (α, θ)-*CRP* $(\Pi_n, n \geq 1)$ is obtained from $(\widetilde{\Pi}_n, n \geq 1)$ by ordering the blocks by least labels. For $n \in \mathbb{N}$, we write $\Pi_n = (\Pi_{n,1}, \ldots, \Pi_{n,K_n})$ and $\widetilde{\Pi}_n = (\widetilde{\Pi}_{n,1}, \ldots, \widetilde{\Pi}_{n,K_n})$ for the blocks of the two partitions of $[n]$, where K_n denotes the number of tables at step n. The block sizes at step n form random *compositions* of n, $n \geq 1$, i.e. sequences of positive integers $(n_1, \ldots, n_k)$ with sum $n = \sum_{j \in [k]} n_j$. The composition related to $\widetilde{\Pi}_n$, $n \geq 1$, can be shown to be *regenerative* in the sense of Gnedin and Pitman [19]. The number of tables K_n at step n, rescaled by n^α, converges a.s., i.e. there is $L_{\alpha,\theta} > 0$ a.s. such that

$$L_{\alpha,\theta} = \lim_{n \to \infty} n^{-\alpha} K_n \quad \text{a.s..} \tag{2.11}$$

The distribution of $L_{\alpha,\theta}$ can be identified as $\mathrm{ML}(\alpha, \theta)$. Furthermore, there are 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 order of least labels, i.e.

$$\lim_{n \to \infty} \left(n^{-1}\#\Pi_{n,1}, \ldots, n^{-1}\#\Pi_{n,K_n}\right) = (P_1, P_2, P_3, \ldots) = \left(V_1, \overline{V}_1 V_2, \overline{V}_1 \overline{V}_2 V_3, \ldots\right) \quad \text{a.s.} \tag{2.12}$$

where $(V_i, i \geq 1)$ are independent with $V_i \sim \mathrm{Beta}(1-\alpha, \theta + i\alpha)$, and $\overline{V}_i := 1 - V_i$. The distribution of the vector $(P_1, P_2, \ldots)$ is a *Griffiths–Engen–McCloskey distribution* $\mathrm{GEM}(\alpha, \theta)$. Ranking $(P_i, i \geq 1)$ in decreasing order we obtain a Poisson–Dirichlet sequence $(P_i^\downarrow, i \geq 1) := (P_i, i \geq 1)^\downarrow \sim \mathrm{PD}(\alpha, \theta)$. Each P_i, $i \geq 1$, is further associated with a position on the limiting interval $[0, L_{\alpha,\theta}]$ induced by the table order, in a way that we now describe.

Consider an *ordered* (α, θ)-CRP $(\widetilde{\Pi}_n = (\widetilde{\Pi}_{n,1}, \ldots, \widetilde{\Pi}_{n,K_n}), n \geq 1)$ for $\alpha \in (0, 1)$, $\theta > 0$. Let $N_{n,j} := \sum_{i \in [j]} \#\widetilde{\Pi}_{n,i}$, $j \in [n]$, be the number of customers at the first j tables from the left. Then by [39, Proposition 6],

$$\lim_{n \to \infty} \left\{n^{-1} N_{n,j}, j \geq 0\right\} = \mathcal{N}_{\alpha,\theta} := \left\{1 - e^{-G_t}, t \geq 0\right\}^{\mathrm{cl}} \quad \text{a.s.} \tag{2.13}$$

with respect to the Hausdorff metric on closed subsets of $[0, 1]$, where cl denotes the closure in $[0, 1]$, and $(G_t, t \geq 0)$ is a subordinator with Laplace exponent $\Phi_{\alpha,\theta}(s) = s\Gamma(s+\theta)\Gamma(1-\alpha)/\Gamma(s+\theta+1-\alpha)$, and 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

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

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 $\mathrm{PD}(\alpha, \theta)$. The inverse local time

$$\mathcal{L}^{-1} : [0, L_{\alpha,\theta}) \to [0, 1), \qquad \mathcal{L}^{-1}(x) := \inf\{u \in [0, 1] : \mathcal{L}(u) > x\}, \tag{2.14}$$

is right-continuous increasing. We equip the random interval $[0, L_{\alpha,\theta}]$ with the Stieltjes measure $d\mathcal{L}^{-1}$.

Definition 2.7 (*String of Beads*)**.** A *string of beads* (I, λ) is an interval I equipped with a discrete mass measure λ. A measure-preserving isometric copy of $([0, L_{\alpha,\theta}], d\mathcal{L}^{-1})$ associated as above with an (α, θ)-regenerative interval partition $[0, 1] \setminus \mathcal{N}_{\alpha,\theta}$ is called an (α, θ)-*string of beads*, for $\alpha \in (0, 1), \theta > 0$.

We can view a string of beads $([0, K], \lambda)$ as a weighted $\mathbb{R}$-tree consisting of one single branch connecting the root 0 with a leaf at distance K.

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 from the vector $(P_i, i \geq 1)$ of the stick-breaking representation (2.12) via

$$L_{\alpha,\theta} = \lim_{i\to\infty} i\,\Gamma(1-\alpha)(P_i^{\downarrow})^{\alpha} = \lim_{k\to\infty} \left(1 - \sum_{i\in[k]} P_i\right)^{\alpha} \alpha^{-\alpha} k^{1-\alpha}, \tag{2.15}$$

which is the so-called *α-diversity* of $(P_i^{\downarrow}, i \geq 1) \sim \mathrm{PD}(\alpha, \theta)$, cf. [38, Lemma 3.11].

One of the key properties of (α, θ)-strings of beads is the regenerative nature inherited from the underlying regenerative interval partition, cf. [19]. Pitman and Winkel [39, Proposition 10] developed a method ('(α, θ)-coin-tossing sampling') 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). This sampling procedure makes precise the idea of choosing the first marked bead seen from the bottom where each bead is marked independently by a probability of $x\theta/(x\theta + (1 - x)\alpha)$, where x is the size of the bead relative to the residual mass of the string at and above this bead. The mass split between the two induced interval components and the selected atom is $\mathrm{Dirichlet}(\alpha, 1 - \alpha, \theta)$, with parameters assigned in their order on the interval $[0, L_{\alpha,\theta}]$. When $\theta = \alpha$, the special sampling reduces to uniform sampling from the mass measure $d\mathcal{L}^{-1}$.

Proposition 2.8 (*[39, Proposition 10/14(b), Corollary 15]*)**.** *Let $(I, \lambda) := ([0, L_{\alpha,\theta}], d\mathcal{L}^{-1})$ be an (α, θ)-string of beads for some $\alpha \in (0, 1), \theta > 0$. Then there is a random variable $J \in (0, L_{\alpha,\theta})$, (which can be constructed explicitly using (α, θ)-coin-tossing) on a suitably enlarged probability space such that the following are independent.*

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

In Section 3 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 mass-sampled, not length-sampled. The following lemma will imply that the algorithms based on masses induce the length versions.

Lemma 2.9. *Let* $(X_1, \ldots, X_n) \sim \text{Dirichlet}(\theta_1, \ldots, \theta_n)$ *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,* $i \in [n]$.

- *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* $B \sim \text{Beta}(1, \theta_j/\alpha)$ *is independent.*

Then the two random vectors $\left(I', L_1, \ldots, L_{I'-1}, L', L_{I'} - L', L_{I'+1}, \ldots, L_n\right)$ *and* $\left(I'', L_1, \ldots, L_{I''-1}, L'', L_{I''} - L'', L_{I''+1}, \ldots, L_n\right)$ *have the same distribution.*

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

$$\begin{aligned}&\mathbb{E}\left[f\left(I', L_1, \ldots, L_{I'-1}, L', L_{I'} - L', L_{I'+1}, \ldots, L_n\right)\right]\\&= \mathbb{E}\left[f\left(I'', L_1, \ldots, L_{I''-1}, L'', L_{I''} - L'', L_{I''+1}, \ldots, L_n\right)\right].\end{aligned} \tag{2.16}$$

Conditioning on $I' = j$, and using Proposition 2.6(iv), the LHS of (2.16) is

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

Conditionally given $I' = j$, we select an atom of the (α, θ_j)-string of beads via (α, θ_j)-coin tossing sampling. By Proposition 2.8 and Proposition 2.6(ii), the mass split $(1 - \lambda_j(L'))^{-1} \left(\lambda_j\left([0, L')\right), \lambda_j\left((L', L_j]\right)\right) \sim \text{Dirichlet}\left(\alpha, \theta_j\right)$ and the (α, α)- and the (α, θ)-strings of beads given by

$$\begin{aligned}&\left(\lambda\left([0, L')\right)^{-\alpha} [0, L'), \lambda\left([0, L')\right)^{-1} \lambda \restriction_{[0,L')}\right),\\&\left(\lambda\left((L', L_j]\right)^{-\alpha} \left(L', L_j\right], \lambda\left((L', L_j]\right)^{-1} \lambda \restriction_{(L',L_j]}\right),\end{aligned}$$

respectively, are independent. By Proposition 2.5, we conclude that the relative length split on $[0, L_j]$ is $L'/L_j \sim \text{Beta}(1, \theta_j/\alpha)$. To see (2.16), proceed likewise with the RHS of (2.16), using that, by Proposition 2.5, $(L_1, \ldots, L_n) \sim \text{Dirichlet}(\theta_1/\alpha, \ldots, \theta_n/\alpha)$. More precisely, note that $\mathbb{P}(I'' = j) = (\theta_j/\alpha)/(\sum_{i\in[n]} \theta_i/\alpha) = \theta_j/\sum_{i\in[n]} \theta_i$, and that, conditionally given $I'' = j$, we have $L''/L_j \sim \text{Beta}(1, \theta_j/\alpha)$, as before. □

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

Proposition 2.10 (*Sampling from PD*(α, θ), *[40, Proposition 34]*)**.** *Let* $(P_i, i \geq 1) \sim \text{PD}(\alpha, \theta)$ *for some* $0 \leq \alpha < 1$ *and* $\theta > -\alpha$, *and let* N *be an index such that*

$$\mathbb{P}\left(N = i \mid P_i, i \geq 1\right) = P_i, \quad i \geq 1.$$

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,* $P_N \sim \text{Beta}(1 - \alpha, \alpha + \theta)$, *and* $(P_i'', i \geq 1) \sim \text{PD}(\alpha, \alpha + \theta)$ *is independent of* P_N.

2.5. Distributional properties of stable trees and ford trees

In this section, we collect some preliminary results on stable trees. Recall the line-breaking construction of the stable tree given by Algorithm 1.1 yielding the sequence of compact $\mathbb{R}$-trees $(\mathcal{T}_k, k \geq 0)$. Leaves and branch points have a natural order induced by the time of appearance in the sequence $(\mathcal{T}_k, k \geq 0)$, 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 list the edges $E_k^{(1)}, \ldots, E_k^{(|\mathcal{T}_k|)}$ of $\mathcal{T}_k$ and their lengths $L_k^{(i)} = \mathrm{Leb}(E_k^{(i)})$, $i \in [|\mathcal{T}_k|]$, in the order encountered on a depth-first search directed by least labels.

Haas et al. [25] 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 index $\beta \in (0, 1/2]$, and let $\Sigma \sim \mu$ be a leaf sampled from μ. Consider the *spine*, i.e. the path $[[\rho, \Sigma]]$ from the root to this leaf. Remove all vertices of degree one or two from this path. This yields a sequence of connected components that can a.s. be ranked in decreasing order of mass, and which we denote by $(\overline{\mathcal{S}}^{(i)}, i \geq 1)$, rooted at vertices $\rho_i \in [[\rho, \Sigma]]$ of a.s. infinite degree, $i \geq 1$, respectively. Each $\overline{\mathcal{S}}^{(i)}$ further separates into a sequence $(\overline{\mathcal{S}}_j^{(i)\downarrow}, j \geq 1)$ when removing ρ_i.

- The *coarse spinal mass partition* is $\big(\overline{P}^{(i)}, i \geq 1\big) := \big(\mu(\overline{\mathcal{S}}^{(i)}), i \geq 1\big)$,
- The *fine spinal mass partition* is $\big(\overline{P}_j^{(i)\downarrow}, j \geq 1, i \geq 1\big)^{\downarrow} := \big(\mu\big(\overline{\mathcal{S}}_j^{(i)\downarrow}\big), j \geq 1, i \geq 1\big)^{\downarrow}$, i.e. the ranked sequence of masses of connected components obtained after removal of the whole spine.

Lemma 2.11 (*Mass Partition in the Stable Tree, [25, 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. $\big(\overline{P}^{(i)}, i \geq 1\big) = \big(\mu\big(\overline{\mathcal{S}}^{(i)}\big), i \geq 1\big) \sim \mathrm{PD}\big(\beta, \beta\big)$.*

(ii) *The fine spinal mass partition is a $(1-\beta, -\beta)$-fragmentation of the coarse spinal mass partition, i.e. for each block $\mu(\overline{\mathcal{S}}^{(i)})$ of the coarse partition, the relative part sizes $(\mu(\overline{\mathcal{S}}_j^{(i)\downarrow})/\mu(\overline{\mathcal{S}}^{(i)}), j \geq 1)$ are independent with distribution $\mathrm{PD}(1-\beta, -\beta)$, $i \geq 1$.*

(iii) *Conditionally given the fine spinal mass partition $(\mu(\overline{\mathcal{S}}_j^{(i)\downarrow}), j \geq 1, i \geq 1)^{\downarrow}$, the rescaled trees equipped with restricted mass measures*

$$\left(\mu\left(\overline{\mathcal{S}}_j^{(i)\downarrow}\right)^{-\beta} \overline{\mathcal{S}}_j^{(i)\downarrow}, \mu(\overline{\mathcal{S}}_j^{(i)\downarrow})^{-1} \mu \restriction_{\overline{\mathcal{S}}_j^{(i)\downarrow}}\right), \quad j \geq 1, i \geq 1, \tag{2.17}$$

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

The α-diversities of $\mathrm{PD}(\alpha, \theta)$ partitions can naturally be interpreted as lengths in trees. 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. The fragmenting $\mathrm{PD}(1-\beta, -\beta)$ random partitions for each block of the coarse spinal mass partition capture important information about the branch points that we relate to sizes of the Ford CRTs by which we replace them in Theorem 1.6.

The following result is essentially [25, Corollary 10(3)]. See also [39, discussion after Corollary 8].

Lemma 2.12. *Let $(\mathcal{T}, \mu)$ be a stable tree of parameter $\beta \in (0, 1/2]$, and let $\Sigma_0 \sim \mu$. Consider the spine $\mathcal{T}_0 = [[\rho, \Sigma_0]]$, 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.*

Finally, our key to identifying Ford trees is the following distributional characterisation:

Proposition 2.13 (*[24, Proposition 18]*). *Consider the tree growth process $(\mathcal{F}_m, m \geq 1)$ from Algorithm 1.2 for some $\beta' \in (0, 1)$. The distribution of $\mathcal{F}_m$ is given in terms of three independent random variables: its shape, the total length $S'_m \sim \mathrm{ML}(\beta', m - \beta')$ and the length split with* Dirichlet $\left(1, \ldots, 1, (1 - \beta')/\beta', \ldots, (1 - \beta')/\beta'\right)$ *distribution, between the edges of $\mathcal{F}_m$, where a parameter of 1 is assigned to each of the $m-1$ internal edges, and a parameter of $(1-\beta')/\beta'$ to each of the m external edges of $\mathcal{F}_m$.*

3. Bead-splitting constructions of marked stable trees

In this section we formulate two enhancements of the line-breaking constructions that, respectively, include mass measures and embed in a limiting CRT. We will formulate these enhancements as bead-splitting constructions in their own right and later show that they reduce to line-breaking constructions.

3.1. The autonomous binary two-colour bead-splitting construction

We present an enhanced version of Algorithm 1.3, which is based on sampling from the mass measure. We will use this enhanced version to prove Theorem 1.4.

The following (1-marked) string of beads will be at the centre of our construction. For $\beta \in (0, 1/2]$, consider $([0, K_1], \lambda_1)$ and $([0, K_2], \lambda_2)$ two independent $(\beta, 1 - 2\beta)$- and (β, β)-strings of beads, respectively, and an independent $B \sim \mathrm{Beta}(1 - 2\beta, \beta)$. Then scale the two strings by B and $1 - B$, as follows: set

$$K := B^{\beta} K_1 + (1 - B)^{\beta} K_2, \qquad K' := B^{\beta} K_1 \tag{3.1}$$

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

$$\lambda\left([0, x]\right) = \begin{cases} B\lambda_1\left(\left[0, B^{-\beta}x\right]\right) & \text{if } x \in \left[0, K'\right], \\ B + (1 - B)\,\lambda_2\left(\left[0, (1 - B)^{-\beta}\left(x - K'\right)\right]\right) & \text{if } x \in \left[K', K\right]. \end{cases} \tag{3.2}$$

The string of beads $([0, K], \lambda)$ is called a β-*mixed* string of beads [42]. We denote the distributions of $([0, K], \lambda)$ and $([0, K], [0, K'], \lambda)$ on $\mathbb{T}_{\mathrm{w}}$ and $\mathbb{T}_{\mathrm{w}}^{[1]}$ by ν_β and $\nu_\beta^{[1]}$, respectively.

Remark 3.1. By Proposition 2.5 with $\theta_1 = 1 - 2\beta, \theta_2 = \beta$, noting that $(B, 1 - B) \sim \mathrm{Dirichlet}(1 - 2\beta, \beta)$, we have

$$\left(B^{\beta} K_1, (1 - B)^{\beta} K_2\right) \stackrel{d}{=} L\left(B', 1 - B'\right) \tag{3.3}$$

where $B' \sim \mathrm{Beta}(1/\beta - 2, 1)$ is independent of L, and $L \sim \mathrm{ML}(\beta, 1 - \beta)$. We conclude that for each β-mixed string of beads $\xi = ([0, K], \lambda)$ we have $(\lambda(x) \colon x \in [0, K], \lambda(x) > 0)^{\downarrow} \sim \mathrm{PD}(\beta, 1 - \beta)$, cf. e.g. [41, Corollary 1.2]. Although the length of a β-mixed string of beads ξ is $\mathrm{ML}(\beta, 1 - \beta)$ and the atom sizes are $\mathrm{PD}(\beta, 1 - \beta)$, we cannot expect that ξ is a $(\beta, 1 - \beta)$-string of beads when $\beta \in (0, 1/2)$. Specifically, at the junction point in a $(\beta, 1 - \beta)$-string of beads, we would expect a $\mathrm{Beta}(\beta, 1 - 2\beta)$ 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 will use the notation $\xi = \left([0, K], \sum_{i\geq 1} P_i \delta_{X_i}\right)$ for any (α, θ)- or β-mixed string of beads where K is the length of the string of beads with ranked atomic masses of sizes $1 > P_1 > P_2 > \cdots > 0$, a.s., in the points $X_i \in [0, K]$, $i \geq 1$, respectively.

Let us now explain how to attach a weighted $\mathbb{R}$-tree onto another weighted $\mathbb{R}$-tree. This clarifies in particular how to construct weighted $\mathbb{R}$-trees by attaching strings of beads as a string of beads can be interpreted as a weighted $\mathbb{R}$-tree consisting of a single branch. For any weighted $\mathbb{R}$-tree $(\mathcal{T}, d, \rho, \mu)$, a parameter $\beta \in (0, 1/2]$, an element $J \in \mathcal{T}$ and another weighted $\mathbb{R}$-tree $(\mathcal{T}^+, d^+, \rho^+, \mu^+)$ with $\mathcal{T} \cap \mathcal{T}^+ = \emptyset$, the tree $(\mathcal{T}', d', \mu')$ created from $(\mathcal{T}, d, \mu)$ by *attaching* to J the tree $(\mathcal{T}^+, d^+, \rho^+, \mu^+)$ with mass measure μ^+ rescaled by $\mu(J)$ and metric d^+ rescaled by $\mu(J)^\beta$ is defined as follows. Set $\mathcal{T}' := \mathcal{T} \setminus \{J\} \sqcup \mathcal{T}^+$ and equip $\mathcal{T}'$ with the metric

$$d'(x, y) := \begin{cases} d(x, y) & \text{if } x, y \in \mathcal{T}, \\ d(x, J) + (\mu(J))^\beta d^+(\rho^+, y) & \text{if } x \in \mathcal{T}, y \in \mathcal{T}^+, \\ (\mu(J))^\beta d^+(x, y) & \text{if } x, y \in \mathcal{T}^+, \end{cases} \tag{3.4}$$

the root $\rho' = \rho$ and the mass measure μ' given by $\mu' \restriction_{\mathcal{T}\setminus\{J\}} = \mu \restriction_{\mathcal{T}\setminus\{J\}}$, $\mu'(J) = 0$, $\mu' \restriction_{\mathcal{T}^+} = \mu(J)\, \mu^+$.

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

$$x \sim y \qquad :\Leftrightarrow \qquad x, y \in \mathcal{R}_k^{(i)} \quad \text{for some } i \geq 1. \tag{3.5}$$

If 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 $k' \geq k$, the equivalence relation $\sim$ is consistent as k varies. This will be the case for us, by construction, and we denote the equivalence class related to $\mathcal{R}_k^{(i)}$ by $\widetilde{v}_i := [\mathcal{R}_k^{(i)}]_\sim$, and let

$$\widetilde{\mathcal{T}}_k := \mathcal{T}_k^* / \sim \tag{3.6}$$

denote the quotient space of $\mathcal{T}_k^*$, $k \geq 0$, with the canonical quotient metric. Furthermore, for $k \geq 0$, let $\widetilde{\mu}_k$ be the push-forward of μ_k^* under the projection map from $\mathcal{T}_k^*$ onto $\widetilde{\mathcal{T}}_k$.

We are now ready to present the two-colour bead-splitting construction, cf. Fig. 3 for an illustration.

Algorithm 3.2 (*Two-Colour Bead-Splitting Construction*)**.** Let $\beta \in (0, 1/2]$. We grow weighted ∞-marked $\mathbb{R}$-trees $(\mathcal{T}_k^*, (\mathcal{R}_k^{(i)}, i \geq 1), \mu_k^*)$ with associated quotient trees $(\widetilde{\mathcal{T}}_k, \widetilde{\mu}_k)$ as in (Eq. (3.6)), $k \geq 0$, as follows (cf. Fig. 3).

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

Given $(\mathcal{T}_j^*, (\mathcal{R}_j^{(i)}, i \geq 1), \mu_j^*)$, $0 \leq j \leq k$, and $r_k = \#\{i \geq 1 : \mathcal{R}_k^{(i)} \neq \{\rho\}\} = \#\{i \geq 1 : \widetilde{v}_i \neq \{\rho\}\}$,

1.–2. select $\widetilde{J}_k$ from $\widetilde{\mu}_k$; if $\widetilde{J}_k \neq \widetilde{v}_i$ for all $i \in [r_k]$, set $J_k^* = \widetilde{J}_k$ and $I_k = r_k + 1$; otherwise, if $\widetilde{J}_k = \widetilde{v}_i$ for some $i \in [r_k]$, let $I_k = i$ and 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 $J_k^* \in E_k^*$;

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 $\mathcal{R}_k^{(I_k)} \setminus \{\rho\}$ the (image under the isometry of) R_k^+ to form $\mathcal{R}_{k+1}^{(I_k)}$; set $\mathcal{R}_{k+1}^{(i)} = \mathcal{R}_k^{(i)}$, $i \neq I_k$.

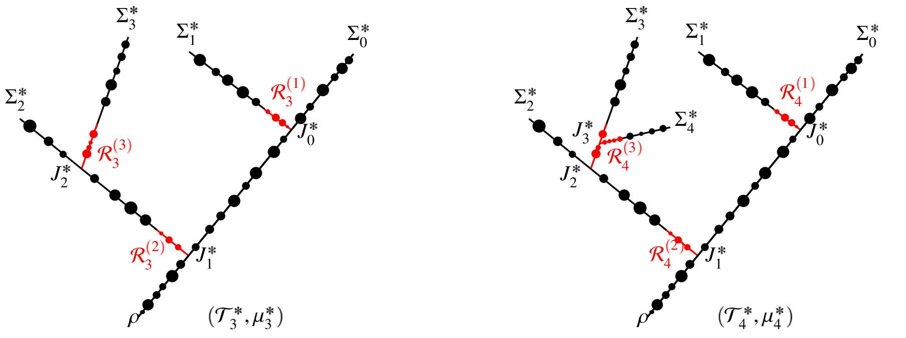

Fig. 3. Example of two-colour bead splitting. To construct $(\mathcal{T}_4^*, (\mathcal{R}_4^{(i)}, i \geq 1), \mu_4^*)$ from $(\mathcal{T}_3^*, (\mathcal{R}_3^{(i)}, i \geq 1), \mu_3^*)$, an element $J_3^* \in \mathcal{T}_3^*$ is selected, and a rescaled independent β-mixed string of beads is attached to J_3^*. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

To analyse Algorithm 3.2, 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

$$k_1^{(i)} := \inf\left\{n \geq 1 : \mathcal{R}_n^{(i)} \neq \mathcal{R}_0^{(i)}\right\} = \inf\left\{n \geq 1 : \mathcal{R}_n^{(i)} \neq \{\rho\}\right\}, \tag{3.7}$$

and, for $m \geq 1$,

$$k_{m+1}^{(i)} := \inf\left\{n \geq k_m^{(i)} : \mathcal{R}_n^{(i)} \neq \mathcal{R}_{k_m^{(i)}}^{(i)}\right\}, \tag{3.8}$$

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 the union$\bigcup_{i\geq 1}\{k_m^{(i)}, m \geq 1\} = \{1, 2, \ldots\}$ is a disjoint union, and that, for any $i \geq 1$, $\mathcal{R}_k^{(i)}$ is a binary tree for any $k \geq 1$. We will also note that $\mathcal{R}_k^{(i)} = \{\rho\}$ for $k < k_1^{(i)}$ by convention, but 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^+), i.e. $\left(E_k^+, R_k^+, \mu_k^+\right)$ is built from independent $\xi_k^{(1)}, \xi_k^{(2)}$ and γ_k in the same way as $([0, K], [0, K'], \lambda)$ is built from independent $([0, K_1], \lambda_1)$, $([0, K_2], \lambda_2)$ and B in (3.1)–(3.2).

We can describe the distribution of the tree $\mathcal{T}_k^*$ as follows.

Proposition 3.3 (*Distribution of $\mathcal{T}_k^*$*)**.** *Let $(\mathcal{T}_k^*, (\mathcal{R}_k^{(i)}, i \geq 1), \mu_k^*, k \geq 0)$ be as in Algorithm 3.2 for some $\beta \in (0, 1/2]$. The distribution of $\mathcal{T}_k^*$ is characterised by the following independent random variables:*

- *the shape T_k^* of $\mathcal{T}_k^*$ obtained from the shape $\widetilde{T}_k$ of $\widetilde{\mathcal{T}}_k$ and the shapes $R_k^{(i)}$ of $\mathcal{R}_k^{(i)}$, $i \geq 1$, are as follows;*
    - *$\widetilde{T}_k$ has the distribution of the shape of a stable tree $\mathcal{T}_k$ reduced to the first k leaves, and*

– *conditionally given that $\widetilde{T}_k$ has ℓ branch points of degrees $d_1, \ldots, d_\ell$, the shapes $R_k^{(1)}, \ldots, R_k^{(\ell)}$ are the shapes of Ford trees with $m_1 := d_1 - 2, \ldots, m_\ell := d_\ell - 2$ leaves, respectively;*

- *the total mass split between the $3k+1$ edges of $\mathcal{T}_k^*$ has a*

$$\text{Dirichlet}\,(\beta, \ldots, \beta, 1-2\beta, \ldots, 1-2\beta) \tag{3.9}$$

  distribution, with parameter β for each internal marked and each unmarked edge, and parameter $1-2\beta$ for each external marked edge with edges ordered according to depth-first search (first run for unmarked and internal marked edges, then for external marked edges);
- *the $3k+1$ independent (β, θ)-strings of beads isometric to*

$$\left(\mu_k^*(E)^{-\beta} E, \mu_k^*(E)^{-1} \mu_k^* \restriction_E\right), \quad E \in \text{Edg}\left(\mathcal{T}_k^*\right), \tag{3.10}$$

  where $\theta = 1-2\beta$ if E is an external marked edge of $\mathcal{R}_k^{(i)}$ for some $i \in [\ell]$, and $\theta = \beta$ otherwise, again listed according to depth-first search.

Proof. This proof is mainly an application of the properties of the Dirichlet distribution, Proposition 2.6, and of coin tossing sampling, Proposition 2.8. We give a brief sketch of the proof via an induction on k.

For $k = 0$, the claim is trivial as $(\mathcal{T}_0^*, \mu_0^*)$ is a (β, β)-string of beads by definition. For the induction step, suppose that the claim holds for some $k \geq 0$.

We first consider the shape transition from T_k^* to T_{k+1}^*. Observe that, given $\widetilde{T}_k$ has ℓ branch points of degrees $d_1, \ldots, d_\ell$, we have a Dirichlet $(\beta, \ldots, \beta, w(d_1), \ldots, w(d_\ell))$ mass split in $\widetilde{\mathcal{T}}_k$ with weight β for each edge and weight $w(d) = (d-2)(1-\beta) - \beta$ for each branch point of degree $d \geq 3$. Hence, the overall edge selection is as required.

Conditionally given that the ith branch point of $\widetilde{T}_k$ is selected, an edge of $\mathcal{R}_k^{(i)}$ is chosen proportionally to the weights assigned by the relative Dirichlet $(\beta, \ldots, \beta, 1-2\beta, \ldots, 1-2\beta)$ mass split in $\mathcal{R}_k^{(i)}$, so that each internal, resp. external, edge is chosen with probability $\beta/((d_i-2)(1-2\beta)+(d_i-3)\beta)$, resp. $(1-2\beta)/((d_i-2)(1-2\beta)+(d_i-3)\beta)$. This corresponds to the shape growth rule in a Ford tree growth process of index $\beta/(1-\beta)$, using obvious cancellations, cf. Algorithm 1.2 and Proposition 2.13.

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 2.6(iv), the parameter for this edge in the Dirichlet split (3.9), 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 2.8, 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 2.6(iii), the claims (3.9) and (3.10) follow, as statements conditionally given tree shapes.

Finally, these conditional distributions of the Dirichlet mass split (3.9) and the independent (β, θ)-strings of beads (3.10) do not depend on the shape T_{k+1}^*, and are hence unconditionally independent. $\square$

Remark 3.4. By Proposition 3.3 and Lemma 2.9 we see that Algorithm 3.2 reduces to Algorithm 1.3. More specifically, the claim can be proved by induction:

- For $k = 0$, the two algorithms clearly give the same tree as the (β, β)-string of beads from Algorithm 3.2 has length $S_0 \sim \mathrm{ML}(\beta, \beta)$ (as in Algorithm 1.3)
- Now assume that the trees obtained by the two algorithms are the same for some $k \geq 0$. By Proposition 3.3, there is a Dirichlet$(\beta, \ldots, \beta, 1-2\beta, \ldots, 1-2\beta)$ mass split between the edges of $\mathcal{T}_k^*$ with a parameter of β for each internal marked and each unmarked edge (which represent rescaled (β, β)-strings of beads), and a parameter of $1 - 2\beta$ for each external marked edge (which represent rescaled $(\beta, 1-2\beta)$-strings of beads). By Lemma 2.9 selecting the attachment point via sampling from the mass-measure in combination with coin tossing (as in Algorithm 3.2) is the same as selecting the attachment point from the length measure in combination with a suitable Beta-variable (as in Algorithm 1.3). Further note that the attached β-mixed string of beads in Algorithm 3.2 has the same length as in Algorithm 1.3 given that the total lengths are the same at step k and evolve in the same way (see Proposition 2.5).

3.2. Two-colour bead-splitting construction using a given stable tree

We will identify Ford trees by carrying out the two-colour bead-splitting construction using a given stable tree $(\mathcal{T}, \mu)$ equipped with a sequence of i.i.d. leaves $(\Sigma_k, k \geq 0)$ sampled from μ, and i.i.d. sequences of i.i.d. ordered $(\beta', 1-\beta')$-Chinese restaurant processes $(\widetilde{\Pi}_n^{(i,m)}, n \geq 1)$, $i \geq 1$, $m \geq 1$, cf. Section 2.4.

Definition 3.5 (*Labelled Bead Tree/string of Beads*)**.** A pair (x, Λ) is called a *labelled bead* if x is a point in a metric space and $\Lambda \subset \mathbb{N}$ is an infinite label set. A weighted $\mathbb{R}$-tree $(\mathcal{R}, \mu_{\mathcal{R}})$ equipped with a point process $\mathcal{P}_{\mathcal{R}} = \sum_{i\geq 1} \delta_{(x_i, \Lambda_i)}$ on some countable subset $\{x_i, i \geq 1\} \subset \mathcal{R}$, $x_i \neq x_j, i \neq j$, is called a *labelled bead tree* if (x_i, Λ_i) is a labelled bead for every $i \geq 1$. If $(\mathcal{R}, \mu_{\mathcal{R}})$ is a string of beads we call $(\mathcal{R}, \mu_{\mathcal{R}}, \mathcal{P}_{\mathcal{R}})$ a *labelled string of beads*.

We will also speak of *labelled* (α, θ)*-strings of beads* for $\alpha \in (0, 1)$, $\theta > 0$, as induced by an ordered (α, θ)-Chinese restaurant process. Specifically, the label sets are the blocks $\Pi_{\infty,i}$, $i \geq 1$, of the limiting partition of $\mathbb{N}$, which we relabel by $\mathbb{N} \setminus \{1\}$ using the increasing bijection $\mathbb{N} \to \mathbb{N} \setminus \{1\}$. The locations X_i are the locations of the corresponding atom of size P_i on the string, $i \geq 1$. A Ford tree growth process of index $\beta' \in (0, 1)$ as in Algorithm 1.2 can be represented in terms of labelled $(\beta', 1-\beta')$-strings of beads $\widehat{\xi}_m$, $m \geq 1$, as follows [39, Corollary 16].

Proposition 3.6 (*Ford Tree Growth Via Labelled Strings of Beads*)**.** *For $\beta' \in (0, 1)$, construct a sequence of labelled bead trees $(\mathcal{F}_m, \nu_m, \mathcal{P}_m, m \geq 1)$ as follows.*

0. *Let $(\mathcal{F}_1, \nu_1, \mathcal{P}_1) = \widehat{\xi}_0$ be a labelled $(\beta', 1-\beta')$-string of beads with label set $\mathbb{N} \setminus \{1\}$.*

Given $(\mathcal{F}_j, \nu_j, \mathcal{P}_j)$, $1 \leq j \leq m$, with $\mathcal{P}_m = \sum_{i\geq 1} \delta_{(X_{m,i}, \Lambda_{m,i})}$, to construct $(\mathcal{F}_{m+1}, \nu_{m+1}, \mathcal{P}_{m+1})$,

1.–2. *select the unique $X_{m,i} \in \mathcal{F}_m$ such that $m+1 \in \Lambda_{m,i}$;*

3. *to obtain $(\mathcal{F}_{m+1}, \nu_{m+1}, \mathcal{P}_{m+1})$, remove $\nu_m(X_{m,i})\delta_{X_{m,i}}$ from ν_m and $\delta_{(X_{m,i}, \Lambda_{m,i})}$ from $\mathcal{P}_m$; attach to $\mathcal{F}_m$ at $X_{m,i}$ an independent copy $\widehat{\xi}_m$ of $\widehat{\xi}_0$ with metric rescaled by $\nu_m(X_{m,i})^{\beta'}$,*

mass measure by $\nu_m(X_{m,i})$, and label sets in $\widehat{\xi}_m$ relabelled by the increasing bijection $\mathbb{N}\setminus\{1\}\to\Lambda_{m,i}\setminus\{m+1\}$.

Then the tree growth process $(\mathcal{F}_m, m\geq 1)$ is a Ford tree growth process of index $\beta'\in(0,1)$.

It will be useful to represent two-colour trees in the space $l^1(\mathbb{N}_0^2)$ as follows. We denote by $e_{a,b}$, $a,b\geq 0$, the unit coordinate vectors. We will use $e_{k,0}$, $k\geq 0$, to embed a given stable tree $(\mathcal{T},d,\rho,\mu)$, using $e_{k,0}$ to embed Σ_k, $k\geq 0$. Indeed, from now on we assume $(\mathcal{T},d,\rho,\mu)=(\mathcal{T},d,0,\mu)\in\mathbb{T}_{\mathrm{w}}^{\mathrm{emb}}$ is this *embedded stable tree*, with embedded leaves Σ_k, $k\geq 0$. We will use $e_{m,i}$, $i\geq 1$, $m\geq 1$, to embed the mth branch of the ith marked component, so the last step of Algorithm 3.2 is:

3. let $([0,L_k],[0,L_kB_k'],\mu_k^+)$ be an independent β-mixed string of beads in the notation of (3.3); denote by M_k the size (number of leaves) of $\mathcal{R}_k^{(I_k)}$; define the scaling factor $c=\mu_k^*(J_k^*)$ and add the scaled string to $\mathcal{T}_k^*$ using new directions e_{M_k+1,I_k} and $e_{k+1,0}$ for the marked and unmarked parts:

$$\mathcal{T}_{k+1}^*:=\mathcal{T}_k^*\cup\big(J_k^*+]0,L_kB_k'c^\beta]e_{M_k+1,I_k}\big)\cup\big(J_k^*+L_kB_k'c^\beta e_{M_k+1,I_k}+]0,L_k(1-B_k')c^\beta]e_{k+1,0}\big).$$

Similarly, add the marked part to the I_k-th marked component and maintain all other components:

$$\mathcal{R}_{k+1}^{(I_k)}:=\mathcal{R}_k^{(I_k)}\cup\big(J_k^*+]0,L_kB_k'c^\beta]e_{M_k+1,I_k}\big),\qquad \mathcal{R}_{k+1}^{(i)}:=\mathcal{R}_k^{(i)},\quad i\neq I_k.$$

Finally, distribute the atom at J_k^* according to the scaled string mass: $\mu_{k+1}^*:=\mu_k^*-c\delta_{J_k^*}+\lambda_k^+$, where

$$\begin{aligned}\lambda_k^+(J_k^*+]c^\beta s,c^\beta t]e_{M_k+1,I_k})&=c\mu_k^+(]s,t]), && 0\leq s<t\leq L_kB_k',\\ \lambda_k^+(J_k^*+L_kB_k'c^\beta+]c^\beta s,c^\beta t]e_{M_k+1,I_k})&=c\mu_k^+(L_kB_k'+]s,t]), && 0\leq s<t\leq L_k(1-B_k').\end{aligned}$$

We will now formulate a modification of Algorithm 3.2 starting from a given stable tree. Let $(\mathcal{T},\mu)$ be a stable tree of index $\beta\in(0,1/2]$ and $(\Sigma_k,k\geq 0)$ an i.i.d. sequence of leaves sampled from μ. Consider the sequence of reduced weighted $\mathbb{R}$-trees $(\mathcal{T}_k,\mu_k,k\geq 0)$ where μ_k captures the masses of the connected components of $\mathcal{T}\setminus\mathcal{T}_k$ projected onto $\mathcal{T}_k$ as in (1.1). Let $(v_i,i\geq 1)$ be the sequence of branch points of $\mathcal{T}$ in order of appearance in $(\mathcal{T}_k,k\geq 0)$, and denote by $(\mathcal{S}_j^{(i)},j\geq 1)$ the subtrees of $\mathcal{T}\setminus\mathcal{T}_{k^{(i)}}$ rooted at v_i, $i\geq 1$, where $k^{(i)}=\inf\{k\geq 0\colon v_i\in\mathcal{T}_k\}$ and where indices are assigned in increasing order of least leaf labels $\min\{\ell\geq k^{(i)}\colon\Sigma_\ell\in\mathcal{S}_j^{(i)}\}$, $j\geq 1$. For $i,j\geq 1$, set $P_j^{(i)}:=\mu(\mathcal{S}_j^{(i)})$,

$$D^{(i)}:=\lim_{n\to\infty}\left(1-\sum_{j\in[n]}P_j^{(i)}/P^{(i)}\right)^{1-\beta}(1-\beta)^{\beta-1}n^\beta,\qquad\text{where }P^{(i)}:=\sum_{j\geq 1}P_j^{(i)}.\quad(3.11)$$

This yields an i.i.d. sequence of $(1-\beta)$-diversities $(D^{(i)},i\geq 1)$ with $D^{(i)}\sim\mathrm{ML}(1-\beta,-\beta)$, cf. Lemma 2.11 and (2.15). In the following algorithm, we build i.i.d. Ford trees in the branch points of the stable tree $(\mathcal{T},\mu)$ from i.i.d. labelled $(\beta',1-\beta')$-strings of beads $\widehat{\xi}_k$, $k\geq 0$, for $\beta'=\beta/(1-\beta)$. To do so, we consider two separate mass measures: the measures $(\widehat{\mu}_k,k\geq 0)$, that equal μ on (shifted) subtrees of the stable tree, and the measures $\widehat{\nu}_k$ on the Ford trees, which, restricted to each Ford tree separately, play the role of the mass measures ν_m, $m\geq 1$, in the construction in Proposition 3.6.

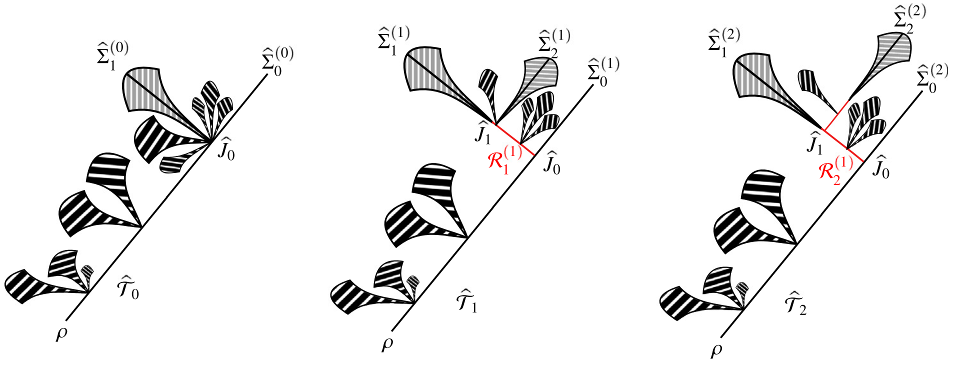

Fig. 4. Example of initial steps in Algorithm 3.7. $\widehat{\mathcal{T}}_0$ is a stable tree, shown as the spine $\mathcal{R}(\widehat{\mathcal{T}}_0, \widehat{\Sigma}_0^{(0)})$ with subtrees planted in the beads. The subtree containing leaf $\widehat{\Sigma}_1^{(0)}$ is selected and a rescaled labelled $(\beta', 1-\beta')$-string of beads is inserted (with subtree order induced by the label sets which are not shown here). Next, the marked subtree containing leaf $\widehat{\Sigma}_2^{(1)}$ is selected and a rescaled $(\beta', 1-\beta')$-string of beads is inserted, leading, in this case, to a growth step within the first marked subtree.

Algorithm 3.7 (*Algorithm* 3.2 *with Subtrees from a Given Stable Tree*)**.** Let $\beta \in (0, 1/2]$ and $\beta' = \beta/(1-\beta)$. We construct a sequence of weighted ∞-marked $\mathbb{R}$-trees $\big(\widehat{\mathcal{T}}_k, \big(\widehat{\mathcal{R}}_k^{(i)}, i \geq 1\big), \widehat{\mu}_k, \widehat{\nu}_k, \big(\widehat{\Sigma}_n^{(k)}, n \geq 0\big), k \geq 0\big)$ embedded in $l^1(\mathbb{N}_0^2)$, each equipped with a leaf sequence $(\widehat{\Sigma}_n^{(k)}, n \geq 0)$ and an additional finite measure $\widehat{\nu}_k$ as follows (cf. Fig. 4).

0. Let $(\widehat{\mathcal{T}}_0, (\widehat{\mathcal{R}}_0^{(i)}, i \geq 1), \widehat{\mu}_0, \widehat{\nu}_0, (\widehat{\Sigma}_n^{(0)}, n \geq 0)) = (\mathcal{T}, (\{\rho\}, i \geq 1), \mu, 0, (\Sigma_n, n \geq 0))$ be a labelled stable tree.

Given $(\widehat{\mathcal{T}}_j, (\widehat{\mathcal{R}}_j^{(i)}, i \geq 1), \widehat{\mu}_j, \widehat{\nu}_j, (\widehat{\Sigma}_n^{(j)}, n \geq 0))$, $0 \leq j \leq k$, let $\widehat{r}_k = \#\{i \geq 1 : \mathcal{R}_k^{(i)} \neq \{\rho\}\}$;

1.–2. let $\widehat{J}_k \in \widehat{\mathcal{T}}_k$ be the closest point to the leaf $\widehat{\Sigma}_{k+1}^{(k)}$ in the reduced tree $\mathcal{R}(\widehat{\mathcal{T}}_k, \widehat{\Sigma}_0^{(k)}, \ldots, \widehat{\Sigma}_k^{(k)})$; if $\widehat{J}_k \in \widehat{\mathcal{R}}_k^{(i)}$ for some $i \in [\widehat{r}_k]$, set $I_k = i$, otherwise let $I_k = \widehat{r}_k + 1$; denote by $M_k \geq 0$ the size of $\widehat{\mathcal{R}}_k^{(I_k)}$;

3. let $\widehat{\xi}_k$ be an independent labelled $(\beta', 1-\beta')$-string of beads with label set $\mathbb{N} \setminus \{1\}$; if $M_k \geq 1$, define the scaling factor $\widehat{c} = \widehat{\nu}_k(\widehat{J}_k)$, otherwise set $\widehat{c} = 1$; write as $([0, K_k], \nu_k, \sum_{j\geq 1} \delta_{(X_{k,j}, \Lambda_{k,j})})$ the string of beads $\widehat{\xi}_k$ with metric rescaled by the factor $\widehat{c}^{\beta'}(P^{(I_k)})^{\beta}(D^{(I_k)})^{\beta'}$ and mass measure rescaled by $\widehat{c}$, where $P^{(I_k)}$ and $D^{(I_k)}$ are as in (Eq. (3.11)); denote by $\mathcal{S}_{k,m}$, $m \in \{0, 1, 2, \ldots; \infty\}$, the connected components of $\widehat{\mathcal{T}}_k \setminus \{\widehat{J}_k\}$, where $\mathcal{S}_{k,\infty}$ contains the root and the other components are ordered by least label; set $\widehat{\mathcal{T}}_{k+1} := \mathcal{S}_{k,\infty} \cup \mathcal{S}_{k,0} \cup \big(\widehat{J}_k + [0, K_k]e_{M_k+1, I_k}\big) \cup \big(K_k e_{M_k+1,I_k} + \mathcal{S}_{k,1}\big) \cup \bigcup_{j\geq 1} \bigcup_{m \in \Lambda_{k,j}} \big(X_{k,j} e_{M_k+1,I_k} + \mathcal{S}_{k,m}\big)$; if $M_k = 0$, let $\widehat{\mathcal{R}}_{k+1}^{(I_k)} = \widehat{J}_k + [0, K_k]e_{M_k+1,I_k}$, otherwise add this shifted string to $\widehat{\mathcal{R}}_k^{(I_k)}$ to form $\widehat{\mathcal{R}}_{k+1}^{(I_k)}$; retain the other marked components, just shifted by the appropriate $X_{k,j}e_{M_k+1,I_k}$ if $\widehat{\mathcal{R}}_k^{(i)} \subset \mathcal{S}_{k,j}$. Finally, let $\widehat{\mu}_{k+1}$ denote the mass measure obtained from $\widehat{\mu}_k$ by appropriate shifting, and similarly for $\widehat{\nu}_{k+1}$, just with the scaled mass measure ν_k of $\widehat{\xi}_k$ shifted onto $\widehat{J}_k + [0, K_k]e_{M_k+1,I_k}$ replacing $\widehat{\nu}_k(\widehat{J}_k)\delta_{\widehat{J}_k}$. Denote by $\widehat{\Sigma}_n^{(k+1)}$, $n \geq 0$, the leaves $\widehat{\Sigma}_n^{(k)}$, $n \geq 0$, after the shifting operation.

Remark 3.8. Note that the scaling factor $(C^{(i)})^{-1} := (P^{(i)})^{\beta}(D^{(i)})^{\beta/(1-\beta)}$ can be rewritten as

$$(C^{(i)})^{-1} = \lim_{n\to\infty}\left(P^{(i)} - \sum_{j\in[n]} P_j^{(i)}\right)^{\beta}(1-\beta)^{-\beta}n^{\beta^2/(1-\beta)} = \lim_{n\to\infty}\left(\sum_{j\geq n+1} P_j^{(i)}\right)^{\beta}(1-\beta)^{-\beta}n^{\beta^2/(1-\beta)},$$

or, alternatively, using (2.15), as $(C^{(i)})^{-1} = \lim_{j\to\infty}(j\Gamma(\beta))^{\beta/(1-\beta)}\left(P_j^{(i)\downarrow}\right)^{\beta}$.

The labels of the labelled strings of beads $\widehat{\xi}_k$, $k\geq 0$, are used to assign the connected components of $\widehat{\mathcal{T}}_k\setminus\{\widehat{J}_k\}$ to the labelled beads of the rescaled string. The reader may want to consider these labels as auxiliary labels, which are matched with those labels of the labelled stable tree that feature as a least label of a connected component (excluding the component containing the root). This matching is induced by the increasing bijection from the set $\mathbb{N}\cup\{0\}$ of auxiliary labels to the set of least component labels. This includes two additional auxiliary labels 0 and 1, which are not assigned to a labelled bead. Indeed, while for each labelled bead $(X_{k,j}, \Lambda_{k,j})$, $j\geq 1$, and each auxiliary label $m\in\Lambda_{k,j}$, the component $\mathcal{S}_{k,m}$ is shifted by $X_{k,j}$ in the direction e_{M_k+1,I_k}, the component $\mathcal{S}_{k,1}$ is shifted by K_k, the full rescaled length of the string, while components $\mathcal{S}_{k,0}$ and the root component $\mathcal{S}_{k,\infty}$ are not shifted. See Fig. 4.

The following result follows from the construction in Algorithm 3.7 and Proposition 3.6.

Proposition 3.9. *In the setting of Algorithm 3.7, there exists a sequence of i.i.d. Ford CRTs $(\widehat{\mathcal{F}}_i, i\geq 1)$ of index $\beta' = \beta/(1-\beta)$ which is independent of the stable tree $(\mathcal{T},\mu)$ such that, for all $i\geq 1$,*

$$\lim_{k\to\infty}\widehat{\mathcal{R}}_k^{(i)} =: \widehat{\mathcal{R}}^{(i)} = \left(C^{(i)}\right)^{-1}\widehat{\mathcal{F}}_i \qquad \textit{a.s. w.r.t. the Gromov–Hausdorff topology.}$$

Proof. Almost surely, the stable tree has infinitely many branch points, all of infinite degree. This means that each of the infinitely many marked components in Algorithm 3.7 undergoes infinitely many growth steps, almost surely.

Now fix any $i\geq 1$ and consider the increasing sequence $(k_m^{(i)}, m\geq 1)$ of (3.8), i.e. let $\{k_m^{(i)}-1, m\geq 1\} = \{k\in\mathbb{N}\cup\{0\}: I_k = i\}$ contain the steps in which the ith marked component grows. For all $m\geq 1$, we enhance $\mathcal{R}_{k_m^{(i)}}^{(i)}$ to a labelled bead tree $(\mathcal{R}_{k_m^{(i)}}^{(i)}, \nu_{k_m^{(i)}}^{(i)}, \mathcal{P}_{k_m^{(i)}}^{(i)})$, as follows. We define $\nu_{k_m^{(i)}}^{(i)}$ as the restriction of $\widehat{\nu}_{k_m^{(i)}}$ to $\mathcal{R}_{k_m^{(i)}}^{(i)}$ for all $m\geq 1$, and we use the labels of labelled strings $(\widehat{\xi}_{k_m^{(i)}-1}, m\geq 1)$ like the strings $(\widehat{\xi}_m, m\geq 1)$ in Proposition 3.6 to define $\mathcal{P}_{k_m^{(i)}}^{(i)}$, $m\geq 1$. Denote by $\mathcal{S}_{m,q}^{(i)}$, $q\geq 0$, the connected components of $\widehat{\mathcal{T}}_{k_m^{(i)}}\setminus\mathcal{R}_{k_m^{(i)}}^{(i)}$, not including the component containing the root, enumerated in the order of least component labels. Inductively, we see that this ensures that for each $m\geq 1$,

- the combined set of (auxiliary!) labels in the beads of $\mathcal{P}_{k_m^{(i)}}^{(i)}$ is $\mathbb{N}\setminus[m]$;
- the components $\mathcal{S}_{m,q}^{(i)}$, $q\in[m]$, are subtrees rooted at the m leaves of $\mathcal{R}_{k_m^{(i)}}^{(i)}$;
- for each labelled bead (X,Λ) in $\mathcal{P}_{k_m^{(i)}}^{(i)}$, the components $\mathcal{S}_{m,q}^{(i)}$, $q\in\Lambda$, are rooted at X;
- leaf $\widehat{\Sigma}_{k_m^{(i)}}^{(k_m^{(i)}-1)}$ of $\widehat{\mathcal{T}}_{k_m^{(i)}-1}$ is in $\mathcal{S}_{m,m+1}^{(i)}$, which is a component of $\widehat{\mathcal{T}}_{k_m^{(i)}-1}\setminus\mathcal{R}_{k_m^{(i)}-1}^{(i)}$ rooted at $\widehat{J}_{k_m^{(i)}-1}\in\mathcal{R}_{k_m^{(i)}-1}^{(i)}$.

Therefore, the growth steps of the ith marked component are Ford growth steps as in Proposition 3.6, just with the metric scaled by $C^{(i)}$. By [39, Theorem 3], this entails the claimed convergence to Ford CRTs, for each $i \geq 1$.

Finally note that the labelled strings of beads $(\widehat{\xi}_k, k \geq 0)$ and the labelled stable tree in Algorithm 3.7 are jointly independent. This entails that the sequences $(\widehat{\xi}_{k_m^{(i)}-1}, m \geq 1)$, $i \geq 1$, are independent sequences independent of the labelled stable tree. As the limiting Ford CRTs $\mathcal{F}_i$, $i \geq 1$, are measurable functions of $(\widehat{\xi}_{k_m^{(i)}-1}, m \geq 1)$, $i \geq 1$, respectively, the claimed joint independence follows. □

We also derive some further auxiliary results about labelled (α, θ)-strings of beads.

Lemma 3.10. *Let $P = (P_i, i \geq 1) \sim \mathrm{GEM}(\alpha, \theta)$ with α-diversity S, and $\widehat{\xi} = ([0, \widehat{K}], \widehat{\mu}, \widehat{\mathcal{P}} = \sum_{j \geq 1} \delta_{(X_j, \widehat{\Lambda}_j)})$ an independent labelled $(\beta', \theta/\alpha)$-string of beads. Use $([0, \widehat{K}], \widehat{\mu}, \widehat{\mathcal{P}})$ to coagulate $(P_i, i \geq 1)$ into $\mu(\{X_j\}) := \sum_{i \in \widehat{\Lambda}_j} P_i$, with relative part sizes $Q_m^{(j)} := P_{\pi_j(m)}/\mu(\{X_j\})$, $m \geq 1$, labelled by the increasing bijection $\pi_j : \mathbb{N} \to \widehat{\Lambda}_j$, $j \geq 1$. Then*

- *the string of beads $([0, S^{\beta'} \widehat{K}], \mu)$ is an $(\alpha\beta', \theta)$-string of beads,*
- *the sequence of fragments $(Q_m^{(j)}, m \geq 1)$ has a $\mathrm{GEM}(\alpha, -\alpha\beta')$ distribution, for each $j \geq 1$,*
- *the string $([0, S^{\beta'} \widehat{K}], \mu)$ and the fragments $(Q_m^{(j)}, m \geq 1)$ of $\mu(\{X_j\})$, $j \geq 1$, are independent.*

Proof. This is an enriched instance of coagulation–fragmentation duality, see e.g. [38, Section 5.5]. We use a combinatorial approach, writing $(x)_{n\uparrow\gamma} = x(x+\gamma)\cdots(x+(n-1)\gamma)$ and using known distributions of (ordered and unordered) Chinese restaurant partitions [38,39]. Fix $n \geq 1$.

What is the probability that an ordered $(\beta', \theta/\alpha)$-coagulation groups the tables of an unordered (α, θ)-Chinese restaurant partition of $[n]$ into m groups $(n_{1,1}, \ldots, n_{1,k_1}), \ldots, (n_{m,1}, \ldots, n_{m,k_m})$? If we denote by ℓ the number of new right-most groups opened, and $(\gamma)_{j\uparrow\delta} := \gamma(\gamma+\delta)\cdots(\gamma+(j-1)\delta)$, then it is

$$\frac{(\theta+\alpha)_{k_1+\cdots+k_m-1\uparrow\alpha} \prod_{i\in[m]} \prod_{j\in[k_i]} (1-\alpha)_{n_{ij}\uparrow 1}}{(1+\theta)_{n-1\uparrow 1}} \frac{(\beta')^{m-\ell-1} (\theta/\alpha)^{\ell} \prod_{i\in[m]} (1-\beta')_{k_i-1\uparrow 1}}{(1+\theta/\alpha)_{k_1+\cdots+k_m-1\uparrow 1}}.$$

What is the probability that an unordered $(\alpha, -\alpha\beta')$-fragmentation of an ordered $(\alpha\beta', \theta)$-Chinese restaurant partition of $[n]$ yields m tables further split into $(n_{1,1}, \ldots, n_{1,k_1}), \ldots, (n_{m,1}, \ldots, n_{m,k_m})$? If we denote by ℓ the number of new right-most tables, then it is

$$\frac{(\alpha\beta')^{m-\ell-1} \theta^{\ell} \prod_{i\in[m]} (1-\alpha\beta')_{n_{i,1}+\cdots+n_{i,k_i}-1\uparrow 1}}{(1+\theta)_{n-1\uparrow 1}} \prod_{i\in[m]} \frac{(\alpha-\alpha\beta')_{k_i-1\uparrow\alpha} \prod_{j\in[k_i]} (1-\alpha)_{n_{ij}-1\uparrow 1}}{(1-\alpha\beta')_{n_{i,1}+\cdots+n_{i,k_i}-1\uparrow 1}}.$$

Elementary cancellations show that these two expressions are equal for all $n \geq 1$. Since these structured partitions can be constructed in a consistent way, as $n \geq 1$ varies, the statement of the lemma merely records different aspects of the limiting arrangement, either asymptotic frequencies in size-biased order of least labels coagulated by a labelled strings of beads, or respectively a string of beads with blocks further fragmented, with fragments in size-biased order of least labels. □

The following result can be proved using the same method.

Lemma 3.11. *Let $P = (P_i, i \geq 1) \sim \mathrm{GEM}(\alpha, \theta)$ and, for $\alpha \in (0,1)$, $\theta > 0$, let $\widehat{\Lambda} = (\widehat{\Lambda}_1, \ldots, \widehat{\Lambda}_r)$ be an independent $\mathrm{Dirichlet}(\theta_1/\alpha, \ldots, \theta_r/\alpha)$ partition of $\mathbb{N}$ with $\sum_{i\in[r]} \theta_i = \theta$. Use $(\widehat{\Lambda}_1, \ldots, \widehat{\Lambda}_r)$ to coagulate $(P_i, i \geq 1)$ into $R_j := \sum_{i\in\widehat{\Lambda}_j} P_i$, with relative part sizes $Q_m^{(j)} := P_{\pi_j(m)}/R_j$, $m \geq 1$, labelled by the increasing bijection $\pi_j : \mathbb{N} \to \widehat{\Lambda}_j$, $j \in [r]$. Then*

- *the vector $(R_1, \ldots, R_r)$ of aggregate masses has a $\mathrm{Dirichlet}(\theta_1, \ldots, \theta_r)$ distribution,*
- *the sequence of fragments $(Q_m^{(j)}, m \geq 1)$ has a $\mathrm{GEM}(\alpha, \theta_j)$ distribution, for each $j \in [r]$,*
- *the vector $(R_1, \ldots, R_r)$ and the fragments $(Q_m^{(j)}, m \geq 1)$ of R_j, $j \in [r]$, are independent.*

We will apply these results to the coagulations by Dirichlet and labelled string of beads partitions of Ford trees applied to GEM sequences of relative frequencies of stable leaf labels in subtrees.

3.3. Embedding of the two-colour bead-splitting construction into a binary compact CRT

In [42] we constructed CRTs recursively based on recursive distribution equations as reviewed by Aldous and Bandyopadhyay [4]. This method applied to a β-mixed string of beads yields a compact CRT $(\mathcal{T}^*, \mu^*)$ in which we can embed the two-colour bead-splitting construction. Let us briefly recall the recursive construction of $(\mathcal{T}^*, \mu^*)$ from [42, Proposition 4.12] including some useful notation. We only outline the constructions without going into the mathematical details for which we refer to [42].

For $\beta \in (0, 1/2]$, consider a sequence of independent strings of beads $(\xi_{\mathbf{i}}, \mathbf{i} \in \mathbb{U})$,

$$\xi_{\mathbf{i}} = \left([0, L_{\mathbf{i}}], \sum_{j\geq 1} P_{\mathbf{i}j}\delta_{X_{\mathbf{i}j}}\right), \quad \mathbf{i} \in \mathbb{U},$$

where $\xi_\varnothing$ is a (β, β)-string of beads independent of the β-mixed strings of beads $\xi_{\mathbf{i}}$, $\mathbf{i} \in \mathbb{U}\setminus\{\varnothing\}$, and $\mathbb{U} := \bigcup_{n\geq 0} \mathbb{N}^n$ is the infinite Ulam-Harris tree. Let $(\check{\mathcal{T}}_0, \check{\mu}_0) = \xi_\varnothing$, and for $n \geq 0$, conditionally given $(\check{\mathcal{T}}_n, \check{\mu}_n)$ with $\check{\mu}_n = \sum_{\mathbf{i}j\in\mathbb{N}^{n+1}} \check{P}_{\mathbf{i}j}\delta_{\check{X}_{\mathbf{i}j}}$, attach to each $\check{X}_{\mathbf{i}j}$ an isometric copy of the string of beads $\xi_{\mathbf{i}j}$

- with metric rescaled by $\check{\mu}_n(\check{X}_{\mathbf{i}j})^\beta$, and mass measure rescaled by $\check{\mu}_n(\check{X}_{\mathbf{i}j})$,
- so that the atom $P_{\mathbf{i}jk}\delta_{X_{\mathbf{i}jk}}$ of $\xi_{\mathbf{i}j}$ is scaled to become an atom of $\check{\mathcal{T}}_{n+1}$ denoted by $\check{P}_{\mathbf{i}jk}\delta_{\check{X}_{\mathbf{i}jk}}$, $k \geq 1$,

for all $\mathbf{i}j \in \mathbb{N}^{n+1}$ respectively. Denote the resulting tree by $(\check{\mathcal{T}}_{n+1}, \check{\mu}_{n+1})$.

By construction, $(\check{\mathcal{T}}_n, \check{\mu}_n)$ only carries mass in the points $\check{X}_{\mathbf{i}j}$, $\mathbf{i}j \in \mathbb{N}^{n+1}$, i.e. $\check{\mu}_n(\check{\mathcal{T}}_n \setminus \check{\mathcal{T}}_{n-1}) = 0$ for $n \geq 0$. Note that, for any $\check{X}_{i_1 i_2 \cdots i_{n+1}} \in \check{\mathcal{T}}_n$, $n \geq 0$,

$$\check{\mu}_n\left(\check{X}_{i_1 i_2\cdots i_{n+1}}\right) = \check{P}_{i_1 i_2\cdots i_{n+1}} = P_{i_1} P_{i_1 i_2} \cdots P_{i_1 i_2 \cdots i_{n+1}}.$$

This induces a recursive description of the trees $(\check{\mathcal{T}}_n, \check{\mu}_n, n \geq 0)$ via the strings of beads $(\xi_{\mathbf{i}}, \mathbf{i} \in \mathbb{U})$.

Theorem 3.12 *([42, Proposition 4.12]).* *Let $\beta \in (0, 1/2]$ and $(\check{\mathcal{T}}_n, \check{\mu}_n, n \geq 0)$ as above. Then there exists a compact CRT $(\mathcal{T}^*, \mu^*)$ such that*

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

with respect to the Gromov–Hausdorff–Prokhorov topology.

We will show that the increasing sequence $(\mathcal{T}_k^*, k \geq 0)$ of compact $\mathbb{R}$-trees from Algorithm 3.2 converges a.s. to a tree with the same distribution as $\mathcal{T}^*$. To do this and handle the marked components, we will embed the sequence of weighted ∞-marked $\mathbb{R}$-trees $(\mathcal{T}_k^*, (\mathcal{R}_k^{(i)}, i \geq 1), \mu_k^*, k \geq 0)$ into a given $(\mathcal{T}^*, \mu^*)$.

Note that the strings of beads $\xi_{\mathbf{i}}$, $\mathbf{i} \in \mathbb{U} \setminus \{\varnothing\}$, are β-mixed strings of beads as used in Algorithm 3.2 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. As we would like to embed into $(\mathcal{T}^*, \mu^*)$ the two-colour line breaking construction which carries colour marks on β-mixed strings of beads, we need to determine $I_1 = [0, K_1] \subset I = [0, K]$ such that $(I, I_1, \lambda) \sim \nu_\beta^{[1]}$ given some $\xi = (I = [0, K], \lambda) \sim \nu_\beta$, where ν_β and $\nu_\beta^{[1]}$ were introduced at the beginning of Section 3 as distributions on one-branch trees in $\mathbb{T}_{\mathrm{w}}$ and $\mathbb{T}_{\mathrm{w}}^{[1]}$, respectively. The existence of the conditional distribution of the point of the colour change K_1 given ξ is stated in the following lemma.

Lemma 3.13. *Let $\xi \sim \nu_\beta$. Then there exists a unique probability kernel κ from $\mathbb{T}_{\mathrm{w}}$ to $\mathbb{R}$ such that*

$$\mathbb{P}(K_1 \in \cdot | \xi) = \kappa(\xi, \cdot) \quad a.s.. \tag{3.12}$$

Proof. This is a special case of Theorem 6.3 in [28], since $\mathbb{R}$ is a Borel space. □

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

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

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 $\xi_{\mathbf{i}}, \mathbf{i} \in \mathbb{U}$, provides the i.i.d. strings of beads needed in Algorithm 3.2, which the colour change kernel (3.12) turns into i.i.d. 1-marked strings of beads.

Algorithm 3.14 (*Two-Colour Embedding*)**.** Let $\beta \in (0, 1/2]$. We embed into the tree $(\mathcal{T}^*, \mu^*)$ of Theorem 3.12 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 $(\overline{\mathcal{T}}_0^*, \overline{\mu}_0^*) = \xi_\varnothing$ be the initial (β, β)-string of beads; 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 $\overline{\mu}_j^* = \sum_{x \in \overline{\mathcal{T}}_j^*} \overline{\mu}_j^*(x)\delta_x$, $0 \leq j \leq k$, 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 $\mathbf{j} \in \mathbb{U}$ such that $\overline{J}_k^* = \check{X}_{\mathbf{j}}$ and $\overline{\mu}_k^*(\overline{J}_k^*) = \check{P}_{\mathbf{j}}$; sample a point $\overline{\Omega}_k$ from $\kappa(\xi_{\mathbf{j}}, \cdot)$; 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 scaled copy of the string of beads $\xi_{\mathbf{j}}$ with $\overline{\Omega}_k$ embedded in $\mathcal{T}^*$; set $\overline{\mathcal{R}}_{k+1}^{(\overline{I}_k)} = \overline{\mathcal{R}}_k^{(\overline{I}_k)} \cup [[\overline{J}_k^*, \overline{\Omega}_k]]$ and $\overline{\mathcal{R}}_{k+1}^{(i)} = \overline{\mathcal{R}}_k^{(i)}$, $i \neq \overline{I}_k$.

4. Identification of bead-splitting and line-breaking constructions

The following collection of results culminates in identifying the desired weight-length transformation, i.e. the branch point weights in Goldschmidt–Haas' stable line-breaking construction (Algorithm 1.1) are indeed as the lengths of the marked subtrees in the two-colour bead-splitting construction (Algorithm 3.2). We begin by related identifications of relevant parts of Algorithm 3.2 with Algorithm 3.7 and Algorithm 3.14.

4.1. Two-colour bead-splitting reduces to stable bead-splitting

Proposition 4.1. *Let the sequence* $(\mathcal{T}_k^*, (\mathcal{R}_k^{(i)}, i \geq 1), \mu_k^*, k \geq 0)$ *be as in Algorithm 3.2, and associate* $(\widetilde{\mathcal{T}}_k, (\widetilde{v}_i = [\mathcal{R}_k^{(i)}]_\sim, i \geq 1), \widetilde{\mu}_k, k \geq 0)$ *as in (3.6). Then the sequence of trees with mass measures from Algorithm 3.2 and (3.6) has the same distribution as the sequence in Algorithm 1.7, i.e.*

$$\left(\widetilde{\mathcal{T}}_k, \widetilde{\mu}_k, k \geq 0\right) \overset{d}{=} (\mathcal{T}_k, \mu_k, k \geq 0)\,. \tag{4.1}$$

Proof. Recall the constructions of $(\mathcal{T}_k^*, \mu_k^*)$ in Algorithm 3.2 and $(\widetilde{\mathcal{T}}_k, \widetilde{\mu}_k)$ in (3.6). We couple $(\mathcal{T}_k, \mu_k, k \geq 0)$ to $(\mathcal{T}_k^*, \mu_k^*, k \geq 0)$ and identify the distribution as required for Algorithm 1.7:

- We couple the initial (β, β)-strings of beads to be equal $(\mathcal{T}_0, \mu_0) = (\widetilde{\mathcal{T}}_0, \widetilde{\mu}_0) = (\mathcal{T}_0^*, \mu_0^*)$.
- Supposing that $(\mathcal{T}_k, \mu_k) = (\widetilde{\mathcal{T}}_k, \widetilde{\mu}_k)$ for some $k \geq 0$, set $J_k := \widetilde{J}_k = [J_k^*]_\sim$, $\xi_k = \xi_k^{(2)}$, and

  $$Q_k := (1 - \gamma_k)\, \mu_k^*(J_k^*)/\widetilde{\mu}_k(\widetilde{J}_k),$$

  where we recall that $(1 - \gamma_k) \sim \text{Beta}(\beta, 1 - 2\beta)$ is the independent scaling factor for $\xi_k^{(2)}$ in the construction of a β-mixed string of beads from $\xi_k^{(1)}$, $\xi_k^{(2)}$ and γ_k, as at the beginning of Section 3. If the selected atom J_k^* is an element of a marked component, Q_k is the proportion of the mass of J_k^* added to this marked component in the form of a rescaled independent (β, β)-string of beads $\xi_k^{(2)}$, while a proportion of $1 - Q_k$ is split into an unmarked rescaled $(\beta, 1 - 2\beta)$-string of beads $\xi_k^{(1)}$.

Since $\widetilde{J}_k$ was sampled from $\widetilde{\mu}_k$, J_k is sampled from μ_k, as required for Algorithm 1.7. It remains to check that the scaling factor $Q_k \widetilde{\mu}_k(\widetilde{J}_k)$ induced by Algorithm 3.2, applied to the (β, β)-string of beads $\xi_k = \xi_k^{(2)}$ that is used in the attachment procedure, is as needed for Algorithm 1.7. We work conditionally given the event that $\widetilde{\mathcal{T}}_k$ has ℓ branch points $\widetilde{v}_j$ of sizes $d_j = \deg(\widetilde{v}_j, \widetilde{\mathcal{T}}_k)$, $j \in [\ell]$, respectively.

- If $\widetilde{J}_k \neq \widetilde{v}_i$ for $i \in [\ell]$, then $J_k = \widetilde{J}_k = J_k^*$, and a new branch point $\widetilde{J}_k$ of degree $\deg(\widetilde{J}_k, \widetilde{\mathcal{T}}_{k+1}) = 3 = 1 + \deg(\widetilde{J}_k, \widetilde{\mathcal{T}}_k)$ is created. The mass $\mu_k^*(J_k^*) = \widetilde{\mu}_k(\widetilde{J}_k)$ is split by the independent random variable $\gamma_k \sim \text{Beta}(1 - 2\beta, \beta)$ into a branch point weight $\widetilde{\mu}_{k+1}(\widetilde{J}_k) = \gamma_k \widetilde{\mu}_k(\widetilde{J}_k)$ and the isometric copy of the (β, β)-string of beads $\xi_k^{(2)} = \xi_k$, scaled by $\widetilde{\mu}_k(\widetilde{J}_k)(1 - \gamma_k) = \widetilde{\mu}_k(\widetilde{J}_k) Q_k$ where $Q_k \sim \text{Beta}(\beta, 1 - 2\beta)$ 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 $\widetilde{J}_k = \widetilde{v}_i$ of degree $\deg(\widetilde{v}_i, \widetilde{\mathcal{T}}_k) = d_i$ for some $i \in [\ell]$, 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. By Proposition 2.8 and Proposition 2.6(iii)–(iv), conditionally given $J_k^* \in E_k^*$, the relative mass split in $\mathcal{R}_k^{(i)}$ is

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

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 2.6(i)–(ii) that, conditionally given $\widetilde{J}_k = \widetilde{v}_i$,

$$\mu_k^*(J_k^*)/\widetilde{\mu}_k(\widetilde{J}_k) \sim \text{Beta}\,(1-\beta, (d_i-2)(1-\beta))$$

is independent of $\widetilde{\mu}_k(\widetilde{J}_k)$, as the internal relative mass split in $\mathcal{R}_k^{(i)}$ is independent of its total mass, see Proposition 3.3 and Proposition 2.6(ii). Overall, still conditionally given $\widetilde{J}_k = \widetilde{v}_i$, we have that

$$\mu_k^*\left(J_k^*\right)(1-\gamma_k) = (1-\gamma_k)\left(\mu_k^*\left(J_k^*\right)\widetilde{\mu}_k\left(\widetilde{J}_k\right)^{-1}\right)\widetilde{\mu}_k\left(\widetilde{J}_k\right) = Q_k\widetilde{\mu}_k\left(\widetilde{J}_k\right)$$

where $Q_k \sim \text{Beta}(\beta, d_i(1-\beta)-1)$, as is easily checked using Proposition 2.6(i)–(iii). Note that Q_k is also conditionally independent of $\widetilde{\mu}_k(\widetilde{J}_k)$ given $\widetilde{J}_k = \widetilde{v}_i$ and $\deg(\widetilde{v}_j, \widetilde{\mathcal{T}}_k) = d_i$. 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 there are ℓ branch points $\widetilde{v}_j$ with $\deg(\widetilde{v}_j, \widetilde{\mathcal{T}}_k) = d_j$, $j \in [\ell]$. □ □

4.2. *Stable bead splitting reduces to stable line-breaking, proof of Theorem 1.8*

Proposition 4.2. *The sequence of weighted $\mathbb{R}$-trees $(\mathcal{T}_k, \mu_k, k \geq 0)$ from Algorithm 1.7 has the same distribution as the sequence of trees in (1.1) equipped with projected subtree masses, i.e. with the mass measures $(\pi_k)_*\mu$, $k \geq 1$, as in (2.4)–(2.5). Furthermore, conditionally given $|\mathcal{T}_k| = k+1+\ell$, 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*

$$\left(\mu_k\left(E_k^{(i)}\right)^{-\beta} E_k^{(i)}, \mu_k\left(E_k^{(i)}\right)^{-1} \mu_k \restriction_{E_k^{(i)}}\right), \quad i \in [k+1+\ell], \tag{4.2}$$

and the total mass distribution is

$$\left(\mu_k\left(E_k^{(1)}\right), \ldots, \mu_k\left(E_k^{(k+1+\ell)}\right), \mu_k(v_1), \ldots, \mu_k(v_\ell)\right) \sim \text{Dirichlet}\,(\beta, \ldots, \beta, w(d_1), \ldots, w(d_\ell))$$

where $v_i, i \in [\ell]$, are the branch points of $\mathcal{T}_k$ of degrees $d_i = \deg(v_i, \mathcal{T}_k)$, $i \in [\ell]$, $w(d_i) = (d_i-3)(1-\beta)+(1-2\beta)$, $i \in [\ell]$, and where we number the edges $E_k^{(i)}, i \in [k+1+\ell]$ by depth-first search.

Proof of Theorem 1.8. Gromov–Hausdorff convergence in the setting of (1.1) follows straight from the compactness of $\mathcal{T}$ and the fact that the support of μ is $\mathcal{T}$. This entails Gromov–Hausdorff–Prokhorov convergence since the approximating trees are equipped with the projections of the limiting mass measure. See e.g. [39, Lemma 17] and [36, Proposition 7]. By Proposition 4.2, this convergence also holds in the setting of Algorithm 1.7. □

Proof of Proposition 4.2. Construction (1.1) and Algorithm 1.7 use the same notation. To avoid confusion in this proof, we denote the sequence of trees of (1.1) by $(\mathcal{T}_k', \mu_k', k \geq 0)$. We will couple the construction of $(\mathcal{T}_k, \mu_k, k \geq 0)$ of Algorithm 1.7 to the given sequence $(\mathcal{T}_k', \mu_k', k \geq 0)$, specifically identifying the sequences $(J_k, k \geq 0)$ of attachment points, and $(Q_k, k \geq 0)$ of update random variables.

The coupling is as follows. Set $(\mathcal{T}_0, \mu_0) = (\mathcal{T}'_0, \mu'_0)$, and, given $(\mathcal{T}_k, \mu_k) = (\mathcal{T}'_k, \mu'_k)$ for some $k \geq 0$, set $J_k := J'_k$ where

$$J'_k := \arg\inf\left\{d(\rho, x) : x \in \mathcal{T}'_{k+1} \setminus \mathcal{T}'_k\right\},$$

let $\xi_k := \left(\mu'_{k+1}(\mathcal{T}'_{k+1} \setminus \mathcal{T}'_k)^{-\beta}\mathcal{T}'_{k+1} \setminus \mathcal{T}'_k,\ \mu'_{k+1}(\mathcal{T}'_{k+1} \setminus \mathcal{T}'_k)^{-1}\mu'_{k+1} \restriction_{\mathcal{T}'_{k+1}\setminus\mathcal{T}'_k}\right)$ and $Q_k = 1 - \mu'_{k+1}(J'_k)/\mu'_k(J'_k)$.

By Lemma 2.12, $(\mathcal{T}'_0, \mu'_0)$ is a (β, β)-string of beads, as required in Algorithm 1.7. Now assume that $(\mathcal{T}_k, \mu_k) = (\mathcal{T}'_k, \mu'_k)$ for some $k \geq 0$ with the distribution claimed in Proposition 4.2. Denote the connected components of $\mathcal{T} \setminus \mathcal{T}'_k$ by $\overline{\mathcal{S}}_j^{(i)\downarrow}$, $j \geq 1$, $i \geq 1$, completed by their root vertices $\rho_i \in \mathcal{T}'_k$, $i \geq 1$, respectively. Note that $\mu'_k(\rho_i) = \sum_{j\geq 1} \mu(\overline{\mathcal{S}}_j^{(i)\downarrow})$.

Since we sample Σ_{k+1} from the mass measure μ on $\mathcal{T}$, the conditional probability that $\Sigma_{k+1} \in \overline{\mathcal{S}}_j^{(i)\downarrow}$, given $(\mathcal{T}, \mu)$, $(\mathcal{T}'_k, \mu'_k)$ and $(\overline{\mathcal{S}}_j^{(i)\downarrow}, j \geq 1, i \geq 1)$, is $\mu(\overline{\mathcal{S}}_j^{(i)\downarrow}) = \mu'_k(\rho_i)(\mu(\overline{\mathcal{S}}_j^{(i)\downarrow})/\mu'_k(\rho_i))$, i.e. we can sample J'_k in two steps: first, select one of the atoms ρ_i of $\mathcal{T}'_k$ proportionally to $\mu'_k(\rho_i)$, and second, select one of the components $\overline{\mathcal{S}}_j^{(i)\downarrow}$ with root ρ_i proportionally to relative mass $\mu(\overline{\mathcal{S}}_j^{(i)\downarrow})/\mu'_k(\rho_i)$. By Lemma 2.11(ii) and Proposition 2.10, we further note that conditionally given $(\mathcal{T}'_k, \mu'_k)$ with $\mu'_k = \sum_{i\geq 1} \mu'_k(\rho_i)\delta_{\rho_i}$, independently for all $i \geq 1$, that $(\mu(\mathcal{S}_j^{(i)})/\mu'_k(\rho_i), j \geq 1)^\downarrow \sim \mathrm{PD}(1-\beta, (d_i-3)(1-\beta)+1-2\beta)$, with $d_i = \deg(\rho_i, \mathcal{T}'_k)$.

We have $J'_k = \rho_i$ with probability $\mu'_k(\rho_i)$, and hence J_k is sampled from μ_k, as required in Algorithm 1.7. By Lemma 2.11(iii), the weighted $\mathbb{R}$-trees

$$\left(\mu\left(\overline{\mathcal{S}}_j^{(i)\downarrow}\right)^{-\beta}\overline{\mathcal{S}}_j^{(i)\downarrow}, \mu\left(\overline{\mathcal{S}}_j^{(i)\downarrow}\right)^{-1}\mu \restriction_{\overline{\mathcal{S}}_j^{(i)\downarrow}}\right), \qquad j \geq 1,\ i \geq 1,$$

are independent copies of $(\mathcal{T}, \mu)$, i.e. conditionally given $\Sigma_{k+1} \in \overline{\mathcal{S}}_j^{(i)\downarrow}$, the sampling procedure of $\Sigma_{k+1} \in \overline{\mathcal{S}}_j^{(i)\downarrow}$ from $\mu(\overline{\mathcal{S}}_j^{(i)\downarrow})^{-1}\mu \restriction_{\overline{\mathcal{S}}_j^{(i)\downarrow}}$ 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 1.7.

Let us consider the distribution of Q_k. Conditionally given $\deg(J'_k, \mathcal{T}'_k) = 2$, Σ_{k+1} is a leaf of a connected component $\overline{\mathcal{S}}_j^{(i)\downarrow}$ of $\mathcal{T} \setminus \mathcal{T}'_k$ with root $\rho_i = J'_k$, which is chosen independently and proportionally to relative mass $\mu(\overline{\mathcal{S}}_j^{(i)\downarrow})/\mu'_k(\rho_i)$. As noted above, the relative mass partition above J'_k is $\mathrm{PD}(1-\beta, -\beta)$, i.e. by Proposition 2.10, $Q_k \sim \mathrm{Beta}(\beta, 1-2\beta)$, as required in Algorithm 1.7.

Conditionally given $\deg(J'_k, \mathcal{T}'_k) = d$ for some $d \geq 3$, Σ_{k+1} is a leaf of a connected component $\overline{\mathcal{S}}_j^{(i)\downarrow}$ of $\mathcal{T} \setminus \mathcal{T}'_k$ with root $\rho_i = J'_k$. Then the relative mass partition of the connected components $\mathcal{T} \setminus \mathcal{T}'_k$ with root ρ_i has distribution $\mathrm{PD}(1-\beta, (d-3)(1-\beta)+1-2\beta)$ where we note that J'_k must have been selected $d-2$ times up step k in order to obtain $\deg(J'_k, \mathcal{T}'_k) = d$. Therefore, by Proposition 2.10, conditionally given $\deg(J'_k, \mathcal{T}'_k) = d$, we have $Q_k \sim \mathrm{Beta}(\beta, (d-3)(1-\beta)+1-2\beta)$, as required in Algorithm 1.7. Also, by Proposition 2.10, Q_k is conditionally independent of $\mu'_k(J'_k)$ given $\deg(J'_k, \mathcal{T}'_k) = d$. The mass split in $(\mathcal{T}'_{k+1}, \mu'_{k+1})$ is easily found from Proposition 2.6, cf. the proof of Proposition 3.3 for a similar elementary Dirichlet argument. □

We also record the following consequence of Algorithm 1.7 and Proposition 4.2.

Corollary 4.3. *Let $(\mathcal{T}, \mu)$ be a stable tree of index $\beta \in (0, 1/2]$, and let $(\mathcal{T}_k, k \geq 0)$ be as in (1.1) with branch points $(v_i, i \geq 1)$ in the order of their appearance in $(\mathcal{T}_k, k \geq 0)$. Let $k_i := \inf\{k \geq 0 : [[\rho, \Sigma_k]] \cap [[\rho, v_i]] = [[\rho, v_i]]\}$ and let $(\mathcal{S}_j^{(i)}, j \geq 1)$ be the subtrees of*

$\mathcal{T} \setminus [[\rho, \Sigma_{k_i}]]$ rooted at v_i in increasing order of smallest leaf labels, $i \geq 1$. Set $P_j^{(i)} := \mu(\mathcal{S}_j^{(i)})$ and $P^{(i)} = \sum_{j \geq 1} \mu(S_j^{(i)})$, $i \geq 1$. Then the sequences $(P_j^{(i)}/P^{(i)}, j \geq 1), i \geq 1$, are i.i.d. with distribution $\mathrm{GEM}(1-\beta, -\beta)$.

Proof. This is a consequence of the stick-breaking representation (2.12) of $\mathrm{GEM}(1-\beta, -\beta)$ and the random variables $(Q_k, k \geq 0)$ splitting branch point mass into subtrees from Algorithm 1.7. Specifically, conditionally given the branch point degrees in the sequence $(\mathcal{T}_k, k \geq 0)$, for each branch point v_i, we can find a sequence of random variables $(Q_m^{(i)}, m \geq 1)$ such that

$$P_j^{(i)} = \mu_{k_1^{(i)}-1}(v_i)\, Q_j^{(i)} \prod_{m \in [j-1]} \left(1 - Q_m^{(i)}\right), \quad j \geq 1,$$

where $Q_m^{(i)} := Q_{k_m^{(i)}} \sim \mathrm{Beta}(\beta, m(1-\beta) - \beta)$ and $k_m^{(i)} = \inf\{k \geq 1 : \deg(v_i, \mathcal{T}_k) = m+1\}$. Note that, for $m_1, \ldots, m_i \geq 1$, the random variables $Q_j^{(i)}$, $j \in [m_i]$, $i \geq 1$, have conditional distributions given $k_j^{(i)}$, $j \in [m_i]$, $i \geq 1$, that do not depend on $k_j^{(i)}$, $j \in [m_i]$, $i \geq 1$, and are hence unconditionally independent. $\square$

4.3. Two-colour bead splitting from a given stable tree reduces to two-colour bead splitting

Proposition 4.4. *Let $(\widehat{\mathcal{T}}_k, (\widehat{\mathcal{R}}_k^{(i)}, i \geq 1), \widehat{\mu}_k, (\widehat{\Sigma}_n^{(k)}, n \geq 0), k \geq 0)$ be as in Algorithm 3.7 and $(\mathcal{T}_k^*, (\mathcal{R}_k^{(i)}, i \geq 1), \mu_k^*, k \geq 0)$ as in Algorithm 3.2, and consider the projection $\widehat{\pi}_k : \widehat{\mathcal{T}}_k \to \mathcal{R}(\widehat{\mathcal{T}}_k, \widehat{\Sigma}_0^{(k)}, \ldots, \widehat{\Sigma}_k^{(k)})$ as in (2.4). Then,*

$$\left(\mathcal{R}\left(\widehat{\mathcal{T}}_k, \widehat{\Sigma}_0^{(k)}, \ldots, \widehat{\Sigma}_k^{(k)}\right), \left(\widehat{\mathcal{R}}_k^{(i)}, i \geq 1\right), (\widehat{\pi}_k)_* \widehat{\mu}_k, k \geq 0\right) \stackrel{d}{=} \left(\mathcal{T}_k^*, \left(\mathcal{R}_k^{(i)}, i \geq 1\right), \mu_k^*, k \geq 0\right). \tag{4.3}$$

Furthermore, $(P_j^{(x)}, j \geq 1)$ with $P_j^{(x)} := \widehat{\mu}_k\left(\mathcal{S}_j^{(x)}\right) / \sum_{\ell \geq 1} \widehat{\mu}_k\left(\mathcal{S}_\ell^{(x)}\right)$, $j \geq 1$, are i.i.d. $\mathrm{GEM}(1-\beta, -\beta)$ *for all $x \in \mathcal{R}(\widehat{\mathcal{T}}_k, \widehat{\Sigma}_0^{(k)}, \ldots, \widehat{\Sigma}_k^{(k)})$ with $(\widehat{\pi}_k)_* \widehat{\mu}_k(x) > 0$, where we denoted by $(\mathcal{S}_j^{(x)}, j \geq 1)$ the connected components of $\widehat{\mathcal{T}}_k \setminus \mathcal{R}(\widehat{\mathcal{T}}_k, \widehat{\Sigma}_0^{(k)}, \ldots, \widehat{\Sigma}_k^{(k)})$ rooted at $x \in \mathcal{R}(\widehat{\mathcal{T}}_k, \widehat{\Sigma}_0^{(k)}, \ldots, \widehat{\Sigma}_k^{(k)})$, ranked in increasing order of least leaf labels.*

The following is a direct consequence of Proposition 4.4.

Corollary 4.5. *In Algorithm 3.2, the tree growth processes $\left(C^{(i)} \mathcal{R}_{k_m^{(i)}}^{(i)}, m \geq 1\right)$, $i \geq 1$, are i.i.d. Ford tree growth processes of index $\beta' = \beta/(1-\beta)$ independent of the stable tree $(\mathcal{T}, \mu) = \lim_{k \to \infty}(\widetilde{\mathcal{T}}_k, \widetilde{\mu}_k)$ of Corollary 4.8, where the scaling factors $(C^{(i)})^{-1} = (P^{(i)})^\beta (D^{(i)})^{\beta/(1-\beta)}$, $i \geq 1$, are as in Remark 3.8 and the random subsequences $(k_m^{(i)}, m \geq 1)$, $i \geq 1$, are defined as in (3.8).*

Proof of Proposition 4.4. As the families of weighted discrete ∞-marked $\mathbb{R}$-trees in (4.3), suitably represented, are consistent and at step k uniquely determine the trees at all previous steps $0, \ldots, k-1$, it suffices to show that for fixed $k \geq 0$

$$\left(\mathcal{R}\left(\widehat{\mathcal{T}}_k, \widehat{\Sigma}_0^{(k)}, \ldots, \widehat{\Sigma}_k^{(k)}\right), \left(\widehat{\mathcal{R}}_k^{(i)}, i \geq 1\right), (\widehat{\pi}_k)_* \widehat{\mu}_k\right) \stackrel{d}{=} \left(\mathcal{T}_k^*, \left(\mathcal{R}_k^{(i)}, i \geq 1\right), \mu_k^*\right). \tag{4.4}$$

We will prove (4.4) by induction on k, showing that the LHS follows the characterisation of the distribution of the two-colour tree on the RHS given in Proposition 3.3. The case $k = 0$ follows from Lemma 2.12 in combination with Corollary 4.3.

For general $k \geq 0$, we obtain the shape T_k of a stable tree $\mathcal{T}_k$ reduced to the first $k + 1$ leaves from the stable tree growth processes with masses naturally embedded in Algorithm 3.7, and conditionally given its shape with ℓ branch points $v_1, \ldots, v_\ell$ of degrees $d_1, \ldots, d_\ell$, a Dirichlet$(\beta, \ldots, \beta, m_1 + (1 - 2\beta), \ldots, m_\ell + (1 - 2\beta))$ mass split between edges and branch points as in Proposition 4.2 where $m_i := d_i - 2$, $i \in [\ell]$. We further obtain rescaled independent (β, β)-strings of beads on the branches of the stable tree, i.e. the unmarked branches of $\mathcal{R}(\widehat{\mathcal{T}}_k, \widehat{\Sigma}_0^{(k)}, \ldots, \widehat{\Sigma}_k^{(k)})$, cf. Proposition 4.2 and Lemma 2.12.

From the stick-breaking representation (2.12) of GEM$(\cdot, \cdot)$ and Algorithm 1.7, the relative masses of the subtrees of $\mathcal{T} \setminus \mathcal{T}_k$ rooted at v_i indexed in increasing order of smallest leaf labels form an infinite vector with distribution GEM$(1 - \beta, m_i(1 - \beta) + (1 - 2\beta))$, independently for each branch point, $i \in [\ell]$.

From the independent Ford tree growth processes via labelled strings of beads built from the $(\widehat{\xi}_k, k \geq 0)$ in Algorithm 3.7, we have the shapes of conditionally independent Ford trees with $m_1, \ldots, m_\ell$ leaves, and for each Ford tree conditionally given the shape, independently a Dirichlet$(\beta', \ldots, \beta', 1-\beta', \ldots, 1-\beta')$ partition of $\mathbb{N}$ obtained by relabelling the edge-partition of labels $\mathbb{N} \setminus [m_i]$ by the increasing bijection $\mathbb{N} \setminus [m_i] \to \mathbb{N}$. These partitions are further split on each internal edge by a labelled (β', β')-string of beads, and on each external edge by a labelled $(\beta', 1 - \beta')$-string of beads, again all labelled by $\mathbb{N}$ and obtained by increasing bijections from $\mathbb{N}$ to the label sets of the edges.

We apply Lemma 3.11 with P as the GEM$(1 - \beta, m_i(1 - \beta) + (1 - 2\beta))$ split into further subtree masses of the ith marked component and $\widehat{\Lambda}$ as the Dirichlet$(\beta', \ldots, \beta', 1-\beta', \ldots, 1-\beta')$ partition of marked Ford labels in the ith component. We note that we eventually place subtrees in their size-biased order in P into the further Ford leaves of the ith component. Therefore, the coagulation of Lemma 3.11 produces a Dirichlet$(\beta, \ldots, \beta, 1-2\beta, \ldots, 1-2\beta)$ mass split onto the edges and independent GEM$(1 - \beta, \beta)$ and GEM$(1 - \beta, 1-2\beta)$ sequences of fragments of these edge masses.

We apply Lemma 3.10 for each edge, with P as the GEM$(1-\beta, \beta)$ or GEM$(1-\beta, 1-2\beta)$ sequence of fragments and with the labelled (β', β')- or $(\beta', 1 - \beta')$-string of beads as $\widehat{\xi}$, independent. Again, we note that we eventually place subtrees in their size-biased order in P according to the positions of the labels in the labelled string of beads. Therefore, the coagulation of Lemma 3.10 produces a mass split according to a (β, β)- or $(\beta, 1 - 2\beta)$-string of beads, respectively.

We obtain two-colour shapes as needed for the distribution of the RHS of (4.4) characterised in Proposition 3.3. Conditionally given the two-colour shape, we obtain independent Dirichlet splits onto edges that combine to a Dirichlet$(\beta, \ldots, \beta, 1-2\beta, \ldots, 1-2\beta)$ split, with parameters β for unmarked and marked internal edges and $1 - 2\beta$ for marked external edges. Again conditionally given the two-colour shape, we obtain, independently of the Dirichlet splits, for each unmarked and marked internal edge an independent (β, β)-string of beads, and for each marked external edge a $(\beta, 1 - 2\beta)$-string of beads. If we arrange the edges in the tree shape suitably by depth first search and sort the Dirichlet vectors and the vectors of strings accordingly, their joint conditional distribution does not depend on the two-colour shape, so the two-colour shape, the overall Dirichlet split and the strings of beads are jointly independent.

Finally, Algorithm 3.7 scales the strings of beads. We write $(P^{(i)})^\beta (D^{(i)})^{\beta'} = (D^{(i)}_{m_i})^{\beta'} (P^{(i)}_{(m_i)})^\beta$, where $D^{(i)}_{m_i}$ is the $(1-\beta)$-diversity of P in the application of Lemma 3.11, independent of the total mass $P^{(i)}_{(m_i)} = \sum_{j\ge m_i+1} P^{(i)}_j$ on the ith component, which is further split according to the Dirichlet distribution found above, as required. Altogether, the distribution is the same as in Proposition 3.3. □

4.4. Embedded two-colour bead splitting reduces to two-colour bead splitting

Proposition 4.6. *The sequences of trees constructed in Algorithm 3.2 and Algorithm 3.14 have the same distribution, i.e.* $(\mathcal{T}^*_k, (\mathcal{R}^{(i)}_k, i\ge 1), \mu^*_k, k\ge 0) \overset{d}{=} (\overline{\mathcal{T}}^*_k, (\overline{\mathcal{R}}^{(i)}_k, i\ge 1), \overline{\mu}^*_k, k\ge 0)$.

Proof. Similarly to the proof of Proposition 4.1, let us couple so that the initial weighted ∞-marked $\mathbb{R}$-trees coincide, i.e. let $(\mathcal{T}^*_0, (\mathcal{R}^{(i)}_0, i\ge 1), \mu^*_0) := (\overline{\mathcal{T}}^*_0, (\overline{\mathcal{R}}^{(i)}_0, i\ge 1), \overline{\mu}^*_0)$. Then, $(\mathcal{T}^*_0, \mu^*_0)$ is a (β,β)-string of beads, and $\mathcal{R}^{(i)}_0 = \{\rho\}$ for all $i\ge 1$, as required for Algorithm 3.2.

Supposing that $(\mathcal{T}^*_k, (\mathcal{R}^{(i)}_k, i\ge 1), \mu^*_k) = (\overline{\mathcal{T}}^*_k, (\overline{\mathcal{R}}^{(i)}_k, i\ge 1), \overline{\mu}^*_k)$ for some $k\ge 0$, set $J^*_k := \overline{J}^*_k$, $I_k := \overline{I}_k$, and if $\overline{J}^*_k = \check{X}_{\mathbf{i}j}$, take as (E^+_k, μ^+_k) the scaled copy of $\xi_{\mathbf{i}j}$ embedded in $\mathcal{T}^*$ and $R^+_k = [[\overline{J}^*_k, \overline{\Omega}_k]]$. We need to check that the induced update step from $(\mathcal{T}^*_k, (\mathcal{R}^{(i)}_k, i\ge 1), \mu^*_k)$ to $(\mathcal{T}^*_{k+1}, (\mathcal{R}^{(i)}_{k+1}, i\ge 1), \mu^*_{k+1})$ is as required in Algorithm 3.2. Selecting $\overline{J}^*_k$ in Algorithm 3.14, 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, and since $\mu^*_k = \overline{\mu}^*_k$, this means that J^*_k is sampled precisely as required for Algorithm 3.2. In particular we have $\mu^*_k(J^*_k) = \overline{\mu}^*_k(\overline{J}^*_k)$. Furthermore, (E^+_k, R^+_k, μ^+_k) is an independent β-mixed string of beads, as it is obtained from $\xi_{\mathbf{i}j}$ and the transition kernel $\kappa(\xi_{\mathbf{i}j}, \cdot)$ of Lemma 3.13. Therefore,

$$\left(\left(\mathcal{T}^*_k, \left(\mathcal{R}^{(i)}_k, i\ge 1\right), \mu^*_k\right), \left(\mathcal{T}^*_{k+1}, \left(\mathcal{R}^{(i)}_{k+1}, i\ge 1\right), \mu^*_{k+1}\right)\right)$$

has the same distribution as $\big((\overline{\mathcal{T}}^*_k, (\overline{\mathcal{R}}^{(i)}_k, i\ge 1), \mu^*_k), (\overline{\mathcal{T}}^*_{k+1}, (\overline{\mathcal{R}}^{(i)}_{k+1}, i\ge 1), \overline{\mu}^*_{k+1})\big)$, which proves Proposition 4.6, as both Algorithm 3.2 and Algorithm 3.14 specify Markov chains. □

4.5. Two-colour bead splitting reduces to stable line-breaking

Theorem 4.7. *In the setting of Proposition 4.1, the sequence of trees with marked component lengths from Algorithm 3.2 and (3.6) has the same distribution as the sequence of trees with weights from Algorithm 1.1, i.e.*

$$\left(\widetilde{\mathcal{T}}_k, \left(\widetilde{W}^{(i)}_k, i\ge 1\right), k\ge 0\right) \overset{d}{=} \left(\mathcal{T}_k, \left(W^{(i)}_k, i\ge 1\right), k\ge 0\right), \tag{4.5}$$

where $\widetilde{W}^{(i)}_k = \mathrm{Leb}(\mathcal{R}^{(i)}_k)$ *is the length of* $\mathcal{R}^{(i)}_k, i\ge 1$*, respectively. In particular, letting* $S^*_k = \mathrm{Leb}(\mathcal{T}^*_k)$ *denote the length of* $\mathcal{T}^*_k$*, the sequence* $(S^*_k, k\ge 0)$ *is a Mittag-Leffler Markov chain starting from* $\mathrm{ML}(\beta,\beta)$*, i.e.* $\left(S^*_k, k\ge 0\right) \overset{d}{=} (S_k, k\ge 0)$.

Corollary 4.8. *In the setting of Proposition 4.1,* $\lim_{k\to\infty}(\widetilde{\mathcal{T}}_k, \widetilde{\mu}_k) = (\mathcal{T}, \mu)$ *a.s. with respect to the Gromov–Hausdorff–Prokhorov distance, where* $(\mathcal{T}, \mu)$ *is a stable tree of index* β.

Proof. Goldschmidt and Haas [20] showed this for the RHS of (4.5), in the Gromov–Hausdorff sense, so this also holds for the LHS. This implies Gromov–Hausdorff–Prokhorov convergence, as $\widetilde{\mu}_k$ is the projection of μ to $\widetilde{\mathcal{T}}_k$. See e.g. [39, Lemma 17]. □

Proof of Theorem 4.7. Recall that the ingredients in Algorithm 1.1 to construct the sequence on the RHS of (4.5) are the Mittag-Leffler Markov chain $(S_k, k \geq 0)$, attachment points $(J_k, k \geq 0)$, and i.i.d. random variables B_k, $k \geq 0$, with $B_1 \sim \mathrm{Beta}(1, 1/\beta - 2)$. We recover these ingredients from the random variables incorporated in the construction of the LHS of (4.5) via the following coupling.

- Set $S_0 = S_0^*$, i.e. $S_0 \sim \mathrm{ML}(\beta, \beta)$ is the length of the initial (β, β)-string of beads $(\mathcal{T}_0^*, \mu_0^*) = (\widetilde{\mathcal{T}}_0, \widetilde{\mu}_0)$. For $k \geq 0$, set S_k equal to the total length of $\mathcal{T}_k^*$, i.e. $S_k = S_k^*$.
- Set $(J_k, k \geq 0) = (\widetilde{J}_k, k \geq 0)$.
- 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^+, μ_k^+) built from $\xi_k^{(1)}$, $\xi_k^{(2)}$ and γ_k. By Remark 3.1, $(B_k, k \geq 0)$ is an i.i.d. sequence with $B_1 \sim \mathrm{Beta}(1, 1/\beta - 2)$, as required.

We will show that

$$\left(\widetilde{\mathcal{T}}_k, \left(\widetilde{W}_j^{(i)}, 0 \leq j \leq k, i \geq 1\right)\right) \stackrel{d}{=} \left(\mathcal{T}_k, \left(W_j^{(i)}, 0 \leq j \leq k, i \geq 1\right)\right) \tag{4.6}$$

for all $k \geq 0$, which implies (4.5) as the families of trees $(\widetilde{\mathcal{T}}_k, k \geq 0)$ and $(\mathcal{T}_k, k \geq 0)$ are consistent, i.e. given the tree $\mathcal{T}_k$ at step k, we can recover the previous steps $\mathcal{T}_{k-1}, \ldots, \mathcal{T}_0$ of the tree sequence.

We prove (4.6) by induction on k. For $k = 0$ the claim is trivial. Suppose that (4.6) holds up to k. 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 $\mathcal{R}_k^{(i)} \neq \{\rho\}$, $i \in [\ell]$, of sizes $d_i - 2$, $i \in [\ell]$, respectively, or, in other words, that $\widetilde{\mathcal{T}}_k$ has ℓ branch points $\widetilde{v}_i$, $i \in [\ell]$, of degrees d_i, $i \in [\ell]$, respectively, and a total of $k + \ell + 1$ edges. By Propositions 4.1 and 4.2, the total mass split in $\widetilde{\mathcal{T}}_k$ is

$$\left(\widetilde{\mu}_k\left(E_k^{(1)}\right), \ldots, \widetilde{\mu}_k\left(E_k^{(k+\ell+1)}\right), \widetilde{\mu}_k(\widetilde{v}_1), \ldots, \widetilde{\mu}_k(\widetilde{v}_\ell)\right) \sim \mathrm{Dirichlet}(\beta, \ldots, \beta, w(d_1), \ldots, w(d_\ell)) \tag{4.7}$$

where $w(d_i) = (d_i - 3)(1 - \beta) + 1 - 2\beta$ for $i \in [\ell]$. We denote the edge lengths and the branch point weights in $\widetilde{\mathcal{T}}_k$ by

$$\widetilde{L}_k = \left(\widetilde{L}_k^{(1)}, \ldots, \widetilde{L}_k^{(k+\ell+1)}\right), \quad \widetilde{W}_k = \left(\widetilde{W}_k^{(1)}, \ldots, \widetilde{W}_k^{(\ell)}\right), \tag{4.8}$$

and, on the event $\{\widetilde{J}_k \in E_k^{(j)}\}$, we denote by $\widetilde{V}_k$ the length proportion of $E_k^{(j)}$ below $\widetilde{J}_k$. We use corresponding notation in $\mathcal{T}_k$. We will show that the joint distributions of edge lengths, weights and the selected attachment points $\widetilde{J}_k$ and J_k in $\widetilde{\mathcal{T}}_k$ and $\mathcal{T}_k$, respectively, are the same in Algorithm 3.2 and Algorithm 1.1, i.e. for any $k \geq 0$, and any continuous and bounded functions $f : \mathbb{R}^{k+2\ell+1} \to \mathbb{R}$ and $g : [0, 1] \to \mathbb{R}$,

$$\mathbb{E}\left[f\left(\widetilde{L}_k, \widetilde{W}_k\right) \mathbb{1}_{\{\widetilde{J}_k \in E_k^{(j)}\}} g(\widetilde{V}_k)\right] = \mathbb{E}\left[f\left(L_k, W_k\right) \mathbb{1}_{\{J_k \in E_k^{(j)}\}} g(V_k)\right] \qquad \text{for any } j \in [k + \ell + 1], \tag{4.9}$$

$$\text{and} \qquad \mathbb{E}\left[f\left(\widetilde{L}_k, \widetilde{W}_k\right) \mathbb{1}_{\{\widetilde{J}_k = v_j\}}\right] = \mathbb{E}\left[f\left(L_k, W_k\right) \mathbb{1}_{\{J_k = v_j\}}\right] \qquad \text{for any } j \in [\ell]. \tag{4.10}$$

Then, together with the coupling, this completes the induction step. It remains to prove (4.9) and (4.10).

- *Proof of* (4.9). Fix some $j \in [k+\ell+1]$, and consider the LHS of (4.9) first. Conditioning on $\widetilde{J}_k \in E_k^{(j)}$, and using the mass split (4.7) and Proposition 2.6(iv), we obtain

$$\mathbb{E}\left[f\left(\widetilde{L}_k, \widetilde{W}_k\right) \mathbb{1}_{\{\widetilde{J}_k \in E_k^{(j)}\}} g(\widetilde{V}_k)\right] = \frac{\beta}{k+\beta} \mathbb{E}\left[f\left(\widetilde{L}_k, \widetilde{W}_k\right) g(\widetilde{V}_k) \middle| \widetilde{J}_k \in E_k^{(j)}\right].$$

By Proposition 2.6(iv) and (4.7), conditionally given $\widetilde{J}_k \in E_k^{(j)}$, the distribution of the mass split

$$\left(X_k^{(1)}, \ldots, X_k^{(j-1)}, X_k^{(j)}, X_k^{(j+1)}, \ldots, X_k^{(k+\ell+1)}, X_k^{(k+\ell+2)}, \ldots, X_k^{(k+2\ell+1)}\right) \tag{4.11}$$

with $X_k^{(i)} = \widetilde{\mu}_k(E_k^{(i)})$ for $i \in [k+\ell+1]$ and $X_k^{(i)} = \widetilde{\mu}_k(\widetilde{v}_{i-(k+\ell+1)})$ for $i \in [k+2\ell+1] \setminus [k+\ell+1]$ is

$$\text{Dirichlet}\,(\beta, \ldots, \beta, 1+\beta, \beta, \ldots, \beta, w(d_1), \ldots, w(d_\ell))\,. \tag{4.12}$$

Furthermore, still conditionally given $\widetilde{J}_k \in E_k^{(j)}$, $\widetilde{J}_k$ is an atom of mass $\widetilde{\mu}_k(\widetilde{J}_k) =: U_k^{(j)} X_k^{(j)}$ 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 $E_{k+1}^{(k+\ell+3)}$ of masses $\widetilde{\mu}_k(E_{k+1}^{(j)}) =: U_k^{(-)} X_k^{(j)}$ and $\widetilde{\mu}_k(E_{k+1}^{(k+\ell+3)}) =: U_k^{(+)} X_k^{(j)}$, respectively. By Proposition 2.8, the relative mass split on $E_k^{(j)}$ is given by

$$\left(U_k^{(-)}, U_k^{(j)}, U_k^{(+)}\right) \sim \text{Dirichlet}\,(\beta, 1-\beta, \beta)\,,$$

and is independent of $X_k^{(j)} = \widetilde{\mu}_k(E_k^{(j)})$, since, by Propositions 4.1 and 4.2, the (β, β)-string of beads

$$\left(\left(X_k^{(j)}\right)^{-\beta} E_k^{(j)}, \left(X_k^{(j)}\right)^{-1} \widetilde{\mu}_k \restriction_{E_k^{(j)}}\right)$$

is independent of the scaling factor $X_k^{(j)}$. We obtain the refined mass split

$$\left(\overline{X}_k^{(1)}, \ldots, \overline{X}_k^{(j-1)}, \overline{X}_k^{(-)}, \overline{X}_k^{(j)}, \overline{X}_k^{(+)}, \overline{X}_k^{(j+1)}, \ldots, \overline{X}_k^{(k+2\ell+1)}\right) \tag{4.13}$$

where $\overline{X}_k^{(i)} = X_k^{(i)}$, $i \in [k+2\ell+1] \setminus \{j, +, -\}$ and $\overline{X}_k^{(-)} = U_k^{(-)} X_k^{(j)}$, $\overline{X}_k^{(j)} = U_k^{(j)} X_k^{(j)}$ and $\overline{X}_k^{(+)} = U_k^{(+)} X_k^{(j)}$. By Proposition 2.6(iii), the distribution of (4.13) is

$$\text{Dirichlet}\,(\beta, \ldots, \beta, \beta, 1-\beta, \beta, \beta, \ldots, \beta, w(d_1), \ldots, w(d_\ell))\,.$$

Furthermore, the atom $\widetilde{J}_k$ induces the two rescaled independent (β, β)-strings of beads

$$\left(\left(\overline{X}_k^{(-)}\right)^{-\beta} E_{k+1}^{(j)}, \left(\overline{X}_k^{(-)}\right)^{-1} \widetilde{\mu}_k \restriction_{E_{k+1}^{(j)}}\right), \quad \left(\left(\overline{X}_k^{(+)}\right)^{-\beta} E_{k+1}^{(k+\ell+3)}, \left(\overline{X}_k^{(+)}\right)^{-1} \widetilde{\mu}_k \restriction_{E_{k+1}^{(k+\ell+3)}}\right)$$

where $\overline{X}_k^{(i)} = U_k^{(i)} X_k^{(j)}$, $i \in \{-, +\}$, i.e. the edge $E_k^{(j)}$ is split by $\widetilde{J}_k$ into parts $E_{k+1}^{(j)}$ and $E_{k+1}^{(k+\ell+3)}$ of lengths

$$\widetilde{L}_k^{(-)} = \left(U_k^{(-)} X_k^{(j)}\right)^{\beta} M_k^{(-)}, \quad \widetilde{L}_k^{(+)} = \left(U_k^{(+)} X_k^{(j)}\right)^{\beta} M_k^{(+)},$$

respectively, where $M_k^{(i)} \sim \text{ML}(\beta, \beta)$, $i \in \{-, +\}$, are independent, see Proposition 2.8. Conditionally given $\widetilde{J}_k \in E_k^{(j)}$, by (4.6) and Proposition 2.5, the weights $\widetilde{W}_k^{(i)}$ of $\widetilde{\mathcal{T}}_k$ and remaining lengths $\widetilde{L}_k^{(i)}$ are therefore

$$\widetilde{W}_k^{(i-(k+\ell+1))} = \left(X_k^{(i)}\right)^{\beta} M_k^{(i)}, \quad i \in [k+2\ell+1] \setminus [k+\ell+1], \quad \widetilde{L}_k^{(i)} = \left(X_k^{(i)}\right)^{\beta} M_k^{(i)}, \quad i \in [k+\ell+1] \setminus \{j\}.$$

for independent random variables

$$M_k^{(i)} \sim \begin{cases} \mathrm{ML}(\beta, \beta), & i \in [k+\ell+1] \setminus \{j\} \cup \{-,+\}, \\ \mathrm{ML}(\beta, w(d_{i-(k+\ell+1)})), & i \in [k+2\ell+1] \setminus [k+\ell+1]. \end{cases}$$

Also note that, by the definition of $(S_k^*, k \geq 0)$ and the attachment procedure,

$$S_{k+1}^* - S_k^* = \widetilde{\mu}_k \left(\widetilde{J}_k\right)^\beta M_k^* = \left(U_k^{(j)} X_k^{(j)}\right)^\beta M_k^*$$

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 2.5 and Proposition 2.6(i)–(ii) that $S_k^* = A_k^* S_{k+1}^* \sim \mathrm{ML}(\beta, k+\beta)$ where $S_{k+1}^* \sim \mathrm{ML}\left(\beta, k+1+\beta\right)$ and $A_k^* \sim \mathrm{Beta}(k/\beta+2, 1/\beta-1)$ are independent, and that, conditionally given $\widetilde{J}_k \in E_k^{(j)}$, we have

$$\left(\widetilde{L}_k^{(1)}, \ldots, \widetilde{L}_k^{(j-1)}, \widetilde{L}_k^{(-)}, \widetilde{L}_k^{(+)}, \widetilde{L}_k^{(j+1)}, \ldots, \widetilde{L}_k^{(k+\ell+1)}, \widetilde{W}_k^{(1)}, \ldots, \widetilde{W}_k^{(\ell)}\right) = S_{k+1}^* A_k^* \widetilde{Z}_k^{\mathrm{split}},$$

where

$$\widetilde{Z}_k^{\mathrm{split}} = \left(\widetilde{Z}_k^{(1)}, \ldots, \widetilde{Z}_k^{(j-1)}, \widetilde{Z}_k^{(-)}, \widetilde{Z}_k^{(+)}, \widetilde{Z}_k^{(j+1)}, \ldots, \widetilde{Z}_k^{(k+\ell+1)}, \widetilde{Z}_k^{(k+\ell+2)}, \ldots, \widetilde{Z}_k^{(k+2\ell+1)}\right)$$

is $\mathrm{Dirichlet}(1, \ldots, 1, 1, 1, 1, \ldots, 1, w(d_1)/\beta, \ldots, w(d_\ell)/\beta)$-distributed and independent of (S_{k+1}^*, A_k^*). Now aggregating $\widetilde{Z}_k^{(-)} + \widetilde{Z}_k^{(+)} = \widetilde{Z}_k^{(j)}$ yields a $\mathrm{Dirichlet}(1, \ldots, 1, 2, 1, \ldots, 1, w(d_1)/\beta, \ldots, w(d_\ell)/\beta)$ vector $\widetilde{Z}_k = (\widetilde{Z}_k^{(1)}, \ldots, \widetilde{Z}_k^{k+2\ell+1})$ and independent $\mathrm{Dirichlet}(1, 1)$ proportions, by Proposition 2.6(ii), so that $\widetilde{V}_k = \widetilde{Z}_k^{(-)}/\widetilde{Z}_k^{(j)} = \widetilde{L}_k^{(-)}/\widetilde{L}_k^{(j)} \sim \mathrm{Unif}(0, 1)$ is independent of $\widetilde{Z}_k$, all jointly independent from S_k^*. Hence,

$$\mathbb{E}\left[f\left(\widetilde{L}_k, \widetilde{W}_k\right) \mathbb{1}_{\{\widetilde{J}_k \in E_k^{(j)}\}} g\left(\widetilde{V}_k\right)\right] = \frac{\beta}{k+\beta} \mathbb{E}\left[f\left(S_k^* \widetilde{Z}_k\right) \middle| \widetilde{J}_k \in E_k^{(j)}\right] \mathbb{E}\left[g\left(\widetilde{V}_k\right)\right].$$

We now consider the RHS of (4.9). We condition on $J_k \in E_k^{(j)}$, and apply [20, Proposition 3.2] and Proposition 2.6(iv) to obtain

$$\begin{aligned} &\mathbb{E}\left[f\left(L_k^{(1)}, \ldots, L_k^{(k+\ell+1)}, W_k^{(1)}, \ldots, W_k^{(\ell)}\right) \mathbb{1}_{\{J_k \in E_k^{(j)}\}} g(V_k)\right] \\ &= \frac{\beta}{k+\beta} \mathbb{E}\left[f\left(S_k Z_k\right) \middle| J_k \in E_k^{(j)}\right] \mathbb{E}\left[g(V_k)\right] \end{aligned}$$

where $V_k \sim \mathrm{Unif}(0, 1)$ and $S_k \sim \mathrm{ML}(\beta, k+\beta)$ are independent and jointly independent of

$$Z_k = \left(Z_k^{(1)}, \ldots, Z_k^{(j-1)}, Z_k^{(j)}, Z_k^{(j+1)}, \ldots, Z_k^{(k+\ell+1)}, Z_k^{(k+\ell+2)}, \ldots, Z_k^{(k+2\ell+1)}\right)$$

and $Z_k \sim \mathrm{Dirichlet}(1, \ldots, 1, 2, 1, \ldots, 1, w(d_1)/\beta, \ldots, w(d_\ell)/\beta)$. Hence, we conclude (4.9).

- *Proof of* (4.10). Consider now the LHS of (4.10). We follow the lines of the proof of (4.9) and only sketch the argument. Conditionally given $\widetilde{J}_k = \widetilde{v}_j$, the mass split (4.11) has distribution

$$\mathrm{Dirichlet}\left(\beta, \ldots, \beta, w(d_1), \ldots, w(d_{j-1}), 1 + w(d_j), w(d_{j+1}), \ldots, w(d_\ell)\right). \tag{4.14}$$

Still conditionally given $\widetilde{J}_k = \widetilde{v}_j$, Algorithm 3.2 samples an atom J_k^* in the jth marked component of $\mathcal{T}_k^*$. By Propositions 2.8 and 3.3, this entails further Dirichlet splits inside the jth marked component into $2d_j - 1$ parts, one of which is further split, at J_k^*, into three parts. By Proposition 2.6(iii), this refines the Dirichlet split to

$$\text{Dirichlet}\left(\beta, \ldots, \beta, w(d_1), \ldots, w(d_{j-1}), 1-\beta, \beta, \ldots, \beta, 1-2\beta, \ldots, 1-2\beta, w(d_{j+1}), \ldots, w(d_\ell)\right), \tag{4.15}$$

first listing the $\widetilde{\mu}_k$-masses of the edges and $j-1$ branch points of $\widetilde{\mathcal{T}}_k$, then the atom $\mu_k^*(J_k^*)$, then μ_k^*-masses of d_j internal and d_j external edges and split edges of $\mathcal{R}_k^{(j)}$, then $\ell - j$ more $\widetilde{\mu}_k$-masses of branch points of $\widetilde{\mathcal{T}}_k$. Specifically, the size of this atom $\mu^*(J_k^*)$ is a proportion $U_k^{(j)} \sim \text{Beta}(1-\beta, w(d_j)+\beta)$ of $\widetilde{\mu}_k(\widetilde{v}_j)$. Denoting again by $M_k^* \sim \text{ML}(\beta, 1-\beta)$ the length of the attached independent β-mixed string of beads, we have $S_{k+1}^* - S_k^* = \mu_k^*(J_k^*)^\beta M_k^* = (U_k^{(j)}\widetilde{\mu}_k(\widetilde{J}_k))^\beta M_k^*$. The rest of the proof of (4.10) is analogous to the proof of (4.9), with each component contributing a Mittag-Leffler variable, and with the aggregation step now aggregating the entries of the Dirichlet length split corresponding to the $2d_j$ edges in the jth marked component. □

5. Proofs of the main results stated in the introduction

5.1. *Weight-length representation, Ford trees, and the proof of Theorem 1.4*

Proof of Theorem 1.4. We noted in Remark 3.4 that the sequence of two-colour trees of Algorithm 3.2 without mass measures has the same joint distribution as the sequence of two-colour trees of Algorithm 1.3. Hence, (4.5) is precisely (1.3) and it suffices to continue in the richer setting of Algorithm 3.2.

It remains to identify the marked tree growth processes $(\mathcal{R}_k^{(i)}, k \geq 1)$, $i \geq 1$, as rescaled i.i.d. Ford tree growth processes of index $\beta' = \beta/(1-\beta)$. Specifically, it follows from Proposition 4.4 and Corollary 4.5 that, in the setting of Algorithm 3.2, there exists a sequence of scaling factors $(C^{(i)}, i \geq 1)$ such that $\lim_{k\to\infty} \mathcal{R}_k^{(i)} = \mathcal{R}^{(i)}$ a.s., for all $i \geq 1$, in the Gromov–Hausdorff topology where $(C^{(i)}\mathcal{R}^{(i)}, i \geq 1)$ is a sequence of i.i.d. Ford CRTs of index $\beta' = \beta/(1-\beta)$. Furthermore, the sequence $(C^{(i)}\mathcal{R}^{(i)}, i \geq 1)$ is independent of the stable tree $(\widetilde{\mathcal{T}}, \widetilde{\mu}) = \lim_{k\to\infty}(\widetilde{\mathcal{T}}_k, \widetilde{\mu}_k)$ obtained from $(\mathcal{T}_k^*, (\mathcal{R}_k^{(i)}, i \geq 1), \mu_k^*, k \geq 0)$ as in Corollary 4.8. □

This identifies the tree growth processes $(\mathcal{R}_k^{(i)}, k \geq 1)$, $i \geq 1$, as consistent families of tree growth processes which obey the growth rules of a Ford tree growth process of index $\beta' = \beta/(1-\beta)$. Rescaling these processes to obtain i.i.d. sequences of Ford trees requires knowledge of the scaling factor which is incorporated in the limiting stable tree. It is, however, possible to approximate this scaling factor using the tree constructed up to step k only. We are further able to obtain i.i.d. marked subtree growth processes obeying the Ford growth rules (but with wrong starting lengths) applying suitable scaling.

Theorem 5.1 (*Embedded Ford trees*). *Let $(\mathcal{T}_k^*, (\mathcal{R}_k^{(i)}, i \geq 1), \mu_k^*, k \geq 0)$ as in Algorithm 3.2.*

(i) *The following normalised tree growth processes in the components, with projected μ-masses, are i.i.d.:*

$$\left(\mathcal{G}_m^{(i)}, \mu_m^{(i)}, m \geq 1\right) = \left(\mu^*_{k_1^{(i)}}\left(\mathcal{R}^{(i)}_{k_1^{(i)}}\right)^{-\beta} \mathcal{R}^{(i)}_{k_m^{(i)}}, \mu^*_{k_1^{(i)}}\left(\mathcal{R}^{(i)}_{k_1^{(i)}}\right)^{-1} \mu_k^* \restriction_{\mathcal{R}_k^{(i)}}, m \geq 1\right), \quad i \geq 1. \tag{5.1}$$

(ii) *The processes* $\big(\mu^*_{k_1^{(i)}}\big(\mathcal{R}^{(i)}_{k_1^{(i)}}\big)^{-\beta}\mathcal{R}^{(i)}_{k_m^{(i)}}, m \geq 1\big)$, *without* μ*-masses are i.i.d. Ford tree growth processes of index* $\beta' = \beta/1-\beta$ *as in Algorithm 1.2,* $i \geq 1$, *but starting from* $ML(\beta, 1-2\beta)$, *not* $ML(\beta', 1-\beta')$.

(iii) *For* $i \geq 1$, *let* $C^{(i)}_m := (1-\beta)^\beta\, m^{-\beta^2/(1-\beta)}\mu^*_{k_m^{(i)}}\big(\mathcal{R}^{(i)}_{k_m^{(i)}}\big)^{-\beta}$. *The processes* $(C^{(i)}_m\mathcal{R}^{(i)}_{k_m^{(i)}}, m \geq 1)$ *with scaling constant depending on* m, $i \geq 1$, *are i.i.d.,* $\lim_{m\to\infty} C^{(i)}_m = \big(H^{(i)}\big)^{-\beta/(1-\beta)}$ $\mu^*_{k_1^{(i)}}\big(\mathcal{R}^{(i)}_{k_1^{(i)}}\big)^{-\beta}$ *a.s., where* $H^{(i)} \sim \mathrm{ML}(1-\beta, 1-2\beta)$, *and* $\lim_{m\to\infty} C^{(i)}_m\mathcal{R}^{(i)}_{k_m^{(i)}} = \mathcal{F}^{(i)}$ *a.s. in the Gromov–Hausdorff topology where* $(\mathcal{F}^{(i)}, i \geq 1)$ *are i.i.d. Ford CRTs of index* β'.

Proof. See the appendix. □

5.2. *Convergence of two-colour trees, and the proof of Theorem 1.5*

Proposition 4.1 and Corollary 4.5 demonstrate that the two-colour bead-splitting construction combines the stable tree growth process and infinitely many rescaled subtree growth processes that build rescaled independent Ford CRTs. We show that the tree growth process $(\mathcal{T}^*_k, k \geq 0)$ converges to a compact CRT with the same distribution as the CRT $(\mathcal{T}^*, \mu^*)$ constructed in the beginning of Section 3.3, using the embedding of Algorithm 3.14 and Proposition 4.6:

Proposition 5.2 (*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 3.2. Then, there is a compact CRT* $(\mathcal{T}^*, \mu^*)$ *such that*

$$\lim_{k\to\infty} d_{\mathrm{GHP}}\left(\left(\mathcal{T}^*_k, \mu^*_k\right), \left(\mathcal{T}^*, \mu^*\right)\right) = 0 \quad \textit{a.s.} \tag{5.2}$$

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 $(\mathcal{T}^*, \mu^*)$ as in Section 3.3. Then (5.2) will follow from Proposition 4.6.

By Proposition 4.1 and Corollary 4.8, we can couple $(\overline{\mathcal{T}}^*_k, \overline{\mu}^*_k, k \geq 0)$ with a stable tree growth process $(\widetilde{\mathcal{T}}_k, \widetilde{\mu}_k) \to (\mathcal{T}, \mu)$ in such a way that $\widetilde{\mu}_k$ is a push-forward of $\overline{\mu}^*_k$. In particular,

$$\max\{\overline{\mu}^*_k(x), x \in \overline{\mathcal{T}}^*_k\} \leq \max\{\widetilde{\mu}_k(x), x \in \widetilde{\mathcal{T}}_k\} \to 0 \qquad \text{a.s.} \tag{5.3}$$

On the other hand, $\overline{\mu}^*_k$ is the pushforward of μ^* under the projection map $\overline{\pi}^*_k : \mathcal{T}^* \to \overline{\mathcal{T}}^*_k$. Now assume, for contradiction that $\overline{\bigcup_{k\geq 0}\overline{\mathcal{T}}^*_k} \neq \mathcal{T}^*$. Since all leaves are limit points of $\mathcal{T}^* \setminus \mathrm{Lf}(\mathcal{T}^*)$ and by Theorem 3.12, $\mathcal{T}^*$ is a CRT, there is $x \in \mathcal{T}^* \setminus \overline{\bigcup_{k\geq 0}\overline{\mathcal{T}}^*_k}$ such that the subtree of $\mathcal{T}^*$ above x has positive mass $c := \mu^*(\mathcal{T}^*_x) > 0$. Since $\overline{\bigcup_{k\geq 0}\overline{\mathcal{T}}^*_k}$ is path-connected, $\mathcal{T}^*_x \cap \overline{\bigcup_{k\geq 0}\overline{\mathcal{T}}^*_k} = \emptyset$, and hence all $\overline{\mu}^*_k$ must have an atom greater than c, which contradicts (5.3).

We conclude that $\overline{\bigcup_{k\geq 0}\overline{\mathcal{T}}^*_k} = \mathcal{T}^*$. Since $\mathcal{T}^*$ is compact and the union is increasing in $k \geq 0$, this implies GH-convergence. The convergence in the GHP sense follows since the mass measure $\overline{\mu}^*_k$ is the projection of μ^* onto $\overline{\mathcal{T}}^*_k$, see the proof of [39, Corollary 23] for details of this argument. □

Corollary 5.3 (*Convergence of Two-Colour Trees*). *Let* $(\mathcal{T}^*_k, (\mathcal{R}^{(i)}_k, i \geq 1), \mu^*_k, k \geq 0)$ *be the two-colour tree growth process from Algorithm 3.2 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 Corollary 4.5 with* $\lim_{k\to\infty} d^{\infty}_{\mathrm{GHP}}\big(\big(\mathcal{T}^*_k, \big(\mathcal{R}^{(i)}_k, i \geq 1\big), \mu^*_k\big), \big(\mathcal{T}^*, \big(\big(C^{(i)}\big)^{-1}\mathcal{F}^{(i)}, i \geq 1\big), \mu^*\big)\big) = 0$ *a.s.*

Proof. This is a direct consequence of Proposition 5.2 and Corollary 4.5. □

It will be convenient to use the representation of Algorithm 3.7. We note the following consequences of the construction, in the light of Proposition 5.2.

Corollary 5.4. *In the setting of Algorithm 3.7*

(i) the closure $\widehat{\mathcal{T}}$ in $l^1(\mathbb{N}_0^2)$ of the increasing union $\bigcup_{k\ge 0}\mathcal{R}(\widehat{\mathcal{T}}_k, \widehat{\Sigma}_0^{(k)}, \ldots, \widehat{\Sigma}_k^{(k)})$ is compact;

(ii) the natural projection of $\widehat{\mathcal{T}}$ onto the subspace spanned by $e_{k,0}$, $k \ge 0$, is the stable tree $\mathcal{T}$;

(iii) the natural projection of $\widehat{\mathcal{T}}$ onto the subspace spanned by $e_{m,i}$, $m \ge 1$, scaled by the scaling factor $C^{(i)}$ of Remark 3.8, is a Ford CRT for each $i \ge 1$.

Proof. (i) It follows from Propositions 4.4 and 5.2, that the closure $\widehat{\mathcal{T}}$ in $l^1(\mathbb{N}_0^2)$ of the increasing union is compact. (ii) holds by construction since all steps of Algorithm 3.7 preserve this projection property for the trees $\widehat{\mathcal{T}}_k$, $k \ge 0$. (iii) holds by Corollary 4.5 since the scaled projections of $\mathcal{R}(\widehat{\mathcal{T}}_k, \widehat{\Sigma}_0^{(k)}, \ldots, \widehat{\Sigma}_k^{(k)})$ are Ford tree growth processes whose mth growth step is for $k = k_m^{(i)}$, $m \ge 1$, $i \ge 1$. □

These two corollaries imply Theorem 1.5, the convergence of the two-colour line-breaking construction.

5.3. Branch point replacement in a stable tree, and the proof of Theorem 1.6

The aim of this section is to replace branch points of the stable tree by rescaled independent Ford CRTs. Let us denote the independent Ford tree growth processes underlying Corollary 5.4(iii) by $(\mathcal{F}_m^{(i)}, m \ge 1)$, and the Ford CRTs with leaf labels by $(\mathcal{F}^{(i)}, \Omega_m^{(i)}, m \ge 1)$, $i \ge 1$, all embedded in the appropriate coordinates. Now fix $i \ge 1$, and focus on the mth subtree of the ith branch point of $\mathcal{T}$, suppose Σ_n is its smallest label. In Algorithm 3.7, each insertion into the ith marked component shifts some subtrees of the ith branch point, and the subtree we consider stops being shifted at the mth insertion.

The branch point replacement algorithm can be viewed as a change of order of the insertions of Algorithm 3.7. The kth step of Algorithm 3.7 gets Σ_k into its final position $\widehat{\Sigma}_k^{(k)}$ by inserting one branch of a marked component. The ith step of the branch point replacement algorithm gets the smallest labelled leaf of all subtrees of the ith branch point into their final positions by making all insertions into the ith component. This amounts to shifting the mth subtree of the ith branch point by $\Omega_m^{(i)}$, $m \ge 1$.

Algorithm 5.5 (*Branch Point Replacement in the Stable Tree*)**.** We construct a sequence of weighted i-marked $\mathbb{R}$-trees $(\mathcal{B}^{(i)}, (\mathcal{R}^{(1)}, \ldots, \mathcal{R}^{(i)}), \mu^{(i)})$. Let $(\mathcal{B}^{(0)}, \mu^{(0)}) = (\mathcal{T}, \mu)$ be the embedded stable tree with leaves $\Sigma_n^{(0)} = \Sigma_n$, $n \ge 0$. For $i \ge 1$, conditionally given $(\mathcal{B}^{(i-1)}, (\mathcal{R}^{(1)}, \ldots, \mathcal{R}^{(i-1)}), \mu^{(i-1)}, (\Sigma_n^{(i-1)}, n \ge 0))$, shift the connected components $\mathcal{S}_m^{(i)}$, $m \in \{0, 1, 2, \ldots; \infty\}$, of $\mathcal{B}^{(i-1)} \setminus v_i^{(i-1)}$ of the ith branch point $v_i^{(i-1)}$:

$$\mathcal{B}^{(i)} := \mathcal{S}_\infty^{(i)} \cup \mathcal{S}_0^{(i)} \cup \left(v_i^{(i-1)} + \left(C^{(i)}\right)^{-1} \mathcal{F}^{(i)}\right) \cup \bigcup_{m\ge 1} \left(\left(C^{(i)}\right)^{-1} \Omega_m^{(i)} + \mathcal{S}_m^{(i)}\right)$$

where $\mathcal{F}^{(i)}$ is the independent Ford CRT with labelled Ford leaves $(\Omega_m^{(i)}, m \geq 1)$. Take as $\mu^{(i)}$ the measure $\mu^{(i-1)}$ shifted with each of the connected components and set $\mathcal{R}^{(i)} := \left(v_i^{(i-1)} + \left(C^{(i)}\right)^{-1} \mathcal{F}^{(i)}\right)$.

Theorem 5.6 (*Branch Point Replacement*). *The $\mathbb{R}$-trees $(\mathcal{B}^{(i)}, (\mathcal{R}^{(1)}, \ldots, \mathcal{R}^{(i)}, \{0\}, \{0\}, \ldots), \mu^{(i)})$, of Algorithm 5.5 converge in $(\mathbb{T}_{\mathrm{w}}^{\infty}, d_{\mathrm{GHP}}^{\infty})$ to a limit with the same distribution as in Corollary 5.3, i.e.*

$$\lim_{i\to\infty} d_{\mathrm{GHP}}^{\infty}\left(\left(\mathcal{B}^{(i)}, \left(\mathcal{R}^{(1)}, \ldots, \mathcal{R}^{(i)}, \{0\}, \ldots\right), \mu^{(i)}\right), \left(\mathcal{T}^*, \left(\left(C^{(i)}\right)^{-1} \mathcal{F}^{(i)}, i \geq 1\right), \mu^*\right)\right) = 0 \quad a.s..$$

Proof. We already know (by Proposition 4.1 and Theorem 1.8) that the trees have the correct distribution. To identify the weights, we provide a coupling argument. By construction, the trees spanned by the first k leaves are the same in Algorithm 3.7 and Algorithm 5.5:

$$\left(\mathcal{R}\left(\widehat{\mathcal{T}}_k, \Sigma_0^{(k)}, \ldots, \Sigma_k^{(k)}\right), \left(\widehat{\mathcal{R}}_k^{(i)}, i \geq 1\right), \widehat{\mu}_k^*, k \geq 0\right) = \left(\mathcal{B}_k^{(k)}, \left(\mathcal{U}_k^{(i)}, i \geq 1\right), \lambda_k, k \geq 0\right) \tag{5.4}$$

where $\mathcal{B}_k^{(k)} := \mathcal{R}(\mathcal{B}^{(k)}, \Sigma_0^{(k)}, \ldots, \Sigma_k^{(k)})$, $\mathcal{U}_k^{(i)} := \mathcal{R}^{(i)} \cap \mathcal{B}_k^{(k)}$, and $\lambda_k = (\pi_k^{\mathcal{B}})_* \mu^{(k)}$ denotes the projected mass measure.

By Proposition 4.4 and Corollary 5.3, we have convergence of reduced trees to the claimed limit. In particular, for all $\varepsilon > 0$, there is $k_0 \geq 0$ such that for all $k \geq k_0$,

$$d_{\mathrm{GHP}}^{\infty}\left(\left(\mathcal{B}_k^{(k)}, \left(\mathcal{U}_k^{(i)}, i \geq 1\right), \lambda_k\right), \left(\widehat{\mathcal{T}}, \left(\left(C^{(i)}\right)^{-1} \mathcal{F}^{(i)}, i \geq 1\right), \widehat{\mu}\right)\right) < \varepsilon/3.$$

But this is only possible if all connected components of $\widehat{\mathcal{T}} \setminus \mathcal{B}_k^{(k)}$ have height less than $2\varepsilon/3$. By construction, the components of $\mathcal{B}^{(k)} \setminus \mathcal{B}_k^{(k)}$ are bounded in height by the corresponding components of height less than $2\varepsilon/3$. Since $\widehat{\mu}$ and $\mu^{(k)}$ have the same projection onto $\widehat{\mathcal{T}}_k = \mathcal{B}_k^{(k)}$, we conclude that also

$$d_{\mathrm{GHP}}^{\infty}\left(\left(\mathcal{B}_k^{(k)}, \left(\mathcal{U}_k^{(i)}, i \geq 1\right), \lambda_k\right), \left(\mathcal{B}^{(k)}, \left(\mathcal{R}^{(1)}, \ldots, \mathcal{R}^{(k)}, \{0\}, \ldots\right), \mu^{(k)}\right)\right) < 2\varepsilon/3.$$

By the triangle inequality, this completes the proof. $\square$

This formalises and proves the branch point replacement claims made in Theorem 1.6.

Acknowledgements

This research was funded in whole or in part by EPSRC, United Kingdom grants EP/P505666/1 and EP/K029797/1. For the purpose of Open Access, the authors have applied a CC BY public copyright licence to any Author Accepted Manuscript (AAM) version arising from this submission.

Appendix. Proof of Theorem 5.1

We present the proof postponed from an earlier part of this paper.

We first consider the evolution of marked subtrees $(\mathcal{R}_k^{(i)}, k \geq 1)$, $i \geq 1$. Recall the notation in Algorithm 3.2. Given that $\mathcal{R}_k^{(i)}$ has size m, 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

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

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

Lemma 6.1 (*Mass Split in Marked Subtrees*). *Let $(\mathcal{T}_k^*, (\mathcal{R}_k^{(i)}, i \ge 1), \mu_k^*, k \ge 0)$ be as in Algorithm 3.2, and fix some $i \ge 1$. Then, for $m \ge 1$, conditionally given $k_m^{(i)} = k$, the relative mass split in $\mathcal{R}_k^{(i)}$ given by*

$$\mu_k^*\left(\mathcal{R}_k^{(i)}\right)^{-1}\left(\mu_k^*\left(E_{m,i}^{(1)}\right), \ldots, \mu_k^*\left(E_{m,i}^{(m-1)}\right), \mu_k^*\left(E_{m,i}^{(m)}\right), \ldots, \mu_k^*\left(E_{m,i}^{(2m-1)}\right)\right) \tag{6.2}$$

has a Dirichlet $(\beta, \ldots, \beta, 1-2\beta, \ldots, 1-2\beta)$ *distribution and is independent of $\mu_k^*(\mathcal{R}_k^{(i)})$ and of the mass split in $\mathcal{T}_k^* \setminus \mathcal{R}_k^{(i)}$. Furthermore, for $j \in [2m-1]$,*

$$\left(\mu_k^*\left(E_{m,i}^{(j)}\right)^{-\beta} E_{m,i}^{(j)}, \mu_k^*\left(E_{m,i}^{(j)}\right)^{-1} \mu_k^* \restriction_{E_{m,i}^{(j)}}\right) \tag{6.3}$$

is a (β, θ)-strings of beads, where $\theta = \beta$ for $j \in [m-1]$ and $\theta = 1-2\beta$ for $j \in$ $[2m-1] \setminus [m-1]$. The strings of beads (6.3) are independent of each other and of the mass split in $\mathcal{R}_k^{(i)}$ given by (6.2). Conditionally given that $k_{m+1}^{(i)} = k'$,

$$\mu_{k'}^*\left(\mathcal{R}_{k'}^{(i)}\right) = \left(1 - Q_m^{(i)}\right) \mu_k^*\left(\mathcal{R}_k^{(i)}\right) \tag{6.4}$$

where $Q_m^{(i)} \sim$ Beta$(\beta, m(1-\beta) + 1 - 2\beta)$ is independent of $\left(\mu_k^*\left(\mathcal{R}_{k'}^{(i)}\right)^{-\beta} \mathcal{R}_{k'}^{(i)}, \mu_k^*\left(\mathcal{R}_{k'}^{(i)}\right)^{-1} \mu_{k'}^* \restriction_{\mathcal{R}_{k'}^{(i)}}\right)$.

Proof. This is a direct consequence of Proposition 3.3, and Proposition 2.6(ii). To see (6.4), note that $\mathcal{R}_{k'}^{(i)} \setminus \mathcal{R}_k^{(i)} = E_{m+1,i}^{(2\ m)}$ and that $\mu_{k'}^*(E_{m+1,i}^{(2\ m)}) = \gamma_k \mu_k^*(J_k^*)$ where $\gamma_k \sim$ Beta$(1-2\beta, \beta)$ is independent, and apply Proposition 2.6(i)–(ii). $\square$

Corollary 6.2 (*Length Split in Marked Subtrees*). *In the setting of Lemma 6.1, let $\widetilde{S}_{m,i} = \sum_{j \in [2m-1]} L_{m,i}^{(j)}$ denote the total length of $\mathcal{R}_{k_m^{(i)}}^{(i)}$, $m \ge 1$. Then, conditionally given $k_m^{(i)} = k$,*

$$\left(L_{m,i}^{(1)}, \ldots, L_{m,i}^{(m-1)}, L_{m,i}^{(m)}, \ldots, L_{m,i}^{(2m-1)}\right) = \mu_k^*\left(\mathcal{R}_k^{(i)}\right)^\beta S_{m,i} \cdot \left(Z_{m,i}^{(1)}, \ldots, Z_{m,i}^{(m-1)}, Z_{m,i}^{(m)}, \ldots, Z_{m,i}^{(2m-1)}\right) \tag{6.5}$$

where $\mu_k^*(\mathcal{R}_k^{(i)})$, $S_{m,i} \sim$ ML$(\beta, (m-1)(1-\beta) + 1 - 2\beta)$ and

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

are independent. In particular, $\widetilde{S}_{m,i} = \mu_k^*(\mathcal{R}_k^{(i)})^\beta S_{m,i}$. Furthermore, for $m \ge 1$,

$$\widetilde{S}_{m,i} = B_{m,i} \widetilde{S}_{m+1,i} \tag{6.6}$$

where $B_{m,i} \sim$ Beta$(m(1/\beta - 1), 1/\beta - 2)$ and $\widetilde{S}_{m+1,i}$ are independent, i.e. the sequence of lengths of each marked subtree is a Markov chain with the same transition rule as the Mittag-Leffler Markov chain with parameter $\beta/(1-\beta)$ starting from ML$(\beta/(1-\beta), (1-2\beta)/(1-\beta))$.

Proof. Fix $i \ge 1$, and set $X_j = \mu_{k_m^{(i)}}^*(E_{m,i}^{(j)})$, $j \in [2m-1]$, so that $\sum_{j \in [2m-1]} X_j = \mu_{k_m^{(i)}}^*(\mathcal{R}_{k_m^{(i)}}^{(i)})$. By Lemma 6.1, the edge lengths $L_{m,i}^{(j)}$, $j \in [2m-1]$, are given by $L_{m,i}^{(j)} = X_j^\beta M_m^{(j)}$ 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 $j \in [2m-1] \setminus [m-1]$, $\sum_{j\in[2m-1]} X_j$ and

$$\Bigg(\sum_{j\in[2m-1]} X_j\Bigg)^{-1} (X_1, \dots, X_{m-1}, X_m, \dots, X_{2m-1}) \sim \text{Dirichlet}\,(\beta, \dots, \beta, 1-2\beta, \dots, 1-2\beta)$$

are independent. We apply Proposition 2.5 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

$$\left(L_{m,i}^{(1)}, \dots, L_{m,i}^{(2m-1)}\right) = \Bigg(\sum_{j\in[2m-1]} X_j\Bigg)^{\beta} \left(\left(\frac{X_1}{\sum_{j\in[2m-1]} X_j}\right)^{\beta} M_m^{(1)}, \dots, \left(\frac{X_{2m-1}}{\sum_{j\in[2m-1]} X_j}\right)^{\beta} M_m^{(2m-1)}\right) \tag{6.7}$$

Then $\theta = (m-1)(1-\beta) + 1 - 2\beta$, and hence (6.5) follows.

To see (6.6), recall that $E_{m+1,i}^{(2\ m)} = \mathcal{R}^{(i)}_{k^{(i)}_{m+1}} \setminus \mathcal{R}^{(i)}_{k^{(i)}_m}$. By (6.7) for $m+1$, and Proposition 2.6(i)–(ii), $\mu^*_{k^{(i)}_m}\big(\mathcal{R}^{(i)}_{k^{(i)}_m}\big)^{\beta} S_{m,i} = B_{m,i}\mu^*_{k^{(i)}_{m+1}}\big(\mathcal{R}^{(i)}_{k^{(i)}_{m+1}}\big)^{\beta} S_{m+1,i}$ 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 $\mu^*_{k^{(i)}_{m+1}}(\mathcal{R}^{(i)}_{k^{(i)}_{m+1}})$, are independent, i.e. $\widetilde{S}_{m,i} = B_{m,i}\widetilde{S}_{m+1,i}$. □

Proof of Theorem 5.1.

(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. [39, Section 3.3] for a formal introduction. We define transition kernels κ_m from $\mathbb{T}_{[m]}$ to $\mathbb{T}_{[m+1]}$, $m \geq 1$: 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 $\gamma \sim \mathrm{Beta}(1-2\beta, \beta)$ 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 (5.1) 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 $\kappa_0(\cdot)$.

More formally, for $\ell \geq 1$ and some $m_i \geq 1$, $i \in [\ell]$, we will show that

$$\begin{aligned}&\mathbb{E}\left[\prod_{i\in[\ell]} f_i\left(\left(\mathcal{G}_m^{(i)}, \mu_m^{(i)}\right), m \in [m_i]\right)\right] \\ &= \prod_{i\in[\ell]} \iint \cdots \int f_i\left(R_1, \dots, R_{m_i}\right) \kappa_{m_i-1}\left(R_{m_i-1}, dR_{m_i}\right) \cdots \kappa_1\left(R_1, dR_2\right) \kappa_0\left(dR_1\right)\end{aligned} \tag{6.8}$$

for any bounded continuous functions $f_i : \mathbb{T}_{[1]} \times \cdots \times \mathbb{T}_{[m_i]} \to \mathbb{R}$, $i \in [\ell]$.

We first show Eq. (6.8) for $\ell = 1$. 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 Beta$(\beta, 1-2\beta)$. We obtain,

$$
\begin{aligned}
\mathbb{E}\left[f\left(\mathcal{G}_1,\ldots,\mathcal{G}_{m_1}\right)\right] = &\sum_{k_1^{(1)}=1,k_2^{(1)},\ldots,k_{m_1}^{(1)}} \int_{\xi_0} \sum_{v\in\xi_0} \mu_0(v) \int_{x_1} B_{\beta,1-2\beta}(x_1) \\
&\int_{\xi_1} (1-\mu_0(v)(1-\overline{x}_1))^{k_2^{(1)}-k_1^{(1)}-1} \mu_0(v)(1-\overline{x}_1) \sum_{w_1\in R_1} p_1(w_1) \int_{x_2} B_{\beta,1-2\beta}(x_2) \\
&\int_{\xi_2} \cdots \left(1-\mu_0(v)\prod_{i\in[m_1-1]}(1-\overline{x}_i)\right)^{k_{m_1}^{(1)}-k_{m_1-1}^{(1)}-1} \mu_0(v) \prod_{i\in[m_1-1]}(1-\overline{x}_i) \\
&\qquad \sum_{w_{m_1-1}\in R_{m_1-1}} p_{m_1-1}\left(w_{m_1-1}\right) \int_{x_{m_1}} B_{\beta,1-2\beta}(x_{m_1}) \int_{\xi_{m_1}} f\left(R_1,\ldots,R_{m_1}\right) \\
&\mathbb{P}\left(\xi_{\beta,1-2\beta}\in d\xi_{m_1}\right)dx_{m_1}\cdots\mathbb{P}\left(\xi_{\beta,1-2\beta}\in d\xi_2\right)dx_2\mathbb{P}\left(\xi_{\beta,1-2\beta}\in d\xi_1\right)dx_1\mathbb{P}\left(\xi_{\beta,\beta}\in d\xi_0\right)
\end{aligned}
$$

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 $\overline{x}_1 = x_1$, $\overline{x}_i = 1 - \frac{\mu_{i-1}^{(1)}(w_{i-1})}{\mu_{i-1}^{(1)}(R_{i-1})}(1-x_i)$, $i = 2, \ldots, m_1$;
- the integrals are taken over the whole ranges of $x_i \in [0, 1]$ and the subspaces of $\xi_i \in \mathbb{T}_{\mathrm{w}}$ that correspond to strings of beads.

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

$$
\begin{aligned}
\mathbb{E}\left[f\left(\mathcal{G}_1,\ldots,\mathcal{G}_{m_1}\right)\right] = &\int_{\xi_0}\sum_{v\in\xi_0}\mu_0(v)\int_{x_1}B_{\beta,1-2\beta}(x_1)\int_{\xi_1}\sum_{w_1\in R_1}p_1(w_1)\int_{x_2}B_{\beta,1-2\beta}(x_2) \\
&\int_{\xi_2}\cdots\sum_{w_{m_1-1}\in R_{m_1-1}}p_{m_1-1}\left(w_{m_1-1}\right)\int_{x_{m_1}}B_{\beta,1-2\beta}(x_{m_1})\int_{\xi_{m_1}}f\left(R_1,\ldots,R_{m_1}\right) \\
&\mathbb{P}\left(\xi_{\beta,1-2\beta}\in d\xi_{m_1}\right)dx_{m_1}\cdots\mathbb{P}\left(\xi_{\beta,1-2\beta}\in d\xi_2\right)dx_2\mathbb{P}\left(\xi_{\beta,1-2\beta}\in d\xi_1\right)dx_1\mathbb{P}\left(\xi_{\beta,\beta}\in d\xi_0\right).
\end{aligned}
$$

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

$$\mathbb{E}\left[f\left(\mathcal{G}_1,\ldots,\mathcal{G}_{m_1}\right)\right]=\int_{x_1}B_{\beta,1-2\beta}(x_1)\int_{\xi_1}\sum_{w_1\in R_1}p_1\left(w_1\right)\int_{x_2}B_{\beta,1-2\beta}(x_2)$$
$$\int_{\xi_2}\cdots\sum_{w_{m_1-1}\in R_{m_1-1}}p_{m_1-1}\left(w_{m_1-1}\right)\int_{x_{m_1}}B_{\beta,1-2\beta}(x_{m_1})\int_{\xi_{m_1}}f\left(R_1,\ldots,R_{m_1}\right)$$
$$\mathbb{P}\left(\xi_{\beta,1-2\beta}\in d\xi_{m_1}\right)dx_{m_1}\cdots\mathbb{P}\left(\xi_{\beta,1-2\beta}\in d\xi_2\right)dx_2\mathbb{P}\left(\xi_{\beta,1-2\beta}\in d\xi_1\right)dx_1.$$

We recognise the definition of the transition kernels κ_m, $m\geq 1$, and rewrite this integral in the form

$$\mathbb{E}\left[f\left(\mathcal{G}_1,\ldots,\mathcal{G}_{m_1}\right)\right]=\iint\cdots\int f\left(R_1,\ldots,R_m\right)\kappa_{m-1}\left(R_{m-1},dR_m\right)\cdots\kappa_1\left(R_1,dR_2\right)\kappa_0\left(dR_1\right)$$

To see (6.8) in the general setting, 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 3.2, as a sum over $k_j^{(i)}$, $j\in[m_i]$, $i\in[\ell]$. Then we can proceed as follows.

- First integrate out irrelevant transitions which affect components $i\geq\ell+1$ and parts of earlier transitions such as unmarked strings of beads after the creation of the ℓth component. These transitions do not affect the marked components $i\in[\ell]$.
- Move the sums over $k_{m_\ell}^{(\ell)},\ldots,k_2^{(\ell)}$ inside the integrals. Notice that there is only one term depending on $k_{m_\ell}^{(\ell)}$, i.e. we obtain the sum over $k_{m_\ell}^{(\ell)}\geq k_{m_\ell-1}^{(l)}$, $k_{m_\ell}^{(l)}\neq k_j^{(i)}$, $j\in[m_i]$, $i\in[\ell-1]$ of the probabilities of selecting the ℓth marked component at step $k_{m_\ell}^{(\ell)}$, skipping indices $k_j^{(i)}$ of insertions into other marked components $i\in[\ell-1]$, i.e.

$$\sum_{k_{m_\ell}^{(\ell)}\geq k_{m_\ell-1}^{(\ell)}+1,k_{m_\ell}^{(l)}\neq k_j^{(i)},j\in[m_i],i\in[\ell-1]}\left(1-\mu_{k_1^{(\ell)}-1}\left(v_\ell\right)\prod_{r\in[m_\ell-1]}\left(1-\overline{x}_r^{(\ell)}\right)\right)^{k(m,\ell)}\mu_{k_1^{(\ell)}-1}\left(v_\ell\right)\prod_{r\in[m_\ell-1]}\left(1-\overline{x}_r^{(\ell)}\right),$$

  where $k(m,\ell):=k_{m_\ell}^{(\ell)}-k_{m_\ell-1}^{(\ell)}-\#\{k_{m_\ell-1}^{(\ell)}<k<k_{m_\ell}^{(\ell)}:k=k_j^{(i)},\,j\in[m_i],i\in[\ell-1]\}$, and where the sequences $(x_i^{(\ell)},i\geq 1)$ and $(\overline{x}_i^{(\ell)},i\geq 1)$ are defined as $(x_i,i\geq 1)$ and $(\overline{x}_i,i\geq 1)$, respectively. Note that

$$\mu_{k_1^{(\ell)}-1}\left(v_\ell\right)\prod_{r\in[m-1]}\left(1-\overline{x}_r^{(\ell)}\right)$$

  is the mass of the ℓth 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 $i\in[\ell-1]$ happen, this sum is 1. We can proceed inductively down to $k_2^{(\ell)}$.
- The sum over the insertion point v_ℓ is just a sum over the bead selection probabilities

$$\mu_{k_1^{(\ell)}-1}(v_\ell),\quad k_1^{(\ell)}\geq k_1^{(\ell-1)}+1,$$

  which sum to the probability of creating the ℓth component (no matter what the sizes of the other components are at this step). The sum over $k_1^{(\ell)}$ 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 ℓth marked component with probability one).
- We can put the integrals over the ingredients for the ℓth subtree growth process in front of the other integrals, as they do not depend on anything else.

- Inductively, for $j = \ell - 1, \ldots, 1$, repeat these steps to lose all sums over insertion times $k_i^{(j)}$ and first insertion points v_i, $i \in [\ell]$.
- 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 (6.8).

(ii) Note that, by Lemma 6.1 (and Proposition 3.3), for each i and $k = k_m^{(i)} - 1$ for some $m \geq 1$, we are in the situation of Lemma 2.9 with $n = 2m - 1$, $\theta_1 = \cdots = \theta_{m-1} = \beta$, $\theta_m = \cdots = \theta_{2m-1} = 1 - 2\beta$, $\alpha = \beta$. We recover Algorithm 1.2 with index $\beta' = \beta/(1-\beta)$ and the "wrong" starting length $\mathrm{ML}(\beta, 1-2\beta)$, cf. Corollary 6.2.

(iii) First, note that, by Corollary 6.2, the lengths of the trees $C_m^{(i)}\mathcal{R}_{k_m^{(i)}}^{(i)}$ do not depend on $\mu_{k_m^{(i)}}^*(\mathcal{R}_{k_m^{(i)}}^{(i)})$. Fix some $i \geq 1$ and recall from Lemma 6.1 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 $m \geq 1$, define $P_m^{(i)} := \overline{Q}_1^{(i)}\overline{Q}_2^{(i)}\cdots\overline{Q}_{m-1}^{(i)}Q_m^{(i)}$, 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 ith marked component.

We recognise the stick-breaking construction (2.12) of $(P_m^{(i)}, m \geq 1)^{\downarrow} \sim \mathrm{PD}(1-\beta, 1-2\beta)$, and obtain the corresponding $(1-\beta)$-diversity $H^{(i)}$ by

$$H^{(i)} = \lim_{m\to\infty}\left(1 - \sum_{j\in[m]} P_j^{(i)}\right)^{1-\beta}(1-\beta)^{-(1-\beta)}m^{\beta} \sim \mathrm{ML}\left(1-\beta, 1-2\beta\right), \tag{6.9}$$

as in (2.15). Now fix some $m_0 \geq 1$ and let $k \geq k_{m_0}^{(i)}$. We consider the reduced tree

$$\mathcal{R}\left(C_m^{(i)}\mathcal{R}_{k_m^{(i)}}^{(i)}, \Omega_1^{(i)}, \ldots, \Omega_{m_0}^{(i)}\right) \tag{6.10}$$

spanned by the root v_i and the leaves $\Omega_1^{(i)}, \ldots, \Omega_{m_0}^{(i)}$ of $\mathcal{R}_k^{(i)}$. Recall from (i), Corollary 6.2 and Proposition 2.13 that the shape and the $\mathrm{Dirichlet}(1, \ldots, 1, 1/\beta - 2, \ldots, 1/\beta - 2)$ length split between the edges $E_{m_0,i}^{(i)}, \ldots, E_{m_0,i}^{(2m_0-1)}$ of $\mathcal{R}_{k_{m_0}^{(i)}}^{(i)}$ are as required for the reduced tree associated with a Ford CRT of index β'. Scaling by $C_m^{(i)}$ only affects the total length of the reduced tree (6.10). We will show that the total length of (6.10) scaled by $C_m^{(i)}$ converges a.s. to some $S'_{m_0} \sim \mathrm{ML}(\beta', m_0 - \beta')$, 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

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

where we will use that

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

as $1-\sum_{j\in[m]} P_j^{(i)} = \mu^*_{k_m^{(i)}}(\mathcal{R}^{(i)}_{k_m^{(i)}})/\mu^*_{k_1^{(i)}}(\mathcal{R}^{(i)}_{k_1^{(i)}})$. Then $\lim_{m\to\infty} C_i(m) = (H^{(i)})^{-\beta/(1-\beta)}\mu^*_{k_1^{(i)}}(\mathcal{R}_k^{(i)})^{-\beta}$ a.s. Note that $H^{(i)}$ is independent of $\mu^*_{k_1^{(i)}}(\mathcal{R}_k^{(i)})$ as it only depends on the sequence of independent random variables $(Q_i, i\geq 1)$ which is independent of $\mu^*_{k_1^{(i)}}(\mathcal{R}_k^{(i)})$.

The shape of $\mathcal{R}^{(i)}_{k_m^{(i)}}$ 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, we already know that the number of edges N_m+2m_0-1, $m\geq m_0$, of the reduced trees (6.10) as a subset of $\mathcal{R}^{(i)}_{k_m^{(i)}}$ behaves like the number of tables in a $(\beta', m_0-\beta')$-CRP, started at m_0, i.e. by (2.11),

$$\lim_{m\to\infty} (m-m_0)^{-\beta/(1-\beta)} N_m = \lim_{m\to\infty} m^{-\beta/(1-\beta)} N_m = S'_{m_0} \quad \text{a.s.} \tag{6.11}$$

where $S'_{m_0}\sim \mathrm{ML}(\beta', m_0-\beta')$. By Lemma 6.1, we conclude that, in the limit, the length of $\mathcal{R}^{(i)}_{k_{m_0}^{(i)}}$ is given by

$$\lim_{m\to\infty} \mu^*_{k_m^{(i)}}\left(\mathcal{R}^{(i)}_{k_m^{(i)}}\right)^{\beta} \sum_{j\in[N_m]} \left(X'_j\right)^{\beta} M^{(j)} = \mu^*_{k_{m_0}^{(i)}}\left(\mathcal{R}^{(i)}_{k_{m_0}^{(i)}}\right)^{\beta} S_{m_0,i} = \widetilde{S}_{m_0,i} \tag{6.12}$$

where $X' := (X'_1,\ldots,X'_{m-1},X'_m,\ldots,X'_{2m-1}) \sim \mathrm{Dirichlet}(\beta,\ldots,\beta,1-2\beta,\ldots,1-2\beta)$, $\mu^*_{k_m^{(i)}}(\mathcal{R}^{(i)}_{k_m^{(i)}})$, $(N_m, m\geq 1)$ and $M^{(j)}\sim \mathrm{ML}(\beta,\beta)$, $j\geq 1$, are independent. Note that we do not consider the lengths of the m_0 external edges leading to the leaves of the reduced tree (6.10) and the initial m_0-1 internal edges, which does not affect the asymptotics. We will use the representation of a Dirichlet vector $X'\sim \mathrm{Dirichlet}(\beta,\ldots,\beta,1-2\beta,\ldots,1-2\beta)$ in terms of independent Gamma variables, i.e.

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

for independent i.i.d. sequences $(Y_j, j\geq 1)$, $(Y'_j, j\geq 1)$ with $Y_1\sim \mathrm{Gamma}(\beta,1)$, $Y_1\sim \mathrm{Gamma}(1-2\beta,1)$, and $Y=\sum_{j\in[m-1]} Y_j + \sum_{j\in[m]} Y'_j \sim \mathrm{Gamma}((m-1)(1-\beta)+1-2\beta,1)$. By (6.12),

$$C_m^{(i)}\widetilde{S}_{m_0,i} = C_m^{(i)}\mu^*_{k_m^{(i)}}\left(\mathcal{R}^{(i)}_{k_m^{(i)}}\right)^{\beta}\left(\sum_{j\in[N_m+(m_0-1)]}\left(X'_j\right)^{\beta} M^{(j)} + \sum_{j=0}^{m_0-1}\left(X'_{j+m}\right)^{\beta}\overline{M}^{(j)}\right)$$

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

$$\frac{N_m(1-\beta)^{\beta}}{m^{\beta/(1-\beta)}}\left(m^{-1}\left(\sum_{j\in[m-1]} Y_j + \sum_{j\in[m]} Y'_j\right)\right)^{-\beta}\left(N_m^{-1}\left(\sum_{j\in[N_m+(m_0-1)]} Y_j^{\beta} M^{(j)} + \sum_{j\in[m_0]} Y_j'^{\beta}\overline{M}^{(j)}\right)\right). \tag{6.13}$$

By the strong law of large numbers, we have $\lim_{m\to\infty} N_m^{-1}\sum_{j\in[N_m]} Y_j^{\beta} M_m^{(j)} = \mathbb{E}[Y_1^{\beta} M_m^{(j)}] = 1$ a.s. since $N_m\to\infty$ a.s., $\mathbb{E}[Y_1^{\beta}] = \Gamma(2\beta)/\Gamma(\beta)$, and where we use the first moment of the Mittag-Leffler distribution (2.9). 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 $m^{-1}(\sum_{j\in[m-1]} Y_j + \sum_{j\in[m]} Y'_j) \to \mathbb{E}[Y''_1] = 1-\beta$ a.s. By (6.11), we conclude that the expression in (6.13) converges to S'_{m_0} a.s. where $S'_{m_0}\sim \mathrm{ML}(\beta', m_0-\beta')$. We already know that $\mathcal{R}^{(i)}_{k_m^{(i)}}$ and the scaling factor $C_m^{(i)}$ converge almost-surely, and hence, by

Proposition 2.13,

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

for $(\mathcal{F}^{(i)}_m, m \geq 1)$ are i.i.d. Ford tree growth processes of index β', i.e. (ii) follows as $m_0 \to \infty$. □

References

[1] D. Aldous, The continuum random tree. I, Ann. Probab. 19 (1) (1991a) 1–28, 1085326.

[2] D. Aldous, The continuum random tree. II. An overview, in: Stochastic Analysis (Durham, 1990), in: London Math. Soc. Lecture Note Ser., vol. 167, Cambridge Univ. Press, Cambridge, 1991b, pp. 23–70, 1166406.

[3] D. Aldous, The continuum random tree. III, Ann. Probab. 21 (1) (1993) 248–289, 1207226.

[4] D.J. Aldous, A. Bandyopadhyay, A survey of max-type recursive distributional equations, Ann. Appl. Probab. 15 (2) (2005) 1047–1110, 2134098.

[5] D. Burago, Y. Burago, S. Ivanov, A Course in Metric Geometry, American Mathematical Society, Providence, RI, 2001.

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

[7] B. Chen, D. Ford, M. Winkel, A new family of Markov branching trees: The alpha-gamma model, Electron. J. Probab. 14 (15) (2009) 400–430.

[8] N. Curien, B. Haas, The stable trees are nested, Probab. Theory Related Fields 157 (2013) 847–883.

[9] N. Curien, B. Haas, Random trees constructed by aggregation, in: Ann. Inst. Fourier, 67, (5) 2017, pp. 1963–2001.

[10] N. Curien, I. Kortchemski, Random stable looptrees, Electron. J. Probab. 19 (108) (2014) 35, 3286462.

[11] D. Dieuleveut, The vertex-cut-tree of galton-watson trees converging to a stable tree, Ann. Appl. Probab. 25 (4) (2015) 2215–2262, 3349006.

[12] B. Duplantier, J. Miller, S. Sheffield, Liouville quantum gravity as a mating of trees, Astérisque 427 (2021) arXiv preprint arXiv:1409.7055.

[13] T. Duquesne, A limit theorem for the contour process of conditioned Galton-Watson trees, Ann. Probab. 31 (2) (2003) 996–1027.

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

[15] T. Duquesne, J.-F. Le Gall, Probabilistic and fractal aspects of Lévy trees, Probab. Theory Related Fields 131 (4) (2005) 553–603, 2147221.

[16] S. Evans, Probability and real trees: École d'été de Probabilités de Saint-Flour XXXV-2005, in: Lecture Notes in Mathematics, Springer, 2008.

[17] S. Evans, J. Pitman, A. Winter, Rayleigh processes, real trees, and root growth with re-grafting, Probab. Theory Related Fields 134 (2006) 81–126.

[18] D. Ford, Probabilities on Cladograms: Introduction To the Alpha Model, Stanford University, 2005.

[19] A. Gnedin, J. Pitman, Regenerative composition structures, Ann. Probab. 33 (2) (2005) 445–479.

[20] C. Goldschmidt, B. Haas, A line-breaking construction of the stable trees, Electron. J. Probab. 20 (16) (2015) 24, 3317158.

[21] C. Goldschmidt, B. Haas, D. Sénizergues, Stable graphs: distributions and line-breaking construction, Ann. H. Lebesgue 5 (2022) 841–904.

[22] B. Haas, G. Miermont, The genealogy of self-similar fragmentations with negative index as a continuum random tree, Electron. J. Probab. 9 (2004) 57–97.

[23] B. Haas, G. Miermont, Scaling limits of Markov branching trees with applications to Galton-Watson and random unordered trees, Ann. Probab. 40 (6) (2012) 2589–2666, 3050512.

[24] B. Haas, G. Miermont, J. Pitman, M. Winkel, Continuum tree asymptotics of discrete fragmentations and applications to phylogenetic models, Ann. Probab. 36 (5) (2008) 1790–1837.

[25] B. Haas, J. Pitman, M. Winkel, Spinal partitions and invariance under re-rooting of continuum random trees, Ann. Probab. 37 (4) (2009) 1381–1411.

[26] L.F. James, Poisson calculus for spatial neutral to the right processes, Ann. Statist. 34 (1) (2006) 416–440.

[27] L.F. James, Generalized Mittag Leffler distributions arising as limits in preferential attachment models, 2015, arXiv Preprint arXiv:1509.07150.

[28] O. Kallenberg, Foundations of Modern Probability, Springer, 2001.
[29] I. Kortchemski, Invariance principles for galton-watson trees conditioned on the number of leaves, Stochastic Process. Appl. 122 (9) (2012) 3126–3172, 2946438.
[30] J.-F. Le Gall, Random real trees, Ann. Fac. Sci. Toulouse Math. 41 (1) (2006) 35–62.
[31] J.-F. Le Gall, Subordination of trees and the Brownian map, Probab. Theory Related Fields 171 (3–4) (2018) 819–864, 3827223.
[32] J.-F. Le Gall, Y. Le Jan, Branching processes in Lévy processes: the exploration process, Ann. Probab. 26 (1) (1998) 213–252.
[33] P. Marchal, A note on the fragmentation of a stable tree, in: Fifth Colloquium on Mathematics and Computer Science, AI, Discrete Mathematics and Theoretical Computer Science, 2008, pp. 489–500.
[34] G. Miermont, Self-similar fragmentations derived from the stable tree I: splitting at heights, Probab. Theory Relat. Fields 127 (3) (2003) 423–454.
[35] G. Miermont, Self-similar fragmentations derived from the stable tree II: splitting at nodes, Probab. Theory Relat. Fields 131 (3) (2005) 341–375.
[36] G. Miermont, Tessellations of random maps of arbitrary genus, Ann. Sci. Éc. Norm. Supér. 42 (5) (2009) 725–781.
[37] J. Miller, S. Sheffield, An axiomatic characterization of the Brownian map, J. Éc. Polytech. Math. 8 (2021) 609–731, 4225028.
[38] J. Pitman, Combinatorial stochastic processes: École d'été de Probabilités de Saint-Flour XXXII-2002, in: Lecture Notes in Mathematics vol. 1875, Springer, 2006.
[39] J. Pitman, M. Winkel, Regenerative tree growth: binary self-similar continuum random trees and Poisson-Dirichlet compositions, Ann. Probab. 37 (5) (2009) 1999–2041.
[40] J. Pitman, M. Yor, The two-parameter Poisson-Dirichlet distribution derived from a stable subordinator, Ann. Probab. 25 (2) (1997) 855–900.
[41] F. Rembart, Branch merging on continuum trees with applications to regenerative tree growth, ALEA Lat. Am. J. Probab. Math. Stat. 13 (2016) 563–603.
[42] F. Rembart, M. Winkel, Recursive construction of continuum random trees, Ann. Probab. 46 (5) (2018) 2715–2748, 3846837.
[43] D. Sénizergues, Random gluing of metric spaces, Ann. Probab. 47 (6) (2019) 3812–3865.
[44] D. Sénizergues, Growing random graphs with a preferential attachment structure, ALEA 19 (2022) 259–309.
[45] D. Sénizergues, S.Ö. Stefánsson, B. Stufler, Decorated stable trees, 2022, arXiv preprint arXiv:2205.02968.
[46] X. Zhang, A very gentle note on the construction of Dirichlet process, 2008, http://webdocs.cs.ualberta.ca/~xinhua2/pubDoc/notes/dirichlet_process.pdf.