

Scaling limits of coalescent processes near time zero

Batı Şengül

Centre of Mathematical Sciences, University of Cambridge, Wilberforce Road, Cambridge CB3 0WA, UK. E-mail: batisengul@gmail.com

Received 17 March 2014; revised 1 October 2015; accepted 5 November 2015

Abstract. In this paper we obtain scaling limits of Λ -coalescents near time zero under a regularly varying assumption. In particular this covers the case of Kingman's coalescent and beta coalescents. The limiting processes are coalescents with infinite mass, obtained geometrically as tangent cones of Evans metric space associated with the coalescent. In the case of Kingman's coalescent we are able to obtain a simple construction of the limiting space using a two-sided Brownian motion.

Résumé. Nous obtenons des limites d'échelle de Λ -coalescents en temps zéro sous une hypothèse de variation régulière. Cette hypothèse inclut notamment le coalescent de Kingman ainsi que la famille des Beta-coalescents. Les processus limites sont des processus de coalescence avec masse infinie, construits de manière géométrique comme cônes tangents de l'espace métrique de Evans. Dans le cas particulier du coalescent de Kingman une construction simple du processus limite est donnée à partir d'un mouvement brownien bidirectionnel.

MSC: 60J80; 60J99; 60F99

Keywords: Regularly varying coalescents; Small time asymptotics; Scaling limits; Random metric space; Tangent cones; Gromov–Hausdorff convergence

1. Introduction

1.1. Statement of the main results

A coalescent process is a particle system in which particles merge into blocks. Coalescent processes have found a variety of applications in physics, chemistry and most notably in genetics where the coalescent process models ancestral relationships as time runs backwards. The work on coalescent theory dates back to the seminal paper [23] where Kingman considered coalescent processes with pairwise mergers. In [24,26] and [14] this was extended to the case where multiple mergers are allowed to happen. We refer to [8] and [9] for an overview of the field.

In this paper we shall consider Λ -coalescents where Λ is a finite strongly regularly varying measure with index $1 < \alpha \leq 2$, see (2). These coalescents encompass a large variety of well known examples such as beta coalescents and Kingman's coalescent. Further, these coalescents have the property that they come down from infinity, that is, when starting with infinitely many particles, the process has finitely many blocks for any time $t > 0$.

Our goal is to gain precise information about the behaviour of the coalescent near time zero under the regularly varying assumption. Prior work on this appears in [7] where the authors establish many results about the behaviour of beta coalescents as $t \downarrow 0$; they show the asymptotic behaviour of the number of blocks and further show results about the largest block and the block containing 1 at time t . These results are generalisations of the results for the case of Kingman's coalescent (see [3]) and concern the behaviour of a single block at a single time. Our aim here is to describe the behaviour of many blocks spanning over a small interval of time.

One central insight of this work is that the correct framework for taking such scaling limits is to view coalescent processes as geometric objects. What follows is an outline of our approach. To any coalescent process $\Pi = (\Pi(t) : t \geq$

0), one can associate a certain ultra-metric space (E, δ) which completely characterises the process Π . This was first suggested in the work of Evans in [16], who introduced this object in the case of Kingman's coalescent and studied some of its properties.

The construction is simple and we describe it now. Let $\Pi = (\Pi(t) : t \geq 0)$ be a coalescent process and define an ultra-metric on \mathbb{N} by

$$\delta(i, j) := \inf\{t > 0 : i \stackrel{\Pi(t)}{\sim} j\}, \quad (1)$$

where $i \stackrel{\Pi(t)}{\sim} j$ if and only if i, j are in the same block of $\Pi(t)$. The metric space (E, δ) is then the completion of (\mathbb{N}, δ) .

Notice that $\delta(i, j)$ gives the time for the most recent common ancestor of i and j . Moreover it is not hard to check that (E, δ) is compact if and only if the coalescent process Π comes down from infinity, that is for all $t > 0$, $\Pi(t)$ has finitely many (non-empty) blocks. We call the space (E, δ) the Evans space associated to the coalescent Π .

Let Λ be a finite measure on $[0, 1]$. We say that Λ is SRV(α) if Λ is strongly regularly varying with index $\alpha \in (1, 2)$. That is when $\Lambda(dp) = f(p) dp$ and there exists a constant $A_\Lambda > 0$ such that

$$f(p) \sim A_\Lambda p^{1-\alpha}, \quad p \rightarrow 0, \quad (2)$$

where the above notation means the quotient of both sides approaches 1. We will abuse notation slightly and say that Λ is SRV(2) when $\Lambda = \delta_{\{0\}}$. It is possible to associate with each finite Λ on $[0, 1]$ a coalescent process called the Λ -coalescent and the case when Λ is SRV(2), the Λ -coalescent is Kingman's coalescent. Note that if Λ is a finite SRV(α), then the Λ -coalescent comes down from infinity if and only if $\alpha \in (1, 2]$.

Finite SRV(α) measures encompass a large variety of measures. A prominent example is the Beta($2 - \alpha, \alpha$) distribution which has density $B(2 - \alpha, \alpha)^{-1} p^{1-\alpha} (1 - p)^{\alpha-1} dp$, where $B(x, y)$ is the beta function. This is a one parameter family which interpolates between the uniform measure ($\alpha = 1$) and δ_0 ($\alpha \rightarrow 2$) for which the corresponding coalescents are the Bolthausen–Sznitman coalescent and Kingman's coalescent respectively. The importance of the SRV(α) condition stems primarily from population genetics where the models correspond to populations in which there is large variability in the offspring distribution, see [8, Section 3.2].

The first theorem of the paper presented below shows convergence of the metric spaces that correspond to the coalescent processes as in (1). Let us briefly discuss the pointed Gromov–Hausdorff topology which we use as our notion of convergence (see Section 2 for the details). A pointed metric space (S, d, p) is called proper if every closed ball is compact and Polish if it is complete and separable. A sequence of proper Polish pointed metric spaces (S_n, d_n, p_n) converges to a proper Polish pointed metric space (S, d, p) under the pointed Gromov–Hausdorff topology if for every $r > 0$, the closed ball of radius r around $p_n \in S_n$ converges in the usual Gromov–Hausdorff sense to the closed ball of radius r around $p \in S$. The space of all proper Polish pointed metric spaces can be equipped with a metric, called the pointed Gromov–Hausdorff metric, which is compatible with the notion of convergence described and further this space is itself a Polish space when equipped with the pointed Gromov–Hausdorff metric.

Theorem 1.1. *Let Λ be a finite measure satisfying (2) for some $\alpha \in (1, 2]$ and (E, δ) be the Evans space associated to the corresponding Λ coalescent via (1). Then for all $i \in \mathbb{N}$, there exists a random pointed ultra-metric space $(\mathbb{S}, d_{\mathbb{S}}, o)$, which is independent of i , such that*

$$(E, \varepsilon^{-1} \delta(\cdot, \cdot), i) \rightarrow (\mathbb{S}, d_{\mathbb{S}}, o)$$

weakly under the pointed Gromov–Hausdorff metric as $\varepsilon \rightarrow 0$.

The limiting spaces in the theorem depend on the value of $\alpha \in (1, 2]$ and A_Λ . We will denote them by $(\mathbb{S}^{(\alpha)}, d_{\mathbb{S}}^{(\alpha)}, o^{(\alpha)})$ if there is a risk of confusion.

Geometrically, the space $(\mathbb{S}, d_{\mathbb{S}}, o)$ in Theorem 1.1 is referred to as tangent cone of (E, δ) at the point i . More precisely a tangent cone of a metric space (X, d) at a point $x \in X$ is given by the pointed Gromov–Hausdorff limit of $(X, r_i^{-1} d, x)$ as $i \rightarrow \infty$ where $\{r_i\}_{i \geq 1}$ is some sequence such that $r_i \downarrow 0$. Tangent cones are generalisations of tangent spaces on manifolds. Indeed, on a Riemannian manifold the tangent cone at any point exists and is isometric to the

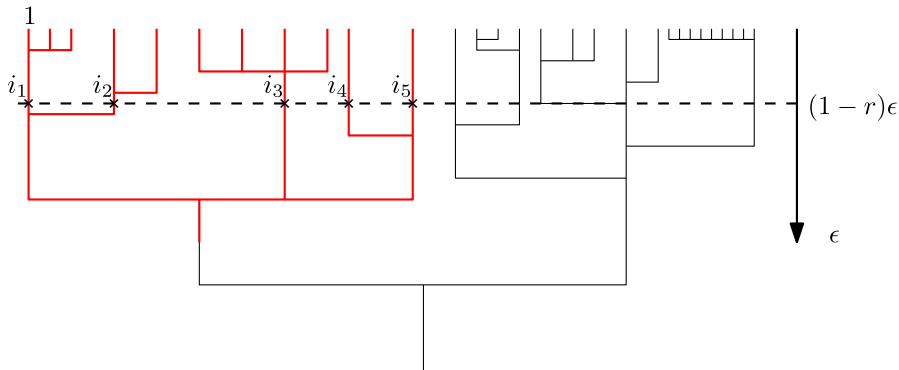


Fig. 1. Picture illustrating the process \mathcal{Z}_ε . The highlighted sub-tree on the left represents the blocks of the coalescent process which eventually merge to form $\Pi_1(\varepsilon)$. Here $\mathcal{Z}_\varepsilon(r) = 5$.

tangent space. Tangent cones have appeared in a variety of contexts ranging from geometric measure theory [27] to a recent paper [13] in which the tangent cones of the Brownian map are identified as the Brownian plane. In our case, tangent cones are the correct objects for describing the scaling limits as they allow us to forget about the mass and ordering imposed on the coalescent.

In [22] the author identifies a homeomorphism between the space of ultra-metric spaces and the space of real trees both equipped with the Gromov–Hausdorff metric. Consequently Theorem 1.1 can be stated in terms of the real trees that correspond to the coalescents. The tangent cones are only of interest at the leaves of a real tree as they can be easily identified at any other point as follows. If (T, d) is a coalescent tree and $x \in T$ such that $T \setminus \{x\}$ has exactly two connected components then the tangent cone $\lim_{r \downarrow 0} (T, r^{-1}d, x)$ exists and is isometric to \mathbb{R} with the Euclidean distance. If $T \setminus \{x\}$ has $k \geq 3$ components then the tangent cone around x exists and is isometric to k disjoint copies of $[0, \infty)$ glued together at the point 0, equipped with the intrinsic metric.

The next result (which is both a crucial step in the proof of Theorem 1.1, and of independent interest) provides a description of the mergers of the block containing 1 at small times. This description will allow us to depict the space $(\mathbb{S}, d_{\mathbb{S}})$. Loosely speaking, this result should be interpreted as a local limit of the coalescent tree, whereas Theorem 1.1 deals with global scaling limits. More precisely for $\varepsilon > 0$ and $r \in [0, 1)$ let $\mathcal{Z}_\varepsilon(r)$ be the number of blocks of $\Pi((1-r)\varepsilon)$ that make up $\Pi_1(\varepsilon)$, the block containing 1 at time ε (see also Figure 1). Thus there exists $1 = i_1 < \dots < i_{\mathcal{Z}_\varepsilon(r)}$ such that

$$\Pi_1(\varepsilon) = \Pi_{i_1}((1-r)\varepsilon) \cup \dots \cup \Pi_{i_{\mathcal{Z}_\varepsilon(r)}}((1-r)\varepsilon),$$

where $\Pi_i(t)$ denotes the i th block of $\Pi(t)$ under the standard infimum ordering. Henceforth we shall be considering the càdlàg modification of the process \mathcal{Z}_ε .

Theorem 1.2. *For $\alpha \in (1, 2]$ and Λ a finite SRV(α) measure let \mathcal{Z}_ε be the process constructed above using a Λ -coalescent. Then as $\varepsilon \rightarrow 0$, $\mathcal{Z}_\varepsilon \rightarrow \mathcal{Z}$ in the Skorokhod sense on $[0, 1)$. The process \mathcal{Z} is an inhomogeneous Markov process with generator*

$$L_r f(i) = A_\Lambda \sum_{j \geq 1} (i+j) \frac{\Gamma(2-\alpha)\Gamma(j-\alpha+1)}{(1-r)\alpha\Gamma(j+2)} [f(i+j) - f(i)]$$

when $\alpha \in (1, 2)$ and

$$L_r f(i) = \frac{(i+1)}{1-r} [f(i+1) - f(i)]$$

when $\alpha = 2$.

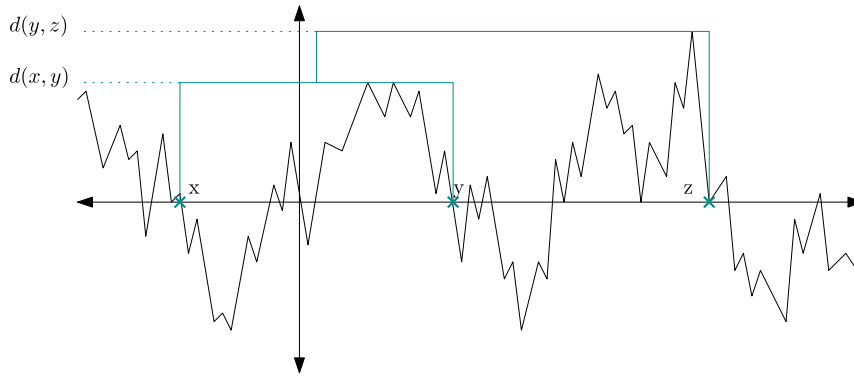


Fig. 2. Construction of the space $(\mathbb{S}^{(2)}, d_{\mathbb{S}}^{(2)}, o^{(2)})$ from a two-sided Brownian motion.

We will now depict how the closed unit ball $B(o, 1) \subset (\mathbb{S}, d_{\mathbb{S}})$ is constructed (see Section 3.2 for a proof of the assertions here). First construct a tree T from a branching process. Start the tree with one particle which does not die and is hence referred to as an immortal particle. The immortal particle produces $j \geq 1$ offspring at time $r \in [0, 1)$ with rate $(j+1)q_{j+1}^{(r)}$ where

$$q_{j+1}^{(r)} = \begin{cases} \frac{\Gamma(2-\alpha)\Gamma(j-\alpha+1)}{(1-r)^{\alpha}\Gamma(j+2)} & \text{if } \alpha \in (1, 2), \\ \mathbb{1}_{\{j=1\}} \frac{1}{1-r} & \text{if } \alpha = 2. \end{cases}$$

For $j \geq 1$, the other particles at time $r \in [0, 1)$ die and are replaced by $j+1$ offspring at rate $q_{j+1}^{(r)}$. Thus for each $r \in [0, 1)$, the number of particles which are at distance r from the root is distributed $\mathcal{Z}(r)$. The process $\mathcal{Z}(r)$ explodes as $r \rightarrow 1$ so we have infinitely many particles at distance one from the root. The space $B(o, 1)$ is the set of particles at distance one from the root and o is the immortal particle that is distance one from the root. For each $v, w \in B(o, 1)$ there exists two unique paths from the root ending at the points v, w and these paths deviate at distance $h_{v,w} \geq 0$ from the root. The distance between two points is given by $d_{\mathbb{S}}(v, w) = 1 - h_{v,w}$. The multi-type branching process described is the spine decomposition of an inhomogeneous Galton–Watson process for which each particle at time $r \in [0, 1)$ dies gives rise to $j+1$ offspring at rate $q_{j+1}^{(r)}$ as introduced in [12].

In the case $\alpha = 2$ we are able to strengthen the convergence in Theorem 1.1 to that of metric measure spaces and explicitly construct the limiting space $(\mathbb{S}, d_{\mathbb{S}}, o) = (\mathbb{S}^{(2)}, d_{\mathbb{S}}^{(2)}, o^{(2)})$ (see Figure 2). To that end construct a measure ν on the space (E, δ) as follows. Let ν be such that the mass it assigns to each closed ball $B(i, t)$ of radius $t > 0$ around i is equal to the asymptotic frequency of the block of $\Pi(t)$ containing i . This extends uniquely to a measure on the whole space by Carathéodory's extension theorem. Our next result shows the tangent cones of the metric space (E, δ) equipped with the measure ν .

Theorem 1.3. *In the case when $\alpha = 2$ in Theorem 1.1, there exists a locally finite measure μ on the space $(\mathbb{S}, d_{\mathbb{S}})$ such that for all $i \in \mathbb{N}$,*

$$(E, \varepsilon^{-1}\delta(\cdot, \cdot), 4\varepsilon^{-1}\nu, i) \rightarrow (\mathbb{S}, d_{\mathbb{S}}, \mu, o)$$

weakly as $\varepsilon \rightarrow 0$ under the Gromov–Hausdorff–Prokhorov topology.

The limiting metric measure space $(\mathbb{S}, d_{\mathbb{S}}, \mu, o)$ is independent from i and can be constructed as follows. Let $W = (W(t) : t \in \mathbb{R})$ be a two-sided Brownian motion on \mathbb{R} and let $\mathcal{N} := \{t \in \mathbb{R} : W(t) = 0\}$. For each $x, y \in \mathcal{N}$ with $x \leq y$, define the pseudo-metric

$$d_{\mathbb{S}}(x, y) := \sup\{W(t) : t \in [x, y]\} \vee 0 \quad (3)$$

and $\mathbb{S} = \mathcal{N} / \sim$ where $x \sim y$ if and only if $d_{\mathbb{S}}(x, y) = 0$ and $o = 0$. The measure μ is the projection of the local time measure on \mathcal{N} .

We delay the exact definition of the Gromov–Hausdorff–Prokhorov metric to Section 2.

Note that Theorem 1.2 in the case $\alpha = 2$ can be obtained from Theorem 1.3 through some routine computations. The factor of 4 appearing on the measure in Theorem 1.3 is for the following reason. For small $\varepsilon > 0$ and $t \leq \varepsilon$, the number of blocks of a Kingman coalescent at time $t\varepsilon$ is roughly $2/(t\varepsilon)$. Hence the typical mass of a ball of radius t in $(E, \varepsilon^{-1}\delta)$ is roughly $t/2$. On the other hand the typical ball of radius t under $(\mathbb{S}, d_{\mathbb{S}})$ is the same as the total local-time at level t obtained by a Brownian excursion conditioned to reach level t . This has an exponential distribution with expectation $2t$. Thus the two measures are off by roughly a factor of 4.

To illustrate the usefulness of the results in the case $\alpha = 2$ we present the following corollary. This is an immediate consequence of Theorem 1.3.

Corollary 1.4. *Let $F(t)$ be the asymptotic frequency of the block containing 1 in Kingman’s coalescent at time $t \geq 0$. Then we have in the sense of weak convergence on the Skorokhod space*

$$(\varepsilon^{-1}F(\varepsilon t) : t \geq 0) \rightarrow (X(t) : t \geq 0)$$

as $\varepsilon \rightarrow 0$.

The process $X = (X(t) : t \geq 0)$ is characterised by the following.

- (i) $X(0) = 0$ and for $t > 0$, $X(t)$ is the sum of two i.i.d. exponential distributions with parameter $1/(2t)$
- (ii) X is an inhomogeneous compound Poisson process where at time $t > 0$ the rate of jumps is given by $2/t$ and the jump distribution is exponential with parameter $1/(2t)$.

Note that Corollary 1.4 extends [4, Corollary 1.3] which shows the above convergence for fixed $t \geq 0$.

1.2. Directions for further research

It would be interesting to see if any of the results presented thus far can be extended to general coalescents. It seems that not every Λ -coalescent which comes down from infinity will admit a scaling limit (see for example the Λ -coalescent constructed in [6, Section 5]). For coalescents which do not come down from infinity we have that the closed balls are not compact. Taking scaling limits of these coalescents would require a new framework of metric space convergence.

We note that similar objects appear in the limit when studying the behaviour of fragmentation process with large mass, near the explosion time, see [19–21]. It would also be interesting to see if there is a relation between coalescents near time zero and fragmentation processes with large mass.

A simpler question is if Theorem 1.1 can be extended to use metric-measure space convergence when $\alpha \in (1, 2)$? This was done in Theorem 1.3 for Kingman’s coalescent but the proof of Theorem 1.3 relies on a very specific construction of Kingman’s coalescent.

1.3. Outline of the paper

In Section 2 we introduce some background on metric geometry. We assume the reader is familiar with the basic concepts in coalescent theory and excursion theory. We refer to [8,9] and [25]. In Section 3.1 we prove Theorem 1.2 for $\alpha \in (1, 2)$ and explain the changes needed for the case $\alpha = 2$. Then in Section 3.2 we prove Theorem 1.1. In Section 4 we shall prove Theorem 1.3.

2. Convergence of metric spaces

In this section we briefly review some basic notions of convergence of metric spaces. For a detailed treatment of the material refer to [11].

Here we introduce the Gromov–Hausdorff metric used in Theorem 1.1 and the Gromov–Hausdorff–Prokhorov metric in Theorem 1.3. We start by defining a metric on certain metric spaces without measures, called the Gromov–Hausdorff metric. We introduce this by first defining the Gromov–Hausdorff metric on compact metric spaces. Consider two compact metric spaces (X, d_X) and (Y, d_Y) . The compact Gromov–Hausdorff distance

$d_{\text{cGH}}((X, d_X), (Y, d_Y))$ is constructed as follows. Let (Z, d_Z) be a metric space such that there exists isometric embeddings $\phi_X : (X, d_X) \rightarrow (Z, d_Z)$, $\phi_Y : (Y, d_Y) \rightarrow (Z, d_Z)$, then

$$d_{\text{cGH}}((X, d_X), (Y, d_Y)) := \inf \{d_H^Z(\phi_X(X), \phi_Y(Y))\}, \quad (4)$$

where the infimum is over all metric spaces (Z, d_Z) with the above property and

$$d_H^Z(A, B) := \inf \{\varepsilon > 0 : B \subset \{z \in Z : \text{dist}_{d_Z}(z, A) < \varepsilon\} \text{ and } A \subset \{z \in Z : \text{dist}_{d_Z}(z, B) < \varepsilon\}\}$$

is the Hausdorff distance in (Z, d_Z) . In particular, $d_{\text{cGH}}((X, d_X), (Y, d_Y)) = 0$ if and only if (X, d_X) and (Y, d_Y) are isometric.

Now we define the Gromov–Hausdorff–Prokhorov metric on compact spaces. Suppose in addition we have two finite measures μ and ν defined on the spaces (X, d_X) and (Y, d_Y) respectively. Again let (Z, d_Z) be a metric space such that there exist isometric embeddings $\phi_X : (X, d_X) \rightarrow (Z, d_Z)$, $\phi_Y : (Y, d_Y) \rightarrow (Z, d_Z)$. Then $\mu^* = \mu \circ \phi_X^{-1}$ and $\nu^* = \nu \circ \phi_Y^{-1}$ are measures on the space (Z, d_Z) . The Prokhorov metric on (Z, d_Z) is given by

$$d_{\text{Pr}}^Z(\mu^*, \nu^*) := \inf \{\varepsilon > 0 : \mu^*(A) \leq \nu^*(A^\varepsilon) + \varepsilon \text{ and } \nu^*(A) \leq \mu^*(A^\varepsilon) + \varepsilon, \forall \text{ measurable } A\},$$

where $A^\varepsilon := \{z \in Z : \text{dist}_{d_Z}(z, A) < \varepsilon\}$. Then the compact Gromov–Hausdorff–Prokhorov distance is given by

$$d_{\text{GHP}}((X, d_X, \mu), (Y, d_Y, \nu)) = \inf \{d_H^Z(\phi_X(X), \phi_Y(Y)) + d_{\text{Pr}}^Z(\mu^*, \nu^*)\},$$

where the infimum is over all metric spaces (Z, d_Z) with the above property.

In this paper we work with non-compact spaces and there are several ways to extend the definition above to a certain class of non-compact metric spaces. We now introduce our notion of the Gromov–Hausdorff and Gromov–Hausdorff–Prokhorov distance on non-compact metric spaces satisfying certain properties. Suppose that (X, d_X, μ, p_X) and (Y, d_Y, ν, p_Y) are proper Polish pointed metric spaces, that is they are complete, separable and every closed ball is compact. Then the (pointed) Gromov–Hausdorff distance is given by

$$d_{\text{GH}}((X, d_X, p_X), (Y, d_Y, p_Y)) = \sum_{n \geq 1} 2^{-n} (1 \wedge d_{\text{cGH}}((B(p_X, n), d_X), (B(p_Y, n), d_Y))),$$

where here and throughout $B(p, r)$ denotes the closed ball of radius r around p .

Denote by $(\mathcal{X}, d_{\text{GH}})$ the space of proper Polish spaces with a distinguished point, up to isometry, equipped with the Gromov–Hausdorff metric. The space $(\mathcal{X}, d_{\text{GH}})$ is a Polish space (see [17]). Some of the properties of metric spaces are preserved under d_{GH} convergence. One such example is when the metric is an ultra-metric. A metric d is called an ultra-metric if

$$d(x, y) \leq d(x, z) \vee d(y, z) \quad \forall x, y, z.$$

It is not hard to check all the metric spaces in this paper are in fact ultra-metric spaces and that this property is preserved under d_{GH} convergence.

Suppose that (X, d_X, μ, p_X) and (Y, d_Y, ν, p_Y) are proper Polish pointed metric spaces which come equipped with two measures μ and ν respectively. Suppose further that both measures are finite on compact sets. Then the (pointed) Gromov–Hausdorff–Prokhorov distance is given by

$$d_{\text{GHP}}((X, d_X, \mu, p_X), (Y, d_Y, \nu, p_Y)) = \sum_{n \geq 1} 2^{-n} (1 \wedge d_{\text{cGHP}}((B(p_X, n), d_X, \mu), (B(p_Y, n), d_Y, \nu))).$$

Denote by $(\mathcal{X}_\mu, d_{\text{GHP}})$ the space of proper Polish spaces with a distinguished point, up to measure preserving isometries, equipped with the Gromov–Hausdorff–Prokhorov metric. Then space $(\mathcal{X}_\mu, d_{\text{GHP}})$ is a Polish space (see for example [1]).

It is not hard to check that the space (E, δ) is compact and the measure ν is finite. Further it can be seen that the limiting space $(\mathbb{S}, d_{\mathbb{S}})$ for $\alpha = 2$ in Theorem 1.3 is a proper Polish space and the measure μ is finite on compact sets.

3. SRV(α) case

3.1. Proof of Theorem 1.2

Throughout this proof we omit the superscript α from the notation and assume that $\alpha \in (1, 2)$. The proof for the case $\alpha = 2$ follows analogously and we remark the only alteration to the proof that is required for $\alpha = 2$.

Let $\Pi = (\Pi(t) : t \geq 0)$ denote a Λ -coalescent such that $\Lambda(dp) = f(p)dp$ with $f(p) \sim A_\Lambda p^{1-\alpha}$ as $p \rightarrow 0$. Let $N = (N(t) : t > 0)$ denote the number of blocks of process Π . Recall that $Z_\varepsilon(r)$ is the number of blocks at time $(1-r)\varepsilon$ that make up the block containing 1 at time ε . We will show the convergence result by showing that $Z_\varepsilon(r)$ almost satisfies a certain martingale problem for small $\varepsilon > 0$.

We will simplify the notation further by writing $\Pi^\varepsilon(r) := \Pi((1-r)\varepsilon)$ and $N^\varepsilon(r) := N((1-r)\varepsilon)$ for $r \in [0, 1)$. We will denote by $\mathcal{F}_r^\varepsilon$ the natural filtration of $Z_\varepsilon(r)$. Before we begin the proof let us describe an outline.

Outline of the proof

Our strategy in proving Theorem 1.2, the convergence $Z_\varepsilon \rightarrow Z$ as $\varepsilon \rightarrow 0$, is separated into two steps: (1) showing that $\{Z_\varepsilon\}_{\varepsilon>0}$ is tight (Lemma 3.6) and (2) showing that every subsequence of Z_ε which converges satisfies the martingale problem for the generator L_r of the limit Z , which given in Theorem 1.2 (Lemma 3.5). The latter statement turns out to be somewhat harder to show.

The natural approach to show (2) would be to identify $\lim_{\delta \downarrow 0} \delta^{-1} \mathbb{P}(Z_\varepsilon(r+\delta) - Z_\varepsilon(r) = j | \mathcal{F}_r^\varepsilon)$ as this corresponds to the Q -matrix of the process. We could then take the limit as $\varepsilon \rightarrow 0$ and conclude the proof by using results from [15]. Unfortunately we are unable to show the existence of $\lim_{\delta \downarrow 0} \delta^{-1} \mathbb{P}(Z_\varepsilon(r+\delta) - Z_\varepsilon(r) = j | \mathcal{F}_r^\varepsilon)$, so instead we rely on estimating $\lim_{\delta \downarrow 0} \delta^{-1} \mathbb{P}(Z_\varepsilon(r+\delta) - Z_\varepsilon(r) = j; A)$ uniformly over $A \in \mathcal{F}_r^\varepsilon$ and r inside a compact interval. For fixed $\varepsilon > 0$ our upper and lower bound will differ however they become arbitrarily close to each other as $\varepsilon \downarrow 0$. This will be sufficient to show that the limit Z satisfies the martingale problem.

Our strategy for obtaining bounds on $\delta^{-1} \mathbb{P}(Z_\varepsilon(r+\delta) - Z_\varepsilon(r) = j; A)$ is by using Bayes' formula to write this probability in terms of events that move forwards in time for the coalescent process. We do this as follows. The event $\{Z_\varepsilon(r+\delta) - Z_\varepsilon(r) = j\}$ means that there was a coalescent event during $((1-r-\delta)\varepsilon, (1-r)\varepsilon)$ merging $j+1$ blocks together, after which the resulting block as well as $Z_\varepsilon(r) - 1$ other blocks merge during the interval $((1-r)\varepsilon, \varepsilon)$ to form the block containing 1 at time ε . Calculating the probability of the first merger is simple. Next, we apply the Markov property after the first merger (at time $(1-r)\varepsilon$) and ask for the conditional probability the event A holds as well as the resulting block together with $Z_\varepsilon(r) - 1$ other blocks merging during the interval $((1-r)\varepsilon, \varepsilon)$ to form the block containing 1 at time ε . It turns out that the unconditional probability of this event is much easier to calculate. Thus firstly in Lemma 3.1 we estimate the difference between the conditional and the unconditional probability of such events. Then in Lemma 3.2 we use the strategy described here to derive bounds on $\lim_{\delta \downarrow 0} \delta^{-1} \mathbb{P}(Z_\varepsilon(r+\delta) - Z_\varepsilon(r) = j; A)$. Finally in Lemma 3.3 we show the convergence of the bounds obtained in Lemma 3.2 as $\varepsilon \downarrow 0$.

Recall that the event $\{Z_\varepsilon(r) = \ell\}$ means that the block containing 1 at time ε is made up of exactly ℓ blocks at time $(1-r)\varepsilon$. It will be very convenient to fix the indices of these blocks. For $\ell \leq n$, define

$$\mathcal{S}_{\ell,n} := \{\mathbf{z} = (z_j)_{j=1}^\ell : 1 = z_1 < \dots < z_\ell \leq n\}$$

which will represent the set indices of the blocks at time $(1-r)\varepsilon$ which eventually merge and form the block containing 1 at time ε . For $\mathbf{z} \in \mathcal{S}_{\ell,n}$ and $r \in [0, 1)$ consider the event

$$\kappa(r, \mathbf{z}) := \left\{ \Pi_1^\varepsilon(0) = \bigcup_{i \in \mathbf{z}} \Pi_i^\varepsilon(r) \right\} \cap \{ \Pi_i^\varepsilon(r) \neq \emptyset, \forall i \in \mathbf{z} \}, \quad (5)$$

where $\Pi_i^\varepsilon(r)$ denotes the i th block of $\Pi^\varepsilon(r)$ where the blocks are ordered by infimum. In words, this is the event that the block containing 1 at time ε is made up of blocks with labels given by \mathbf{z} at time $(1-r)\varepsilon$. For example if $\Pi^\varepsilon(0) = \{\{1, 3, 5\}, \{2, 4\}\}$ and $\Pi^\varepsilon(r) = \{\{1, 3\}, \{2\}, \{4\}, \{5\}\}$, then the event $\kappa(r, \{1, 4\})$ would hold since

$\Pi_1^\varepsilon(0) = \{1, 3, 5\} = \{1, 3\} \cup \{5\} = \Pi_1^\varepsilon(r) \cup \Pi_4^\varepsilon(r)$. Thus for fixed $r \in [0, 1)$, $\kappa(r, \mathbf{z}) \cap \kappa(r, \mathbf{z}') = \emptyset$ for $\mathbf{z} \neq \mathbf{z}'$, further

$$\{\mathcal{Z}_\varepsilon(r) = \ell\} \cap \{N^\varepsilon(r) = n\} = \bigcup_{\mathbf{z} \in \mathcal{S}_{\ell,n}} \kappa(r, \mathbf{z}) \cap \{N^\varepsilon(r) = n\}. \quad (6)$$

For the next lemma let R_n be the map which maps a partition on \mathbb{N} to a partition on $[n]$ by projection. The next lemma estimates the effects of a merger of $j + 1$ blocks during $((1 - r - \delta)\varepsilon, (1 - r)\varepsilon)$ on the probability of the event A .

Lemma 3.1. *For any $\varepsilon > 0$, $r \in [0, 1)$, $A \in \mathcal{F}_r^\varepsilon$ and $j < n$, it holds that*

$$|\mathbb{P}(A|N^\varepsilon(r) = n) - \mathbb{P}(A|N^\varepsilon(r) = n - j)| \leq j(1 - e^{-\varepsilon}). \quad (7)$$

Let us give heuristics about the proof before we begin. The event A asks questions concerning the particles which merge with block containing 1 during the time period $[(1 - r)\varepsilon, \varepsilon]$. Now we have two configurations for $N^\varepsilon(r) = N((1 - r)\varepsilon)$, one of which has j extra particles. On the event that these extra j particles do not interact with the block containing 1 during the time interval $[(1 - r)\varepsilon, \varepsilon]$, the event A has the same probability in each of these two configurations. The probability on the right hand side of (7) is an estimate for the event that the extra j particles do interact with the block containing 1 during the time interval $[(1 - r)\varepsilon, \varepsilon]$.

Proof of Lemma 3.1. Fix $r \in [0, 1)$, $n \in \mathbb{N}$ and $j < n$ throughout. In order to make the sketch above rigorous we will use a monotone class argument. Let \mathcal{M} be the set of $A \in \mathcal{F}_r^\varepsilon$ such that (7) holds. Suppose that $A_1, A_2, \dots \in \mathcal{M}$ such that $A_1 \subset A_2 \subset \dots$ and let $A = \bigcup_{i \geq 1} A_i$. Then

$$|\mathbb{P}(A|N^\varepsilon(r) = n) - \mathbb{P}(A|N^\varepsilon(r) = n - j)| \leq \limsup_{i \rightarrow \infty} |\mathbb{P}(A_i|N^\varepsilon(r) = n) - \mathbb{P}(A_i|N^\varepsilon(r) = n - j)| \leq j(1 - e^{-\varepsilon})$$

and hence \mathcal{M} is a monotone class. All that remains to be shown is that \mathcal{M} contains a π -system which generates the σ -algebra $\mathcal{F}_r^\varepsilon$ since once this is shown the result follows from monotone class theorem.

Condition on $N^\varepsilon(r) = n$, so that $\Pi^\varepsilon(r) = (\Pi_1^\varepsilon(r), \dots, \Pi_n^\varepsilon(r))$. Let \mathcal{A} denote the set of events of the form

$$\bigcup_{\mathbf{z} \in \mathcal{S}_{\ell,n}} \kappa(u, \mathbf{z}), \quad u \leq r, \ell \leq n.$$

Let \mathcal{A}' denote the π -system generated by \mathcal{A} . We will show that $\mathcal{A}' \subset \mathcal{M}$ by showing that every $A \in \mathcal{A}'$ satisfies (7).

Note that conditioning on $N^\varepsilon(r) = n$ implies that for each $\ell \leq n$ and $u \leq r$, $\{\mathcal{Z}_\varepsilon(u) = \ell\} = \bigcup_{\mathbf{z} \in \mathcal{S}_{\ell,n}} \kappa(u, \mathbf{z})$. Hence \mathcal{A}' generates the σ -algebra $\mathcal{F}_r^\varepsilon$. Henceforth let $A \in \mathcal{A}'$ be fixed and denote by $A_1, \dots, A_m \in \mathcal{A}$ the elements such that $A = A_1 \cap \dots \cap A_m$.

For any $u \leq r$, $\ell \leq n$, applying the Markov property we have

$$\begin{aligned} & \mathbb{P}\left(\bigcup_{\mathbf{z} \in \mathcal{S}_{\ell,n}} \kappa(u, \mathbf{z}) \mid N^\varepsilon(r) = n\right) \\ &= \mathbb{P}\left(\bigcup_{\mathbf{z} \in \mathcal{S}_{\ell,n}} \{R_n \Pi_1(r\varepsilon) = R_n \Pi_{z_1}((r - u)\varepsilon) \cup \dots \cup R_n \Pi_{z_\ell}((r - u)\varepsilon)\}\right), \end{aligned} \quad (8)$$

where $R_n \Pi_i(t)$ is the i th block of $R_n \Pi$ at time $t > 0$. For each A_i , $i = 1, \dots, m$, we denote by \tilde{A}_i^n the event on the right hand side of (8) so that $\tilde{A}_i^n \in \sigma(R_n \Pi(s) : s \leq r\varepsilon)$ and

$$\mathbb{P}(A_i|N^\varepsilon(r) = n) = \mathbb{P}(\tilde{A}_i^n), \quad i \leq m. \quad (9)$$

We will show that $\mathbb{P}(\bigcap_{i \leq m} \tilde{A}_i^n \Delta \bigcap_{i \leq m} \tilde{A}_i^{n-j}) \leq j(1 - e^{-\varepsilon})$, which will show (7) for all $A \in \mathcal{A}'$.

For any $u \leq r$, $\ell \leq n$ and $\mathbf{z} \in \mathcal{S}_{\ell,n}$,

$$\begin{aligned} \{R_n \Pi_1(r\varepsilon) = R_n \Pi_{z_1}((r-u)\varepsilon) \cup \dots \cup R_n \Pi_{z_\ell}((r-u)\varepsilon)\} \cap \bigcap_{k=n-j+1}^n \{k \notin R_n \Pi_1(r\varepsilon)\} \\ = \{R_{n-j} \Pi_1(r\varepsilon) = R_{n-j} \Pi_{z_1}((r-u)\varepsilon) \cup \dots \cup R_{n-j} \Pi_{z_\ell}((r-u)\varepsilon)\} \cap \bigcap_{k=n-j+1}^n \{k \notin R_n \Pi_1(r\varepsilon)\}. \end{aligned} \quad (10)$$

Indeed, consider the blocks at time $(r-u)\varepsilon$ that coalesce by time $r\varepsilon$ form the block containing 1. The only way these can differ between the restrictions to $\{1, \dots, n-j\}$ and to $\{1, \dots, n\}$ is if at least one of $\{n-j+1, \dots, n\}$ coalesced with 1 by time $r\varepsilon$.

Recall that for each $i \in m$, \tilde{A}_i^n and \tilde{A}_i^{n-j} are given by (9). Similar to (10),

$$\bigcap_{i=1}^m \tilde{A}_i^n \cap \bigcap_{k=n-j+1}^n \{k \notin R_n \Pi_1(r\varepsilon)\} = \bigcap_{i=1}^m \tilde{A}_i^{n-j} \cap \bigcap_{k=n-j+1}^n \{k \notin R_n \Pi_1(r\varepsilon)\}.$$

The event $\bigcap_{k=n-j+1}^n \{k \notin R_n \Pi_1(r\varepsilon)\}$ does not depend on u , it suffices to show that this event has probability at least $1 - j(1 - e^{-\varepsilon})$ as

$$\left(\bigcap_{i=1}^m \tilde{A}_i^n \right) \Delta \left(\bigcap_{i=1}^m \tilde{A}_i^{n-j} \right) \subset \left(\bigcap_{k=n-j+1}^n \{k \notin R_n \Pi_1(r\varepsilon)\} \right)^c.$$

Now notice that $\mathbb{P}(\{k \in R_n \Pi_1(r\varepsilon)\}) = 1 - e^{-r\varepsilon} \leq 1 - e^{-\varepsilon}$ and hence

$$\mathbb{P}\left(\bigcup_{k=n-j+1}^n \{k \in R_n \Pi_1(r\varepsilon)\}\right) \leq j(1 - e^{-\varepsilon}).$$

Thus it follows that (7) holds for every $A \in \mathcal{A}'$. □

Let $\lambda_{n,k}$ be the rate at which a collision involving exactly k fixed blocks occurs when there are currently n blocks present. Define

$$\gamma_{n,k} := \binom{n}{k} \lambda_{n,k}$$

which is the total rate of mergers of k blocks when n blocks are present. Let us fix $M > 1$ and define the process $\mathcal{Z}_\varepsilon^M$ by

$$\mathcal{Z}_\varepsilon^M(r) := \mathcal{Z}_\varepsilon(r) \wedge M, \quad r \in [0, 1).$$

The next lemma gives is the central lemma of the proof and gives estimates on the jump probabilities of \mathcal{Z}_ε . The expectation of the random variable which these probabilities are close to will be identified in Lemma 3.3.

Lemma 3.2. *Let $K \subset [0, 1)$ be a compact set, then for any $j \leq M$,*

$$\begin{aligned} \sup_{r \in K} \sup_{A \in \mathcal{F}_r^\varepsilon} \limsup_{\delta \rightarrow 0} \left| \frac{1}{\delta} \mathbb{P}(A; \mathcal{Z}_\varepsilon^M(r+\delta) - \mathcal{Z}_\varepsilon^M(r) = j) - \mathbb{E} \left[\varepsilon (j + \mathcal{Z}_\varepsilon^M(r)) \frac{\gamma_{N^\varepsilon(r), j+1}}{N^\varepsilon(r)} \mathbb{1}_{A; \mathcal{Z}_\varepsilon^M(r) < M-j} \right] \right| \\ \leq j(1 - e^{-\varepsilon}) \frac{3M(M+1)}{2} \mathbb{E} \left[\sup_{r \in K} \varepsilon \frac{\gamma_{N^\varepsilon(r), j+1}}{N^\varepsilon(r)} \right]. \end{aligned}$$

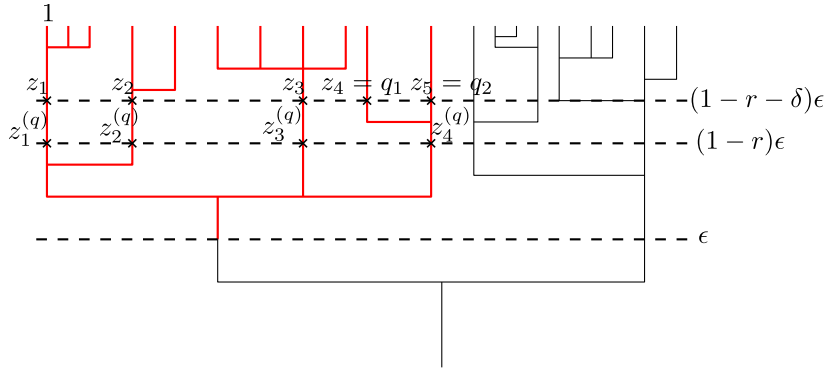


Fig. 3. Example showing how to evaluate interpret $\mathbb{P}(\mathcal{Z}_\varepsilon(r + \delta) = \ell; \mathcal{Z}_\varepsilon(r) = \ell - j | N^\varepsilon(r + \delta) = n)$. In this figure $\mathbf{z} = \{1, 2, 3, 4, 5\}$, $\mathbf{q} = \{4, 5\}$ and $\mathbf{z}^{(\mathbf{q})} = \{1, 2, 3, 4\}$.

Before we begin let us give a small outline of the proof. Suppose for simplicity that we want to work out $\mathbb{P}(\mathcal{Z}_\varepsilon(r + \delta) = \ell; \mathcal{Z}_\varepsilon(r) = \ell - j | N^\varepsilon(r + \delta) = n)$. This means that there are ℓ blocks at time $(1 - r - \delta)\varepsilon$ with labels $\mathbf{z} \in \mathcal{S}_{\ell, n}$ which form the block containing 1 at time ε . Furthermore during the time interval $[(1 - r - \delta)\varepsilon, (1 - r)\varepsilon]$ there must have been a merger involving some labels $\mathbf{q} \subset \mathbf{z}$ with $|\mathbf{q}| = j + 1$, see Figure 3.

The key to unravelling this statement is to successively use Bayes' formula and the Markov property of the coalescent. First we can work out the probability that there is merger which merges the blocks with labels \mathbf{q} during the time interval $[(1 - r - \delta)\varepsilon, (1 - r)\varepsilon]$. Conditionally on this, we work out the probability that indices $\mathbf{z}^{(\mathbf{q})}$ of size $\ell - j$ at time $(1 - r)\varepsilon$ must make up the block containing 1 at time ε . The indices $\mathbf{z}^{(\mathbf{q})}$ are the indices which correspond to the positions of indices \mathbf{z} after the merger involving blocks with labels \mathbf{q} has taken place (see again Figure 3).

Proof of Lemma 3.2. Let $n \in \mathbb{N}$, $\ell < n \wedge M$, $j \leq (n - \ell) \wedge M$, $r \in [0, 1)$ and $A \in \mathcal{F}_r^\varepsilon$. Suppose further that $\pi \in \mathcal{P}_\infty$ with n blocks and that $\delta > 0$ is small. Let $\tilde{\mathbb{P}}$ be the measure \mathbb{P} conditioned on having at most one coalescent event during the time interval $((1 - r - \delta)\varepsilon, (1 - r)\varepsilon)$. Then it suffices to show the result for $\tilde{\mathbb{P}}$ since for any event A' , $\mathbb{P}(A') = \tilde{\mathbb{P}}(A') + o(\delta)$ uniformly over $r \in K$ a compact set K .

For $\mathbf{q} \subset \mathbf{z}$ let $M_\delta(\mathbf{q})$ denote the event that there is a coalescent event in the interval $((1 - r - \delta)\varepsilon, (1 - r)\varepsilon)$ which merges the blocks of $\Pi^\varepsilon(r + \delta)$ with labels \mathbf{q} . We claim that for any $\mathbf{z} \in \mathcal{S}_{\ell+j, n}$

$$\begin{aligned} & \tilde{\mathbb{P}}(A; \kappa(r + \delta, \mathbf{z}); \mathcal{Z}_\varepsilon(r) = \ell; \Pi^\varepsilon(r + \delta) = \pi) \\ &= \sum_{\substack{\mathbf{q} \subset \mathbf{z} \\ |\mathbf{q}|=j+1}} \tilde{\mathbb{P}}(A; M_\delta(\mathbf{q}); \kappa(r, \mathbf{z}^{(\mathbf{q})}); \Pi^\varepsilon(r + \delta) = \pi), \end{aligned} \quad (11)$$

where $\mathbf{z}^{(\mathbf{q})} \in \mathcal{S}_{\ell, n-j}$ represents the position of the indices \mathbf{z} after the merger involving indices \mathbf{q} has occurred (see Figure 3). Indeed, suppose there is only one coalescent event during the interval $((1 - r - \delta)\varepsilon, (1 - r)\varepsilon)$. On the event $\kappa(r + \delta, \mathbf{z}) \cap \{\Pi^\varepsilon(r + \delta) = \pi\}$, $\mathcal{Z}_\varepsilon(r) = \ell$ if and only if during this coalescent event exactly $j + 1$ blocks $\pi_{q_1}, \dots, \pi_{q_{j+1}}$ merge with $q_i \in \mathbf{z}$ for each $i \leq j + 1$. We set $\mathbf{q} = \{q_1, \dots, q_{j+1}\}$. After this merger the blocks of π with labels given by \mathbf{z} now have new labels $\mathbf{z}^{(\mathbf{q})} \in \mathcal{S}_{\ell, n-j}$. Then we require that the blocks of $\Pi^\varepsilon(r)$ with labels $\mathbf{z}^{(\mathbf{q})}$ eventually merge to give $\Pi_1^\varepsilon(1)$, i.e. $\kappa(r, \mathbf{z}^{(\mathbf{q})})$ holds.

Similar to $\mathbf{z}^{(\mathbf{q})}$, let $\pi^{(\mathbf{q})} \in \mathcal{P}_\infty$ be the partition obtained from π by merging the blocks with labels \mathbf{q} . Markov property of the coalescent implies that

$$\tilde{\mathbb{P}}(A; \kappa(r, \mathbf{z}^{(\mathbf{q})}) | \Pi^\varepsilon(r + \delta) = \pi; M_\delta(\mathbf{q})) = \tilde{\mathbb{P}}(A; \kappa(r, \mathbf{z}^{(\mathbf{q})}) | \Pi^\varepsilon(r) = \pi^{(\mathbf{q})}).$$

The above equation and (11) gives

$$\begin{aligned}
& \tilde{\mathbb{P}}(A; \kappa(r + \delta, \mathbf{z}); \mathcal{Z}_\varepsilon(r) = \ell; \Pi^\varepsilon(r + \delta) = \pi) \\
&= \sum_{\substack{\mathbf{q} \subset \mathbf{z} \\ |\mathbf{q}|=j+1}} \tilde{\mathbb{P}}(M_\delta(\mathbf{q}) | \Pi^\varepsilon(r + \delta) = \pi) \tilde{\mathbb{P}}(A; \kappa(r, \mathbf{z}^{(\mathbf{q})}) | \Pi^\varepsilon(r + \delta) = \pi; M_\delta(\mathbf{q})) \tilde{\mathbb{P}}(\Pi^\varepsilon(r + \delta) = \pi) \\
&= \sum_{\substack{\mathbf{q} \subset \mathbf{z} \\ |\mathbf{q}|=j+1}} \tilde{\mathbb{P}}(M_\delta(\mathbf{q}) | \Pi^\varepsilon(r + \delta) = \pi) \tilde{\mathbb{P}}(A; \kappa(r, \mathbf{z}^{(\mathbf{q})}) | \Pi^\varepsilon(r) = \pi^{(\mathbf{q})}) \tilde{\mathbb{P}}(\Pi^\varepsilon(r + \delta) = \pi). \tag{12}
\end{aligned}$$

Now we compute each term inside the sum in (12). Using the exchangeability of the coalescent for the first term gives that

$$\tilde{\mathbb{P}}(M_\delta(\mathbf{q}) | \Pi^\varepsilon(r + \delta) = \pi) = \tilde{\mathbb{P}}(M_\delta(\{1, \dots, j+1\}) | N^\varepsilon(r + \delta) = n). \tag{13}$$

For the second term notice first that by exchangeability for any $\mathbf{z}' \in \mathcal{S}_{\ell, n-j}$ and any partition π' with $n-j$ blocks,

$$\tilde{\mathbb{P}}(A; \kappa(r, \mathbf{z}^{(\mathbf{q})}) | \Pi^\varepsilon(r) = \pi^{(\mathbf{q})}) = \tilde{\mathbb{P}}(A; \kappa(r, \mathbf{z}') | \Pi^\varepsilon(r) = \pi').$$

Thus it follows that

$$\begin{aligned}
\mathbb{P}(A; Z_\varepsilon(r) = \ell; N^\varepsilon(r) = n-j) &= \sum_{\# \pi = n-j} \sum_{\mathbf{z}' \in \mathcal{S}_{\ell, n-j}} \tilde{\mathbb{P}}(A; \kappa(r, \mathbf{z}') | \Pi^\varepsilon(r) = \pi') \tilde{\mathbb{P}}(\Pi^\varepsilon(r) = \pi') \\
&= |\mathcal{S}_{\ell, n-j}| \tilde{\mathbb{P}}(A; \kappa(r, \mathbf{z}^{(\mathbf{q})}) | \Pi^\varepsilon(r) = \pi^{(\mathbf{q})}) \tilde{\mathbb{P}}(N^\varepsilon(r) = n-j).
\end{aligned}$$

Now $|\mathcal{S}_{\ell, n-j}| = \binom{n-j-1}{\ell-1}$, thus

$$\tilde{\mathbb{P}}(A; \kappa(r, \mathbf{z}^{(\mathbf{q})}) | \Pi^\varepsilon(r) = \pi^{(\mathbf{q})}) = \tilde{\mathbb{P}}(A; Z_\varepsilon(r) = \ell | N^\varepsilon(r) = n-j) \left(\frac{n-j-1}{\ell-1} \right)^{-1}. \tag{14}$$

Plugging (13) and (14) into (12) and summing gives

$$\begin{aligned}
& \tilde{\mathbb{P}}(A; Z_\varepsilon(r + \delta) = \ell + j; \mathcal{Z}_\varepsilon(r) = \ell; N^\varepsilon(r + \delta) = n) \\
&= \sum_{\# \pi = n} \sum_{\mathbf{z} \in \mathcal{S}_{\ell+j, n}} \tilde{\mathbb{P}}(A; \kappa(r + \delta, \mathbf{z}); \mathcal{Z}_\varepsilon(r) = \ell; \Pi^\varepsilon(r + \delta) = \pi) \\
&= \sum_{\# \pi = n} \sum_{\mathbf{z} \in \mathcal{S}_{\ell+j, n}} \sum_{\substack{\mathbf{q} \subset \mathbf{z} \\ |\mathbf{q}|=j+1}} \tilde{\mathbb{P}}(M_\delta(\mathbf{q}) | \Pi^\varepsilon(r + \delta) = \pi) \tilde{\mathbb{P}}(A; \kappa(r, \mathbf{z}^{(\mathbf{q})}) | \Pi^\varepsilon(r) = \pi^{(\mathbf{q})}) \tilde{\mathbb{P}}(\Pi^\varepsilon(r + \delta) = \pi) \\
&= \sum_{\# \pi = n} \sum_{\mathbf{z} \in \mathcal{S}_{\ell+j, n}} \sum_{\substack{\mathbf{q} \subset \mathbf{z} \\ |\mathbf{q}|=j+1}} \frac{\tilde{\mathbb{P}}(M_\delta(\{1, \dots, j\}) | N^\varepsilon(r + \delta) = n) \tilde{\mathbb{P}}(A; Z_\varepsilon(r) = \ell | N^\varepsilon(r) = n-j) \tilde{\mathbb{P}}(\Pi^\varepsilon(r + \delta) = \pi)}{\binom{n-j-1}{\ell-1}} \\
&= \frac{\binom{n-1}{\ell+j-1} \binom{\ell+j}{j}}{\binom{n-j-1}{\ell-1}} \tilde{\mathbb{P}}(M_\delta(\{1, \dots, j+1\}) | N^\varepsilon(r + \delta) = n) \tilde{\mathbb{P}}(A; Z_\varepsilon(r) = \ell | N^\varepsilon(r) = n-j) \tilde{\mathbb{P}}(N^\varepsilon(r + \delta) = n) \\
&= \frac{j+\ell}{n} \binom{n}{j+1} \tilde{\mathbb{P}}(M_\delta(\{1, \dots, j+1\}) | N^\varepsilon(r + \delta) = n) \tilde{\mathbb{P}}(A; Z_\varepsilon(r) = \ell | N^\varepsilon(r) = n-j) \tilde{\mathbb{P}}(N^\varepsilon(r + \delta) = n),
\end{aligned}$$

where $\#\pi$ denotes the number of blocks of π . Diving by δ and taking limits gives

$$\begin{aligned} \lim_{\delta \downarrow 0} \frac{1}{\delta} \tilde{\mathbb{P}}(A; Z_\varepsilon(r + \delta) = \ell + j; Z_\varepsilon(r) = \ell; N^\varepsilon(r + \delta) = n) \\ = \frac{j + \ell}{n} \varepsilon \gamma_{n, j+1} \tilde{\mathbb{P}}(A; Z_\varepsilon(r) = \ell | N^\varepsilon(r) = n - j) \tilde{\mathbb{P}}(N^\varepsilon(r) = n), \end{aligned} \quad (15)$$

where we have used the fact that

$$\binom{n}{j+1} \lim_{\delta \downarrow 0} \frac{1}{\delta} \tilde{\mathbb{P}}(M_\delta(\{1, \dots, j+1\}) | N^\varepsilon(r + \delta) = n) = \varepsilon \gamma_{n, j+1}.$$

We wish to sum (15) over n and then apply dominated convergence to swap the limit with the summation. For $\delta' < 1 - r$ there exists a constant $C > 0$ such that

$$\begin{aligned} \sup_{\delta < \delta'} \frac{1}{\delta} \tilde{\mathbb{P}}(A; Z_\varepsilon(r + \delta) = \ell + j; Z_\varepsilon(r) = \ell; N^\varepsilon(r + \delta) = n) \\ \leq C \sup_{\delta < \delta'} \frac{1}{\delta} \tilde{\mathbb{P}}(A; Z_\varepsilon(r + \delta) = \ell + j; Z_\varepsilon(r) = \ell; N^\varepsilon(r) = n) \\ \leq C \sup_{\delta < \delta'} \frac{1}{\delta} C \mathbb{P}(\text{there is a coalescent event during } ((1 - r - \delta)\varepsilon, (1 - r)\varepsilon); N^\varepsilon(r) = n) \\ = C \sup_{\delta < \delta'} \frac{1}{\delta} (1 - e^{-\delta \varepsilon \lambda_n}) \tilde{\mathbb{P}}(N^\varepsilon(r) = n) \\ \leq C \varepsilon \lambda_n \mathbb{P}(N^\varepsilon(r) = n), \end{aligned}$$

where $\lambda_n = \sum_{k \leq n} \binom{n}{k} \lambda_{n, k}$ is the total rate of coalescent events and in the final inequality we have used the fact that for $x \geq 0$, $1 - e^{-x} \leq x$. The final term on the line above is summable hence by (15) and dominated convergence theorem,

$$\begin{aligned} \lim_{\delta \downarrow 0} \frac{1}{\delta} \tilde{\mathbb{P}}(A; Z_\varepsilon(r + \delta) = \ell + j; Z_\varepsilon(r) = \ell) \\ = \sum_{n=\ell+j}^{\infty} \frac{j + \ell}{n} \varepsilon \gamma_{n, j+1} \tilde{\mathbb{P}}(A; Z_\varepsilon(r) = \ell | N^\varepsilon(r) = n - j) \tilde{\mathbb{P}}(N^\varepsilon(r) = n). \end{aligned} \quad (16)$$

Next Lemma 3.1 gives

$$|\tilde{\mathbb{P}}(A; Z_\varepsilon(r) = \ell | N^\varepsilon(r) = n - j) - \tilde{\mathbb{P}}(A; Z_\varepsilon(r) = \ell | N^\varepsilon(r) = n)| \leq j(1 - e^{-\varepsilon}). \quad (17)$$

Using (17) together with (16) we get that

$$\begin{aligned} \limsup_{\delta \downarrow 0} \left| \frac{1}{\delta} \tilde{\mathbb{P}}(A; Z_\varepsilon^M(r + \delta) - Z_\varepsilon^M(r) = j) - \sum_{\ell=1}^M \sum_{n=\ell+j}^{\infty} \varepsilon \gamma_{n, j+1} \frac{j + \ell}{n} \tilde{\mathbb{P}}(A; Z_\varepsilon(r) = \ell; N^\varepsilon(r) = n) \right| \\ \leq j(1 - e^{-\varepsilon}) \sum_{\ell=1}^M \sum_{n=\ell+j}^{\infty} \varepsilon \gamma_{n, j+1} \frac{j + \ell}{n} \tilde{\mathbb{P}}(N^\varepsilon(r) = n) \\ \leq j(1 - e^{-\varepsilon}) \frac{3M(M+1)}{2} \sum_{n=j+1}^{\infty} \varepsilon \frac{\gamma_{n, j+1}}{n} \tilde{\mathbb{P}}(N^\varepsilon(r) = n) \end{aligned}$$

and the result follows. \square

Now we can identify the limiting behaviour of the expectations in Lemma 3.2 when $\varepsilon \downarrow 0$.

Lemma 3.3. *For any $M > 1$, $j < M$ and $K \subset [0, 1)$ compact,*

$$\lim_{\varepsilon \rightarrow 0} \mathbb{E} \left[\sup_{r \in K} \left| \varepsilon \frac{\gamma_{N^\varepsilon(r), j+1}}{N^\varepsilon(r)} - A_\Lambda \frac{\Gamma(2-\alpha)\Gamma(j-\alpha+1)}{(1-r)\alpha\Gamma(j+2)} \right| \right] = 0. \quad (18)$$

Let us present the idea of the proof of this lemma. Firstly [5, Theorem 2] implies that $\varepsilon/N^\varepsilon(r) \approx (1-r)^{-1}N^\varepsilon(r)^{-\alpha}$ thus we are effectively seeking to show that

$$\lim_{n \rightarrow \infty} \frac{\gamma_{n, j+1}}{n^\alpha} = A_\Lambda \frac{\Gamma(2-\alpha)\Gamma(j-\alpha+1)}{\alpha\Gamma(j+2)}.$$

This limit is not hard to take and has appeared in literature before in [10, Lemma 4] and [7, (10)]. The proof presented below is somewhat more delicate since we would like to obtain a strong notion of convergence.

Proof of Lemma 3.3. Fix $M > 1$, $j < M$ and $K \subset [0, 1)$ compact. Firstly from [5, Theorem 2] we have that there exists a constant $C_\alpha > 0$ such that

$$\lim_{\varepsilon \rightarrow 0} \mathbb{E} \left[\sup_{r \in K} \left| \frac{(1-r)\varepsilon}{C_\alpha N^\varepsilon(r)^{1-\alpha}} - 1 \right|^2 \right] = 0. \quad (19)$$

One can show (see [18, XIII.6]) that this constant is given by

$$C_\alpha = \frac{\alpha}{\Gamma(2-\alpha)}.$$

Using the Cauchy–Schwarz inequality we have

$$\begin{aligned} & \mathbb{E} \left[\sup_{r \in K} \left| \varepsilon \frac{\gamma_{N^\varepsilon(r), j+1}}{N^\varepsilon(r)} - \frac{\alpha}{(1-r)\Gamma(2-\alpha)} \frac{\gamma_{N^\varepsilon(r), j+1}}{N^\varepsilon(r)^\alpha} \right| \right]^2 \\ &= \mathbb{E} \left[\sup_{r \in K} \left| \frac{\gamma_{N^\varepsilon(r), j+1}}{N^\varepsilon(r)^\alpha} \left(\frac{(1-r)\varepsilon\Gamma(2-\alpha)}{\alpha N^\varepsilon(r)^{1-\alpha}} - 1 \right) \right| \right]^2 \\ &\leq \mathbb{E} \left[\sup_{r \in K} \left| \frac{\gamma_{N^\varepsilon(r), j+1}}{N^\varepsilon(r)^\alpha} \right|^2 \right] \mathbb{E} \left[\sup_{r \in K} \left| \frac{(1-r)\varepsilon\Gamma(2-\alpha)}{\alpha N^\varepsilon(r)^{1-\alpha}} - 1 \right|^2 \right]. \end{aligned} \quad (20)$$

The second term on the last line converges to zero by (19), thus we focus on the first term in the last line.

Recall that for $n \in \mathbb{N}$ we have that

$$\gamma_{n, j+1} = \binom{n}{j+1} \lambda_{n, j+1} = \frac{\Gamma(n+1)}{\Gamma(j+2)\Gamma(n-j)} \int_0^1 p^{j-1} (1-p)^{n-j-1} \Lambda(dp). \quad (21)$$

Moreover $\Lambda(dp) = f(p) dp$ where $f(p) \sim A_\Lambda p^{1-\alpha}$ as $p \rightarrow 0$. Fix $\eta_0 > 0$, then there exist a $p_0 \in (0, 1)$ such that whenever $p < p_0$ we have $|f(p) - A_\Lambda p^{1-\alpha}| \leq \eta p^{1-\alpha}$. Thus

$$\begin{aligned} & \left| \int_0^{p_0} p^{j-1} (1-p)^{n-j-1} \Lambda(dp) - A_\Lambda \int_0^{p_0} p^{j-\alpha} (1-p)^{n-j-1} dp \right| \\ &\leq \eta A_\Lambda \int_0^{p_0} p^{j-\alpha} (1-p)^{n-j-1} dp \leq \eta A_\Lambda \int_0^1 p^{j-\alpha} (1-p)^{n-j-1} dp. \end{aligned} \quad (22)$$

Then combining (21) and (22) we have

$$\begin{aligned}
& \left| \gamma_{n,j+1} - \binom{n}{j+1} A_\Lambda \int_0^1 p^{j-\alpha} (1-p)^{n-j-1} dp \right| \\
& \leq \eta \binom{n}{j+1} A_\Lambda \int_0^1 p^{j-\alpha} (1-p)^{n-j-1} dp + \int_{p_0}^1 p^{j-1} (1-p)^{n-j-1} |A_\Lambda p^{1-\alpha} - f(p)| dp \\
& \leq \eta \binom{n}{j+1} A_\Lambda \int_0^1 p^{j-\alpha} (1-p)^{n-j-1} dp + (1-p_0)^{n-j-1} (A_\Lambda p_0^{1-\alpha} + \Lambda[0, 1]).
\end{aligned} \tag{23}$$

From the definition of the Beta function we have that

$$\binom{n}{j+1} \int_0^1 p^{j-\alpha} (1-p)^{n-j-1} dp = \frac{\Gamma(j-\alpha+1)\Gamma(n+1)}{\Gamma(j+2)\Gamma(n-\alpha+1)}. \tag{24}$$

Thus using (23) and (24) we have,

$$\begin{aligned}
& \mathbb{E} \left[\sup_{r \in K} \left| \frac{\gamma_{N^\varepsilon(r), j+1}}{N^\varepsilon(r)^\alpha} - A_\Lambda \frac{\Gamma(j-\alpha+1)\Gamma(N^\varepsilon(r)+1)}{N^\varepsilon(r)^\alpha \Gamma(j+2)\Gamma(N^\varepsilon(r)-\alpha+1)} \right|^2 \right] \\
& \leq \eta A_\Lambda \mathbb{E} \left[\sup_{r \in K} \left| \frac{\Gamma(j-\alpha+1)\Gamma(N^\varepsilon(r)+1)}{N^\varepsilon(r)^\alpha \Gamma(j+2)\Gamma(N^\varepsilon(r)-\alpha+1)} \right|^2 \right] \\
& \quad + (A_\Lambda p_0^{1-\alpha} + \Lambda[0, 1])^2 \mathbb{E} \left[\sup_{r \in K} (1-p_0)^{2(N^\varepsilon(r)-j-1)} \right].
\end{aligned} \tag{25}$$

The final term on the right hand side converges to 0 as $\varepsilon \rightarrow 0$. For the penultimate term an application of Stirling's formula yields that

$$\mathbb{E} \left[\sup_{r \in K} \left| \frac{\Gamma(j-\alpha+1)\Gamma(N^\varepsilon(r)+1)}{N^\varepsilon(r)^\alpha \Gamma(j+2)\Gamma(N^\varepsilon(r)-\alpha+1)} - \frac{\Gamma(j-\alpha+1)}{\Gamma(j+2)} \right|^2 \right] \rightarrow 0$$

as $\varepsilon \rightarrow 0$. Thus the term on the left hand side of (25) goes to zero as $\varepsilon \rightarrow 0$. Moreover using the triangle inequality we have that

$$\begin{aligned}
& \mathbb{E} \left[\sup_{r \in K} \left| \frac{\gamma_{N^\varepsilon(r), j+1}}{N^\varepsilon(r)^\alpha} - A_\Lambda \frac{\Gamma(j-\alpha+1)}{\Gamma(j+2)} \right|^2 \right] \\
& \leq \mathbb{E} \left[\sup_{r \in K} \left| \frac{\gamma_{N^\varepsilon(r), j+1}}{N^\varepsilon(r)^\alpha} - A_\Lambda \frac{\Gamma(j-\alpha+1)\Gamma(N^\varepsilon(r)+1)}{N^\varepsilon(r)^\alpha \Gamma(j+2)\Gamma(N^\varepsilon(r)-\alpha+1)} \right|^2 \right] \\
& \quad + \mathbb{E} \left[\sup_{r \in K} \left| A_\Lambda \frac{\Gamma(j-\alpha+1)\Gamma(N^\varepsilon(r)+1)}{N^\varepsilon(r)^\alpha \Gamma(j+2)\Gamma(N^\varepsilon(r)-\alpha+1)} - \frac{\Gamma(j-\alpha+1)}{\Gamma(j+2)} \right|^2 \right] \\
& \rightarrow 0
\end{aligned} \tag{26}$$

as $\varepsilon \rightarrow 0$. A final application of the triangle inequality and using (20) gives

$$\begin{aligned}
& \mathbb{E} \left[\sup_{r \in K} \left| \varepsilon \frac{\gamma_{N^\varepsilon(r), j+1}}{N^\varepsilon(r)} - A_\Lambda \frac{\Gamma(2-\alpha)\Gamma(j-\alpha+1)}{(1-r)^\alpha \Gamma(j+2)} \right| \right] \\
& \leq \mathbb{E} \left[\sup_{r \in K} \left| \varepsilon \frac{\gamma_{N^\varepsilon(r), j+1}}{N^\varepsilon(r)} - \frac{\alpha}{(1-r)\Gamma(2-\alpha)} \frac{\gamma_{N^\varepsilon(r), j+1}}{N^\varepsilon(r)^\alpha} \right| \right] + \mathbb{E} \left[\sup_{r \in K} \left| \frac{\gamma_{N^\varepsilon(r), j+1}}{N^\varepsilon(r)^\alpha} - \frac{\Gamma(j-\alpha+1)}{\Gamma(j+2)} \right| \right].
\end{aligned}$$

The proof now follows from (26) and (25). \square

Remark 3.4. In the case when $\alpha = 2$ all the arguments in this section apply apart from Lemma 3.3. In its place we have that for $\alpha = 2$ as $\varepsilon \rightarrow 0$,

$$\mathbb{E} \left[\sup_{r \in K} \left| \varepsilon \frac{\gamma_{N^\varepsilon(r), j+1}}{N^\varepsilon(r)} - \frac{1}{1-r} \mathbb{1}_{j=1} \right| \right] \rightarrow 0.$$

Indeed, this follows from the fact that for $\alpha = 2$, $\gamma_{n,2} = \binom{n}{2}$ and $\gamma_{n,j+1} = 0$ for $j > 1$, together with [5, Theorem 2].

The preceding two lemmas directly imply that for each $M > 1$, $j \leq M$ and $K \subset [0, 1)$, there exists a constant $C = C(M, K) > 0$ such that

$$\limsup_{\delta \rightarrow 0} \left| \frac{1}{\delta} \mathbb{P}(A; \mathcal{Z}_\varepsilon^M(r + \delta) - \mathcal{Z}_\varepsilon^M(r) = j) - A_\Lambda \frac{\Gamma(2-\alpha)\Gamma(j-\alpha+1)}{(1-r)\alpha\Gamma(j+2)} \mathbb{E}[(j + \mathcal{Z}_\varepsilon^M(r))\mathbb{1}_A] \right| \leq C\varepsilon \quad (27)$$

uniformly over all $r \in K$ and $A \in \mathcal{F}_r^\varepsilon$.

For $f : \mathbb{N} \rightarrow \mathbb{R}$ recall that

$$L_r f(i) := A_\Lambda \sum_{j \geq 1} (i+j) \frac{\Gamma(2-\alpha)\Gamma(j-\alpha+1)}{(1-r)\alpha\Gamma(j+2)} [f(i+j) - f(i)]. \quad (28)$$

Using the last two lemmas we are able to show that $\mathcal{Z}_\varepsilon^M$ almost solves a martingale problem. This will enable us to show that the limiting process satisfies the martingale problem.

Lemma 3.5. Let $u < r \in [0, 1)$ then for any $f : \mathbb{N} \rightarrow \mathbb{R}$ with support in $\{1, \dots, \lfloor M \rfloor\}$,

$$\lim_{\varepsilon \rightarrow 0} \sup_{A \in \mathcal{F}_u^\varepsilon} \left| \mathbb{E} \left[\left(f(\mathcal{Z}_\varepsilon^M(r)) - \int_0^r L_s f(\mathcal{Z}_\varepsilon^M(s)) ds \right) \mathbb{1}_A \right] - \mathbb{E} \left[\left(f(\mathcal{Z}_\varepsilon^M(u)) - \int_0^u L_s f(\mathcal{Z}_\varepsilon^M(s)) ds \right) \mathbb{1}_A \right] \right| = 0.$$

Proof. Fix $u < r \in [0, 1)$, $A \in \mathcal{F}_s^\varepsilon$ and $f : \mathbb{N} \rightarrow \mathbb{R}$ with support in $\{1, \dots, \lfloor M \rfloor\}$. Suppose that $\delta > 0$ is small. Suppose that $\{s_\ell\}_{\ell=0}^m$ is such that $s_0 = u$, $s_m = r$ and $s_\ell - s_{\ell-1} \leq \delta$ for each $\ell = 1, \dots, m$. Now

$$\begin{aligned} & \mathbb{E}[(f(\mathcal{Z}_\varepsilon^M(r)) - f(\mathcal{Z}_\varepsilon^M(u)))\mathbb{1}_A] \\ &= \sum_{\ell=1}^m \mathbb{E}[(f(\mathcal{Z}_\varepsilon^M(s_\ell)) - f(\mathcal{Z}_\varepsilon^M(s_{\ell-1})))\mathbb{1}_A] \\ &= \sum_{\ell=1}^m \sum_{j=1}^M \sum_{i=1}^M [f(i+j) - f(i)] \mathbb{P}(\mathcal{Z}_\varepsilon^M(s_\ell) - \mathcal{Z}_\varepsilon^M(s_{\ell-1}) = j; \mathcal{Z}_\varepsilon^M(s_{\ell-1}) = i; A). \end{aligned} \quad (29)$$

Now by (27), for small enough $\delta > 0$, there exists a constant $C > 0$ independent of i, j, ℓ and A (but depending on M) such that

$$\begin{aligned} & \left| [f(i+j) - f(i)] \frac{\mathbb{P}(\mathcal{Z}_\varepsilon^M(s_\ell) - \mathcal{Z}_\varepsilon^M(s_{\ell-1}) = j; \mathcal{Z}_\varepsilon^M(s_{\ell-1}) = i; A)}{\delta} - L_{s_{\ell-1}} f(i) \mathbb{P}(\mathcal{Z}_\varepsilon^M(s_{\ell-1}) = i; A) \right| \\ & \leq C\varepsilon. \end{aligned} \quad (30)$$

Suppose also that $\delta > 0$ is small enough so that

$$\mathbb{E} \left[\left| \int_u^r L_s f(\mathcal{Z}_\varepsilon^M(s)) ds - \delta \sum_{\ell=1}^m L_{s_{\ell-1}} f(\mathcal{Z}_\varepsilon^M(s_{\ell-1})) \right| \right] < \varepsilon. \quad (31)$$

Thus from (29), (30) and (31) and using the fact that f is bounded,

$$\left| \mathbb{E} \left[\left(f(\mathcal{Z}_\varepsilon^M(r)) - f(\mathcal{Z}_\varepsilon^M(u)) - \int_u^r L_s f(\mathcal{Z}_\varepsilon^M(s)) ds \right) \mathbb{1}_A \right] \right| \leq C' \varepsilon$$

for some new constant $C' > 0$. The result follows by taking limits. \square

Next we show a tightness result.

Lemma 3.6. *For each $M > 1$, the sequence of processes $\{\mathcal{Z}_\varepsilon^M\}_{\varepsilon>0}$ is tight in the Skorokhod sense.*

Proof. Let $M > 1$ be fixed. To prove this lemma we will verify the conditions of [2, Corollary 2]. Note that $\mathcal{Z}_\varepsilon^M$ is uniformly bounded by M . Hence it suffices to check that for each $s \in [0, 1]$, there exists a deterministic constant $\alpha(\varepsilon, \delta)$, possibly depending on s , such that

$$\sup_{r \leq s} \mathbb{P}(\mathcal{Z}_\varepsilon^M(r + \delta) - \mathcal{Z}_\varepsilon^M(r) > 0 | \mathcal{F}_r^\varepsilon) \leq \alpha(\varepsilon, \delta) \quad \text{a.s.} \quad (32)$$

and

$$\lim_{\delta \rightarrow 0} \limsup_{\varepsilon \rightarrow 0} \alpha(\varepsilon, \delta) = 0. \quad (33)$$

Fix $s \in [0, 1]$, let $r \leq s$ and let $\delta > 0$. Let $M_\varepsilon^2(a, b)$ denote the event that there are two or more coalescent events during the time interval $((1-a)\varepsilon, (1-b)\varepsilon)$. Then by (15) there exists a constant $C > 0$ such that gives that for each $A \in \mathcal{F}_r^\varepsilon$,

$$\begin{aligned} & \mathbb{P}(A; \mathcal{Z}_\varepsilon^M(r + \delta) - \mathcal{Z}_\varepsilon^M(r) > 0) \\ & \leq \sum_{\ell=1}^M \sum_{j=1}^{\ell} \sum_{n=j}^{\infty} \mathbb{P}(A; Z_\varepsilon(r + \delta) = \ell + j; Z_\varepsilon(r) = \ell; N^\varepsilon(r + \delta) = n | M_\varepsilon^2(r, r + \delta)^c) + \mathbb{P}(M_\varepsilon^2(r, r + \delta)) \\ & \leq C \delta \varepsilon \sum_{\ell=1}^M \sum_{j=1}^{\ell} \sum_{n=j}^{\infty} \frac{j + \ell}{n} \gamma_{n, j+1} \mathbb{P}(N^\varepsilon(r) = n | M_\varepsilon^2(r, r + \delta)^c) + \mathbb{P}(M_\varepsilon^2(r, r + \delta)) \\ & \leq C' \delta \sum_{j=1}^M \mathbb{E} \left[\varepsilon \frac{\gamma_{N^\varepsilon(r), j+1}}{N^\varepsilon(r)} \right], \end{aligned}$$

where $C' = C'(M, s) > 0$ is a constant and in the final inequality we have used the fact that $\sup_{r \leq s} \mathbb{P}(M_\varepsilon^2(r, r + \delta)) = o(\delta)$.

Hence we have that (32) holds with

$$\alpha(\varepsilon, \delta) = \delta C' \sup_{r \leq s} \sum_{j=1}^M \mathbb{E} \left[\varepsilon \frac{\gamma_{N^\varepsilon(r), j+1}}{N^\varepsilon(r)} \right].$$

Using Lemma 3.3,

$$\begin{aligned} \limsup_{\varepsilon \rightarrow 0} \alpha(\varepsilon, \delta) &= \delta C' \limsup_{\varepsilon \rightarrow 0} \sup_{r \leq s} \sum_{j=1}^M \mathbb{E} \left[\varepsilon \frac{\gamma_{N^\varepsilon(r), j+1}}{N^\varepsilon(r)} \right] \\ &= \delta C' \sum_{j=1}^M A_\Lambda \frac{\Gamma(2 - \alpha) \Gamma(j - \alpha + 1)}{(1 - s) \alpha \Gamma(j + 2)}. \end{aligned}$$

Taking limits as $\delta \rightarrow 0$ in the above equation we obtain (33). \square

Now we can show the convergence of \mathcal{Z}_ε in the Skorokhod sense to a Markov process $\mathcal{Z} = (\mathcal{Z}(r) : r \in [0, 1])$ with generator given by (28).

Proof of Theorem 1.2. We are done if we can show that for any $M > 1$ we have that

$$\mathcal{Z}_\varepsilon^M \rightarrow (\mathcal{Z}(r) \wedge M : r \in [0, 1])$$

in the Skorokhod sense as $\varepsilon \rightarrow 0$.

Fix $M > 1$. The sequence of processes $\{\mathcal{Z}_\varepsilon^M\}_{\varepsilon > 0}$ is tight by Lemma 3.6. Suppose now that for some sequence $\varepsilon' \rightarrow 0$ we have that

$$\mathcal{Z}_{\varepsilon'}^M \rightarrow \mathcal{Z}^M$$

in the Skorokhod sense as $\varepsilon' \rightarrow 0$, to some process $\mathcal{Z}^M = (\mathcal{Z}^M(r) : r \in [0, 1])$. It is enough to show that \mathcal{Z}^M has the same law as $(\mathcal{Z}(r) \wedge M : r \in [0, 1])$. We will show this by showing that \mathcal{Z}^M satisfies a martingale problem. Let $f : \mathbb{N} \rightarrow \mathbb{R}$ have support in $\{1, \dots, \lfloor M \rfloor\}$. Then to prove the lemma, it is enough to show that

$$M_r^f := f(\mathcal{Z}^M(r)) - \int_0^r L_s f(\mathcal{Z}^M(s)) ds \quad (34)$$

is a martingale. Let $u \in [0, 1)$ be fixed and let $\mathbb{D}([0, u], \mathbb{N})$ denote the Skorokhod space of càdlàg functions $g : [0, u] \rightarrow \mathbb{N}$. Suppose that $F : \mathbb{D}([0, u], \mathbb{N}) \rightarrow \mathbb{R}$ is a continuous and bounded function. We will show (34) by showing that for $u < r < 1$,

$$\mathbb{E}[M_r^f F((\mathcal{Z}^M(s) : s \leq u))] = \mathbb{E}[M_u^f F((\mathcal{Z}^M(s) : s \leq u))]. \quad (35)$$

Fix $u < r < 1$ and $\eta > 0$. The Skorokhod convergence implies that there exist an $\varepsilon_0 > 0$ such that

$$\left| \mathbb{E}[M_r^f F((\mathcal{Z}^M(s) : s \leq u))] - \mathbb{E}\left[\left(f(\mathcal{Z}_{\varepsilon_0}^M(r)) - \int_0^r L_s f(\mathcal{Z}_{\varepsilon_0}^M(s)) ds\right) F((\mathcal{Z}_{\varepsilon_0}^M(s) : s \leq u))\right] \right| < \eta. \quad (36)$$

On the other hand by Lemma 3.5, we have that there exists an $\varepsilon_1 > 0$ such that

$$\begin{aligned} & \left| \mathbb{E}\left[\left(f(\mathcal{Z}_{\varepsilon_1}^M(r)) - \int_0^r L_s f(\mathcal{Z}_{\varepsilon_1}^M(s)) ds\right) F((\mathcal{Z}_{\varepsilon_1}^M(s) : s \leq u))\right] \right. \\ & \quad \left. - \mathbb{E}\left[\left(f(\mathcal{Z}_{\varepsilon_1}^M(r)) - \int_0^r L_s^{\varepsilon_1} f(\mathcal{Z}_{\varepsilon_1}^M(s)) ds\right) F((\mathcal{Z}_{\varepsilon_1}^M(s) : s \leq u))\right] \right| < \eta. \end{aligned} \quad (37)$$

Applying the Skorokhod convergence once more yields that there exists an $\varepsilon_2 > 0$ such that

$$\left| \mathbb{E}[M_u^f F((\mathcal{Z}^M(s) : s \leq u))] - \mathbb{E}\left[\left(f(\mathcal{Z}_{\varepsilon_2}^M(u)) - \int_0^u L_s f(\mathcal{Z}_{\varepsilon_2}^M(s)) ds\right) F((\mathcal{Z}_{\varepsilon_2}^M(s) : s \leq u))\right] \right| < \eta.$$

Combining this with (36) and (37) gives that

$$|\mathbb{E}[M_r^f F((\mathcal{Z}^M(s) : s \leq u))] - \mathbb{E}[M_u^f F((\mathcal{Z}^M(s) : s \leq u))]| < 3\eta.$$

As $\eta > 0$ is arbitrary this shows (35) which concludes the proof. \square

3.2. Proof of Theorem 1.1

The proof of Theorem 1.1 follows from Theorem 1.2 in a straight forward manner. We first begin by describing a sufficient and necessary condition for the convergence of compact ultra-metric spaces. Let recall that a metric space (X, d) is called an ultra-metric space if

$$d(x, y) \leq d(x, z) \vee d(z, y) \quad \forall x, y, z \in X.$$

Compact ultra-metric spaces have the property that for any $\eta > 0$, there exists disjoint closed balls B_1, \dots, B_k of radius η which cover X .

Definition 3.7. For a compact ultra-metric space (X, d) , the η -cover space (X^η, d^η) is defined as follows. Let B_1, \dots, B_k be the disjoint closed balls of radius η which cover X . Then $X^\eta = \{1, \dots, k\}$ and

$$d^\eta(i, j) := \inf\{d(x, y) : x \in B_i, y \in B_j\}, \quad i, j \in X^\eta.$$

The following lemma is not hard to show and we leave the proof out.

Lemma 3.8. Let $(X_1, d_1), (X_2, d_2), \dots$ be a sequence of compact ultra-metric spaces and for $\eta > 0$ let $(X_1^\eta, d_1^\eta), (X_2^\eta, d_2^\eta), \dots$ denote the respective η -cover spaces.

Suppose that for every $\eta > 0$, $(X_k^\eta, d_k^\eta) \rightarrow (X^\eta, d^\eta)$ under the compact Gromov–Hausdorff metric as $k \rightarrow \infty$, then there exists a compact ultra-metric space (X, d) such that $\lim_{k \rightarrow \infty} (X_k, d_k) = (X, d)$ in the compact Gromov–Hausdorff sense.

Henceforth fix $\eta \in (0, 1)$. Here and throughout we let $B_\varepsilon(i, r)$ denote the closed ball of radius r around i in the space $(E, \varepsilon^{-1}\delta)$. Then to show the theorem it suffices to show that $(B_\varepsilon(1, 1), \varepsilon^{-1}\delta)$ converges weakly. Indeed we may assume that $i = 1$ by exchangeability of the coalescent and the proof for general $r > 0$ follows with more cumbersome notation.

Notice that $(B_\varepsilon(1, 1), \varepsilon^{-1}\delta)$ is a compact ultra-metric space. Let $(S_\varepsilon, r_\varepsilon)$ denote its η -cover space. Thus to show the theorem, it suffices to verify Lemma 3.8 by showing the convergence of $(S_\varepsilon, r_\varepsilon)$ as $\varepsilon \downarrow 0$.

The cardinality of S_ε is precisely the number of blocks of Π at time $(1 - \eta)\varepsilon$ that make up the block containing 1 at time ε and hence $|S_\varepsilon| = \mathcal{Z}_\varepsilon(1 - \eta)$. Furthermore we will now see that we can discover the exact structure of $(S_\varepsilon, r_\varepsilon)$ using the process \mathcal{Z}_ε .

Let $B_1, \dots, B_{\mathcal{Z}_\varepsilon(1-\eta)}$ denote the disjoint closed balls of radius η that cover the space $(B_\varepsilon(1, 1), \varepsilon^{-1}\delta)$. Without a loss of generality we will assume that $1 \in B_1$. Given the process \mathcal{Z}_ε we can construct the space $(S_\varepsilon, r_\varepsilon)$ as follows. First construct a tree T from a branching process. Start the tree with one immortal particle which will not die. We call all other particles mortal. Suppose now that for some $i \in \mathbb{N}$ and $j \geq 1$, $\mathcal{Z}_\varepsilon(r-) = i$ and $\mathcal{Z}_\varepsilon(r+) = i + j$. Then there is a birth at height r of the tree. With probability $(j + 1)/(i + j)$, the immortal particle gives birth to j offspring and with probability $(i - 1)/(i + j)$ a uniformly chosen mortal individual dies gives birth to $j + 1$ offspring. Thus the tree has exactly $\mathcal{Z}_\varepsilon(1 - \eta)$ leaves and height $1 - \eta$. These leaves form the space S_ε and the distance r_ε between two leaves is the genealogical distance i.e. half of the length of the unique path between the two leaves. Use the same procedure but with the process \mathcal{Z} to obtain a space (S, r) . It is clear that (S, r) is a compact metric space.

Using Theorem 1.2 and Skorokhod's representation theorem suppose henceforth that $\mathcal{Z}_\varepsilon \rightarrow \mathcal{Z}$ as $\varepsilon \rightarrow 0$ almost surely under the Skorokhod topology. Let $J_1^\varepsilon, \dots, J_n^\varepsilon$ be the jumps of the process \mathcal{Z}_ε before time $1 - \eta$ and similarly let J_1, \dots, J_m be the jumps of the process \mathcal{Z} before time $1 - \eta$. Then from the almost sure convergence of the process \mathcal{Z}_ε to \mathcal{Z} we can conclude the following for $\varepsilon > 0$ small enough. Firstly $m = n$ and $\mathcal{Z}_\varepsilon(J_i^\varepsilon) = \mathcal{Z}(J_i)$ for each $i \leq n = m$. Second, $\max_{i \leq n} |J_i^\varepsilon - J_i|$ is small.

Thus for $\varepsilon > 0$ small enough this gives a coupling between the spaces $(S_\varepsilon, r_\varepsilon)$ and (S, r) such that $S = S_\varepsilon$ and further

$$\max_{i, j \in S} |r_\varepsilon(i, j) - r(i, j)| \leq 2\mathcal{Z}_\varepsilon(1 - \eta) \max_{i \leq n} |J_i^\varepsilon - J_i|$$

which is small. Hence $(S_\varepsilon, r_\varepsilon) \rightarrow (S, r)$ almost surely under the compact Gromov–Hausdorff topology as $\varepsilon \rightarrow 0$ and the theorem now follows from Lemma 3.8.

4. Kingman case

In this section we will prove Theorem 1.3. As before we write $B_\varepsilon(i, r)$ to mean the closed ball of radius r around i in the space $(E, \varepsilon^{-1}\delta)$. Recall the construction of the space $(\mathbb{S}, d_\mathbb{S}, \mu, 0)$ given in the statement of Theorem 1.3. Again we will only show that

$$(B_\varepsilon(1, 1), \varepsilon^{-1}\delta, 4\varepsilon^{-1}\nu) \rightarrow (B(0, 1), d_\mathbb{S}, \mu)$$

weakly under the compact Gromov–Hausdorff metric as $\varepsilon \rightarrow 0$, where $B(0, 1) \subset (\mathbb{S}, d_\mathbb{S})$ is the closed ball of radius 1 around 0.

We first show how to construct metric measure spaces using excursions. We term this the Evans metric space associated to an excursion due to the similarities of the Evans metric space associated to coalescent processes. We describe this process in generality and then use it to construct the spaces $(B_\varepsilon(1, 1), \varepsilon^{-1}\delta, 4\varepsilon^{-1}\nu)$, $(B(0, 1), d_\mathbb{S}, \mu)$ as well as an auxiliary space.

Constructing Evans metric measure space from an excursion

Let $e = (e(t) : 0 \leq t \leq \zeta(e))$ be an excursion that has height greater than 1 meaning $e : [0, \zeta(e)] \rightarrow [0, \infty)$ is a continuous path such that e hits 1 and further $e(t) = 0$ if and only if $t \in \{0, \zeta(e)\}$. In all the applications we consider, e will be Brownian. In this case there exists a jointly continuous local time $(L(t, x) : t \in [0, \zeta(e)], x \in \mathbb{R})$ with the property that

$$L(t, x) = \lim_{\varepsilon \rightarrow 0} \frac{1}{2\varepsilon} \int_0^t \mathbb{1}_{\{e(s) \in (x-\varepsilon, x+\varepsilon)\}} ds, \quad t \in [0, \zeta(e)], x \in \mathbb{R},$$

see for example [25, Chapter IV] and in particular Corollary 1.8 and Corollary 1.9. We now show how to obtain a metric measure space (S, d, π) from the excursion e . Let $\{\varepsilon_i\}_{i=1}^\infty$ be the positive excursions of the excursion e above level 1 and we order them as follows. Since $t \mapsto L(t, 1)$ is increasing the Stieltjes measure $dL(\cdot, 1)$ exists and we let U_1, U_2, \dots be i.i.d. random variables with the law given by $dL(\cdot, 1)/Z_1$ where $Z_1 = L(\zeta(e), 1)$. For each $k \in \mathbb{N}$, U_k is the local time corresponding to a unique excursion at level 1 (see [25, Chapter IV, Proposition 2.5]) which may be positive or negative. Let U_{k_1}, U_{k_2}, \dots denote the local times corresponding to positive excursions, then $\{\varepsilon_i\}_{i=1}^\infty$ is ordered such that for each $i \in \mathbb{N}$, ε_i is the excursion which starts at local time U_{k_i} .

For $i, j \in \mathbb{N}$ with $i \neq j$ we define $1 - d(i, j)$ to be the first height at which ε_i and ε_j are a part of the same excursion (see Figure 4). In other words let $t(\varepsilon_i)$ and $t(\varepsilon_j)$ denote the start time of the excursions ε_i and ε_j and suppose without a loss of generality that $t(\varepsilon_i) < t(\varepsilon_j)$. Then

$$d(i, j) = 1 - \inf\{e(t) : t(\varepsilon_i) \leq t \leq t(\varepsilon_j)\}. \quad (38)$$

By definition, the space (S, d) is the completion of (\mathbb{N}, d) . We also define a measure π on (S, d) as follows. For each $i \in \mathbb{N}$ and $r \in (0, 1]$, every closed ball $B(i, r) \subset (S, d)$ corresponds to an excursion e of e above level $1 - r$ that hits level 1. We define $\pi(B(i, r))$ to be the total local time $\ell_1(e)$ the excursion e attains at level 1. Note that this defines

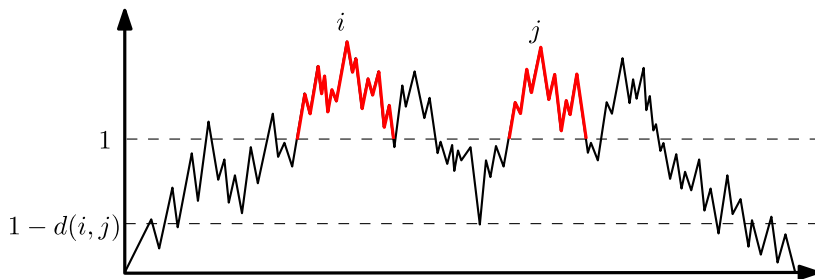


Fig. 4. Visual interpretation of the construction of the metric space (S, d) .

the measure uniquely by Carathéodory's extension theorem and thus we obtain a metric measure space (S, d, π) . We remark that the total mass $\pi(S) = Z_1$ is the total local time spent at level 1 by the excursion e and thus is finite. Lastly in all of our applications the excursion e will have unique local minima which is enough to conclude that (S, d) is compact.

Definition 4.1. *The metric measure space (S, d, π) is called the Evans metric measure space associated to the excursion e .*

Remark 4.2. *For the reader who is familiar with continuum real trees, the Evans metric measure space associated to the excursion e can be constructed as follows. Let (T, d_T, o) be the rooted real tree that is encoded by the excursion e where $o \in T$ is the root of the tree. Then S is the set of points $x \in T$ such that $d_T(o, x) = 1$, and the metric d is given by $d = (1/2)d_T$. Let \tilde{T} be the subtree of T spanned by S and the root o so that S is the set of leaves of \tilde{T} . Then π is the uniform measure on (\tilde{T}, d_T) which is supported on S .*

Constructing (E, δ, ν) from an excursion

Let $X = (X_t : 0 \leq t \leq \zeta(X))$ be a Brownian excursion conditioned to hit level 1. Let $(\tilde{E}, \tilde{\delta}, \tilde{\nu})$ be the Evans metric measure space associated to the excursion X . Then [4, Theorem 1.1] gives a construction of (E, δ, ν) in terms of $(\tilde{E}, \tilde{\delta}, \tilde{\nu})$ which we now describe. For $x \geq 0$ let Z_x be the total local time attained at level $x \geq 0$ by the process X . For $t \in [0, 1]$ define

$$V(t) := \int_{1-t}^1 \frac{4}{Z_v} dv. \quad (39)$$

Then $E = \tilde{E}$ and for $x, y \in E$, $\delta(x, y) = V(\tilde{\delta}(x, y))$. Lastly ν is the renormalisation of $\tilde{\nu}$, that is $\nu(\cdot) = \tilde{\nu}(\cdot)/Z_1$ where Z_1 is the total local time that the excursion X attains at level 1.

Constructing an intermediary space $(\tilde{B}_\varepsilon(1, 1), T_\varepsilon \tilde{\delta}, T_\varepsilon \tilde{\nu})$

For $\varepsilon > 0$ let

$$T_\varepsilon := \frac{4}{\varepsilon Z_{1-\sqrt{\varepsilon}}} \vee \frac{1}{\sqrt{\varepsilon}}. \quad (40)$$

The maximum in the definition is for technical reasons and for small $\varepsilon > 0$ it will be the case that $T_\varepsilon = 4(\varepsilon Z_{1-\sqrt{\varepsilon}})^{-1}$ with high probability. Thus for small $\varepsilon > 0$, it is not hard to see that $V(t\varepsilon) \approx tT_\varepsilon$. We will now construct a space $(\tilde{B}_\varepsilon(1, 1), T_\varepsilon \tilde{\delta}, T_\varepsilon \tilde{\nu})$ which will turn out to be close to $(B_\varepsilon(1, 1), \varepsilon^{-1}\tilde{\delta}, 4\varepsilon^{-1}\tilde{\nu})$ and has the advantage that it is easier to show convergence to the limiting space $(B(0, 1), d_S, \mu)$.

Let $\tilde{B}_\varepsilon(1, 1) \subset (\tilde{E}, \tilde{\delta})$ be the closed ball of radius $1/T_\varepsilon$ around 1 and consider $(\tilde{B}_\varepsilon(1, 1), T_\varepsilon \tilde{\delta}, T_\varepsilon \tilde{\nu})$ as a compact metric measure space in its own right. Thus the space $(\tilde{B}_\varepsilon(1, 1), T_\varepsilon \tilde{\delta}, T_\varepsilon \tilde{\nu})$ has diameter 1. It will be convenient for us to obtain this space as the Evans metric measure space of some excursion. Since $(\tilde{E}, \tilde{\delta}, \tilde{\nu})$ is obtained from a Brownian excursion X conditioned to reach level 1, we can obtain $(\tilde{B}_\varepsilon(1, 1), T_\varepsilon \tilde{\delta}, T_\varepsilon \tilde{\nu})$ by scaling X . To be more precise, define

$$X_\varepsilon(t) := 1 + [X(tT_\varepsilon^{-2}) - 1]T_\varepsilon, \quad 0 \leq t \leq \zeta(X)T_\varepsilon^2. \quad (41)$$

Then $(\tilde{B}_\varepsilon(1, 1), T_\varepsilon \tilde{\delta}, T_\varepsilon \tilde{\nu})$ is the Evans metric measure space obtained using a sub-excursion Y_ε of X above level 0 that hits level 1.

To describe the law of Y_ε fully, let $e_1^{(\varepsilon)}, \dots, e_{M(\varepsilon)}^{(\varepsilon)}$ be the excursions of X_ε above level 0 that reach level 1 where $M(\varepsilon)$ is the total number of such excursions. Though we have obtained the process X_ε by performing diffusive scaling on X , the scaling factor T_ε is random so it is not obvious that $e_1^{(\varepsilon)}, \dots, e_{M(\varepsilon)}^{(\varepsilon)}$ are themselves Brownian excursions conditioned to reach level 1. We will see that this is nevertheless the case in Lemma 4.3. From the construction of the

Table 1

Table of the spaces used in the proof

Evans mm-space	Excursion	Description of excursion
$(\tilde{E}, \tilde{\delta}, \tilde{\nu})$	X	Brownian excursion conditioned to reach level 1
$(\tilde{B}_\varepsilon(1, 1), T_\varepsilon \tilde{\delta}, T_\varepsilon \tilde{\nu})$	Y_ε	Local-time biased pick of excursions of X_ε above level 0 that hit level 1, where $X_\varepsilon(t) := 1 + [X(tT_\varepsilon^{-2}) - 1]T_\varepsilon$
$(B(0, 1), d_S, \mu)$	Y	Excursion of a two-sided Brownian motion above level -1 straddling the origin

Evans metric measure space above we see that given the excursions $e_1^{(\varepsilon)}, \dots, e_{M(\varepsilon)}^{(\varepsilon)}$, the excursion Y_ε is selected by biasing on the total local time obtained at level 1. Precisely, Y_ε can be described by the following

$$\mathbb{P}(Y_\varepsilon = e_i^{(\varepsilon)} | e_1^{(\varepsilon)}, \dots, e_{M(\varepsilon)}^{(\varepsilon)}) = \frac{\ell_1(e_i^{(\varepsilon)})}{\sum_{j=1}^{M(\varepsilon)} \ell_1(e_j^{(\varepsilon)})}, \quad 1 \leq i \leq M(\varepsilon), \quad (42)$$

where $\ell_1(e_i^{(\varepsilon)})$ is the total local time at 1 that the excursion $e_i^{(\varepsilon)}$ accumulates.

Constructing $(B(0, 1), d_S, \mu)$ from an excursion

Finally we show how to construct $(B(0, 1), d_S, \mu)$ directly from an excursion. Let $W = (W_t : t \in \mathbb{R})$ be a two-sided Brownian motion and let $Y = (Y_t : t \in [0, \zeta(Y)])$ be the excursion of W above level -1 straddling the origin. That is, let $\tau_+ = \inf\{t > 0 : W_t = -1\}$ and $\tau_- = \sup\{t < 0 : W_t = -1\}$, then $Y_t = W_{t+\tau_-} + 1$ for $t \leq \tau_+ - \tau_-$. It is not hard to check that the space $(B(0, 1), d_S, \mu)$ can be constructed as the Evans metric measure space associated to the excursion Y .

We sum up the various spaces defined here with Table 1.

Outline of the proof

The proof consists of (1) showing the spaces $(\tilde{B}_\varepsilon(1, 1), T_\varepsilon \tilde{\delta}, T_\varepsilon \tilde{\nu})$ and $(B_\varepsilon(1, 1), \varepsilon^{-1} \delta, 4\varepsilon^{-1} \nu)$ are close and (2) showing $\lim_{\varepsilon \downarrow 0} (\tilde{B}_\varepsilon(1, 1), T_\varepsilon \tilde{\delta}, T_\varepsilon \tilde{\nu}) = (B(0, 1), d_S, \mu)$ in the compact Gromov–Hausdorff–Prokhorov sense. Showing (1) is relatively straight forward and relies on analytical estimates on $V(t)$ as well as controls on Z_X . We show (1) at the end with equation (48).

Showing (2) on the other hand is a little bit more complicated and will take up large portion of the proof. First in Lemma 4.3 we show that the excursions that Y_ε is picked from are actually Brownian excursions conditioned to reach level 1. An important observation is that the excursion Y can be picked from an infinite sequence of i.i.d. Brownian excursions conditioned to reach level 1, where each excursion is weighted by the total local it accumulates at level 1. The difference is that Y_ε is picked from a finite set of excursions, however the number of excursions that Y_ε is picked from goes to ∞ as $\varepsilon \downarrow 0$. We will use this in Lemma 4.5 to give a suitable coupling of Y_ε to Y which will imply $\lim_{\varepsilon \downarrow 0} (\tilde{B}_\varepsilon(1, 1), T_\varepsilon \tilde{\delta}, T_\varepsilon \tilde{\nu}) = (B(0, 1), d_S, \mu)$.

We begin the proof with the following lemma.

Lemma 4.3. *Let $e_1^{(\varepsilon)}, \dots, e_{M(\varepsilon)}^{(\varepsilon)}$ be the excursions of X_ε above level 0 that reach level 1. Then conditionally on $M(\varepsilon)$, $e_1^{(\varepsilon)}, \dots, e_{M(\varepsilon)}^{(\varepsilon)}$ are i.i.d. Brownian excursions conditioned to reach level 1.*

Proof. Fix $\varepsilon > 0$. Observe that the excursions $e_1^{(\varepsilon)}, \dots, e_{M(\varepsilon)}^{(\varepsilon)}$ correspond to excursions of X above level $u := 1 - 1/T_\varepsilon$ that hit level 1. Note that since $T_\varepsilon \geq 1/\sqrt{\varepsilon}$ we have that $u \geq u_0 := 1 - \sqrt{\varepsilon}$. Define the σ -algebra $\mathcal{H} = \sigma(X_{\alpha(s)} : s \geq 0)$ where

$$\alpha(s) := \inf \left\{ t \geq 0 : \int_0^t \mathbb{1}_{\{X(v) \leq u_0\}} dv > s \right\}.$$

In words \mathcal{H} contains all the information about the excursions of X below level $u_0 = 1 - \sqrt{\varepsilon}$. The total local time Z_{u_0} of the process X at level u_0 satisfies (see [25, Chapter VI Corollary (1.9)])

$$Z_{u_0} = \lim_{\eta \rightarrow 0} \frac{1}{\eta} \int_0^{\zeta(X)} \mathbb{1}_{\{X_s \in (u_0 - \eta, u_0)\}} ds.$$

Thus Z_{u_0} is measurable with respect to \mathcal{H} and consequently so is T_ε and $u = 1 - 1/T_\varepsilon$.

It is well known that after hitting level u_0 , the law of X is that of a Brownian motion started at level u_0 , killed the first time it hits 0 and conditioned to reach level 1 before hitting level 0. Itô's description of Brownian motion ([25, Chapter XII Theorem (2.4)]) tells us that conditionally on $Z_{u_0} = z$ the excursions of the process X above level u_0 form a Poisson point process on the local time interval $[0, z]$, conditioned to have at least one excursion of height greater than $\sqrt{\varepsilon}$. Further, the excursions above level $1 - \sqrt{\varepsilon}$ are independent of the excursions below level $1 - \sqrt{\varepsilon}$, and hence independent of the σ -algebra \mathcal{H} .

On the other hand Z_{u_0} and u are measurable with respect to \mathcal{H} . Thus conditionally on \mathcal{H} the excursions of the process X above level $u = 1 - 1/T_\varepsilon$ that hit level 1 are i.i.d. Brownian excursions conditioned to have height greater than $1/T_\varepsilon$. By Brownian scaling and (41) it follows that conditionally on \mathcal{H} and $M(\varepsilon) = m$, $e_1^{(\varepsilon)}, \dots, e_m^{(\varepsilon)}$ are i.i.d. Brownian excursions conditioned to reach level 1. In other words let e be a Brownian excursion conditioned to reach level 1 and ϕ_1, \dots, ϕ_m be continuous bounded functions mapping the set of excursions to \mathbb{R} , then we have just shown that

$$\mathbb{E}[\phi_1(e_1^{(\varepsilon)}) \cdots \phi_m(e_m^{(\varepsilon)}) | \mathcal{H}, M(\varepsilon) = m] = \prod_{i=1}^m \mathbb{E}[\phi_i(e)].$$

Taking expectations conditionally on $M(\varepsilon) = m$ on both sides above finishes the proof. \square

The next lemma shows the convergence of the local time at 1 of Y_ε . Recall that Y is the excursion of two-sided Brownian motion straddling the origin which is used in the construction of the space $(B(0, 1), d_S, \mu)$.

Lemma 4.4. *We have that*

$$\ell_1(Y_\varepsilon) \rightarrow \ell_1(Y)$$

in distribution as $\varepsilon \rightarrow 0$, where $\ell_1(\cdot)$ denotes the total local time attained by the excursion at level 1.

Proof. Let $e_1^{(\varepsilon)}, \dots, e_m^{(\varepsilon)}$ be the excursions of X_ε above level 0 that reach level 1. Condition on $M(\varepsilon) = m$. Then from Lemma 4.3 it follows that $e_1^{(\varepsilon)}, \dots, e_m^{(\varepsilon)}$ are i.i.d. Brownian excursions conditioned to reach level 1. Thus it follows that $\ell_1(e_1^{(\varepsilon)}), \dots, \ell_1(e_m^{(\varepsilon)})$ are i.i.d. exponential random variables with parameter 1/2 (see [25, Chapter VI Proposition (4.6)]). Let E_1, E_2, \dots be i.i.d. exponential random variables with parameter 1/2, then by (42) it follows that

$$\mathbb{P}(\ell_1(Y_\varepsilon) \in \cdot | M(\varepsilon) = m) = \mathbb{E} \left[\frac{m E_1}{\sum_{i=1}^m E_i} \mathbb{1}_{\{E_1 \in \cdot\}} \right].$$

On the other hand $\ell_1(Y)$ has the same law as $E_1 + E_2$, the sum of two independent exponential random variables. This is a sized biased exponential random variable and by bounded convergence and the law of large numbers we have

$$\mathbb{P}(\ell_1(Y) \in \cdot) = \lim_{n \rightarrow \infty} \mathbb{E} \left[\frac{n E_1}{\sum_{i=1}^n E_i} \mathbb{1}_{\{E_1 \in \cdot\}} \right].$$

The lemma now follows from the fact that $M(\varepsilon) \rightarrow \infty$ in probability as $\varepsilon \rightarrow 0$. \square

From Lemma 4.3 for Y_ε and the definition of Y we can deduce the following. Conditionally on $\ell_1(Y_\varepsilon) = \ell_\varepsilon$ and $\ell_1(Y) = \ell$, the excursions Y_ε and Y are both Brownian excursions conditioned on attaining total local time ℓ_ε and ℓ ,

respectively, at level 1. Thus at this point it is immediate that $Y_\varepsilon \rightarrow Y$ in the Skorokhod sense as $\varepsilon \rightarrow 0$. Unfortunately, this is not enough to show the weak convergence of the space $(\tilde{B}_\varepsilon(1, 1), T_\varepsilon \tilde{\delta}, T_\varepsilon \tilde{\nu})$, under the Gromov–Hausdorff–Prokhorov metric, to the space $(B(0, 1), d_{\mathbb{S}}, \mu)$. Instead, in the next lemma we construct a coupling between Y_ε and Y under which the paths of the two processes agree up to a time. This will in turn enable us to show that the spaces $(\tilde{B}_\varepsilon(1, 1), T_\varepsilon \tilde{\delta}, T_\varepsilon \tilde{\nu})$ and $(B(0, 1), d_{\mathbb{S}}, \mu)$ are close to each other.

Lemma 4.5. *We have that*

$$(\tilde{B}_\varepsilon(1, 1), T_\varepsilon \tilde{\delta}, T_\varepsilon \tilde{\nu}) \rightarrow (B(0, 1), d_{\mathbb{S}}, \mu)$$

weakly as $\varepsilon \rightarrow 0$ under the Gromov–Hausdorff–Prokhorov topology.

Proof. We first present a coupling between Y_ε and Y . By Lemma 4.4 and Skorokhod representation theorem we can suppose that $\ell_1(Y_\varepsilon)$ and $\ell_1(Y)$ are coupled such that $\ell_1(Y_\varepsilon) \rightarrow \ell_1(Y)$ almost surely as $\varepsilon \rightarrow 0$. Fix $\varepsilon > 0$ and condition on $\ell_1(Y_\varepsilon) = \ell_\varepsilon$ and $\ell_1(Y) = \ell$. Suppose further that $\ell_\varepsilon \leq \ell$ (the other case is similar). Note that both Y_ε and Y are Brownian excursions conditioned to have ℓ_ε and ℓ total local time at level 1. Hence we can couple Y_ε and Y such that they have the same path until their local time at level 1 reaches ℓ_ε .

Excursions of the process Y_ε above level 1 correspond to the dense subset \mathbb{N} of the space $(\tilde{B}_\varepsilon(1, 1), T_\varepsilon \tilde{\delta})$. Similarly for Y and $(B(0, 1), d_{\mathbb{S}})$. Thus the coupling of the processes Y_ε and Y gives us a coupling of the spaces such that $(\tilde{B}_\varepsilon(1, 1), T_\varepsilon \tilde{\delta}) \subset (B(0, 1), d_{\mathbb{S}})$. Under this coupling of the spaces it is immediate that

$$d_{\text{Pr}}(T_\varepsilon \tilde{\nu}, \mu) = |\ell - \ell_\varepsilon|, \quad (43)$$

where d_{Pr} denotes the Prokhorov distance.

Fix $\eta \in (0, 1)$ and recall that $d_H((\tilde{B}_\varepsilon(1, 1), T_\varepsilon \tilde{\delta}), (B(0, 1), d_{\mathbb{S}})) < \eta$ if and only if the η -enlargement of the space $(\tilde{B}_\varepsilon(1, 1), T_\varepsilon \tilde{\delta})$ contains $(B(0, 1), d_{\mathbb{S}})$. Clearly it suffices to show the latter condition for the dense subspaces of $(\tilde{B}_\varepsilon(1, 1), T_\varepsilon \tilde{\delta})$ and $(B(0, 1), d_{\mathbb{S}})$ corresponding to the excursions above level 1 of the processes Y_ε and Y respectively. Consider the excursions of Y above level 1 which appear before local time ℓ_ε . Call these excursions matched; these are also excursions of Y_ε above level 1. Then $d_H((\tilde{B}_\varepsilon(1, 1), T_\varepsilon \tilde{\delta}), (B(0, 1), d_{\mathbb{S}})) < \eta$ if and only if every excursion of Y above level 1 – η that hits level 1 contains a matched excursion. This is the same as the event that every excursion of Y below 1 with local time in the interval $[\ell_\varepsilon, \ell)$ has infimum greater than $-\eta$. Thus from standard excursion theory (see [25, Chapter XII Exercise (2.10)]) we have that

$$\mathbb{P}(d_H((\tilde{B}_{T_\varepsilon}(1, 1), T_\varepsilon \tilde{\delta}), (\tilde{B}(0, 1), d_{\mathbb{S}})) \leq \eta) = e^{-|\ell - \ell_\varepsilon|/\eta}. \quad (44)$$

The equations (43) and (44) hold by the same argument when $\ell < \ell_\varepsilon$. Hence in conclusion we have constructed a coupling where

$$\begin{aligned} & \mathbb{P}(d_H((\tilde{B}_\varepsilon(1, 1), T_\varepsilon \tilde{\delta}), (B(0, 1), d_{\mathbb{S}})) + d_{\text{Pr}}(T_\varepsilon \tilde{\nu}, \mu) > 2\eta) \\ & \leq \mathbb{P}(d_H((\tilde{B}_\varepsilon(1, 1), T_\varepsilon \tilde{\delta}), (B(0, 1), d_{\mathbb{S}})) > \eta) + \mathbb{P}(d_{\text{Pr}}(T_\varepsilon \tilde{\nu}, \mu) > \eta) \\ & = 1 - \mathbb{E}\left[\exp\left(-\frac{|\ell_1(Y_\varepsilon) - \ell_1(Y)|}{\eta}\right)\right] + \mathbb{P}(|\ell_1(Y_\varepsilon) - \ell_1(Y)| > \eta). \end{aligned}$$

Taking the limit as $\varepsilon \rightarrow 0$ above and using bounded convergence finishes the proof. \square

Using the previous lemma we can now prove Theorem 1.3.

Proof of Theorem 1.3. For $t \in [0, 1]$ let $U(t)$ denote the inverse of $V(t)$ in (39), that is

$$U(t) = V^{-1}(t) = \inf\left\{s > 0 : \int_{1-s}^1 \frac{4}{Z_v} dv > t\right\}.$$

Fix $\eta \in (0, 1)$ and define

$$A_\varepsilon^\eta := \{(1 - \eta)Z_{1-\sqrt{\varepsilon}} \leq Z_s \leq (1 + \eta)Z_{1-\sqrt{\varepsilon}} \ \forall s \in [1 - U(\varepsilon), 1]\},$$

$$B_\varepsilon := \left\{ T_\varepsilon = \frac{4}{\varepsilon Z_{1-\sqrt{\varepsilon}}} \right\} = \{Z_{1-\sqrt{\varepsilon}} \leq \varepsilon^{-1/2}\},$$

$$\mathcal{E}_\varepsilon^\eta := A_\varepsilon^\eta \cap B_\varepsilon.$$

We claim that $\mathbb{P}(\mathcal{E}_\varepsilon^\eta) \rightarrow 1$ as $\varepsilon \rightarrow 0$. Indeed, $x \mapsto Z_x$ is uniformly continuous (this follows from the Ray–Knight theorems, see [25, Chapter XI Theorem (2.2)]). Further it is elementary to check that $\lim_{\varepsilon \rightarrow 0} U(\varepsilon) = 0$. This shows that $\mathbb{P}(A_\varepsilon^\eta) \rightarrow 1$ as $\varepsilon \rightarrow 0$. The convergence of $\mathbb{P}(B_\varepsilon)$ follows from the fact that $Z_{1-\sqrt{\varepsilon}}$ is distributed exponentially with parameter $1/(2 - 2\sqrt{\varepsilon})$ (see [25, Chapter VI, Proposition (4.6)]). Thus $\mathbb{P}(\mathcal{E}_\varepsilon^\eta) \rightarrow 1$ as $\varepsilon \rightarrow 0$.

On the event $\mathcal{E}_\varepsilon^\eta$ we have that for each $t \leq \varepsilon$,

$$\frac{U(t)\varepsilon T_\varepsilon}{1 + \eta} = \frac{4U(t)}{(1 + \eta)Z_{1-\sqrt{\varepsilon}}} \leq t \leq \frac{4U(t)}{(1 - \eta)Z_{1-\sqrt{\varepsilon}}} = \frac{U(t)\varepsilon T_\varepsilon}{1 - \eta}. \quad (45)$$

Recall that $E = \tilde{E}$ and that for each $x, y \in E$, $\tilde{\delta}(x, y) = U(\delta(x, y))$. Hence from (45) it follows that on the event $\mathcal{E}_\varepsilon^\eta$ for any $x, y \in E$ such that $\delta(x, y) \leq \varepsilon$ we have

$$\frac{1}{1 + \eta} T_\varepsilon \tilde{\delta}(x, y) \leq \varepsilon^{-1} \delta(x, y) \leq \frac{1}{1 - \eta} T_\varepsilon \tilde{\delta}(x, y). \quad (46)$$

A brief computation shows that on the event $\mathcal{E}_\varepsilon^\eta$, for each $x \in E$ and $r \in (0, 1]$

$$4\varepsilon^{-1} \nu(B_\varepsilon(x, r)) = \frac{T_\varepsilon Z_{1-\sqrt{\varepsilon}}}{Z_1} \tilde{\nu}(\tilde{B}_\varepsilon(x, T_\varepsilon U(r)\varepsilon)),$$

where $\tilde{B}_\varepsilon(p, r) \subset (\tilde{E}, T_\varepsilon \tilde{\delta})$ is the closed ball of radius $r > 0$ around p . Thus it follows that on the event $\mathcal{E}_\varepsilon^\eta$ for each $x \in E$,

$$\frac{1}{1 + \eta} T_\varepsilon \tilde{\nu}(\tilde{B}_\varepsilon(x, (1 - \eta)r)) \leq 4\varepsilon^{-1} \nu(\tilde{B}_\varepsilon(x, r)) \leq \frac{1}{1 - \eta} T_\varepsilon \tilde{\nu}(\tilde{B}_\varepsilon(x, (1 + \eta)r)). \quad (47)$$

Using the fact that $\mathbb{P}(\mathcal{E}_\varepsilon^\eta) \rightarrow 1$ as $\varepsilon \rightarrow 0$, Lemma 4.5 and that $\eta > 0$ is arbitrary, an easy pinching argument using (46) and (47) shows that

$$\lim_{\varepsilon \rightarrow 0} d_{\text{GHP}}((B_\varepsilon(1, 1), \varepsilon^{-1} \delta, 4\varepsilon^{-1} \nu), (\tilde{B}_\varepsilon(1, 1), T_\varepsilon \tilde{\delta}, T_\varepsilon \tilde{\nu})) = 0 \quad (48)$$

almost surely. The theorem now follows from Lemma 4.5. \square

Acknowledgements

I would like to thank my supervisor Nathanaël Berestycki for introducing me to the problem and his continuous assistance throughout. My thanks to James Norris and Jean Bertoin who kindly pointed out a mistake in a previous version of this paper and the two anonymous referees who helped to improve on the paper substantially.

This work was supported by the UK Engineering and Physical Sciences Research Council (EPSRC) grant EP/H023348/1 for the University of Cambridge Centre for Doctoral Training, the Cambridge Centre for Analysis.

References

- [1] R. Abraham, J.-F. Delmas and P. Hoscheit. A note on the Gromov–Hausdorff–Prokhorov distance between (locally) compact metric measure spaces. *Electron. J. Probab.* **18** (2013) no. 14, 21. [MR3035742](#)

- [2] D. Aldous. Stopping times and tightness. *Ann. Probab.* **6** (2) (1978) 335–340. [MR0474446](#)
- [3] D. J. Aldous. Deterministic and stochastic models for coalescence (aggregation and coagulation): A review of the mean-field theory for probabilists. *Bernoulli* **5** (1) (1999) 3–48. [MR1673235](#)
- [4] J. Berestycki and N. Berestycki. Kingman’s coalescent and Brownian motion. *ALEA Lat. Am. J. Probab. Math. Stat.* **6** (2009) 239–259. [MR2534485](#)
- [5] J. Berestycki, N. Berestycki and V. Limic. The Λ -coalescent speed of coming down from infinity. *Ann. Probab.* **38** (1) (2010) 207–233. [MR2599198](#)
- [6] J. Berestycki, N. Berestycki and V. Limic. A small-time coupling between Λ -coalescents and branching processes. *Ann. Appl. Probab.* **24** (2) (2014) 449–475. [MR3178488](#)
- [7] J. Berestycki, N. Berestycki and J. Schweinsberg. Small-time behavior of beta coalescents. *Ann. Inst. Henri Poincaré Probab. Stat.* **44** (2) (2008) 214–238. [MR2446321](#)
- [8] N. Berestycki. *Recent Progress in Coalescent Theory. Ensaios Matemáticos [Mathematical Surveys]* **16**. Sociedade Brasileira de Matemática, Rio de Janeiro, 2009. [MR2574323](#)
- [9] J. Bertoin. *Random Fragmentation and Coagulation Processes. Cambridge Studies in Advanced Mathematics* **102**. Cambridge University Press, Cambridge, 2006. [MR2253162](#)
- [10] J. Bertoin and J.-F. Le Gall. Stochastic flows associated to coalescent processes. III. Limit theorems. *Illinois J. Math.* **50** (1–4) (2006) 147–181 (electronic). [MR2247827](#)
- [11] D. Burago, Y. Burago and S. Ivanov. *A Course in Metric Geometry. Graduate Studies in Mathematics* **33**. American Mathematical Society, Providence, RI, 2001. [MR1835418](#)
- [12] B. Chauvin and A. Rouault. KPP equation and supercritical branching Brownian motion in the subcritical speed area. Application to spatial trees. *Probab. Theory Related Fields* **80** (2) (1988) 299–314. [MR0968823](#)
- [13] N. Curien and J.-F. Le Gall. The Brownian plane. *J. Theoret. Probab.* **27** (2014) 1249–1291. [MR3278940](#)
- [14] P. Donnelly and T. G. Kurtz. Particle representations for measure-valued population models. *Ann. Probab.* **27** (1) (1999) 166–205. [MR1681126](#)
- [15] S. N. Ethier and T. G. Kurtz. *Markov Processes. Wiley Series in Probability and Mathematical Statistics: Probability and Mathematical Statistics*. Wiley, New York, 1986. [MR0838085](#)
- [16] S. N. Evans. Kingman’s coalescent as a random metric space. In *Stochastic Models (Ottawa, ON, 1998)* 105–114. *CMS Conf. Proc.* **26**. Amer. Math. Soc., Providence, RI, 2000. [MR1765005](#)
- [17] S. N. Evans. *Probability and Real Trees. Lecture Notes in Mathematics* **1920**. Springer, Berlin, 2008. Lectures from the 35th Summer School on Probability Theory held in Saint-Flour, July 6–23, 2005. [MR2351587](#)
- [18] W. Feller. *An Introduction to Probability Theory and Its Applications. Vol. II*, 2nd edition. Wiley, New York, 1971. [MR0270403](#)
- [19] C. Goldschmidt and B. Haas. Behavior near the extinction time in self-similar fragmentations. I. The stable case. *Ann. Inst. Henri Poincaré Probab. Stat.* **46** (2) (2010) 338–368. [MR2667702](#)
- [20] C. Goldschmidt and B. Haas. Behavior near the extinction time in self-similar fragmentations. II. Finite dislocation measures. ArXiv preprint, 2013. Available at [arXiv:1309.5816](#).
- [21] B. Haas. Fragmentation processes with an initial mass converging to infinity. *J. Theoret. Probab.* **20** (4) (2007) 721–758. [MR2359053](#)
- [22] B. Hughes. Trees and ultrametric spaces: A categorical equivalence. *Adv. Math.* **189** (1) (2004) 148–191. [MR2093482](#)
- [23] J. F. C. Kingman. The coalescent. *Stochastic Process. Appl.* **13** (3) (1982) 235–248. [MR0671034](#)
- [24] J. Pitman. Coalescents with multiple collisions. *Ann. Probab.* **27** (4) (1999) 1870–1902. [MR1742892](#)
- [25] D. Revuz and M. Yor. *Continuous Martingales and Brownian Motion*, 3rd edition. *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]* **293**. Springer, Berlin, 1999. [MR1725357](#)
- [26] S. Sagitov. The general coalescent with asynchronous mergers of ancestral lines. *J. Appl. Probab.* **36** (4) (1999) 1116–1125. [MR1742154](#)
- [27] L. Simon. *Lectures on Geometric Measure Theory. Proceedings of the Centre for Mathematical Analysis, Australian National University* **3**. Australian National University Centre for Mathematical Analysis, Canberra, 1983. [MR0756417](#)