Inhomogeneous continuum random trees and the entrance boundary of the additive coalescent*

by David Aldous and Jim Pitman

Technical Report No. 525

Department of Statistics
University of California
367 Evans Hall # 3860
Berkeley, CA 94720-3860

May 11, 1999

Abstract

Regard an element of the set of ranked discrete distributions $\Delta := \{(x_1, x_2, \ldots) : x_1 \geq x_2 \geq \ldots \geq 0, \sum_i x_i = 1\}$ as a fragmentation of unit mass into clusters of masses x_i. The additive coalescent is the Δ-valued Markov process in which pairs of clusters of masses $\{x_i, x_j\}$ merge into a cluster of mass $x_i + x_j$ at rate $x_i + x_j$. Aldous and Pitman (1998) showed that a version of this process starting from time $-\infty$ with infinitesimally small clusters can be constructed from the Brownian continuum random tree of Aldous (1991,1993) by Poisson splitting along the skeleton of the tree. In this paper it is shown that the general such process may be constructed analogously from a new family of inhomogeneous continuum random trees.

Key words and phrases. Continuum random tree, random forest, splitting, stochastic coalescence, stochastic fragmentation.

AMS 1991 subject classifications. 05C05, 60C05, 60J25, 60J50

*Research supported in part by N.S.F. Grants DMS 96-22859 and 97-03961

1 Introduction

Markov models of stochastic coalescence of N particles into clusters, and systems of differential equations representing evolution of relative frequencies of cluster sizes in the $N \to \infty$ limit, have a lengthy history and literature, surveyed in Aldous [5]. A more recent reformulation (Evans and Pitman [13]) is to regard an element of the set

$$\Delta := \{\boldsymbol{x} = (x_1, x_2, \ldots) : x_1 \geq x_2 \geq \ldots \geq 0, \sum_i x_i = 1\} \tag{1}$$

as a fragmentation of unit mass into clusters of masses x_i and to consider the Markov process on state space Δ whose transitions are described informally by

> each pair of clusters, of masses $\{x_i, x_j\}$ say, merges into one cluster of mass $x_i + x_j$ at rate $\kappa(x_i, x_j)$

where κ is a specified rate kernel. This is the *general stochastic coalescent.* In this paper we specialize to the case $\kappa(x, y) = x + y$, the *additive coalescent.*

Settings where coalescence with additive kernels has been studied include

- droplet formation in clouds (Golovin [14])
- algorithms for merging sets (Yao [22], Pitman [19])
- gravitational clustering in an expanding universe (Sheth and Pitman [21])
- block lengths in hashing with linear probing (Chassaing and Louchard [10])

and for general kernels see [5, 11].

There is a remarkable relationship between the additive coalescent and *continuum random trees* (CRTs). In brief, a realization of a CRT is a connected set of vertices, with a unique path of length $d(v_1, v_2)$ between any two vertices v_1, v_2 (giving a *length* measure $\ell(\cdot)$ on the *skeleton*, i.e. the subset of those vertices inside such a path), and with a probability measure μ (the *mass*) on the set of leaves of the CRT. A particular *Brownian* CRT was constructed in Aldous [2], and more general CRTs were studied in Aldous [3]. Given $0 < \lambda < \infty$, a Poisson process of cuts along the skeleton of the Brownian CRT, with intensity $\lambda\ell(\cdot)$, will fragment the tree into subtrees, and the ranked μ-masses of the subtrees define a random element $\boldsymbol{Y}(\lambda)$ of Δ. Varying λ gives a *fragmentation process* $(\boldsymbol{Y}(\lambda), 0 \leq \lambda < \infty)$. The remarkable relationship, developed in Aldous and Pitman [7], is that the deterministic time-reversal $\boldsymbol{X}(t) := \boldsymbol{Y}(e^{-t})$ yields a version $(\boldsymbol{X}(t), -\infty < t < \infty)$ of the additive coalescent, for which the mass $X_1(t)$ of the largest cluster satisfies $X_1(t) \to 0$ as $t \to -\infty$. Call this the *standard* additive coalescent. Note the time interval is $-\infty < t < \infty$. We call such a process *eternal*.

Bertoin [8] recently gave a different construction of the fragmentation process $\boldsymbol{Y}$ based on excursions of Brownian motion with varying drift. While conceptually simpler than the construction via CRTs, it does not appear so useful for the generalizations developed in this paper.

Let $l^2_\downarrow$ be the subset of the unit ball of l^2 consisting of non-negative ranked vectors $\boldsymbol{\theta}$. In other words, an element $\boldsymbol{\theta} := (\theta_1, \theta_2, \ldots)$ of $l^2_\downarrow$ has $\theta_1 \geq \theta_2 \geq \theta_3 \geq \ldots \geq 0$ and $\sum_i \theta_i^2 \leq 1$. Now define

$$\Theta := \{\boldsymbol{\theta} \in l^2_\downarrow : \sum_i \theta_i^2 < 1 \text{ or } \sum_i \theta_i = \infty\}.$$

The first purpose of this paper is to construct, for each $\boldsymbol{\theta} \in \Theta$, a particular *inhomogeneous continuum random tree* (ICRT) $\mathcal{T}^{\boldsymbol{\theta}}$. The construction of $\mathcal{T}^{\boldsymbol{\theta}}$ in Section 2 is an extension of the line-breaking construction of the Brownian CRT made in [2] in the special case $\boldsymbol{\theta} = \mathbf{0} := (0, 0, \ldots)$. The novel feature for $\boldsymbol{\theta} \neq \mathbf{0}$ is that in this case the ICRT has the extra structure of distinguished vertices, which we call *hubs*, labeled by the i with $\theta_i > 0$. In Section 3.3 we construct a mass measure μ on $\mathcal{T}^{\boldsymbol{\theta}}$ using the general theory of CRTs from [3] (this is the only use of such general theory). We then show in Theorem 10 that, as with the Brownian CRT, a time-reversal of the fragmentation process on $\mathcal{T}^{\boldsymbol{\theta}}$ yields an eternal version $\boldsymbol{X}^{\boldsymbol{\theta}} = (\boldsymbol{X}^{\boldsymbol{\theta}}(t), -\infty < t < \infty)$ of the additive coalescent. According to Theorem 15, the processes $\{\boldsymbol{X}^{\boldsymbol{\theta}},\ \boldsymbol{\theta} \in \Theta\}$ together with the trivial process (one cluster of mass 1 for all t) make up the entire entrance boundary at time $-\infty$ for the additive coalescent. The entrance boundary for the *multiplicative* coalescent ($\kappa(x, y) = xy$) was studied by Aldous and Limic [6]. The two cases are compared and contrasted in Section 6.1.

The central result of Theorem 10, that time-reversing the fragmentation process on $\mathcal{T}^{\boldsymbol{\theta}}$ gives a version of the additive coalescent, is deduced in Sections 3.2 and 4 from corresponding discrete-space results via a weak convergence argument, using the Feller property [13] of the additive coalescent. Part of the preparatory work for the weak convergence argument is done in a companion paper, Camarri and Pitman [9]. Another companion paper, Aldous and Pitman [1], develops combinatorial and distributional aspects of the ICRT $\mathcal{T}^{\boldsymbol{\theta}}$.

To summarize, the contributions of this paper are

- the construction of the ICRT $\mathcal{T}^{\boldsymbol{\theta}}$
- the construction of eternal additive coalescents by fragmentation and time reversal of $\mathcal{T}^{\boldsymbol{\theta}}$

- the proof that *every* extreme eternal additive coalescent may be obtained in this manner.

The proofs use diverse techniques:

- discrete combinatorial structure relating tree-fragmentation and coalescence (Proposition 1, from [13])
- discrete "birthday trees" (Proposition 2, from [9])
- weak convergence
- a little general theory of CRTs (Section 3.3)
- Kingman's [15] theory of exchangeable random partitions (Section 4.1)
- calculations with the combinatorial structure of the ICRT (Section 5.1)
- stochastic calculus to analyze asymptotic properties of a general additive coalescent (Section 5.2).

Terminology. We deal with several kinds of trees, defined in detail later, but let us record here some general terminology. We regard trees as unrooted, and edges as undirected. A *discrete* tree has a countable (finite or countably infinite) number of vertices, with edges of length 1. If instead the edges are assigned positive real numbers as lengths, we call it a *discrete tree with edge lengths*. If we regard each edge of such a tree as a continuous set of vertices, we get one instance of a continuum tree, where the tree has a countable set of leaves. More interesting continuum trees, such as realizations of $\mathcal{T}^{\boldsymbol{\theta}}$ for $\boldsymbol{\theta} \in \Theta$, have an uncountable set of leaves, and the mass measure μ on the set of leaves is non-atomic.

2 The line-breaking construction of the ICRT

This construction generalizes a construction in [2, Process 3], which is the special case of the present construction with $\boldsymbol{\theta} = 0$. The construction is motivated as a weak limit of a discrete construction, as will become clear in Section 3.2. Our focus in this paper is on the explicit construction; how it fits into the more abstract framework [3] of continuum random trees will be described briefly in Section 3.3.

Fix $\boldsymbol{\theta} := (\theta_1, \theta_2, \ldots) \in l^2_\downarrow$ and define $a = 1 - \sum_i \theta_i^2$. So $0 \le a \le 1$. If $a > 0$ let $((U_j, V_j), 1 \le j < \infty)$ be the points of a Poisson point process of rate a per unit area on the octant $\{(u,v) : 0 < v < u < \infty\}$, labeled so that $0 < U_1 < U_2 < \ldots$. In the case $a = 0$, ignore subsequent mentions of U_j and V_j. For each i such that $\theta_i > 0$, let $0 < \xi_{i,1} < \xi_{i,2} < \ldots$ be the points of a Poisson point process on $(0,\infty)$ of rate θ_i per unit length. The ICRT $\mathcal{T}^{\boldsymbol{\theta}}$ will be constructed as a function of the points of these Poisson processes, which are assumed to be independent. The construction is illustrated in Figures 1-3. In outline, we use the Poisson points to cut the line $[0,\infty)$ into finite-length segments, then assemble the segments to form the branches of a tree, where each point of the tree is labeled by some $0 \le x < \infty$; finally, we pass to a metric space completion. Here are the details.

Call each point U_j a 0-*cutpoint*, and say that V_j is the corresponding *joinpoint*. Call each point $\xi_{i,j}$ with $\theta_i > 0$ and $j \ge 2$ (note the 2) an i-*cutpoint*, and say that $\xi_{i,1}$ is the corresponding *joinpoint*. Note that there are (with probability 1, a qualification in effect throughout the construction) only finitely many cutpoints in any finite interval $[0,x]$, because for $i \ge 1$ the mean number of i-cutpoints in that interval equals $\theta_i x - (1 - \exp(-\theta_i x)) \le \theta_i^2 x^2$. We may therefore order the cutpoints as $0 < \eta_1 < \eta_2 < \ldots$, where $\eta_k \to \infty$ as $k \to \infty$. Figure 1 illustrates a typical realization of the cutpoints, with each η_k identified as some U_j or $\xi_{i,j}$.

We build the tree by starting with the branch $[0,\eta_1]$ and then, inductively on $k \ge 1$, attaching the branch $(\eta_k, \eta_{k+1}]$ to the joinpoint η_k^* corresponding to the cutpoint η_k. Figure 2 illustrates the attachment of the first 8 branches, using the realization in Figure 1. The reader will find it helpful to work through the construction in Figure 2: the sequence of attachments of branches is

$$[0,U_1],\ \ (V_1,U_2],\ \ (V_2,\xi_{1,2}],\ \ (\xi_{1,1},\xi_{4,2}],\ \ (\xi_{4,1},U_3],\ \ (V_3,\xi_{2,2}],\ \ (\xi_{2,1},\xi_{1,3}],\ \ (\xi_{1,1},U_4].$$

After all the branches $(\eta_k,\eta_{k+1}]$ are attached we obtain a tree, say $\mathcal{T}_0^{\boldsymbol{\theta}}$. By construction, if $\theta_i > 0$ then an infinite number of branches (the line segments starting at each $\xi_{i,k}, k \ge 2$) are attached at $\xi_{i,1}$. Note also that the points $(V_j, 1 \le j < \infty)$ at which line segments $(U_j = \eta_k, \eta_{k+1}]$ are attached are all distinct and are distinct from the $\xi_{i,1}$, because V_j conditionally given U_j is uniform on $[0,U_j]$.

Formally, a realization of $\mathcal{T}_0^{\boldsymbol{\theta}}$ is just the halfline $[0,\infty)$ with an unusual metric d determined by the realization of the Poisson point processes. To see this, for $x \in (0,\infty)$ write $\chi(x) := \max\{\eta_k \le x : \eta_k \text{ is a cutpoint }\}$ and let $\chi^*(x)$ be the joinpoint corresponding to $\chi(x)$. Then for each $x \in [0,\infty)$ the path from x to 0 in the tree consists of branch segments

$$[x = x_0, \chi(x_0)),\ [x_1 = \chi^*(x_0), \chi(x_1)),\ [x_2 = \chi^*(x_1), \chi(x_2)),\ \ldots [x_u = \chi^*(x_{u-1}), \chi(x_u) = 0)$$

where $[b,a)$ for $a < b$ indicates that the path traverses the interval $(a,b]$ from right to left. So the distance $d(x,0)$ equals $\sum_{m=0}^{u}(x_u - \chi(x_u))$, and then

$$d(x,y) = (d(x,0) - d(b,0)) + (d(y,0) - d(b,0))$$

where b is the branchpoint of the paths from 0 to x and to y. Because a realization of $\mathcal{T}_0^{\boldsymbol{\theta}}$ is a metric space, we may define $\mathcal{T}^{\boldsymbol{\theta}}$ to be the completion of the metric space, and call the elements v of $\mathcal{T}^{\boldsymbol{\theta}}$ *vertices*. It is straightforward to verify directly the following properties of almost all realizations of $\mathcal{T}^{\boldsymbol{\theta}}$.

(i) $\mathcal{T}^{\boldsymbol{\theta}}$ is a *tree* in the sense that there is a unique (non self-intersecting) path $[[v,w]]$ between each pair v,w of vertices.

(ii) Define the *skeleton* skel$(\mathcal{T}^{\boldsymbol{\theta}})$ to be the set of vertices which are interior to some path $[[v,w]]$. Then skel$(\mathcal{T}^{\boldsymbol{\theta}}) = [0,\infty) \setminus \cup_k\{\eta_k\}$.

(iii) Lebesgue measure on $[0,\infty)$ induces a σ-finite *length measure* $\ell(\cdot)$ on $\mathcal{T}^{\boldsymbol{\theta}}$, which is null outside skel$(\mathcal{T}^{\boldsymbol{\theta}})$, such that the distance $d(v,w)$ between any pair of vertices equals the ℓ-measure of the path $[[v,w]]$.

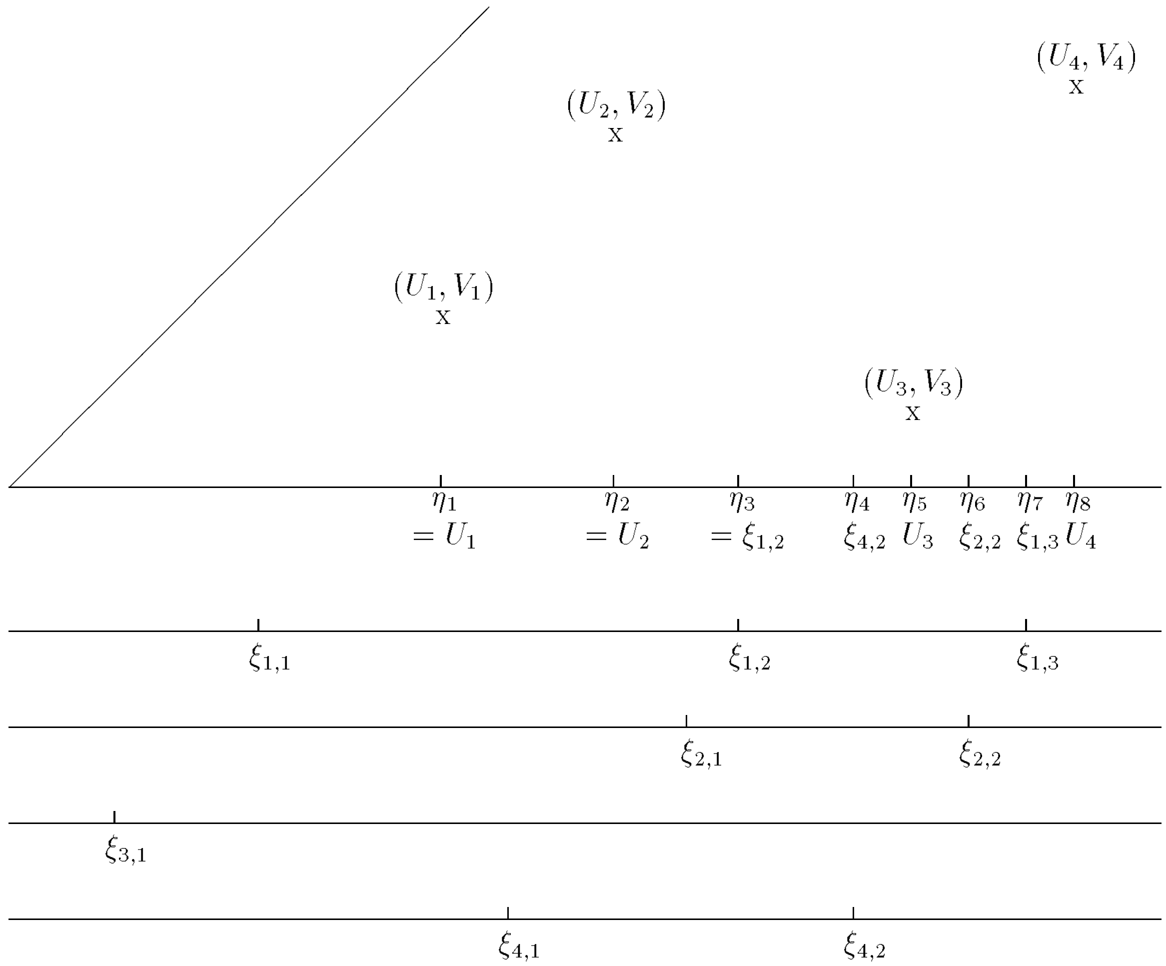

(U_4, V_4)
(U_2, V_2)
(U_1, V_1)
(U_3, V_3)
η_1 $= U_1$
η_2 $= U_2$
η_3 $= \xi_{1,2}$
η_4 $\xi_{4,2}$
η_5 U_3
η_6 $\xi_{2,2}$
η_7 $\xi_{1,3}$
η_8 U_4
$\xi_{1,1}$
$\xi_{1,2}$
$\xi_{1,3}$
$\xi_{2,1}$
$\xi_{2,2}$
$\xi_{3,1}$
$\xi_{4,1}$
$\xi_{4,2}$

Figure 1

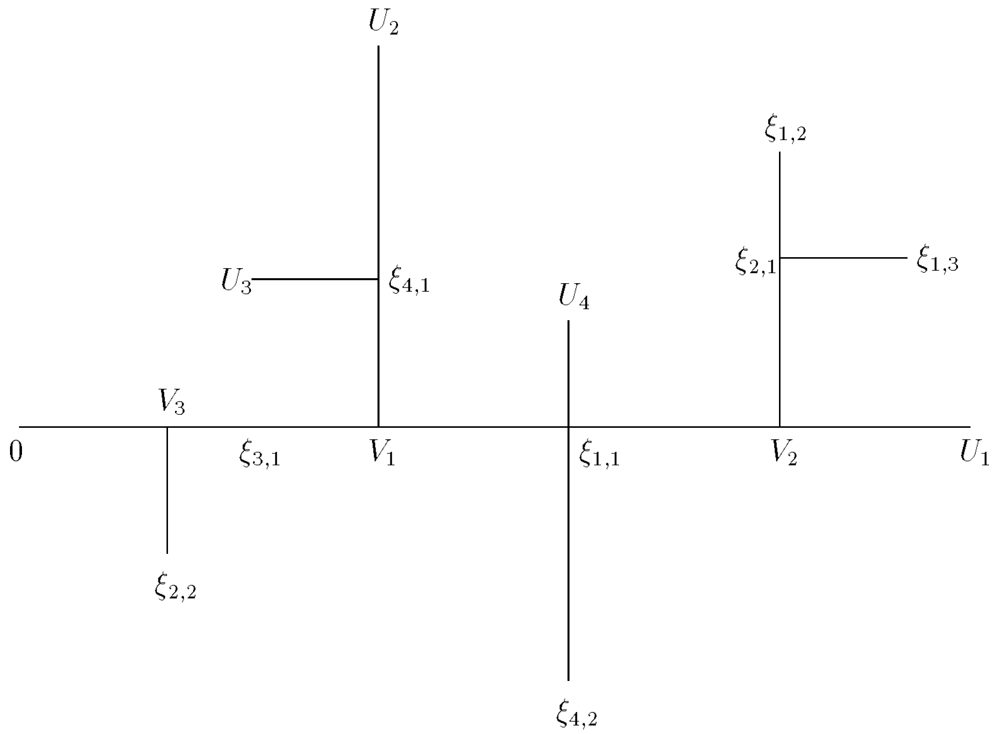


Figure 2

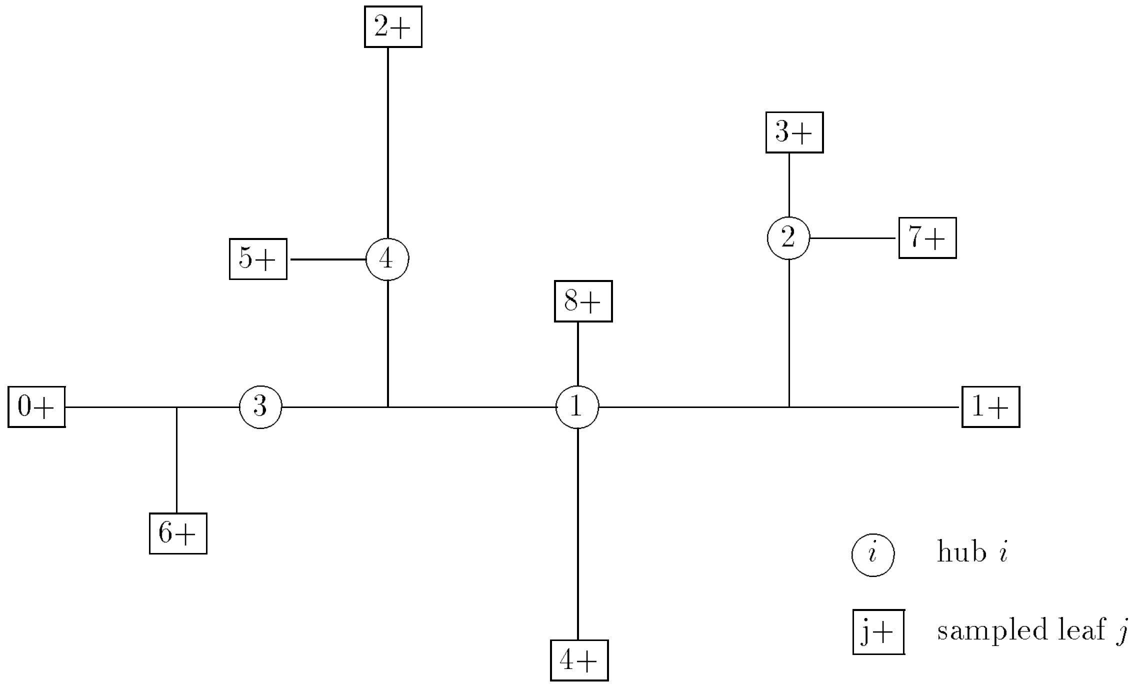


Figure 3

(iv) Define the *branchpoints* $\mathrm{br}(\mathcal{T}^{\boldsymbol{\theta}})$ to be the set of vertices v such that there exist vertices $w_1, w_2, w_3 \neq v$ such that the paths $[[v, w_u]], u = 1, 2, 3$ are disjoint except for v. Then $\mathrm{br}(\mathcal{T}^{\boldsymbol{\theta}}) = \{\xi_{i,1} : \theta_i > 0\} \cup \{V_j : j \geq 1\}$.

The fact that the skeleton is constructed as a subset of $[0, \infty)$ is rather an artifact of the construction. To discuss what we regard as more intrinsic properties of $\mathcal{T}^{\boldsymbol{\theta}}$, we use a relabeling, illustrated in Figure 3. When $\theta_i > 0$, call vertex $\xi_{i,1}$ *hub* i. The hubs are branchpoints of infinite degree. If $a = 0$ there are no other branchpoints, while if $a > 0$ there are a countably infinite number of other branchpoints $V_j, j \geq 1$, all of degree 3. Setting $\eta_0 = 0$, for each $j \geq 0$ call vertex η_j *sampled leaf* j, written symbolically as $j+$ to distinguish it from hub j. Motivation for the name comes from Proposition 5 below.

The construction above requires only that $\boldsymbol{\theta} \in l^2_{\downarrow}$. To complete the construction of the ICRT $\mathcal{T}^{\boldsymbol{\theta}}$ we need to specify the mass measure μ on a realization of $\mathcal{T}^{\boldsymbol{\theta}}$. This specification, which requires $\boldsymbol{\theta} \in \Theta$, will be done in Proposition 5, which describes μ as the almost sure weak limit of the discrete uniform distribution on $\{0+,\ 1+,\ \ldots\ J+\}$ as $J \to \infty$. The existence of this limit is not easy to prove directly from the construction above.

Remark. Abstractly, a *continuum tree* is a metric space with certain regularity properties. In this paper we view the line-breaking construction as yielding a (random) metrization of $[0, \infty)$. An alternative formulation ([3] section 2.2) views a continuum tree as a subset of sequence space l_1, built by attaching the successive line-segments orthogonally. Continuum trees are closely related to objects in general topology called *dendrites* (Nadler [17] Chapter 10), or $\mathbf{R}$-*trees* (Mayer and Oversteegen [16]).

2.1 Spaces of trees

For later use we set up notation for spaces of trees. For $I \geq 0$ and $J \geq 1$ let $\mathbf{T}_{IJ}$ be the space of trees such that

(i) there are exactly $J + 1$ leaves, labeled $0+,\ 1+,\ \ldots, J+$;

(ii) there may be extra labeled vertices, with distinct labels in $\{1, \ldots, I\}$;

(iii) there may be unlabeled branchpoints, of degree ≥ 3;

(iv) each edge e has a length l_e, where l_e is a strictly positive real number.

From $\mathcal{T}^{\boldsymbol{\theta}}$ we can now define a $\mathbf{T}_{IJ}$-valued *reduced tree* $r_{IJ}(\mathcal{T}^{\boldsymbol{\theta}})$ as follows. First take the subtree of $\mathcal{T}^{\boldsymbol{\theta}}$ spanned by $0+,\ 1+, \ldots, J+$, in other words the part of $\mathcal{T}^{\boldsymbol{\theta}}$ constructed from the interval $[0, \eta_J]$. Then for each hub i appearing in the subtree, if $i \leq I$ we retain the label i, and if $i > I$ we remove the label. Thus Figure 3 shows a possible realization of $r_{4,8}(\mathcal{T}^{\boldsymbol{\theta}})$. To illustrate further, the reader should check that

(a) the corresponding realization of $r_{5,8}(\mathcal{T}^{\boldsymbol{\theta}})$ will be either the same tree as in Figure 3 (if $\xi_{5,1} > \eta_8$) or will have some point in the tree identified as hub 5;

(b) the corresponding realization of $r_{4,9}(\mathcal{T}^{\boldsymbol{\theta}})$ will be the tree in Figure 3, with an extra edge to a leaf 9+.

Later arguments will use weak convergence for $\mathbf{T}_{IJ}$-valued random trees. This presupposes some topology on $\mathbf{T}_{IJ}$. Each tree $\mathbf{t} \in \mathbf{T}_{IJ}$ has a *shape* shape($\mathbf{t}$), which is the discrete tree obtained by ignoring edge-lengths. The set $\mathbf{T}_{IJ}^{\text{shape}}$ of possible shapes is finite. One can formally regard $\mathbf{t}$ as a vector (shape($\mathbf{t}$); l_e, e an edge of shape($\mathbf{t}$)) and thereby $\mathbf{T}_{IJ}$ inherits a topology from the discrete topology on $\mathbf{T}_{IJ}^{\text{shape}}$ and the usual product topology on $\mathbb{R}^d$.

While we informally think of $\mathcal{T}^{\boldsymbol{\theta}}$ as a random element of some space of continuum trees, it seems complicated to provide a satisfactory formalization of such a space. For this reason we avoid talking about "the distribution of $\mathcal{T}^{\boldsymbol{\theta}}$", and instead make distributional statements about the reduced trees $r_{IJ}(\mathcal{T}^{\boldsymbol{\theta}})$.

3 Birthday trees

3.1 Background results

The material below on discrete trees is developed further in [9], and we give only what is needed for the present paper.

Let $\boldsymbol{p} = (p_i, 1 \leq i < \infty)$ be a ranked discrete distribution on the positive integers. That is $p_1 \geq p_2 \geq \ldots \geq 0$ and $\sum_i p_i = 1$. The *support* of $\boldsymbol{p}$ is $S := \{i \geq 1 : p_i > 0\}$. Let $(W_0, W_1, \ldots)$ be i.i.d.($\boldsymbol{p}$), that is independent with distribution $\boldsymbol{p}$. Define a random discrete tree $\mathcal{T}^{\boldsymbol{p}} := \mathcal{T}(W_0, W_1, \ldots)$ to have vertex-set S and undirected edges

$$\{\{W_{m-1}, W_m\} : W_m \notin \{W_0, \ldots, W_{m-1}\}, m \geq 1\}.$$

Here's a mental picture. Mark the vertices $s \in S$ as dots on a piece of paper, and use a pencil to draw edges between the vertices according to the rule

> After step $m-1$ the pencil is at vertex W_{m-1}. If W_m has not been previously visited, draw an edge from W_{m-1} to W_m; otherwise move the pencil to W_m without drawing an edge.

We are abusing notation by using the same symbol $\mathcal{T}$ for the continuum tree $\mathcal{T}^{\boldsymbol{\theta}}$ and for the birthday tree $\mathcal{T}^{\boldsymbol{p}}$. But the meaning should be clear from context.

Here is why we are interested in $\mathcal{T}^{\boldsymbol{p}}$. To the edges e of $\mathcal{T}^{\boldsymbol{p}}$ associate independent random variables ζ_e with distribution $U(0,1)$, that is uniform on the interval $(0,1)$. For $0 \leq q \leq 1$ let $\mathcal{F}^{\boldsymbol{p}}(q)$ be the forest on vertices S with edge-set $\{e \in \mathcal{T}^{\boldsymbol{p}} : \zeta_e > q\}$. Let $\boldsymbol{Y}^{\boldsymbol{p}}(q)$ be the ranked $\boldsymbol{p}$-measures of the tree-components of $\mathcal{F}^{\boldsymbol{p}}(q)$. Call $\boldsymbol{Y}^{\boldsymbol{p}}(\cdot)$

the *fragmentation process* associated with $\mathcal{T}^{\boldsymbol{p}}$. So $\boldsymbol{Y}^{\boldsymbol{p}}(q)$ records the $\boldsymbol{p}$-measures of components obtained when each edge is cut with probability q.

Proposition 1 ([13, Construction 5]) *Define* $\boldsymbol{X}^{\boldsymbol{p}}(t) = \boldsymbol{Y}^{\boldsymbol{p}}(e^{-t}), 0 \leq t < \infty$. *Then* $(\boldsymbol{X}^{\boldsymbol{p}}(t), 0 \leq t < \infty)$ *is an additive coalescent with initial state* $\boldsymbol{p}$.

The state space of $\boldsymbol{X}^{\boldsymbol{p}}(t)$ or $\boldsymbol{Y}^{\boldsymbol{p}}(q)$ is the set Δ at (1), equipped with the topology it inherits as a subset of l_1. Foundational aspects of the additive coalescent are discussed in [13], but all we really need is Proposition 1 and the Feller property quoted in Section 5 below. Proposition 1 gives an explicit construction of the additive coalescent as a Δ-valued process, starting at time 0 from an arbitrary point $\boldsymbol{p} \in \Delta$. But such a "discrete" construction will not serve to construct an additive coalescent starting at time $-\infty$. To do the latter we need to pass to a limit continuum tree, and this is the central idea of the paper.

A convenient way of studying asymptotic behavior of random trees is by studying subtrees spanned by a random finite set of vertices. The next proposition implies that such subtrees appear automatically within the construction above. For $k \geq 1$ let R_k be the index of the k'th repeat in the sequence (W_m), in other words the smallest r such that $\{W_0, W_1, \ldots, W_r\}$ contains exactly $1 + r - k$ distinct vertices.

Proposition 2 ([9, Theorem 2],[18, Lemma 11]) *The subsequence* $(W_0, W_{R_1-1}, W_{R_2-1}, \ldots)$ *is i.i.d.*$(\boldsymbol{p})$ *and this subsequence is independent of* $\mathcal{T}^{\boldsymbol{p}} := \mathcal{T}(W_0, W_1, \ldots)$.

In [9, Theorem 2] it is shown that the tree $\mathcal{T}(W_0, W_1, \ldots)$, when regarded as a tree rooted at W_0, is independent of the subsequence $(W_{R_1-1}, W_{R_2-1}, \ldots)$. Proposition 2 combines this result with the fact [18, Lemma 11] that W_0 is independent of $\mathcal{T}(W_0, W_1, \ldots)$ regarded as an unrooted tree.

According to Proposition 2, the tree generated by the pencil construction up to step R_k has the distribution of the subtree of $\mathcal{T}^{\boldsymbol{p}}$ spanned by $k+1$ vertices picked independently of $\mathcal{T}^{\boldsymbol{p}}$ with distribution $\boldsymbol{p}$. Proposition 5 will give an analogous result for the ICRT $\mathcal{T}^{\boldsymbol{\theta}}$.

3.2 Weak convergence of birthday trees

We now work toward Proposition 3, which gives one sense in which $\mathcal{T}^{\boldsymbol{\theta}}$ is a limit of birthday trees. Fix $I \geq 0, J \geq 1$. Recall the earlier definition of the space $\mathbf{T}_{IJ}$ of trees. We shall define a *reduced tree* $r_{IJ}(\mathcal{T}^{\boldsymbol{p}})$ using only part of the construction of $\mathcal{T}^{\boldsymbol{p}} = \mathcal{T}(W_0, W_1, \ldots)$. We would like to say that $r_{IJ}(\mathcal{T}^{\boldsymbol{p}})$ takes values in $\mathbf{T}_{IJ}$, and we handle this by appending to $\mathbf{T}_{IJ}$ a conventional state ∂ and declaring $r_{IJ}(\mathcal{T}^{\boldsymbol{p}}) = \partial$ when it is not in $\mathbf{T}_{IJ}$. First, consider the subtree of $\mathcal{T}^{\boldsymbol{p}}$ obtained by stopping drawing edges at the time R_J of the J'th repeat. This subtree has edges

$$\{\{W_{m-1}, W_m\} : W_m \notin \{W_0, \ldots, W_{m-1}\}, 1 \leq m < R_J\}.$$

Make this subtree into a "tree with edge-lengths" by assigning length σ to each edge, where

$$\sigma := \sqrt{\sum_i p_i^2}. \tag{2}$$

Relabel vertex W_0 as vertex $0+$ and, for each $1 \leq j \leq J$, relabel vertex $W_{R_j - 1}$ as vertex $j+$. Of the remaining vertices, those with labels $1 \leq i \leq I$ retain the label, and the others are unlabeled. Finally, unlabeled vertices of degree 2 are deleted. More precisely, as illustrated in Figure 4, each maximal l-edge path joining such vertices is replaced by a single edge of length $l\sigma$.

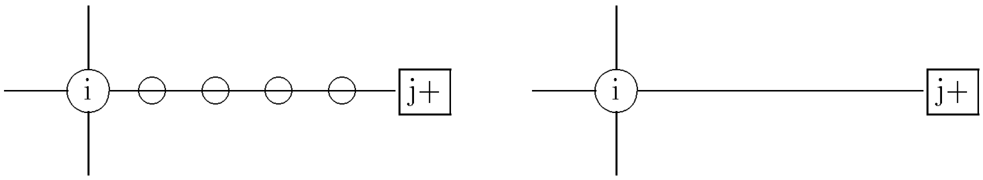


Figure 4

Call the resulting tree $r_{IJ}(\mathcal{T}^{\boldsymbol{p}})$. As mentioned above, the tree might not satisfy the requirements of $\mathbf{T}_{IJ}$ (e.g. vertices may be multiply labeled, or some $j+$ might not be a leaf), in which case we set $r_{IJ}(\mathcal{T}^{\boldsymbol{p}}) = \partial$. Note that $r_{IJ}(\mathcal{T}^{\boldsymbol{p}})$ is actually a function not just of $\mathcal{T}^{\boldsymbol{p}}$ but also of the variables W_0 and W_{R_j-1} for $1 \leq j \leq J$, which according to Proposition 2 are i.i.d.($\boldsymbol{p}$) independent of $\mathcal{T}^{\boldsymbol{p}}$.

For each $n = 1, 2, \ldots$ let $\boldsymbol{p}_n := (p_{ni}, 1 \leq i < \infty)$ be a ranked discrete probability distribution, and write W_{nm}, R_{nk}, $\mathcal{T}^{\boldsymbol{p}_n} = \mathcal{T}(W_{n0}, W_{n1}, \ldots)$ for the associated quantities from Section 3.1. We shall be concerned throughout the paper with the asymptotic regime

$$\lim_{n\to\infty} \sigma_n = 0, \quad \lim_{n\to\infty} \frac{p_{ni}}{\sigma_n} = \theta_i, \ \ i \geq 1, \ \text{where } \sigma_n := \sqrt{\sum_i p_{ni}^2}. \tag{3}$$

Note that $\boldsymbol{\theta} = (\theta_i)$ is automatically a ranked vector with $\sum_i \theta_i^2 \leq 1$, and so we are in the setting of Section 2 and the ICRT $\mathcal{T}^{\boldsymbol{\theta}}$ exists. In fact the construction of the ICRT was motivated by the following result.

Proposition 3 ([9, Corollary 15]) *Fix $I \geq 0, J \geq 1$. Under the asymptotic regime* (3)*,*

$$r_{IJ}(\mathcal{T}^{\boldsymbol{p}_n}) \xrightarrow{d} r_{IJ}(\mathcal{T}^{\boldsymbol{\theta}}) \ \textit{on}\ \mathbf{T}_{IJ}.$$

Part of the assertion of Proposition 3 is that

$$\sigma_n R_{nJ} \xrightarrow{d} \eta_J.$$

Due to obvious geometric bounds on the distribution of R_{nJ} it follows that also

$$\sigma_n E R_{nJ} \to E\eta_J. \tag{4}$$

Proposition 3 may be understood as follows. The construction of $\mathcal{T}^{\boldsymbol{p}}$ (with edges rescaled to have length σ) from $(W_0, W_1, W_2, \ldots)$ can be pictured as a line-breaking construction, based on the marked point process in which the point in $[0, \infty)$ at position $m\sigma$ is given mark W_m. Under the asymptotic regime (3), the Poisson(θ_i) processes $(\xi_{i,k}, k \geq 1)$ and $((U_i, V_i), i \geq 1)$ featuring in the construction of $\mathcal{T}^{\boldsymbol{\theta}}$ arise as weak limits of $(\xi^n_{i,k}, k \geq 1)$ and $((U^n_i, V^n_i), i \geq 1)$, where

$\xi^n_{i,k} =$ position of k'th occurrence of mark i derived from $(W_{nu}, u \geq 1)$

(U^n_i, V^n_i) are the positions of pairs (u, v) such that $W_{nu} = W_{nv} \geq m(n)$ and $v < u$, ordered so that $(U^n_i, i \geq 1)$ is increasing, for $m(n) \to \infty$ slowly enough that $\lim_{n\to\infty} \sum_{i=1}^{m(n)} (p_{ni}/\sigma_n)^2 = \sum_i \theta_i^2$.

Later we shall use Proposition 3 to derive properties of $\mathcal{T}^{\boldsymbol{\theta}}$, so we need to check that approximating sequences $(\boldsymbol{p}_n)$ exist.

Lemma 4 *For each $\boldsymbol{\theta} \in l^2_\downarrow$ there exists $(\boldsymbol{p}_n, n \geq 1)$ satisfying (3) with limit $\boldsymbol{\theta}$.*

Proof. The following construction works only for $\boldsymbol{\theta} \in \Theta$, which is the case needed later. The case $\boldsymbol{\theta} \in l^2_\downarrow \setminus \Theta$ is simpler and left to the reader.

Define $m_n = \max\{i : \theta_i > n^{-1/2}\}$. Because $0 \leq \sum_{i=1}^{m_n} \theta_i^2 < 1$ (the upper bound by definition of Θ) there is a positive solution z_n of

$$z_n^2 = n + z_n^2 \sum_{i=1}^{m_n} \theta_i^2$$

and $n^{1/2} \leq z_n < \infty$. So $z_n\theta_i > 1$ for $i \leq m_n$. We may therefore define a ranked probability measure $\boldsymbol{p}_n$ by

$$p_{ni} = \begin{cases} z_n\theta_i/s_n, & 1 \leq i \leq m_n \\ 1/s_n, & m_n < i \leq m_n + n \end{cases}$$

where $s_n := z_n \sum_{i=1}^{m_n} \theta_i + n$. So

$$\sigma_n := \sqrt{\sum_i p_{ni}^2} = \frac{1}{s_n} \sqrt{z_n^2 \sum_{i=1}^{m_n} \theta_i^2 + n} = \frac{z_n}{s_n}.$$

We first need to show that $\sigma_n \to 0$, equivalently that

$$\frac{s_n}{z_n} = \sum_{i=1}^{m_n} \theta_i + \frac{n}{z_n} \to \infty.$$

But if $\sum_i \theta_i^2 = 1$ the first term $\to \infty$ by definition of Θ, and if not then $z_n = O(n^{1/2})$ by definition of z_n and so the second term $\to \infty$. Finally, note that

$$\frac{p_{ni}}{\sigma_n} = \begin{cases} \theta_i, & 1 \le i \le m_n \\ 1/z_n, & m_n < i \le m_n + n. \end{cases}$$

For each i with $\theta_i > 0$ we have $m_n \ge i$ ultimately, so (since $z_n \to \infty$) we see $p_{ni}/\sigma_n \to \theta_i \ \forall i$, establishing (3).

3.3 The mass measure on $\mathcal{T}^{\boldsymbol{\theta}}$

Fundamental to this paper is the idea that for $\boldsymbol{\theta} \in \Theta$ there exists a mass measure μ on $\mathcal{T}^{\boldsymbol{\theta}}$, which we view informally as a limit of the probability measures $\boldsymbol{p}_n$ on $\mathcal{T}^{\boldsymbol{p}_n}$ in the setting of Proposition 3. But this limit relationship is not easy to formulate or prove starting from Proposition 3, because the $\boldsymbol{p}_n$-measure of the vertices of $\mathcal{T}^{\boldsymbol{p}_n}$ involved in $r_{IJ}(\mathcal{T}^{\boldsymbol{p}_n})$ is asymptotically negligible. Instead, we use existing general CRT theory to establish existence of μ (Proposition 5). The asymptotic relationship appears implicitly later, in Proposition 13.

Associated with a realization of $\mathcal{T}^{\boldsymbol{\theta}}$ is the probability measure μ_J on $\mathcal{T}^{\boldsymbol{\theta}}$ (considered as a metric space) defined to be the discrete uniform distribution on the $J+1$ vertices $0+, \ 1+, \ldots, J+$.

Proposition 5 *Let* $\boldsymbol{\theta} \in \Theta$.
(a) *For almost all realizations of* $\mathcal{T}^{\boldsymbol{\theta}}$*, there exists a probability measure* μ *on* $\mathcal{T}^{\boldsymbol{\theta}}$ *such that* $\mu_J \to \mu$ *weakly as* $J \to \infty$.
(b) *For each* $I \ge 0, J \ge 1$*, the reduced tree* $r_{IJ}(\mathcal{T}^{\boldsymbol{\theta}})$ *has the same unconditional distribution as the* $\mathbf{T}_{IJ}$*-valued random tree defined as follows. Given* $\mathcal{T}^{\boldsymbol{\theta}}$ *and* μ*, let* $\{0*, 1*, \ldots, J*\}$ *be* $J+1$ *vertices chosen independently from distribution* μ*, take the spanning subtree of* $\{0*, 1*, \ldots, J*\}$ *and remove any labels* $i > I$.

Part (b) motivates our "sampled leaves" terminology for the vertices $j+$.

To see why we need $\boldsymbol{\theta} \in \Theta$ in Proposition 5, consider the particular case $\boldsymbol{\theta} = (1, 0, 0, \ldots)$. In this case $\mathcal{T}^{\boldsymbol{\theta}}$ contains hub 1 and, for each $0 \leq j < \infty$, sampled vertex $j+$ is linked to hub 1 by a different edge whose length has exponential(1) distribution, independently for each j. Clearly Proposition 5(a) fails for this $\boldsymbol{\theta}$.

The proof of Proposition 5 occupies the remainder of this section. We start with a technical lemma, which is the first place where the definition of Θ comes into play. Recall $d(\cdot)$ is distance on $\mathcal{T}^{\boldsymbol{\theta}}$.

Lemma 6 *(a) Let $\boldsymbol{\theta} \in \Theta$. For almost all realizations of $\mathcal{T}^{\boldsymbol{\theta}}$,*

$$\inf_{1 \leq j < \infty} d(0+, j+) = 0. \tag{5}$$

(b) Let $\boldsymbol{\theta} \in l^2_{\downarrow} \setminus \Theta$. For almost all realizations of $\mathcal{T}^{\boldsymbol{\theta}}$, the intersection $\cap_{j \geq 1}[[0+, j+]]$ of the paths from $0+$ to $j+$ has strictly positive ℓ-measure.

Proof. (a) Recall the construction of $\mathcal{T}^{\boldsymbol{\theta}}$. Fix $\varepsilon > 0$. It is enough to prove that there exists (almost surely) an interval $(\eta_j, \eta_{j+1}]$ of length $\leq \varepsilon$ which gets joined to some point in the interval $[0, \varepsilon]$. In the case $a > 0$ this is clear, because the desired event will occur whenever the Poisson point process on the octant contains points $(U_m, V_m), (U_{m+1}, V_{m+1})$ with $U_{m+1} < U_m + \varepsilon$ and $V_m < \varepsilon$. Now consider the case $a = 0$. By definition of Θ we have $\sum_i \theta_i = \infty$, and so we can define some (random) hub i as the smallest i such that $\xi_{i,1} \leq \varepsilon$. For $2 \leq k < \infty$ the branches $(\xi_{i,k} = \eta_{a_k}, \eta_{a_k+1}]$ are attached to hub i, so it is enough to show that the lengths L_k of these branches satisfy $L_k \to_P 0$. (We write $\to_P$ for convergence in probability.) But conditional on all the ξ-values in $[0, \xi_{i,k}]$, the distribution of L_k is stochastically smaller than exponential(s_k), where $s_k = \sum\{\theta_j : \xi_{j,1} \leq \xi_{i,k}\}$. Clearly $s_k \to_P \infty$ and so $L_k \to_P 0$ as required.

For (b), by assumption $a = 0$ and $\sum_i \theta_i < \infty$. From the construction of $\mathcal{T}^{\boldsymbol{\theta}}$,

$$\ell(\cap_{j \geq 1}[[0+, j+]]) \geq \inf_{i \geq 1} \xi_{i,1}.$$

But the right side has exponential (rate $\sum_i \theta_i$) distribution and hence is a.s. strictly positive. $\square$

For a permutation π of $\{0, 1, \ldots, J\}$, the operation "for each j, relabel leaf $j+$ as leaf $\pi(j)+$" defines a map from $\mathbf{T}_{IJ}$ to $\mathbf{T}_{IJ}$. (Declare the map to take the conventional state ∂ to itself.) Call a probability distribution on $\mathbf{T}_{IJ}$ *leaf-exchangeable* if it is invariant under this map, for each π. Proposition 2 easily implies

Corollary 7 *$r_{IJ}(\mathcal{T}^{\boldsymbol{p}})$ has leaf-exchangeable distribution, for any birthday tree $\mathcal{T}^{\boldsymbol{p}}$.*

The invariance property is preserved under weak convergence, so Proposition 3 and Lemma 4 imply that Corollary 7 can be passed to the limit:

Corollary 8 *$r_{IJ}(\mathcal{T}^{\boldsymbol{\theta}})$ has leaf-exchangeable distribution, for any $\boldsymbol{\theta} \in l^2_{\downarrow}$.*

Note that this property is not at all obvious from the construction of $\mathcal{T}^{\boldsymbol{\theta}}$. Note also that for fixed I the family $(r_{IJ}(\mathcal{T}^{\boldsymbol{\theta}}), J \geq 1)$ has a consistency property: the subtree of $r_{IJ}(\mathcal{T}^{\boldsymbol{\theta}})$ spanned by leaves $\{0+, 1+, \ldots, (J-1)+\}$ is $r_{I,J-1}(\mathcal{T}^{\boldsymbol{\theta}})$. We now appeal to a general result on CRTs. Recall that, as in Section 2, a realization of a CRT is a metric space. There are minor differences between the hypotheses on trees in this paper and in [3] (we may have vertices of degree greater than 3, and we have the additional structure of labeled hubs), but these differences make no essential change to the proof.

Theorem 9 ([3] Theorem 3 and Lemma 7) *Fix $I \geq 0$. Let $(\nu_J, J \geq 1)$ be leaf-exchangeable probability distributions on $\mathbf{T}_{IJ}$ satisfying the consistency condition*

for each $J \geq 1$ the subtree of a ν_J-distributed tree spanned by

vertices $0+, 1+, \ldots, (J-1)+$ is distributed as ν_{J-1}.

Suppose also that property (5) holds.
(a) There exists a CRT $\mathcal{T}$, and a probability measure μ on each realization of $\mathcal{T}$, such that for each $J \geq 1$ the subtree of $\mathcal{T}$ spanned by $V_0, V_1, \ldots, V_J$ has unconditional distribution ν_J, where $V_0, V_1, \ldots, V_J$ are (conditionally on $\mathcal{T}$) independent with distribution μ.
(b) Let μ_J be the empirical distribution of $\{V_0, V_1, \ldots, V_J\}$. For almost all realizations of $\mathcal{T}$,

$$\mu_J \to \mu \text{ weakly as } J \to \infty. \tag{6}$$

The argument in [3] uses the line-breaking construction for a general family satisfying the consistency condition (in [3] we viewed a continuum tree as a subset of sequence space l_1, obtained by attaching the successive line-segments orthogonally, but we may re-interpret the argument in terms of a metrization of $[0, \infty)$). With this re-interpretation, the continuum tree obtained by applying Theorem 9(a) to the consistent family in Corollary 8 is $\mathcal{T}^{\boldsymbol{\theta}}$. Then Proposition 5(a) follows from Theorem 9(b).

4 The fragmentation process

Analogous to the fragmentation process $\boldsymbol{Y}^{\boldsymbol{p}}(q)$ associated with the birthday tree (Section 3.1) is the idea of cutting the ICRT $\mathcal{T}^{\boldsymbol{\theta}}$ according to a Poisson(λ) process of cuts along its skeleton. Recall that associated with a realization of $\mathcal{T}^{\boldsymbol{\theta}}$ is the σ-finite length measure ℓ on its skeleton. So for $0 < \lambda < \infty$ we can construct a Poisson point process (and call the points "cuts") of mean measure $\lambda\ell(\cdot)$. The cuts partition the tree $\mathcal{T}^{\boldsymbol{\theta}}$ into a forest $\mathcal{F}^{\boldsymbol{\theta}}(\lambda)$ in which vertices v, w are in the same tree-component if the path $[[v, w]]$ does not contain any cut point. Define $\boldsymbol{Y}^{\boldsymbol{\theta}}(\lambda) = (Y_i^{\boldsymbol{\theta}}(\lambda), i \geq 1)$ to be the ranked μ-masses of the tree-components. For now we can assert only that $\boldsymbol{Y}^{\boldsymbol{\theta}}(\lambda)$ takes values in

$$\bar{\Delta} := \{(x_1, x_2, \ldots) : x_1 \geq x_2 \geq \ldots \geq 0, \sum_i x_i \leq 1\} \supset \Delta$$

because in principle there might be uncountably many components, each of μ-mass 0; Lemma 12 will show that in fact $\boldsymbol{Y}^{\boldsymbol{\theta}}(\lambda)$ is Δ-valued. We now define the *fragmentation process* $(\boldsymbol{Y}^{\boldsymbol{\theta}}(\lambda), 0 < \lambda < \infty)$ of the ICRT $\mathcal{T}^{\boldsymbol{\theta}}$ by coupling the cut-processes in the natural way. That is, we use a marked point process on skel($\mathcal{T}^{\boldsymbol{\theta}}$), with mark-space $[0, \infty)$, which is Poisson with mean intensity $\ell(\cdot)\times$ (Lebesgue measure); then in the definition of $\mathcal{F}^{\boldsymbol{\theta}}(\lambda)$ we require that the path $[[v, w]]$ does not contain any cut with mark $\hat{\lambda}$ which is *strictly* less than λ. This convention will give the desired path-continuity property – see end of section 4.3. We wish to use weak convergence techniques to deduce the continuum analog of Proposition 1.

Theorem 10 *For each $\boldsymbol{\theta} \in \Theta$ let $(\boldsymbol{Y}^{\boldsymbol{\theta}}(\lambda), 0 < \lambda < \infty)$ be the fragmentation process of the ICRT $\mathcal{T}^{\boldsymbol{\theta}}$. Define $\boldsymbol{X}^{\boldsymbol{\theta}}(t) = \boldsymbol{Y}^{\boldsymbol{\theta}}(e^{-t})$, $-\infty < t < \infty$. Then $(\boldsymbol{X}^{\boldsymbol{\theta}}(t), -\infty < t < \infty)$ is an additive coalescent.*

Theorem 10 generalizes the $\boldsymbol{\theta} = \boldsymbol{0}$ case which was the focus of [7]. The proof is given in Section 4.3, after recalling some background theory in the next section.

We are abusing notation by using the same symbols ($\mathcal{T}$ and $\boldsymbol{Y}$) for the continuum tree $\mathcal{T}^{\boldsymbol{\theta}}$ and its fragmentation process $\boldsymbol{Y}^{\boldsymbol{\theta}}(\cdot)$ as we used for the birthday tree $\mathcal{T}^{\boldsymbol{p}}$ and its fragmentation process $\boldsymbol{Y}^{\boldsymbol{p}}(\cdot)$. But the meaning should be clear from context.

4.1 Exchangeable random partitions

An equivalence relation $\sim$ on the set $\{0, 1, 2, \ldots\}$ can be identified with a *partition* of the set into equivalence classes. So a random equivalence relation may be identified with a random partition. Write Π for such a random partition, and Π_J for its restriction to $\{0, 1, \ldots, J\}$. There is a natural notion of an *exchangeable random partition* (the distribution of Π_J is invariant under permutations of $\{0, 1, \ldots, J\}$, for each J) going back to Kingman [15]. Kingman [15] essentially established the following results, in slightly different language (he gives (iii) in the context of Δ-valued frequencies, but the more general setting is similar).

Theorem 11 ([15]) (i) *Let Π be an exchangeable random partition. Then the limiting ranked frequencies*

$$F_i := \lim_{J\to\infty} \frac{\textit{size of } i\textit{'th largest class of } \Pi_J}{J+1}, \ i = 1, 2, \ldots$$

exist a.s. and $(F_i, i \geq 1)$ is a random element of $\bar{\Delta}$.
(ii) $P((F_i, i \geq 1) \in \Delta) = 1$ *iff* $P(\{0\}$ *is a class of* $\Pi) = 0$.
(iii) *Let $(\Pi^n, 1 \leq n \leq \infty)$ be a sequence of exchangeable random partitions. Then as $n \to \infty$*

$$\Pi^n_J \xrightarrow{d} \Pi^\infty_J \textit{ for each } J$$

if and only if

$$(F^n_i, i \geq 0) \xrightarrow{d} (F^\infty_i, i \geq 0) \textit{ on } \bar{\Delta}$$

where $\bar{\Delta}$ is given the topology of co-ordinatewise convergence.

This setup provides a different way of viewing the fragmentation process. Fix $\boldsymbol{\theta} \in l^2_\downarrow$ and $0 < \lambda < \infty$, and consider the ICRT $\mathcal{T}^{\boldsymbol{\theta}}$ with a Poisson process of cuts with rate λ per unit ℓ-length. Define a random equivalence relation on $\{0, 1, 2, \ldots\}$ by:

$$i \sim j \text{ iff there is no cut on the path from } i+ \text{ to } j+ \text{ in } \mathcal{T}^{\boldsymbol{\theta}}. \tag{7}$$

By leaf-exchangeability (Corollary 8) the associated partition $\Pi^{\boldsymbol{\theta}}$ into equivalence classes is an exchangeable random partition. The vector

$$\left(\frac{\text{size of } i\text{'th largest class of } \Pi_J}{J+1}, \ i = 1, 2, \ldots\right)$$

is the ranked vector

$$(\mu_J(A) : A \text{ a tree-component of } \mathcal{F}^{\boldsymbol{\theta}}(\lambda))$$

for μ_J the empirical distribution on $\{0+, 1+, \ldots, J+\}$. By Proposition 5(a), when $\boldsymbol{\theta} \in \Theta$ we have $P(\mu_J \to \mu \text{ weakly}) = 1$, and it easily follows that we can identify the limit ranked frequencies in Theorem 11(i) as the vector $\boldsymbol{Y}^{\boldsymbol{\theta}}(\lambda)$ of ranked μ-masses in the fragmentation of $\mathcal{T}^{\boldsymbol{\theta}}$.

Remark. In fact we could avoid introducing μ by taking this limit to be the *definition* of $\boldsymbol{Y}^{\boldsymbol{\theta}}(\lambda)$, and indeed for $\boldsymbol{\theta} \notin \Theta$ we use this definition in the following lemma. Avoiding discussion of μ would make the argument shorter, but introducing μ makes the fragmentation process easier to visualize.

Lemma 12 *$P(\boldsymbol{Y}^{\boldsymbol{\theta}}(\lambda) \in \Delta) = 1$ iff $\boldsymbol{\theta} \in \Theta$.*

Proof. Because $P(0 \sim j) = \exp(-\lambda d(0+, j+))$, Lemma 6 implies

$$P(\{0\} \text{ is a class of } \Pi^{\boldsymbol{\theta}}) = 0 \text{ iff } \boldsymbol{\theta} \in \Theta.$$

Apply Theorem 11(iii). □

We can make analogous exchangeable random partitions in the discrete setting of Section 3.1. Fix $\boldsymbol{p}$ and q. Recall that $\mathcal{F}^{\boldsymbol{p}}(q)$ is the random forest obtained by independently cutting each edge of the birthday tree $\mathcal{T}^{\boldsymbol{p}}$ with probability q. Recall also the i.i.d.($\boldsymbol{p}$) sequence of vertices $(W_{R_j-1}, 0 \le j < \infty)$, with $R_0 = 1$, featuring in Proposition 2. Write $\Pi^{\boldsymbol{p}}$ for the exchangeable random partition associated with the equivalence relation

$$i \sim j \text{ iff there is no cut edge on the path from } W_{R_i-1} \text{ to } W_{R_j-1} \text{ in } \mathcal{T}^{\boldsymbol{\theta}}.$$

The limit ranked frequencies are now the vector $\boldsymbol{Y}^{\boldsymbol{p}}(q)$ in the discrete fragmentation process.

4.2 Weak convergence with mass measures

Now consider a sequence $(\boldsymbol{p}_n)$ of discrete probability distributions satisfying the asymptotic regime (3), that is

$$\lim_{n\to\infty} \sigma_n = 0, \quad \lim_{n\to\infty} \frac{p_{ni}}{\sigma_n} = \theta_i, \; i \ge 1, \text{ where } \sigma_n := \sqrt{\sum_i p_{ni}^2}.$$

Applying Proposition 3 with $I = 0$ (the hub-labels are irrelevant here) gives

$$r_{0J}(\mathcal{T}^{\boldsymbol{p}_n}) \xrightarrow{d} r_{0J}(\mathcal{T}^{\boldsymbol{\theta}}) \text{ on } \mathbf{T}_{0J}.$$

Recall we have fixed λ, and take q_n such that $q_n/\sigma_n \to \lambda$. Consider the discrete fragmentation $\boldsymbol{Y}^{\boldsymbol{p}_n}(q_n)$. The Bernoulli (rate q_n per edge of length σ_n) process of cuts of the edges of $r_{0J}(\mathcal{T}^{\boldsymbol{p}_n})$ converges to the Poisson (rate λ per unit ℓ-length) process of cuts of $r_{0J}(\mathcal{T}^{\boldsymbol{\theta}})$.

It follows that $\Pi_J^{\boldsymbol{p}_n} \xrightarrow{d} \Pi_J^{\boldsymbol{\theta}}$ as $n \to \infty$. Theorem 11(iii) then implies $\boldsymbol{Y}^{\boldsymbol{p}_n}(q_n) \xrightarrow{d} \boldsymbol{Y}^{\boldsymbol{\theta}}(\lambda)$ on $\bar{\Delta}$. If $\boldsymbol{\theta} \in \Theta$ then Lemma 12 shows $\boldsymbol{Y}^{\boldsymbol{\theta}}(\lambda)$ is Δ-valued, implying that in fact convergence holds on Δ. We have thus proved the first assertion of the following Proposition; the second assertion is similar, using coupled Bernoulli processes of cuts converging to coupled Poisson processes of cuts.

Proposition 13 *Under the asymptotic regime* (3) *with* $\boldsymbol{\theta} \in \Theta$*, if* $q_n/\sigma_n \to \lambda \in (0, \infty)$ *then* $\boldsymbol{Y}^{\boldsymbol{p}_n}(q_n) \xrightarrow{d} \boldsymbol{Y}^{\boldsymbol{\theta}}(\lambda)$ *on* Δ*. More generally, if* $0 < \lambda_1 < \lambda_2 < \ldots < \lambda_d$ *and* $q_{n,i}/\sigma_n \to \lambda_i$ *for each* $1 \le i \le d$ *then* $(\boldsymbol{Y}^{\boldsymbol{p}_n}(q_{n,1}), \ldots, \boldsymbol{Y}^{\boldsymbol{p}_n}(q_{n,d})) \xrightarrow{d} (\boldsymbol{Y}^{\boldsymbol{\theta}}(\lambda_1), \ldots, \boldsymbol{Y}^{\boldsymbol{\theta}}(\lambda_d))$.

A converse to Proposition 13 will be needed in Section 5.3.

Lemma 14 *Suppose* $(\boldsymbol{p}_n)$ *satisfies the asymptotic regime* (3) *for some* $\boldsymbol{\theta} \in l^2_\downarrow$*. Suppose, for some* $\phi_n \to 1$*,*

$$\boldsymbol{Y}^{\boldsymbol{p}_n}(\phi_n \sigma_n) \xrightarrow{d} \boldsymbol{Y} \text{ (say) on } \Delta.$$

Then $\boldsymbol{\theta} \in \Theta$*.*

Proof. As in the argument above, $\Pi_J^{\boldsymbol{p}_n} \xrightarrow{d} \Pi_J^{\boldsymbol{\theta}}$ as $n \to \infty$, and then Theorem 11(iii) implies that $\boldsymbol{Y}$ is the vector of ranked frequencies of the classes of $\Pi^{\boldsymbol{\theta}}$. Since $\boldsymbol{Y}$ is Δ-valued, Theorem 11(ii) shows $P(\{0\}$ is a class of $\Pi^{\boldsymbol{\theta}}) = 0$. Then Lemma 6(b) implies $\boldsymbol{\theta} \in \Theta$. $\square$

4.3 Proof of Theorem 10.

Fix $\boldsymbol{\theta} \in \Theta$. By Lemma 4 there exists a sequence $(\boldsymbol{p}_n)$ satisfying the asymptotic regime (3) with limit $\boldsymbol{\theta}$. Proposition 13 implies that

$$(\boldsymbol{Y}^{\boldsymbol{p}_n}(\lambda\sigma_n),\ 0 \le \lambda \le 1/\sigma_n) \xrightarrow{d} (\boldsymbol{Y}^{\boldsymbol{\theta}}(\lambda),\ 0 \le \lambda < \infty) \tag{8}$$

in the sense of convergence of f.d.d.'s. Proposition 1 showed that, for any ranked $\boldsymbol{p}$, $(\boldsymbol{Y}^{\boldsymbol{p}}(e^{-t}), 0 \le t < \infty)$ is the additive coalescent started at state $\boldsymbol{p}$ at time 0. So $(\boldsymbol{Y}^{\boldsymbol{p}_n}(e^{-t}\sigma_n), \log \sigma_n \le t < \infty)$ is the additive coalescent started at state $\boldsymbol{p}_n$ at time $\log \sigma_n$. Note that $\log \sigma_n \to -\infty$. Then by (8) and the Feller property ([13] Theorem 10) of the additive coalescent, the limit process $\boldsymbol{X}^{\boldsymbol{\theta}}(t) = \boldsymbol{Y}^{\boldsymbol{\theta}}(e^{-t}),\ -\infty < t < \infty$ is indeed an additive coalescent. Using the "strictly less" convention in the definition of $\mathcal{F}^{\boldsymbol{\theta}}(\lambda)$, it is not hard to check that, after modifying on a null set, the sample paths $\lambda \to \boldsymbol{Y}^{\boldsymbol{\theta}}(\lambda)$ are left-continuous with right limits. In other words

$$\text{the sample paths } t \to \boldsymbol{X}^{\boldsymbol{\theta}}(t) \text{ are càdlàg .} \tag{9}$$

5 The entrance boundary of the additive coalescent

Call an additive coalescent defined for $-\infty < t < \infty$ *eternal*. General Markov process theory (see e.g. [12, §10] for a concise treatment) says that any eternal additive coalescent is a mixture of *extreme* eternal additive coalescents, and the extreme ones (which form the *entrance boundary*) are characterized by the property that the tail σ-field at time $-\infty$ is trivial. Our main theorem gives a complete description of the entrance boundary. The first part repeats the assertions of Theorem 10, with the extra assertion of extremality.

Theorem 15 *For each $\boldsymbol{\theta} \in \Theta$ let $(\boldsymbol{Y}^{\boldsymbol{\theta}}(\lambda), 0 < \lambda < \infty)$ be the fragmentation process of the ICRT $\mathcal{T}^{\boldsymbol{\theta}}$. Define $\boldsymbol{X}^{\boldsymbol{\theta}}(t) = \boldsymbol{Y}^{\boldsymbol{\theta}}(e^{-t}),\ -\infty < t < \infty$. Then for each real t_0 the process $(\boldsymbol{X}^{\boldsymbol{\theta}}(t - t_0), -\infty < t < \infty)$ is an extreme additive coalescent.*

Conversely, if $\boldsymbol{X} = (\boldsymbol{X}(t), -\infty < t < \infty)$ is an extreme additive coalescent then either $\boldsymbol{X} \stackrel{d}{=} (\boldsymbol{X}^{\boldsymbol{\theta}}(t - t_0),\ -\infty < t < \infty)$ for some $\boldsymbol{\theta} \in \Theta$ and $-\infty < t_0 < \infty$ or else $\boldsymbol{X}$ is the constant process $\boldsymbol{X}(t) = (1, 0, 0, \ldots)\ \forall t$.

Remarks. Evans and Pitman [13] show that the additive coalescent can be taken to be a càdlàg process, while (9) shows that the fragmentation construction of $\boldsymbol{X}^{\boldsymbol{\theta}}$ in Theorem 15 gives a càdlàg version. So we may just assume processes in Section 5 are càdlàg .

The proof of Theorem 15 rests upon an analysis of the $t \to -\infty$ behavior of $\boldsymbol{X}^{\boldsymbol{\theta}}(t)$ using explicit calculations in Section 5.1, and an analysis of the $t \to -\infty$ behavior of a general additive coalescent using stochastic calculus in Section 5.2. The proof is completed in Section 5.3.

5.1 Behavior of the fragmentation process as $\lambda \to \infty$

Recall $Y_i^{\boldsymbol{\theta}}(\lambda)$ is the ith largest μ-measure of the components of $\mathcal{F}^{\boldsymbol{\theta}}(\lambda)$; we may also write $Y_{(i)}^{\boldsymbol{\theta}}(\lambda)$ for the μ-measure of the component of $\mathcal{F}^{\boldsymbol{\theta}}(\lambda)$ containing hub i. Recall $\to_P$ denotes convergence in probability.

Proposition 16 *For each $\boldsymbol{\theta} \in \Theta$*
(a) $\lambda^2 \sum_i (Y_i^{\boldsymbol{\theta}}(\lambda))^2 \ \to_P\ 1$ *as* $\lambda \to \infty$;
(b) $\lambda Y_{(i)}^{\boldsymbol{\theta}}(\lambda) \ \to_P\ \theta_i$ *as* $\lambda \to \infty$.
(c) *Writing $Q_k(\lambda)$ for the largest μ-measure of a component of $\mathcal{F}^{\boldsymbol{\theta}}(\lambda)$ which does not contain hub i for any $1 \le i \le k$,*

$$\lim_{k\to\infty} \limsup_{\lambda\to\infty} P(\lambda Q_k(\lambda) > \varepsilon) = 0, \quad \varepsilon > 0.$$

Since any two different hubs are ultimately (as $\lambda \to \infty$) in different components, (b) and (c) easily imply

$$\lambda Y_i^{\boldsymbol{\theta}}(\lambda) \ \to_P \ \theta_i \text{ as } \lambda \to \infty.$$

For future reference we rewrite this in terms of the associated additive coalescent $\boldsymbol{X}^{\boldsymbol{\theta}}(t) := \boldsymbol{Y}^{\boldsymbol{\theta}}(e^{-t})$.

Corollary 17

$$\begin{aligned} e^{-2t}\sum_i (X_i^{\boldsymbol{\theta}}(t))^2 \quad &\to_P \quad 1 \ \textit{as } t \to -\infty \\ e^{-t}X_i^{\boldsymbol{\theta}}(t) \quad &\to_P \quad \theta_i \ \textit{as } t \to -\infty. \end{aligned}$$

The proof of Proposition 16 occupies the rest of the section. We need a few preliminaries. First, an easy lemma in analysis.

Lemma 18 *Let $f : [0,\infty) \to [0,\infty)$ be measurable with $\int_0^\infty f(t)dt < \infty$.*
(a) $\quad \int_0^\infty \int_0^\infty \frac{u^3}{6} e^{-\lambda u} f(t+u)\ dt\ du \sim \lambda^{-4}\int_0^\infty f(t)dt$ *as* $\lambda \to \infty$.
(b) If f is continuous at 0 then $\int_0^\infty \frac{u^q}{q!}e^{-\lambda u}f(u)du \sim \lambda^{-q-1}f(0)$ as $\lambda \to \infty$, $q = 1,2,\ldots$.

Next we describe a probability law (Figure 5 and (10)) which will arise later as a limit (specifically, as a limit of the distance between branchpoints in the spanning tree on $\{0+, 1+, 2+, 3+\}$: see Figures 6 - 8). For distinct $i, j \geq 1$ write $h_{ij}(t)$ for the density function of the distance between hub i and hub j in $\mathcal{T}^{\boldsymbol{\theta}}$. Write $h_{0i}(t)$ for the density function of the distance between sampled leaf $0+$ and hub i, and write $h_{00}(t)$ for the density function of the distance between sampled leaf $0+$ and sampled leaf $1+$. Recall that the mass measure μ assigns no mass to the hubs of $\mathcal{T}^{\boldsymbol{\theta}}$ (else sampled leaves would coincide with hubs with non-zero probability). Because $a + \sum_i \theta_i^2 = 1$, we can define a new probability law ν on each realization of $\mathcal{T}^{\boldsymbol{\theta}}$ by: ν is the superposition of $a\mu(\cdot)$ and the measure putting mass θ_i^2 on each hub i. Now consider picking independently two points from law ν on the same realization of $\mathcal{T}^{\boldsymbol{\theta}}$, noting whether the points are hubs or unlabeled vertices, and drawing the spanning tree on these two points with the first-picked point on the left. Figure 5 illustrates the possibilities: t denotes the edge-length and the formulas give the probability density functions (in the last case, the probability).

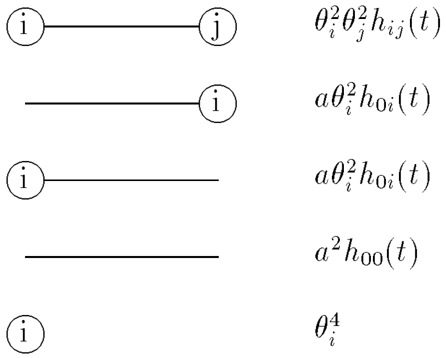


Figure 5 (the edge has length t)

By construction this is a *probability* law, that is

$$\sum_{i\geq 1}\sum_{j\geq 1, j\neq i}\int_0^\infty \theta_i^2\theta_j^2 h_{ij}(t)dt + 2\sum_{i\geq 1}\int_0^\infty a\theta_i^2 h_{0i}(t)dt + \int_0^\infty a^2 h_{00}(t)dt \ + \sum_{i\geq 1}\theta_i^4 = 1. \quad (10)$$

Now define

$$H(t) = e^{-at^2/2}\prod_{i\geq 1}\left(e^{-\theta_i t}(1+\theta_i t)\right) \quad (11)$$

$$h(t) = -\frac{d}{dt}H(t) = \left(a + \sum_i \tfrac{\theta_i^2}{1+\theta_i t}\right)\ tH(t). \quad (12)$$

From the line-breaking construction (Section 2) of the ICRT $\mathcal{T}^{\boldsymbol{\theta}}$, the distance ($\eta_1$ in the construction) between sampled leaves 0+ and 1+ has distribution function $1 - H(t)$ and thus has density function $h_{00}(t) = h(t)$. More generally, it can be deduced from [1, Corollary 3 and Proposition 14] that the density function $h_{ij}(t)$ for the distance between hubs i and j is

$$h_{ij}(t) = \frac{1}{(1+\theta_i t)(1+\theta_j t)}\left(h(t) + \left(\tfrac{\theta_i}{1+\theta_i t} + \tfrac{\theta_j}{1+\theta_j t}\right) H(t)\right) \quad (13)$$

and this formula gives also the densities h_{0i} and h_{00} involving sampled leaves, by setting $\theta_0 = 0$.

Turning to the proof of Proposition 16(a), write $S(\lambda) = \sum_i (Y_i(\lambda))^2$. We want to prove

$$\lambda^2 S(\lambda) \ \to_P \ 1 \text{ as } \lambda \to \infty. \quad (14)$$

Write $A_{01}(\lambda)$ for the event that sampled leaves 0+ and 1+ are in the same component of the forest $\mathcal{F}^{\boldsymbol{\theta}}(\lambda)$ underlying $\boldsymbol{Y}(\lambda)$, and write $A_{23}(\lambda)$ for the event that sampled leaves 2+ and 3+ are in the same component of that forest. Using Proposition 5(b) we see

$$\begin{aligned} ES(\lambda) &= P(A_{01}(\lambda)) \\ ES^2(\lambda) &= P(A_{01}(\lambda) \cap A_{23}(\lambda)). \end{aligned}$$

In terms of the distance $d(0+,1+)$ between sampled leaves 0+ and 1+,

$$P(A_{01}(\lambda)) = E\exp(-\lambda d(0+,1+)) = \int_0^\infty e^{-\lambda t} h(t) dt$$

$$= \int_0^\infty t e^{-\lambda t} \left(a + \sum_i \frac{\theta_i^2}{1+\theta_i t} \right) H(t) dt \sim \lambda^{-2} H(0+) = \lambda^{-2} \text{ as } \lambda \to \infty,$$

the asymptotic equivalence by Lemma 18(b).

Thus to prove (14) via Chebyshev's inequality, it is enough to prove

$$P(A_{01}(\lambda) \cap A_{23}(\lambda)) \sim \lambda^{-4} \text{ as } \lambda \to \infty. \tag{15}$$

Consider the spanning tree on sampled leaves $\{0+, 1+, 2+, 3+\}$. Figure 6 illustrates two possible shapes for this spanning tree.

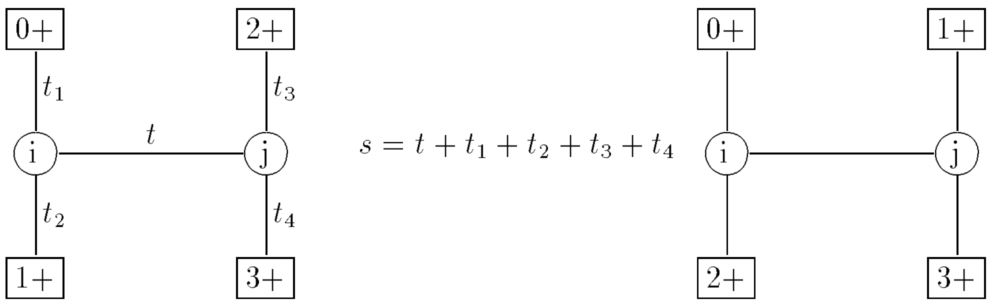


Figure 6

On the sets of trees with these shapes, we can calculate the density function for edge-lengths (t, t_1, t_2, t_3, t_4) from the line-breaking construction. To get the tree on the left side of Figure 6 we need (in the notation of Section 2)

$$\xi_{i,1} = t_1, \ \xi_{i,2} = t_1 + t_2, \ \xi_{j,1} = t_1 + t_2 + t, \ \xi_{j,2} = t_1 + t_2 + t + t_3 \tag{16}$$

$$\min(U_1, \xi_{i,3}, \xi_{j,3}, \xi_{m,2}, m \neq i, j) = s. \tag{17}$$

The density function is

$$\theta_i^2 e^{-\theta_i s}\ \theta_j^2 e^{-\theta_j s}\ e^{-as^2/2}\ \times$$

$$\left((as+\theta_i+\theta_j) \prod_{m\neq i,j} \left(e^{-\theta_m s}(1+\theta_m s)\right) \ + \ \sum_{k\neq i,j} \theta_k^2 s e^{-\theta_k s} \prod_{m\neq i,j,k} \left(e^{-\theta_m s}(1+\theta_m s)\right)\right).$$

Here the first line is the density corresponding to (16) and the event $\min(U_1, \xi_{i,3}, \xi_{j,3}) \geq s$ and the second line is the conditional density of (17): the term $(as+\theta_i+\theta_j)$ reflects the possibility that the minimum is attained by U_1 or $\xi_{i,3}$ or $\xi_{j,3}$, while the term $\theta_k^2 s e^{-\theta_k s}$ reflects the possibility that the minimum is attained by $\xi_{k,2}$. Rewriting this density in terms of H gives

$$\frac{\theta_i^2}{1+\theta_i s}\ \frac{\theta_j^2}{1+\theta_j s}\ \left(as+\theta_i+\theta_j+\sum_{k\neq i,j} \frac{\theta_k^2}{1+\theta_k s}\right)\ H(s).$$

Comparing this with the formula (12) for $h(t)$, and noting that $c - \frac{c^2 s}{1+cs} = \frac{c}{1+cs}$, we can rewrite it as

$$\frac{\theta_i^2}{1+\theta_i s}\ \frac{\theta_j^2}{1+\theta_j s} \left(h(s) \ + \ \left(\frac{\theta_i}{1+\theta_i s} + \frac{\theta_j}{1+\theta_j s}\right) H(s)\right) = \theta_i^2\theta_j^2 h_{ij}(s)$$

for $h_{ij}(s)$ at (13). Now consider the contribution to $P(A_{01}(\lambda)\cap A_{23}(\lambda))$ from trees of this shape. The events happen if none of the four edges to the leaves contain a point from the Poisson(λ) process of cuts, and so the contribution is

$$\int \ldots \int \exp(-\lambda(t_1+t_2+t_3+t_4))\ \theta_i^2\theta_j^2 h_{ij}(t+t_1+t_2+t_3+t_4)\ dt dt_1 dt_2 dt_3 dt_4$$

$$= \theta_i^2\theta_j^2 \int\int e^{-\lambda u} h_{ij}(t+u) \tfrac{u^3}{6} du dt \quad \sim \ \lambda^{-4}\theta_i^2\theta_j^2 \ \int h_{ij}(t)\ dt \text{ by Lemma 18.}$$

If the spanning tree has the shape on the *right* side of Figure 6, then it would be required that none of the *five* edges contains a cut, and this chance works out as $O(\lambda^{-5})$.

Thus our strategy for proving (15) is to show that the coefficients of λ^{-4} in the contributions to $P(A_{01}(\lambda)\cap A_{23}(\lambda))$ from different shapes of the spanning tree on sampled leaves $\{0+, 1+, 2+, 3+\}$ are the terms of the probability law (10), arising from the different possibilities for the "edge between branchpoints" in Figure 5. The argument above shows that, for spanning trees with the shape shown in Figure 6, the coefficient is as stated in the top line of Figure 5. Minor modifications of the argument above show that, for spanning trees with the shapes shown in Figure 7, the coefficients are as stated in the middle three lines of Figure 5.

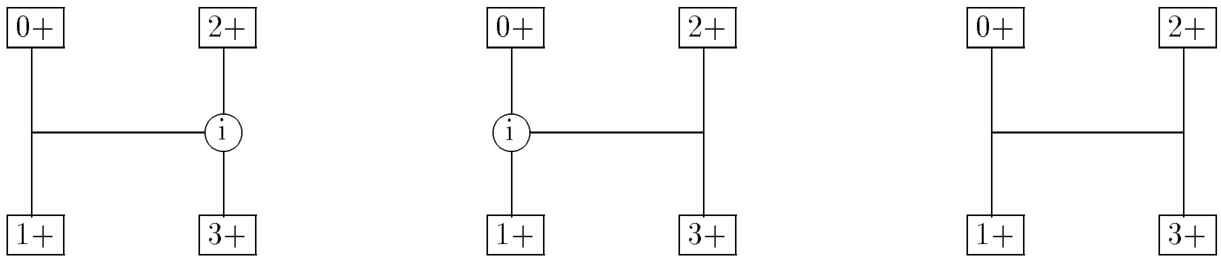
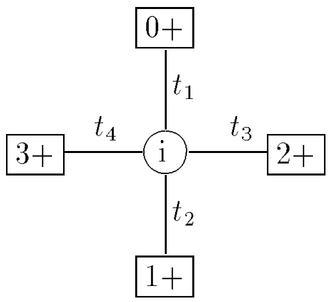


Figure 7

The remaining case is as shown in Figure 8.



$$s = t_1 + t_2 + t_3 + t_4$$

Figure 8

Here the density function on edge-lengths turns out to be $f(t_1 + t_2 + t_3 + t_4)$ for

$$f(s) = \frac{\theta_i^3}{1 + \theta_i s} \left(h(s) + \tfrac{\theta_i}{1+\theta_i s} H(s) \right).$$

So the contribution to $P(A_{01}(\lambda) \cap A_{23}(\lambda))$ from trees of this shape is

$$\int \dots \int e^{-\lambda(t_1+t_2+t_3+t_4)} \ f(t_1 + t_2 + t_3 + t_4) \ dt_1 dt_2 dt_3 dt_4$$

$$= \int \tfrac{u^3}{6} f(u) e^{-\lambda u} du \sim \lambda^{-4} f(0) = \theta_i^4 \lambda^{-4} \text{ by Lemma 18 .}$$

This matches the corresponding term in (10), completing the proof of part (a) of Proposition 16.

The proof of part (b) is similar but easier; we just give an outline.

$$EY_{(i)}(\lambda) \quad = \quad E \exp(-\lambda d(0+, i))$$

$$\begin{aligned} &= \int e^{-\lambda t} h_{0i}(t)dt \\ &\sim \lambda^{-1} h_{0i}(0) = \theta_i \lambda^{-1}. \end{aligned}$$

To prove (b) via Chebyshev's inequality, the required variance calculation becomes the following. Writing L for the length of the spanning tree on sampled leaves $0+$ and $1+$ and hub i, we need to show

$$E \exp(-\lambda L) \sim \theta_i^2 \lambda^{-2}.$$

To prove this it suffices to show

$$P(L \le t, \Omega) \sim \theta_i^2 t^2/2 \text{ as } t \to 0 \tag{18}$$

$$P(L \le t, \Omega^c) = o(t^2) \text{ as } t \to 0 \tag{19}$$

where Ω is the event that the path from $0+$ to $1+$ goes through hub i. Proving (18) is straightforward, because the line-breaking construction gives a formula for the density $P(L \in dt, \Omega)$. For (19), on Ω^c the spanning tree on sampled leaves $0+$ and $1+$ and hub i has a branchpoint; write $\tilde{L}$ for the distance from the branchpoint to hub i. Because

$$P(d(0+, 1+) \le t) = 1 - H(t) \sim \tfrac{1}{2} t^2 \text{ as } t \to 0$$

to prove (19) it suffices to prove

$$P(\tilde{L} \le s, \Omega^c | d(0+, 1+) \le t) \to \int_0^s \tilde{h}(u)\, du \text{ as } t \to 0 \tag{20}$$

for some sub-probability density $\tilde{h}$. But we can take this limit within the line-breaking construction. There $d(0+, 1+) = \eta_1$ and as $t \to 0$

$$\begin{aligned} P(\eta_1 = U_1 | \eta_1 = t) &\to a \\ P(\eta_1 = \xi_{j,2} | \eta_1 = t) &\to \theta_j^2. \end{aligned}$$

It is not hard to deduce that (20) holds with

$$\tilde{h} = a h_{00} + \sum_{j \ne i} \theta_j^2 h_{0j}.$$

To prove (c), observe first that

$$E Q_k^3(\lambda) \le P(B_k(\lambda))$$

where $B_k(\lambda)$ is the event

0+ and 1+ and 2+ are in the same component of the forest $\mathcal{F}(\lambda)$, but this component does not contain hub i for any $1 \leq i \leq k$.

In the notation of the line-breaking construction of $\mathcal{T}^{\boldsymbol{\theta}}$,

$$P(B_k(\lambda)) = E\exp(-\lambda\eta_2)\ I(\eta_2 < \min_{1\leq i\leq k} \xi_{i,1}) = \int_0^\infty e^{-\lambda t} dV_k(t)$$

where $I(\cdot)$ denotes indicator r.v. and

$$V_k(t) := P(\eta_2 \leq t, \eta_2 < \min_{1\leq i\leq k} \xi_{i,1}).$$

Then

$$\begin{aligned} V_k(t) &\leq P(U_2 \leq t) + P(U_1 \leq t, \min_{i>k} \xi_{i,2} \leq t) + P(\min_{i>k} \xi_{i,3} \leq t) \\ &\leq (at^2/2)^2 + (at^2/2)(\sum_{i>k} \theta_i^2 t^2) + \sum_{i>k} \theta_i^3 t^3 \\ &= \sum_{i>k} \theta_i^3 t^3 + O(t^4) \text{ as } t \to 0. \end{aligned}$$

Then

$$P(B_k(\lambda)) = \int e^{-\lambda t} dV_k(t) = \lambda \int e^{-\lambda t} V_k(t) dt = O(\lambda^{-3} \sum_{i>k} \theta_i^3) \text{ as } \lambda \to \infty.$$

So

$$\lim_{k\to\infty} \limsup_{\lambda\to\infty} \lambda^3 E Q_k^3(\lambda) = 0$$

establishing part (c).

5.2 Stochastic analysis of eternal additive coalescents

This section gives a "stochastic calculus" analysis of an eternal additive coalescent, analogous to (but technically simpler than) the analysis of eternal multiplicative coalescents in [6, §3].

Notation. Write $E(dZ(t)|\mathcal{G}(t)) = a(t)dt$ and var $(dZ(t)|\mathcal{G}(t)) = b(t)dt$ to mean $M(t) := Z(t) - \int_0^t a(s)ds$ is an $\mathcal{G}(t)-$martingale with quadratic variation $\langle M(t), M(t)\rangle = \int_0^t b(s)ds$. We may also write for instance var $(dZ(t)|\mathcal{G}(t)) \leq \beta(t)dt$ to indicate that var $(dZ(t)|\mathcal{G}(t)) = b(t)dt$ for some $b(t)$ with $0 \leq b(t) \leq \beta(t)$.

We quote a version of the L^2 convergence theorem for (reversed) martingales.

Lemma 19 *Let $(Z(t); -\infty < t \leq 0)$ be a càdlàg process adapted to $(\mathcal{G}(t))$ and satisfying*

$$|E(dZ(t)|\mathcal{G}(t))| \leq \alpha(t)\ dt, \quad \mathrm{var}\ (dZ(t)|\mathcal{G}(t)) \leq \beta(t)\ dt.$$

If

$$\int_{-\infty}^{0} \alpha(t)\ dt < \infty \text{ a.s. and } \int_{-\infty}^{0} \beta(t)\ dt < \infty \text{ a.s.}$$

then $\lim_{t\to-\infty} Z(t)$ exists and is finite a.s.

Write $\boldsymbol{X}(t) = (X_i(t), i \geq 1)$ for a ranked additive coalescent parameterized by t in some interval I and let $(\mathcal{G}(t), t \in I)$ denote the filtration generated by $(\boldsymbol{X}(t), t \in I)$. Write

$$Q(t) = \sum_i X_i^2(t)$$

$$S_3(t) = \sum_i X_i^3(t).$$

Lemma 20 *Writing $Y(t)$ for the size of cluster containing a specified atom,*

$$E(dQ(t)|\mathcal{G}(t)) = 2(Q(t) - S_3(t))\ dt \tag{21}$$

$$E(dY(t)|\mathcal{G}(t)) = (Y(t) + Q(t) - 2Y^2(t))\ dt \tag{22}$$

$$\mathrm{var}\ (dQ(t)|\mathcal{G}(t)) \leq 4Q(t)S_3(t)\ dt \tag{23}$$

$$\mathrm{var}\ (dY(t)|\mathcal{G}(t)) \leq (Y(t)Q(t) + S_3(t))\ dt \tag{24}$$

Proof. The argument is similar to arguments in [13, Section 6.2]. Because coalescence of clusters of masses x and y causes $Q(\cdot)$ to increase by $2xy$,

$$\begin{aligned} E(dQ(t)|\mathcal{G}(t)) &= \sum_i \sum_{j>i} (2X_i(t)X_j(t))\ (X_i(t) + X_j(t))\ dt \\ &= 2\sum_i \sum_{j\neq i} X_i^2(t)X_j(t)\ dt \\ &= 2\sum_i X_i^2(t)(1 - X_i(t))\ dt \end{aligned}$$

giving (21). And

$$\begin{aligned} \mathrm{var}\ (dQ(t)|\mathcal{G}(t)) &= \sum_i \sum_{j>i} (2X_i(t)X_j(t))^2\ (X_i(t) + X_j(t))\ dt \\ &= 4\sum_i \sum_{j\neq i} X_i^3(t)X_j^2(t)\ dt \\ &\leq 4S_3(t)Q(t)\ dt \end{aligned}$$

giving (23). Similar calculations give (22,24). $\square$

By combining the estimates in Lemma 20 with the convergence criteria in Lemma 19 we shall prove

Proposition 21 *Let $(\boldsymbol{X}(t), -\infty < t < \infty)$ be an extreme eternal additive coalescent which is not the constant process $\boldsymbol{X}(t) = (1, 0, 0, \ldots)$. Then as $t \to -\infty$*

$$e^{-2t} \sum_i X_i^2(t) \;\; \to \;\; \eta^2 \;\; a.s. \tag{25}$$

$$e^{-t} X_i(t) \;\; \to \;\; \eta_i \;\; a.s., \; each \; i \geq 1 \tag{26}$$

where $\eta > 0$ and $\eta_1 \geq \eta_2 \geq \ldots \geq 0$ are constants and $\sum_i \eta_i^2 \leq \eta^2 < \infty$.

Proof. We will prove this for possibly random limits η, η_i, but then by extremality the η's must be constants.

Any jump $\Delta Q(t) = Q(t) - Q(t-)$ satisfies $\Delta Q(t) \leq 2X_1^2(t-) \leq 2Q(t-)$. It follows that

$$\frac{\Delta Q(t)}{Q(t-)} \geq \Delta \log Q(t) \geq \frac{\Delta Q(t)}{Q(t-)} - c\left(\frac{\Delta Q(t)}{Q(t-)}\right)^2 \tag{27}$$

for some constant c. Now consider

$$Z(t) = -2t + \log Q(t).$$

Combining (27) with (21) and the bound $S_3(t) \leq X_1(t)Q(t)$ gives

$$0 \geq E(dZ(t)|\mathcal{G}(t)) \geq -2X_1(t)\; dt - cR(t)\; dt$$

where $R(t)$ is the contribution from the $(\cdot)^2$ term of (27). And

$$\begin{aligned} R(t) \;\; &= \;\; \frac{1}{Q^2(t)} \sum_i \sum_{j>i} (2X_i(t)X_j(t))^2 \; (X_i(t) + X_j(t)) \\ &\leq \;\; \frac{1}{Q^2(t)} \, 4S_3(t)Q(t) \;\; \leq \;\; 4X_1(t). \end{aligned}$$

So

$$0 \geq E(dZ(t)|\mathcal{G}(t)) \geq -(2 + 4c)X_1(t)\; dt. \tag{28}$$

And using (23)

$$\text{var}\; (dZ(t)|\mathcal{G}(t)) \leq \frac{\text{var}\; (Q(t)|\mathcal{G}(t))}{Q^2(t)} \;\leq 4X_1(t)\; dt. \tag{29}$$

Now by Lemma 19, to prove (25) it is enough to prove

$$\int_{-\infty}^{0} X_1(t)\ dt < \infty \text{ a.s.} \tag{30}$$

By (22), the size of cluster containing a specified atom satisfies

$$E(dY(t)|\mathcal{G}(t)) \geq (Y(t) - Y^2(t))\ dt \geq (1-a)Y(t)\ dt \text{ on } \{Y(t) \leq a\}$$

for fixed $0 < a < 1$. A moment's thought indicates that $X_1(t)$ must satisfy the same inequality. By nontriviality, $\lim_{t\to-\infty} X_1(t) = a'$ a.s. for some $0 \leq a' < 1$, Choose $a > a'$ and consider $T = \inf\{t : X_1(t) > a\} > -\infty$ a.s. Then

$$1 \geq EX_1(T) - a' = E\int_{-\infty}^{T} E(dX_1(t)|\mathcal{G}(t))\ dt \geq (1-a)E\int_{-\infty}^{T} X_1(t)\ dt$$

establishing (30).

To prove (26), the essential idea (see discussion later) is to show that the size $Y(t)$ of cluster containing any specified atom satisfies

$$e^{-t}Y(t) \to \eta_* \text{ a.s. as } t \to -\infty \tag{31}$$

for some $\eta_* \geq 0$. Write $Z(t) = e^{-t}Y(t)$. Then by (22)

$$E(dZ(t)|\mathcal{G}(t)) \leq e^{-t}(Y(t) + Q(t))\ dt - e^{-t}Y(t)\ dt \leq e^{-t}Q(t)\ dt$$

$$E(dZ(t)|\mathcal{G}(t)) \geq -2e^{-t}Y^2(t)\ dt.$$

And by (23)

$$\begin{aligned} \text{var}\ (dZ(t)|\mathcal{G}(t)) &= e^{-2t}\text{var}\ (dY(t)|\mathcal{G}(t)) \\ &\leq e^{-2t}(Y(t)Q(t) + S_3(t))\ dt \\ &\leq 2e^{-2t}Q(t)X_1(t)\ dt. \end{aligned}$$

By (25) we have $Q(t) = O(e^{2t})$ and $X_1(t) = O(e^t)$ a.s. as $t \to \infty$, so the bounds on $E(dZ(t)|\mathcal{G}(t))$ and var $(dZ(t)|\mathcal{G}(t))$ are $O(e^t)$, and Lemma 19 implies (31).

This isn't quite rigorous, because it's hard to make precise the idea of "selecting an atom at time $-\infty$". But we may rephrase as follows. Fix i_0. Write $(Y_i(t), t \geq t_0, 1 \leq i \leq i_0)$ for the post-t_0 evolution of the i_0 largest clusters at time t_0. By applying the argument above with a quantitative version of Lemma 19, then letting $t_0 \to -\infty$, we establish (26) for $i \leq i_0$. Details of this argument are written out in [6, §3.4], where exactly the same issue arises. Once (26) is established, the fact $\eta_1 \geq \eta_2 \geq \ldots \geq 0$ holds by ranking and the fact $\sum_i \eta_i^2 \leq \eta^2$ holds by Fatou's lemma.

5.3 Proof of Theorem 15

Suppose $(\boldsymbol{X}(t), -\infty < t < \infty)$ is a non-constant extreme additive coalescent. The limits in Proposition 21 are constants, so after replacing $\boldsymbol{X}(t)$ by $\boldsymbol{X}(t - t_0)$ for some t_0, as $t \to -\infty$

$$e^{-2t} \sum_i X_i^2(t) \quad \to \quad 1 \text{ a.s.} \tag{32}$$

$$e^{-t} X_i(t) \quad \to \quad \theta_i \text{ a.s., each } i \geq 1 \tag{33}$$

where $\boldsymbol{\theta} = (\theta_i) \in l_\downarrow^2$. We want to apply Lemma 14 and Proposition 13 to

$$\boldsymbol{p}_n = \boldsymbol{X}(-n), \ \sigma_n = \sqrt{\sum_i X_i^2(-n)}, \ \phi_n = e^{-n}/\sigma_n$$

for which we have $p_{ni}/\sigma_n \to \theta_i$ a.s., $\sigma_n \to 0$ a.s. and $\phi_n \to 1$ a.s. (Lemma 14 and Proposition 13 were stated for deterministic $\boldsymbol{p}_n$, but extend unchanged to the present random setting). By Proposition 1 the associated fragmentation processes $(\boldsymbol{Y}^{\boldsymbol{p}_n}(q), 0 \leq q \leq 1)$ have

$$\boldsymbol{Y}^{\boldsymbol{p}_n}(\phi_n \sigma_n) = \boldsymbol{Y}^{\boldsymbol{p}_n}(e^{-n}) \stackrel{d}{=} \boldsymbol{X}(0)$$

and so Lemma 14 implies $\boldsymbol{\theta} \in \Theta$. Now, as in the proof of Theorem 10, apply Proposition 13 to this sequence $(\boldsymbol{p}_n)$. Proposition 13 shows

$$(\boldsymbol{Y}^{\boldsymbol{p}_n}(\lambda \sigma_n \phi_n), \ 0 \leq \lambda \leq \tfrac{1}{\sigma_n \phi_n}) \stackrel{d}{\to} (\boldsymbol{Y}^{\boldsymbol{\theta}}(\lambda), \ 0 \leq \lambda < \infty)$$

in the sense of convergence of f.d.d.'s. The left side is $(\boldsymbol{Y}^{\boldsymbol{p}_n}(\lambda e^{-n}), 0 \leq \lambda \leq e^n)$, so setting $\lambda = e^{-t}$

$$(\boldsymbol{X}(t), -n \leq t < \infty) \stackrel{d}{\to} (\boldsymbol{X}^{\boldsymbol{\theta}}(t), -\infty < t < \infty).$$

We deduce that $\boldsymbol{X} \stackrel{d}{=} \boldsymbol{X}^{\boldsymbol{\theta}}$, for $\boldsymbol{\theta}$ defined by (33).

It remains to prove that each $\boldsymbol{X}^{\boldsymbol{\theta}}$ is extreme. If $\boldsymbol{X}^{\boldsymbol{\theta}}$ were not extreme, then it would have a decomposition as a mixture of extreme processes, that is $\boldsymbol{X}^{\boldsymbol{\theta}}(\cdot) \stackrel{d}{=} \boldsymbol{X}^{\boldsymbol{\theta}^*}(\cdot - t_0)$ for some random $(\boldsymbol{\theta}^*, t_0)$. But then we can apply Corollary 17 to both sides and the conclusion of that corollary implies $t_0 = 0$ and $\boldsymbol{\theta}^* = \boldsymbol{\theta}$ a.s.. In other words, $\boldsymbol{X}^{\boldsymbol{\theta}}$ is extreme.

6 Final remarks

6.1 Comparisons with the standard multiplicative coalescent

In many ways, the treatment of the standard additive coalescent in [7] and the entrance boundary in this paper parallel the treatment of the standard multiplicative coalescent in [4] and its entrance boundary in [6]. One difference is that the multiplicative coalescent takes values in l_2 rather than l_1; its total mass is infinite. Here we proved the additive coalescent entrance boundary was essentially (neglecting the distinction between $l^2_\downarrow$ and Θ) parametrized by $\mathbb{R} \times l^2_\downarrow$; in [6] it is shown that (with similar neglect) the multiplicative coalescent entrance boundary was essentially parametrized by $\mathbb{R} \times \mathbb{R}^+ \times l^3_\downarrow$. Our discrete construction (Proposition 1) of the additive coalescent is analogous to the "random graph" construction of the multiplicative coalescent. The broad outline of the proof of Theorem 15 (the weak convergence in Proposition 13 and the stochastic calculus in Proposition 21) is paralleled by Propositions 7 and 18 of [6]. Despite these parallels, we know of no argument which allows results for one process to be deduced from results for the other process. A final distinction is that there is no "multiplicative" analog of the central fact that the whole additive coalescent can be obtained by fragmenting the ICRT.

6.2 Other representations of the ICRT

Given an "excursion" function $f : [0,1] \to [0,\infty)$ satisfying certain conditions, one can define an associated continuum tree $\mathcal{S}_f$. Theorem 15 of [3] gives intrinsic conditions under which a CRT may be obtained as $\mathcal{S}_f$ for some random function f. This is a very useful way of looking at the $\boldsymbol{\theta} = \mathbf{0}$ case of $\mathcal{T}^{\boldsymbol{\theta}}$, in which case the random function is (up to a scaling constant) just standard Brownian excursion. The hypotheses of ([3] Theorem 15) allow only degree-3 branchpoints, but one could modify the result to allow more general branchpoints, and then show that for general $\boldsymbol{\theta} \in \Theta$ the ICRT $\mathcal{T}^{\boldsymbol{\theta}}$ can be represented by some random excursion-type function $f_{\boldsymbol{\theta}}$. In general there seems no simple description of $f_{\boldsymbol{\theta}}$, so we have not pursued the general case.

6.3 The additive coalescent with immigration

Consider the setting of Proposition 13, but take $\boldsymbol{\theta} = (1, 0, 0, 0, \ldots)$ (cf. remark below Proposition 5). In this case, instead of convergence $\boldsymbol{Y}^{\boldsymbol{p}_n}(q_n) \xrightarrow{d} \boldsymbol{Y}^{\boldsymbol{\theta}}(\lambda)$ on Δ we have only coordinatewise convergence, and the limit is the deterministic process $\boldsymbol{Y}(\lambda) = (e^{-\lambda}, 0, 0, \ldots)$ with corresponding $\boldsymbol{X}(t) = (\exp(-e^{-t}), 0, 0, \ldots)$. We have not pursued the details, but it seems that for general $\boldsymbol{\theta} \in l^2_\downarrow \setminus \Theta$ there is a process $\boldsymbol{X}^{\boldsymbol{\theta}}(t)$ whose total mass increases from 0 to 1 over $-\infty < t < \infty$. Informally, this process evolves as the additive coalescent, but instead of the unit total mass of infinitesimally small clusters being all present at time $-\infty$, the mass "immigrates" over time $(-\infty, \infty)$ according to some density function. See Pitman [20, §3.7] for an example of a similar phenomenon involving another coalescent process.

References

[1] D. Aldous and J. Pitman. A family of random trees with random edge lengths. Technical Report 526, Dept. Statistics, U.C. Berkeley, 1998. To appear in *Random Structures and Algorithms*. Available via http://www.stat.berkeley.edu/users/pitman.

[2] D.J. Aldous. The continuum random tree I. *Ann. Probab.*, 19:1–28, 1991.

[3] D.J. Aldous. The continuum random tree III. *Ann. Probab.*, 21:248–289, 1993.

[4] D.J. Aldous. Brownian excursions, critical random graphs and the multiplicative coalescent. *Ann. Probab.*, 25:812–854, 1997.

[5] D.J. Aldous. Deterministic and stochastic models for coalescence (aggregation and coagulation): a review of the mean-field theory for probabilists. *Bernoulli*, 5:3–48, 1999.

[6] D.J. Aldous and V. Limic. The entrance boundary of the multiplicative coalescent. *Electron. J. Probab.*, 3:1–59, 1998.

[7] D.J. Aldous and J. Pitman. The standard additive coalescent. *Ann. Probab.*, 26:1703–1726, 1998.

[8] J. Bertoin. A fragmentation process connected to Brownian motion. Technical Report 487, Laboratoire de Probabilités, Université Paris VI, 1999.

[9] M. Camarri and J. Pitman. Limit distributions and random trees derived from the birthday problem with unequal probabilities. Technical Report 523, Dept. Statistics, U.C. Berkeley, 1998. To appear in *Electronic J. Probab.*. Available via http://www.stat.berkeley.edu/users/pitman.

[10] P. Chassaing and G. Louchard. Phase transition for parking blocks, Brownian excursion and coalescence. Technical report, Université Nancy, 1999.

[11] R.L. Drake. A general mathematical survey of the coagulation equation. In G.M. Hidy and J.R. Brock, editors, *Topics in Current Aerosol Research (Part 2)*, volume 3 of *International Reviews in Aerosol Physics and Chemistry*, pages 201–376. Pergammon, 1972.

[12] E.B. Dynkin. Sufficient statistics and extreme points. *Ann. Probab.*, 6:705–730, 1978.

[13] S.N. Evans and J. Pitman. Construction of Markovian coalescents. *Ann. Inst. Henri Poincaré*, 34:339–383, 1998.

[14] A.M. Golovin. The solution of the coagulating equation for cloud droplets in a rising air current. *Izv. Geophys. Ser.*, 5:482–487, 1963.

[15] J. F. C. Kingman. The representation of partition structures. *J. London Math. Soc.*, 18:374–380, 1978.

[16] J.C. Mayer and L.G. Oversteegen. A topological characterization of **R**-trees. *Trans. Amer. Math. Soc.*, 320:395–415, 1990.

[17] S.B. Nadler. *Continuum Theory.* Dekker, New York, 1992.

[18] J. Pitman. Abel-Cayley-Hurwitz multinomial expansions associated with random mappings, forests and subsets. Technical Report 498, Dept. Statistics, U.C. Berkeley, 1997. Available via http://www.stat.berkeley.edu/users/pitman.

[19] J. Pitman. Coalescent random forests. *J. Comb. Theory A.*, 85:165–193, 1999.

[20] J. Pitman. The SDE solved by local times of a Brownian excursion or bridge derived from the height profile of a random tree or forest. *Ann. Probab.*, 27:261–283, 1999.

[21] R.K. Sheth and J. Pitman. Coagulation and branching process models of gravitational clustering. *Mon. Not. R. Astron. Soc.*, 289:66–80, 1997.

[22] A. Yao. On the average behavior of set merging algorithms. In *Proc. 8th ACM Symp. Theory of Computing*, pages 192–195, 1976.