Global divergence of spatial coalescents

Omer Angel · Nathanaël Berestycki · Vlada Limic

Abstract We study several fundamental properties of a class of stochastic processes called spatial Λ-coalescents. In these models, a number of particles perform independent random walks on some underlying graph G. In addition, particles on the same vertex merge randomly according to a given coalescing mechanism. A remarkable property of mean-field coalescent processes is that they may come down from infinity, meaning that, starting with an infinite number of particles, only a finite number remains after any positive amount of time, almost surely. We show here however that, in the spatial setting, on any infinite and bounded-degree graph, the total number of particles will always remain infinite at all times, almost surely. Moreover, if $G = \mathbb{Z}^d$, and the coalescing mechanism is Kingman's coalescent, then starting with N particles at the origin, the total number of particles remaining is of order $(\log^* N)^d$ at any fixed positive time (where $\log^*$ is the inverse tower function). At sufficiently large times the total number of particles is of order $(\log^* N)^{d-2}$, when $d > 2$. We provide parallel results in the recurrent case $d = 2$. The spatial Beta-coalescents behave similarly, where $\log \log N$ is replacing $\log^* N$.

O. Angel
Department of Mathematics, University of British Columbia, Vancouver, BC V6T 1Z2, Canada
e-mail: angel@math.ubc.ca

N. Berestycki (✉)
Statistical Laboratory, DPMMS, University of Cambridge,
Wilberforce Rd., Cambridge CB3 0WB, UK
e-mail: n.berestycki@statslab.cam.ac.uk

V. Limic
UMR 6632, LATP, CMI, CNRS-Université de Provence, Technopôle de Château-Gombert,
39, rue F. Joliot Curie, 13453 Marseille cedex 13, France
e-mail: vlada@cmi.univ-mrs.fr

Mathematics Subject Classification (2000)   60K35 · 92D10

Contents

1 Introduction

1.1 Motivation and main results

The theory of stochastic coalescent processes has expanded considerably in the last decade, as a consequence of their deep connections to population genetics, spin glass models and polymers. In theoretical population genetics, coalescents arise as natural models of merging of ancestral lineages (see, for example, [9,18]). A particular Λ-coalescent, usually called the Bolthausen–Sznitman coalescent, is thought to be an important object for describing the conjectured universal ultrametric structure of numerous mean-field spin glass models including the Sherrington-Kirkpatrick model (see [12,13,38]). The same coalescent has also been recently linked in [15] to scaling limits of directed polymers.

The Λ-coalescents are stochastic processes taking values in $\mathcal{P}$, the space of partitions of $\mathbb{N} = \{1, 2, \ldots\}$. In the current context, each class (or block) in $\Pi \in \mathcal{P}$ can and will be thought of as a particle. For any coalescent process $(\Pi_t, t \geq 0)$ it is true that Π_{t+s} is a coarsening of Π_t, for any $t, s > 0$. There is a natural semi-group structure $(\mathcal{P}, \star)$, where $\Pi \star \Pi'$ is the result of merging the blocks of Π "according to" the partition Π'. The Λ-coalescents can be canonically characterized as the Lévy processes in $(\mathcal{P}, \star)$, with the property that no two coagulation events occur simultaneously. The above Lévy property corresponds to the fact that $\Pi_{t+s} = \Pi_t \star \Pi'_s$, for all $t, s \geq 0$, where $(\Pi'_t, t \geq 0)$ is an independent, identically distributed process (see, e.g. [6, Sect. 3.1.3] for details). A direct construction of Λ-coalescents from [34] is now considered standard in the probability literature. As will be discussed in more detail below, each Λ-coalescent corresponds uniquely to a finite measures Λ on [0, 1]: for instance, the case where $\Lambda = \delta_{\{0\}}$ gives the well-known *Kingman* coalescent from mathematical population genetics, while the uniform measure on [0, 1] gives the Bolthausen–Sznitman coalescent. These two processes, as well as the more general Beta-coalescents are especially interesting and amenable for analysis, since they are self-similar in a certain sense made precise by the results of [4].

The present work is devoted to the study of several fundamental properties of a more general class of models, introduced in [32] and called the spatial Λ-coalescents. In this setting, particles (i.e., partition classes) are positioned on some underlying

locally finite graph $G = (V, E)$. The dynamics of the process is enriched in the presence of the geographical structure in two ways. First, the particles move as independent continuous time simple random walks on G. Second, stochastically independent Λ-coalescence takes place on each site of G. More precisely, at any given time, only particles that are on a same site can coagulate. Moreover, at every site the coalescence mechanism is that of the original (mean-field) Λ-coalescent. The spatial Kingman coalescent is a natural model of an interacting particle system where particles perform independent random walks on an underlying graph, and any pair of particles coalesce at rate 1, as long as they are located at the same site.

The spatial Λ-coalescent processes are particularly well suited to model merging of ancestral lineages for a population that is evolving in a geographical space G, where the spatial motion of individuals is taken into account. In this way, geographical factors such as isolation and overpopulation can influence the dynamics, making it a more realistic model for long-term population behavior. In the above interpretation, the vertices of the graph are referred to as *demes* and represent a discretization of physical space. Each edge represents potential migratory routes between two adjacent demes.

While the mean-field Λ-coalescent processes are relatively well understood at this point, even basic properties of their spatial counterparts are much more delicate to analyze. Intuitively, the difficulty comes from the fact that the two ingredients in the dynamics, the coalescence and the migration, affect the particles in the system in opposite directions: the spatial motion makes particles diffuse away from one another, and the coalescence keeps them together. Indeed, our results show that the competition between these two forces can be very tight. Our main results provide information about the limiting behavior of spatial Λ-coalescents as the initial number of particles n tends to infinity, at both small and large time-scales. We consider the case where initially all the particles are located at the origin o of G. For some of our results, the only assumptions on G are that it is connected and has bounded degree $\Delta = \max_{v \in V} \text{degree}(v) < \infty$. However, several of our more precise results on the asymptotic behavior are restricted to the setting where G is the d-dimensional lattice $\mathbb{Z}^d$.

Define the function $\log^* n$ as the inverse $\log^* n := \inf\{m \geq 1 : \text{Tow}(m, 1) \geq n\}$ of the tower function, where $\text{Tow}(0, x) = x$ and

$$\text{Tow}(n, x) = e^{\text{Tow}(n-1,x)} = \underbrace{e^{e^{\cdot^{\cdot^{\cdot^{e^x}}}}}}_{n \text{ iterations}}. \tag{1}$$

Theorem 1.1 *Fix $\varepsilon > 0$, and consider the spatial Kingman coalescent on a graph G with bounded degrees. Start with n particles located at $o \in G$, and let $N^n(t)$ be the total number of particles at time $t > 0$. There are constants $C, c > 0$ depending only on t and the degree bound such that*

$$\mathbb{P}\left(N^n(t) \geq c \, \text{Vol} \, B(o, (1 - \varepsilon) \log^* n)\right) \xrightarrow[n \to \infty]{} 1, \quad \text{and}$$

$$\mathbb{P}\left(N^n(t) \leq C \, \text{Vol} \, B(o, (1 + \varepsilon) \log^* n)\right) \xrightarrow[n \to \infty]{} 1.$$

A more concise way of stating Theorem 1.1 is that $N^n(t) \asymp \text{Vol } B(o, (1+o(1)) \log^* n)$ with high probability (see the paragraph on "Other notations" at the end of Sect. 1.3 for $\asymp$ and other notations related to asymptotic behavior).

Remark The function $\log^* n$ tends to infinity with n, but at a very slow rate: $\log^* n \leq 4$ for $n \leq 10^{1656520}$. Thus while this function diverges to infinity from the mathematically rigorous point of view, for all practical sample sizes it takes value 3 or 4.

Remark The behavior in Theorem 1.1 contrasts that of the mean-field case, where $N^n(t)$ converges (without renormalization) to a finite random variable $N(t)$ for all $t > 0$, due to well-known properties of Kingman coalescent. In the lattice case $G = \mathbb{Z}^d$, we see that $N^n(t)$ diverges as $(\log^* n)^d$, i.e. extremely slowly. Even on a regular tree, where balls have maximal volume given the degree, $N^n(t)$ diverges only as $e^{C \log^* n}$.

The mean-field Λ-coalescent processes can be classified according to the *coming down from infinity* (CDI) property. For the partition-valued process $(\Pi_t, t \geq 0)$, this means that the initial configuration $\Pi_0 = \{\{i\} : i \in \mathbb{N}\}$ is countably infinite, but that Π_t contains only finitely many classes at any time $t > 0$, almost surely. The Kingman and the Beta-coalescents with parameter in a certain range (see below) come down from infinity, while the Bolthausen–Sznitman coalescent does not. If the spatial coalescent is viewed as an interacting particle system, then CDI is the property that the total number of particles in the system at any fixed positive time remains bounded (tight), as the initial number of particles tends to ∞. It turns out that for the mean-field model (as well as the spatial model with G a finite graph), either the total number of particles in the system converges almost surely to a finite random variable, or it diverges at any given time.

A natural question, and one of the motivations of this work, is whether a similar dichotomy occurs in spatial coalescents. It is known that if the mean-field Λ-coalescent comes down from infinity, then the number of particles in its spatial counterpart will be locally finite (see Sect. 1.3). It is natural to ask whether the total number of particles can nevertheless be infinite on infinite graphs. We answer this question for general spatial Λ-coalescents. It is remarkable that the answer is universal, in that it does not depend on the driving measure Λ nor on the geometry of the underlying infinite graph G.

Theorem 1.2 *For any measure Λ on (0,1) and any infinite graph G, consider the spatial Λ-coalescent on G started with n particles at $o \in G$. If $N^n(t)$ denotes the total number of particles at time t, then $N^n(t) \to \infty$ almost surely, as $n \to \infty$.*

In particular, for Λ such that the mean-field coalescent comes down from infinity, the number of particles will be locally finite, but globally infinite. For this reason we call this phenomenon the *global divergence* of the spatial Λ-coalescent.

Our next result on the fixed time asymptotics of N^n concerns a setting that is particularly relevant for some biological applications, where the coalescence mechanism is given by the Beta-coalescent with parameter $\alpha \in (1, 2)$ (as defined in the next section).

Theorem 1.3 *Fix $\alpha \in (1, 2)$, and consider the spatial Beta$(2 - \alpha, \alpha)$ coalescent on a graph G with bounded degree. Start with n particles located at $o \in G$, and let $N^n(t)$*

be the total number of particles at time t. There are constants $C, c > 0$ depending only on t, α and the degree bound such that

$$\mathbb{P}\left(N^n(t) \geq c \operatorname{Vol} B(o, c \log \log n)\right) \xrightarrow[n \to \infty]{} 1, \quad \text{and}$$

$$\mathbb{P}\left(N^n(t) \leq C \operatorname{Vol} B(o, C \log \log n)\right) \xrightarrow[n \to \infty]{} 1.$$

Similarly to Theorem 1.1, this may be stated more concisely as $N_n(t) = \Theta(\operatorname{Vol} B(o, \Theta(\log \log n)))$. When the graph has growth $\operatorname{Vol} B(o, R) \asymp R^d$, this translates to $N^n(t) \asymp (\log \log n)^d$.

All the previous theorems describe the state of the system at a fixed time t. We also provide estimates for the number of particles that survive for a long time. Here the diffusion of particles plays a more important role, hence the results depend in a more fundamental way on the underlying graph. We focus on Euclidean lattices $G = \mathbb{Z}^d$, $d \geq 2$. Since by Theorem 1.1, particles are quickly spread on a region of size roughly $m = \log^* n$, it is natural to look at a time scale of order m^2 in order to see this diffusion.

Theorem 1.4 *Assume that the coalescence mechanism is Kingman's coalescent. Let $G = \mathbb{Z}^d$, let $m = \log^* n$, and fix $\delta > 0$. Then there exist some constants $c > 0$ and $C > 0$ (depending only on d, δ) such that, if $d > 2$,*

$$\mathbb{P}\left(cm^{d-2} < N^n(\delta m^2) < Cm^{d-2}\right) \xrightarrow[n \to \infty]{} 1,$$

while, if $d = 2$, then

$$\mathbb{P}\left(c \log m < N^n(\delta m^2) < C \log m\right) \xrightarrow[n \to \infty]{} 1.$$

If the coalescence mechanism is a Beta-coalescent with parameter $\alpha \in (1, 2)$, then the same statement holds with $m = \log \log n$.

One interpretation of this theorem is that, when the underlying graph G is $\mathbb{Z}^d$, the resulting random particle system may also be thought of as a microscopic description of the small-time evolution of a solution to the parabolic nonlinear partial differential equation:

$$\partial_t u = \frac{1}{2}\Delta u - \beta u^2, \quad (\beta > 0), \tag{2}$$

starting from a singular initial condition, such as a Dirac delta measure at a given spatial location. We refer the reader to [22, 23] for a discussion of this equation, and to the end of Sect. 1.2 for intuition about the relation between (2) and our model.

As suggested by Theorem 1.4, for the study of the long-term particle system behavior it is natural to rescale the particle system's time by a factor of m^2, while rescaling space by a factor of m. Theorem 1.4 indicates that the rescaled system should exhibit a Boltzmann-Grad limiting behavior, i.e., the number of interactions (intersections and

coalescences) between a particle and all others over any finite time interval is tight and does not tend to 0 as $n \to \infty$. The behavior of the rescaled system mirrors the system of Brownian coagulating particles studied in [22] and [23], in which the PDE (2) is derived as the governing macroscopic behavior, and is obtained as a particular case of the Smoluchowski system of PDEs. It is worth pointing out that in the current case, the discrete structure of the lattice remains important in determining the frequency of coalescence events even after space and time have been rescaled, which would alter the formula fixing the reaction coefficient in the limiting PDE.

Remark Greven, Limic and Winter [20, Theorem 3] have shown that when $d = 2$, $(N^n(m^{2\delta}), \delta \geq 1)$ (i.e. the process seen at an exponential time scale) converges in distribution to a time-changed mean-field Kingman coalescent, as $n \to \infty$. Thus the diffusivity of the particles is observed at time scale m^2, but the recurrence of particles starts to affect the behaviour in a significant way only at later time scales.

Remark The log* function featured in Theorem 1.1 might remind the reader of a result of Kesten [24,25], who studied the number of allelic types in a Wright-Fisher model with small mutation probability.

In Kesten's model, allelic types take values in $\mathbb{Z}$, and the type of an offspring is identical to that of its parent, except on a mutation event of a small probability (inversely proportional to the total population size). When a mutation occurs, the offspring's type is chosen by adding an independent $\mathbb{Z}$-valued random variable (with some given, bounded, distribution) to the parent's type, that is, by making a random walk step from the parent's type. It turns out that the number of types (and, in fact, their relative positions in space) has an equilibrium distribution. It is shown in [24,25] that the number of observed types at equilibrium is of order log* n, where n is the sample size. The above Fisher-Wright model may seem closely related to the one-dimensional spatial Kingman coalescent, but on a closer look one realizes that the dynamics of the two models are quite different, and there is no direct relation between the results.

Kesten's result may be phrased as follows. Let T_n be the tree generated by Kingman's (non-spatial) coalescent started with n particles, and consider a branching random walk indexed by T_n. Then the number of distinct values at the leaves is of order log* n. A variation of the strategy used in Sect. 3 applies in this setting, and can lead to an alternate proof of Kesten's result.

1.2 Heuristics and proof ideas

It is evident from Theorems 1.1 and 1.3 that the long term behavior of the number of particles in the spatial coalescent depends delicately on the precise nature of the coalescent. We now describe the approximate behavior of the spatial coalescent started with a large number of particles, all located at o. The proofs are mostly a detailed treatment of the following heuristic observations.

To understand the finite initial condition, we turn to the infinite one. Consider a given Λ-coalescent which comes down from infinity (see below). Let N_t be the number of particles in the (non-spatial) coalescent started with $N_0 = \infty$. For Kingman's coalescent it is the case that $N_t \sim 2/t$, whereas for Beta-coalescents with parameters

$(2 - \alpha, \alpha)$ with $1 < \alpha < 2$, we have $N_t \sim c_\alpha t^{1-\alpha}$ [4,5]. The rough description that follows applies to both of these, as well as more general coalescents. In general, one would expect N_t to be concentrated (for small t) around some function $g(t)$ (such a function is found in [3]). The coalescent started with N particles is similar to the infinite coalescent observed from time $g^{-1}(N)$ onward.

Consider now the non-spatial coalescent with emigration, where each particle also disappears at some rate ρ. In fact, the parameter ρ may depend on the size of the population, as long as $n\rho(n)$ is non-decreasing. It turns out that for coalescents that come down from infinity, the emigration does not influence N_t so much, and N_t is still close to $g(t)$. The total number of particles that emigrate when starting with N particles is then close to a Poisson variable with mean

$$f(N) := \int_{g^{-1}(N)} g(t)\rho(g(t))dt. \tag{3}$$

(The upper bound of integration is some arbitrary constant.)

Now comes the key observation: if N is large, the number of particles migrating back into o is negligible (under a technical condition that holds for most spatial coalescents), and in fact, an overwhelming proportion of those particles that emigrate will have emigrated by time $g^{-1}(f(N))$. Thus we find that at this time, the number of particles at o and each of its neighbors is of order $f(N)$. A second observation is that the resulting populations can be approximated by independent spatial coalescents, when observed from time $g^{-1}(f(N))$ onward. In particular, at time $g^{-1}(f \circ f(N))$ there are of the order of $f \circ f(N)$ particles at each vertex in $B(o, 2)$. This "cascading onto neighbors" continues until step m, where m is such that $f \circ \cdots \circ f(N)$ (m repeated iterations of f) is of order 1. Note that in these m steps a ball of radius m has been roughly filled.

Applying this heuristics to the case of Kingman's coalescent and the Beta-coalescents with parameters $(2 - \alpha, \alpha)$ and $1 < \alpha < 2$, gives the following. For Kingman's coalescent and constant ρ, we have $f(n) \sim 2\rho \log n$, and for Beta-coalescents we have $f(n) \sim C_\alpha \rho n^{2-\alpha}$, for some constant $C_\alpha > 0$. Thus in the first case, $m = \log^* N$. In the second case, we find $m \sim c \log \log N$. In general, this gives $m \sim f^*(n)$, where

$$f^*(n) = \inf \left\{ m \geq 1 : \underbrace{f \circ \cdots \circ f(n)}_{m \text{ iterations}} \leq 1 \right\}. \tag{4}$$

Note that if $\rho(n)$ decreases fast enough so that $f(n)$ is bounded, then it follows from this heuristic analysis that the spatial coalescent will come down from infinity globally. However, when ρ is constant, it can be proved that f is always unbounded, which in turn implies the result about global divergence of any spatial Λ-coalescent.

Turning to the long time asymptotics, by the above reasoning we may start from a configuration consisting of a tight number of particles at each site of the ball of radius m around the origin. Since the number of particles per site is tight, the coalescent dynamics influences the evolution less than the diffusion. In particular, the structure

of the underlying graph becomes important for the asymptotic behavior of the process. For simplicity, let us restrict ourselves to d-dimensional Euclidean lattices with $d \geq 3$. Let $\rho(t)$ denote the average number of particles per site in the ball of radius m at time t. Then at time $t_0 = 1$ we have $\rho(t_0) \asymp 1$ and $\lim_{t \to \infty} \rho(t) = 0$. Each particle present in the configuration at time t coalesces with another particle at an average rate approximately $\rho(t)$, so that $\frac{d}{dt} N(t) = -N(t)\rho(t)/2$. Dividing by the volume of the ball, one arrives to the ODE

$$\frac{d}{dt}\rho(t) = -\frac{1}{2}\rho(t)^2, \tag{5}$$

whose solution is given by $\rho(t) = 2/(t + c)$ for some $c > 0$.

The approximation (5) should be valid as long as the diffusion of particles away from the initial region (i.e., $B(o, m)$) is negligible. The influence of diffusion should start to be visible at times of order m^2. In particular, at time m^2, the density $\rho(m^2)$ is of order m^{-2}, so the total number of remaining particles is of order m^{d-2}. Assuming the plausible claim that the remaining particles are approximately uniformly distributed over a ball of radius order m, a simple calculation (using hitting probabilities for random walks) now implies that each of them has a positive probability of never meeting any other particle again, and so the number of particles that survive indefinitely is of order m^{d-2}.

We wish to point out that van den Berg and Kesten [7,8] have shown a density decay similar to (5) for a related model of coalescing random walks. However their results differ in two ways. On the one hand, the coalescence mechanism which they analyze is different. On the other hand, and more importantly, their initial condition is initially homogeneous in space, and not restricted to a large ball. This restriction is the cause of much of the difficulty in the current setting—see Sect. 7 for more details.

1.3 Definitions and background on spatial coalescents

Kingman's coalescent. Suppose that we are given an integer $n \geq 1$. *Kingman's n-coalescent* is the Markov process $(\Pi_t^n, t \geq 0)$, with values in the set $\mathcal{P}_n$ of partitions of $[n] := \{1, \ldots, n\}$, such that $\Pi_0^n = \{\{1\}, \{2\}, \ldots, \{n\}\}$, and such that each pair of blocks merges at rate 1, and these are the only transitions of the process. Blocks of the partition Π_t^n may be viewed as indistinguishable particles, and we often refer to the number of blocks of Π_t^n as the number of particles alive at time t. A simple but essential property of Kingman's n-coalescent is the so-called *sampling consistency* property: the restriction of $(\Pi_t^{n+1}, t \geq 0)$ to $[n]$ has the same distribution as an n-coalescent. This enables one to construct a Markov process $(\Pi_t, t \geq 0)$ with state space $\mathcal{P}$, the set of partitions of $\mathbb{N}$, such that the law of Π when restricted to $[n]$ equals the law of Π^n. In particular, the initial state of this process is the trivial partition $\Pi_0 = \{\{1\}, \{2\}, \ldots\}$. The process Π is called *Kingman's coalescent*. For background reading, see for instance [6,18,35].

Λ-coalescents. Let Λ be a finite measure on [0, 1]. A *coalescent with multiple collisions*, or *Λ-coalescent*, is a Markov process $(\Pi_t, t \geq 0)$ with values in the set of

partitions of $\mathbb{N}$ characterized by the following properties. If $n \in \mathbb{N}$, then the restriction of $(\Pi_t, t \geq 0)$ to $[n]$ is a Markov chain $(\Pi_t^{(n)}, t \geq 0)$, where $\Pi_0^n = \{\{1\}, \{2\}, \ldots, \{n\}\}$, and where the only possible transitions are mergers of blocks (it is possible to merge several blocks simultaneously into one block, but no two mergers of this kind can occur simultaneously) so that whenever the current configuration consists of b blocks, any given k-tuple of blocks merges at rate

$$\lambda_{b,k} = \int_{[0,1]} x^{k-2}(1-x)^{b-k} \Lambda(dx). \tag{6}$$

Note that 0^0 is interpreted as 1, so that an atom of Λ at 0 causes each pair of particles to coalesce at a finite positive rate $\Lambda(\{0\})$. In this way any Λ-coalescent can be thought of as a superposition of a "pure" coalescent with multiple collisions driven by measure $\Lambda(dx)\mathbf{1}_{(0,1]}(x)$, and a time-changed Kingman's coalescent. An atom of Λ at 1 causes all the particles to coalesce at some positive fixed rate. Such Λ-coalescent may be viewed as a killed Λ'-coalescent where $\Lambda'(dx) = \Lambda(dx)\mathbf{1}_{[0,1)}(x)$. Kingman's coalescent is a particular Λ-coalescent, obtained when the measure Λ equals δ_0, the unit Dirac mass at 0. Any Λ-coalescent Π is sampling consistent, that is, if $m < n$ then the restriction of Π^n to $[m]$ is equal in law to Π^m. It is this observation that allows one to construct an infinite version of the process. It is interesting to note the following fact shown by Pitman [34]: Λ-coalescents are the only exchangeable Markov coalescent processes without simultaneous collisions. We refer the reader to [34] for definitions and further properties.

As already mentioned, if $\Lambda(dx) = dx\mathbf{1}_{[0,1]}$, the corresponding Λ-coalescent is usually called the Bolthausen–Sznitman coalescent, and more generally if Λ is the Beta$(2 - \alpha, \alpha)$ distribution where $\alpha \in (0, 2)$ is a fixed parameter, that is,

$$\Lambda(dx) = \frac{1}{\Gamma(2-\alpha)\Gamma(\alpha)} x^{1-\alpha}(1-x)^{\alpha-1}\, dx, \tag{7}$$

the corresponding Λ-coalescent is called *Beta-coalescent* with parameter α. The Bolthausen–Sznitman coalescent is the special case $\alpha = 1$, and it does not come down from infinity. For $\alpha \in (1, 2)$, the corresponding Beta-coalescents come down from infinity, and they are important processes from the theoretical evolutionary biology perspective, due to the following result from [36]: the Beta-coalescent with parameter $\alpha \in (1, 2)$ arises in the scaling limit of population models where the offspring distribution of a typical individual is in the domain of attraction of a stable law with index α. Apart from the Kingman coalescent, the Beta-coalescents with parameter $\alpha \in (1, 2)$ are the most-studied class of Λ-coalescents (see, e.g., [4,5,11]).

Spatial coalescents. As informally described above, spatial coalescents are processes which combine spatial motion of individual particles with coalescence of particles located on the same site of a given graph of bounded degree. Let Λ be a given finite measure on $[0, 1]$. A spatial Λ-coalescent, as defined in [32], is a Markov process $(\Pi_t^\ell, t \geq 0)$ with values in the space $\mathcal{P}^\ell = \mathcal{P} \times V^{\{1,2,\ldots\}}$ of partitions of $\{1, 2, \ldots\}$

indexed by spatial locations. That is, an element $x = (\pi, \ell) \in \mathcal{P}^\ell$ consists of a partition $\pi = \{A_1, A_2, \ldots\}$, and a sequence $\ell = (\ell_1, \ell_2, \ldots)$, where ℓ_i specifies the location of the block A_i. There are only two types of transitions possible for $\Pi_t^\ell = (\Pi_t, \ell_t)$: (i) provided there are b blocks at a location $v \in V$, then any given k-tuple of them will merge at rate $\lambda_{b,k}$ given by (6), independently over v; and (ii) independently of the coalescent mechanism, each block A_k of π migrates at rate θ. This means that if the block is at v, then some vertex w is chosen according to the distribution $p(v, \cdot)$, where $p(v, w)$ is a given Markov kernel. When this happens, ℓ_k is changed from v to w. To simplify the discussion, we will assume unless otherwise specified, that $p(x, y)$ is the transition kernel for the simple random walk on the underlying graph G.

If π is a partition let $i \sim_\pi j$ mean that the particles labeled i and j belong to the same block of π. For $(\pi, \ell) \in \mathcal{P}^\ell$ and $v \in V$, denote by $\#_v(\pi, \ell)$ the number of blocks in π with label (location) v.

Spatial Λ-coalescents inherit the sampling consistency directly from Λ-coalescents. Namely, if we consider a spatial coalescent started from $n + 1$ particles (that is, blocks) and consider its restriction to the first n particles, the new process has the law of a spatial coalescent started from n particles. This simple property will be used on several occasions. In particular, it implies that if (π^1, ℓ^1) and (π^2, ℓ^2) are such that $\#_v(\pi^1, \ell^1) \leq \#_v(\pi^2, \ell^2)$, for all v, then there exists a coupling of two spatial coalescents $((\Pi_t^1, \ell_t^1), (\Pi_t^2, \ell_t^2)), t \geq 0)$ such that $(\Pi_0^i, \ell_0^i) = (\pi^i, \ell^i), i = 1, 2$ and $\#_v(\Pi_t^1, \ell_t^1) \leq \#_v(\Pi_t^2, \ell_t^2)$ for all v, almost surely. The same property guarantees the existence of spatial coalescents started with infinitely many particles on an infinite graph (see Theorem 1 in [32] for a particular construction).

Spatial Λ-coalescents may be started from configurations containing countably infinitely many particles at each site of G, see [32]. However, our main results concern spatial Λ-coalescents started from the following initial condition:

$$\Pi_0^\ell = (\{\{1\}, \{2\}, \ldots\}, (o, o, \ldots)), \tag{8}$$

where o is some given reference vertex called the *origin* of G. In words, all the infinitely many particles are initially located at the origin o.

From now on we abbreviate

$$X_v(t) = \#_v(\Pi_t, \ell_t) \qquad \text{and} \qquad X_v^n(t) = \#_v(\Pi_t^n, \ell_t). \tag{9}$$

We denote the *total number of blocks* by $N^*(t) = \sum_{v \in V} X_v(t)$ (resp. $N^n(t) = \sum_{v \in V} X_v^n(t)$). When not in risk of confusion, we will drop the superscript n to simplify notations. It is clear from the definitions that $(\{X_v(t)\}_{v \in G}, t \geq 0)$ has Markovian transitions, with respect to the filtration generated by the coalescent process Π. All these processes carry only partial information about the evolution of the corresponding spatial coalescent, in particular, they do not determine the evolving partition structure.

In the language of theoretical population biology, a sample of n individuals is selected from the population at the present time, and their ancestral lineages are followed in reversed time. The transition rules (i)–(ii) given above reflect the idea that individuals typically reproduce within their own colony (so that only lineages on the same site may coalesce), and occasionally there is a rare migration event, which corre-

sponds to the random walk transitions. In the case where the coalescence mechanism is simply Kingman's coalescent, we note that this model may be viewed as the ancestral partition process associated with Kimura's *stepping-stone model* [26,27].

Previous works on spatial processes of coalescence include the papers of Cox and Griffeath [17] who study instantaneous coalescence on $\mathbb{Z}^2$; Cox and Durrett [16] and Zähle, Cox and Durrett [39], who study Kingman's coalescent on a large but finite two-dimensional torus and with a fixed sample size; Limic and Sturm [32] who introduce the spatial Λ-coalescent and study asymptotics on a torus with a fixed sample size; and Greven, Limic and Winter [20] who study Kingman's coalescent with a large number of particles on a torus on an exponential (rather than linear) time scale. An informal survey of some of these results may be found in Chapter 5 of [6].

Coming down from infinity. Let $(\Pi_t, t \geq 0)$ be Kingman's coalescent. As already mentioned, Kingman [28,29] realized that while Π starts with an infinite number of blocks at $t = 0$, its number of blocks becomes finite for all $t > 0$, almost surely. A coalescent with multiple collisions may or may not have the same property, depending on the measure Λ. More precisely, there are only two possibilities as shown in [34]: let E (resp. F) denote the event that for all $t > 0$ there are infinitely (resp. finitely) many blocks. Then, if $\Lambda(\{1\}) = 0$, either $P(E) = 1$ or $P(F) = 1$. When $P(F) = 1$, the process Π is said to *come down from infinity*. For instance, a Beta-coalescent comes down from infinity if and only if $1 < \alpha < 2$, henceforth we make this an assumption whenever working with Beta-coalescents.

In the context of spatial coalescents, assuming that $\Lambda(\{1\}) = 0$, Proposition 11 in [32] implies that when the initial number of particles is infinite, then $X_v(t)$ becomes finite for all $v \in V$ and $t > 0$ with probability 1, if and only if the underlying measure Λ is such that the mean-field (i.e., non-spatial) Λ-coalescent comes down from infinity. In this situation, we may say that the spatial coalescent comes down from infinity *locally*. Naturally, this stays true if the initial condition is (8).

Other notations. Unless specified otherwise, c, C (and variations c_1, C_2, ...) will henceforth denote positive constants that depend only on the underlying graph, and that may change from line to line. Typically, c, c_1, ... denote sufficiently small, whereas C, C_1, ... denote sufficiently large constants. We also use the symbols $a_n \sim b_n$ and $a_n \asymp b_n$ to denote respectively that $a_n/b_n \to 1$, and a_n/b_n is bounded away from 0 and ∞, as $n \to \infty$.

Organization of the paper. The rest of the paper is organized as follows. Section 2 starts with some preliminary remarks and observations concerning large deviation estimates for Kingman's coalescent and Ewens's sampling formula, as well as several couplings between the spatial Λ-coalescents and the corresponding (mean-field) Λ-coalescents, which will be used throughout the paper. Section 3 contains a proof of Theorem 1.1 on the behavior of the spatial Kingman coalescent in finite time. As many of the subsequent results in the paper build on this, we recommend reading this section prior to any of the following sections. Section 4 contains the proof of Theorem 1.3 on the finite-time behavior of the spatial Beta-coalescents. Section 5 returns to the general case of spatial Λ-coalescents and arbitrary graphs with bounded degree, and

contains the proof of global divergence (Theorem 1.2). In final Sects. 6 and 7 we study respectively the lower bound and the upper bound for long term behavior of Kingman's coalescent (as stated in Theorem 1.4). The lower bound obtained in Sect. 6.2 is true for general Λ-coalescents, but we provide in Sect. 6.3 an alternate shorter proof for the special case of Kingman's coalescent, that also gives tighter bounds. The proof of the upper bound in Sect. 7 turns out to be the most technical part of the paper, and it is based on a delicate multi-scale analysis.

Sections 4–6 may be read in any order, depending on the interest of the reader. We recommend reading Sect. 6 prior to Sect. 7.

2 Preliminary lemmas

2.1 Some large deviation estimates

We begin with an easy Chernoff type bound for a sum of exponential random variables, which we prefer to state in an abstract form now so as to refer to it on several occasions later. In our applications, $\mathbb{E}S$ will typically be small.

Lemma 2.1 *Let $\{E_i\}_{i\in I}$ be independent exponential random variables with $\mathbb{E}E_i = \mu_i$. Let $S = \sum_{i\in I} E_i$. Then for any $0 < \varepsilon < 1$*

$$\mathbb{P}\left(S < (1-\varepsilon)\mathbb{E}S\right) \leq \exp\left(-\frac{\varepsilon^2(\mathbb{E}S)^2}{4\,\mathrm{Var}\,S}\right).$$

Additionally, for $0 < \varepsilon < \frac{\mathrm{Var}\,S}{\mathbb{E}S\,\sup\{\mu_i\}}$,

$$\mathbb{P}\left(S > (1+\varepsilon)\mathbb{E}S\right) \leq \exp\left(-\frac{\varepsilon^2(\mathbb{E}S)^2}{4\,\mathrm{Var}\,S}\right).$$

Remark If $I = \{n, n+1, \dots\}$ and $\mu_i \sim ci^{-\alpha}$ for some $\alpha > 1$, then as $n \to \infty$, $\frac{\mathrm{Var}\,S}{\mathbb{E}S\,\sup\{\mu_i\}}$ is bounded away from 0, hence the second bound holds for all $\varepsilon > 0$ small enough, for all n.

Proof Using Markov's inequality, for any $0 < \lambda \leq \frac{1}{2}\inf\{\mu_i^{-1}\}$

$$
\begin{aligned}
\mathbb{P}\left(S > (1+\varepsilon)\mathbb{E}S\right) &\leq e^{-\lambda(1+\varepsilon)\mathbb{E}S}\mathbb{E}e^{\lambda S}\\
&= e^{-\lambda(1+\varepsilon)\mathbb{E}S}\prod \frac{1}{1-\lambda\mu_i}\\
&< e^{-\lambda(1+\varepsilon)\mathbb{E}S}\exp\left(\sum \lambda\mu_i + \lambda^2\mu_i^2\right)\\
&= e^{-\lambda\varepsilon\mathbb{E}S + \lambda^2\,\mathrm{Var}\,S},
\end{aligned}
$$

where we have used that for $x \in (0, 1/2)$ we have $-\ln(1-x) < x + x^2$. Taking $\lambda = \frac{\varepsilon\mathbb{E}S}{2\,\mathrm{Var}\,S}$, which is allowed since $\varepsilon < \frac{\mathrm{Var}\,S}{\mathbb{E}S\,\sup\{\mu_i\}}$, yields the upper bound.

The lower bound follows from a similar argument with $\lambda = -\frac{\varepsilon\mathbb{E}S}{2\,\mathrm{Var}\,S}$. □

We now apply this to get a large deviation estimate for Kingman's coalescent. This uses a simple idea which can already be found in Aldous [1], who used it to prove a central limit theorem for the number of particles at time t. Denote by $\mathbb{P}^n$ the law of the (non-spatial) Kingman coalescent started with n blocks. Let $N(t)$ be the number of blocks at time t.

Lemma 2.2 *Let* $t = t(n) \to 0$ *in such a way that* $t(n)^{-1} = o(n)$. *For any* $0 < \varepsilon < 1/4$, *for* n *large enough,*

$$\mathbb{P}^n \left(1 - \varepsilon < \frac{N(t)}{2/t} < 1 + \varepsilon \right) > 1 - \exp \left(-\frac{\varepsilon^2}{t} \right).$$

Proof For the upper bound, let $m = \lceil (1 + \varepsilon)2/t \rceil$. The time it takes the process to get from n to m particles is a sum of independent exponential random variables with means $\binom{k}{2}^{-1}$ for $k = m + 1, \ldots, n$. Call this sum S. Note that $N(t) > m$ if and only if $S > t$. We have

$$\mathbb{E}S = \sum_{k=m+1}^{n} \binom{k}{2}^{-1} \sim 2m^{-1} \sim t/(1 + \varepsilon)$$

provided $m = o(n)$. Similarly,

$$\text{Var } S = \sum_{k=m+1}^{n} \binom{k}{2}^{-2} \sim (4/3)m^{-3}.$$

Thus, for $\varepsilon < 1/3 + o(1)$,

$$\mathbb{P}(S > t) < \exp \left(-\frac{(3 + o(1))\varepsilon^2}{2t} \right),$$

by Lemma 2.1. The lower bound is similar using the upper bound on S.   □

We now consider Kingman's coalescent with spatial migration. Let $\mathbb{P}^n$ be the law of a simplified process where n particles initially located at a single site o coalesce according to Kingman's dynamics, while each particle (or block of particles) migrates at rate ρ, and any block that migrates away from o is ignored from that time onwards. Denote by Z_n the total number of blocks that ever migrate away from o.

One can think of each migration event as of a "unique mutation on the genealogical tree", by giving it for example the label equal to its occurrence time. Since migrations happen at rate ρ for each block present in the configuration at site o, one quickly realizes that Z_n is a realization from a well-known distribution arising in mathematical population genetics. Namely, set $\theta = 2\rho$, and suppose that on the (non-spatial) Kingman coalescent tree mutation marks occur at a Poisson rate of $\theta/2$ per unit length. Using the language of mathematical population genetics, assume the infinite alleles models (all mutations create a different allele, and so different individuals in the

Fig. 1 The random partition
generated by mutations
(*squares*). Here $Z_5 = 4$
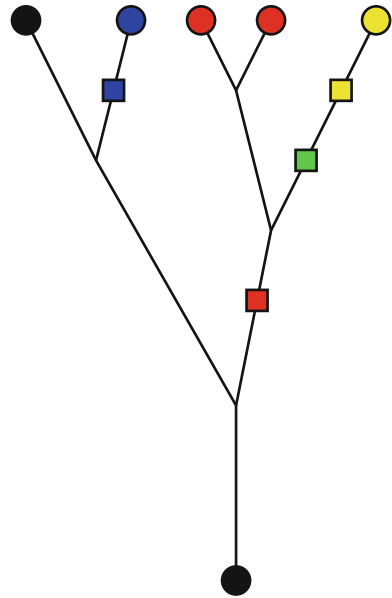
original sample of n are in the same family if and only if they descend from the same
mutation and there has been no other mutation between this common ancestor and the
present individuals). The marks of the mutation process generate a random partition
Π_θ on the leaves of the tree by declaring that i and j are in the same block of Π_θ if
and only if there is no mutation mark on the shortest path that connects i and j. In
Figure 1 different blocks of this partition are represented by different colors. Then it
is easy to see that Z_n has the law of the number of blocks in Π_θ. It is well-known (see,
e.g., (3.24) in Pitman [35]) that Z_n is of order $\theta \log n$ for large n. The following large
deviation estimate is part of the folklore, but we could not find a precise reference for
it in the literature.

Lemma 2.3 *Fix $\varepsilon > 0$. There are $c, C > 0$ such that*

$$\mathbb{P}^n \left(\left| \frac{Z_n}{\log n} - \theta \right| > \varepsilon \right) < Cn^{-c}.$$

Furthermore, for any U,

$$\mathbb{P}^n(Z_n > U) < Cn^C e^{-U}.$$

The proof is based on the Chinese restaurant process representation of Ewens's
sampling formula. Let $K_{n,i}$ be the number of blocks of size i in Π_θ, where $i =
1, \ldots, n$. Then the distribution of $(K_{n,1}, \ldots, K_{n,n})$ is given by Ewens's sampling
formula $ESF(\theta)$:

$$P(K_{n,i} = a_i, \ i = 1, \ldots, n) = \frac{n!}{\theta(\theta+1)\cdots(\theta+n-1)} \prod_{i=1}^{n} \frac{\theta^{a_i}}{i^{a_i} a_i!}, \qquad (10)$$

for any given collection $a_1, \ldots, a_n$ of non-negative integers such that $\sum_{i=1}^{n} i a_i = n$.

The Chinese restaurant process representation of (10) (see [35, Sect. 3.1]), states that the number of blocks in Π_θ satisfies

$$Z_n \stackrel{d}{=} \sum_{i=1}^{n} \zeta_i \qquad (11)$$

where ζ_i are independent Bernoulli random variables with mean

$$P(\zeta_i = 1) = \frac{\theta}{i+\theta}.$$

Proof of Lemma 2.3. By (11) we have for $\lambda > 0$

$$\mathbb{E}e^{-\lambda Z_n} = \prod_{i \le n} \left(1 - \frac{\theta}{i+\theta}(1 - e^{-\lambda}) \right)$$

$$< \exp\left(\sum_{i \le n} -\frac{\theta}{i+\theta}(1 - e^{-\lambda}) \right)$$

$$< \exp\left(-\theta(1 - e^{-\lambda})(C + \log n) \right).$$

By Markov's inequality

$$\mathbb{P}(Z_n < (1-\varepsilon)\theta \log n) < \exp\left(\lambda(1-\varepsilon)\theta \log n - \theta(1 - e^{-\lambda})(C + \log n) \right)$$

$$< \exp\left((-\lambda\varepsilon\theta + O(\lambda^2)) \log n + C \right)$$

For small positive λ the coefficient of $\log n$ is strictly negative.

Similarly, for $\lambda > 0$

$$\mathbb{E}e^{\lambda Z_n} = \prod_{i=1}^{n} \left(1 + \frac{\theta}{i+\theta}(e^\lambda - 1) \right)$$

$$< \exp\left(\sum_{i=1}^{n} \frac{\theta}{i}(e^\lambda - 1) \right)$$

$$< \exp\left(\theta(e^\lambda - 1)(C + \log n) \right).$$

Thus, by Markov's inequality,

$$\mathbb{P}(Z_n > U) < \exp\left(\theta(e^\lambda - 1)(C + \log n) - \lambda U \right) \qquad (12)$$

Taking $U = (1 + \varepsilon)\theta \log n$ and $\lambda = \lambda(\varepsilon, \theta)$ small enough gives the first upper bound. Taking $\lambda = 1$ gives the second claim.   □

A similar computation can be found in Greven et al. [20, Lemma 3.3].

2.2 Coupling and comparison

The final general tool we use is a coupling of the spatial process with simpler coalescent processes such as non-spatial ones. While it is usual in coalescent theory to keep track of the entire partition structure as time evolves, here we are only interested in the number of particles remaining in the system at any given time, so accounting for the full partition data is cumbersome.

However, the detailed *Poisson process construction* (see e.g., [32, Theorem 1]) becomes useful in the context of coupling. Its advantage comes from the following fact: all the information on both the merging and the migration is given by the Poisson clocks (ringing for jumps and for mergers), hence one builds the spatial coalescent process path (keeping track of the particle labels, and of the partition structure) by applying a deterministic function ω-by-ω to this data. Therefore, it is straightforward to append another deterministic ingredient (as the "coloring procedure" in the following lemma) to the construction.

Henceforth, it is convenient to consider the following simpler variation, where the partition structure is ignored. Label the n initial particles by $1, \ldots, n$, and let $x_1, \ldots, x_n$ be their initial locations. Let $S^1, \ldots, S^n$ be n i.i.d. simple random walks on G in continuous time with jump rate ρ started at $x_1, \ldots, x_n$, respectively. To each k-tuple of labels $i_1 < \cdots < i_k$, and each $b \geq k$, corresponds an independent Poisson process $M^b_{i_1, \ldots, i_k}$ with intensity $\lambda_{b,k}$. The particles labeled $i_1 < \cdots < i_k$ coalesce at a jump time t of $M^b_{i_1, \ldots, i_k}$ if and only if they are all located at the same site v at time t^-, and there are a total of b particles at v. In this case, the newly created particle inherits the minimal label i_1. (Subsequently there are no particles with labels $i_2, \ldots, i_k$.) In particular, its trajectory starting from time t will be $(S^{i_1}_s, s \geq t)$.

Suppose that the n initial particles are grouped into classes according to a partition $\pi = (B_1, \ldots, B_r)$ of $\{1, \ldots, n\}$, where $r \geq 1$. Informally, we give a color to each particle initially, and B_i is the set of particles with color i. We wish to compare the system $X = \{X_v(t), t \geq 0\}_{v \in V}$ to the one which consists only of the particles that belong to a particular class B of π. More precisely, for each $B \in \pi$, denote by X^B the spatial coalescent process whose initial configuration contains only the particles from B. Denote by $N^*(t)$ the total number of particles of $X(t)$ and by $N^B(t)$ the total number of particles of $X^B(t)$.

Lemma 2.4 *There is a coupling $(X, X^{B_1}, \ldots, X^{B_r})$ such that, almost surely,*

$$\forall v \in V \quad X^B_v(t) \leq X_v(t) \leq \sum_{i=1}^{r} X^{B_i}_v(t),$$

for each block B of π, and hence

$$\mathbb{N}^B(t) \leq N^*(t) \leq \sum_{i=1}^{r} N^{B_i}(t).$$

Note that in the coupling given below, the processes X^{B_i} are not independent. In fact, under weak assumptions on the coalescent and when all the blocks are "small", there is a coupling including independence, see Lemma 7.2.

Proof Fix a realization of the process $(X(t), t \geq 0)$ as described above. Note that at any time t, each particle in the current configuration can be identified with a set of particles from the original configuration, that have been merging (possibly in several steps) to form this particle (this is the partition-valued realization of the spatial coalescent). In the rest of the argument, we say that a particle intersects $B \in \pi$ if its corresponding set intersects B.

For each $i = 1, \ldots, r$, and $t \geq 0$, let $X^{B_i}(t)$ be the configuration obtained from $X(t)$ by restricting to only those particles that intersect B_i. The consistency property of Λ-coalescents implies that for each i the law of X^{B_i} is that of the spatial coalescent started from the initial configuration restricted to elements of B_i. Thus this is a coupling of the processes.

In this construction we have $X_v^{B_i}(t) \leq X_v(t)$, since X_v may contain particles that do not intersect B_i. Moreover, any particle contributing to $X_v(t)$ intersects at least one B_i, giving the bound $X_v(t) \leq \sum_i X_v^{B_i}(t)$. The inequalities relating $N^*(t)$ and $N^{B_i}(t)$ are an immediate consequence. □

A second type of coupling we will need is between a spatial Λ-coalescent and its mean-field (i.e., non-spatial) counterpart. Fix a vertex $u \in V$ of the graph, and consider a spatial Λ-coalescent $\{X_v(t), t \geq 0\}_{v \in V}$ started with a finite number of particles and such that initially $X_u(0) = n$. Let $M(t)$ denote the number of particles on u at time t that have always stayed at u, and let $Z(t)$ denote the number of particles that jumped out of u prior to time t. In parallel, let $(N(t), t \geq 0)$ denote the number of particles at time t in a mean-field Λ-coalescent started with n particles.

Lemma 2.5 *There exists a coupling of X and N such that:*

$$M(t) \leq N(t) \leq M(t) + Z(t), \qquad \text{a.s. for all } t \geq 0, \tag{13}$$

and

$$N(t) - Z(t) \leq X_u(t) \leq N(t) + Z(t), \quad \text{a.s. for all } t \geq 0. \tag{14}$$

Proof The process $M(t)$ may be realized as a mean-field coalescent where, in addition, particles are killed at rate ρ. In that case, if we let $Z(t)$ denote the total number of particles that have been killed, we see immediately that on the one hand, $M(t) \leq N(t)$, and on the other hand, $N(t) = M(t) + \bar{Z}(t)$ where $\bar{Z}(t) \leq Z(t)$. Indeed, $M(t) + Z(t)$ counts the number of particles if we freeze particle instead of killing them. However, in $N(t)$

these particles keep coalescing, and so the difference $\bar{Z}(t) = N(t) - M(t) \leq Z(t)$. This proves (13). For (14), note first that $X_u(t) \geq M(t) = N(t) - \bar{Z}(t) \geq N(t) - Z(t)$. Finally, the last inequality in (14) is obtained by observing that $X_u(t)$ is made of particles that never jumped out of u (there are $M(t)$ such particles) and of particles that have jumped out of u and have come back at some time later, potentially coalescing in the meantime. There can never be more than $Z(t)$ such particles, since this is the total number of particles that jump out of u.   □

In fact, one can be slightly more precise than the above estimate by using the following observation. Define two processes

$$S(t) = Z(t) - \int_0^t \rho M(s) ds \quad V(t) = S(t)^2 - \int_0^t \rho M(s) ds. \tag{15}$$

Let $\mathbb{P}^n$ denote the law of the above coupling (thus $N(0) = n$ and $X_u(0) = n$ and $X_v(0) = 0$ if $v \neq u$), and let $\mathcal{F}$ denote the filtration that it generates. It is a standard (and easy) fact that both processes are continuous time martingales under $\mathbb{P}^n$ with respect to $\mathcal{F}$. In fact, if we define $\mathcal{G} = \sigma\{N(u), u \geq 0\}$ to be the σ-algebra generated by N, and let $\mathcal{F}_t^* = \sigma\{\mathcal{G}, \mathcal{F}_t\}$, then the processes $S(\cdot)$ and $V(\cdot)$ are continuous-time martingales with respect to the filtration $\mathcal{F}^*$.

Lemma 2.6 *For each time interval $[a, b]$, we have the stochastic domination*

$$\mathbb{P}(Z(b) - Z(a) \geq x | \mathcal{G}) \leq \mathbb{P}\left(\text{Poisson}\left(\rho \int_a^b N(s)\, ds\right) \geq x \,\middle|\, \mathcal{G}\right).$$

Proof Given $\mathcal{G}$, Z is a pure jump process with jumps of size 1 that arrive at rate $\rho M(t) \leq \rho N(t)$ at time t, almost surely.   □

Finally, a global comparison with mean-field coalescents can be obtained in the case of the spatial Kingman coalescent as follows (see also [21, Sect. 6.1]). Let S be an arbitrary subset of vertices and consider the restriction of X to S.

Lemma 2.7 *Fix a time $\tau \leq 2$, and vertex set S, and assume that all particles are in S at time 0. Let $Z = Z(\tau)$ be the number of distinct particles that exit S by time τ, and let $N_S^*(t)$ be the number of particles in S at time t. Then, for some $c = c(\varepsilon) > 0$,*

$$\mathbb{P}\left(N_S^*(\tau) > Z + \frac{(4 + \varepsilon)|S|}{\tau}\right) < e^{-c|S|/\tau}.$$

Note that the bound is independent of the starting configuration. This lemma is a precursor to Lemma 7.5.

Proof Let Q_t be the number of particles in S that have survived until time t but have not left S. We have then that $N_S^*(t) \leq Z + Q_t$. The rate of coalescence inside S at time t is

$$\sum_{v \in S} \binom{X_v(t)}{2} \geq |S| \binom{Q_t/|S|}{2}$$

(by Jensen's inequality for $\binom{x}{2}$.) If $Q_t < 2|S|$ for some $t \leq \tau$ then we are done (since $\tau \leq 2$.) Otherwise, $|S| \cdot \binom{Q_t/|S|}{2} \geq \frac{1}{2|S|}\binom{Q_t}{2}$, and so Q_t is stochastically dominated by the block counting process of a Kingman coalescent slowed down by a factor of $2|S|$. Lemma 2.2 completes the proof. $\qquad\qquad\square$

3 Finite time behaviour of the spatial Kingman coalescent

3.1 An induction

The first step of the argument is to show that for some m (close to $\log^* n$) there are no particles outside $B(o, m)$ at some specified time, and to provide lower and upper bounds (both polynomial in the volume of the ball) on the number of particles at each site inside the ball at the same time, on an event of high probability. This can be done for any n, but it is easier to consider initially a sub-sequence of n's, and then interpolate to get the result for all n. With this in mind, for a given integer m denote $V_m = \text{Vol } B(o, m)$, and define $n = n(m) := \text{Tow}(m, V_m^2)$. Note that $\log^* n = m + \log^* V_m^2$ is very close to m, as $m \to \infty$.

Define the sequence of times $t_k = (\text{Tow}(m - k, V_m^2))^{-3}$, where $k = 0, \ldots, m$. This sequence is increasing from $t_0 = n^{-3}$ to $t_m = V_m^{-6}$. Moreover, $t_{k+1} \gg t_k$ (in particular $t_{k+1} > 2t_k$). Recall that $\theta = 2\rho$, where ρ is the jump rate of particles, and that Δ is the maximal degree in the graph. Define the events B_k by

$$B_k = \{X_v(t_k) = 0 \text{ for all } v \notin B(o, k)\} \bigcap \Big\{X_v(t_k)$$
$$\in \Big[\min(1, \tfrac{\rho}{\Delta})t_k^{-1/3}, 4t_k^{-1}\Big] \text{ for all } v \in B(o, k)\Big\}. \qquad (16)$$

We are particularly interested in the event B_m which states that at time t_m each site of $B(o, m)$ has between cV_m^2 and CV_m^6 particles, with no remaining particles outside $B(o, m)$.

Lemma 3.1 *With the above notations, $\mathbb{P}(B_m) \to 1$ as $m \to \infty$.*

The idea is to prove a bound on $\mathbb{P}(B_k^c)$ by induction on k. For $k = 0$ we have $\mathbb{P}^n(B_0) \geq 1 - n^{-1}$, since the probability of a pair of particles coalescing by time $t_0 = n^{-3}$ is at most $\binom{n}{2}t_0$, and the probability of a particle jumping by that time is at most $t_0 n$. The key to the induction step is the following

Lemma 3.2 *Fix constants $a_0, a_1, \varepsilon > 0$ with $\varepsilon < 1/2$. Consider the coalescent started with n particles, all located at $u \in V$, that is, $X_v(0) = n\delta_u(v)$. Let $\tau = a(\log n)^{-3}$ for some $a \in [a_0, a_1]$, and define the event*

$$A = \cap_v \{X_v(\tau) \in [(1 - \varepsilon)Q_v, (1 + \varepsilon)Q_v]\},$$

where, letting d_u be the degree of u,

$$Q_v = \begin{cases} 2/\tau & v = u, \\ (\theta/d_u)\log n & |v - u| = 1, \\ 0 & |v - u| > 1. \end{cases}$$

Then there exists a C depending on $\varepsilon, a_0, a_1, d_u$ only such that

$$\mathbb{P}^n(A^c) < \frac{C}{\log n}.$$

Proof In this argument, the expression *with high probability (w.h.p.)* stands for "with probability greater or equal to $1 - \frac{C}{\log n}$". Let $Z(t)$ be the number of distinct labels corresponding to particles that exit u during $[0, t]$ (where each label is counted at most once). Let $N(t)$ denote the total number of particles in the coupling with the mean-field coalescent of Lemma 2.5. Thus we have:

$$N(t) - Z(t) \leq X_u(t) \leq N(t) + Z(t), \text{ almost surely.} \tag{17}$$

Therefore one needs to estimate $N(t)$ and $Z(t)$. For any fixed ε, by Lemma 2.2 we have

$$\mathbb{P}^n(|\tau N(\tau)/2 - 1| > \varepsilon) < Ce^{-c/\tau} < Cn^{-1}. \tag{18}$$

So the event $\left\{\frac{N(\tau)}{2/\tau} \in (1 - \varepsilon, 1 + \varepsilon)\right\}$ happens with high probability. By Lemma 2.3 we have

$$\mathbb{P}^n\left(\left|\frac{Z(\infty)}{\log n} - \theta\right| > \varepsilon\right) < Cn^{-c}. \tag{19}$$

In the rest of the argument consider the process on the event $B := \left\{\frac{N(\tau)}{2/\tau} \in (1 - \varepsilon, 1 + \varepsilon)\right\} \cap \left\{\frac{Z(\infty)}{\log n} \in (\theta - \varepsilon, \theta + \varepsilon)\right\}$ that occurs with high probability. Note that on B, $Z(\tau) \leq Z(\infty) \leq (\theta + \varepsilon)\log n \ll 2/\tau$ and (17) imply the required bounds for $X_u(\tau)$.

Moreover, on $\{Z(\tau) \leq (1 + \varepsilon)\log n\} \supset B$, the probability that at least one particle jumps more than once before time τ is bounded by $\rho\tau(1 + \varepsilon)\log n = C/\log^2 n$. On the event that no particle jumps more than once, there cannot be any particle located at a distance strictly greater than 1 from u at time τ.

Similarly, on $\{Z(\tau) \leq (1 + \varepsilon)\log n\} \supset B$, the probability of at least one coalescence event involving particles located at site $v \neq u$ before time τ is at most $\tau\binom{(1+\varepsilon)\log n}{2}$, again bounded by $C/\log n$. We conclude that w.h.p. there is no coalescence outside of u before time τ.

This implies that w.h.p. the particles located at a neighbor v of u at time τ are precisely those that made a (single) jump from u to v. To show that their number is close to $(\theta/d_u)\log n$, it suffices to show that $Z(\tau)$ is concentrated around $\theta\log n$ (which is already known for $Z(\infty)$). Namely, since (on the event of high probability) each jump is to made from u to a random neighbor of u, and there are no further moves

or coalescence events involving the particles outside of u, $X_v(\tau)$ is concentrated near $Z(\tau)/d_u$ for any $v \sim u$, due to a law of large numbers argument. Indeed, the number of particles jumping from u to any particular of its neighbors has variance of order $\log n$, and using a normal approximation to binomial random variables, the probability of deviating by $\varepsilon \log n$ from the mean is no more than Cn^{-c}.

Thus it remains to show that $Z(\infty) - Z(\tau) \leq \varepsilon \log n$ with high probability. To this end, note that $Z(\infty) - Z(\tau)$ is the number of particles that exit u after time τ. Denote by $\mathcal{F}_\tau$ the σ-field generated by the evolution of the process up to time τ. By Lemma 2.3, monotonicity and the Markov property at time τ,

$$\mathbb{P}^n(Z(\infty) - Z(\tau) > \varepsilon \log n | \mathcal{F}_\tau)\mathbf{1}_{\{N(\tau)<3/\tau\}} < \mathbb{P}^{3/\tau}(Z(\infty) > \varepsilon \log n)$$
$$< C(3/\tau)^C e^{-\varepsilon \log n}$$
$$\ll 1/\log n,$$

and since $\{N(\tau) < 3/\tau\} \supset B$ occurs w.h.p. (because $\varepsilon < 1/2$), this concludes the argument. $\qquad\square$

Proof of Lemma 3.1 We have $\mathbb{P}(B_m^c) \leq \mathbb{P}(B_m^c \cap B_{m-1}) + \mathbb{P}(B_{m-1}^c) \leq \mathbb{P}(B_0^c) + \sum_{k \leq m} \mathbb{P}(B_k^c | B_{k-1})$. Noting that $\mathbb{P}(B_0^c) \leq 1/n$, we turn to estimating $\mathbb{P}(B_k^c | B_{k-1})$.

Given $\mathcal{F}_{t_{k-1}}$, consider now the coupling of Lemma 2.4 applied to the process observed on $[t_{k-1}, t_k]$, where the partition π is $\mathcal{F}_{t_{k-1}}$-measurable and where two labels belong to the same equivalence class of π if and only if their corresponding particles have the same position at time t_{k-1}. On the event B_{k-1} we have that $(-1/4)\log t_{k-1} < \log X_u(t_{k-1}) < -4\log t_{k-1}$ for $u \in B(o, k-1)$, and note that by definition, $-\log t_{k-1} = 3t_k^{-1/3}$. Hence Lemma 3.2 applies to each of the corresponding processes, where we may take $a_1 = 12^3$ and $a_0 = (1/2)(3/4)^3$. We conclude that with probability at least $1 - \frac{C}{\log X_u(t_{k-1})} \geq 1 - Ct_k^{1/3}$ the following occurs: during $[t_{k-1}, t_k]$ (i) No particle from u jumps more than once; (ii) At most $3/t_k$ particles remain at u; (iii) Each neighbor of u receives between $\frac{\rho}{\Delta}t_k^{-1/3}$ and $3\rho t_k^{-1/3}$ particles. Say that a vertex $u \in B(o, k-1)$ is *bad at stage* k on the complement of the above event.

Applying the right hand inequality of Lemma 2.4, on the event $B_{k-1} \cap \{$there are no bad vertices at stage $k\}$, we have that $X_u(t_k) = 0$ outside $B(o, k)$ (since no particle jumps twice and at time t_{k-1} all the particles are inside $B(o, k-1)$. Moreover, each site $v \in B_k$ has at least $\frac{\rho}{\Delta}t_k^{-1/3}$ particles jumping to it from some neighbor u of v, hence $X_v(t_k) \geq \frac{\rho}{\Delta}t_k^{-1/3}$. Finally, for each $v \in B(o, k)$ we have $X_v(t_k) \leq 3t_k^{-1} + 3\rho\Delta t_k^{-1/3} < 4t_k^{-1}$ (here we may assume that $t_k^{2/3} < 1/(3\rho\Delta)$), since it receives at most $3\rho t_k^{-1/3}$ from each of its (at most Δ) neighbors. Hence $B_{k-1} \cap \{$there is no bad vertex at stage $k\} \subset B_k$.

Therefore $\mathbb{P}(B_k^c | B_{k-1}) \leq \mathbb{P}(\exists$ a bad vertex at stage $k) \leq CV_{k-1}t_k^{1/3} \leq \frac{CV_m}{\text{Tow}(m-k, V_m^2)}$. It follows that

$$\mathbb{P}(B_m^c) \leq \frac{1}{n} + \sum_{k \leq m} \frac{CV_m}{\text{Tow}(m-k, V_m^2)} \leq \frac{C}{V_m},$$

since the term for $k = m$ overwhelmingly dominates all the others. □

3.2 Lower bound estimates

Lemma 3.1 gives us a fairly accurate description of the spatial coalescent up to positive times of order $o(1)$. Additional estimates are needed for understanding the behavior up to a constant time t. We begin with the lower bound, since it is simpler. Henceforth, we let $t > 0$ be a fixed time. Recall that the initial configuration of the spatial coalescent consists of $n = \text{Tow}(m, V_m^2)$ particles located at o.

Lemma 3.3 *Fix $t > 0$. The collection $(X_t(v), v \in B(o, m))$ can be coupled with the family $(\zeta_v, v \in B(o, m))$ of i.i.d. Bernoulli variables with mean $e^{-\rho t}$, so that*

$$\mathbb{P}\left(\forall v \in B(o, m), X_v(t) \geq \zeta_v\right) \xrightarrow[m \to \infty]{} 1.$$

Proof Assume that m is sufficiently large so that $t_m < t$. By Lemma 3.1, with probability tending to 1, each site in $B(o, m)$ is not empty at time t_m. On this event, fix one particle at each $v \in B(o, m)$, and color it red. Consider the evolution with coloring (see the proof of Lemma 2.4 for a similar construction), so that if a red particle coalesces with another particle, the newly formed particle retains the red color. Now, it is obvious that between time t_m and t, each red particle has probability $e^{-\rho(t-t_m)} > e^{-\rho t}$ of not migrating, independently of all other red particles, so the claim holds. □

3.3 Upper bound estimates

After time t_m, the bounds in the definition of B_m (cf. 16) still hold for most vertices, but will begin to fail for some vertices. As the number of particles per vertex decreases, the probability of failure increases. We overcome this by combining the second part of Lemma 2.3 with Lemma 2.7.

Lemma 3.4 *Fix $\varepsilon, t > 0$, and start with $n = \text{Tow}(m, V_m^2)$ particles at o. With high probability there is no particle outside $B(o, (1 + \varepsilon)m)$ at or before time t, and the total number of particles at time t is at most $CV_{(1+\varepsilon)m}$.*

Proof By Lemma 3.1, with high probability at time t_m there are no particles outside B_m, and the number of particles inside $B(o, m)$ is at most $4V_m/t_m = 4V_m^7$. We ignore all coalescence transitions after time t_m, so that each particle performs a simple random walk independently of all the others. By ignoring coalescence events, the number of particles located at any particular site at any later time can only become larger. Between times t_m and t, each particle makes an additional $\text{Poisson}(\rho(t - t_m))$ steps, so the probability that at least one of these particles makes at least εm steps is bounded by $4V_m^7 C^{\varepsilon m}/\lfloor \varepsilon m \rfloor!$. This last quantity tends to 0, since $V_m \leq C\Delta^m$.

Thus with an overwhelming probability, there are no particles outside $B(o, (1+\varepsilon)m)$ at time t. By Lemma 2.7 and the above observation, the number of particles within $B(o, (1+\varepsilon)m)$ is at most a constant multiple of $V_{(1+\varepsilon)m}$, again with an overwhelming probability. $\square$

3.4 Interpolation

Proof of Theorem 1.1 If $n = n(m) = \text{Tow}(m, V_m^2)$ for some m, then Lemmas 3.3 and 3.4 imply that with high probability the number of particles at time t is between cV_m and $CV_{(1+\varepsilon)m}$. Since $\log^* n = m + \log^* V_m^2 \sim m$, this implies the claim.

For intermediate n, we use the monotonicity of the process in n. Note that $\log^* n(m+1) - \log^* n(m) \leq 2$ (since $V_{m+1} < e^{V_m}$), so that the sequence $n(m)$ is sufficiently dense to imply the theorem. $\square$

Remark Since $m = \log^* n(m) - \log^* m + O(1)$, the proof above gives the lower bound $cV_{m-\log^* m}$. As for the upper bound, the proof of Lemma 3.4 works with radius $m + Cm/\log m$ in general, and $m + \log m$ for graphs with polynomial growth.

It is possible to get both lower and upper bounds that are closer to Vol $B(o, \log^* n)$. For the lower bound, one way would be to argue that most vertices continue to behave typically (as in Lemma 3.2) even up to constant times.

The upper bound is more delicate. One way of improving it is by considering the evolution of the total number of particles in $B(o, k)$ for $k > m$, similarly to the argument of Sect. 6. Under additional growth assumptions on the graph, both bounds are of order Vol $B(o, \log^* n)$.

4 Results for spatial Beta-coalescents

We now turn to the proof of Theorem 1.3. In fact, we prove a slightly more general result. Suppose that Λ has a sufficiently regular density near 0: $\Lambda(dx) = g(x)dx$, where for some $B > 0$ and $\alpha \in (1, 2)$ we have

$$g(x) \sim Bx^{1-\alpha}, \qquad x \to 0. \tag{20}$$

This includes the case where Λ is the Beta$(2 - \alpha, \alpha)$ distribution. A consequence of (20) is the following standard estimate for the rate of coalescence events when there are n particles remaining:

Lemma 4.1 *The sequence $(\lambda_n)_{n \geq 2}$ is increasing in n. Furthermore, there exists $c_0 > 0$ which depends only on α, B, such that if Λ satisfies (20), then $\lambda_n \sim c_0 n^\alpha$.*

Proof The monotonicity of λ_n in n is a consequence of the natural consistency of Λ-coalescents. The second part of the statement is a consequence of (20) and Tauberian theorems. See, e.g., [10, Lemma 4] for more details. $\square$

4.1 Lower bound in Theorem 1.3

As in the Kingman case, we first consider the mean-field coalescent satisfying (20). Define the following parameters

$$\beta = \frac{\alpha - 1}{2} \qquad \gamma = \min\left\{\frac{2-\alpha}{4}, \frac{\beta}{2}\right\}, \tag{21}$$

and observe that both $\gamma > 0$ and $\alpha - 2 + \gamma \leq -\gamma$. We shall consider $\tau = n^{-\beta}$. We first consider the quantity

$$Y_n = \int_0^\tau N(s)\, ds.$$

Lemma 4.2 *Assume that Λ satisfies (20). Then for $c_1 = 1/(4c_0)$, for some constant C depending only on Λ,*

$$\mathbb{P}(Y_n \geq n^{2-\alpha+\gamma}) \leq Cn^{-\gamma}, \quad \forall n \geq 2, \tag{22}$$

and

$$\mathbb{P}(Y_n \leq c_1 n^{1-\alpha/2}) \leq Cn^{-\gamma}, \quad \forall n \geq 2. \tag{23}$$

Remark It follows from Theorem 5 in [3] that $Y_n \sim cn^{2-\alpha}$, almost surely as $n \to \infty$, for some $c > 0$. However this result does not provide any estimate on the deviation probability.

Proof The key fact is that if the process $N(t)$ attains some value k, then it stays at k for an exponentially distributed time with mean $1/\lambda_k$. Since the probability of hitting k is at most 1,

$$\mathbb{E}Y_n \leq \sum_{k\leq n} \frac{k}{\lambda_k} \leq Cn^{2-\alpha}$$

by Lemma 4.1. The upper bound (22) follows by Markov's inequality.

The lower bound is more delicate. We argue that with high probability $N(t) \geq n/2$ for all $t < 1/(2c_0)n^{-\alpha/2} = o(\tau)$, yielding (23). In order for $N(t)$ to decrease to below $n/2$ quickly, there must either be an unusually large number of jumps, or else there must be unusually large jumps. We proceed to define these two events and bound their probability.

Let $M = n^{\alpha/2}$, and consider first the event that the first M jumps occur before time $n^{-\alpha/2}/(2c_0)$. By monotonicity, the rate of each jump is at most λ_n, and so the expected time t_M to make M jumps stochastically dominates $\sum_{i=1}^M E_i$, where E_i are

i.i.d. exponentials with rate λ_n. Now, from Lemma 4.1 we know that

$$\mathbb{E} \sum_{i=1}^{M} E_i = M/\lambda_n \sim n^{-\alpha/2}/c_0,$$

and by Lemma 2.1 with $\varepsilon = 1/2$,

$$\mathbb{P}(t_M \le n^{-\alpha/2}/(2c_0)) \le \mathbb{P}\left(\sum_{i=1}^{M} E_i < n^{-\alpha/2}/(2c_0) \right) \le e^{-M/16} \le Cn^{-\gamma}$$

as needed.

Let B_m be the decrease in the number of particles due to the next coalescence event, when the current configuration consists of m particles. It is known [5, Lemma 7.1] that there exists $C > 0$ such that

$$\mathbb{P}(B_m > k) \le Ck^{-\alpha} \quad \text{for all } m, k \ge 1. \tag{24}$$

In particular, $\mathbb{E}B_m \le C$ for some constant depending only on Λ. Hence the expected total decrease in the number of particles due to the first M jumps is at most CM. By Markov's inequality, the probability that at least $n/2$ particles are lost in the first M jumps is at most $MC/(n/2) = Cn^{\alpha/2-1} < Cn^{-\gamma}$. □

We now turn to the spatial setting. The next result gives a lower bound on the number of particles that exit the origin. This complements the upper bound of Lemma 2.6. Recall the coupling of Lemma 2.5, and that $Z(t)$ is the number of particles that exit the origin by time t and $N(t)$ counts the number of particles in a coupled mean-field coalescent with $N(0) = n$. The idea is that as long as Z is small, the true behavior is close to the upper bound.

Lemma 4.3 *With τ, γ as above, there is constants C, c so that $\mathbb{P}(Z(\tau) < \rho n^\gamma/2) \le Cn^{-c}$.*

Proof Recall the notations introduced in Lemma 2.6. We introduce the random time T defined by $T = \inf\{t > 0 : Z(t) \ge \frac{1}{4}N(t)\}$, and note that T is a stopping time with respect to the filtration $\mathcal{F}^*$. Recall $Y_n = \int_0^\tau N(s)ds$, and define the events $A = \{Z(\tau) < \rho n^\gamma/2\}$, and

$$B = \left\{ \frac{1}{2} < \frac{Z_\tau}{\rho Y_n} < \frac{5}{4} \right\}.$$

We bound the event A by

$$A \subset \{A \cap B\} \bigcup \{B^c \cap \{T \ge \tau\}\} \bigcup \{A \cap \{T < \tau\}\},$$

and proceed to bound the probability of each.

For the first, $A \cap B$ implies $Y_n \leq n^\gamma \leq n^{1-\alpha/2}/(4c_0)$ for n large enough, so by (23), $\mathbb{P}^n(A \cap B) < Cn^{-\gamma}$ for n large enough.

Denote by $\tilde{\mathbb{P}}^n$ the law

$$\tilde{\mathbb{P}}^n(\cdot) = \mathbb{P}^n(\cdot|\mathcal{G}),$$

of all processes, conditioned on the entire evolution of N. Consider the martingale S_t stopped at time τ. By Doob's inequality, we find that for any $\delta > 0$

$$\tilde{\mathbb{P}}^n\left(\sup_{s\leq\tau}|S_s| \geq \delta \int_0^\tau \rho N(s)ds\right) \leq \frac{4\rho\tilde{\mathbb{E}}^n\left(\int_0^\tau M_u du\right)}{\delta^2\left(\int_0^\tau \rho N(s)ds\right)^2} \wedge 1$$

$$\leq \frac{4}{\rho\delta^2\int_0^\tau N(s)ds} \wedge 1. \tag{25}$$

The last inequality follows from the first bound of (13), which implies that $\tilde{\mathbb{E}}^n(\int_0^\tau M(u)du) \leq \int_0^\tau N(u)du$. Until time T we have $M(t) \geq \frac{3}{4}N(t)$, and so (13) and (25) with $\delta = 1/4$ imply

$$\tilde{\mathbb{P}}^n(B^c; T \geq \tau) \leq \frac{16}{\rho\int_0^\tau N(s)\,ds} \wedge 1.$$

By (23), $\mathbb{P}^n(B^c; T \geq \tau) \leq Cn^{-\gamma}$.

Finally, note that $A \cap \{T < \tau\} \subset \{N(\tau) \leq 2\rho n^\gamma\}$ since $Z(\tau) \geq Z(T) \geq \frac{1}{4}N(T) \geq \frac{1}{4}N(\tau)$. By (24), there is a probability at most $Cn^{-\gamma\alpha} < Cn^{-\gamma}$ that $N(s)$ jumps from above $3\rho n^\gamma$ directly to below $2\rho n^\gamma$. Otherwise there is some s such that $N(s) \in [2\rho n^\gamma, 3\rho n^\gamma]$. The time $N(\cdot)$ stays at each visited value in this range is exponential with rate at most $\lambda_{3\rho n^\gamma}$. Thus the probability that N stays less than τ at such a value is bounded by $\tau\lambda_{3\rho n^\gamma} < Cn^{\gamma\alpha-\beta}$. Since $\gamma\alpha - \beta < 0$, this finishes the proof. □

We are now ready to start proving the lower-bound of Theorem 1.3. Let $t_0 > 0$ be a fixed time.

Lemma 4.4 *Consider the spatial coalescent started with n particles, all located at $u \in V$, that is, $X_v(0) = n\delta_u(v)$. Let $\tau = n^{-\beta}$, and define the event A by*

$$A = \bigcap_{v\sim u}\left\{X_v(\tau) \geq \frac{\rho}{4d_u}n^\gamma\right\}.$$

There are constants c, C depending on d_u only such that $\mathbb{P}(A^c) < Cn^{-c}$.

Proof Lemma 4.3 gives a bound on the probability of too few particles leaving the origin; thus with high probability at least $(\rho/2)n^\gamma$ particles leave u. It is highly probable that a proportion close to $1/d_u$ of these particles jumps to v. By consistency it suffices to select $\rho n^\gamma/(4d_u)$ of them and check that with high probability they have neither the time to coalesce nor to move an extra step before time τ.

If all these particles that migrate to v would do so immediately at time 0, so that they would have strictly more chance to coalesce, the number of particles remaining at v at time τ would still be sufficiently large. Indeed, it would then take Y amount of time, where Y is an exponential random variable with parameter $\lambda_{\rho n^\gamma/(4d_u)}$, before the first coalescence. Since $\lambda_m \leq cm^\alpha$ for all $m \geq 1$, we deduce that $\mathbb{E}(Y) \geq cn^{-\gamma\alpha}$. However, since $\gamma \leq \beta/2$ and $\alpha < 2$, we have $\tau = n^{-\beta} \ll cn^{-\gamma\alpha}$, hence $\mathbb{P}(Y < \tau) \leq cn^{\alpha\gamma-\beta}$.

In addition, note that by Lemma 4.1, the total jump rate of $\rho n^\gamma/(4d_u)$ particles is smaller than the total coalescence rate (since $\alpha > 1$), so the probability of an extra jump before time τ for any of the particles that have already jumped to v is smaller than $cn^{\alpha\gamma-\beta}$. It follows (by consistency) that there are at least $\rho n^\gamma/(4d_u)$ particles located at v at time τ, with probability greater than $1 - Cn^{-c}$. □

Proof of Theorem 1.3 *(lower bound)* Let $f_k(n) = f \circ f \ldots \circ f(n)$ (k iterations) where $f(n) = \rho n^\gamma/(4\Delta)$. Define the sequence of times $(\tau_k)_{k=1}^\infty$

$$\tau_k = \tau_{k-1} + f_{k-1}(n)^{-\beta}.$$

It is easy to check that if we take $k = k(n) = \log\log n/(-2\log\gamma)$, then

$$f_k(n) \geq c\exp(\sqrt{\log n})$$

Let A be the event that at each site within radius k there was at some time $t \leq \tau_k$ more than $f_k(n)$ particles. On A, reasoning as in Lemma 3.3, (at each site of this ball at least one particle remains with positive probability until time t_0), we see that $N^n(\tau) \geq c\,\text{Vol}\,B(o, k) \geq c\,\text{Vol}\,B(o, c\log\log n)$ for some $c > 0$. Thus to obtain the lower bound of Theorem 1.3, it suffices to compute the cumulative error probability in the iterated application of Lemma 4.4. However, it is easy to check that

$$\mathbb{P}(A^c) \leq \sum_{i=1}^k C\,\text{Vol}\,B(o, i)f_i(n)^{-\gamma} \leq Ck\,\text{Vol}\,B(o, k)f_k(n)^{-\gamma}.$$

Since $\text{Vol}\,B(o, k) < \Delta^k$, where Δ is the degree of the graph, this converges to 0 as $n \to \infty$. □

4.2 Upper bound in Theorem 1.3

The proof of the upper-bound in Theorem 1.3 requires a few additional estimates.

Consider the spatial coalescent on any graph G. Given some subset $A \subset V$ of the vertices, denote by Q_t the number of particles that are present in A throughout the time interval $[0, t]$.

Lemma 4.5 *There are constants $c, C > 0$ which depend on Δ only, so that*

$$\mathbb{P}\left(Q_{t_0} > Ct_0^{-1/(\alpha-1)}|A|\right) < e^{-c|A|}.$$

Proof Ignoring the particles after they exit A, one may assume that any particle leaving A is immediately killed. The main reason for Q_{t_0} being small is the coalescence. The total rate of coalescence at a site v holding X_v particles is $\lambda_{X_v} \sim c(X_v)^\alpha$. At each such event at least one particle disappears, and therefore the total rate of decrease of Q_t at time $t \leq t_0$ is at least

$$\sum_{v \in A} c(X_v(t))^\alpha \geq c|A|^{1-\alpha} Q_t^\alpha,$$

due to Jensen's inequality, since $\alpha > 1$. (This is similar to [32, Theorem 12], but the above inequality is stronger). Thus $(Q_t, t \leq t_0)$ is stochastically dominated by a pure death chain where the rate of decrease from $i - 1$ to $i - 1$ is $c|A|^{1-\alpha} i^\alpha$.

One concludes the argument using Lemma 2.1. Let E_k be independent exponential random variables with mean $\mu_k = c|A|^{\alpha-1} k^{-\alpha}$, and define $S_K = \sum_{k>K} E_k$. Then we have

$$\mathbb{P}(Q_{t_0} > K) < \mathbb{P}(S_K > t_0).$$

To apply Lemma 2.1 to S_K we need to estimate $\mathbb{E} S_K$ and $\text{Var}\, S_K$: note that for suitable constants, as $K \to \infty$,

$$\mathbb{E} S_K = \sum_{k>K} \mu_k \sim c_1 |A|^{\alpha-1} K^{1-\alpha} \tag{26}$$

and

$$\text{Var}\, S_K = \sum_{k>K} \mu_k^2 \sim c_2 |A|^{2\alpha-2} K^{1-2\alpha}. \tag{27}$$

In particular $\frac{\text{Var}\, S_K}{\mathbb{E} S_K \mu_K}$ is asymptotically constant and we may apply Lemma 2.1 with some constant ε. Thus for some $c_3 > 0$,

$$\mathbb{P}(S_K > 2\mathbb{E} S_K) \leq \exp\left(-c \frac{(\mathbb{E} S_K)^2}{\text{Var}\, S_K}\right)$$
$$< e^{-c_3 K}.$$

Now, if K is such that $\mathbb{E} S_K < t_0/2$ we may conclude that

$$\mathbb{P}(Q_{t_0} > K) < e^{-c_3 K}.$$

From (26) we see that $K = C t_0^{-1/(\alpha-1)} |A|$ works for C large enough.   □

Lemma 4.6 *Fix $\varepsilon > 0$. Consider the coalescent started with n particles, all located at $u \in V$, that is, $X_v(0) = n\delta_u(v)$. Let $\tau = n^{-\beta}$, and define the event A by*

$$A = \bigcap_v \{X_v(\tau) \leq C_1 Q_v\},$$

with

$$Q_v = \begin{cases} n^{1-(\alpha-1)/4} & \text{if } v = u, \\ n^{2-\alpha+\gamma} & \text{if } |v-u| \le r := \lceil 4/(\alpha-1) \rceil, \\ 0 & \text{otherwise.} \end{cases}$$

Then there are constants C, C_1 depending only on Λ such that $\mathbb{P}(A^c) < Cn^{-\gamma}$.

Proof We have that $A = A_u \cap A_{\le r} \cap A_{>r}$, where $A_u = \{X_u(\tau) \le C_1 Q_u\}$ and

$$A_{\le r} = \bigcap_{0 < |v-u| \le r} \{X_v(\tau) \le C_1 Q_v\}, \qquad A_{>r} = \bigcap_{|v-u| > r} \{X_v(\tau) = 0\}.$$

With sufficiently high probability, at most $n^{2-\alpha+\gamma}$ particles leave the origin by time τ (due to Lemma 2.6 and 22). This implies the required bound for $\mathbb{P}(A^c_{\le r})$.

We claim that except on an event of polynomially small probability, none of the (at most $n^{2-\alpha+\gamma}$ many) particles that leave u before time τ jump more than r times before time τ. Indeed, the probability any given particle makes more than r jumps before time τ is smaller than $(\rho\tau)^r$. Since there are at most n particles in total, we conclude that the probability of any particle reaching distance r is less than $Cn^{-\gamma}$ whenever r is such that $n^{1-r\beta} < n^{-\gamma}$. This implies the required bound for $\mathbb{P}(A^c_{>r})$.

It remains to prove a similar bound for $\mathbb{P}(A^c_u)$. Here we invoke Lemma 4.5 with an arbitrary set $A \ni u$ of size $c \log n$. If c is large enough then, except on an event of probability bounded by $n^{-\gamma}$, we have $Q_\tau < Cn^{1/2}|A| \ll n^{3/4} \le n^{1-(\alpha-1)/4}$. However, $X_u(\tau) < Q_\tau + Z_\tau$. It thus suffices to bound Z_τ. By Lemma 2.6, note that if $\delta = 2 - \alpha + \gamma$,

$$\mathbb{P}(Z_\tau > 2\rho n^\delta | \mathcal{G}) \le \mathbf{1}_{\{Y_n > n^\delta\}} + \mathbb{P}(\text{Poisson}(\rho n^\delta) > 2\rho n^\delta)\mathbf{1}_{\{Y_n \le n^\delta\}}.$$

Using standard Chernoff estimates for Poisson random variables, taking expectations and using (22), we deduce $Z_\tau \le 2\rho n^\delta$ with probability at least $1 - Cn^{-\gamma}$. It is easy to see from (21) that for all $\alpha \in (1, 2)$ we have $\delta < 1 - (\alpha - 1)/4$, so we deduce that $X_u(\tau) \le n^{1-(\alpha-1)/4}$ with probability at least $1 - Cn^{-\gamma}$. $\square$

Proof of Theorem 1.3 *Upper bound* Let $C_2 = \Delta C_1 \times \text{Vol } B(o, r)$, where C_1 and r are the constants in Lemma 4.6. Let $c_3 = 1 - (\alpha - 1)/4$, and note that $c_3 \in (2 - \alpha + \gamma, 1)$. Let $f(n) = C_2 n^{c_3}$, and as before set $f_k(n) = f \circ \cdots \circ f(n)$ (k iterations). Also set $\tau_1 = \tau = n^{-\beta}$, and

$$\tau_k = \tau_{k-1} + f_{k-1}(n)^{-\beta}.$$

Finally, let A_i be the event that at time τ_i there are no particles outside $B(o, ir)$ intersected with

$$\bigcap_{v, |v| \le ir} \{X_v(\tau_i) \le f_i(n)\}.$$

Choose $k = k(n)$ to be the maximal k so that $f_k(n) > \log n$. It is clear that $f_k(n) < (\log n)^2$. It is also straightforward to check that $k \sim c \log \log n$, and that $\tau_k = o(1)$.

Applying Lemma 4.6 iteratively, we see that

$$\mathbb{P}(A_k^c) \leq \sum_{i<k} Cf_i(n)^{-\gamma} \operatorname{Vol}(B(o, ir)) \leq C \operatorname{Vol}(B(o, kr)) f_k(n)^{-\gamma} \xrightarrow[n \to \infty]{} 0. \quad (28)$$

Consequently, at time τ_k the total number of remaining particles is at most $Cf_k(n) \operatorname{Vol}(B(o, kr))$, and these particles are all located in $B(o, kr)$, with high probability.

Consider now the set $B' = B(o, M \log \log n)$ for some large M to be specified soon. In order for any particle to exit B' by time t it must survive to time τ_k and jump at least $M \log \log n - kr$ times by time t. Thus the expected number of particles that exit B' by time t is at most

$$Cf_k(n) \operatorname{Vol}(B(o, kr)) e^{-c(M \log \log n - kr)} < C(\log n)^2 (\log \log n)^d e^{-(cM-c') \log \log n}.$$

Fix M large enough that the last expression tends to 0 as $n \to \infty$.

Finally note that if no particle leaves B' then $\sum_v X_v(t) = Q_t$. By Lemma 4.5, with high probability the number of particles that remain in B' throughout $[0, t]$ is at most $O(\operatorname{Vol} B')$. □

5 Global divergence of spatial Λ-coalescents

5.1 Infinite tree length for Λ-coalescents

Fix an arbitrary probability measure Λ on $[0, 1]$. Consider the corresponding mean-field Λ-coalescent that starts from a configuration consisting of infinitely many blocks, and let $(K^n(s), s \geq 0)$ be the number of blocks process of its restriction to the first n particles. Define:

$$X_n(t) \equiv X_n = \int_0^t (K^n(s) - 1) ds. \quad (29)$$

The notation K^n might be suggestive of the Kingman coalescent, so we wish to point out that the measure Λ in the following calculation is quite general.

We are interested in the quantity X_n due to the following observation: if K^n is a good approximation for the number of blocks at the origin of the spatial Λ-coalescent at small times s, then for t small, ρX_n approximates well the number of particles that emigrate from the origin up to time t (see, for instance, Lemma 2.6). The key ingredient in the proof of Theorem 1.2 is the following result.

Lemma 5.1 *For any fixed $t > 0$ we have $X_n \xrightarrow[n \to \infty]{} \infty$ almost surely.*

Proof Denote by $\sim^t$ the equivalence relation on the labels generated by the coalescent blocks at time t. For $n \geq 2$ let

$$\tau_n := \min\{t > 0 : \exists j < n \text{ s.t. } n \sim^t j\}$$

be the first time that the particle labeled n coalesces with any of the particles with smaller labels. We have that

$$K^n(s) = K^{n-1}(s) + \mathbf{1}_{\{s < \tau_n\}},$$

and therefore

$$X_n = X_{n-1} + (\tau_n \wedge t),$$

i.e. the contribution to X_n of particle n is $\tau_n \wedge t$.

Define $\mathcal{F}_n$ to be the σ-algebra generated by $\{K_s^j\}_{j \leq n, s > 0}$. Conditioned on $\mathcal{F}_{n-1}$, the infinitesimal rate of coalescence of particle n with particles with smaller labels at time s is given by

$$\int_{[0,1]} \frac{1}{x^2} \cdot x \cdot (1 - (1-x)^{K^{n-1}(s)}) \, d\Lambda(x).$$

Applying $(1-x)^k \geq 1 - kx$ (for $x \in [0, 1]$) we find that the rate of coalescence of particle n is at most $K^{n-1}(s)$ (with equality if and only if Λ is the point mass at 0, in which case the coalescent is Kingman's coalescent). Thus

$$
\begin{aligned}
\mathbb{E}(\tau_n \wedge t | \mathcal{F}_{n-1}) &= \int_0^t \mathbb{P}(\tau_n > s | \mathcal{F}_{n-1}) \, ds \\
&\geq \int_0^t \exp\left(-\int_0^s K^{n-1}(u) \, du\right) ds \\
&\geq \int_0^t \exp\left(-s - \int_0^t (K^{n-1}(u) - 1) \, du\right) ds \\
&= e^{-X_{n-1}} \int_0^t e^{-s} ds = (1 - e^{-t})e^{-X_{n-1}}.
\end{aligned}
$$

Note that X_n is increasing and consider the martingale

$$M_n = X_n - \sum_{k=2}^n \mathbb{E}(\tau_k \wedge t | \mathcal{F}_{k-1}).$$

On the event that X_n is bounded, the last calculation implies that $\mathbb{E}(\tau_k \wedge t | \mathcal{F}_{k-1})$ is bounded from below, hence $M_n \to -\infty$. Since M is a martingale, the last event has probability 0. □

Note that a different proof of Lemma 5.1 follows from Corollary 3 in [3], although the arguments there are significantly more involved.

5.2 Proof of Theorem 1.2

We now consider the spatial coalescent corresponding to some fixed Λ as in the previous section, on an arbitrary locally finite graph G. As usual, let n denote the initial size of the population, with all particles initially located at o, a fixed vertex of G. Recall the definitions of the processes M and Z in Lemma 2.5. Both processes M and Z depend implicitly on n, omitted from the notation. We consider the usual coupling of coalescents that correspond to different n.

Lemma 5.2 *For any $t > 0$ we have that $Z(t) \to \infty$ almost surely as $n \to \infty$.*

Proof We follow the argument of Lemma 4.3, except that we are only interested in showing that Z diverges, which simplifies the argument. Since $Z(t)$ is non-decreasing in n it suffices to show that for any fixed m we have $\mathbb{P}(Z(t) < m) \xrightarrow[n \to \infty]{} 0$.

Recall the martingales (15). On the event $\{Z(t) \leq m\}$, we have for all $s \leq t$ that $M_s \geq N(s) - Z_s \geq N(s) - m$, and therefore $S_t \leq m + \rho m t - \int_0^t \rho N(t) dt$. Due to Lemma 5.1, for any fixed m, t and any sufficiently large n, on the event $\{Z(t) \leq m\}$ (this event also depends on n)

$$ S_t \leq -\frac{1}{2} \int_0^t \rho N(s) ds. $$

As in (25), Doob's maximal inequality yields that for large enough n

$$ \mathbb{P}(Z_t < m | \mathcal{G}) \leq \mathbb{P}\left(\sup_{s \leq t} |S_s| \geq \frac{1}{2} \int_0^t \rho N_s ds \,\Big|\, \mathcal{G} \right) \leq \frac{16}{\rho \int_0^t N_s ds}. $$

By Lemma 5.1 the right-hand side tends to 0 almost surely, so $\mathbb{P}(Z(t) < m) \xrightarrow[n \to \infty]{} 0$. □

Fix $\varepsilon > 0$ and a vertex v of the graph, and let $E_{m,\varepsilon,v} = \{\sup_{t \in [0,\varepsilon]} X_v^n(t) \geq m\}$ be the event that at some time $t < \varepsilon$ there are at least m particles located at site v.

Lemma 5.3 *We have $\mathbb{P}^n(E_{m,\varepsilon,v}) \xrightarrow[n \to \infty]{} 1$.*

Proof Note that the claim is trivially true if $v = o$. We prove it first for v a neighbor of o. Take

$$ t_0 = \eta \min\{\varepsilon, \lambda_m^{-1}, (\rho m)^{-1}\}, $$

where η is an arbitrarily small number. Now, choose $n_0 = n_0(t_0)$ large enough that $Z_{t_0} > 2dm$ with probability at least $1 - \eta$, where $d = \deg(o)$. By the weak law of large numbers, one can choose $n_0 = n_0(t_0, h)$ large enough that on the event $\{Z_{t_0} > 2dm\}$, v receives at least m particles from o with probability at least $1 - \eta$. We concentrate on this event of high probability, and on these m particles, ignoring any further particles that might visit v.

Jumps from v occur at rate ρ per particle, so the probability that any of the m particles above leave v before time t_0 is at most η. Since a coalescent event involving any k-tuple of particles occurs at total rate λ_k (increasing in k), and since at a given time $s < t_0$ there are up to m of the above particles located at v, the probability of a coalescent event before time t_0 in which two or more of the m above particles participate is at most η. It follows that $\mathbb{P}(X_v^n(t) \geq m) \geq 1 - 4\eta$, for all $n > n_0$. Since η can be made arbitrarily small, this proves our claim for v a neighbor of o.

For other v we use induction in the distance $|v|$ to o. Indeed, such v has a neighbor u satisfying $|u| < |v|$. For any fixed m', η, and n sufficiently large, we have $\mathbb{P}^n(E_{m', \varepsilon, u}) \geq 1 - \eta$. Given this, and using the strong Markov property, one can repeat the previous argument with m' sufficiently large to conclude that with probability at least $1 - 2\eta$ there will be at least m particles at v (arriving from u) at some time $t < 2\varepsilon$. $\qquad\square$

Proof of Theorem 1.2 Again, due to monotonicity in n and t, it suffices to show that for any $m < \infty$ and any $t > 0$, we have $\lim_{n\to\infty} \mathbb{P}(N_t^n > m) = 1$.

Let $\eta, \varepsilon > 0$ be small numbers. Fix $m < \infty$, and choose a subgraph $G^m \subset G$ of size m such that the distance between any two vertices of G^m is larger than $1/\eta$. By Lemma 5.3 we have that

$$\mathbb{P}^n(E_{1, \varepsilon, v}) \xrightarrow[n\to\infty]{} 1, \ \forall v \in G^m.$$

Moreover, if $A_{v, \varepsilon}$ is the event that the first (if any) particle that enters v before time ε stays at v up to time ε (while it may possibly coalesce with other particles), note that $\mathbb{P}(A_{v, \varepsilon} | E_{1, \varepsilon, v}) \geq e^{-\rho\varepsilon}$. By choosing ε sufficiently small we arrive to

$$\lim_{n\to\infty} \mathbb{P}(\cap_{v \in G^m} A_{v, \varepsilon}) \geq 1 - \eta.$$

However, given $\cap_{v \in G^m} A_{v, \varepsilon}$, the probability that any pair of the above particles (located at mutual distance greater than $1/\eta$ at time ε) will coalesce before time t tends to 0 as $\eta \to 0$. $\qquad\square$

6 Lower bound for the long time asymptotics

We now turn to the large time asymptotic behavior of spatial coalescents. The underlying measure Λ does not play an important role here as it did for the behavior at constant times (provided that the mean-field coalescent comes down from infinity). The reason for this is that, as explained in the introduction (Sect. 1.2), at the beginning of this phase, say at constant time $t > 0$, the number of particles at each site is tight

with respect to n. When the number of particles at a site is small, the coalescents corresponding to different choices of Λ behave similarly. In fact, the density of particles quickly decays, and once it is small enough, it rarely happens that more than two particles are at the same location. With at most two particles at each site, any spatial Λ-coalescent is equivalent to spatial Kingman's coalescent.

An important quantity in this setting is the radius m of the region (ball) which is initially "filled" with particles. As we have seen, for Kingman's coalescent the radius of this ball is $m = \log^* n$, while in the case of Beta-coalescents it is approximately $m = \log \log n$ up to constants. In the general case, the radius m should be a certain function of both n and Λ, namely $m = f^*(n)$ where f^* is defined in (4). This was rigorously established only for Kingman's coalescent and those with "regular variation" (i.e., satisfying (20)). However, the results which we present in this section and the next one, are valid for essentially arbitrary coalescence mechanisms (subject to (48) for the upper bound in Sect. 7), and assume that the spatial Λ-coalescent starts with a possibly random but tight number of particles per site in a large ball of radius m. See Theorem 6.4 for the full statement. Note that in this result as in the rest of the paper, we will be taking limits as m tends to ∞, recalling that m is itself a function of n when applying these results to get Theorem 1.4.

Let us first present some further heuristic arguments for the lower bound in Theorem 1.4. Consider for the moment the case $d > 2$, so that the random walk migration process is transient. The first heuristic comes from the first moment calculation and simple Green function estimates: label the particles in an arbitrary way and let S_i be the total number of particles that ever coalesce with the ith particle (including the ith particle itself). Observing that a typical particle is at initial distance of order k away from an order k^{d-1} particles, where k ranges from 1 up to a number of order m, gives for a typical i

$$\mathbb{E}(S_i) \asymp \sum_{k=1}^{m} k^{d-1} \frac{1}{k^{d-2}} \asymp m^2,$$

where we use the fact that the probability that two particles ever coalesce is proportional to the probability that their corresponding walks intersect (visit the same site *at the same time*) (the constant comes from the delayed coalescence dynamics). The fact that this probability is approximately k^{2-d} is a well-known Green function estimate. Since $N = \sum_{i=1}^{M} 1/S_i$ gives the total number of clusters that survive forever (with M being the initial number of particles, of order m^d), and since $\mathbb{E}(1/S_i) \geq 1/\mathbb{E}(S_i)$ we arrive at

$$\mathbb{E}(N) \geq cm^{d-2}. \tag{30}$$

While Jensen's inequality may seem crude, this does give the correct exponents because the distribution of S_i is sufficiently concentrated. The next section contains results confirming this heuristic.

6.1 Technical random walk lemmas

We begin with technical results concerning random walks. Most of these are standard yet difficult to "pinpoint" in the random walk literature. Let $(S_n, n \geq 0)$ be sim-

ple symmetric random walk on $\mathbb{Z}^2$, started from a point $X_0 \in B(o, 2m)$ which will later be chosen in a certain random fashion (very roughly speaking, close to uniform) and recall that $m \to \infty$. Let $(X_t, t \geq 0)$ be a continuous time random walk on $\mathbb{Z}^2$ obtained as $X_t := S_{N_t^*}$ where $(N_t^*, t \geq 0)$ is a Poisson process with rate 1, independent of X. Let S' be a lazy version of S, with $S'_0 = S_0$ and step distribution given by $\mathbb{P}(S'_{n+1} - S'_n = 0) = 1/2$ and $\mathbb{P}(S'_{n+1} - S'_n = \pm \mathbf{e}_i) = 1/8$, where $\mathbf{e}_1, \mathbf{e}_2$ are the coordinate vectors, then $(S'_{N^*_{2t}}, t \geq 0)$ has the same law as X. We write $\mathbb{P}_x$ for the corresponding probability measures when $X_0 = x$.

Define $\tau'_x := \inf\{n \geq 0 : S'_n = x\}$, $\tau_x := \inf\{s > 0 : X_s = x\}$ to be the hitting times of x. Similarly, let $\tau'^+_x := \inf\{n \geq 1 : S'_n = x\}$ denote the *positive* hitting time of x. We abbreviate $\tau' = \tau'_0$, $\tau = \tau_0$ etc.

The next result is a variation of an Erdős-Taylor formula [19] (see also [17], p. 354). We assume that as $m \to \infty$,

$$\mathbb{E}\left(\frac{1}{\|X_0\|_+^2}\right) = O\left(\frac{1}{\log m}\right), \qquad \mathbb{E}\left(\log \frac{m}{\|X_0\|_+}\right) = O(1), \qquad (31)$$

where for any $y \in \mathbb{Z}$ we abbreviate $\|y\|_+ := \|y\| \vee 1$.

Lemma 6.1 *Assume $d = 2$ and fix $s > 4$, assume a random $\|X_0\| \leq 2m$ satisfies* (31). *Then*

$$\mathbb{P}(\tau < sm^2) \asymp \frac{\log s}{\log m + \log s}, \qquad (32)$$

where the constants implicit in the $\asymp$ notation depend only on those implicit in (31) *(and not on s or m).*

It is easy to check that X_0 drawn from a uniform on $B(o, 2m)$ or from a difference of two independent uniforms on $B(o, m)$ will satisfy the hypotheses of Lemma 6.1 and therefore (32) with universal constants (not depending on m) for any fixed $s > 4$. Also note that if $\mathbb{P}(X_0 \notin B(o, 2m)) = 1$, under no further restriction on the distribution of X_0, the upper bound on the probabilities $\mathbb{P}(\tau < sm^2)$ holds with the same constant as in Lemma 6.1. Indeed, in order for τ to happen, the walk needs first to enter $B(o, 2m)$ at a location close to its boundary, for which the argument below gives the required estimate.

Proof We estimate the above probability for any given $x \in B(o, 2m)$, and then integrate over the law of X_0. Without loss of generality, assume that sm^2 is an integer. Use the "last-exit decomposition":

$$\mathbf{1}_{\{\tau' < sm^2\}} = \sum_{k=1}^{sm^2-1} \mathbf{1}_{\{S'_k=0\}} \prod_{j=k+1}^{sm^2-1} \mathbf{1}_{\{S'_j \neq 0\}}$$

together with the Markov property, to obtain

$$\mathbb{P}_x(\tau' < sm^2) = \sum_{k=1}^{sm^2-1} \mathbb{P}_x(S'_k = 0)\mathbb{P}_0(\tau'^+ > sm^2 - k - 1).$$

We now apply a local central limit theorem and an estimate on the distribution of excursion length [37], statement E1 on p. 167, (or [30, Prop. 4.2.4]) and [37] statement P10 on p. 79 (or [30, Theorem 2.1.1]). We find that for some universal sequence $e_n \xrightarrow[n\to\infty]{} 0$

$$\mathbb{P}_x(\tau' < sm^2) = \sum_{k=1}^{sm^2-1} \frac{1}{k} e^{-\frac{2\|x\|^2}{k}} \frac{1 + e_{sm^2-k+1}}{\log(sm^2 - k + 1)}$$

$$+ O\left(\sum_{k=1}^{sm^2-1} \frac{1}{k\|x\|_+^2} \frac{1}{\log(sm^2 - k + 1)}\right)$$

$$= \sum_{k=1}^{sm^2-1} \frac{1}{k} e^{-\frac{2\|x\|^2}{k}} \frac{1 + e_{sm^2-k+1}}{\log(sm^2 - k + 1)} + O\left(\frac{1}{\|x\|_+^2}\right).$$

Split this sum in three: For $k \leq \|x\|^2$ use $e^{-x} < x^{-1}$ to get a total contribution of $O(1/\log(sm^2))$. For $k > sm^2 - \sqrt{sm^2}$ each term is at most C/k so the total contribution is $O(1/(\sqrt{s} \cdot m))$. Finally, for the intermediate k's each term is of order $1/k \log sm^2$, so the total contribution is of order $\log(sm^2/\|x\|^2)/\log(sm^2)$. Thus

$$\mathbb{P}_x(\tau' < sm^2) \asymp \frac{\log s + 2\log(m/\|x\|_+) + O(1)}{\log(sm^2)} + O\left(\frac{1}{\|x\|_+^2}\right),$$

uniformly over $x \in B(o, sm/2)$. Taking expectation with respect to X_0 while using (31) and $s > 2$, yields $\mathbb{P}(\tau' < sm^2) \asymp \frac{\log s}{\log(sm^2)}$ as $m \to \infty$.

Going back to the continuous time random walk, we have $\mathbb{P}(|N_{2t}^* - 2t| > \varepsilon t) \leq e^{-c(\varepsilon)t}$, $t \geq 0$ for some $c(\varepsilon) > 0$, accounting for an additional error of $O(e^{-c(\varepsilon)sm^2}) = o(1/\log(sm))$ in the corresponding estimate for τ.   □

We will also need later a simpler result which goes along the same lines.

Lemma 6.2 *Assume $d = 2$ and $X_0 = x$ is such that $\|x\| = m$. For all $c_1 > 0$, there exists $c_2 > 0$ which depends only on c_1 such that $\mathbb{P}_x(\tau_0 < c_1 m^2) \geq c_2/\log m$.*

Proof First we note that by easy large deviations on Poisson random variables, it suffices to prove the same inequality with τ_0 replaced by the discrete time τ'_0. By the strong Markov property, note that if $K(t)$ counts the number of hits of 0 by time t, then for all $c > 0$,

$$\mathbb{E}_x\big(K(cm^2)\big) \leq \mathbb{P}_x(\tau'_0 \leq cm^2)\mathbb{E}_0(K(cm^2)).\tag{33}$$

By the local central limit theorem,

$$\mathbb{E}_0(K(cm^2)) = \sum_{k=0}^{cm^2} \mathbb{P}_0(S'_k = 0) \asymp \sum_{k=1}^{cm^2} \frac{1}{k} \sim 2c \log m. \tag{34}$$

Also,

$$\mathbb{E}_x(K(cm^2)) = \sum_{k=0}^{cm^2} \mathbb{P}_0(S'_k = x) \asymp \sum_{k=1}^{cm^2} \frac{e^{-c'\|x\|^2/(2k)}}{k} \geq \sum_{k=cm^2/2}^{cm^2} \frac{e^{-c'm^2/(2k)}}{k} \geq c''. \tag{35}$$

Combining (33)–(35), we complete the proof. $\qquad\square$

If $d \geq 3$, we denote by G_X the Green function of a d-dimensional walk X. It is well-known (see e.g. [37]) that

$$G_X(x) \sim c\|x\|^{2-d}, \quad \text{as } \|x\| \to \infty, \tag{36}$$

for some constant c that depends on d (here $\|x\|$ denotes the Euclidean norm in $\mathbb{Z}^d$).

Let $X(\cdot)$, $Y(\cdot)$ be independent continuous random walks in $\mathbb{Z}^d$ $d \geq 2$ with jump rate 1 and with starting points uniform in $B(o, m)$. Denote by $\sigma_t\{X\} := \sigma\{X(s), 0 \leq s \leq t\}$ the natural filtration of X and let $\sigma\{X\} := \sigma_\infty\{X\} = \sigma\{X(s), 0 \leq s < \infty\}$. Define the stopping time $\tau := \inf\{t : X(t) = Y(t)\}$. Define the collision events by

$$H_s \equiv H_s(X, Y) := \{\tau \leq s\}, \qquad H \equiv H(X, Y) := \{\tau < \infty\}.$$

Lemma 6.3 *Let X, Y be independent continuous time random walks in $\mathbb{Z}^d$ starting at uniform points at $B(o, m)$. For any $d > 2$ we have*

$$\mathbb{P}(H) \asymp m^{2-d}, \tag{37}$$
$$\mathrm{Var}\,(\mathbb{P}(H|\sigma\{X\})) \leq Cm^{2(2-d)}, \tag{38}$$

while if $d = 2$, for any $t > 4$ we have

$$\mathbb{P}\left(H_{tm^2}\right) \asymp \frac{\log t}{\log m + \log t}, \tag{39}$$

$$\mathrm{Var}\,\left(\mathbb{P}(H_{tm^2}|\sigma\{X\})\right) \leq C\left(\frac{\log t}{\log m + \log t}\right)^2, \tag{40}$$

where C and the constants in $\asymp$ relation depend only on d.

Proof Assume first that $d \geq 3$. Note that the difference $X(t) - Y(t)$ is also a continuous time simple random walk (with a doubled rate of jumps), and abbreviate $G_{X-Y} = G$.

It is well-known and easy to check that

$$\mathbb{P}(H|X_0 = x_0, Y_0 = y_0) = \frac{G(x_0 - y_0) - \mathbf{1}_{\{x_0 = y_0\}}}{G(0)}. \tag{41}$$

Since $x \mapsto \|x\|^{2-d}$ is integrable near 0 as a function on $\mathbb{R}^d$, then (36) implies that

$$\frac{1}{\mathrm{Vol}\, B(o, m)} \sum_{y \in B(o,m)} G(x - y) \le Cm^{2-d}, \quad \text{for any } x \in \mathbb{Z}^d. \tag{42}$$

If $x \in B(o, m)$, then a corresponding lower bound holds since a positive fraction of points in $B(o, m)$ is at distance order m from x. Hence, for any $x \in B(o, m)$,

$$\frac{1}{\mathrm{Vol}\, B(o, m)} \sum_{y \in B(o,m)} G(x - y) \asymp m^{2-d},$$

where the constants implicit in $\asymp$ depend only on d. Due to (41), averaging over $x \in B(o, m)$ gives that $\mathbb{P}(H) \asymp m^{2-d}$ as claimed. (It is not hard to show similarly that $\mathbb{P}(H) \sim cm^{2-d}$ for some c.)

In order to show (38), introduce a third random walk Y' independent from, and identically distributed as, X and Y. In analogy to H define $H' = \{\exists t, X(t) = Y'(t)\}$. Given $\sigma\{X\}$, the events H, H' are independent and have the same probability. Thus

$$\begin{aligned}
\mathrm{Var}\, \mathbb{P}(H|\sigma\{X\}) &\le \mathbb{E}\big[\mathbb{P}(H|\sigma\{X\})^2\big] \\
&= \mathbb{E}\big[\mathbb{P}(H|\sigma\{X\})\mathbb{P}(H'|\sigma\{X\})\big] \\
&= \mathbb{E}\big[\mathbb{P}(\exists t, s : X(t) = Y(t), X(s) = Y'(s)|\sigma\{X\})\big] \\
&\le 2\mathbb{P}(\exists t, s : t \le s, X(t) = Y(t), X(s) = Y'(s)),
\end{aligned}$$

where for the last inequality we use the symmetry between Y and Y'. Denote by $\mathcal{F}_\tau$ the standard σ-field generated by processes X and Y up to time τ. On the event $\{\tau < \infty\}$, due to the strong Markov property and (41),

$$\mathbb{P}\left(\exists s \ge \tau : X(s) = Y'(s) \mid \mathcal{F}_\tau\right) \le c\mathbb{E}\big[G(X(\tau) - Y'(\tau)) \mid \mathcal{F}_\tau\big].$$

Let $Z = X(\tau) - (Y'(\tau) - Y'(0))$. Noting that $Y'(0)$ is independent from both $\mathcal{F}_\tau$ and Z, we have $\mathbb{E}(G(Z - Y'(0))|\mathcal{F}_\tau, Z) \le Cm^{2-d}$, almost surely, and therefore

$$\mathbb{E}\left(G(X(\tau) - Y'(\tau))|\mathcal{F}_\tau\right) = \mathbb{E}[\mathbb{E}\left(G(Z - Y'(0))|\mathcal{F}_\tau, Z\right)|\mathcal{F}_\tau] \le Cm^{2-d}.$$

In view of the discussion above this yields a uniform bound on $\mathrm{Var}\, \mathbb{P}(H|\sigma\{X\})$.

If $d = 2$, we proceed similarly, with H replaced by H_{tm^2}. In particular, Lemma 6.1 gives the asymptotics of $\mathbb{P}(H_{tm^2})$. For the conditional variance estimate, one obtains as above

$$\mathrm{Var}\, \mathbb{P}(H_{tm^2}(X, Y)|\sigma\{X\}) \le 2\mathbb{P}[\mathbf{1}_{\{\tau < t\}}\mathbb{P}(H_{tm^2}(X'', Y'')|\mathcal{F}_\tau)],$$

where X'', Y'' are independent random walks started from $X(\tau)$ and $Y'(\tau)$, respectively, and otherwise independent of $\mathcal{F}_\tau$. The result follows as before, since by Lemma 6.1, $\mathbb{P}(H_{tm^2}(X'', Y'')|\mathcal{F}_\tau) \asymp \frac{\log t}{\log m + \log t}$. □

6.2 Proof of the lower bound

We return to the spatial coalescent. Let Λ be an arbitrary finite measure on $(0,1)$. Consider a spatial coalescent with initial configuration $X(0)$ that stochastically dominates i.i.d. Bernoulli random variables with mean $p > 0$ in $B(o, m)$ (we make no assumptions on the initial configuration outside of $B(o, m)$). With a slight abuse of notation, we write $N^m(t)$ in this section for the total number of particles at time t, and we define $N \equiv N^m = \lim_{t\to\infty} N^m(t)$ be the number of particles that survive to time ∞.

Theorem 6.4 *Consider the spatial coalescent with initial state dominating Bernoulli variables in $B(o, m)$. If $d > 2$, then there exist a constant $a > 0$ such that*

$$\mathbb{P}(N > am^{d-2}) \xrightarrow[m\to\infty]{} 1.$$

If $d = 2$, then there exists a constant $a > 0$ such that, for any $t > 4$,

$$\mathbb{P}\left(N^m(tm^2) > a\,\frac{\log m}{\log t}\right) \xrightarrow[m\to\infty]{} 1.$$

Note that, since the total number of particles is non-increasing, the lower bound in the $d = 2$ case holds for any $t > 1$ with modified constant a (or with $\log(2 + t)$ in place of $\log t$ for any positive t).

We begin with a lemma stating a similar result for a simpler initial condition and with an "instantaneous" coalescent mechanism, where two particles coalesce as soon as they visit the same site. This model is called *coalescing random walks* (CRW). Afterwards we couple the two models to obtain Theorem 6.4.

Lemma 6.5 *Consider a system of s coalescing random walks, such that their initial positions are i.i.d. uniform points in $B(o, m)$, where*

$$s \equiv s(a) = \begin{cases} am^{d-2}, & d \geq 3, \\ a \log m, & d = 2. \end{cases}$$

Let $Z(t)$ denote the total number of particles at time t and let $Z = \lim_{t\to\infty} Z(t)$. If $d > 2$, then for some $a > 0$ we have $\mathbb{P}(Z > am^{d-2}/4) \xrightarrow[m\to\infty]{} 1$.

If $d = 2$, then for some a and all $t > 4$, we have $\mathbb{P}\left(Z(tm^2) > a\,\frac{\log m + \log t}{4 \log t}\right) \xrightarrow[m\to\infty]{} 1$.

Proof We use the following explicit construction of the CRW model with the given initial condition: Let $(X_i(t), t \geq 0)$, $i = 0, 1, \ldots, s - 1$ be a family of i.i.d. (non-coalescing) random walks, such that for each i, $X_i(0)$ is uniform in $B(o, m)$. At time 0, each block contains a single particle that is assigned a unique label in $\{0, 1, \ldots, s-1\}$.

While present in the system, the particle (or block of particles) carrying label i follows the trajectory of X_i. If the trajectories of blocks labeled i and j ever intersect, they instantaneously merge into a new block that inherits the smaller label $i \wedge j$.

Consider first the case $d > 2$. For each pair i, j let $A_{i,j} := \{\forall u \geq 0 : X_i(u) \neq X_j(u)\} = H(X_i, X_j)^c$. Then on $A_{i,j}$ the blocks carrying labels i and j cannot merge as a consequence of a single coalescence event, but might merge due to a collection of coalescence events involving lower indexed particles. However, on the event

$$A_k := \bigcap_{i<k} A_{k,i}, \tag{43}$$

the block carrying label k stays in the system indefinitely.

Consider the filtration $\mathcal{F}_k = \sigma\{X_i(\cdot), i \leq k\}$. Define $p_k = \mathbb{P}(A_k|\mathcal{F}_{k-1})$, and note that $p_0 = 1$. The random variables $\{p_k\}$, are a non-increasing sequence of random variables. To see this we use the fact that the random walks are independent and so

$$p_k \leq \mathbb{P}\left(\bigcap_{i=0}^{k-2} A_{k,i}|\mathcal{F}_{k-1}\right) = \mathbb{P}\left(\bigcap_{i=0}^{k-2} A_{k-1,i}|\mathcal{F}_{k-2}\right) = p_{k-1}, \text{ almost surely.}$$

Next, define events

$$B_k = A_k \cup \{p_k < 1/2\},$$

and note that

$$\mathbb{P}(B_k|\mathcal{F}_{k-1}) = \begin{cases} 1, & p_k < 1/2, \\ p_k, & p_k \geq 1/2. \end{cases}$$

Consider the martingale

$$M_k = \sum_{i=0}^{k} \mathbf{1}_{B_i} - \mathbb{P}(B_i|\mathcal{F}_{i-1}).$$

Note that M_k has increments with variance bounded (crudely) by 1. Thus $\text{Var } M_s < s$ (here s is the initial total number of blocks) and, by Markov's inequality,

$$\mathbb{P}(|M_s| > s/4) \leq \frac{s}{(s/4)^2} = \frac{16}{s}.$$

However, $\mathbb{P}(B_k|\mathcal{F}_{k-1}) \geq 1/2$, so by the definition of M, we find

$$\mathbb{P}\left(\sum_{i<s} \mathbf{1}_{B_i} < s/4\right) \leq \frac{16}{s} \xrightarrow[m\to\infty]{} 0. \tag{44}$$

Since p_k is non-increasing and since on the event $\{p_k \geq 1/2\}$ the events A_k and B_k coincide, we realize that on the event $\{p_s \geq 1/2\}$

$$\sum_{i<s} \mathbf{1}_{A_i} = \sum_{i<s} \mathbf{1}_{B_i}.$$

Thus if we prove that

$$\mathbb{P}(p_s < 1/2) \xrightarrow[m\to\infty]{} 0, \tag{45}$$

then (44) would imply the lemma. To this end we show that p_s is bounded below by a random quantity that is concentrated above $1/2$, via second moment estimates. Specifically, from the definition (43) we have

$$1 - p_s \leq \sum_{i<s} \mathbb{P}(A_{s,i}^c | \mathcal{F}_{s-1}) = \sum_{i<s} \mathbb{P}(A_{s,i}^c | \sigma\{X_i\}),$$

where the last identity is due to independence of $\sigma\{X_i\}$ for different i's. Moreover, $\{\mathbb{P}(A_{s,i}^c | \sigma\{X_i\}), i = 0, \ldots, s-1\}$ is an i.i.d. family of random variables. Using (37),

$$\mathbb{E}\left(\sum_{i<s} \mathbb{P}(A_{s,i}^c | \sigma\{X_i\})\right) < s \cdot Cm^{2-d} \leq Ca.$$

We choose $a = 1/(4C)$ so that this expectation is at most $1/4$. Due to (38),

$$\mathrm{Var}\left(\sum_{i<s} \mathbb{P}(A_{s,i}^c | \sigma\{X_i\})\right) \leq s \cdot Cm^{2(2-d)} \to 0.$$

so the sum is concentrated near its mean, and (45) follows.

In the case $d = 2$, the proof is almost identical. We take $s = a \log m$ and $a < 1/(4C \log t)$, where C is the constant that appears in (39). The event $A_{i,j}$ is accordingly redefined as $A_{i,j} := H_{tm^2}(X_i, X_j)^c$. Otherwise, the argument proceeds exactly as above, with (39), (40) used in place of (37), (38). □

Proof of Theorem 6.4 The idea is to couple the spatial coalescent X with a system of coalescing random walks, denoted $X^\diamond$, with an initial state of s particles at i.i.d. sites, uniform in $B(o, m)$. We first argue that it is possible to couple the initial states so that w.h.p. $X_v^\diamond(0) \leq X_v(0)$ (at every vertex). Indeed, in $X^\diamond$, there are $N^\diamond(0) \leq s$ occupied sites (since there may be repetitions) and given $N^\diamond(0)$, these sites are uniformly sampled from the ball $B(0, m)$ without replacement. On the other hand, $X(0)$ dominates a Bernoulli configuration on $B(o, m)$, hence $X(0)$ has at least $\mathrm{Bin}(\#B(o, m), p)$ particles sampled without replacement. Since $\mathbb{P}(\mathrm{Bin}(\#B(o, m), p) > s) \to 1$, this holds.

The second step of the proof is that if the initial configurations satisfy $X_v^\diamond(0) \leq X_v(0)$ for all v, then there is a coupling of the processes so that

$$X_v(t) \geq X_v^\diamond(t), \quad t \geq 0, v \in V. \tag{46}$$

To see this, observe that by the consistency property of spatial Λ-coalescent it suffices to prove the result assuming that $X_v(0) = X_v^\diamond(0)$ for all $v \in V$. In this case, (46) follows easily by induction on the number of particles: Just apply the consistency property of spatial Λ-coalescents, after the first time that two particles occupy the same site. (This idea is further exploited in Lemma 7.2.)

Finally, Theorem 6.4 follows by Lemma 6.5. □

6.3 Concentration of the number of particles

The main result of this section is a concentration result for the number of particles alive at a certain time. This provides a soft alternate route for the lower-bound on the long-time behavior of the spatial coalescent, as we briefly explain.

Theorem 6.6 *Fix $t > 0$, and consider a spatial Kingman coalescent started from some arbitrary configuration containing a finite number of particles. Then we have*

$$\mathrm{Var}(N(t)) \leq \mathbb{E}N(t).$$

Proof The tool used here again is a comparison to the coalescing random walk model, where particles coalesce immediately upon meeting. We denote by $(X^\diamond(t), t \geq 0)$ a system of instantaneously coalescing random walks started from a certain set of vertices A in a graph $G = (V, E)$, to be chosen suitably later, and let $N^\diamond(t)$ denote the total number of particles at time t. The proof is based on Arratia's correlation inequality [2, Lemma 1], which states that

$$\mathbb{E}X_x^\diamond(t)X_y^\diamond(t) \leq \mathbb{E}X_x^\diamond(t) \cdot \mathbb{E}X_y^\diamond(t). \tag{47}$$

Thus at any time, any two sites are negatively correlated. This inequality holds not just for the process on $\mathbb{Z}^d$, but on any edge weighted graph.

We now remark that the spatial Kingman coalescent on $\mathbb{Z}^d$ can be approximated by a system of instantly coalescing random walks on a larger graph. For any integer N such that $N > n$ (the initial number of particles), consider the graph $G_N = (V, E)$ with vertices $V = \mathbb{Z}^d \times \{1, \ldots, N\}$. The edges of G_N are of two types. If $x \sim y$ in $\mathbb{Z}^d$ then there is an edge between (x, i) and (y, j) with weight ρ/N. Additionally, there is an edge with weight $1/2$ between (x, i) and (x, j) for any x, i, j. Call the set $x \times \{1, \ldots, N\}$ a cluster. Clusters correspond to vertices of $\mathbb{Z}^d$ in a natural way. The $\mathbb{Z}^d$ coordinate of a continuous time random walk on G_N is a continuous time random walk on $\mathbb{Z}^d$ with jump rate ρ. However, two walks may be present in the same cluster and not meet. It is clear that as long as two random walks are in the same cluster they will meet at rate one (since each may jump into the vertex occupied by the other).

The probability of two random walks meeting when one jumps from one cluster to another is of order $1/N$. Thus as long as the number of particles is negligible compared to N, the projection onto $\mathbb{Z}^d$ of the coalescing random walks $X_N^\diamond$ on G_N is close to the spatial Kingman coalescent on $\mathbb{Z}^d$. As $N \to \infty$, the projection of $X_N^\diamond(t)$ converges to $X(t)$ (in the sense of vague convergence, identifying X_v and the projection of $X_N^\diamond$ to point measures on $\mathbb{Z}^d$). More precisely, for an initial configuration $X(0)$ of particles on $\mathbb{Z}^d$, we define a set $A \subset V_N$ by choosing for each v (arbitrarily) $X_v(0)$ particles from the cluster of v. Let $X_N^\diamond(t)$ be the process of coalescing random walks on G_N started with this configuration. Then if $M_N^\diamond(t)$ denote the total number of particles of $X_N^\diamond(t)$,

$$\mathbb{E}(M_N^\diamond(t)^2) = \sum_{x \in V_N} \mathbb{E} X_{N,t}^\diamond(x) + \sum_{x \neq y \in V_N} \mathbb{E} X_{N,t}^\diamond(x) X_{N,t}^\diamond(y)$$
$$\leq \mathbb{E} M_N^\diamond(t) + \sum_{x \neq y \in V_N} \mathbb{E} X_{N,t}^\diamond(x) \mathbb{E} X_{N,t}^\diamond(y)$$
$$\leq \mathbb{E} M_N^\diamond(t) + (\mathbb{E} M_N^\diamond(t))^2.$$

Thus for any N we have $\operatorname{Var} M_N^\diamond(t) \leq \mathbb{E} M_N^\diamond(t)$. By dominated convergence (since all processes have at most n particles) we see that

$$\lim_{N \to \infty} \mathbb{E} M_N^\diamond(t) = \mathbb{E} N(t) \qquad \lim_{N \to \infty} \mathbb{E} M_N^\diamond(t)^2 = \mathbb{E} N(t)^2,$$

and the result follows. □

As a simple corollary of this result, we obtain an alternate proof of Theorem 6.4. We have already seen in (30) that $\mathbb{E}(N(\infty)) \geq cm^{d-2}$ for some $c > 0$ if $d \geq 3$ (this argument is a simple Green function estimate, and is easy to adapt to the case $d = 2$). Applying Theorem 6.6 concludes the proof. Indeed, letting $t \to \infty$ in Theorem 6.6 by the monotone convergence theorem shows that $\operatorname{Var}(N(\infty)) \leq \mathbb{E}(N(\infty))$. Thus, by Chebyshev's inequality, letting $\mu = \mathbb{E}(N(\infty))$,

$$\mathbb{P}\left(N(\infty) \leq \tfrac{c}{2} m^{d-2}\right) \leq \mathbb{P}\left(|N(\infty) - \mu| \geq |\mu - \tfrac{c}{2} m^{d-2}|\right)$$
$$\leq \frac{\mu}{(\mu - \tfrac{c}{2} m^{d-2})^2} \leq \frac{4}{\mu} \to 0.$$

It would be also possible to derive a lower-bound on the expected number of particles in a system of instantaneously coalescing random walks at time tm^2, starting from a set A which dominates i.i.d. Bernoulli random variables with mean $p > 0$, using technology from coalescing random walks. We briefly outline the steps needed to do this. First, starting from a configuration where there is a particle at every site of $\mathbb{Z}^d$, and using a famous result of Bramson and Griffeath [14] on the asymptotic density of particles, we conclude that about cm^{d-2} such particles are in a region of volume Cm^d for some large $C > 0$ to be chosen suitably. If we treat the particles that

started outside of A as ghosts, we are then led to estimate the number of ghost particles among those cm^{d-2}. For this, one can use the duality with the voter model (see [31]) and [33, Lemma 4], which gives good control on the probability that the voter model escapes a ball of radius $\sqrt{t}$, for large t.

7 Upper bound for the number of survivors

Assume that Λ is a finite measure on $[0, 1]$ such that for some $a_0 > 0$, we have

$$\lambda_n \geq a_0 n \quad \text{for all } n \geq 2, \tag{48}$$

where $\lambda_n = \sum_{k=2}^n \lambda_{b,k}$ is the total merger rate when there are n particles. Note that most coalescents which come down from infinity satisfy (48), in particular, if $\Lambda = \delta_{\{0\}}$ (the Kingman case) then (48) holds since $\lambda_n = \binom{n}{2}$, and if Λ has the regular variation property of (20), then (48) holds by Lemma 4.1.

Our goal here is to prove the following result.

Theorem 7.1 *Fix $C_0 \in (0, \infty)$ and $\delta > 0$, and consider the spatial Λ-coalescent in $\mathbb{Z}^d$ satisfying (48), started from a configuration of at most $C_0 m^d$ particles located in $B(o, m)$, and no particles in $\mathbb{Z}^d \setminus B(o, m)$. There exists $C = C(\delta, C_0)$, such that if $d > 2$ then*

$$\mathbb{P}\left(N^*(\delta m^2) < Cm^{d-2}\right) \xrightarrow[m \to \infty]{} 1,$$

while, if $d = 2$,

$$\mathbb{P}\left(N^*(\delta m^2) < C \ln m\right) \xrightarrow[m \to \infty]{} 1.$$

Note that when $d > 2$ this order of magnitude bound is sharp, since Theorem 6.4 shows $N^*(\infty) \geq cm^{d-2}$. For $d = 2$, due to recurrence, $N^*(\infty) = 1$, almost surely.

The idea behind the proof is a comparison of the spatial system to a mean field approximation. The actual argument is based on a somewhat technical construction so we start with a non-technical overview. Recall the comparison with ODE described in (5): if at time t the density of particles averaged over some ball is $\rho(t)$ (typically small), then we approximate the spatial coalescent with the mean-field model where the coalescence rate per particle is $\rho(t)$ at time t, leading to the differential equation

$$\frac{d}{dt}\rho(t) = -\rho^2(t)/2, \quad t \geq s.$$

Hence $\rho(t)^{-1} = \rho(s)^{-1} + (t - s)/2$ and therefore $\rho(t) = \frac{2}{t - s + 2\rho(s)^{-1}}$, $t \geq s$. Provided that all the particles in the spatial coalescent configuration are located in the ball of radius m during the whole interval $[s, t]$ (and that the above approximation is valid) then their total number is approximately $Cm^d \rho(t)$. In turn, this approximation

remains valid as long as the particles remain inside a ball centered at the origin with radius of order m, i.e. up to time of order m^2. At times of order m^2, the number of remaining particles is of order m^{d-2}.

A key difficulty of the approach outlined above comes from the fact that some particles diffuse away from the densest regions relatively early in the evolution, which might enable them to survive longer. To account for such "runaways", we adopt a *multi-scale approach*, bounding at each stage the number of particles that "escape". This is done in Lemma 7.8. Lemma 7.5 provides the estimates on the number of non-escaping particles at each stage.

To justify the comparison of the spatial process with the mean field process we average over small time intervals (cf. Lemma 7.4 below). This is necessary since at any given time it is possible that no vertex contains more than a single particle, in which case the immediate rate of coalescence is 0. However, the system is unlikely to stay in such states long enough to hinder the approximation. Indeed, Lemma 7.4 implies that the average rate of coalescence is (up to constants) as predicted by the mean field heuristic. The multiplicative constants are inherent to the spatial structure, and it seems difficult to compute them.

7.1 Preparatory lemmas

Our first step is a comparison lemma between the spatial Λ-coalescent X and a slower spatial coalescent. We then consider a possibly more general spatial coalescent process $\{(\bar{X}_v(t), t \geq 0)\}_{v \in V}$. If the process consists initially of n particles labeled by $[n] = \{1, \ldots, n\}$, a configuration consists as usual of labeled partitions of $[n]$, where the label of a block corresponds to its location on V. Equivalently, a configuration $\bar{x} = (\bar{x}_v)_{v \in V}$ may be thought of as giving the list of blocks (referred to as particles) present at each particular site $v \in V$. We will also sometimes abuse notation and denote by $\bar{X}_v(t)$ the number of particles (i.e., blocks) present at time t and at position v. We assume that particles perform independent continuous-time simple random walks with jump rate ρ, and that there exists a family of real numbers $\bar{\lambda}_{\bar{x}, S}$ such that for all configuration $\bar{x} = (\bar{x}_v)_{v \in V}$, all $v \in V$, any particular subset S of all blocks present at $v \in V$ coalesces at an instantaneous rate $\bar{\lambda}_{\bar{x}, S}$, if the current configuration is $\bar{x}$. Moreover, coalescence events at different sites occur independently of one another, and are independent of the migration. We now make the following assumption on the family of rates $\bar{\lambda}_{\bar{x}, S}$: if $v \in V$ and $\bar{x}_v$ contains $n \geq 2$ particles, then for every $2 \leq k \leq n$, we have:

$$\sum_{S: |S| \geq k} \bar{\lambda}_{\bar{x}, S} \leq \sum_{\ell \geq k} \binom{n}{\ell} \lambda_{n, \ell}, \tag{49}$$

where $\lambda_{n,k}$ is the coalescence rate of any particular subset of size k in a Λ-coalescent. The idea behind (49) is that if X and $\bar{X}$ have the same number of particles at time t, then $X(t + \varepsilon)$ is stochastically dominated by $\bar{X}(t + \varepsilon)$.

Lemma 7.2 *Consider a Λ-coalescent X and a coalescent process $\bar{X}$ such that* (49) *holds, and $X_v(0) \leq \bar{X}_v(0)$ for all v. Then there is a coupling of the processes X and $\bar{X}$ such that $X_v(t) \leq \bar{X}_v(t)$ holds for all $v \in V$ and $t \geq 0$.*

Proof By the consistency of spatial Λ-coalescents, it suffices to prove the result when $X_v(0) = \bar{X}_v(0)$ for all $v \in V$. We associate each particle of X with a particle of $\bar{X}$ and let them perform the same random walks as long as there are no coalescence events. A consequence of (49) is that it is possible to couple the processes so that if $X_v = \bar{X}_v$ then the coalescence events of X dominate those of $\bar{X}$, that is, any coalescence event in $\bar{X}$ occurs at the same time as an event in X involving at least as many particles.

The proof now proceeds by induction on the total number of particles, which are allowed to be distributed arbitrarily. By the above remark, we may couple the processes X and $\bar{X}$ so that the domination holds up to and including the first time t_0 of a coalescence event, which involves particles from X and possibly from $\bar{X}$. Assume that $\bar{X}$ also experiences a coagulation event at this time. (Else, we can artificially retain particles in X that were supposed to coagulate at time t_0. By the consistency property, this may only increase the process X stochastically.)

We now use the induction hypotheses to construct processes $(X'(t), t \geq t_0)$ and $(\bar{X}'(t), t \geq t_0)$ with initial configuration $X'(t_0) = \bar{X}'(t_0) = \bar{X}(t_0)$ such that $X'_u(t) \leq \bar{X}'_u(t)$ for all $t \geq t_0$. We can define $\bar{X}(t) = \bar{X}'(t)$ for $t > t_0$, and by consistency of the spatial Λ-coalescents, we extend the coupling to X for $t > t_0$ so that $X_u(t) \leq X'_u(t)$ for all $u \in V$, which proves the claim. $\square$

Remark This lemma holds for more general spatial coalescents: e.g., the instantaneous coalescence rates $\lambda_{\bar{x},S}$ could be allowed to be arbitrary path-dependent (i.e., $\mathcal{F}_t$-measurable at time t), almost surely nonnegative and finite random variables. The only crucial assumption is that (49) holds uniformly.

We now apply Lemma 7.2 to the situation which is particularly useful in our setting. Recall that we are considering a spatial Λ-coalescent for which (48) holds. Assume that initially there are N particles, and let $\{X_v(t), t \geq 0\}_{v \in V}$ denote the number of particles of this process as a function of time and space.

Let π be a partition of $\{1, \ldots, N\}$. We refer to the blocks of π as *classes*. Let $\{\bar{X}_v(t), t \geq 0\}_{v \in V}$ denote a process where classes evolve independently of one another, and particles within each class evolve according to a spatial $(\bar{\Lambda})$-coalescent, where $\bar{\Lambda}$ will be specified soon. That is, particles move as continuous-time simple random walks with rate ρ and coalesce when they are on the same site and from the same class according to a $\bar{\Lambda}$-coalescent.

Lemma 7.3 *Assume that the blocks of π are all of size 1 or 2, and that $\bar{\Lambda} = (a_0/\lambda_2)\Lambda$, where a_0 is the constant of* (48) *and $\lambda_2 = \lambda_{2,2}$ is the pairwise coalescence rate. Assume also that $X_v(0) \leq \bar{X}_v(0)$ for all v. Then there is a coupling of the processes X and $\bar{X}$ such that $X_v(t) \leq \bar{X}_v(t)$ holds for all $v \in V$ and $t \geq 0$.*

Proof Observe first that our process $\bar{X}$ is of the type described above Lemma 7.2, so that it suffices to establish (49). Note however that if a configuration $\bar{x}$ contains n particles at site v, and S is a subset of particles with $|S| = k$ and $2 \leq k \leq n$, we have

$\bar{\lambda}_{\bar{x},S} = 0$ for $k \geq 3$, while if $k = 2$, $\lambda_{\bar{x},S} = 0$ when the particles of S are not of the same class, and if they are of the same class, $\lambda_{\bar{x},S} = (a_0/\lambda_2)\lambda_2 = a_0$. Since there are at most n subsets of particles that are allowed to coalesce, we have

$$\sum_{S:|S|\geq 2} \bar{\lambda}_{\bar{x},S} \leq na_0 \leq \lambda_n = \sum_{k=2}^{n} \binom{n}{k}\lambda_{n,k},$$

which proves (49), and completes the proof. $\qquad\square$

Lemma 7.4 *Fix* c_0, C_0, *and consider a spatial* Λ-*coalescent satisfying* (48) *with* $N^*(0)$ *particles all inside* $B(o, R)$. *Let* $\rho(t) = \frac{N^*(t)}{R^d}$ *be the density, and assume* $\rho(0) \in (c_0 R^{-2}, C_0)$. *Denote* $\tau = \rho(0)^{-2/d}$. *Then for* $d > 2$ *we have*

$$\mathbb{P}\left(\rho(\tau)^{-1} < \rho(0)^{-1} + c_1\tau\right) < \exp\left(-cR^{(d-2)^2/d}\right),$$

where c_1 *depends only on* d, c_0, C_0, a_0.
    If $d = 2$ *we have*

$$\mathbb{P}\left(\rho(\tau)^{-1} < \rho(0)^{-1} + \frac{c_1\tau}{\log \tau}\right) < \exp\left(-\frac{cR^2}{\tau \log \tau}\right).$$

Proof We first argue that for some $C = C(d)$, it is possible to find at least $N^*(0)/4$ disjoint pairs in the set of initial particles, so that for each pair the initial distance between its particles is at most $C\rho(0)^{-1/d}$ (for large $\rho(0)$, the particles forming such a pair are initially located at the same site). To achieve this, cover $B(o, R)$ with $N^*(0)/2$ (disjoint) boxes of diameter $CRN^*(0)^{-1/d} = C'\rho^{-1/d}$ (this is possible for some C). Within each box match as many pairs as possible in an arbitrary manner. This leaves at most one unpaired particle in each ball, so at least $N^*(0)/2$ are matched, with all distances bounded as claimed. Refer to two particles forming a pair as "partners".

Consider the coupling from Lemma 7.3, where π corresponds to the partitioned formed by identifying particles with their partners (which therefore contains only singletons or doubletons). Let Z' be the total number of coalescence events in the process Π' where coalescence events involving members of different classes are not allowed and occur at rate a_0 when they are. Lemma 7.3 implies that $Z' \preceq Z$, in the sense of stochastic domination. Hence, it suffices to prove the claimed bounds for Z'. The advantage of considering Π' instead of Π is that different pairs of partners evolve independently.

From this point on, the arguments for the cases $d = 2$ and $d > 2$ differ. In dimensions $d > 2$, by our assumptions, $\tau > c_0$ for some c_0. The probability that random walkers started at distance at most $\rho^{-1/d}$ meet before time $\tau/2 = \rho^{-2/d}/2$ is at least $c\rho^{(d-2)/d}$. On this event, there is probability bounded from 0 that they coalesce before time τ. Thus the number of partners that coalesce by time τ dominates a $\text{Bin}(N^*(0)/4, c\rho^{(d-2)/d})$ random variable. This random variable has expectation $cN^*(0)\rho^{(d-2)/d} \geq cR^{(d-2)^2/d}$. The bound in the lemma is the probability that this random variable is less than half its expectation.

Finally, if the number of coalescence events is at least $cN^*(0)\rho(0)^{(d-2)/d} = cN^*(0)\rho(0)\tau$ then

$$\rho(t)^{-1} = \frac{R^d}{N^*(t)} \geq \frac{R^d}{N^*(0)(1-c\rho(0)\tau)} \geq \frac{R^d}{N^*(0)}(1+c\rho(0)\tau) = \rho(0)^{-1}+c\tau.$$

In the case $d = 2$, each pair coalesces with probability at least $c/\log\tau$ (along the same lines) by Lemma 6.2. As above, the number of coalesce events is at least $cR^2/\tau\log\tau$ except with probability $e^{-cR^2/\tau\log\tau}$. On this event, a similar computation gives

$$\rho_t^{-1} \geq \rho_0^{-1} + \frac{c}{\tau\log\tau}.$$

$\square$

Lemma 7.5 *Let S_t denote the number of particles (in the spatial coalescent) that remain in $B(o,R)$ during the whole interval $[0,t]$. In particular, $S_0 = \sum_{v\in B(o,R)} X_v(0)$. Fix $C_0 > 0$, and assume that $S_0 = n < C_0 R^d$ and that $2 < t < R^2$. Then for some C depending only on d,*

$$\mathbb{P}\left(\frac{S_t}{R^d} > \frac{C}{t}\right) < cR^4 \exp\left(-cR^{(d-2)^2/d}\right) \quad \text{if } d > 2,$$

and

$$\mathbb{P}\left(\frac{S_t}{R^2} > \frac{C\log t}{t}\right) < (\log t)^{-1} \quad \text{if } d = 2.$$

Proof As usual, the case $d > 2$ is considered first. Assume that $S_0 > R^{d-2}$, else there is nothing to prove. The previous lemma can then be formulated as follows: The process $\{\rho_t^{-1}\}$ is unlikely to spend more than $u^{-2/d}$ units of time in the interval $[u, u + c_1 u^{-2/d}]$. Note that S_t can only decrease faster than $N^*(t)$, so this will also hold for the modified density $\rho(t) = \frac{S_t}{R^d}$.

We apply this to the following sequence of intervals. Let $u_0 = \rho(0)$ and $u_{k+1} = u_k + c_1 u_k^{-2/d}$. Let K be minimal with $u_K > t/c_1$. As long as $u_j < t/c_1$ the increment is at least $ct^{-2/d}$. It follows that $K < Ct^{1+2/d} < R^4$. If the process does not spend more than $u_k^{-2/d}$ time in $[u_k, u_{k+1}]$ then the time before ρ^{-1} exceeds t/c_1 is at most t. The probability that this fails to hold is at most $R^4 \exp\left(-cR^{(d-2)^2/d}\right)$.

This works provided $K > 1$, or equivalently $t \geq \rho(0)^{-2/d}$. If $t < \rho(0)^{-2/d}$ then we have

$$S_t \leq S_0 = \rho(0)R^d \leq \frac{R^d}{t^{d/2}} < \frac{R^d}{t}.$$

In the case $d = 2$, we instead have $u_{k+1} = u_k + \frac{c_1 u_k}{\log u_k}$. It is not hard to see that

$$u_k \asymp e^{\sqrt{2c_1 k + c}} \tag{50}$$

for some c depending on u_0. To this end, note that u is increasing and hence is dominated by the solution of the ODE $f' = c_1 f / \log f$ (at least once u is large enough that $u / \log u$ is increasing). This ODE is solved by $f = e^{\sqrt{2c_1 x + x}}$, giving the upper bound on u. For the other direction, note that once u_k is large u_{k+1}/u_k is close to 1. This implies that u dominates a solution of $f' = (c_1 - \varepsilon) f / \log f$.

Lemma 7.4 tells us that ρ_t^{-1} is unlikely to spend more than u_k units of time in $[u_k, u_{k+1}]$, and the probability of this unlikely event is at most

$$p_k = e^{-cR^2 / u_k \log u_k}.$$

Let K be such that $u_K > \alpha t / \log t$, with α small to be determined soon. Note that now the failure probability for the last intervals is of order 1, so a union bound does not work. However, Lemma 7.4 tells us more. If the process ρ_t^{-1} fails to exceed u_{k+1} in the next u_k units of time, then by the Markov property and Lemma 7.4 again, it gets a fresh chance to do so in the next u_k units of time. Therefore, the number of attempts is smaller than a geometric random variable with success probability $1 - p_k$. It follows that the total time spent in $[u_0, u_K]$ is stochastically dominated by $Q := \sum_{k=1}^{K-1} u_k \, \text{Geom}(1 - p_k)$, where the geometric random variables are independent. The probability we wish to bound is therefore at most $\mathbb{P}(Q > t)$. By making α small we can guarantee $p_k < 1/2$ for all k, so that the geometric variables are typically small.

More precisely, from (50) it follows that $K \asymp \log^2 t$ and therefore that

$$\sum_{k < K} u_k \asymp \int_0^K e^{\sqrt{2c_1 t}} dt \asymp \sqrt{K} u_K \asymp \alpha t,$$

hence for small enough α we have $\mathbb{E} Q \leq 2 \sum_{k=1}^{K-1} u_k \leq t/2$. Similarly, we can compute

$$\text{Var } Q \leq C \sum_{k < K} u_k^2 < \frac{C \alpha^2 t^2}{\log t}.$$

The lemma now follows from Chebyshev's inequality and choosing α small enough that $4\alpha^2 C \leq 1$. □

The following is a fairly standard fact which follows easily from the optional stopping theorem and Doob's inequality:

Lemma 7.6 *If X_t is a continuous time random walk on $\mathbb{Z}^d$ then for all $x \geq 0$,*

$$\mathbb{P}\left(\sup_{s \leq t} \|X_s\|_2 \geq x\right) \leq Ce^{-cx^2/t},$$

where c, C depend only on d.

Lemma 7.7 *If $W = \mathrm{Bin}(n, p)$, and $\Delta > 0$, then*

$$\mathbb{P}(W > 2np + \Delta) < \Delta^{-1/2}. \tag{51}$$

Proof If $\Delta > (np)^2$, then Markov's inequality gives $\mathbb{P}(W > \Delta) \leq np/\Delta < \Delta^{-1/2}$. If $\Delta \leq (np)^2$, one can use Chebyshev's inequality to obtain $\mathbb{P}(W > 2np + \Delta) \leq \mathbb{P}(W - \mathbb{E}W > np) \leq (1 - p)/(np) < \Delta^{-1/2}$. $\qquad\square$

7.2 Completing the proof

Lemma 7.5 is almost sufficient to deduce Theorem 7.1. The missing piece is to account for the particles that "escape" from the ball under observation. We accomplish this by partitioning the time interval $[0, m^2]$ into several segments and applying Lemma 7.5 to each segment. More precisely, let $K = K(m) = \lfloor \log \log m \rfloor$, and consider the process at a particular sequence of times given by

$$t_k = \begin{cases} 0 & k = 0, \\ e^{k-K}m^2 & k = 1, \ldots, K. \end{cases}$$

Thus $t_1 \asymp m^2/\ln m$, and the sequence increases geometrically up to $t_K = m^2$. At each time t_k, we will consider the behavior of the process with respect to the ball $B(o, R_k)$, where the radii R_k are defined by:

$$R_k = \begin{cases} 0 & k = 0, \\ \gamma(m + \sqrt{t_k(K + 1 - k)}) & k > 0, \end{cases}$$

where $\gamma > 1$ is some constant to be determined during the proof of Lemma 7.8. Note that R_k is increasing in k, and that $R_K = 2\gamma m$.

With the above notations in mind, let X_k (resp. Y_k) be the number of particles inside (resp. outside) $B(o, R_k)$ at time t_k. Let $A_m = A_m(\alpha, \beta, \gamma)$ be the event

$$A_m = \left\{ X_k < \frac{\alpha R_k^d}{t_k} \quad \text{and} \quad Y_k < \frac{\beta R_k^d}{t_k} \quad \text{for all } k \leq K \right\} \qquad \text{when } d > 2.$$

and

$$A_m = \left\{ X_k < \frac{\alpha R_k^2 \log m}{t_k} \quad \text{and} \quad Y_k < \frac{\beta R_k^d \log m}{t_k} \quad \text{for all } k \leq K \right\} \qquad \text{when } d = 2.$$

Lemma 7.8 *Assume the initial conditions of Theorem 7.1. Then for some choice of α, β, γ we have $\mathbb{P}(A_m) \xrightarrow[m \to \infty]{} 1$.*

Proof The idea is to inductively bound X_{k+1}, Y_{k+1} in terms of X_k, Y_k. The bound on X_{k+1} is mostly an application of Lemma 7.5. However, to bound X_{k+1} we need to also account for particles that are outside the ball $B(o, R_k)$ at time t_k, or particles that exit the ball $B(o, R_{k+1})$ at some time before time t_{k+1} and re-enter it. These quantities can be bounded in terms of Y_k and X_k as well as auxiliary quantities introduced soon. A delicate point in the proof is that the number of steps of the induction is not fixed ($\ln \ln m$), so we make sure that constants do not grow with m. Thus all constants below depend only on d.

At time t_0 our assumptions are that $Y_0 = 0$ and $X_0 \leq C_0 m^d$. For the induction step we define two additional quantities: S_k and Z_k. Let S_k be the number of particles that remain in $B(o, R_k)$ throughout the time interval $[t_{k-1}, t_k]$. We wish to apply Lemma 7.5 to S_k. The conditions are clearly satisfied (recall $S_k \leq C_0 m^d < C_0 R_k^d$, since $\gamma > 1$). Since $t_k - t_{k-1} \geq ct_k$ this will imply that with high probability

$$S_k < C_1 \frac{R_k^d}{t_k} \qquad \text{for all } k \leq K. \tag{52}$$

(The probability of failure at each of $\log \log m$ steps is exponentially small.)

Let Z_k be the number of particles located inside $B(o, R_k)$ at time t_k that exit $B(o, R_{k+1})$ before time t_{k+1}. Lemma 7.6, bounds the escape probability for each of X_k particles inside $B(o, R_k)$. Coalescence can only reduce the number of escaping particles, so given X_k,

$$Z_k \preceq Z_k' = \text{Bin}\left(X_k, C_2 \exp\left(-c_2 \frac{(R_{k+1} - R_k)^2}{t_{k+1} - t_k}\right)\right).$$

Here c_2, C_2 depend only on d, and "$\preceq$" denotes stochastic domination.

If $k = 0$ this implies

$$\mathbb{E}Z_0 \leq C_0 m^d C_2 e^{-c_2 \gamma^2 m^2 / t_1}.$$

Since $m^2/t_1 \asymp \ln m$, by making γ large enough we obtain $\mathbb{E}Z_0 = o(1)$. Then in particular $Z_0 = 0$ with probability tending to 1. For $k \geq 1$ we use Lemma 7.7, with $\Delta = K^4$ to find

$$\mathbb{P}\left(\bigcup_{k=1}^{K} \left\{Z_k' \geq K^4 + 2\mathbb{E}Z_k'\right\}\right) \leq K/\sqrt{K^4} = 1/K.$$

Thus with probability at least $1 - K^{-1} \to 1$ (as $m \to \infty$), we have

$$Z_k < K^4 + 2C_2 X_k \exp\left(-c_2 \frac{(R_{k+1} - R_k)^2}{t_{k+1} - t_k}\right) \qquad \text{for all } k.$$

Using $t_{k+1} = et_k$, $k \geq 1$, together with $\sqrt{t_{k+1}(K-k)} \geq \sqrt{e/2}\sqrt{t_k(K+1-k)}$, we conclude that with high probability, $Z_0 = 0$ and

$$Z_k < K^4 + 2C_2 X_k e^{-c_3\gamma^2(K-k)} \qquad \text{for all } k \leq K. \tag{53}$$

With these preparations in place, we are ready for the induction. Assume from here on that $Z_0 = 0$ and (52), (53) hold. We have, for each k, the deterministic bounds

$$X_k \leq S_k + Y_{k-1} + Z_{k-1},$$
$$Y_k \leq Y_{k-1} + Z_{k-1}.$$

To see this, note that particles in $B(o, R_k)$ either stayed inside (S), started outside (Y), or exited and returned (Z). The bound on Y is similar.

We now carry out an induction over k to bound X_k, Y_k for all $k = 1, \ldots, K$. Suppose that the bound (from the event A_m) on X_j and Y_j hold for all $j < k$. It follows from (53) and the inductive hypothesis that

$$Y_k \leq \sum_{j<k} Z_j < kK^4 + 2C_2\alpha \sum_{1\leq j<k} \frac{R_j^d}{t_j} e^{-c_3\gamma^2(K-j)}$$

$$< kK^4 + 2C_2\alpha \frac{R_k^d}{t_k} \sum_{j<k} \frac{t_k}{t_j} e^{-c_3\gamma^2(K-j)}$$

$$= kK^4 + 2C_2\alpha \frac{R_k^d}{t_k} e^{-c_3\gamma^2(K-k)} \sum_{j<k} e^{(1-c_3\gamma^2)(k-j)}.$$

We require γ to be large enough that $c_3\gamma^2 > 2$. Then the last sum is at most $1/(e-1) < 1$ and so

$$Y_k < kK^4 + 2C_2\alpha \frac{R_k^d}{t_k} e^{-c_3\gamma^2(K-k)}. \tag{54}$$

This proves the induction step for Y_k with any choice of $\beta > 2C_2\alpha$ (since $kK^4 \leq K^5 \ll R_k^d/t_k$, the last inequality coming from the fact that $K = \lfloor \log\log m \rfloor$, while $t_k \geq t_1 \asymp m^2/\ln m$ and $R_k \leq R_K \leq 2\gamma m$), on the event from (53), for all sufficiently large m.

It remains to bound X_k, for which we will use the bounds on S_k, Y_{k-1} and Z_{k-1}. We already have

$$Y_{k-1} < (k-1)K^4 + 2C_2\alpha \frac{R_{k-1}^d}{t_{k-1}} e^{-c_3\gamma^2(K+1-k)} < (k-1)K^4 + 2eC_2\alpha \frac{R_k^d}{t_k} e^{-c_3\gamma^2}.$$

Using the induction hypothesis and (53) (with $k - 1$ replacing k) one finds

$$Z_{k-1} < K^4 + 2C_2\alpha \frac{R_{k-1}^d}{t_{k-1}} e^{-c_3\gamma^2(K+1-k)} < K^4 + 2eC_2\alpha \frac{R_k^d}{t_k} e^{-c_3\gamma^2}.$$

Thus we have

$$X_k \leq S_k + Y_{k-1} + Z_{k-1} < C_1 \frac{R_k^d}{t_k} + kK^4 + 4eC_2\alpha \frac{R_k^d}{t_k} e^{-c_3\gamma^2}$$

$$= kK^4 + \left(C_1 + 4eC_2\alpha e^{-c_3\gamma^2}\right) \frac{R_k^d}{t_k}. \qquad (55)$$

To finish the proof it remains to select α and γ (and $\beta > 2C_2\alpha$) so that

$$C_1 + 4eC_2\alpha e^{-c_3\gamma^2} < \alpha.$$

This is done by requiring γ to satisfy $e^{c_3\gamma^2} > 4eC_2$ and taking any sufficiently large α.

Turning to the case $d = 2$, we proceed along the same lines. Lemma 7.5 gives with probability with high probability

$$S_k < C_1 \frac{R_k^2 \log m}{t_k} \qquad \text{for all } k \leq K. \qquad (56)$$

The failure probability is $\log^{-1} m$ at each of $\log\log m$ steps. Note that $\log t_i \sim \log m$ for all i, so we are not giving much away here. Furthermore, w.h.p. $Z_0 = 0$ and (53) holds (the proof of these facts does not depend on d).

We now repeat the induction. Given (53) and the induction hypothesis bounds on X we get (as above, with an extra $\log m$ factor)

$$Y_k < kK^4 + 2C_2\alpha \frac{R_k^2 \log m}{t_k} e^{-c_3\gamma^2(K-k)}.$$

Since the bounds for S_k, Y_{k-1} and Z_{k-1} differ from the general case only by a $\log m$ on the R_k^2/t_k term, the bound for X_k gets the same factor as well. □

References

1. Aldous, D.J.: Deterministic and stochastic models for coalescence (aggregation and coagulation): a review of the mean-field theory for probabilists. Bernoulli **5**(1), 3–48 (1999)

2. Arratia, R.: Limiting point processes for rescalings of coalescing and annihilating random walks on $\mathbf{Z}^d$. Ann. Probab. **9**(6), 909–936 (1981)
3. Berestycki, J., Berestycki, N., Limic, V.: The Λ-coalescent speed of coming down from infinity. Ann. Probab. (2010)
4. Berestycki, J., Berestycki, N., Schweinsberg, J.: Beta-coalescents and continuous stable random trees. Ann. Probab. **35**(5), 1835–1887 (2007)
5. Berestycki, J., Berestycki, N., Schweinsberg, J.: Small-time behavior of beta-coalescents. Ann. Inst. H. Poincaré Probab. Statist. **44**(2), 214–238 (2008)
6. Berestycki, N.: Recent progress in coalescent theory. Ensaios Matematicos **16**, 1–193 (2010)
7. Berg, J.v.d., Kesten, H.: Asymptotic density in a coalescing random walk model. Ann. Probab. **28**(1), 303–352 (2000)
8. Berg, J.v.d., Kesten, H.: Randomly coalescing random walks in dimension $\geq$ 3. In: Sidoravicius, V. (ed.) In and Out of Equilibrium, Volume 51 of Progress in Probability, pp. 1–45. Birkhauser, Boston (2002)
9. Bertoin, J.: Random Fragmentation and Coagulation Processes, Volume 102 of Cambridge Studies in Advanced Mathematics. Cambridge University Press, Cambridge (2006)
10. Bertoin, J., Le Gall, J.-F.: Stochastic flows associated to coalescent processes. III. Limit theorems. Illinois J. Math. **50**(1–4):147–181 (electronic) (2006)
11. Birkner, M., Blath, J., Capaldo, M., Etheridge, A., Möhle, M., Schweinsberg, J., Wakolbinger, A.: Alpha-stable branching and beta-coalescents. Electron. J. Probab. **10**(9), 303–325 (electronic) (2005)
12. Bolthausen, E., Sznitman, A.-S.: On Ruelle's probability cascades and an abstract cavity method. Commun. Math. Phys. **197**(2), 247–276 (1998)
13. Bovier, A., Kurkova, I.: Much ado about Derrida's GREM. In: Bolthausen, E., Bovier, A. (eds.) Spin Glasses, vol. 1900 of Lecture Notes in Mathematics, pp. 81–113. Springer, Berlin (2007)
14. Bramson, M., Griffeath, D.: Asymptotics for interacting particle systems on $\mathbf{Z}^d$. Z. Wahrsch. Verw. Gebiete **53**(2), 183–196 (1980)
15. Brunet, E., Derrida, B., Simon, D.: Universal tree structures in directed polymers and models of evolving populations. Phys. Rev. E **78**, 061102 (2008)
16. Cox, J., Durrett, R.: The stepping stone model: new formulae expose old myths. Ann. Appl. Probab. **12**, 1348–1377 (2002)
17. Cox, J.T., Griffeath, D.: Diffusive clustering in the two dimensional voter model. Ann. Probab. **14**, 347–370 (1986)
18. Durrett, R.: Probability Models for DNA Sequence Evolution. 2nd edn. Springer, New York (2008)
19. Erdös, P., Taylor, S.J.: Some problems concerning the structure of random walk paths. Acta Math. Acad. Sci. Hungar. **11**, 137–162 (1960)
20. Greven, A., Limic, V., Winter, A.: Coalescent processes arising in the study of diffusive clustering. http://arxiv.org/abs/math/0703875
21. Greven, A., Limic, V., Winter, A.: Representation theorems for interacting moran models, interacting fisher-wrighter diffusions and applications. Electron. J. Probab. **7**, 74 (2005)
22. Hammond, A., Rezakhanlou, F.: Kinetic limit for a system of coagulating planar Brownian particles. J. Stat. Phys. **124**(2–4), 997–1040 (2006)
23. Hammond, A., Rezakhanlou, F.: The kinetic limit of a system of coagulating Brownian particles. Arch. Ration. Mech. Anal. **185**(1), 1–67 (2007)
24. Kesten, H.: The number of alleles in electrophoretic experiments. Theoret. Pop. Biol. **18**(2), 290–294 (1980)
25. Kesten, H.: The number of distinguishable alleles according to the ohta-kimura model of neutral mutation. J. Math. Biol. **10**, 167–187 (1980)
26. Kimura, M.: "Stepping stone" model of population genetics. Ann. Rept. Nat. Inst. Genetics. Jpn **3**, 62–63 (1953)
27. Kimura, M., Weiss, G.: The stepping stone model of population structure and the decrease of genetic correlations with distance. Genetics **49**, 561–576 (1964)
28. Kingman, J.F.C.: The coalescent. Stochastic Process. Appl. **13**(3), 235–248 (1982)
29. Kingman, J.F.C.: On the genealogy of large populations. J. Appl. Probab., Special vol. **19A**:27–43 (1982) (Essays in statistical science)
30. Lawler, G.F., Limic, V.: Random Walk: a modern introduction. Book in preparation. http://www.math.uchicago.edu/~lawler/books.html

31. Liggett, T.M.: Interacting particle systems, volume 276 of Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]. Springer, New York (1985)
32. Limic, V., Sturm, A.: The spatial Λ-coalescent. Electron. J. Probab. **11**(15), 363–393 (electronic) (2006)
33. Merle, M.: Hitting probability of a distant point for the voter model started with a single one. Ann. Probab. **36**(3), 807–861 (2008)
34. Pitman, J.: Coalescents with multiple collisions. Ann. Probab. **27**(4), 1870–1902 (1999)
35. Pitman, J.: Combinatorial stochastic processes, volume 1875 of Lecture Notes in Mathematics. Springer, Berlin, 2006. Lectures from the 32nd Summer School on Probability Theory held in Saint-Flour, July 7–24 (2002)
36. Schweinsberg, J.: Coalescent processes obtained from supercritical Galton-Watson processes. Stochastic Process. Appl. **106**(1), 107–139 (2003)
37. Spitzer, F.: Principles of Random Walk, volume 34 of Graduate Texts in Mathematics. 2nd edn. Springer, New York (1976)
38. Talagrand, M.: The Parisi formula. Ann. Math. **16**(1), 221–263 (2006)
39. Zähle, I., Cox, J.T., Durrett, R.: The stepping stone model II: Genealogies and the infinite sites model. Ann. Appl. Probab. **15**, 671–699 (2005)