The Annals of Probability
2026, Vol. 54, No. 3, 1258–1319
https://doi.org/10.1214/25-AOP1789

SPATIAL PARTICLE PROCESSES WITH COAGULATION: GIBBS-MEASURE APPROACH, GELATION, AND SMOLUCHOWSKI EQUATION

BY LUISA ANDREIS[1,a], WOLFGANG KÖNIG[2,b], HEIDE LANGHAMMER[3,c] AND ROBERT I. A. PATTERSON[3,d]

[1]*Dipartimento di Matematica "Giuseppe Peano", Università degli Studi di Torino,* [a]*luisa.andreis@unito.it*
[2]*Weierstrass Institute for Applied Analysis and Stochastics Berlin and Institute of Mathematics, Technische Universität Berlin,* [b]*wolfgang.koenig@wias-berlin.de*
[3]*Weierstrass Institute for Applied Analysis and Stochastics Berlin,* [c]*heide.langhammer@gmx.de,* [d]*riap@cantab.net*

We study a spatial Markovian *particle system with pairwise coagulation*, a spatial version of the Marcus–Lushnikov process: according to a *coagulation kernel* K, particle pairs merge into a single particle, and their masses are united. We introduce a *statistical-mechanics approach* to the study of this process. We derive an explicit formula for the empirical process of the particle configuration at a given fixed time T in terms of a reference Poisson point process, whose points are trajectories that coagulate into one particle by time T. The noncoagulation between any two of them induces an exponential pair-interaction, which turns the description into a *many-body system with a Gibbsian pair-interaction*.

Based on this, we first give a *large-deviation principle* for the joint distribution of the particle histories (conditioning on an upper bound for particle sizes), in the limit as the number N of initial atoms diverges and the kernel scales as $\frac{1}{N}K$. We characterise the minimiser(s) of the rate function, we give criteria for its uniqueness and prove a *law of large numbers* (unconditioned). Furthermore, we use the unique minimiser to construct a solution of the *Smoluchowski equation* and give a criterion for the occurrence of a *gelation phase transition*.

CONTENTS

Received January 2024; revised September 2024.
MSC2020 subject classifications. Primary 82C22, 60J25, 60F10; secondary 60G55, 60K35, 35Q70.
Key words and phrases. Spatial coagulation process, spatial Marcus–Lushnikov process, empirical measures of particles, coagulation trajectories, monodispersed initial condition, Gibbsian representation, gelation phase transition, Smoluchowski equation, large deviations.

1. Introduction.

In this paper we investigate a Markovian process of spatially distributed particles that coagulate pairwise. Our main interest is in asymptotics and phase transitions in the limit of many particles. To investigate this, we adopt a typical statistical mechanics approach: we describe each particle at time T as a trajectory in space-time that lead to the formation of such particle (via the multiple coagulation events), and we write the joint distribution of all the particles as a many-body system with a pair-interaction. Starting from this, we apply the tools from large deviation theory to study the limits, and we use the results to derive other properties of the process. This section serves as an introduction to the work. In particular, in Section 1.1 we summarise the goals of this paper. In Section 1.2 we introduce the model, a spatial Markovian coagulation process. Then we turn in Section 1.3 to an explanation of our approach, which lies in a decomposition of the process into distinct parts, each leading to one of the single particles that can be observed at a fixed time.

1.1. *Summary of our results.*

The model. We consider a spatial version of the *Marcus–Lushnikov model* for coagulating particles. Each particle has a mass (an integer) and a location (a point in a Polish space $\mathcal{S}$). We start with a *monodispersed* configuration, where each particle is an atom with mass one. Coagulations of pairs occur independently over the pairs after exponential random holding times, whose parameters depend only on the locations and masses of the two particles and are assumed symmetric in the two particles. The collection of all these parameters forms what is called the *coagulation kernel*, $K : (\mathcal{S} \times \mathbb{N})^2 \to [0, \infty)$. The mass of the new particle is the sum of the two old ones, its location is picked randomly according to a *placement kernel*, Υ. This leads to a continuous time Markov process whose dynamics are determined by K and Υ.

Basic properties and the process Ξ. The total amount of mass in the system is kept constant, but all the particle sizes are nondecreasing, and the number of particles decreases by one at each coagulation event. After each coagulation event, the parameters of the exponential random times are updated to accommodate the pairwise interaction between the newly formed particle and all other ones. This process is a special case of what is called a *cluster coagulation process* in [38]. One can write the configuration of the process at time t as a point measure Ξ_t, which registers the statistics of the particles according to their positions and masses. Hence, Ξ_t lies in the set $\mathcal{M}_{\mathbb{N}_0}(\mathcal{S} \times \mathbb{N})$ of all point measures on $\mathcal{S} \times \mathbb{N}$, and the natural state space of the process $(\Xi_t)_{t\in[0,T]}$ is the set Γ_T of measure valued trajectories $[0, T] \to \mathcal{M}_{\mathbb{N}_0}(\mathcal{S} \times \mathbb{N})$ that are piecewise constant and are such that in each jump two particles are lost and one (larger) particle is gained.

Initial condition. In the present paper, we restrict to an initial distribution of atoms (i.e., the monodispersed situation of single-atom particles), taken as a Poisson process with intensity measure $N\mu$ for some probability measure μ, as the answers that we find are particularly transparent; the study of the model under deterministic initial atom configurations is deferred to future work.

The overall goals. First, we are interested in the distribution of $(\Xi_t)_{t\in[0,T]}$, where T is fixed. Then we study the behaviour of $(\frac{1}{N}\Xi_t)_{t\in[0,T]}$ in the limit of large total mass N in the system, when the coagulation kernel K is replaced by $\frac{1}{N}K$. This is often called the *hydrodynamic limit*. Here we would like to understand the limiting distribution of $(\frac{1}{N}\Xi_t)_{t\in[0,T]}$ and the question of a *gelation phase transition*, that is, the emergence of large particles (i.e., with diverging size depending on N) containing in total a macroscopic amount of atoms. Furthermore, we want to derive a characteristic partial differential equation for the limit.

Our new approach. While all the contributions to the analysis of the Marcus–Lushnikov model that we are aware of use the generator of the process, martingale arguments and characteristic equations, like the Smoluchowski and the Flory equation for finding answers (see the literature survey in Section 3.2), our approach is fundamentally different: it follows patterns that are known from statistical physics, large-deviations analysis and variational calculus.

Our contributions. Our main contributions are the following:

(1) We derive an explicit formula for the distribution of $(\Xi_t)_{t\in[0,T]}$ in terms of an *interacting Poisson point process*, whose points are the histories (trajectories) of each particle present at time T.

(2) We prove a *large-deviation principle* for $(\frac{1}{N}\Xi_t)_{t\in[0,T]}$ in the limit $N\to\infty$ of a large total mass $\asymp N$ in the system (for K replaced by $\frac{1}{N}K$).

(3) We give criteria for the occurrence or nonoccurrence of large particles with macroscopic total atom mass (i.e., a *gelation phase transition*) using properties of the large-deviation rate function under bounds on the coagulation kernel. In particular, we give criteria for the convergence of $(\frac{1}{N}\Xi_t)_{t\in[0,T]}$ as $N\to\infty$.

(4) As another by-product of our large-deviation analysis, we derive that in the subcritical regime, that is, when there is no gelation, the process $(\frac{1}{N}\Xi_t)_{t\in[0,T]}$ converges as $N\to\infty$ to a solution of a *spatial version of the famous Smoluchowski equation*.

History trees. Our ansatz for our goal (1) and for everything that follows is a decomposition of the entire configuration process $(\Xi_t)_{t\in[0,T]}$ on the time interval $[0,T]$ into the trajectories that lead to the particles that can be observed at time T. For this we introduce a labelled version of the process, where each particle present at time T is described as an element C of a (random) partition P_T. The set C contains the labels of the atoms that coagulated during $[0,T]$ into precisely one particle at time T and induces a (sub)process $\Xi^{(T,C)}$ that only registers the evolution of atoms with labels in C. We call these subprocesses *history trees* of the particles; they are characterised by terminating in just one particle. In Figure 1 we illustrate the decomposition. We register the history trees via their *empirical measure*

$$\mathcal{V}_N^{(T)}=\frac{1}{N}\sum_{C\in P_T}\delta_{\Xi^{(T,C)}}. \tag{1.1}$$

This object is much more comprehensive and detailed than just the process $(\Xi_t)_{t\in[0,T]}$. All the randomness of the process (holding times and placement decisions) is attached to these one-particle trajectories (coagulation events that lead to the particle) and their mutual interaction (the noncoagulation between the trajectories).

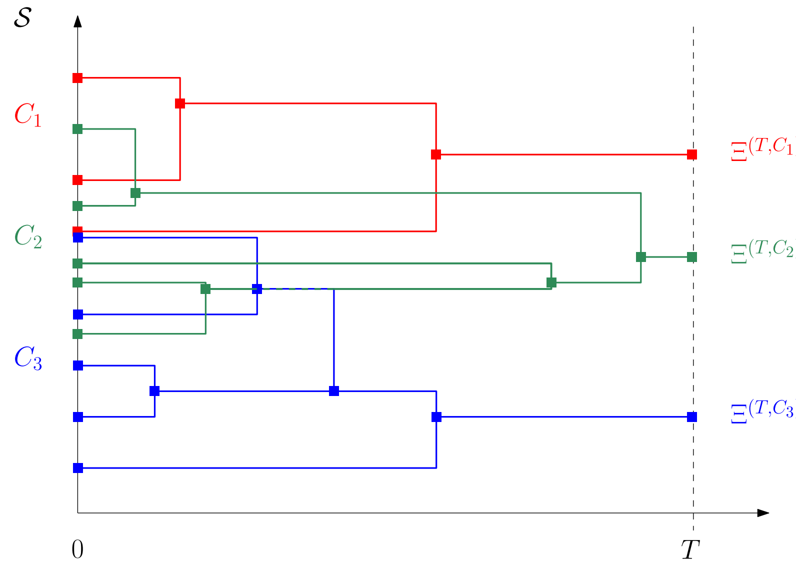

FIG. 1. *An illustration of the decomposition of* $(\Xi_t)_{t\in[0,T]}$ *into three subprocesses* $(\Xi_t^{(T,C_i)})_{t\in[0,T]}$, $i = 1, 2, 3$, *that are distinguished by their colour.*

A crucial Poisson process. Indeed, in our first main result, Theorem 2.1, we introduce a *Poisson point process* (PPP), $Y = \sum_i \delta_{\Xi_i}$. whose points are history trees. The intensity measure of Y is, up to a multiple, equal to the distribution of the coagulation process restricted to history trees. The probability of the event that any two of them do not coagulate during $[0, T]$ is expressed in terms of a pairwise interaction of the form

$$e^{-\frac{1}{2}\sum_{i\neq j} R^{(T)}(\Xi_i,\Xi_j)}.$$

In other words, we write the distribution of the configuration of the coagulation process as an expectation over many independent history trees with an exponential interaction term expressing the noncoagulations. This gives our representation the structure of a many-body system (better: a many-trajectory system): a *Gibbsian ensemble* of many independent trajectories with exponential pair interaction. This turns the Markovian coagulation process into a static model of statistical mechanics with the underlying reference measure as the law of a PPP on the set of history trees.

Large-deviation principle. Let us turn now to our goal (2). We use our Poissonian representation for deriving asymptotics of the empirical process $\mathcal{V}_N^{(T)}$ for diverging system size N and rescaled coagulation kernel $\frac{1}{N}K$ with the help of *large-deviations analysis*, one of the ubiquitous approaches to Gibbsian systems. In particular, in our case a useful large-deviations principle (LDP) for the reference PPP Y is known, and it has an explicit rate function. Since the state space of $\mathcal{V}_N^{(T)}$ is huge, we need to employ a conditioning on an event that generates compactness and continuity properties of certain functionals. This conditioning basically excludes the occurrence of large particles, which rules out the possibility to observe the gelation phase transition right away. In this conditional setting, we obtain in our second main result, Theorem 2.3, a full LDP for $\mathcal{V}_N^{(T)}$ and an explicit formula for the rate function of the coagulation process.

Consequences: *Euler–Lagrange equation and convergence*. An important innovation of our approach is the description of the limit of the process via an Euler–Lagrange equation. This can be seen as an alternative formulation of the famous Smoluchowski equation in terms of a variational problem. Starting with the rate function of the LDP, we prove that every minimiser of the rate function (i.e., every possible accumulation point of $(\mathcal{V}_N^{(T)})_{N\in\mathbb{N}}$ under the conditional measure) satisfies the Euler–Lagrange equation and formulate assumptions under which the solutions in turn are unique, employing an argument known from Banach's fixed point theorem. One obviously wishes to remove the conditioning, but unfortunately, it cannot be removed on an exponential scale. Nevertheless, we succeed in formulating assumptions under which we can prove that the probability of the conditional event converges to one, and we derive tightness of the unconditioned distribution of $\mathcal{V}_N^{(T)}$ and the convergence $\mathcal{V}_N^{(T)}$ toward the minimiser of the large-deviation rate function. This is our third main result, Theorem 2.8, 1, together with Proposition 2.10, 1.

Our assumptions. For the LDP we work under the sole condition that the coagulation kernel K is continuous and that there exists $H < \infty$ such that

$$K\big((x,m),(x',m')\big) \le Hmm' \quad \text{for all } (x,m),(x',m') \in \mathcal{S}\times\mathbb{N}. \tag{1.2}$$

This assumption is equivalent to the upper bound that Norris uses in the definition of an *approximately multiplicative* kernel in [38], where it is needed to obtain existence and uniqueness of the *cluster coagulation equation*.

Gelation and nongelation. Our criteria for nongelation will require that T is sufficiently small with respect to the constant H from assumption (1.2). The results about convergence of $\mathcal{V}_N^{(T)}$ hold, for example, if $TH < \frac{1}{e^2}\frac{\pi}{1+\pi}$. Apart from that, more abstract criteria will be given. To formulate criteria for gelation, we additionally need the "dual"condition that there exists $h > 0$ such that

$$K\big((x,m),(x',m')\big) \ge hmm' \quad \text{for all } (x,m),(x',m') \in \mathcal{S}\times\mathbb{N}. \tag{1.3}$$

We prove in Theorem 2.8, 2, jointly with Proposition 2.10, 2, the occurrence of a gel for all sufficiently large T. We define gelation via the existence of large particles (with a diverging size that can be on any scale) that build a macroscopic part of the configuration (the gel), but do not specify the scale of the size of the large particles. The existence of such particles is identified by the fact that, when considering the total mass of atoms sitting in all the particles at time T of sizes $\le L$, after letting $N\to\infty$, followed by $L\to\infty$, we see that the mass is strictly less than the total mass at time 0. In future work we plan to analyse also the macroscopic and mesoscopic part of the configuration to obtain a more detailed understanding of the gelation phase transition and the gel itself.

Results for the ML-process. So far, all our results are formulated in terms of the empirical trajectory measure, $\mathcal{V}_N^{(T)}$, but one is also highly interested in the Marcus–Lushnikov process Ξ itself. While the process $(\Xi_t)_{t\in[0,T]}$ takes values in the set Γ_T of trajectories $[0,T]\to\mathcal{M}_{\mathbb{N}_0}(\mathcal{S}\times\mathbb{N})$, the empirical process $\mathcal{V}_N^{(T)}$ is contained in a much more comprehensive and more abstract space. Nevertheless, $(\frac{1}{N}\Xi_t)_{t\in[0,T]}$ is a relatively simple functional of $\mathcal{V}_N^{(T)}$, which turns out in Lemma 2.5 to be continuous in a sufficient sense, such that a great deal of our results about $\mathcal{V}_N^{(T)}$ have a consequence for $(\frac{1}{N}\Xi_t)_{t\in[0,T]}$.

LDP for the ML-process. Our first observation is in Corollary 2.6 that, via the well-known contraction principle, $(\frac{1}{N}\Xi_t)_{t\in[0,T]}$ also satisfies an LDP (under the same conditioning as $\mathcal{V}_N^{(T)}$) with an explicit rate function. This nice fact opens up a completely new path to LDPs for the trajectories of the Marcus–Lushnikov process, which is totally different from and independent of the much-used Freidlin–Wentzell theory that uses a toolbox from operator theory and stochastic processes (see the literature survey in Section 3.2). It appears to have the great advantage to be successful also in the current setting of a pretty abstract and huge state space, in contrast with existing works using the mentioned toolbox.

The Smoluchowski equation. Furthermore, and now we are turning to our goal (4) of deriving a partial differential equation for the limiting objects, by the continuity of the map that maps $\mathcal{V}_N^{(T)}$ onto $(\frac{1}{N}\Xi_t)_{t\in[0,T]}$, we also obtain, under the mentioned assumptions, the convergence of the latter to a deterministic process, $\rho = (\rho_t)_{t\in[0,T]}$. From the Euler–Lagrange equation for the minimisers of our rate function, we derive in Lemma 2.12 that ρ is a solution to a natural spatial version of the famous Smoluchowski equation on $[0, T]$. In our approach this equation is only a by-product and provides a nice additional information but is not a vital part of the proof of convergence of $(\frac{1}{N}\Xi_t)_{t\in[0,T]}$, as it is in many previous investigations of related processes.

Advantages and disadvantages. Let us briefly discuss advantages and disadvantages of our ansatz in comparison to earlier ansatzes; see Section 3.2 for an extensive literature comparison. We present the first LDP for the spatial ML-process. Previous results have focused exclusively on the convergence to solutions of the Smoluchowski equation. In [38] this convergence is shown to occur exponentially fast, which suggests that an LDP might hold. However, the result in [38] applies solely to a special case (we conduct an extensive comparison with [38] in Section 3.2). A byproduct of our large-deviations analysis is that we characterize the limits of the spatial ML-process through a variational problem, offering a new interpretation of the solutions to the Smoluchowski equation. From this variational problem, we derive a lower bound on the gelation time that is comparable to the results in [38]. We obtain an upper bound on this time as well, which is worse than the known one from [38]; however, the latter is also not sharp, as it does not account for the role of the spatial component.

A uniqueness criterion. Let us finally mention (see Remark 2.7) that our approach gives also uniqueness of the minimiser of our rate function under the condition that the kernel K is nonnegative definite; however, we believe that this criterion is not too helpful, since it is in general difficult to check this property.

1.2. *The spatial Marcus–Lushnikov process.* Let us enter into the details of our model. As mentioned, particles live in a Polish space $\mathcal{S}$, that is, a separable, complete metric space. Each particle carries a mass $m \in \mathbb{N}$ and sits at a site $x \in \mathcal{S}$; we sometimes also say then that the particle sits at (x, m). Initially, each particle has mass one; that is, we consider what is usually called a monodisperse initial condition. The unit-mass particles at time zero are also called *atoms*, since all later particles are composed out of them by merging and since they are never split anymore into smaller units. The particle process is a Markov process in continuous time. The dynamics of the process depends on a *coagulation kernel* and a *placement kernel*,

$$K\colon (\mathcal{S}\times\mathbb{N})^2 \to [0,\infty) \quad \text{and} \quad \Upsilon\colon (\mathcal{S}\times\mathbb{N})^2\times\mathcal{B}(\mathcal{S}) \to [0,1], \tag{1.4}$$

where $\mathcal{B}(\mathcal{S})$ denotes the Borel-σ-field on $\mathcal{S}$. We assume that K is symmetric and measurable and that Υ is a Markov kernel from $(\mathcal{S}\times\mathbb{N})^2$ into $\mathcal{S}$, that is, measurable in the argument $(\mathcal{S}\times\mathbb{N})^2$ and a probability measure in the last argument. We also assume that Υ is symmetric in the first two arguments, that is, $\Upsilon((x,m),(x',m'),\cdot) = \Upsilon((x',m'),(x,m),\cdot)$.

THE RANDOM MECHANISM. At each time $t \in [0,\infty)$, for each unordered pair of particles in the current configuration, located at x and $x' \in \mathcal{S}$ with masses m and m', respectively, there is an exponential random time with parameter $K((x,m),(x',m'))$ running. When it elapses, the pair is replaced by a single particle with mass $m+m'$, located at a random site that is picked according to $\Upsilon((x,m),(x',m'),\cdot)$. The exponential random times and the placement locations are independent over all particle pairs and over all times.

Hence, after each coagulation event, the parameters of all exponential random times involving any of the two coagulating particles are updated. Since we are starting with only finitely many particles, this excludes explosion and the total number of coagulation events during a time interval $[0, T]$ is finite. Hence, the entire process can be decomposed into finitely many time intervals during which the configuration remains constant. (The process that we study is the special case of what is called a *cluster coagulation process* in [38], with state space $E = \mathcal{S} \times \mathbb{N}$ and mass-preserving function $m((x, m)) = m$.)

The above mechanism defines a Markov chain in continuous time

$$\Xi = (\Xi_t)_{t\in[0,\infty)}, \quad \text{where } \Xi_t = \sum_{i\in[n(t)]} \delta_{(X_i(t),M_i(t))} \in \mathcal{M}_{\mathbb{N}_0}(\mathcal{S} \times \mathbb{N}), \tag{1.5}$$

where $(X_i(t), M_i(t)) \in \mathcal{S} \times \mathbb{N}$ is the location and the mass of the ith particle at time t, and $n(t)$ is the number of particles at time t (we abbreviate $[n] = \{1, \dots, n\}$). We denote by $\mathcal{M}_{\mathbb{N}_0}$ the set of measures with values in $\mathbb{N}_0$, that is, the set of finite point measures. We will treat elements $\phi \in \mathcal{M}_{\mathbb{N}_0}(\mathcal{S} \times \mathbb{N})$ as partially discrete measures and we will write, for example, $\phi(x, m)$ instead of $\phi(\{(x, m)\})$. We write also $\Xi_t(A, m)$ for the number of particles in $A \subset \mathcal{S}$ with mass $m \in \mathbb{N}$ at time t in the configuration.

The process $(n(t))_{t\in[0,\infty)}$ is nonincreasing in t; it actually decreases by one at every coagulation time. We include also the trivial case $n(t) = 0$ for all t in which case there is no atom and hence the process is empty. As usual for point processes, the index $i \in [n(t)]$ is arbitrary and does not specify the ith particle. In particular, if there are multiple particles with the same location and mass, then Ξ does not give information about which one of them are involved in a coagulation.

The process Ξ is a Markov process with jumps of type

$$\phi \mapsto \phi - \delta_{(x,m)} - \delta_{(x',m')} + \delta_{(z,m+m')}, \quad \phi \in \mathcal{M}_{\mathbb{N}_0}(\mathcal{S} \times \mathbb{N}) \tag{1.6}$$

(as long as the right-hand side is nonnegative) that happen with rate

$$\begin{aligned} \mathbf{K}_\phi\big((x,m),(x',m'),\mathrm{d}z\big) &= \mathbf{K}\big((x,m),(x',m'),\mathrm{d}z\big) \\ &\quad \times \begin{cases} \phi(x,m)\phi(x',m') & \text{if } (x,m) \neq (x',m'), \\ \phi(x,m)\big(\phi(x,m)-1\big)/2 & \text{otherwise,} \end{cases} \end{aligned} \tag{1.7}$$

and we abbreviated

$$\mathbf{K}\big((x,m),(x',m'),\mathrm{d}z\big) = K\big((x,m),(x',m')\big)\Upsilon\big((x,m),(x',m'),\mathrm{d}z\big). \tag{1.8}$$

The counting factor in the second line of (1.7) is the number of unordered pairs of particles of types (x, m) and (x', m'). In the case that $\mathcal{S}$ is uncountable, formally, there are uncountably many (x, x') that give rise to a step, but actually only finitely many have a positive rate, since ϕ has a finite support.

We always start with a monodispersed situation; that is, we fix $n(0) \in \mathbb{N}$, $M_i(0) = 1$ for any $i \in [n(0)]$ and some configuration of $(X_i(0))_{i\in[n(0)]}$. Hence, $\Xi_0 = \sum_{i\in[n(0)]} \delta_{(X_i(0),1)}$, and we often identify Ξ_0 with the configuration $\sum_{i\in[n(0)]} \delta_{(X_i(0))} \in \mathcal{M}_{\mathbb{N}_0}(\mathcal{S})$. The trajectories of Ξ lie in the set

$$\begin{aligned} \Gamma = \big\{\xi = (\xi_t)_{t\in[0,\infty)} \in \mathbb{D}\big(\mathcal{M}_{\mathbb{N}_0}(\mathcal{S} \times \mathbb{N})\big) &: \xi_0 \text{ is concentrated on } \mathcal{M}_{\mathbb{N}_0}\big(\mathcal{S} \times \{1\}\big) \\ &t \mapsto \xi_t \text{ is piecewise constant and makes steps as in (1.6)}\big\}, \end{aligned} \tag{1.9}$$

where $\mathbb{D}(\mathcal{X})$ denotes the set of càdlàg-paths $[0, \infty) \to \mathcal{X}$, that is, paths that are right-continuous in $[0, \infty)$ and have left limits everywhere in $(0, \infty)$ (we give $\mathcal{M}_{\mathbb{N}_0}(\mathcal{S} \times \mathbb{N})$ the weak topology induced by integrals against continuous bounded test functions). In particular, the initial configuration of any $\xi \in \Gamma$ is finite, and the collected mass of all particles at time t, which we indicate with $\|\xi_t\|_1 = \sum_{m\in\mathbb{N}} m\xi_t(\mathcal{S}\times\{m\})$, is a constant in t.

For any $k \in \mathcal{M}_{\mathbb{N}_0}(\mathcal{S})$, we write $\mathbb{P}_k$ and $\mathbb{E}_k$ for probability and expectation with respect to the process Ξ when started from $\Xi_0 = k$. For $k = 0$, we define $\mathbb{P}_0$ as the measure that is concentrated on the constant zero point measure.

Initial configurations. In this paper we will formulate and prove all our results for the distribution, $\mathbb{P}_{\mathrm{Poi}_{N\mu}}$, of the process under poissonised initial conditions so that

$$\mathbb{P}_{\mathrm{Poi}_{N\mu}} = \int_{\mathcal{M}_{\mathbb{N}_0}(\mathcal{S})} \mathrm{Poi}_{N\mu}(\mathrm{d}k)\mathbb{P}_k \tag{1.10}$$

where $\mathrm{Poi}_{N\mu}$ is the law of a Poisson point process (PPP) on $\mathcal{S}$ with intensity measure $N\mu$ for an arbitrary probability measure μ on $\mathcal{S}$. Note that the number $n(0)$ of initial particles is not deterministic but is Poi_N-distributed, and the empty configuration appears with probability $\mathrm{e}^{-N} > 0$. The analogous statements and proofs for the deterministic initial configuration $\mathbb{P}_{k_N}$ with $k_N \in \mathcal{M}_{\mathbb{N}_0}(\mathcal{S})$ satisfying $\frac{1}{N}k_N \to \mu$ for some $\mu \in \mathcal{M}_1(\mathcal{S})$ are deferred to future work.

REMARK 1.1 (Special choices). A natural choice of Υ is the deterministic choice $z = \frac{xm+x'm'}{m+m'}$ (if $\mathcal{S}$ is convex), which keeps the centre of mass of the two particles (and hence also the centre of mass of the entire configuration) constant. Another one is the random one, where z is put equal to x or x' with probability $\frac{m}{m+m'}$ and $\frac{m'}{m+m'}$, respectively; this choice keeps the centre of mass fixed on average. An important special case in the nonspatial setting (i.e., $\mathcal{S}$ is a singleton) is the product kernel $K(m, m') = Hmm'$ for some $H \in (0, \infty)$. Then the coagulation process can be mapped one-to-one onto the process of the family of connected component sizes of the well-known Erdős–Rényi graph in its dynamic version. We go into the details of this connection with graphs in Section 3.3.

1.3. *Decomposition into history trees.* We will describe the distribution of $(\Xi_t)_{t\in[0,T]}$ for a fixed $T > 0$ via a decomposition into subprocesses that coagulate into one particle by time T, which we will call *history trees* on the interval $[0, T]$. This section provides the necessary framework to properly define the empirical measure of trajectories $\mathcal{V}_N^{(T)}$ introduced in Section 1.1.

Observe that the notation in (1.5) only counts particles at a given site, that is, with a given location and mass. It is not rich enough to differentiate between multiple particles that sit on the same site and can, therefore, not express information about which particles coagulate into which particle. As a consequence, it cannot grasp the full history of a particle. In order to express the evolution explicitly we need to introduce an alternative version of the coagulation process, that assigns a label to every atom at time 0 such that we can follow its path through the coagulation process. To this end, we define a Markov process $(P_t)_{t\in[0,\infty)}$ on the set of all partitions of $[n(0)]$ starting from $P_0 = \{\{i\} : i \in [n(0)]\}$, together with the locations of the particles (elements of the partition).

As a preparation we need some notation. For any finite, nonempty set A, we denote the set of partitions of A by

$$\mathcal{P}(A) = \Big\{\{C_j\}_{j\in J} : J \text{ an index set}, \dot{\bigcup}_{j\in J} C_j = A, \forall j : \varnothing \neq C_j \subset A\Big\}. \tag{1.11}$$

We define a process $(P_t)_{t\in[0,\infty)}$ on $\mathcal{P}(A)$ in such a way that $t \mapsto P_t$ makes discrete steps by joining two of the partition sets and is otherwise constant; in particular, $t \mapsto |P_t|$ decreases by one in each step. We refer to each partition set $C \in P_t$ as a particle that exists at time t and attach to it the site $X_C^{(t)} \in \mathcal{S}$ at which the particle sits. More precisely, we define the *labelled coagulation process*

$$Z = (Z_t)_{t\in[0,\infty)}, \quad \text{with } Z_t = \big(X_C^{(t)}, C\big)_{C\in P_t}, \tag{1.12}$$

as a Markov process with the mechanism

$$\begin{aligned}&\big((X_C, C), (X_D, D)\big)\\ &\quad\mapsto (X_{C\cup D}, C\cup D) \quad \text{with rate } \mathbf{K}\big((X_C, |C|), (X_D, |D|), \mathrm{d}X_{C\cup D}\big),\end{aligned} \tag{1.13}$$

where our short-hand notation $((X_C, C), (X_D, D)) \mapsto (X_{C\cup D}, C\cup D)$ includes the fact that for all $\widetilde{C} \notin \{C, D\}$ the values $(X_{\widetilde{C}}, \widetilde{C})$ are unchanged in the transition. In contrast to Ξ, the process Z contains for every particle C present at time t the information from which of the atoms at time 0 it stems. Given $\mathbf{x} = (x_i)_{i\in A} \in \mathcal{S}^A$, we denote by $\mathbb{P}_\mathbf{x}$ the distribution of this Markov process starting from $Z_0 = (X_C^{(0)}, C)_{C\in P_0} = (x_i, \{i\})_{i\in A}$. We call $\mathbf{x}$ the *initial atom configuration* of the process.

Now we fix a time $T \in (0,\infty)$. For each particle present at time T, we want to define a subprocess that describes the history of the particle. Now let $(Z_t)_{t\in[0,T]}$ with $Z_t = (X_C^{(t)}, C)_{C\in P_t}$ be the process defined in (1.12) on the time interval $[0, T]$. Fix a set C such that either $C \in P_T$ (which is the case we consider most of the time) or such that C can be written as a union of sets in P_T. Define

$$\Xi^{(T,C)} = \big(\Xi_t^{(T,C)}\big)_{t\in[0,T]}, \quad \text{where } \Xi_t^{(T,C)} = \sum_{\widetilde{C}\in P_t :\, \widetilde{C}\subset C} \delta_{(X_{\widetilde{C}}^{(t)}, |\widetilde{C}|)} \in \mathcal{M}_{\mathbb{N}_0}(\mathcal{S}\times\mathbb{N}). \tag{1.14}$$

One can easily check that, as a consequence of the mechanism (1.13), the sum on $\widetilde{C} \in P_t$ satisfying $\widetilde{C} \subset C$ is nonempty. $\Xi^{(T,C)}$ is the (sub)process that only keeps track of the atoms with labels in C, that is, $\Xi_t^{(T,C)}$ is the number of particles with a given location and mass at time t that have emerged from atoms with labels in C. In particular, it holds that $\Xi_0^{(T,C)} = \sum_{i\in C} \delta_{(X_{\{i\}}^{(0)}, 1)}$. In the following we mainly choose $C \in P_T$ (i.e., C is one particle at time T), and in that case we note that $\Xi^{(T,C)}$ is an element of the *set of history trees*,

$$\Gamma_T^{(1)} = \big\{\xi \in \Gamma_T : \xi_T(\mathcal{S}\times\mathbb{N}) = 1\big\}, \tag{1.15}$$

where

$$\Gamma_T = \pi_{[0,T]}^{-1}(\Gamma), \quad \text{where } \pi_{[0,T]}(\xi) = (\xi_t)_{t\in[0,T]}, \tag{1.16}$$

is the set of trajectories on the time interval $[0, T]$. Thus, if $C \in P_T$, then $\Xi^{(T,C)}$ tracks history of the particle $(X_C^{(T)}, |C|)$, ending in a single-particle configuration $\delta_{(X_C^{(T)}, |C|)}$ at time T. Recall Figure 1, where we illustrate how the process $(\Xi_t)_{t\in[0,T]}$ decomposes into the subprocesses $\Xi^{(T,C)}$, $C \in P_T$.

Now we introduce the main object of this paper: the normalised empirical measure of the history trees of all the particles present at time T,

$$\mathcal{V}_N^{(T)} = \frac{1}{N} \sum_{C\in P_T} \delta_{\Xi^{(T,C)}} \in \mathcal{M}\big(\Gamma_T^{(1)}\big). \tag{1.17}$$

For the definition of $\mathcal{V}_N^{(T)}$, we need the process $(Z_t)_{t\in[0,T]}$ and need to work a priori on the extended probability space with measure $\mathbb{P}_\mathbf{x}$. However, observe that $\mathcal{V}_N^{(T)}$ does not depend on the labels, even though its definition uses the labels in order to decompose the process into its subprocesses $\Xi^{(T,C)}$, $C \in P_T$. More precisely, assume that we fix $k \in \mathcal{M}_{\mathbb{N}_0}(\mathcal{S})$ and some vector $\mathbf{x} = (x_i)_{i=1}^{|k|} \in \mathcal{S}^A$ that is *compatible* with k, that is, it satisfies $k = \sum_{i=1}^{|k|} \delta_{x_i}$.

Then $\mathbb{P}_{\mathbf{x}}(\mathcal{V}_N^{(T)} \in \cdot)$ does not depend on the labelling given by $\mathbf{x}$, but only on k. Hence, in a small abuse of notation, we will consider $\mathcal{V}_N^{(T)}$ under the measure $\mathbb{P}_k$ and mean its distribution under $\mathbb{P}_{\mathbf{x}}$ for some $\mathbf{x}$ that is compatible with k. Hence, we can assume that Ξ and $\mathcal{V}_N^{(T)}$ are defined on the same probability space with measure $\mathbb{P}_k$. In particular, we will speak of its distribution under the Poisson initial configuration, $\mathbb{P}_{\mathrm{Poi}_{N\mu}}$, and mean the distribution with an initial configuration vector whose length is Poi_N-distributed and has i.i.d. μ-distributed entries.

Let us explain how we can obtain the process $(\frac{1}{N}\Xi_t)_{t\in[0,T]}$ from $\mathcal{V}_N^{(T)}$. Let us denote the projection $\pi_t \colon \Gamma_T \to \mathcal{M}_{\mathbb{N}_0}(\mathcal{S} \times \mathbb{N})$ by $\pi_t(\xi) = \xi_t$, and let us write $\nu_t = \nu \circ \pi_t^{-1}$ for $\nu \in \mathcal{M}(\Gamma_T)$ and $t \in [0, T]$. Then, denoting

$$\Xi_t(Z) = \sum_{\widetilde{C}\in P_t} \delta_{(X_{\widetilde{C}}^{(t)}, |\widetilde{C}|)}, \tag{1.18}$$

we have that

$$\begin{aligned}
\frac{1}{N}\Xi_t &= \frac{1}{N}\Xi_t(Z) \\
&= \frac{1}{N}\sum_{\widetilde{C}\in P_t} \delta_{(X_{\widetilde{C}}^{(t)}, |\widetilde{C}|)} \\
&= \frac{1}{N}\sum_{C\in P_T}\sum_{\widetilde{C}\in P_t \colon \widetilde{C}\subset C} \delta_{(X_{\widetilde{C}}^{(t)}, |\widetilde{C}|)} \\
&= \frac{1}{N}\sum_{C\in P_T} \Xi_t^{(T,C)} \\
&= \int \mathcal{V}_N^{(T)}(\mathrm{d}\xi)\xi_t,
\end{aligned} \tag{1.19}$$

where the first equality can be checked easily by comparing the transition rates of the two processes. The connection in (1.19) is crucial, since it shows that $(\frac{1}{N}\Xi_t)_{t\in[0,T]}$ is a function of $\mathcal{V}_N^{(T)}$ and allows us to understand the dynamics of $(\frac{1}{N}\Xi_t)_{t\in[0,T]}$ by studying $\mathcal{V}_N^{(T)}$.

The organisation of the remainder of this paper is as follows. Section 2 presents the main results of this paper, which we discuss and compare to the literature in Section 3. In Section 4 we prove the first main result (identification of the distribution); in Section 5 we prepare for the proofs of the others, which we complete in Sections 6 (large-deviation principle) and Section 7 (convergence and gelation criteria, and the Smoluchowski equation). Technical parts of the proofs are collected in the Appendix.

2. Results. In this section we formulate and comment our four main results: a representation of the empirical measure of history trees in Section 2.1, a large-deviation principle for this empirical measure in Section 2.2, criteria for the occurrence of the gelation phase transition and convergence in Section 2.3, and the validity of the Smoluchowski equation for the limit in Section 2.4.

First, let us introduce some notation that we use throughout the paper. For any measure $\mathfrak{m}$ on some measure space $\mathcal{X}$ and any measurable function f on $\mathcal{X}$, we write $\langle \mathfrak{m}, f\rangle$ for the integral of f with respect to the measure $\mathfrak{m}$. We write $|\mathfrak{m}| = \mathfrak{m}(\mathcal{X})$ for the total mass of $\mathfrak{m}$. We denote by $\mathcal{M}_1(\mathcal{X})$ the set of probability measures on $\mathcal{X}$. For measures $v \in \mathcal{M}(\mathcal{S} \times \mathbb{N})$, we use the fact that v is partially discrete and write $v(\mathrm{d}x, m)$ instead of $v(\mathrm{d}(x, m))$ and define $\|v\|_1 = \sum_{m\in\mathbb{N}} \int_{\mathcal{S}} v(\mathrm{d}x, m)m$.

Considering Poisson point processes, we use the following notation. For $\gamma \in (0,\infty)$, we denote with Poi_γ the Poisson distribution on $\mathbb{N}_0$ with parameter γ, and for a measure $\mu \in \mathcal{M}(\mathcal{S})$, we denote with $\mathrm{Poi}_\mu \in \mathcal{M}_1(\mathcal{M}_{\mathbb{N}_0}(\mathcal{S}))$ the distribution of a Poisson point process on $\mathcal{S}$ with intensity measure μ. By this we mean a finite random collection $\sum_i \delta_{r_i}$ of points $r_i \in \mathcal{S}$ such that $\#\{i : r_i \in A\}$ has the distribution $\mathrm{Poi}_{\mu(A)}$ for any measurable $A \subset \mathcal{S}$ and is independent over mutually disjoint sets A. Similarly, for any measure M on $\Gamma_T^{(1)}$, by Poi_M, we denote a finite collection $\sum_i \delta_{\Xi_i}$ of point measures on nonempty point configurations $\Xi_i \in \Gamma_T^{(1)}$ such that $\#\{i : \Xi_i \in B\}$ is Poisson-distributed with parameter $M(B)$ for any measurable $B \subset \Gamma_T^{(1)}$.

2.1. *First main result*: *Identification of the distribution*. In this section we present our first main result: an identification of the distribution of the empirical measure $\mathcal{V}_N^{(T)}$ of the history trees of the particles present at time T. This is in terms of a Poisson point process (PPP) on $\Gamma_T^{(1)}$ and an exponential pair-interaction term. We fix $\mu \in \mathcal{M}_1(\mathcal{S})$ and will be considering the Marcus–Lushnikov process under the poissonised initial condition $\mathbb{P}_{\mathrm{Poi}_{N\mu}}$.

In the following the reference measure, $M_{\mu,N}^{(T)} \in \mathcal{M}(\Gamma_T^{(1)})$ will play an important role, which we define as

$$M_{\mu,N}^{(T)}(\mathrm{d}\xi) = N^{|k|-1}\mathrm{e}\mathbb{P}_{\mathrm{Poi}_\mu}(\Xi \in \mathrm{d}\xi), \quad \xi \in \Gamma_T^{(1)}, k = \xi_0, \tag{2.1}$$

that is, the restriction of $\mathbb{P}_{\mathrm{Poi}_\mu}$ to $\Gamma_T^{(1)}$ with some factors that turned out convenient. Note that $M_{\mu,N}^{(T)}$ does not weight empty configurations: while Poi_μ gives positive measure to $k = 0$, the restriction of $\mathbb{P}_0$ to $\Gamma_T^{(1)}$ is the zero measure, since $\Gamma_T^{(1)}$ does not contain the zero point measure trajectory. Furthermore, the total mass of $M_{\mu,N}^{(T)}$ is finite, since $\mathbb{P}_k$ has mass ≤ 1 on $\Gamma_T^{(1)}$ and Poi_μ has exponential moments of all orders.

We now introduce a reference PPP $Y_N = \sum_{i\in I} \delta_{\Xi_i}$ on $\Gamma_T^{(1)}$ with intensity measure $NM_{\mu,N}^{(T)}$. We write expectation with respect to Y_N as $\mathrm{E}_{NM_{\mu,N}^{(T)}}$.

Let us express the probability of noncoagulation between any two history trees for a fixed time interval $[0, T]$ for fixed $T > 0$. In Lemma 4.1 we prove that, for any two history trees $\xi, \xi' \in \Gamma_T$, the probability that they do not coagulate takes the form $\mathrm{e}^{-R^{(T)}(\xi,\xi')}$, where we have the explicit formula

$$R^{(T)}(\xi, \xi') = \int_0^T \mathrm{d}t \langle \xi_t, K\xi_t' \rangle, \quad \xi, \xi' \in \Gamma_T, \tag{2.2}$$

and we write $K\phi(x,m) = \int_{\mathcal{S}} \sum_{m\in\mathbb{N}} \phi(\mathrm{d}x', m') K((x,m),(x',m'))$ for any $(x,m) \in \mathcal{S} \times \mathbb{N}$ and any $\phi \in \mathcal{M}_{\mathbb{N}_0}(\mathcal{S} \times \mathbb{N})$.

In our first main result, we assume nothing else than what we stated around (1.4).

THEOREM 2.1 (Poissonian description of the empirical measure). *Fix $\mu \in \mathcal{M}_1(\mathcal{S})$ and $T > 0$ and $N \in \mathbb{N}$ and a measurable bounded function $f : \mathcal{M}(\Gamma_T^{(1)}) \to [0,\infty)$. Then*

$$\mathbb{E}_{\mathrm{Poi}_{N\mu}}\big(f(\mathcal{V}_N^{(T)})\big) = \mathrm{E}_{NM_{\mu,N}^{(T)}}\left[\mathrm{e}^{-\frac{1}{2}\sum_{i,j:\, i\neq j} R^{(T)}(\Xi_i,\Xi_j)} f\left(\frac{1}{N} Y_N\right)\right]\mathrm{e}^{N(|M_{\mu,N}^{(T)}|-1)}, \tag{2.3}$$

where $Y_N = \sum_i \delta_{\Xi_i} \sim \mathrm{Poi}_{NM_{\mu,N}^{(T)}}$ is a PPP with intensity measure $NM_{\mu,N}^{(T)}$.

The proof of Theorem 2.1 is in Section 4. Let us give an interpretation of the formula: the distribution of $\mathcal{V}_N^{(T)}$ is given via the distribution of $\frac{1}{N} Y_N$ with the additional density term

$\exp\{-\frac{1}{2}\sum_{i\neq j} R^{(T)}(\Xi_i, \Xi_j)\}$ that takes into account the noncoagulation between any pair of history trees.

Actually, Theorem 2.1 gives the coagulation process the meaning of a spatial *interacting many-body system* with Gibbsian interaction; see Remark 3.1. We underline that this is not a prerogative of the system when starting from a Poisson initial condition, as an analogue representation can be obtained for any initial condition. However, since in the following we focus only on Poisson initial condition, we omit it here and defer the treatment of such setting to future work.

2.2. *Second main result*: *A large-deviations principle*. Here is our second main result, which is about asymptotics as $N \to \infty$. In view of our Gibbsian representation in Theorem 2.1, limiting statements are most naturally approached in terms of a large-deviation principle for the empirical history tree measure, as we will formulate in Theorem 2.3. We consider initial configurations distributed as $\mathrm{Poi}_{N\mu}$ for some $\mu \in \mathcal{M}_1(\mathcal{S})$ such that N is the order of the number of particles at time zero. In order to see interesting phenomena, we replace now the kernel K by $\frac{1}{N}K$. We indicate this by adding an additional superscript $^{(N)}$. Having done this change, the formula for the particle distribution from Theorem 2.1 receives a structure that is exponential in N; hence, any kind of limiting assertions as $N \to \infty$ based on that formula will naturally involve a large-deviation principle (LDP). It is known that the concept of an LDP depends strongly on the topologies used, and we will be working on sets $\mathcal{Y} = \mathcal{M}(\mathcal{X})$ of bounded Borel measures on a Polish space $\mathcal{X}$, equipped with the weak topology; see Section A.1 for specific remarks on our particular case.

For all limiting assertions in this paper, we will be under the assumption formulated in (1.2). Note that product kernels of the form $K((x,m),(x',m')) = \varphi(x,x')mm'$ satisfy (1.2) if φ is bounded.

In preparation for formulating our main result, we formulate the convergence of the reference measure $M^{(T)}_{\mu,N}$, defined in (2.1), with K replaced by $\frac{1}{N}K$. In a slight abuse, we use the notation $\mathbb{Q}|_A$ for the restriction of a measure $\mathbb{Q}$ to a measurable set A, that is, $\mathbb{Q}|_A(B) = \mathbb{Q}(A \cap B)$. Both measures that we are introducing in the next lemma are crucial for our LDP.

LEMMA 2.2 (Convergence of $N^{|k|-1}\mathbb{P}^{(N)}_k(\Xi \in \cdot)|_{\Gamma^{(1)}_T}$). *Assume that the kernel K satisfies* (1.2). *Fix $\mu \in \mathcal{M}_1(\mathcal{S})$ and $T \in (0,\infty)$ and $k \in \mathcal{M}_{\mathbb{N}_0}(\mathcal{S})$. Replace the coagulation kernel K by $\frac{1}{N}K$, and write $\mathbb{P}^{(N)}_k$ for the probability measure. Then the following limiting measure exists in the weak sense*:

$$\mathbb{Q}^{(T)}_k(\cdot) = \lim_{N\to\infty} N^{|k|-1}\mathbb{P}^{(N)}_k(\Xi \in \cdot)|_{\Gamma^{(1)}_T} \in \mathcal{M}(\Gamma^{(1)}_T). \tag{2.4}$$

In particular, the measure $M^{(T,N)}_{\mu,N}$, defined in (2.1) *w.r.t.* $\mathbb{P}^{(N)}$, *converges toward the measure*

$$M^{(T)}_\mu = \mathrm{e}\,\mathrm{Poi}_\mu \otimes \mathbb{Q}^{(T)} \in \mathcal{M}(\Gamma^{(1)}_T). \tag{2.5}$$

In (2.5) we used the well-known measure-theoretic notation

$$\mathrm{Poi}_\mu \otimes \mathbb{Q}^{(T)}(\mathrm{d}(k,\xi)) = \mathrm{Poi}_\mu(\mathrm{d}k)\mathbb{Q}^{(T)}_k(\mathrm{d}\xi), \qquad k \in \mathcal{M}_{\mathbb{N}_0}(\mathcal{S}), \xi \in \Gamma^{(1)}_T \text{ s.t. } \xi_0 = k. \tag{2.6}$$

The proof of Lemma 2.2 is in Section 5.1. An explicit formula for $\mathbb{Q}^{(T)}_k$ and more information appear in Lemma 5.2. Informally speaking, $\mathbb{Q}^{(T)}_k$ assigns to a trajectory ξ the product of the rates over all coagulation events but drops all terms coming from the exponential densities of the coagulation times. It is an important reference measure for our further analysis.

As we have already said in Section 1.1, in the present paper we study only the microscopic part of the process, that is, those particles that are of finite-order size in N. We leave the study of other particles, with size growing in N, to future work. Note that the space of history trees $\Gamma_T^{(1)}$ does include the history of particles with arbitrarily large sizes. For this reason we encounter the problem of lack of compactness in $\mathcal{M}(\Gamma_T^{(1)})$, the state space of $\mathcal{V}_N^{(T)}$, and we are forced to condition on a set that induces compactness. Fix any $\beta > 0$ and some function $f : \mathbb{N} \to [0, \infty)$ that grows at infinity faster than linear, that is, $f(r)/r \to \infty$ as $r \to \infty$, and satisfes that $f(r) \geq r$ for all $r \in \mathbb{N}$. Define

$$\mathcal{A}_{f,\beta} = \left\{ \nu \in \mathcal{M}(\Gamma_T^{(1)}) : \int_{\Gamma_T^{(1)}} \nu(\mathrm{d}\xi) f(|\xi_0|) \leq \beta \right\}, \tag{2.7}$$

and note that for any $\xi \in \Gamma_T^{(1)}$ it holds that its collected mass $\|\xi_t\|_1 \in \mathbb{N}$ is constant in $t \in [0, T]$ and equal to $|\xi_0|$, the number of initial atoms of ξ. Hence, the condition in $\mathcal{A}_{f,\beta}$ is an higher-moment integrability condition on the sizes/masses of the particles at time T. On the event $\{\mathcal{V}_N^{(T)} \in \mathcal{A}_{f,\beta}\}$, we will be able to derive compactness/continuity of important objects that allow for a smooth proof of the LDP (see Sections A.1 and A.2). See Remark 3.4 below for an explanation that, on this event, the configuration cannot develop a macroscopically large particle by time T in the limit $N \to \infty$, and so the conditioning is more than a purely technical step. Since $f(r) \geq 1$ for all $r \in \mathbb{N}$, $|\nu| \leq \beta$ for all $\nu \in \mathcal{A}_{f,\beta}$. We will later mainly use $f(r) = r^2$.

We are going to introduce the rate function of our LDP.

Recall from (2.2) the noncoagulation functional $R^{(T)}$, and introduce the operator with kernel $R^{(T)}$,

$$\mathfrak{R}^{(T)}(\nu)(\xi) = \int_{\Gamma_T^{(1)}} R^{(T)}(\xi, \xi')\nu(\mathrm{d}\xi'), \quad \xi \in \Gamma_T^{(1)}, \nu \in \mathcal{M}(\Gamma_T^{(1)}). \tag{2.8}$$

See Section A.2 for a number of properties of $\mathfrak{R}^{(T)}$.

We fix $\mu \in \mathcal{M}_1(\mathcal{S})$, and we define the function $I_\mu^{(T)} : \mathcal{M}(\Gamma_T^{(1)}) \to [0, \infty]$ by

$$I_\mu^{(T)}(\nu) = \begin{cases} \left\langle \nu, \log \dfrac{\mathrm{d}\nu}{\mathrm{d}M_\mu^{(T)}} \right\rangle + \dfrac{1}{2}\langle \nu, \mathfrak{R}^{(T)}(\nu)\rangle + 1 - |\nu| & \text{if } \nu \ll M_\mu^{(T)}, \\ \infty & \text{otherwise.} \end{cases} \tag{2.9}$$

Here is our main result: an LDP for the collection $\mathcal{V}_N^{(T)}$ of all the components of the coagulation process with N particles and kernel $\frac{1}{N}K$, restricted to the event that no infinite component appears in the limit.

THEOREM 2.3 (LDP for $\mathcal{V}_N^{(T)}$). *Assume that the kernel K is continuous and satisfies the upper bound in* (1.2). *Replace the kernel K by $\frac{1}{N}K$. Pick $T \in (0, \infty)$ and $\mu \in \mathcal{M}_1(\mathcal{S})$. Pick $f : \mathbb{N} \to [0, \infty)$ satisfying $\lim_{r\to\infty} f(r)/r = \infty$ and $f(r) \geq r$ for any r. Then for any $\beta > 0$, the distribution of $\mathcal{V}_N^{(T)}$ under $\mathbb{P}^{(N)}_{\mathrm{Poi}_{N\mu}}(\cdot|\mathcal{V}_N^{(T)} \in \mathcal{A}_{f,\beta})$ satisfies an LDP on $\mathcal{M}(\Gamma_T^{(1)})$ with speed N and rate function*

$$\mathcal{M}(\Gamma_T^{(1)}) \to [0, \infty], \qquad \nu \mapsto \begin{cases} I_\mu^{(T)}(\nu) - \chi_\beta & \textit{if } \nu \in \mathcal{A}_{f,\beta}, \\ \infty & \textit{otherwise}, \end{cases} \tag{2.10}$$

where $\chi_\beta = \inf_{\nu \in \mathcal{A}_{f,\beta}} I_\mu^{(T)}(\nu)$. The sublevel sets of this rate function are compact.

The proof of Theorem 2.3 is in Section 6.

A standard conclusion from Theorem 2.3 is the following.

COROLLARY 2.4 (Accumulation points). *In the situation of Theorem 2.3, the distribution of* $\mathcal{V}_N^{(T)}$ *under* $\mathbb{P}^{(N)}_{\mathrm{Poi}_{N\mu}}(\cdot|\mathcal{V}_N^{(T)}\in\mathcal{A}_{f,\beta})$ *is tight in* N, *and each limit point of this distribution along any subsequence is concentrated on the set of minimisers of* $I_\mu^{(T)}|_{\mathcal{A}_{f,\beta}}$.

We are not only interested in the empirical process $\mathcal{V}_N^{(T)}$ but also in the Marcus–Lushnikov process $(\Xi_t)_{t\in[0,T]}$ itself. Because of (1.19), this is a function of $\mathcal{V}_N^{(T)}$, more precisely, $(\frac{1}{N}\Xi_t)_{t\in[0,T]}=\rho(\mathcal{V}_N^{(T)})$, where we define

$$\begin{aligned}&\rho\colon \mathcal{M}\big(\Gamma_T^{(1)}\big)\to\mathbb{D}_T\big(\mathcal{M}(\mathcal{S}\times\mathbb{N})\big),\\ &\rho(\nu)=\big(\rho_t(\nu)\big)_{t\in[0,T]}=\bigg(\int_{\mathcal{M}(\Gamma_T^{(1)})}\nu(\mathrm{d}\xi)\xi_t\bigg)_{t\in[0,T]},\end{aligned}\tag{2.11}$$

where we write $\mathbb{D}_T$ for the set of càdlàg functions on $[0,T]$.

The continuity of the map ρ is handled in the following lemma. Details about the topologies are given in Section A.1.

LEMMA 2.5 (Continuity of $\nu\mapsto\rho(\nu)$). *Fix any* $\beta>0$ *and some function* $f:\mathbb{N}\to[0,\infty)$ *that grows at infinity faster than linear, that is,* $f(r)/r\to\infty$ *as* $r\to\infty$. *Let* $(\nu_n)_{n\in\mathbb{N}}$ *be a sequence in* $\mathcal{A}_{f,\beta}$ *that converges toward some* ν *such that* $I_\mu^{(T)}(\nu)<\infty$. *Then* $\rho(\nu_n)\to\rho(\nu)$ *as* $n\to\infty$.

The proof of Lemma 2.5 is in Section 5.2. As a consequence of this and the LDP of Theorem 2.3, we obtain also an LDP for the Marcus–Lushnikov process:

COROLLARY 2.6 (LDP for $(\frac{1}{N}\Xi_t)_{t\in[0,T]}$). *In the situation of Theorem 2.3, the distribution of* $(\frac{1}{N}\Xi_t)_{t\in[0,T]}$ *satisfies an LDP on* $\mathbb{D}_T(\mathcal{M}(\mathcal{S}\times\mathbb{N}))$ *with rate function*

$$\rho\mapsto\inf\big\{I_\mu^{(T)}(\nu)-\chi_\beta\colon\nu\in\mathcal{A}_{f,\beta},\rho(\nu)=\rho\big\}.$$

This immediately follows from the contraction principle, more precisely from Remark (c) on Theorem 4.2.1 in [19].

Let us close this section with a remark on a handy criterion for uniqueness of minimisers for the rate function $I_\mu^{(T)}$.

REMARK 2.7 (Convexity of $I_\mu^{(T)}$ by nonnegative definiteness of K). The map $\nu\mapsto\langle\nu,\mathfrak{R}^{(T)}(\nu)\rangle$ is a priori not convex. But under the additional assumption that K be nonnegative definite, it is. Then $I_\mu^{(T)}$ is strictly convex. Recall that

$$K\text{ is nonnegative definite}\iff\langle v,Kv\rangle\ge 0,\quad v\in\mathcal{M}_{\mathbb{R}}(\mathcal{S}\times\mathbb{N}),\tag{2.12}$$

where we denote by $\mathcal{M}_{\mathbb{R}}(\mathcal{S}\times\mathbb{N})$ the set of signed finite measures on $\mathcal{S}\times\mathbb{N}$. The convexity of the map under the assumption of nonnegative definiteness is clear from the fact that, for positive measures ν_1, ν_2 on $\Gamma_T^{(1)}$ and $\alpha\in(0,1)$,

$$\begin{aligned}&\big\langle\alpha\nu_1+(1-\alpha)\nu_2,\mathfrak{R}^{(T)}\big(\alpha\nu_1+(1-\alpha)\nu_2\big)\big\rangle\\ &\quad=\alpha\big\langle\nu_1,\mathfrak{R}^{(T)}(\nu_1)\big\rangle+(1-\alpha)\big\langle\nu_2,\mathfrak{R}^{(T)}(\nu_2)\big\rangle-\alpha(1-\alpha)\big\langle\nu_1-\nu_2,\mathfrak{R}^{(T)}(\nu_1-\nu_2)\big\rangle.\end{aligned}\tag{2.13}$$

Nonnegative definiteness yields a handy criterion of uniqueness of minimisers of $I_\mu^{(T)}$, but since it is difficult to check in practical examples, we will not rely on it.

2.3. *Third main result*: *Criteria for convergence and gelation*. One of the main questions in the Marcus–Lushnikov model is about the existence or nonexistence of a *gelation phase transition*. That is, the question about the existence of a deterministic critical time threshold $T_{\text{gel}} \in (0,\infty)$ such that there are only microscopically sized particles (i.e., particles of size of finite order, not depending on N) before time T_{gel}, and after this time, a positive fraction of the total mass (i.e., $\asymp N$) lies in large particles (i.e., particles of N-depending diverging size). If this phenomenon occurs, then we call T_{gel} the *gelation time*, the group of all the macroscopic particles the *gel* and the coagulation kernel K a *gelling kernel*. We stick to the convention that we use $\frac{1}{N}K$ instead of K (indicated by an additional superscript $^{(N)}$).

Let us coin a rigorous definition of the occurrence of gelation. We introduce the notation $\|v\|_{1,\le L} = \int_{\mathcal{S}} \sum_{m=1}^{L} m v(\mathrm{d}x, m)$ for $v \in \mathcal{M}(\mathcal{S}\times\mathbb{N})$; then $\|\Xi_T\|_{1,\le L}$ is the total amount of mass in particles with size $\le L$ at time T. Recall that $t \mapsto \|\Xi_t\|_1$ is constant under $\mathbb{P}^{(N)}_{\mathrm{Poi}_{N\mu}}$, and $\|\frac{1}{N}\Xi_t\|_1$ is equal to $\frac{1}{N}$ times a Poi_N-distributed random variable, that is, it converges to one as $N\to\infty$ almost surely and in L^1-sense. One can call the difference $\|\frac{1}{N}\Xi_T\|_1 - \|\frac{1}{N}\Xi_T\|_{1,\le L}$ the *L-gel* of the process at time T. Then

$$\mathrm{NG}^{(\mu)}_T = \lim_{L\to\infty} \limsup_{N\to\infty} \mathbb{E}^{(N)}_{\mathrm{Poi}_{N\mu}}\left[\left\|\frac{1}{N}\Xi_T\right\|_{1,\le L}\right] \tag{2.14}$$

is the limiting expected nongel mass, that is, the mass outside the gel. The map $T \mapsto \mathrm{NG}^{(\mu)}_T$ is nonincreasing with initial value $\mathrm{NG}^{(\mu)}_0 = 1$. If $\mathrm{NG}^{(\mu)}_T < 1$, then we say that there is a gel at time T, and we define the *gelation time* by

$$T^{(\mu)}_{\text{gel}} = \inf\{T \in (0,\infty) : \mathrm{NG}^{(\mu)}_T < 1\} \in [0,\infty]. \tag{2.15}$$

This is the time at which the gelation phase transition occurs, if it is finite. If $T^{(\mu)}_{\text{gel}} < \infty$, we also speak of the phenomenon of *loss of mass* and say that *gelation occurs*. The interpretation is that some positive fraction of all the atoms sits in particles of sizes that depend on N and diverge as $N\to\infty$ such that their total mass goes lost when looking only at particles of finite size, regardless how large this finite-size window is. We think this notion is (one of) the most natural notions of gelation and gelation times; see Section 3.2 for other notions of gelation used in the literature on coagulation processes.

Note that the total mass of the L-gel can also be expressed in terms of our process $\mathcal{V}^{(T)}_N$. Indeed, for any measure λ on point measures on $\mathcal{S}$, introduce its *atom-distribution measure*,

$$c_\lambda(\cdot) = \int_{\mathcal{M}_{\mathbb{N}_0}(\mathcal{S})} \lambda(\mathrm{d}k)k(\cdot) \in \mathcal{M}(\mathcal{S}), \quad \lambda \in \mathcal{M}(\mathcal{M}_{\mathbb{N}_0}(\mathcal{S})), \tag{2.16}$$

and its L-restriction $c^{(\le L)}_\lambda(\cdot) = \int \lambda(\mathrm{d}k)k(\cdot)\mathbb{1}\{|k|\le L\}$. Recall that $(\mathcal{V}^{(T)}_N)_0$ is a measure concentrated on $\mathcal{M}_{\mathbb{N}_0}(\mathcal{S}\times\{1\})$ and can hence be identified with a measure on $\mathcal{M}_{\mathbb{N}_0}(\mathcal{S})$, which allows us to write $c_{(\mathcal{V}^{(T)}_N)_0}$. Recalling the connection between $\mathcal{V}^{(T)}_N$ and $\frac{1}{N}\Xi_t$ from (1.19), we can write the non-L-gel of the process at time T as

$$\begin{aligned} \left|c^{(\le L)}_{(\mathcal{V}^{(T)}_N)_0}\right| &= \int_{\Gamma^{(1)}_T} \mathcal{V}^{(T)}_N(\mathrm{d}\xi)|\xi_0|\mathbb{1}\{|\xi_0|\le L\} \\ &= \int_{\Gamma^{(1)}_T} \mathcal{V}^{(T)}_N(\mathrm{d}\xi)\|\xi_T\|_{1,\le L} = \left\|\frac{1}{N}\Xi_T\right\|_{1,\le L}. \end{aligned} \tag{2.17}$$

Especially from the second expression, it is clearly seen that this quantity is a continuous functional of $\mathcal{V}^{(T)}_N$ in the weak topology. This will be important in our proofs, since they deal with $\mathcal{V}^{(T)}_N$.

Introduce

$$q_\mu^{(T)} = \limsup_{n\to\infty} \big(M_\mu^{(T)}(\{\xi \in \Gamma_T^{(1)} : |\xi_0| = n\})\big)^{1/n} \in (0,\infty). \tag{2.18}$$

This quantity controls the finiteness of all the moments of $|\xi_0|$ under the reference measure $M_\mu^{(T)}$; the threshold for the finiteness of all moments is $q_\mu^{(T)} = 1$.

Here is our main result on the existence and nonexistence of gelation. For all our sufficient criteria for gelation, we will assume that the lower bound from (1.3) holds. Like our main upper bound in (1.2), one main example for a kernel that satisfies (1.3) is $K((x,m),(x',m')) = \varphi(x,x')mm'$ with φ bounded away from zero.

THEOREM 2.8 (Criteria for nongelation and for gelation). *Fix $\mu \in \mathcal{M}_1(\mathcal{S})$ and $T \in (0,\infty)$, and assume that* (1.2) *holds.*

1. Criterion for nongelation: *Assume that $q_\mu^{(T)} < 1$. Then the following hold*:
   (i) *$I_\mu^{(T)}$ has compact sublevel sets and hence possesses minimisers.*
   (ii) $\mathrm{NG}_T^{(\mu)} = 1$, *that is, there is no gelation at time T.*
   (iii) *Any minimising $\nu^{(T)}$ satisfies the Euler–Lagrange equation*

$$\nu(\mathrm{d}\xi) = M_\mu^{(T)}(\mathrm{d}\xi)\mathrm{e}^{-\mathfrak{R}^{(T)}(\nu)(\xi)}, \quad \xi \in \Gamma_T^{(1)} : \tag{2.19}$$

   (iv) *The distributions of $\mathcal{V}_N^{(T)}$ and $c_{(\mathcal{V}_N^{(T)})_0}$ under $\mathbb{P}^{(N)}_{\mathrm{Poi}_{N\mu}}$ are tight in N.*
   (v) *Let $\mathbb{P}$ be a limit point of $\mathbb{P}^{(N)}_{\mathrm{Poi}_{N\mu}}(\mathcal{V}_N^{(T)} \in \cdot)$, $N \in \mathbb{N}$. Denote by $\mathcal{V}$ a random variable with distribution $\mathbb{P}$. Then $|c_{\mathcal{V}_0}| = 1$ almost surely with respect to $\mathbb{P}$.*

2. Criterion for gelation: *In addition to* (1.2), *assume that* (1.3) *holds and that* $\inf I_\mu^{(T)} > 0$. *Then gelation occurs, that is,*

$$\mathrm{NG}_T^{(\mu)} < 1.$$

The proof of Theorem 2.8 consists of several steps and is spread over the entire Section 7. The main argument for 1(ii) is in Section 7.3 and for 2 in Section 7.4.

REMARK 2.9 (Interpretation of the EL-equation). The EL-equation in (2.19) characterises minimisers ν of the rate function in terms of a self-referencing equation for the (nonnormalised) distribution ν on history trees: they are equal to a characteristic reference distribution $M_\mu^{(T)}$, with a term that weights the sampled history tree with the probability that it does not coagulate with another independent sample under the same distribution ν. Via (1.19), this can also be turned into a characteristic equation for all accumulation points of $(\frac{1}{N}\Xi_t)_{t\in[0,T]}$; see Section 2.4.

Now we state simple estimates on the kernel K that imply that the criteria from Theorem 2.8 are satisfied.

PROPOSITION 2.10 (Bounds on T that imply (non)gelation or convergence). *Fix $\mu \in \mathcal{M}_1(\mathcal{S})$ and $T \in (0,\infty)$, and assume that* (1.2) *holds.*

1. Criterion for nongelation:
   (i) *If $TH < \frac{1}{\mathrm{e}^2}$, then $q_\mu^{(T)} < 1$, and thus the statements from Theorem* 2.8, 1 *apply.*

(ii) *If* $TH < \frac{1}{\mathrm{e}^2}\frac{\pi}{1+\pi}$, *then* (2.19) *has at most one solution, and the distribution of* $(\frac{1}{N}\Xi_t)_{t\in[0,T]}$ *under* $\mathbb{P}^{(N)}_{\mathrm{Poi}_{N\mu}}$ *converges to* $(\int \nu^{(T)}(\mathrm{d}\xi)\xi_t)_{t\in[0,T]}$ *with* $\nu^{(T)}$ *the unique solution to* (2.19).

2. Criterion for gelation: *In addition to* (1.2), *assume that* (1.3) *holds. Then*

$$\inf_{\nu\in\mathcal{M}(\Gamma_T^{(1)})} I_\mu^{(T)}(\nu) \ge 1 - \frac{1}{2T}\left(\frac{\mathrm{e}}{\pi H} + \frac{(\log(2TH\mathrm{e}^2))^2}{h}\right), \tag{2.20}$$

that is, the criterion of Theorem 2.8, 2 *applies for all* T *such that the right-hand side is strictly positive.*

The proof of assertion 1 of Proposition 2.10 is at the end of Section 7.3, and the proof for 2 is in Section 7.2.

COROLLARY 2.11 (Bounds on the gelation time). *Under the assumption in* (1.2), $T_{\mathrm{gel}}^{(\mu)} \ge \frac{1}{H\mathrm{e}^2}$. *If additionally* (1.3) *is assumed, then*

$$T_{\mathrm{gel}}^{(\mu)} \le \inf\left\{T : \frac{1}{2T}\left(\frac{\mathrm{e}}{\pi H} + \frac{(\log(2TH\mathrm{e}^2))^2}{h}\right) < 1\right\} < \infty.$$

The correct interpretation of the loss of mass phenomenon in terms of a measure ν on $\Gamma_T^{(1)}$ is that the total mass of all the history trees decays in the limit $N\to\infty$ as a function of T. Indeed, $\nu(\mathrm{d}\xi)$ registers only those history trees that correspond to finite-size particles at time T. All the other (larger) particles that are present at time T do not appear in ν anymore, together with their respective history trees. The reason is that the state space $\Gamma_T^{(1)}$ cannot accommodate nonmicroscopic trees. In future work we plan to extend the description of the history trees by some enlarged space that is able to describe also the larger particles and their history. It seems as if it is not possible to formulate (not to mention, prove) an LDP or a law of large numbers without detailed knowledge about the large particles, since each of them has a nontrivial influence on the dynamics of the entire coagulation process after they appear.

2.4. *Fourth main result*: *The Smoluchowski equation*. Let us now come to the fourth main result of this paper. To introduce it, let us recall that the rigorous analysis of coagulation processes started in 1916 with a formulation and analysis of the—by now famous—Smoluchowski equation [44]. This is a partial differential equation for the evolution of the concentration of particles sitting in any site. It consists of a positive term that describes the formation of new particles in a certain site via coagulation of a pair of smaller particles and a negative term describing the loss of particles in that site due to coagulation with any other particle. In our situation, where we include space, we can formulate the equation in the form

$$\begin{aligned}
&\partial_t\rho_t(\mathrm{d}x^*, m^*)\\
&\quad= \frac{1}{2}\sum_{m,m'\in\mathbb{N}:\ m+m'=m^*}\int_{\mathcal{S}}\int_{\mathcal{S}}\rho_t(\mathrm{d}x,m)\rho_t(\mathrm{d}x',m')\mathbf{K}((x,m),(x',m'),\mathrm{d}x^*)\\
&\qquad - \rho_t(\mathrm{d}x^*,m^*)K\rho_t(x^*,m^*),\qquad x^*\in\mathcal{S}, m^*\in\mathbb{N}.
\end{aligned}\tag{2.21}$$

This equation and its variants play a fundamental role in the investigation of the limiting behaviour of the Marcus–Lushnikov process as well as being of interest in its own right as a deterministic coagulation model. Indeed, in previous probabilistic investigations, proofs of the convergence of the process $(\frac{1}{N}\Xi_t)_{t\in[0,\infty)}$ with kernel $\frac{1}{N}K$ often (if not always) follow the route that: (1) tightness arguments are employed, (2) it is shown that every accumulation point satisfies the Smoluchowski equation, and (3) criteria for the uniqueness of the solution are given; see Section 3.2. If the kernel is such that gelation occurs, then one can only expect this convergence before the gelation time, whereas the full process $(\frac{1}{N}\Xi_t)_{t\in[0,\infty)}$ has a limit that solves an extension of the Smoluchowski equation, called the Flory equation, which captures the influence of the gel on microscopic particles after the gelation time.

The main results of this paper on convergence and gelation so far have nothing to do with the Smoluchowski equation. Nevertheless, this equation is so important that we decided to show that it is satisfied by all limit points of the process $(\frac{1}{N}\Xi_t)_{t\in[0,T]}$, if T is small enough such that we have no gelation. The novelty here is that we derive it from our main characterization, the Euler–Lagrange equation in (2.19).

LEMMA 2.12 (The ML process converges to a solution to the Smoluchowski equation). *Assume that* $\mu\in\mathcal{M}_1(\mathcal{S})$ *and that* K *satisfies* (1.2) *and that* $TH<\frac{1}{e^2}\frac{\pi}{1+\pi}$. *Then, under* $\mathbb{P}^{(N)}_{\mathrm{Poi}_{N\mu}}$, *the process* $(\frac{1}{N}\Xi_t)_{t\in[0,T]}$ *converges to a solution* ρ *of the Smoluchowski equation in* (2.21).

The proof is in Section 7.5. We use Proposition 2.10 to ensure that we are in a regime, where gelation does not occur and to get convergence of $\mathcal{V}_N^{(T)}$ to the (unique) solution of the Euler–Lagrange equation (2.19).

3. Background discussion.

3.1. *Comments on the main results.* In this section we outline our key heuristics and explain the nature of some crucial difficulties in the proofs. We also highlight the benefits and shortcomings of our approach.

REMARK 3.1 (The coagulation process as a Gibbsian many-body system). In our first main theorem (Theorem 2.1), the expectation on the right-hand side of (2.3) is with respect to a Poissonian reference measure on point measures Y_N of history trees on $[0,T]$. The time marginal at time zero is the spatial distribution of the initial particles (atoms) that coagulate during $[0,T]$ into one particle, and the time marginal at time T is the particle distribution at the end. By the Poisson nature, these trajectories are a priori mutually independent, but they are under an exponential pair interaction term that expresses that they do not coagulate by time T. Note that this interaction is mutually repellent. Hence, the right-hand side is a many-body system of points in $\mathcal{S}$ with marks (the mark being their history tree) with Gibbsian pair interaction with a PPP as the underlying reference measure. Putting $f=1$, we can identify the last exponential term $(e^{N(|M^{(T)}_{\mu,N}|-1)})$ in terms of the first one and obtain the formula

$$\mathbb{E}_{\mathrm{Poi}_{N\mu}}\big(f(\mathcal{V}_N^{(T)})\big)=\frac{\mathrm{E}_{NM^{(T)}_{\mu,N}}[e^{-\frac{1}{2}\sum_{i,j:\,i\neq j}R^{(T)}(\Xi_i,\Xi_j)}f(\frac{1}{N}Y_N)]}{\mathrm{E}_{NM^{(T)}_{\mu,N}}[e^{-\frac{1}{2}\sum_{i,j:\,i\neq j}R^{(T)}(\Xi_i,\Xi_j)}]}. \tag{3.1}$$

This shows that $\mathcal{V}_N^{(T)}$ has the distribution of $\frac{1}{N}Y_N$ under exponential transformation with the density $e^{-\frac{1}{2}\sum_{i\neq j}R^{(T)}(\Xi_i,\Xi_j)}$, properly normalised.

REMARK 3.2 (Plausibility of the LDP). On the basis of the Poissonian description in Theorem 2.1, one can easily guess that there might be an LDP valid for $\mathcal{V}_N^{(T)}$ with rate function as in (2.9), after replacing K by $\frac{1}{N}K$ and hence $R^{(T)}$ by $\frac{1}{N}R^{(T)}$ in (2.3). The main point is that $\frac{1}{N}Y_N$ satisfies an LDP under $\mathrm{E}_{NM_{\mu,N}^{(T,N)}}$ with a rate function of entropy form, since $M_{\mu,N}^{(T,N)}$ converges weakly toward $M_\mu^{(T)}$. The interaction term in the exponent (the double-sum on $i \neq j$) is approximated by the sum on *all* i, j and directly leads to the double-integral of $\mathfrak{R}^{(T)}$ with respect to $\nu \otimes \nu$. Now collect all the exponential terms on the right-hand side of (2.3) to see that they lead directly to the formula for $I_\mu^{(T)}$ in (2.9).

REMARK 3.3 (Difficulties). The heuristics of Remark 3.2 suggest that $I_\mu^{(T)}$ governs an LDP for $(\mathcal{V}_N^{(T)})_{N\in\mathbb{N}}$ under $\mathbb{P}_{\mathrm{Poi}_{N\mu}}^{(N)}$ without any conditioning, but this is not true in general: A couple of problems arise when attempting to prove an unconditional LDP, and they turn out to be substantial issues, not merely technical difficulties. Indeed, these problems have a lot to do with the gelation phase transition that we discuss in Section 2.3. One problem is that the approximation of the sum on $i \neq j$ in the interaction term by the sum on all i, j (i.e., the addition of all the self-interactions) fails if the particles are too large, more precisely, if some of them are of a size proportional to N. Another problem is that we do not see an argument for compact sublevels sets of $I_\mu^{(T)}$ (not even for lower semicontinuity) on the entire set $\mathcal{M}(\Gamma_T^{(1)})$ if $|M_\mu^{(T)}| = \infty$; hence, existence of minimisers is not certain. A third problem is that we see a priori no argument for the fact that the infimum of $I_\mu^{(T)}$ over $\mathcal{M}(\Gamma_T^{(1)})$ should be equal to zero; in fact, we disprove it in Section 2.3 under certain assumptions on K and T.

As mentioned, the above issues are not only technical. The reason why the unconditioned process $(\mathcal{V}_N^{(T)})_{N\in\mathbb{N}}$ fails to satisfy an LDP with rate function $I_\mu^{(T)}$ is the possible emergence of a macroscopic particle by time T. One can see that, under the assumptions in Theorem 2.8, 2, actually a different scenario arises, the emergence of a gel, and this makes an LDP with rate function $I_\mu^{(T)}$ impossible. Let us recall that, in the special case of an inhomogeneous Erdős–Rényi graph (see Section 3.3), it turned out that a nontrivial contribution to the true rate function without conditioning comes from the macroscopic part of the configuration, which we neglect in Theorem 2.3. We plan to incorporate this part into our analysis in forthcoming work.

REMARK 3.4 (The role of conditioning on $\mathcal{A}_{f,\beta}$). In Theorem 2.3 we condition on $\mathcal{A}_{f,\beta}$ to avoid the problems in deriving an LDP for $(\mathcal{V}_N^{(T)})_{N\in\mathbb{N}}$ described in Remark 3.3. It is simple to see that the map $\nu \mapsto \int \nu(\mathrm{d}\xi)|\xi_0|$ is continuous and bounded on the set $\mathcal{A}_{f,\beta}$. Furthermore, this enables us to show that $I_\mu^{(T)}$ has compact sublevel sets on it, and the diagonal sum (i.e., the sum on $i = j$) in the exponent can be shown to be small (also using the assumption in (1.2)). This makes the proof of the LDP of Theorem 2.3 practical.

The compactness of the level sets of the rate function indeed hinges on the restriction to the set $\mathcal{A}_{f,\beta}$. In an alternative characterisation of the rate function that we provide in (A.12), at least for large T, the additional parameter b needs to be taken small, and then the one-but-last term is not lower semicontinuous in ν. However, on $\mathcal{A}_{f,\beta}$ it is easily seen to be even continuous. Then the compactness of the sublevel sets of the rate function follows easily from the lower semicontinuity of $\nu \mapsto \langle \nu, \mathfrak{R}^{(T)}(\nu)\rangle$ (see Lemma A.2) and the compactness of the level sets of the entropy.

However, on the event $\{\mathcal{V}_N^{(T)} \in \mathcal{A}_{f,\beta}\}$, there can be no loss of mass in the limit $N \to \infty$ for the random coagulation process $\mathcal{V}_N^{(T)}$, since the total mass of atoms present at time T in particles larger than L is bounded for any $L \in (0,\infty)$. Indeed, using the majorizing function $f(r) = r^2$ implies that

$$\int (\mathcal{V}_N^{(T)})(\mathrm{d}\xi)\|\xi_0\|_1 \mathbb{1}\{\|\xi_0\|_1 > L\} \le \frac{\beta}{L}.$$

This means that the conditioning rules out any occurrence of mesoscopic or macroscopic particles with total mass of order $\asymp N$. In this way, f acts like a majorant that induces tightness. In our arguments in Section 7, we will use only $f(r) = r^2$ and large β, since this is easy to handle.

One might think that, instead of conditioning on $\{\mathcal{V}_N^{(T)} \in \mathcal{A}_{f,\beta}\}$, in the proof of the LDP, one can use a decomposition into this set and its complement and try to show that the probability of the complement is exponentially small in N with a very large rate if β is large. This argument, if it could be carried through, would indeed lead to a proof of the LDP without conditioning. However, it is not successful, since the probability of such complementary event is exponentially small only when gelation is super exponentially unlikely to occur before time T and, therefore, it requires an a priori knowledge about the occurrence of gelation.

REMARK 3.5 (Difficulties in proving (non)gelation). If $\mathcal{V}_N^{(T)}$ were to satisfy an LDP under $\mathbb{P}^{(N)}_{\mathrm{Poi}_{N\mu}}$ with rate function $I_\mu^{(T)}$, as discussed in Remark 3.2, then one would expect that it converges (at least along subsequences) to a minimiser $\nu^{(T)}$ of $I_\mu^{(T)}$, which then satisfies the EL-equations in (2.19). In this setting, one might hope that the definition of gelation via (2.14) would be equivalent to mass loss in the minimiser, that is, $\int \nu^{(T)}(\mathrm{d}\xi)|\xi_0| < 1$.

However, as we already pointed out in Remark 3.3, the truth is much more delicate. Nevertheless, we succeeded in proving that there is no mass loss in either sense under the assumption $q_\mu^{(T)} < 1$, implying no gelation. However, the condition $q_\mu^{(T)} < 1$ is only sufficient, but not necessary; we dropped the exponential term in (2.19) when estimating in Lemma 7.1. Actually, the criterion $q_\mu^{(T)} < 1$ for nongelation (and convergence and characterization of limits) is good only for T small enough, depending on the upper bound on the kernel K in (1.2). Unfortunately, even for kernels that are known never to produce a gel, we are currently not able to derive $q_\mu^{(T)} < 1$ from our criterion; for example, for the constant kernel $K \equiv H$, we can currently deduce only that $q_\mu^{(T)} \le \frac{1}{2}HT$. This shows that our nongelation criterion is only sufficient and may be far from sharp.

On the other hand, under (1.3), we did prove that $\int \nu^{(T)}(\mathrm{d}\xi)|\xi_0| < 1$ for T sufficiently large; however, we were not able to use this to prove gelation. The reason is that we found no argument to show that $\mathcal{V}_N^{(T)}$ converges to $\nu^{(T)}$ as $N \to \infty$ under any measure that satisfies an LDP with rate function $I_\mu^{(T)}$. Instead, the main point in our proof of gelation for large T in Section 7.4 is the fact that $I_\mu^{(T)}$ is bounded away from zero for large T.

3.2. *Literature survey.* In this section we wish to give an overview on related literature.

Spaceless coagulation models. *Spaceless* models for coagulation have been studied for decades, and there are a number of works that derive criteria for the occurrence of gelation. A review by Aldous [4] gives a general overview, covering deterministic and stochastic points of view. He also suggests many open questions, several of which have since been resolved.

Mathematical modelling of coagulation began with Smoluchowski [44] in connection with his work on diffusion. He wrote down a deterministic model in the form of a coupled set of ODEs, (together known as Smoluchowski equation), which he informally derived from an underlying stochastic model of Brownian particles. The original Smoluchowski equation is the spaceless version of (2.21). A very natural stochastic (Markovian) model, which may be viewed as spatially homogenised limit of the stochastic model used by Smoluchowski, has been introduced by Marcus [35], again by Gillespie [25], and later studied by Lushnikov [34]. It is called the Marcus–Lushnikov model, and we study a spatial version of it in this paper.

The stochastic setting is well connected with the deterministic one. Indeed, replacing the coagulation kernel K by $\frac{1}{N}K$, several later authors prove, in the spaceless setting, the convergence of the normalised Marcus–Lushnikov process (written $\frac{1}{N}\Xi$ in our setting) toward the solution of the Smoluchowski equation, under various assumptions on the kernel and on the initial condition. Some authors [24, 37] prove convergence toward a more general version of the Smoluchowski version, the Flory equation [23], which characterises the evolution also in the presence of a gelation phase transition after the gelation time.

Much literature focuses on the deterministic setting, proving existence, uniqueness, and other properties of solutions of the Smoluchowski equation and its variants, under more and more general assumptions [9, 17, 21, 22]. In this setting the phenomenon of gelation was initially interpreted as an explosion of moments of solutions at a finite time, later as the existence of a time at which the solution loses mass, that is, the first moment strictly decreases. The first rigorous treatment of gelation in the stochastic model comes with Jeon's work [29] (see also [39] for extensions and generalisations), where several notions of gelation are considered. One main notion is in terms of some kind of boundedness in N (either in probability or in expectation) of the first time at which the total mass of the configuration that sits in particles larger than ψ_N is larger than δN for some $\delta > 0$ and for some scale function ψ_N. The choice of $\psi_N \asymp N$ (i.e., the appearance of a macroscopic particle) leads to what is called a strong gelation. It is clear that all these notions a priori depend on the parameter δ and the scale function ψ_N and on the sense of boundedness that is required. The notion that we chose refers to $\psi_N \equiv L$ and an additional limit as $L \to \infty$, and we take expectation of the total mass, a notion that is not handled in [29] or in [39]. Our gelation time defined in (2.15) is—so to speak—the "earliest"of all possible gelation times.

It is intuitively clear that sufficient criteria for gelation to occur should consist of lower bounds for the coagulation kernel K. The most interesting yet simple example is the product kernel, $K(m, m') = Cmm'$, which admits a one-to-one map of the coagulation process onto a natural growing sparse Erdős–Rényi graph (see Section 3.3), where gelation has a natural analogue with the famous emergence of a giant component. In this special case, one can identify the gelation time, and it turns out that at that time a macroscopic particle arises; hence, it is a strong gelation time in the above notion.

In [29] Jeon proved that a sufficient condition for gelation is that $K(m, m') \ge \epsilon(mm')^q$ for all $m, m' \in \mathbb{N}$, with some positive ϵ and $q > \frac{1}{2}$. Later, Rezakhanlou [39] proved that kernels satisfying $K(m, n) \ge m^q + n^q$ with $q > 1$ produce instantaneous gelation, that is, the time at which a giant particle appears tends to zero. Let us mention that it is, in general, believed in the applied mathematics areas that all homogeneous kernels, that is, kernels that satisfy $K(cm, cm') = c^\gamma K(m, m')$ for all $m, m' \in \mathbb{N}$, are gelling for $\gamma > 1$, but as far as we know there is not a proof for this statement.

Cluster coagulation model: *Comparison with* [38]. A very general coagulation model that is able to incorporate space or particle features other than mass is introduced and studied by Norris [38]. He introduces what he calls a *cluster coagulation model*, where each particle is registered as a cluster that is an element of some measurable space E, which is equipped with some mass function $m \colon E \to (0, \infty)$. The coagulation of a pair of particles is a replacement of two with one single particle such that the sum of the masses before and after coagulation is preserved. Choosing $E = \mathcal{S} \times \mathbb{N}$ makes our model a special case of this. In a nutshell, our analysis does not offer a more detailed picture of the gelation phenomenon than [38]; our new interpretation in terms of a PPP nevertheless sheds new light on the process and its phase transition.

Norris focuses on the deterministic side, proving existence and uniqueness of solutions to the cluster version of the Smoluchowski equation, in our setting (2.21), and conservation of mass in some cases, that is, when $K(x, y) \le \varphi(m(x))\varphi(m(y))$ for all $x, y \in E$, where φ is at most linear. Moreover Norris calls the kernel K *approximately multiplicative* if

$$\varepsilon m(x)m(y) \le K(x, y) \le M\big(1 + m(x)\big)\big(1 + m(y)\big), \quad x, y \in E,$$

for some $\varepsilon > 0$ and $M < \infty$; this is equivalent to our assumption that (1.2) and (1.3) hold for some H and h. In Theorem 2.2 he proves that equation (2.21) in this case has a unique mass-preserving solution up to a certain time T, which can be upper and lower bounded by functions of the initial condition. Note that this is not a probabilistic result but concerns the deterministic differential equation. In order to tie this result to the (stochastic) ML-process a convergence result is needed. Indeed, Norris studies convergence of $\frac{1}{N}\Xi$ (in our notation) to the solution of (2.21) in his Theorem 4.2. However, his result requires restrictive assumptions on the cluster space E (these are formulated in [38] as equations (3.1) and (3.2)). In particular, the first of these assumptions (his (3.1)) implies a handy discretisation of the process, and it essentially requires that the cluster space is induced by a discrete set of *shapes*. More precisely, the clusters are discrete shapes that are equipped with real-valued weights, which are needed to define the mass of a cluster, which in Norris' work is generally allowed to be real-valued. With other words, the masses are the only ingredient that is allowed to be continuous. Hence, our understanding is that Norris' convergence result does not hold for the case $E = \mathcal{S} \times \mathbb{N}$, if $\mathcal{S}$ is continuous. That being said, let us mention that his Theorem 4.2 states also that the convergence holds exponentially fast in N. This is a first indication that a large deviation principle might hold, even if it comes only in form of a (nonsharp) upper bound.

Let us compare now the estimates on the gelation time from [38] to ours. This is only possible under the restrictive assumptions from Theorem 4.2, since Norris' defines the gelation time via the solution of the (limiting) differential equation (2.21), while we define it via the stochastic process. In this framework the lower estimates on the gelation time are both of the order $1/H$ (or $1/M$ in Norris' notation), while the precise prefactors differ slightly. Regarding upper bounds for the gelation time, Theorem 2.2 from [38] suggests that the gelation time is not larger than $1/(\varepsilon\langle m^2, \mu_0\rangle)$, where $\langle m^2, \mu_0\rangle$ is the second moment of the initial masses. After that time, any solution fails to conserve mass or ceases to exist. Since Norris' definition of a gelation time relies on an existing solution to (2.21), he cannot formulate an explicit bound for the gelation time. It's an advantage of the ML-process that the gelation time can always be defined from the perspective of the particle system. While our quantitative upper bound from Corollary 2.11 is far from optimal, the connection between the infimum of our rate function and the gelation time that we find in Theorem 2.8 provides a new tool that has not been explored yet.

With our new approach, we derive an explicit and autonomous representation of the distribution of the process in terms of a PPP. This leads to a characterisation of convergence and of the limit in terms of an energy–entropy minimisation, which leads to a characteristic Euler–Lagrange equation rather than a differential equation. Furthermore, we derive the explicit large-deviation rate of the speed of convergence. Drawbacks of our ansatz are, in contrast to [38], that our convergence result of the Marcus–Lushnikov process is restricted to a time-interval whose length is determined by some (nonoptimal) analytic conditions on compactness issues and does not reach the gelation time, and our sufficient criteria for gelation are far from optimal and cannot cover some well-known examples.

Other related works on spatial coagulation models. Recently, in [5] the authors focus on gelation for the cluster coagulation model, providing criteria for gelation involving lower bounds on the coagulation kernel and the convergence of a series. The approach is similar to the one in [29, 39] for nonspatial kernels, based on estimates on the generator of the process. It gives upper bounds on the expected gelation time of the stochastic proess. The bounds obtained are similar to the one from Norris [38] and, therefore, very different from the one in our Corollary 2.11. In [6] the authors extend Norris' convergence results to convergence toward the solution(s) of a generalised Flory equation, when E is a σ-compact metric space. It is important to mention that, in general, the uniqueness of such solutions is quite a delicate question. Under relatively strong assumptions, uniqueness has been established both before and after gelation [36] in spaceless models. However, see [37] for a (quite involved) example of a spaceless kernel for which multiple solutions of the Smoluchowski equation are possible. It is not clear whether the introduction of spatial positions for the particles influences uniqueness in a positive or in a negative way. Our work does not shed any light on this question, since our uniqueness assertions are only for the minimiser of the rate function.

There are few other models of coagulation, where the coagulating particles move in Euclidean space: [27, 31] treat the case where the particles move as independent Brownian motions in $\mathbb{R}^d$ and [26] where the Brownian motions are replaced by random walks. The works [13, 46] deal with diffusive particles with very general interactions of which coagulation is only a special case. All these investigations proceed under assumptions that exclude gelation, whereas [41, 45] restrict to particles moving as independent Markov chains on some finite state space in order to investigate gelation in detail.

Large deviations for jump Markov processes. We turn now to large deviation results. In particular, we would like to draw comparisons between our Corollary 2.6 and existing results of dynamical LDPs, that is, LDPs for the path of the empirical measure of weakly interacting jump Markov processes over a compact time interval $[0, T]$. Results of this type have been proved in the case that the empirical measure takes value in a finite dimensional space, starting from [40] and then under increasingly general assumptions; see, for example, [2, 3] and references therein. The classical approach follows Freidlin–Wentzell theory; it consists in a tilting argument for which a law of large numbers for the transformed dynamics is needed.

Our model does not fall in the above category, as our main object of interest is the path of an infinite dimensional empirical measure. The case of LDP for infinite dimensional empirical measure has been under investigation mainly in kinetic theory, in relation to large deviations for Kac type of particle systems (stochastic microscopic models for Boltzmann equation). In this case, particles are characterised by a spatial location and a velocity; they interact pairwise by changing their velocities and preserving kinetic energy. Here the empirical measure is a measure on the space of locations and velocities; therefore, it is a truly infinite dimensional object. For this type of model, a rate function of the classical Freidlin–Wentzell type was suggested by Léonard [33] via a large deviation upper bound. However, matching lower bounds were proved only when restricting to classes of sufficiently good paths [11, 28, 43]; the difficulty lies in the lack of a law of large numbers for the perturbed dynamics. In [10, 12] two models of Kac type are studied, and a new rate function is suggested, which assigns a nontrivial rate to paths with increasing energy. The matching lower bound is then proved for a (larger) class of paths, but not yet for all paths. We are facing similar problems in our infinite dimensional setting, the role of conservation of energy for Kac type of models being played in our case by the conservation of the mass. To deal with this, we prove a conditional LDP; that is, we prove matching upper and lower bounds only for paths in a certain subset of the space (the image of the subset $\mathcal{A}_{f,\beta}$ under ρ). Exactly as predicted in [10, 28] for Kac type of models, we expect the true rate function (without ruling out gelation) to be different but we defer this to future work.

Other Poisson process approaches. Our Poisson-process description of Theorem 2.1 is novel in connection with coagulation processes and opens up a multitude of further research directions, in particular a comprehensive analysis of the entire trajectory configuration on the time-interval $[0, T]$. It also makes the model amenable to an asymptotic analysis, like hydro- or thermodynamic limits and large-deviations, in combination with the powerful LDP of Lemma A.4. Related Poisson representations have been found and employed for a few models in statistical mechanics, like the interacting Bose gas [1, 18]. Also, in a recent paper [42], the joint distribution of all the components of the Erdős–Rényi graph has been identified in terms of a Poisson point process (here a compound one). It is used there to derive moderate-deviations results for three variables (size of largest component, number of components of a given fixed size and total number of components). Previous large-deviations results for Erdős–Rényi graph with and without spatial component (see [7, 8]) can be obtained as well with this approach; see also Section 3.3 for an account on the latter subject. It is clear that the idea of Theorem 2.1 is amenable to all sorts of variants of coagulation processes, for example, with movement or other features. It seems to us that the idea can and should equally be applied in future to other models of statistical mechanics that are related to partitionings in any way, for example, to hard-core gas dynamics or billiards. In this respect, let us mention the work [15], which summarises a series of impressive works on hard-core gas dynamics. In this framework, even though a decomposition into collision trees has been identified as a crucial tool, no Poisson description has been used, and it will be natural to check whether this approach might give new insight or introduce alternate tools there.

3.3. *Comparison to inhomogeneous Erdős–Rényi graph.* There is a special case in which our spatial coagulation process can be realised as the connected component size process of an inhomogeneous Erdős–Rényi graph, which was analysed in [7]. The graph dynamics differs from the coagulation mechanism studied here only in that, for the graph, we add an edge between atoms (and leave their locations untouched) instead of replacing two particles by with a single particle at a new location. The connected components of the graph then play the role of the particles of the coagulation system.

Here is the special form of the coagulation and placement kernels that admit a representation as a graph process provided that $\mathcal{S}$ is a convex subset of a linear space. Consider

$$K\big((x,m),(x',m')\big)=\kappa\big(x,x'\big)mm',\quad x,x'\in\mathcal{S},m,m'\in\mathbb{N}, \tag{3.2}$$

with some symmetric and bilinear function $\kappa:\mathcal{S}\times\mathcal{S}\to[0,\infty)$, together with the deterministic placement kernel

$$\Upsilon\big((x,m),(x',m'),\cdot\big)=\delta_{\frac{mx+m'x'}{m+m'}}. \tag{3.3}$$

Recall from Remark 1.1 that this kernel determines the new location of the particle in such a way that the center of mass is preserved. This implies that the center of mass of a component (the convex combination of its vertices) is equal to the location of the particle in the coagulation process, for any component (resp., particle), at any time.

Since κ is linear in each of its arguments

$$|I||I'|\kappa\bigg(\frac{1}{|I|}\sum_{i\in I}x_i,\frac{1}{|I'|}\sum_{i\in I'}x_i\bigg)=\sum_{i\in I}\sum_{i\in I'}\kappa(x_i,x_i),\quad I,I'\subset[N].$$

Assuming that I and I' are disjoint, on the right-hand side, we see the rate (in the graph model) of putting a bond between any two vertices of the groups $\{x_i:i\in I\}$ and $\{x_i:i\in I'\}$, which corresponds, in the coagulation model, to the rate of the coagulation between two particles at the locations $\frac{1}{|I|}\sum_{i\in I}x_i$ and $\frac{1}{|I'|}\sum_{i\in I'}x_i$ with mass $|I|$, respectively, $|I'|$, which we see on the left-hand side.

The rate function that we derived in [7] shows how the macroscopic and the mesoscopic part of the configuration influence the large deviations of the microscopic part. We strongly expect to see corresponding effects for the more general coagulation process studied in this work, but this issue is deferred to future work. Our present work introduces a major improvement to the proof technique from [7] by introducing in Theorem 2.1 a Poisson point process that enables us to prove large-deviations results without the projective limits that are at the core of the proof in [7].

4. Proof of Theorem 2.1: Distribution of the configuration. In this section we prove our first main result, Theorem 2.1. That is, given an arbitrary probability measure μ on $\mathcal{S}$, we identify the distribution of the empirical measure $\mathcal{V}_N^{(T)}$ defined in (1.17) for any fixed $T \in (0, \infty)$ and $N \in \mathbb{N}$ under the poissonised initial condition $\mathbb{P}_{\mathrm{Poi}_{N\mu}}$ given in (1.10). We assume that the coagulation kernel K from Section 1.2 is just measurable.

The main idea is the decomposition of the process into all the subprocesses that coagulate into one single particle by time T. This is done via a decomposition in terms of all the coagulation trajectories, which we will also call *history trees*, that is, the parts of the coagulation process that coagulate into one single particle by time T. We are going to rewrite the joint distribution of all these trees in terms of a self-interacting Poisson point process. The first result is an explicit characterization of the self-interaction term.

Recall the partition process $Z = (Z_t)_{t\in[0,T]}$ from definition (1.12), that is, $Z_t = (X_C^{(t)}, C)_{C\in P_t}$, which consists of a process of partitions $P = (P_t)_{t\in[0,T]} \in \mathcal{P}(A)$ for some $A \subset \mathbb{N}$ with a location $X_C^{(t)} \in \mathcal{S}$ for every particle $C \in P_t$ at every $t \in [0, T]$. The Markovian dynamics is given in (1.13), and the initial state is $Z_0 = (x_i, \{i\})_{i\in A}$ for some vector $\mathbf{x} \in \mathcal{S}^A$. We denote the state space of Z by $\mathcal{Z}_A$, and we indicate with $\mathbb{P}_{\mathbf{x}}$ the law of such process. Fix disjoint finite label sets A, B and vectors $\mathbf{x} \in \mathcal{S}^A$, $\mathbf{y} \in \mathcal{S}^B$. Suppressing the dependence on T in the notation, we consider the process $Z = (Z_t)_{t\in[0,T]}$ under $\mathbb{P}_{(\mathbf{x},\mathbf{y})}$, where the partition process $(P_t)_{t\in[0,T]}$ takes values in $\mathcal{P}(A \cup B)$ with initial atom configuration $(\mathbf{x}, \mathbf{y}) \in \mathcal{S}^{A\cup B}$. We denote by $\{A \nleftrightarrow B\}$ the event that (up to time T) there is no coagulation between A and B, that is, no coagulation between any pair of particles (i.e., sets) C, D with $C \subset A$ and $D \subset B$. Now we define

$$R^{(T)}(\xi, \xi') = -\log \frac{\mathrm{d}\mathbb{P}_{(\mathbf{x},\mathbf{y})}(A \nleftrightarrow B, (\Xi^{(T,A)}, \Xi^{(T,B)}) \in \cdot)}{\mathrm{d}\mathbb{P}_{\mathbf{x}} \circ (\Xi(Z))^{-1} \otimes \mathbb{P}_{\mathbf{y}} \circ (\Xi(Z))^{-1}}(\xi, \xi'), \quad \xi, \xi' \in \Gamma_T. \tag{4.1}$$

In words, $e^{-R^{(T)}(\xi,\xi')}$ is the probability that the subprocesses $\xi = \Xi^{(T,A)}$ and $\xi' = \Xi^{(T,B)}$ do not coagulate with each other by time T. This probability is the probability that all the exponential holding times of pairs of particles between ξ and ξ' never elapse during $[0, T]$.

LEMMA 4.1 (Identification of $R^{(T)}$). *For any $\xi, \xi' \in \Gamma_T$, the density $R^{(T)}(\xi, \xi')$ in (4.1) exists, and the formula (2.2) holds.*

From formula (2.2) one sees that $R^{(T)}$ does not depend on the placement kernel Υ and is symmetric, since K is. Since for any $t \in [0, T]$ we have that ξ_t and ξ'_t are discrete in space, the integrals $\langle \xi_t, K\xi'_t\rangle$ are indeed sums.

PROOF. As for the definition (4.1) for $R^{(T)}(\xi, \xi')$, let A and B be disjoint finite sets, and fix $\mathbf{x} = (x_i)_{i\in A} \in \mathcal{S}^A$ and $\mathbf{y} = (y_i)_{i\in B} \in \mathcal{S}^B$ such that $\xi_0 = \sum_i \delta_{x_i}$ and $\xi'_0 = \sum_i \delta_{y_i}$. By $\mathbb{P}_{(\mathbf{x},\mathbf{y})}$ we denote the distribution of the process Z as above with partition process in $\mathcal{P}(A \cup B)$ and initial configuration $(\mathbf{x}, \mathbf{y}) \in \mathcal{S}^{A\cup B}$. On the event $\{A \nleftrightarrow B\}$, for any $t \in [0, T]$, P_t can be decomposed into a partition of A and a partition of B. In particular, we can decompose any state $z \in \{A \nleftrightarrow B\}$ of Z as $z = (z_A, z_B)$, where $z_A \in \mathcal{Z}_A$, $z_B \in \mathcal{Z}_B$. On the other hand, to any pair $z = (z_A, z_B)$ with $z_A \in \mathcal{Z}_A$, $z_B \in \mathcal{Z}_B$, we can associate a state $z \in \{A \nleftrightarrow B\}$. Then

$$\begin{aligned} &\int_{\mathcal{Z}_{A\cup B}} \mathbb{P}_{(\mathbf{x},\mathbf{y})}(A \nleftrightarrow B, Z_A \in \mathrm{d}z_A, Z_B \in \mathrm{d}z_B) \\ &\quad = \int_{\mathcal{Z}_A\times\mathcal{Z}_B} \mathbb{P}_{\mathbf{x}}(Z \in \mathrm{d}z_A)\mathbb{P}_{\mathbf{y}}(Z \in \mathrm{d}z_B)p(z_A, z_B), \end{aligned} \tag{4.2}$$

where $p(z_A, z_B) = \mathbb{P}_{(\mathbf{x},\mathbf{y})}(z_A \not\leftrightarrow z_B)$ is equal to the probability that no exponential time elapsed that would result in a coagulation between subsets of A and subsets of B. More precisely, given any two states $z_A \in \mathcal{Z}_A$, $z_B \in \mathcal{Z}_B$, we can iteratively, for $i = 1, 2, \dots$, construct maximal intervals $I_i = [t_{i-1}, t_i)$ such that the corresponding partition processes $(P_t^{(A)})_{t\in[0,T]}$ and $(P_t^{(B)})_{t\in[0,T]}$ are both constant on $[t_{i-1}, t_i)$. Then we indicate with $\{z_A \not\leftrightarrow z_B\}$ the event that, for all i and for all $C \in P_{t_{i-1}}^{(A)}$, $D \in P_{t_{i-1}}^{(B)}$ the exponentially distributed times with parameter $K((X_C, |C|), (X_D, |D|))$ are larger that $t_i - t_{i-1}$. Since for any i we can get the number of sets/particles $C \in P_{t_{i-1}}^{(A)}$ sitting in $(x, m) \in \mathcal{S} \times \mathbb{N}$ via the empirical measure $\Xi_{t_{i-1}}(z_A)(\mathrm{d}(x, m))$, as defined in (1.18) (and the same for B), we have that

$$p(z_A, z_B) = \prod_i \mathrm{e}^{-(t_i - t_{i-1})\langle \Xi_{t_{i-1}}(z_A), K\,\Xi_{t_{i-1}}(z_B)\rangle} = \exp\biggl(-\int_0^T \langle \Xi_t(z_A), K\,\Xi_t(z_B)\rangle\,\mathrm{d}t\biggr),$$

which only depends on $\Xi(z_A)$, $\Xi(z_B)$. Given that these exponential clocks did not ring, by definition the exponential clocks running between elements of $\mathcal{P}(A)$ are independent from the ones running between elements of $\mathcal{P}(B)$. Therefore, we have

$$\mathbb{P}_{(\mathbf{x},\mathbf{y})}(Z_A \in \mathrm{d}z_A, Z_B \in \mathrm{d}z_B | z_A \not\leftrightarrow z_B) = \mathbb{P}_{\mathbf{x}}(Z \in \mathrm{d}z_A)\mathbb{P}_{\mathbf{y}}(Z \in \mathrm{d}z_B),$$

which justifies (4.2). Now, going to the distributions of $\Xi(z_A)$, $\Xi(z_B)$ under $\mathbb{P}_{\mathbf{x}}$ and $\mathbb{P}_{\mathbf{y}}$ gives the result. $\square$

Now we express the probability of the event that the coagulation process decomposes into two pieces, not necessarily terminating with precisely one particle at time T.

LEMMA 4.2. *Let A and B be disjoint finite sets, and fix $\mathbf{x} = (x_i)_{i\in A} \in \mathcal{S}^A$ and $\mathbf{y} = (y_i)_{i\in B} \in \mathcal{S}^B$. By $\mathbb{P}_{(\mathbf{x},\mathbf{y})}$ we denote the distribution of the process Z as above with partition process in $\mathcal{P}(A \cup B)$ and initial configuration $(\mathbf{x}, \mathbf{y}) \in \mathcal{S}^{A\cup B}$. Then for any measurable $\mathcal{N} \subset \mathcal{M}_{\mathbb{N}_0}(\Gamma_T^{(1)})$,*

$$\begin{aligned}\mathbb{P}_{(\mathbf{x},\mathbf{y})}\bigl(A \not\leftrightarrow B, N\mathcal{V}_N^{(T)} \in \mathcal{N}\bigr) \\ = \bigl\langle \mathbb{P}_{\mathbf{x}} \otimes \mathbb{P}_{\mathbf{y}}, \mathrm{e}^{-R^{(T)}(\Xi^{(T,A)}, \Xi^{(T,B)})} \mathbb{1}\bigl\{N\bigl[\mathcal{V}_N^{(T)}(Z_A) + \mathcal{V}_N^{(T)}(Z_B)\bigr] \in \mathcal{N}\bigr\}\bigr\rangle,\end{aligned} \tag{4.3}$$

where we recall definition (1.14) *and* (1.17). (*We write the random processes under $\mathbb{P}_{\mathbf{x}}$ and $\mathbb{P}_{\mathbf{y}}$ as Z_A, respectively, Z_B and used* (1.17) *for each separately.*)

PROOF. On the event $\{A \not\leftrightarrow B\}$, for any $t \in [0, T]$, P_t can be decomposed into a partition of A and a partition of B. In particular, we can decompose any state $z \in \{A \not\leftrightarrow B\}$ of Z as $z = (z_A, z_B)$, where $z_A \in \mathcal{Z}_A$, $z_B \in \mathcal{Z}_B$. The measure $\mathcal{V}_N^{(T)}$ decomposes accordingly into the sum of $\mathcal{V}_N^{(T)}(Z_A)$ and $\mathcal{V}_N^{(T)}(Z_B)$. This gives us

$$\begin{aligned}&\mathbb{P}_{\mathbf{x},\mathbf{y}}\bigl(A \not\leftrightarrow B, N\mathcal{V}_N^{(T)} \in \mathcal{N}\bigr) \\ &\quad = \int_{\mathcal{Z}_A}\int_{\mathcal{Z}_B} \frac{\mathrm{d}\mathbb{P}_{\mathbf{x},\mathbf{y}}(A \not\leftrightarrow B, (Z_A, Z_B) \in \cdot)}{\mathrm{d}\mathbb{P}_{\mathbf{x}} \otimes \mathbb{P}_{\mathbf{y}}}(z_A, z_B)\mathbb{1}\bigl\{N\mathcal{V}_N^{(T)} \in \mathcal{N}\bigr\}\mathbb{P}_{\mathbf{x}}(\mathrm{d}z_A)\mathbb{P}_{\mathbf{y}}(\mathrm{d}z_B).\end{aligned}$$

Hence, we can insert the density which is equal to $\mathrm{e}^{-R^{(T)}(\Xi^{(T,A)}(z_A), \Xi^{(T,B)}(z_B))}$ and arrive at (4.3). $\square$

Now we formulate the decomposition of the coagulation process into pieces that end up with just one particle each at time T. We denote the restriction of $\mathbb{P}_{\mathbf{x}}$ to $\{\Xi(Z) \in \Gamma_T^{(1)}\}$ by $\mathbb{P}_{\mathbf{x}}^{(1)}$ and note that this is a subprobability measure.

LEMMA 4.3. *Fix $M \in \mathbb{N}$, and let $P = \{C_j : j \in [m]\} \in \mathcal{P}([M])$. Fix $\mathbf{x} = (x_i)_{i=1}^M \in \mathcal{S}^M$, and denote $\mathbf{x}^{(j)} = (x_i)_{i \in C_j}$. Then for any measurable $\mathcal{N} \subset \mathcal{M}_{\mathbb{N}_0}(\Gamma_T^{(1)})$*

$$\begin{aligned} &\mathbb{P}_{\mathbf{x}}\big(P_T = P, N\mathcal{V}_N^{(T)} \in \mathcal{N}\big) \\ &\quad = \Big\langle \bigotimes_{j=1}^m \mathbb{P}^{(1)}_{\mathbf{x}^{(j)}}, \mathrm{e}^{-\frac{1}{2}\sum_{j,j' :\, j \neq j'} R^{(T)}(\Xi_j, \Xi_{j'})} \mathbb{1}\Big\{\sum_{j=1}^m \delta_{\Xi_j} \in \mathcal{N}\Big\}\Big\rangle, \end{aligned} \tag{4.4}$$

where for each $j = 1, \dots, m$, we denote the random variables under $\mathbb{P}_{\mathbf{x}^{(j)}}$ by Z_j and put $\Xi_j = \Xi(Z_j)$.

PROOF. Note that $\{P_T = P\}$ implies that $\{C_1 \nleftrightarrow \bigcup_{j=2}^m C_j\}$, and use Lemma 4.2. Iterating this argument gives formula (4.4), since $\sum_{C, \tilde{C} \in P :\, C \neq \tilde{C}} R^{(T)}(\Xi^{(T,C)}, \Xi^{(T,\tilde{C})}) = \frac{1}{2}\sum_{j,j' :\, j \neq j'} R^{(T)}(\Xi_j, \Xi_{j'})$ $\square$

PROOF OF THEOREM 2.1. Without loss of generality, it suffices to show the statement of the Theorem for functions of the form $f(\nu) = \mathbb{1}_{\mathcal{N}}(N\nu)$ for some measurable set $\mathcal{N} \in \mathcal{M}_{\mathbb{N}_0}(\Gamma_T^{(1)})$. Denote by L the set of collections of numbers $\ell = (\ell_n)_{n \in \mathbb{N}} \in \mathbb{N}_0^{\mathbb{N}}$. For any point measure $\nu = \sum_{i \in I} \delta_{\xi^{(i)}} \in \mathcal{M}_{\mathbb{N}_0}(\Gamma_T^{(1)})$, there exists an $\ell \in L$ such that $\ell_n = \#\{i \in I : |\xi_0^{(i)}| = n\}$, for any $n \in \mathbb{N}$, that is, ν consists of exactly ℓ_n (coagulation) trees of of size n, for each $n \in \mathbb{N}$ (where the size of a tree $\xi \in \Gamma_T^{(1)}$ is given as the number of atoms, i.e., by $|\xi_0|$). In that case we say that the tree sizes of ν are given by ℓ. We can decompose any measurable set $\mathcal{N} \subset \mathcal{M}_{\mathbb{N}_0}(\Gamma_T^{(1)})$ as $\mathcal{N} = \bigcup_{\ell \in L} \mathcal{N}(\ell)$ with

$$\mathcal{N}(\ell) = \{\nu \in \mathcal{N} : \text{the tree sizes of } \nu \text{ are given by } \ell\}, \quad \text{for any } \ell \in L,$$

where the sets $\mathcal{N}(\ell)$, for $\ell \in L$, are disjoint. In the following we will assume that $\mathcal{N} = \mathcal{N}(\ell)$ for some $\ell \in L$. Recall that $N\mathcal{V}_N^{(T)} = \sum \delta_{\Xi^{(T,C)}}$, where the sum extends over $C \in P_T$, which we will leave out in the notation. We want study the event $\{\sum \delta_{\Xi^{(T,C)}} \in \mathcal{N}\}$. We abbreviate $M = \sum_n \ell_n n$ and $m = \sum_n \ell_n$ and note that on the event $\{\sum \delta_{\Xi^{(T,C)}} \in \mathcal{N}\}$ it holds that M is the total number of atoms and m is the total number of coagulation trees. Recall that under $\mathbb{P}_{\mathrm{Poi}_{N\mu}}$ we want to consider the empirical measure $\mathcal{V}_N^{(T)}$ under $\mathbb{P}_{\mathbf{x}}$, where the initial condition $\mathbf{x} = (x_i)_{i \in I}$ is such that $\sum_{i \in I} \delta_{x_i}$ is $\mathrm{Poi}_{N\mu}$ distributed. More precisely, we choose the length $|I|$ of the initial vector $\mathbf{x}$ as Poi_N-distributed and sample the $(x_i)_{i \in I}$ i.i.d. and with distribution μ. Therefore, we have that

$$\begin{aligned} &\mathbb{P}_{\mathrm{Poi}_{N\mu}}\Big(\sum \delta_{\Xi^{(T,C)}} \in \mathcal{N}\Big) \\ &\quad = \sum_{M'=0}^{\infty} \mathrm{Poi}_N(M') \int_{\mathcal{S}^{M'}} \mu^{\otimes M'}(\mathrm{d}(x_1, \dots, x_{M'}))\mathbb{P}_{\mathbf{x}}\Big(\sum \delta_{\Xi^{(T,C)}} \in \mathcal{N}\Big). \end{aligned}$$

The sum reduces to the summand $M' = M$, since otherwise the probability of the event is 0. Now,

$$\mathbb{P}_{\mathbf{x}}\Big(\sum \delta_{\Xi^{(T,C)}} \in \mathcal{N}\Big) = \sum_{\substack{P \in \mathcal{P}([M]) \\ P \text{ compatible with } \ell}} \mathbb{P}_{\mathbf{x}}\Big(P_T = P, \sum \delta_{\Xi^{(T,C)}} \in \mathcal{N}\Big),$$

where we say that a partition $P \in \mathcal{P}([M])$ is compatible with ℓ if for each $n \in \mathbb{N}$ we have that $\#\{C \in P\colon |C| = n\} = \ell_n$. Using the formula from Lemma 4.3, we get that

$$
\begin{aligned}
&\mathbb{P}_{\mathrm{Poi}_{N\mu}}\Big(\sum \delta_{\Xi^{(T,C)}} \in \mathcal{N}\Big)\\
&\quad = \mathrm{Poi}_N(M)\\
&\qquad \times \sum_{\substack{P \in \mathcal{P}(M)\\ P \text{ compatible with } \ell}} \int_{\mathcal{S}^M} \mu^{\otimes M}(\mathrm{d}\mathbf{x}) \Big\langle \bigotimes_{j=1}^m \mathbb{P}^{(1)}_{\mathbf{x}^{(j)}}, \mathrm{e}^{-\frac{1}{2}\sum_{j,j'\colon j\neq j'} R(\Xi_j,\Xi_{j'})}\\
&\qquad \times \mathbb{1}\Big\{\sum_{j=1}^m \delta_{\Xi_j} \in \mathcal{N}\Big\}\Big\rangle.
\end{aligned}
\tag{4.5}
$$

For $n \in \mathbb{N}$, we denote $\mu^{\otimes n} \otimes \mathbb{P}^{(1)}(\mathrm{d}(\mathbf{x}, Z)) = \mu^{\otimes n}(\mathrm{d}\mathbf{x}) \otimes \mathbb{P}^{(1)}_{\mathbf{x}}(\mathrm{d}Z)$ (i.e., we conceive $\mathbb{P}^{(1)}$ as the kernel $(\mathbf{x}, A) \mapsto \mathbb{P}^{(1)}_{\mathbf{x}}(A)$). Then for any $P \in \mathcal{P}(M)$ that is compatible with ℓ, we have that each summand in the second line of (4.5) equals

$$
\Big\langle \bigotimes_{j=1}^m \big(\mu^{\otimes |C_j|} \otimes \mathbb{P}^{(1)}\big), \mathrm{e}^{-\frac{1}{2}\sum_{j,j'\colon j\neq j'} R^{(T)}(\Xi_j,\Xi_{j'})} \mathbb{1}\Big\{\sum_{j=1}^m \delta_{\Xi_j} \in \mathcal{N}\Big\}\Big\rangle,
\tag{4.6}
$$

where we put $P = \{C_j\colon j \in [m]\}$. The right-hand side of (4.6) depends only on the cardinalities n_j of the partition sets C_j. Given ℓ, we can uniquely fix a collection of numbers $n_1, \dots, n_m \in \mathbb{N}$ such that $\#\{j\colon n_j = n\} = \ell_n$ for each $n \in \mathbb{N}$. Hence, we need to find the number of partitions $P = \{C_j\colon j \in [m]\}$ such that the cardinalities n_j of the partition sets satisfy this, in other words, such that P is compatible with ℓ. Indeed, we observe that

$$
\#\big\{P \in \mathcal{P}(M)\colon P \text{ compatible with } \ell\big\} = \frac{M!}{m! \prod_{j=1}^m n_j!} = \frac{M!}{m! \prod_n n!^{\ell_n}}.
\tag{4.7}
$$

To see that, recall that the multinomial factor $\frac{M!}{\prod_{j=1}^m n_j!}$ is the number of possibilities of putting the indices from $[M]$ into boxes labelled by $j = 1, \dots, m$ such that box j has exactly n_j many indices. Each of those possibilities gives us an ordered partition of $[M]$ that is compatible with $(n_j)_{j=1}^m$ (in the sense that its jth set has n_j many elements). Any (nonordered) partition of $[M]$ that is compatible with ℓ corresponds to precisely $m!$ many ordered partitions that are compatible with $(n_j)_{j=1}^m$. Thus, formula (4.7) holds.

Summarizing the previous steps, we have shown that

$$
\begin{aligned}
&\mathbb{P}_{\mathrm{Poi}_{N\mu}}\Big(\sum \delta_{\Xi^{(C)}} \in \mathcal{N}\Big)\\
&\quad = \mathrm{Poi}_N(M) \frac{1}{m!} \frac{M!}{\prod_{j=1}^m n_j!}\\
&\qquad \times \Big\langle \bigotimes_{j=1}^m \big(\mu^{\otimes n_j} \otimes \mathbb{P}^{(1)}\big), \mathrm{e}^{-\frac{1}{2}\sum_{j,j'\colon j\neq j'} R^{(T)}(\Xi_j,\Xi_{j'})} \mathbb{1}\Big\{\sum_j \delta_{\Xi_j} \in \mathcal{N}\Big\}\Big\rangle,
\end{aligned}
\tag{4.8}
$$

where we recall that $\mathcal{N} = \mathcal{N}(\ell)$ and $\#\{j\colon n_j = n\} = \ell_n$ for each $n \in \mathbb{N}$ and that Ξ_j is short for $\Xi(Z_j)$ (see (1.18)), where Z_j is the random variable under $\mu^{\otimes n_j} \otimes \mathbb{P}^{(1)}$.

Now we make the connection with the reference measure $M^{(T)}_{\mu,N}$ defined in (2.1). Note that, on $\{\xi \in \Gamma^{(1)}_T\colon |\xi_0| = n_j\}$, we have

$$
\frac{1}{n_j!}\big(\mu^{\otimes n_j} \otimes N^{n_j - 1}\mathbb{P}^{(1)}\big) \circ \Xi^{-1}(\mathrm{d}\xi) = M^{(T)}_{\mu,N}(\mathrm{d}\xi),
$$

where we conceive Ξ as the map defined in (1.18).

Inserting this into (4.8) and using that $\sum_j n_j = M$ (and hence $\prod_j N^{n_j-1} = N^{M-m}$), we get

$$\begin{aligned}
&\mathbb{P}_{\mathrm{Poi}_{N\mu}}\Big(\sum \delta_{\Xi(C)} \in \mathcal{N}\Big) \\
&\quad = \mathrm{Poi}_N(m)\Big\langle \bigotimes_{j=1}^{m}\Big(\frac{1}{n_j!}\mu^{\otimes n_j} \otimes N^{n_j-1}\mathbb{P}^{(1)}\Big), \mathrm{e}^{-\frac{1}{2}\sum_{j,j'\colon j\neq j'} R^{(T)}(\Xi_j,\Xi_{j'})}\mathbb{1}\Big\{\sum_j \delta_{\Xi_j} \in \mathcal{N}\Big\}\Big\rangle \\
&\quad = \mathrm{e}^{-N}\frac{1}{m!}\Big\langle \big(NM_{\mu,N}^{(T)}\big)^{\otimes m}, \mathrm{e}^{-\frac{1}{2}\sum_{j,j'\colon j\neq j'} R^{(T)}(\Xi_j,\Xi_{j'})}\mathbb{1}\Big\{\sum_j \delta_{\Xi_j} \in \mathcal{N}\Big\}\Big\rangle \\
&\quad = \mathrm{e}^{N(|M_{\mu,N}^{(T)}|-1)}\mathrm{E}_{NM_{\mu,N}^{(T)}}\Big[\mathrm{e}^{-\frac{1}{2}\sum_{j,j'\colon j\neq j'} R^{(T)}(\Xi_j,\Xi_{j'})}\mathbb{1}\Big\{\sum_j \delta_{\Xi_j} \in \mathcal{N}\Big\}\Big],
\end{aligned}$$

where the Ξ_j denote in the second line $\Xi(Z_j)$, in the third line the random variables with distribution $M_{\mu,N}^{(T)}$ (even though this is not normalised), and in the fourth line the points of the PPP.

We have arrived at the assertion. $\square$

5. Preparations. In this section we prepare for the proof of our remaining main results by doing the following:

- We prove the convergence of $\mathbb{Q}_k^{(T,N,N)}$ and introduce the measure $\mathbb{Q}_k^{(T)}$ in Section 5.1.
- We prove the continuity of the map $\nu \mapsto \rho(\nu)$ in Section 5.2.

5.1. *The measure* $\mathbb{Q}_k^{(T)}$. In this section we study the limit of $\mathbb{Q}_k^{(T,N,N)} := N^{|k|-1}\mathbb{P}_k^{(N)} \times (\Xi \in \cdot)|_{\Gamma_T^{(1)}}$ as $N \to \infty$ for $k \in \mathcal{M}_{\mathbb{N}_0}(\mathcal{S})$, that is, the restriction of the distribution of the coagulation process to the space $\Gamma_T^{(1)}$ of history trees. Recall that the (additional) super-index "N"indicates that we have replaced the kernel K by $\frac{1}{N}K$. The limiting measure $\mathbb{Q}_k^{(T)}$ will play an important role in the description of the limiting behaviour of the coagulation process. We will identify $\mathbb{Q}_k^{(T)}$ in terms of an explicit formula, using a kind of chart, that is, a push-forward measure under some explicit measure on the space $[(\mathcal{S}\times\mathbb{N})^2 \times \mathcal{S} \times (0,\infty)]^{|k|-1}$ that carries all the data needed to understand which transition is happening in each of the $|k|-1$ coagulation steps.

Fix an initial configuration $k \in \mathcal{M}_{\mathbb{N}_0}(\mathcal{S})$ and a coagulation kernel K as in Section 1.2. Let us identify the distribution of the coagulation process $\Xi = (\Xi_t)_{t\in[0,T]}$ under $\mathbb{P}_k$ on $\Gamma_T^{(1)}$, more precisely, on the set

$$\Gamma_{T,k}^{(1)} = \{\xi \in \Gamma_T^{(1)} \colon \xi_0 = k\}. \tag{5.1}$$

Recall that a history tree $\xi \in \Gamma_{T,k}^{(1)}$ is a trajectory that is only allowed to perform steps as in (1.6). For a pair of particles at $(x,m), (x',m') \in \mathcal{S}\times\mathbb{N}$ that coagulates into a particle with location $z \in \mathcal{S}$, this step is given by the addition of the signed measure

$$W_{(x,m),(x',m')}^{(z)} = -\delta_{(x,m)} - \delta_{(x',m')} + \delta_{(z,m+m')}. \tag{5.2}$$

For a fixed tuple $(y_i, y_i', z_i)_{i=1,\dots,|k|-1}$ with $y_i, y_i' \in \mathcal{S}\times\mathbb{N}$ and $z_i \in \mathcal{S}$, we define

$$\phi_0(\cdot,m) = k(\cdot)\delta_1(m) \quad \text{and} \quad \phi_i = \phi_0 + \sum_{j=1}^{i} W_{y_j,y_j'}^{(z_j)}, \quad \text{for } i = 1,\dots,|k|-1, \tag{5.3}$$

and say that $((y_i, y_i', z_i)_{i=1,\dots,|k|-1})$ is compatible with k if

$$\phi_i \geq 0, \quad \text{for } i = 0, \dots, |k| - 1. \tag{5.4}$$

Let $\mathfrak{X}_k \subset [(\mathcal{S} \times \mathbb{N})^2 \times \mathcal{S}]^{|k|-1}$ denote the set of such tuples. Furthermore, introduce the set of admissable time tuples,

$$\mathfrak{F}_k = \left\{(s_1, \dots, s_{|k|-1}) \in [0, \infty)^{|k|-1} \colon \sum_{i=1}^{|k|-1} s_i \leq T\right\},$$

and define

$$\Psi_k \colon \mathfrak{X}_k \times \mathfrak{F}_k \to \Gamma_{T,k}^{(1)}, \big(y_i, y_i', z_i, s_i\big)_{i=1,\dots,|k|-1} \mapsto \xi = (\xi_t)_{t\in[0,T]},$$

by

$$\xi_t = \sum_{i=1}^{|k|} \mathbb{1}\{t \in I_i\}\phi_{i-1}, \text{where } I_i = \begin{cases} [\sum_{j=1}^{i-1} s_j, \sum_{j=1}^{i} s_j) & \text{for } i < |k|, \\ \left[\sum_{j=1}^{|k|-1} s_j, T\right] & \text{for } i = |k|, \end{cases} \tag{5.5}$$

where $\phi = (\phi_i)_{i=1,\dots,|k|-1}$ is defined in (5.3). In words, if, starting from ϕ_0, for $i = 1, \dots, |k| - 1$, iteratively after a time elapse of s_i time units, two particles at $y_i = (x_i, m_i)$ and $y_i' = (x_i', m_i')$ coagulate into a particle at $(z_i, m_i + m_i')$, then at each time $t \in [0, T]$, the configuration is equal to ξ_t. Note that ξ_T is a delta-measure; that is, after time $\sum_{j=1}^{|k|-1} s_j$, there is no coagulation possible anymore.

It is not hard to see that the mapping Ψ_k is a bijection. We will leave the details to the reader.

In the following lemma, we derive a formula for the distribution of history trees under $\mathbb{P}_k$. Recall the definition of $\mathbf{K}_\phi$ from (1.7). Let us abbreviate $K_\phi((x, m), (x', m')) = \mathbf{K}_\phi((x, m), (x', m'), \mathcal{S})$, for any $(x, m), (x', m') \in \mathcal{S} \times \mathbb{N}$. Recall that we conceive $\mathbf{K}_\phi$ as a measure on $(\mathcal{S} \times \mathbb{N})^2 \times \mathcal{S}$. In the first four arguments, $\mathbf{K}_\phi$ is indeed point a measure. With a slight abuse of notation, one can conceive $\mathbf{K}_{\phi_{i-1}}(\mathrm{d}(x, m), \mathrm{d}(x', m'), \mathrm{d}z)$ as a Markov kernel, since $\phi_i = \phi_{i-1} + W_{y_i,y_i'}^{(z_i)}$. Their product $\bigotimes_{i=1}^{|k|-1} \mathbf{K}_{\phi_{i-1}}$ is concentrated on $\mathfrak{X}_k$.

LEMMA 5.1 (The distribution of a history tree). *Fix $k \in \mathcal{M}_{\mathbb{N}_0}(\mathcal{S})$, $k \neq 0$. Then we have, for any measurable bounded test function $f \colon \Gamma_{T,k}^{(1)} \to \mathbb{R}$,*

$$\mathbb{E}_k\big(f(\Xi)\mathbb{1}\{\Xi \in \Gamma_{T,k}^{(1)}\}\big) = \int_{\mathfrak{X}_k\times\mathfrak{F}_k} \left(\bigotimes_{i=1}^{|k|-1} \mathbf{K}_{\phi_{i-1}} \otimes \bigotimes_{i=1}^{|k|-1} \mathrm{d}s_i\right)(\mathrm{d}\Theta) f\big(\Psi_k(\Theta)\big)\mathrm{e}^{-\widetilde{\varphi}_k(\Theta)}, \tag{5.6}$$

where

$$\widetilde{\varphi}_k(\Theta) = \frac{1}{2} \sum_{i=1}^{|k|-1} s_i\big[\langle\phi_{i-1}, K\phi_{i-1}\rangle - \langle\phi_{i-1}, K^{(\mathrm{diag})}\rangle\big], \quad \Theta = \big(y_i, y_i', z_i, s_i\big)_{i=1,\dots,|k|-1}, \tag{5.7}$$

where we introduced $K^{(\mathrm{diag})}(y) = K(y, y)$ for $y \in \mathcal{S} \times \mathbb{N}$.

The right-hand side of (5.6) is the image measure under Ψ_k of the restriction of the measure $\bigotimes_{i=1}^{|k|-1}(\mathbf{K}_{\phi_{i-1}} \otimes \mathrm{d}s_i)$ to its support $\mathfrak{X}_k \times \mathfrak{F}_k$ with density $\mathrm{e}^{-\widetilde{\varphi}_k}$.

PROOF. Fix $i \in \{1, \dots, |k|-1\}$. During the time interval I_i, for any unordered pair $\{y, y'\}$ of elements of $\mathcal{S} \times \mathbb{N}$, there are $\phi_{i-1}(y)\phi_{i-1}(y')$ if $y \neq y'$, respectively, $\frac{1}{2}\phi_{i-1}(y)(\phi_{i-1}(y) - 1)$ if $y = y'$ independent exponential holding times running with parameter $K(y, y')$, one of which elapses at the respective coagulation time, namely, one with unordered pair $\{y_i, y_i\}$. There are only finitely many such exponential clocks involved, since ϕ_{i-1} has a finite support. The Lebesgue density for the event that the unordered pair of particles $\{y_i, y_i'\}$ coagulates at time s_i as the first pair in the configuration is

$$
\begin{aligned}
s_i \mapsto K_{\phi_{i-1}}(y_i, y_i')\mathrm{e}^{-s_i K_{\phi_{i-1}}(y_i, y_i')}\Bigg(\prod_{\{y,y'\}\neq\{y_i,y_i'\}:\, y\neq y'} \mathrm{e}^{-s_i \phi_{i-1}(y)K(y,y')\phi_{i-1}(y')}\Bigg) \\
\times \prod_{y:\, \{y\}\neq\{y_i,y_i'\}} \mathrm{e}^{-s_i \frac{1}{2}\phi_{i-1}(y)K(y,y)(\phi_{i-1}(y)-1)} \qquad (5.8)\\
= K_{\phi_{i-1}}(y_i, y_i')\mathrm{e}^{-s_i \eta^{(i)}(y_i,y_i')},
\end{aligned}
$$

where, for $y_i \neq y_i'$,

$$
\begin{aligned}
\eta^{(i)}(y_i, y_i') &= \frac{1}{2}\sum_{y,y'} \phi_{i-1}(y)K(y, y')\phi_{i-1}(y') - \frac{1}{2}\sum_{y} \phi_{i-1}(y)K(y, y) \\
&= \sum_{\{y,y'\}:\, y\neq y'} \phi_{i-1}(y)K(y, y')\phi_{i-1}(y') \\
&\quad + \sum_{y:\, \{y\}\neq\{y_i,y_i'\}} \frac{1}{2}\phi_{i-1}(y)K(y, y)\big(\phi_{i-1}(y) - 1\big),
\end{aligned}
$$

and for $y_i = y_i'$

$$
\eta^{(i)}(y_i, y_i') = \sum_{\{y,y'\}:\, y\neq y'} \phi_{i-1}(y)K(y, y')\phi_{i-1}(y') + \sum_{y} \frac{1}{2}\phi_{i-1}(y)K(y, y)\big(\phi_{i-1}(y) - 1\big),
$$

that is, the same formula.

The probability that the new particle is placed at z_i is expressed by multiplying (5.8) with $\Upsilon((x_i, m_i), (x_i', m_i'), \mathrm{d}z_i)$, which turns the first factor into $\mathbf{K}_{\phi_{i-1}}(y_i, y_i', \mathrm{d}z_i)$.

Because of the Markov property of the coagulation process, the probability $\mathbb{P}_k(\Xi \in \mathrm{d}\xi)$ is equal to the product over $i = 1, \dots, |k| - 1$ of (5.8). Noting that $\widetilde{\varphi}_k(\Theta) = \sum_{i=1}^{|k|} \eta^{(i)}(y_i, y_i')$, this implies (5.6). □

Now we introduce the measure

$$
\mathbb{Q}_k^{(T,N)}(\cdot) = N^{|k|-1}\mathbb{P}_k(\Xi \in \cdot)|_{\Gamma_T^{(1)}} \in \mathcal{M}\big(\Gamma_T^{(1)}\big). \qquad (5.9)
$$

The next question that we consider is the identification of the limit of $\mathbb{Q}_k^{(T,N,N)}$ as $N \to \infty$ for fixed $k \in \mathcal{M}_{\mathbb{N}_0}(\mathcal{S})$, where we recall that we add a superindex "N"when we replace the kernel K by $\frac{1}{N}K$. Using the description of ξ that was given in Lemma 5.1, we define a measure $\mathbb{Q}_k^{(T)}$ on $\Gamma_{T.k}^{(1)}$ by dropping just the density in (5.6),

$$
\begin{aligned}
\mathbb{Q}_k^{(T)}(\mathrm{d}\xi) &= \Bigg(\bigotimes_{i=1}^{|k|-1} \mathbf{K}_{\phi_{i-1}} \otimes \bigotimes_{i=1}^{|k|-1} \mathrm{d}s_i\Bigg) \circ \Psi_k^{-1}(\mathrm{d}\xi) \qquad (5.10)\\
&= \mathrm{e}^{\varphi_k(\xi)}\mathbb{P}_k\big(\xi \in \Gamma_{T,k}^{(1)}; \Xi \in \mathrm{d}\xi\big),
\end{aligned}
$$

where

$$\varphi_k(\xi) = \frac{1}{2}\int_0^T \big[\langle \xi_t, K\xi_t\rangle - \langle \xi_t, K^{(\mathrm{diag})}\rangle\big]\,\mathrm{d}t, \quad k = \xi_0. \tag{5.11}$$

Note that $\varphi_k(\Psi_k(\Theta)) = \widetilde{\varphi}_k(\Theta)$ for $\Theta \in \mathfrak{X}_k$.

LEMMA 5.2 (Limit of $\mathbb{P}_k^{(N)}(\Xi \in \cdot)|_{\Gamma_T^{(1)}}$). *Fix any $k \in \mathcal{M}_{\mathbb{N}_0}(\mathcal{S})$, $k \neq 0$. Then*

$$\frac{\mathrm{d}\mathbb{Q}_k^{(T,N,N)}}{\mathrm{d}\mathbb{Q}_k^{(T)}}(\xi) = \mathrm{e}^{-\frac{1}{N}\varphi_k(\xi)}, \quad \xi \in \Gamma_{T,k}^{(1)}. \tag{5.12}$$

If K satisfies (1.2), *then the density satisfies*

$$0 \le \varphi_k(\xi) \le \frac{1}{2}HT|k|^2, \quad \xi \in \Gamma_{T,k}^{(1)}, \tag{5.13}$$

and in particular,

$$\mathbb{Q}_k^{(T)} = \lim_{N\to\infty} \mathbb{Q}_k^{(T,N,N)}, \tag{5.14}$$

weakly and in total variation.

PROOF. We apply formula (5.6) to the coagulation model with kernel $\frac{1}{N}K$ (instead of K). Hence, the $(|k|-1)$-fold product of the K-terms receives a prefactor $N^{-(|k|-1)}$, which is compensated by the prefactor $N^{|k|-1}$ in (5.9). Hence, on $\Gamma_{T,k}^{(1)}$ we have that

$$\begin{aligned}\mathbb{Q}^{(T,N,N)}(\mathrm{d}\xi) &= \left(\bigotimes_{i=1}^{|k|-1} \mathbf{K}_{\phi_{i-1}} \otimes \bigotimes_{i=1}^{|k|-1} \mathrm{d}s_i\right) \circ \Psi_k^{-1}(\mathrm{d}\xi)\exp\Big(-\frac{1}{N}\varphi_k\big(\Theta^{-1}(\xi)\big)\Big)\\ &= \exp\Big(-\frac{1}{N}\varphi_k(\xi)\Big)\mathbb{Q}_k^{(T)}(\mathrm{d}\xi).\end{aligned}$$

This shows (5.12).

To show (5.13), we use that $\langle \phi, K\phi\rangle \le H\|\phi\|_1^2$ for any nontrivial $\phi \in \mathcal{M}_{\mathbb{N}_0}(\mathcal{S}\times\mathbb{N})$ and that $\|\xi_t\|_1 = \|\xi_0\|_1 = |k|$ for all $t \in [0,T]$ and for any $\xi \in \Gamma_{T,k}^{(1)}$. Therefore, we get

$$\varphi_k(\xi) \le \frac{1}{2}\int_0^T \langle \xi_t, K\xi_t\rangle\,\mathrm{d}t \le \frac{1}{2}HT|k|^2. \tag{5.15}$$

The bound (5.13) obviously implies convergence in total variation as well as weak convergence. □

LEMMA 5.3. *We define $\sigma : \mathcal{M}_{\mathbb{N}_0}(\mathcal{S}\times\mathbb{N}) \to [0,\infty)$ recursively via $\sigma(\delta_{(x,m)}) = 1$ for any $(x,m) \in \mathcal{S}\times\mathbb{N}$ and*

$$\sigma(\phi) = \sum_{(x,m),(x',m')} \int_{\mathcal{S}} \mathbf{K}_\phi\big((x,m),(x',m'),\mathrm{d}z\big)\sigma\big(\phi - \delta_{(x,m)} - \delta_{(x',m')} + \delta_{(z,m+m')}\big),$$

where the sum is taken over the support of ϕ. Then the measure $\mathbb{Q}_k^{(T)}$ has total mass equal to

$$\mathbb{Q}_k^{(T)}(\Gamma_{T,k}^{(1)}) = \frac{T^{|k|-1}}{(|k|-1)!}\sigma(\phi_0), \quad \textit{where } \phi_0(\mathrm{d}x,m) = k(\mathrm{d}x)\delta_1(m). \tag{5.16}$$

PROOF. Recall definition (5.10) and the fact that $\Psi_k\colon \mathfrak{X}_k \times \mathfrak{F}_k \to \Gamma^{(1)}_{T,k}$ is a bijection. Then

$$\mathbb{Q}_k^{(T)}(\Gamma^{(1)}_{T,k}) = \left(\bigotimes_{i=1}^{|k|-1} \mathbf{K}_{\phi_{i-1}}\right)(\mathfrak{X}_k)\left(\bigotimes_{i=1}^{|k|-1} \mathrm{d}s_i\right)(\mathfrak{F}_k).$$

Note that

$$\left(\bigotimes_{i=1}^{|k|-1} \mathrm{d}s_i\right)(\mathfrak{F}_k) = \int_{[0,T]^{|k|-1}} \big(\mathrm{d}(s_1,\dots,s_{|k|-1})\big)\mathbb{1}\left\{\sum_{i=1}^{|k|-1} s_i \le T\right\} = \frac{T^{|k|-1}}{(|k|-1)!}.$$

We now generalise the definition of $\mathfrak{X}_k$ in order to write down the recursion. We fix a point measure $\phi \in \mathcal{M}_{\mathbb{N}_0}(\mathcal{S}\times\mathbb{N})$ and define for any tuple $((y_i, y_i', z_i)_{i=1,\dots,|\phi|-1}$ the finite sequence

$$\phi_i = \phi + \sum_{j=1}^{i} W^{(z_j)}_{y_j, y_j'}, \quad \text{for } i = 1,\dots,|\phi|-1,$$

where we recall the definition of $W^{(z)}_{(y,y')}$ given in (5.2). We denote by $\mathfrak{X}_\phi$ the set of all tuples $((y_i, y_i', z_i)_{i=1,\dots,|\phi|-1}$ that are compatible, that is, that $\phi_i \ge 0$ for all $i = 1,\dots,|\phi|-1$. Observe that

$$\mathfrak{X}_\phi = \bigcup_{(y,y',z)\in \operatorname{supp}(\phi)^2\times\mathcal{S}} \{(y,y',z)\} \times \mathfrak{X}_{\phi-\delta_y-\delta_{y'}+\delta_{(z,m(y)+m(y'))}},$$

where for any $y = (x,m) \in \mathcal{S}\times\mathbb{N}$ we abbreviated $m(y) = m$. This implies that

$$\begin{aligned}\sigma(\phi) &:= \left(\bigotimes_{i=1}^{|\phi|-1} \mathbf{K}_{\phi_{i-1}}\right)(\mathfrak{X}_\phi) \\ &= \sum_{(x,m),(x',m')} \int_{\mathcal{S}} \mathbf{K}_\phi\big((x,m),(x',m'),\mathrm{d}z\big) \\ &\quad \times \left(\bigotimes_{i=2}^{|\phi|-1} \mathbf{K}_{\phi_{i-1}}\right)(\mathfrak{X}_{\phi-\delta_{(x,m)}-\delta_{(x',m')}+\delta_{(z,m+m')}}),\end{aligned}$$

which is the claimed recursion and implies the result. □

LEMMA 5.4 (Bounds on the total mass of $\mathbb{Q}^{(T)}$). *Assume that the kernel K satisfies* (1.2) *with constant H. Then for any $k \in \mathcal{M}_{\mathbb{N}_0}(\mathcal{S})$, $k \ne 0$, we have that*

$$\mathbb{Q}_k^{(T)}(\Gamma^{(1)}_{T,k}) \le \frac{(TH)^{|k|-1}}{(|k|-1)!}|k|^{2(|k|-1)}. \tag{5.17}$$

PROOF. We use Lemma 5.3 and show via induction over $|\phi|$ that

$$\sigma(\phi) \le H^{|\phi|-1}\|\phi\|_1^{2(|\phi|-1)}. \tag{5.18}$$

For $|\phi| = 1$, both sides of (5.18) are equal to 1. Fix any $\phi \in \mathcal{M}(\mathcal{S}\times\mathbb{N})$ with $|\phi| \ge 2$. We use the recursion from Lemma 5.3, the induction hypothesis, and the fact that for any $\widetilde{\phi} = \phi - \delta_{(x,m)} - \delta_{(x',m')} + \delta_{(z,m+m')}$ we have that $|\widetilde{\phi}| = |\phi| - 1$ and $\|\widetilde{\phi}\|_1 = \|\phi\|_1$. Also, recall from (1.2) that the definition of H implies that $\langle\phi, K\phi\rangle \le H\|\phi\|_1^2$. Then

$$\begin{aligned}\sigma(\phi) &\le H^{|\phi|-2}\|\phi\|_1^{2(|\phi|-2)} \sum_{(x,m),(x',m')} K_\phi\big((x,m),(x',m')\big) \\ &\le H^{|\phi|-2}\|\phi\|_1^{2(|\phi|-2)}\langle\phi, K\phi\rangle \le H^{|\phi|-1}\|\phi\|_1^{2(|\phi|-1)}.\end{aligned}$$

This implies (5.17). □

REMARK 5.5 (Examples of explicit expressions for $\mathbb{Q}_k^{(T)}(\Gamma_{T,k}^{(1)})$). In some specific cases, the recursive definition of $\mathbb{Q}_k^{(T)}(\Gamma_{T,k}^{(1)})$ in (5.16) can be solved, and the quantity can be expressed explicitly for every $k \in \mathcal{M}_{\mathbb{N}_0}(\mathcal{S}) \setminus \{0\}$. Notable cases are the nonspatial kernels of multiplicative and of additive type. When $K((x,m),(x',m')) = mm'$, then $\mathbb{Q}_k^{(T)}(\Gamma_{T,k}^{(1)}) = \frac{T^{|k|-1}}{(|k|-1)!}|k|^{2(|k|-1)}$. When $K((x,m),(x',m')) = m+m'$, then $\mathbb{Q}_k^{(T)}(\Gamma_{T,k}^{(1)}) = \frac{T^{|k|-1}}{2^{|k|-1}}|k|!$.

5.2. *Proof of Lemma* 2.5: *Convergence of* ρ. In this section we prove Lemma 2.5, that is, the continuity of the map $\nu \mapsto \rho(\nu)$ defined in (2.11). We use the characterisation of the rate function from Section A.4 in terms of the relative entropy, which implies that finiteness of $I_\mu^{(T)}(\nu)$ is equivalent to ν having a density with respect to $M_{b\mu}^{(T)}$ for some $b > 0$, where $M_{b\mu}^{(T)}$ is defined as in (A.9). Let us start by showing the continuity of every marginal.

LEMMA 5.6 (Continuity of $\nu \mapsto \rho_t(\nu)$). *Fix any $\beta > 0$ and some function $f : \mathbb{N} \to [0,\infty)$ that grows at infinity faster than linear, that is, $f(r)/r \to \infty$ as $r \to \infty$. Let $(\nu_n)_{n\in\mathbb{N}}$ be a sequence in $\mathcal{A}_{f,\beta}$ that converges toward some ν that has a density with respect to $M_{b\mu}^{(T)}$ for some $b > 0$. Then for any $t \in [0,T]$, $\rho_t(\nu_n) \to \rho_t(\nu)$ as $n \to \infty$.*

PROOF. Fix $A \subset \mathcal{S} \times \mathbb{N}$ with $\rho_t(\nu)(\partial A) = 0$. It suffices to show that $\rho_t(\nu_n)(A) \to \rho_t(\nu)(A)$ as $n \to \infty$. Note that, for any $L \in \mathbb{N}$,

$$\rho_t(\nu_n)(A) = \int \nu_n(\mathrm{d}\xi)\xi_t(A)\mathbb{1}\{|\xi_0| \le L\} + \int \nu_n(\mathrm{d}\xi)\xi_t(A)\mathbb{1}\{|\xi_0| > L\}.$$

The last term vanishes uniformly in A and n as $L \to \infty$, since

$$\int \nu_n(\mathrm{d}\xi)\xi_t(A)\mathbb{1}\{|\xi_0| > L\} \le \int \nu_n(\mathrm{d}\xi)|\xi_0|\mathbb{1}\{|\xi_0| > L\} \le \varepsilon_L \beta, \tag{5.19}$$

where $\varepsilon_L = \inf_{r>L} r/f(r)$ vanishes as $L \to \infty$.

Concerning the first term, we now show that the map $\xi \mapsto \xi_t(A)\mathbb{1}\{|\xi_0| \le L\}$ is continuous in each ξ that satisfies $\xi_t(\partial A) = 0$ and does not jump in t. Let ξ be such a point, and pick a sequence $(\xi^{(n)})_{n\in\mathbb{N}}$ that converges to ξ. Then $|\xi_0^{(n)}| \to |\xi_0|$ as $n \to \infty$. If $|\xi_0| > L$, then we have $\lim_{n\to\infty} \xi_t^{(n)}(A)\mathbb{1}\{|\xi_0^{(n)}| \le L\} \to 0 = \xi_t(A)\mathbb{1}\{|\xi_0| \le L\}$. Otherwise, for any sufficiently large n (recall that $|\xi_s| \in \mathbb{N}$ for any s), we have $\xi_t^{(n)}(A)\mathbb{1}\{|\xi_0^{(n)}| \le L\} = \xi_t^{(n)}(A) \to \xi_t(A) = \xi_t(A)\mathbb{1}\{|\xi_0| \le L\}$ because $\xi_t^{(n)} \to \xi_t$ weakly (since ξ is continuous in t) since $\xi_t(\partial A) = 0$.

Finally we need to show that the set of considered ξ exhausts all ξ, that is, that $\nu(\{\xi : \xi_t(\partial A) > 0\}) = 0$ and $\nu(\{\xi : \xi \text{ jumps at } t\}) = 0$. The first holds since $\xi_t(\partial A)$ is $\mathbb{N}_0$-valued, and hence $\nu(\{\xi : \xi_t(\partial A) > 0\}) \le \int \nu(\mathrm{d}\xi)\xi_t(\partial A) = \rho_t(\nu)(\partial A) = 0$. The second holds since ν has a density with respect to $M_{b\mu}^{(T)}$, and the latter has a density with respect to the distribution of the Marcus–Lushnikov process, which does not jump with positive probability at time t. $\square$

Now we prove the continuity of the map $\nu \mapsto \rho(\nu)$ defined in (2.11).

PROOF OF LEMMA 2.5.. We are going to show that $\lim_{n\to\infty} \sup_{t\in[0,T]} \mathfrak{d}(\rho_t(\nu_n), \rho_t(\nu)) = 0$, where $\mathfrak{d}$ is the Lévy–Prohorov metric on $\mathcal{M}(\mathcal{S} \times \mathbb{N})$ defined in (A.1). This implies convergence of $\rho(\nu_n)$ toward $\rho(\nu)$ with respect to the J_1-topology on $\mathbb{D}_T(\mathcal{M}(\mathcal{S} \times \mathbb{N}))$. With a small parameter $\varepsilon > 0$, we decompose $[0,T]$ into pieces $I_i^{(\varepsilon)} = [t_{i-1}, t_i]$ of length $\le \varepsilon$ and use the triangle inequality to estimate

$$\sup_{t\in[0,T]} \mathfrak{d}(\rho_t(\nu_n), \rho_t(\nu)) \le \max_i \Big[\mathfrak{d}(\rho_{t_i}(\nu_n), \rho_{t_i}(\nu)) + \sup_{t\in I_i^{(\varepsilon)}} \mathfrak{d}(\rho_t(\nu_n), \rho_{t_i}(\nu_n)) + \sup_{t\in I_i^{(\varepsilon)}} \mathfrak{d}(\rho_t(\nu), \rho_{t_i}(\nu))\Big]. \tag{5.20}$$

The first of the three terms on the right vanishes as $n\to\infty$, according to Lemma 5.6. We estimate now the second. More generally, we give an upper bound for

$$\sup_{s,t\in[0,T]:\ |s-t|\le\varepsilon} \mathfrak{d}(\rho_t(\nu_n), \rho_s(\nu_n)). \tag{5.21}$$

Let $s,t\in[0,T]$ be such that $s<t$ and $|s-t|\le\varepsilon$. According to the definition of the Lévy–Prohorov metric, if we can find some $\eta>0$ such that

$$\rho_t(\nu_n)(A) \le \rho_s(\nu_n)(A^\eta)+\eta \quad \text{and} \quad \rho_s(\nu_n)(A) \le \rho_t(\nu_n)(A^\eta)+\eta$$

holds for all measurable $A\subset\mathcal{S}\times\mathbb{N}$, then $\mathfrak{d}(\rho_t(\nu_n),\rho_s(\nu_n))\le\eta$. For any $\eta>0$, we have the estimate

$$\rho_t(\nu_n)(A) \le \rho_t(\nu_n)(A^\eta) \le \rho_s(\nu_n)(A^\eta) + \int \nu_n(\mathrm{d}\xi)\big|\xi_t(A^\eta)-\xi_s(A^\eta)\big|$$

and the same for s and t exchanged. For the latter term, we can estimate

$$\int \nu_n(\mathrm{d}\xi)\big|\xi_t(A^\eta)-\xi_s(A^\eta)\big| \le 2\int \nu_n(\mathrm{d}\xi) J_{[s,t]}(\xi),$$

where $J_I(\xi)$ is the number of jumps of ξ in the interval $I\subset[0,T]$. This is true since all steps of ξ are of the form of adding $\delta_{(z,m+m')}-\delta_{(x,m)}-\delta_{(y,m')}$ for some $x,y,z\in\mathcal{S}$ and $m,m'\in\mathbb{N}$.

Now, for any interval $I\subset[0,T]$ and $L\in\mathbb{N}$, we define

$$\mathcal{J}_I^{(L)}(\nu) = \int \nu(\mathrm{d}\xi) J_I(\xi) \mathbb{1}\{|\xi_0|\le L\}, \quad \nu\in\mathcal{M}(\Gamma_T^{(1)}). \tag{5.22}$$

Splitting the integrals into the cases $|\xi_0|\le L$ and $|\xi_0|>L$ and proceeding as in (5.19), we get, for any $\varepsilon,\eta>0$,

$$\begin{aligned}\sup_{s,t\in[0,T]:\ |s-t|\le\varepsilon} \mathfrak{d}(\rho_t(\nu_n),\rho_s(\nu_n)) &\le 2\sup_{s,t\in[0,T]:\ |s-t|\le\varepsilon} \mathcal{J}_{[s,t]}^{(L)}(\nu_n) + 2\beta\varepsilon_L \\ &= 2\mathcal{J}_{[s_n,t_n]}^{(L)}(\nu_n) + 2\beta\varepsilon_L,\end{aligned}$$

where (s_n,t_n) are picked as a maximising pair in $[0,T]^2$ with $|s_n-t_n|\le\varepsilon$.

Now, along subsequences, we may assume that $(s_n,t_n)\to(s,t)\in[0,T]^2$ such that $|s-t|\le\varepsilon$. For a given $\delta>0$ and all sufficiently large n in this subsequence, we have $[s_n,t_n]\subset[s-\delta,t+\delta]\cap[0,T]$. Furthermore, observe that $\xi\mapsto J_{[s-\delta,t+\delta]\cap[0,T]}(\xi)$ is upper semicontinuous. Indeed, if $\xi_n\to\xi$ in $\Gamma_T^{(1)}$, each jump of ξ_n converges to a jump of ξ, and since each element is right-continuous, we have that $J_{[s-\delta,t+\delta]\cap[0,T]}(\xi_n)\le J_{[s-\delta,t+\delta]\cap[0,T]}(\xi)$. Hence we have, by [19], Theorem D.12,

$$\limsup_{n\to\infty} \mathcal{J}_{[s_n,t_n]}^{(L)}(\nu_n) \le \mathcal{J}_{[s-\delta,t+\delta]\cap[0,T]}^{(L)}(\nu). \tag{5.23}$$

We show now that the right-hand side is not larger than $C_L(\varepsilon+2\delta)$ for some $C_L>0$ that does not depend on s or on t. For doing this, we note that

$$M_{b\mu}^{(T)}(A) < \mathrm{e}^{-1} \quad\Longrightarrow\quad \nu(A) \le \frac{H(\nu|M_{b\mu}^{(T)})+\mathrm{e}^{-1}}{-1-\log M_{b\mu}^{(T)}(A)}, \quad A\subset\mathcal{A}_{f,\beta}. \tag{5.24}$$

Indeed, note that $x \mapsto x \log x + 1 - x$ is nonnegative and convex in $(0, \infty)$, and, therefore,

$$
\begin{aligned}
H\big(\nu|M_{b\mu}^{(T)}\big) &\geq \int \Big(\frac{\mathrm{d}\nu}{\mathrm{d}M_{b\mu}^{(T)}} \log \frac{\mathrm{d}\nu}{\mathrm{d}M_{b\mu}^{(T)}} + 1 - \frac{\mathrm{d}\nu}{\mathrm{d}M_{b\mu}^{(T)}}\Big) \mathbf{1}_A \, \mathrm{d}M_{b\mu}^{(T)} \\
&\geq M_{b\mu}^{(T)}(A) \Big(\frac{\nu(A)}{M_{b\mu}^{(T)}(A)} \log \frac{\nu(A)}{M_{b\mu}^{(T)}(A)} + 1 - \frac{\nu(A)}{M_{b\mu}^{(T)}(A)}\Big) \\
&\geq \nu(A) \log \nu(A) + \nu(A)\big(-1 - \log M_{b\mu}^{(T)}(A)\big) \\
&\geq -\mathrm{e}^{-1} + \nu(A)\big(-1 - \log M_{b\mu}^{(T)}(A)\big),
\end{aligned}
$$

where the second inequality is obtained thanks to Jensen's inequality, the third because of $M_{b\mu}^{(T)}(A) \geq 0$, and the last one thanks to the fact that $x \log x \geq -\mathrm{e}^{-1}$. We see that, when $M_{b\mu}^{(T)}(A) < \mathrm{e}^{-1}$, the bracket on the last line is positive, and we obtain (5.24).

Now note that the right-hand side of (5.23) may be bounded as

$$
\mathcal{J}_{[s-\delta,t+\delta]\cap[0,T]}^{(L)}(\nu) \leq \sum_{j=1}^{L} \nu\big(\{\xi : J_{[s-\delta,t+\delta]\cap[0,T]}(\xi) \geq j, |\xi_0| \leq L\}\big). \tag{5.25}
$$

Before bounding the right-hand side, we are going to argue now that

$$
M_{b\mu}^{(T)}\big(\{\xi : J_{[s-\delta,t+\delta]\cap[0,T]}(\xi) \geq j, |\xi_0| \leq L\}\big) \leq \widetilde{C}_L(\varepsilon + 2\delta)^j \tag{5.26}
$$

for all $j \in \{1, \dots, L\}$ and some $\widetilde{C}_L > 0$ that does not depend on s or on t. Recall that $M_{b\mu}^{(T)}(\mathrm{d}\xi) = \mathrm{e}^{2-b} \operatorname{Poi}_\mu \otimes \mathbb{Q}^{(T)}(\mathrm{d}\xi) b^{|\xi_0|}$ and the formula for $\mathbb{Q}^{(T)}$ from (5.10). For $\xi \in \Gamma_T^{(1)}$, let $k = \xi_0$, and let $(y_i, y_i', z_i, s_i)_{i=1,\dots,|k|-1} \in \mathfrak{X}_k \times \mathfrak{F}_k$ be such that $\xi = \Psi_k((y_i, y_i', z_i, s_i)_{i=1,\dots,|k|-1})$. Abbreviate $I = [s - \delta, t + \delta] \cap [0, T]$, and note that the value of $J_I(\xi)$ only depends on the intercoagulation times $(s_i)_{i=1,\dots,|k|-1} \in \mathfrak{F}_k$, which are i.i.d. with distribution given by the Lebesgue measure (see (5.10)). More precisely, we can recover the jump times of ξ via $t_i = s_1 + \cdots + s_i$, for $i = 1, \dots, |k| - 1$, and have that $J_I(\xi) = \#\{i : t_i \in I\}$. Then

$$
\begin{aligned}
&\int_{\mathfrak{F}_k} \bigotimes_{i=1}^{|k|-1} \mathrm{d}s_i \mathbb{1}\big\{\#\{i : t_i \in I\} \geq j\big\} \\
&\quad = \int_{[0,T]^{|k|-1}} \bigotimes_{i=1}^{|k|-1} \mathrm{d}t_i \mathbb{1}\big\{\#\{i : t_i \in I\} \geq j\big\} \mathbb{1}\{t_1 < t_2 < \cdots < t_{|k|-1}\} \\
&\quad = \frac{1}{(|k|-1)!} \int_{[0,T]^{|k|-1}} \bigotimes_{i=1}^{|k|-1} \mathrm{d}t_i \mathbb{1}\big\{\#\{i : t_i \in I\} \geq j\big\} = \sum_{\ell \geq j} \frac{|I|^\ell (T - |I|)^{|k|-1-\ell}}{\ell!(|k| - 1 - \ell)!},
\end{aligned}
$$

where the last term is smaller than $|I|^j (2T)^{|k|-1}/(|k| - 1)!$, if $|I|$ is small, which we can assume without loss of generality. The terms of $\mathbb{Q}^{(T)}(\mathrm{d}\xi)$ that depend on $(y_i, y_i', z_i)_{i=1,\dots,|k|-1} \in \mathfrak{X}_k$ can be estimated as in Lemma 5.4. Altogether, we get that

$$
\begin{aligned}
&M_{b\mu}^{(T)}\big(\{\xi : J_{[s-\delta,t+\delta]\cap[0,T]}(\xi) \geq j, |\xi_0| \leq L\}\big) \\
&\qquad \leq (\varepsilon + 2\delta)^j \sum_{n=1}^{L} b^n (2HT)^{n-1} \frac{n^{2(n-1)}}{n!(n-1)!}.
\end{aligned} \tag{5.27}
$$

Via (5.24) and the assumption that $H(\nu|M_{b\mu}^{(T)}) < \infty$, this implies that also the right-hand side of (5.23) is not larger than $C_L(\varepsilon + 2\delta)$ for some $C_L > 0$ that does not depend on s or on t.

Summarizing, we have shown that, for any $L \in \mathbb{N}$ and $\varepsilon, \delta > 0$,

$$\limsup_{n\to\infty} \sup_{s,t\in[0,T]:\, |s-t|\le\varepsilon} \mathfrak{d}\big(\rho_t(\nu_n), \rho_s(\nu_n)\big) \le 2C_L(\varepsilon+\delta) + 2\beta\varepsilon_L.$$

We first pick L large enough such that ε_L is small enough. Since the left-hand side does not depend on δ, we may make $\delta \downarrow 0$ on the right-hand side. In an analogous way we derive the same bound for $\sup_{t,s\in[0,T]:\, |s-t|\le\varepsilon} \mathfrak{d}(\rho_t(\nu), \rho_s(\nu))$. This implies via (5.20) our assertion. □

As a byproduct of the above proof, we have the following result.

COROLLARY 5.7 ($\rho(\nu)$ is a (uniformly) continuous path). *Fix any $\beta > 0$ and some function $f : \mathbb{N} \to [0, \infty)$ that grows at infinity faster than linear, that is, $f(r)/r \to \infty$ as $r \to \infty$. Then if $\nu \in \mathcal{M}(\Gamma_T^{(1)})$ is such that $H(\nu|M_{b\mu}^{(T)}) < \infty$, then $[0, T] \ni t \mapsto \rho_t(\nu) \in \mathcal{M}(\mathcal{S} \times \mathbb{N})$ is uniformly continuous.*

6. Proof of Theorem 2.3: The LDP for $\mathcal{V}_N^{(T)}$. In this section we prove our second main result, the LDP of Theorem 2.3, that is, the LDP for $\mathcal{V}_N^{(T)}$ with Poissonised initial distribution under conditioning on $\mathcal{V}_N^{(T)} \in \mathcal{A}_{f,\beta}$, where we fixed a majorizing function f such that $f(r) \ge r$ for any r and $f(r)/r \to \infty$ as $r \to \infty$ and a constant $\beta \in (0, \infty)$. We rely on the representation of Theorem 2.1 in terms of the PPP Y_N. We consider the empirical measure $\mathcal{V}_N^{(T)}$ with underlying kernel $\frac{1}{N}K$ (instead of K) and indicate this replacement of the kernel by adding the superindex ${}^{(N)}$ to the probability measure; that is, we consider the distribution of $\mathcal{V}_N^{(T)}$ under $\mathbb{P}^{(N)}_{\mathrm{Poi}_{N\mu}}$. This change also affects the intensity measure of the PPP, when applying Theorem 2.1 which is now given as $NM_{\mu,N}^{(T,N)}$, that is,

$$M_{\mu,N}^{(T,N)}(\mathrm{d}\xi) = N^{|\xi_0|-1}\mathrm{e}\mathbb{P}^{(N)}_{\mathrm{Poi}_\mu}(\Xi \in \mathrm{d}\xi), \quad \text{for } \xi \in \Gamma_T^{(1)}. \tag{6.1}$$

Our proof relies on deriving an LDP for $\frac{1}{N}Y_N$ and then to apply Varadhan's lemma. The statement of the LDP for $\frac{1}{N}Y_N$ can be found in the Appendix, Section A.3. There are several technical problems to solve. The first one is that the total mass of the N-limit of $M_{\mu,N}^{(T,N)}$ might be infinite, and in that case we do not have an LDP for $\frac{1}{N}Y_N$; see also Section A.4 on some issues related to this. Hence, we will replace $M_{\mu,N}^{(T,N)}$ by $M_{b\mu,N}^{(T,N)}$ for a small constant $b \in (0, 1)$, as defined in (6.1), with μ replaced by $b\mu$. Note that

$$\frac{\mathrm{d}\,\mathrm{Poi}_\mu}{\mathrm{d}\,\mathrm{Poi}_{b\mu}}(k) = \mathrm{e}^{b-1}b^{-|k|}, \quad k \in \mathcal{M}_{\mathbb{N}_0}(\mathcal{S}).$$

It is then easy to see that $\mathrm{e}^{(b-1)|Y_{N,0}|}b^{-\int Y_{N,0}(\mathrm{d}k)|k|}\mathrm{e}^{N(|M_{b\mu,N}^{(T,N)}|-|M_{\mu,N}^{(T,N)}|)}$ is the density of the PPP Y_N when switching from $\mathtt{P}_{NM_{\mu,N}^{(T,N)}}$ to $\mathtt{P}_{NM_{b\mu,N}^{(T,N)}}$.

This implies for any measurable set $E \subset \Gamma_T^{(1)}$ that

$$\begin{aligned} &\mathbb{P}^{(N)}_{\mathrm{Poi}_{N\mu}}(\mathcal{V}_N^{(T)} \in E) \\ &\quad = \mathtt{E}_{NM_{b\mu,N}^{(T,N)}}\Big[\mathrm{e}^{-\frac{1}{2N}\sum_{i\ne j} R^{(T)}(\Xi_i,\Xi_j)}\mathrm{e}^{(b-1)|Y_{N,0}|}b^{-\int Y_{N,0}(\mathrm{d}k)|k|}\mathbb{1}\Big\{\frac{1}{N}Y_N \in E\Big\}\Big] \\ &\qquad \times \mathrm{e}^{N(|M_{b\mu,N}^{(T,N)}|-1)}, \end{aligned} \tag{6.2}$$

where $Y_N = \sum_i \delta_{\Xi_i}$ is a PPP with intensity measure $NM_{b\mu,N}^{(T,N)}$.

Now we use the following corollary, where the density of $M_{b\mu,N}^{(T,N)}$ with respect to $M_{b\mu}^{(T)}$ is given.

COROLLARY 6.1. *Fix* $\mu \in \mathcal{M}_1(\mathcal{S})$ *and* $b \in (0,\infty)$ *and* $N \in \mathbb{N}$, *and replace* K *by* $\frac{1}{N}K$ (*adding a superscript* ${}^{(N)}$). *Then*

$$M_{b\mu,N}^{(T,N)}(\mathrm{d}\xi) = M_{b\mu}^{(T)}(\mathrm{d}\xi)\mathrm{e}^{-\frac{1}{N}\varphi_{\xi_0}(\xi)}. \tag{6.3}$$

We rewrite the right-hand side of (6.2) in such a way that the intensity measure of the reference PPP does not depend on N (up to the prefactor N): we just carry out a change of measure from $NM_{b\mu,N}^{(T,N)}$ to $NM_{b\mu}^{(T)}$ in the Poissonian expectation. This gives, for any measurable set $E \subset \Gamma_T^{(1)}$,

$$\begin{aligned}&\mathbb{P}^{(N)}_{\mathrm{Poi}_{N\mu}}\big(\mathcal{V}_N^{(T)} \in E\big)\\ &\quad= \mathbb{E}_{NM_{b\mu}^{(T)}}\Big[\mathrm{e}^{N\phi_b(\frac{1}{N}Y_N)}\mathbb{1}\Big\{\frac{1}{N}Y_N \in E\Big\}\mathrm{e}^{\frac{1}{2N}\sum_i R^{(T)}(\Xi_i,\Xi_i)}\mathrm{e}^{-\int\varphi_{\xi_0}(\xi)\frac{1}{N}Y_N(\mathrm{d}\xi)}\Big],\end{aligned} \tag{6.4}$$

where

$$\phi_b(\nu) = -\frac{1}{2}\big\langle \nu, \mathfrak{R}^{(T)}(\nu)\big\rangle + |\nu|(b-1) - |c_{\nu_0}|\log b + \big|M_{b\mu}^{(T)}\big| - 1. \tag{6.5}$$

Note that we also used here that

$$\frac{1}{2N}\sum_{i\neq j} R^{(T)}(\Xi_i,\Xi_j) = \frac{N}{2}\Big\langle \frac{1}{N}Y_N, \mathfrak{R}^{(T)}\Big(\frac{1}{N}Y_N\Big)\Big\rangle - \frac{1}{2N}\sum_i R^{(T)}(\Xi_i,\Xi_i). \tag{6.6}$$

Now we observe that the last two terms in the expectation on the right-hand side of (6.4) are $\mathrm{e}^{o(N)}$, as $N\to\infty$, uniformly on $\{\frac{1}{N}Y_N \in \mathcal{A}_{f,\beta}\}$ by Lemmas A.3 and 5.2. Furthermore, it is not hard to see that ϕ_b is bounded and continuous on $\mathcal{A}_{f,\beta}$ w.r.t. the weak topology: We show in the Appendix in Lemma A.2(ii) that $\nu \mapsto \langle \nu, \mathfrak{R}^{(T)}(\nu)\rangle$ is continuous on $\mathcal{A}_{f,\beta}$. The continuity of $\nu \mapsto |\nu|$ and $\nu \mapsto |c_{\nu_0}|$ follows similarly, and both maps are bounded by β. Furthermore, note that $\mathcal{A}_{f,\beta}$ is closed, due to the continuity of $\xi \mapsto |\xi_0|$ and nonnegativity of f, using Fatou's lemma.

Now the LDP for $\frac{1}{N}Y_N$ under $\mathrm{P}_{NM_{b\mu}^{(T)}}$ from Lemma A.4, together with Varadhan's lemma (Lemma 4.3.6 in [19]), implies, for any closed set $F \subset \mathcal{M}(\Gamma_T^{(1)})$ (implying that also $F \cap \mathcal{A}_{f,\beta}$ is closed), that

$$\begin{aligned}&\limsup_{N\to\infty}\frac{1}{N}\log\mathbb{P}^{(N)}_{\mathrm{Poi}_{N\mu}}\big(\mathcal{V}_N^{(T)} \in F\cap\mathcal{A}_{f,\beta}\big)\\ &\qquad\le -\inf\big\{H\big(\nu|M_{b\mu}^{(T)}\big) - \phi_b(\nu) \colon \nu \in F\cap\mathcal{A}_{f,\beta}\big\}.\end{aligned} \tag{6.7}$$

Observe from (A.12) that $H(\nu|M_{b\mu}^{(T)}) - \phi_b(\nu) = I_\mu^{(T)}(\nu)$ for any $\nu \in \mathcal{M}(\Gamma_T^{(1)})$. In particular, recalling that $\chi_\beta = \inf_{\nu\in\mathcal{A}_{f,\beta}} I_\mu^{(T)}(\nu)$,

$$\limsup_{N\to\infty}\frac{1}{N}\log\mathbb{P}^{(N)}_{\mathrm{Poi}_{N\mu}}\big(\mathcal{V}_N^{(T)} \in \mathcal{A}_{f,\beta}\big) \le -\chi_\beta. \tag{6.8}$$

In order to finish the proof of Theorem 2.3, we need to argue that also the complementary lower bound holds for (6.7) for F replaced by some open set $G \subset \mathcal{M}(\Gamma_T^{(1)})$. This will then imply the corresponding lower bound in (6.8), which finishes the proof.

In this point, there is a technical problem, since the set $\mathcal{A}_{f,\beta}$ is not open in $\mathcal{M}(\Gamma_T^{(1)})$. An obvious idea is to go to the set

$$\mathcal{A}_{f,<\beta} = \Big\{\nu \in \mathcal{M}(\Gamma_T^{(1)}) \colon \int \nu(\mathrm{d}\xi) f\big(|\xi_0|\big) < \beta\Big\}. \tag{6.9}$$

However, $\mathcal{A}_{f,<\beta}$ is still not an open set, since the map $\nu \mapsto \int \nu(\mathrm{d}\xi) f(|\xi_0|)$ is not continuous (the map $\xi \mapsto f(|\xi_0|)$ not bounded). We solve this by applying some restriction argument. Indeed, for a large cutting parameter $L \in \mathbb{N}$, we insert an indicator on the event that the PPP Y_N has no particles that are larger than L, that is, that it is concentrated on

$$\Gamma^{(1)}_{T,\le L} = \{\xi \in \Gamma^{(1)}_T : \xi_t \in \mathcal{M}_{\mathbb{N}_0}(\mathcal{S} \times [L]) \; \forall t \in [0, T]\}. \tag{6.10}$$

Then we will condition on this event and make a change of measure from the PPP Y_N with intensity measure $NM^{(T)}_{b\mu}$ to the intensity measure $NM^{(T,\le L)}_{b\mu}$, the restriction of $NM^{(T)}_{b\mu}$ to $\Gamma^{(1)}_{T,\le L}$.

In the following we often identify measures ν on $\Gamma^{(1)}_T$ that satisfy $\nu(\Gamma^{(1)}_T \setminus \Gamma^{(1)}_{T,\le L}) = 0$ with measures on $\Gamma^{(1)}_{T,\le L}$. We introduce the restriction operator $\Pi_L : \mathcal{M}(\Gamma^{(1)}_T) \to \mathcal{M}(\Gamma^{(1)}_T)$, which maps ν to the measure $\Pi_L(\nu) = \nu^{(\le L)}$, defined by $\nu^{(\le L)}(\cdot) = \nu(\cdot \cap \Gamma^{(1)}_{T,\le L})$. Note that the mapping Π_L is continuous with respect to weak convergence. One easily sees, by definition of PPP, that the distribution of Y_N under $\mathtt{E}_{NM^{(T)}_{b\mu}}$, conditioned on $\{Y_N(\Gamma^{(1)}_T \setminus \Gamma^{(1)}_{T,\le L}) = 0\}$, is equal to its distribution under $\mathtt{E}_{NM^{(T,\le L)}_{b\mu}}$. This implies that

$$\begin{aligned}
&\mathbb{P}^{(N)}_{\mathrm{Poi}_{N\mu}}(\mathcal{V}^{(T)}_N \in G \cap \mathcal{A}_{f,\beta}) \\
&\quad \ge \mathbb{P}^{(N)}_{\mathrm{Poi}_{N\mu}}(\mathcal{V}^{(T)}_N \in G \cap \mathcal{A}_{f,<\beta}) \\
&\quad \ge \mathrm{e}^{o(N)} \mathtt{E}_{NM^{(T)}_{b\mu}}\Big[\mathrm{e}^{N\phi_b(\frac{1}{N}Y_N)} \mathbb{1}\Big\{\frac{1}{N}Y_N \in G \cap \mathcal{A}_{f,<\beta}\Big\}\Big| Y_N\big(\Gamma^{(1)}_T \setminus \Gamma^{(1)}_{T,\le L}\big) = 0\Big] \\
&\qquad \times \mathrm{Poi}_{NM^{(T)}_{\mu}}\big(Y_N\big(\Gamma^{(1)}_T \setminus \Gamma^{(1)}_{T,\le L}\big) = 0\big) \\
&\quad = \mathrm{e}^{o(N)} \mathtt{E}_{NM^{(T,\le L)}_{b\mu}}\Big[\mathrm{e}^{N\phi_b(\frac{1}{N}Y_N)} \mathbb{1}\Big\{\frac{1}{N}Y_N \in G \cap \mathcal{A}_{f,<\beta}\Big\}\Big] \\
&\qquad \times \mathrm{Poi}_{NM^{(T)}_{\mu}}\big(Y_N\big(\Gamma^{(1)}_T \setminus \Gamma^{(1)}_{T,\le L}\big) = 0\big).
\end{aligned} \tag{6.11}$$

Note that

$$-\delta_L = \liminf_{N\to\infty} \frac{1}{N} \log \mathrm{Poi}_{NM^{(T)}_{\mu}}\big(Y_N\big(\Gamma^{(1)}_T \setminus \Gamma^{(1)}_{T,\le L}\big) = 0\big) \tag{6.12}$$

increases to zero as $L \to \infty$. Indeed, this void probability is equal to $\mathrm{e}^{-NM^{(T)}_{b\mu}(\Gamma^{(1)}_T \setminus \Gamma^{(1)}_{T,\le L})}$, and the rate vanishes as $L \to \infty$.

Recall that Y_N in the expectation on the right-hand side of (6.11) is a PPP with intensity measure $M^{(T,\le L)}_{b\mu}$, and hence $Y_N = Y^{(\le L)}_N$ almost surely. Hence, we can rewrite the condition in the indicator on the right-hand side of (6.11) as $\{\frac{1}{N}Y_N \in \Pi^{-1}_L(G \cap \mathcal{A}_{f,<\beta})\}$. Now, $\Pi^{-1}_L(G)$ is open, since Π_L is continuous and G is open. Further, $\Pi^{-1}_L(\mathcal{A}_{f,<\beta})$ is equal to the set of all $\nu \in \mathcal{M}(\Gamma^{(1)}_T)$ that satisfy $\int \nu(\mathrm{d}\xi) f(|\xi_0|)\mathbb{1}\{|\xi_0| \le L\} < \beta$, which is an open set in $\mathcal{M}(\Gamma^{(1)}_T)$, since the map $\xi \mapsto f(|\xi_0|)\mathbb{1}\{|\xi_0| \le L\}$ is dominated by a continuous and bounded function.

Hence, we can apply now the lower-bound part of Varadhan's lemma (Lemma 4.3.4 in [19]) and the LDP for $\frac{1}{N}Y_N$ from Lemma A.4 with $\mathfrak{m} = M^{(T,\le L)}_{b\mu}$ to obtain that

$$\begin{aligned}
&\liminf_{N\to\infty} \frac{1}{N} \log \mathbb{P}^{(N)}_{\mathrm{Poi}_{N\mu}}(\mathcal{V}^{(T)}_N \in G \cap \mathcal{A}_{f,\beta}) \\
&\quad \ge -\inf\big\{H\big(\nu | M^{(T,\le L)}_{b\mu}\big) - \phi_b(\nu) : \nu \in \Pi^{-1}_L(G \cap \mathcal{A}_{f,<\beta})\big\} - \delta_L
\end{aligned} \tag{6.13}$$

$$\geq -\inf\{H(\nu^{(\leq L)}|M_{b\mu}^{(T,\leq L)}) - \phi_b(\nu^{(\leq L)}) \colon \nu \in G \cap \mathcal{A}_{f,<\beta}\} - \delta_L,$$

where for the last equality we used that $H(\nu|M_{b\mu}^{(T,\leq L)}) = \infty$ if $\nu(\Gamma_T^{(1)} \setminus \Gamma_{T,\leq L}^{(1)}) > 0$, and hence it suffices to take the infimum over all ν that satisfy $\nu = \nu^{(\leq L)}$. It is easy to see that

$$\begin{aligned}
&\liminf_{L\to\infty} \inf\{H(\nu^{(\leq L)}|M_{b\mu}^{(T,\leq L)}) - \phi_b(\nu^{(\leq L)}) \colon \nu \in G \cap \mathcal{A}_{f,<\beta}\} \\
&\quad \leq \inf\{H(\nu|M_{b\mu}^{(T)}) - \phi_b(\nu) \colon \nu \in G \cap \mathcal{A}_{f,<\beta}\} \qquad (6.14)\\
&\quad = \inf\{H(\nu|M_{b\mu}^{(T)}) - \phi_b(\nu) \colon \nu \in G \cap \mathcal{A}_{f,\beta}\},
\end{aligned}$$

where the last step follows from a simple approximation step (approach ν satisfying $\int \nu(\mathrm{d}\xi) f(|\xi_0|) = \beta$ by $(1-\varepsilon)\nu$ with $\varepsilon \downarrow 0$).

This shows that the complementary lower bound in (6.7) holds for F replaced by an open set G and finishes the proof of Theorem 2.3.

7. Analysis of $I_\mu^{(T)}$, (non)gelation, and the Smoluchowski equation. In this section we analyse the minimiser(s) of the rate function $I_\mu^{(T)}$ appearing in Theorem 2.3 (defined in (2.9)) and prove Theorem 2.8. In particular, in Section 7.1 we derive bounds on moments of the reference measure $M_\mu^{(T)}$ and lower bounds on $I_\mu^{(T)}$ and give criteria for the existence of minimisers of $I_\mu^{(T)}$. In Section 7.2 we derive the Euler–Lagrange equations for these minimisers and use them to prove some estimates for its moments. Then Section 7.3 is devoted to the proof of nongelation at small times (finishing the proof of Theorem 2.8, 1) and Section 7.4 to the proof of loss of mass (i.e., existence of gelation) at a late time (finishing the proof of Theorem 2.8, 2). On the way we also prove Proposition 2.10, 2 in Section 7.2 and Proposition 2.10, 1 at the end of Section 7.3; furthermore, we derive the Smoluchowski equation in Section 7.5.

For the remainder of this section, we keep $\mu \in \mathcal{M}_1(\mathcal{S})$ and $T \in (0,\infty)$ fixed and assume only that the kernel K is nonnegative and measurable in its four arguments. Recall the reference measure $M_\mu^{(T)}$ from (2.5). We recall from (2.9) that, for $\nu \in \mathcal{M}(\Gamma_T^{(1)})$ that is absolutely continuous with respect to $M_\mu^{(T)}$ (otherwise, $I_\mu^{(T)}(\nu) = \infty$),

$$\begin{aligned}
I_\mu^{(T)}(\nu) &= \left\langle \nu, \log \frac{\mathrm{d}\nu}{\mathrm{d}M_\mu^{(T)}} \right\rangle + \frac{1}{2}\langle \nu, \mathfrak{R}^{(T)}(\nu)\rangle + 1 - |\nu| \\
&= H(\nu|M_{b\mu}^{(T)}) + 1 - |M_{b\mu}^{(T)}| + (1-b)|\nu| + \int \nu_0(\mathrm{d}k)|k| \log b \qquad (7.1)\\
&\quad + \frac{1}{2}\langle \nu, \mathfrak{R}^{(T)}(\nu)\rangle,
\end{aligned}$$

where for the second line we used the alternative characterisation from (A.12) and assumed that $b \in (0,\infty)$ is so small that $|M_{b\mu}^{(T)}| < \infty$. (A sufficient criterion for this is given in Lemma A.5.)

If K is positive definite, then Remark 2.7 implies the strict convexity of $I_\mu^{(T)}$ and hence the uniqueness of the minimiser, since the domain of $I_\mu^{(T)}$ is convex. However, we are not going to use this assumption in this section; therefore, we might have several minimisers.

7.1. *Bounds on $M_\mu^{(T)}$ and on $I_\mu^{(T)}$.* Now we can state conditions under which $I_\mu^{(T)}$ has a minimiser. Recall $q_\mu^{(T)}$ from (2.18) and that the sublevel sets of $H(\cdot|M)$ are compact for any finite measure M (see the proof in Section A.3).

LEMMA 7.1 (Moments of $M_\mu^{(T)}$ under $q_\mu^{(T)} < 1$). *Fix $T \in (0,\infty)$. If $q_\mu^{(T)} < 1$, then $\int M_\mu^{(T)}(\mathrm{d}\xi)|\xi_0|^\alpha < \infty$ for any $\alpha \in [0,\infty)$. Furthermore, the sublevel sets of $I_\mu^{(T)}$ are then compact, that is, $I_\mu^{(T)}$ has at least one minimiser on $\mathcal{M}(\Gamma_T^{(1)})$.*

PROOF. Just note that

$$\begin{aligned}\int M_\mu^{(T)}(\mathrm{d}\xi)|\xi_0|^\alpha &= \sum_{n\in\mathbb{N}} M_\mu^{(T)}(\{\xi : |\xi_0| = n\})n^\alpha \\ &= \sum_{n\in\mathbb{N}} (q_\mu^{(T)})^{n+o(n)} n^\alpha,\end{aligned}$$

and a comparison to the geometric series gives the result.

Now we may use the second line of (7.1) for $b = 1$ and see that $I_\mu^{(T)}$ is equal to the sum of $H(\cdot|M_\mu^{(T)})$ (which has compact sublevel sets) and $1 - |M_\mu^{(T)}| + \frac{1}{2}\langle \nu, \mathfrak{R}^{(T)}(\nu)\rangle$, which is lower semicontinuous in ν by Lemma A.2. Hence, $I_\mu^{(T)}$ has compact sublevel sets as well and possesses, therefore, a minimiser. □

The following lower bound will be used in Section 7.4 for proving gelation for large T, more precisely, for all T such that $I_\mu^{(T)}$ is bounded away from zero on $\mathcal{M}(\Gamma_T^{(1)})$.

LEMMA 7.2 (Lower bound on $I_\mu^{(T)}$). *Under the assumptions in (1.2) and (1.3),*

$$\inf_{\nu\in\mathcal{M}(\Gamma_T^{(1)})} I_\mu^{(T)}(\nu) \ge 1 - \frac{1}{2T}\left(\frac{\mathrm{e}}{\pi H} + \frac{(\log(2TH\mathrm{e}^2))^2}{h}\right). \tag{7.2}$$

In particular, the infimum is positive for all sufficiently large T and tends to one as $T \to \infty$.

PROOF. For $T \in (0,\infty)$, pick some $b \in (0,1)$ such that $bTH\mathrm{e}^2 < 1$. Then from Lemma A.5 with $\alpha = 0$, we have (dropping the factor n^{-2} in the sum)

$$|M_{b\mu}^{(T)}| \le \frac{\mathrm{e}^3 b\mathrm{e}^{-b}}{2\pi(1 - bTH\mathrm{e}^2)} < \infty. \tag{7.3}$$

We derive from the second line of (7.1) and the nonnegativity of the relative entropy that, for any $\nu \in \mathcal{M}(\Gamma_T^{(1)})$, the following holds:

$$\begin{aligned} I_\mu^{(T)}(\nu) &\ge H(\nu|M_{b\mu}^{(T)}) + 1 - |M_{b\mu}^{(T)}| + D\log b + \frac{1}{2}\langle \nu, \mathfrak{R}^{(T)}(\nu)\rangle \\ &\ge 1 - |M_{b\mu}^{(T)}| + D\log b + \frac{1}{2}\langle \nu, \mathfrak{R}^{(T)}(\nu)\rangle, \end{aligned} \tag{7.4}$$

where we abbreviated $D = \int \nu_0(\mathrm{d}k)|k| = |c_{\nu_0}|$. With the help of (1.3), we obtain

$$\begin{aligned} \langle \nu, \mathfrak{R}^{(T)}(\nu)\rangle &= \int_0^T \mathrm{d}t \int \nu(\mathrm{d}\xi) \int \nu(\mathrm{d}\xi')\langle \xi_t, K\xi_t'\rangle \\ &\ge \int_0^T \mathrm{d}t \int \nu(\mathrm{d}\xi) \int \nu(\mathrm{d}\xi') h\|\xi_t\|_1\|\xi_t'\|_1 \\ &= h\int_0^T \mathrm{d}t \int \nu(\mathrm{d}\xi)\int \nu(\mathrm{d}\xi')|\xi_0||\xi_0'| = hTD^2. \end{aligned} \tag{7.5}$$

The polynomial $D \mapsto D \log b + \frac{1}{2} hT D^2$ assumes its minimal value at $D = \frac{-\log b}{hT}$ with value $-\frac{(\log b)^2}{2hT}$. Hence,

$$\inf_{\nu \in \mathcal{M}(\Gamma_T^{(1)})} I_\mu^{(T)}(\nu) \geq 1 - \frac{\mathrm{e}^3 b \mathrm{e}^{-b}}{2\pi(1 - bTH\mathrm{e}^2)} - \frac{(\log b)^2}{2hT}, \quad b \in \left(0, \frac{1}{TH\mathrm{e}^2}\right).$$

Picking $b = 1/(2TH\mathrm{e}^2)$, we get

$$\begin{aligned}\inf_{\nu \in \mathcal{M}(\Gamma_T^{(1)})} I_\mu^{(T)}(\nu) &\geq 1 - \frac{1}{2T}\left(\frac{\mathrm{e}^{1-(2TH\mathrm{e}^2)^{-1}}}{\pi H} + \frac{(\log(2TH\mathrm{e}^2))^2}{h}\right) \\ &\geq 1 - \frac{1}{2T}\left(\frac{\mathrm{e}}{\pi H} + \frac{(\log(2TH\mathrm{e}^2))^2}{h}\right).\end{aligned}$$

□

7.2. *Euler–Lagrange equations.* In this section we characterise minimisers of $I_\mu^{(T)}$ via the variational equalities, which we also call *Euler–Lagrange equations*. This will also lead to a proof of Proposition 2.10, 2.

LEMMA 7.3. *For any $T \in (0, \infty)$, any minimiser $\nu^{(T)}$ of $I_\mu^{(T)}$ on the set $\mathcal{M}(\Gamma_T^{(1)})$ satisfies the Euler–Lagrange equation*

$$\nu^{(T)}(\mathrm{d}\xi) = M_\mu^{(T)}(\mathrm{d}\xi)\mathrm{e}^{-\mathfrak{R}^{(T)}(\nu^{(T)})(\xi)}, \quad \xi \in \Gamma_T^{(1)}. \tag{7.6}$$

PROOF. We drop the superscript $^{(T)}$ and the index μ from the notation.

Let ν be a minimiser of I. Since $I(\nu)$ is finite, ν has a nonnegative density φ with respect to M. We show that φ is positive M-almost surely. Indeed, if there is a measurable set $B \subset \mathcal{M}(\Gamma_T^{(1)})$ with positive M-measure such that $\varphi = 0$ on B, then we are going to see that $\nu_\varepsilon(\mathrm{d}\xi) = M(\mathrm{d}\xi)(\varphi(\xi) + \varepsilon \mathbb{1}_B(\xi))$ has a strictly smaller I-value, in contradiction to the minimality of ν. Indeed, observe that

$$\begin{aligned}I(\nu_\varepsilon) - I(\nu) = \frac{1}{2}\int\int M(\mathrm{d}\xi)M(\mathrm{d}\widetilde{\xi})R(\xi, \widetilde{\xi})\big[2\varepsilon \mathbb{1}_B(\xi)\varphi(\widetilde{\xi}) + \varepsilon^2 \mathbb{1}_B(\xi)\mathbb{1}_B(\widetilde{\xi})\big] \\ + M(B)\varepsilon \log \varepsilon - \varepsilon M(B).\end{aligned}$$

Since this is $\leq M(B)\varepsilon \log \varepsilon + O(\varepsilon)$ for $\varepsilon \downarrow 0$, it is negative for sufficiently small $\varepsilon > 0$. Hence, φ is positive M-almost surely.

Now we calculate the directional derivative of I in $\nu(\mathrm{d}\xi) = M(\mathrm{d}\xi)\varphi(\xi)$ in direction of $\nu_\varepsilon(\mathrm{d}\xi) = M(\mathrm{d}\xi)(\varphi(\xi) + \varepsilon\gamma(\xi))$ (with $\varepsilon \in \mathbb{R}$) for a large class of measurable and bounded functions $\gamma \colon \Gamma_T^{(1)} \to \mathbb{R}$. We fix $\delta > 0$ and $L \in \mathbb{N}$ and assume that $\gamma = 0$ on $\{\xi \colon |\xi_0| > L, \varphi(\xi) \leq \delta\}$. Then $\varphi + \varepsilon\gamma > 0$ for all $\varepsilon \in \mathbb{R}$ with sufficiently small $|\varepsilon|$. By minimality in $\varepsilon = 0$, we have

$$\begin{aligned}0 &= \frac{\mathrm{d}}{\mathrm{d}\varepsilon} I(\nu_\varepsilon)|_{\varepsilon=0} \\ &= \frac{\mathrm{d}}{\mathrm{d}\varepsilon}\bigg(\langle M(\varphi + \varepsilon\gamma), \log(\varphi + \varepsilon\gamma)\rangle + \frac{1}{2}\langle M(\varphi + \varepsilon\gamma), \mathfrak{R}(M(\varphi + \varepsilon\gamma))\rangle \\ &\qquad - \langle M(\varphi + \varepsilon\gamma), \mathbb{1}\rangle\bigg)\bigg|_{\varepsilon=0} \\ &= \langle M\gamma, \log\varphi\rangle + \langle M\gamma, \mathbb{1}\rangle + \langle M\gamma, \mathfrak{R}(\nu)\rangle - \langle M\gamma, \mathbb{1}\rangle \\ &= \langle M\gamma, \log\varphi + \mathfrak{R}(\nu)\rangle.\end{aligned} \tag{7.7}$$

Since this holds for any bounded measurable function γ with $\mathrm{supp}(\gamma) \subset \{\xi : |\xi_0| \leq L \text{ or } \varphi(\xi) > \delta\}$, we obtain that

$$0 = \log \varphi(\xi) + \mathfrak{R}(\nu)(\xi), \qquad M\text{-almost surely},$$

first only on the set $\{\xi : |\xi_0| \leq L \text{ or } \varphi(\xi) > \delta\}$, and hence on

$$\bigcup_{L \in \mathbb{N}} \{\xi : |\xi_0| \leq L\} \cup \bigcup_{\delta > 0} \{\xi : \varphi(\xi) > \delta\} = \Gamma_T^{(1)} \cup \{\varphi > 0\},$$

which is equal to $\Gamma_T^{(1)}$ M-almost surely.

This implies the claim in (7.6). □

PROOF OF PROPOSITION 2.10, 2.. This is a consequence of Lemma 7.2. □

We have prove some additional properties of the minimiser of $I_\mu^{(T)}$, in the following lemma we list some bounds on its moments.

LEMMA 7.4 (Bounds on $\nu^{(T)}$). *Assume that $\nu^{(T)}$ is a minimiser of $I_\mu^{(T)}$ on $\mathcal{M}(\Gamma_T^{(1)})$:*

(i) *Under the assumption in* (1.2) *and if $T < 1/\mathrm{e}^2 H$,*

$$\int \nu^{(T)}(\mathrm{d}\xi)|\xi_0|^2 \leq \frac{\mathrm{e}^2}{2\pi(1 - \mathrm{e}^2 T H)}. \tag{7.8}$$

(ii) *Under the assumptions in* (1.2) *and* (1.3), *for any $T > 0$,*

$$\int \nu^{(T)}(\mathrm{d}\xi)|\xi_0| \leq \max\left\{\frac{1}{hT}\log(2TH\mathrm{e}^2), \frac{1}{2\pi HT}\right\}. \tag{7.9}$$

PROOF. Again, we write ν instead of $\nu^{(T)}$ and M instead of $M_\mu^{(T)}$ and $\mathfrak{R}$ instead of $\mathfrak{R}^{(T)}$.

We start with (i). From the EL-equations in (7.6), we see that $\int \nu(\mathrm{d}\xi)|\xi_0|^2 \leq \int M(\mathrm{d}\xi)|\xi_0|^2$. We assume only (1.2). Under the assumption (1.2), applying Lemma A.5 for $\alpha = 2$ and $b = 1$ finishes the proof of (7.8), recalling (7.3).

We continue with (ii). We abbreviate $D = \int \nu(\mathrm{d}\xi)|\xi_0| = \int M(\mathrm{d}\xi)\mathrm{e}^{-\mathfrak{R}(\nu)(\xi)}|\xi_0|$ (using (7.6)). If $D \leq \frac{1}{hT}\log(2TH\mathrm{e}^2)$, then we are done. If the converse is true, then $TH\mathrm{e}^2\mathrm{e}^{-ThD} < \frac{1}{2}$. We begin by noting that

$$\mathfrak{R}(\nu)(\xi) \geq Th|\xi_0|D$$

holds under the assumption in (1.3) and is derived using the same steps as in (7.5) (without the additional integration over $\nu(\mathrm{d}\xi)$). This already implies that

$$D \leq \int M(\mathrm{d}\xi)\mathrm{e}^{-Th|\xi_0|D}|\xi_0|.$$

This upper bound can be further estimated from above. Indeed, using the same arguments as in (A.11), we obtain that

$$\begin{aligned} D &\leq \frac{1}{2\pi HT}\sum_{n=1}^{\infty}\frac{1}{n}\left(TH\mathrm{e}^2\mathrm{e}^{-ThD}\right)^n \\ &= -\frac{1}{2\pi HT}\log\left(1 - TH\mathrm{e}^2\mathrm{e}^{-ThD}\right) \\ &\leq \frac{1}{2\pi HT}\frac{TH\mathrm{e}^2\mathrm{e}^{-ThD}}{1 - TH\mathrm{e}^2\mathrm{e}^{-ThD}} \leq \frac{1}{2\pi HT}, \end{aligned} \tag{7.10}$$

where we used that, due to $THe^2e^{-ThD} < \frac{1}{2} < 1$, we can apply the formula $\sum_{n=1}^{\infty}\frac{1}{n}q^n = -\log(1-q)$ that holds for $q \in [0,1)$. Then we used the estimate $-\log(1-x) \le \frac{x}{1-x}$, for $x < 1$ and after that we used the monotonicity of $x \mapsto \frac{x}{1-x}$. Hence, we proved (7.9). □

We see also from (7.8) that, for any minimiser $\nu^{(T)}$ of $I_\mu^{(T)}$, we have that $\nu^{(T)} \in \mathcal{A}_{f,\beta}$ for $f(r) = r^2$, all $T \in (0, 1/e^2H)$, and any sufficiently large β. Here is another benefit from (7.8).

LEMMA 7.5 (Uniqueness of solutions to EL-equations). *Assume that K satisfies* (1.2). *For any $T \in (0, \frac{1}{He^2}\frac{\pi}{1+\pi})$, there is at most one solution ν to* (7.6).

PROOF. Assume that ν and $\widetilde{\nu}$ are two solutions to the EL equation in (7.6). Using the estimate $|e^{-x} - e^{-y}| \le |x-y|\min\{e^{-x}, e^{-y}\} \le |x-y|(e^{-x} + e^{-y})$ for $x, y \in \mathbb{R}$, we obtain that

$$\begin{aligned}\int |\nu - \widetilde{\nu}|(d\xi)|\xi_0| &= \Bigl|\int M_\mu^{(T)}(d\xi)|\xi_0|\bigl(e^{-\mathfrak{R}^{(T)}(\nu)(\xi)} - e^{-\mathfrak{R}^{(T)}(\widetilde{\nu})(\xi)}\bigr)\Bigr| \\ &\le \int M_\mu^{(T)}(d\xi)|\xi_0|\bigl|\mathfrak{R}^{(T)}(\nu - \widetilde{\nu})(\xi)\bigr|\bigl(e^{-\mathfrak{R}^{(T)}(\nu)(\xi)} + e^{-\mathfrak{R}^{(T)}(\widetilde{\nu})(\xi)}\bigr).\end{aligned}$$

Now we use (1.2) to get

$$\bigl|\mathfrak{R}^{(T)}(\nu - \widetilde{\nu})(\xi)\bigr| \le \mathfrak{R}^{(T)}\bigl(|\nu - \widetilde{\nu}|\bigr)(\xi) \le HT\int |\nu - \widetilde{\nu}|(d\widetilde{\xi})|\widetilde{\xi}_0||\xi_0|.$$

From now on, we assume that $e^2TH < 1$. Then we can combine that last two estimates, use the EL equations again, afterward (7.8), and obtain

$$\begin{aligned}\int |\nu - \widetilde{\nu}|(d\xi)|\xi_0| &\le HT\int |\nu - \widetilde{\nu}|(d\widetilde{\xi})|\widetilde{\xi}_0|\Bigl(\int \nu(d\xi)|\xi_0|^2 + \int \widetilde{\nu}(d\xi)|\xi_0|^2\Bigr) \\ &\le 2HT\int |\nu - \widetilde{\nu}|(d\xi)|\xi_0|\frac{e^2}{2\pi(1-e^2TH)}.\end{aligned}$$

If T is so small that $e^2TH < 1$ and $\frac{e^2HT}{\pi(1-e^2TH)} < 1$, this implies that $\int |\nu - \widetilde{\nu}|(d\xi)|\xi_0| = 0$, which implies that $\nu = \widetilde{\nu}$. The condition on $x = HT$ reads $0 < x < 1/e^2$ and $x < \frac{\pi}{e^2(1+\pi)}$. Hence, the latter inequality holds for $TH \in (0, \frac{1}{e^2}\frac{\pi}{1+\pi})$. This implies the assertion. □

7.3. *Subcritical phase*: *Convergence and nongelation.* In this section, we provide the proofs of Theorem 2.8, 1 and Proposition 2.10, 1. Throughout the section, we fix $T > 0$ and $\mu \in \mathcal{M}_1(\mathcal{S})$. Note that Lemma 7.1 already covers Theorem 2.8, 1(i) about the compact sublevel sets of $I_\mu^{(T)}$ and the existence of minimisers, and Lemma 7.3 implies assertion (iii) about the validity of the Euler–Lagrange equations for minimisers.

The outline of this section is as follows. The tightness assertion about $\mathcal{V}_N^{(T)}$ under $\mathbb{P}_{\mathrm{Poi}_{N\mu}}^{(N)}$ in Theorem 2.8, 1(iv) is proved in Lemma 7.7, and the tightness of $c_{\mathcal{V}_{N,0}^{(T)}}$ as well as Theorem 2.8, 1(v) is proved in Corollary 7.8. Finally, the asssertion about nongelation in Theorem 2.8, 1(ii) is proved in Corollary 7.9. The proof of Proposition 2.10, 1 is finished at the end of this section.

Let us start by explaining the strategy for proving the tightness result. Usually, tightness is directly implied by the LDP, as we stated in Corollary 2.4. The problem is that, in our LDP from Theorem 2.3, we conditioned on $\{\mathcal{V}_N^{(T)} \in \mathcal{A}_{f,\beta}\}$. However, we want to prove tightness for the unconditioned distribution of $\mathcal{V}_N^{(T)}$. Note that we are free to choose $f(r) = r^2$. For that particular choice, we can argue that the probability of the event $\{\mathcal{V}_N^{(T)} \notin \mathcal{A}_{f,\beta}\}$ vanishes (see Lemma 7.6), for large β. Further, we can show that the minimisers of $I_\mu^{(T)}$ are also minimisers of the β-dependent rate function from (2.10), if β is large enough. Finally, we will use this to argue that the unconditioned distribution of $\mathcal{V}_N^{(T)}$ converges to a distribution that is concentrated on minimisers of $I_\mu^{(T)}$.

LEMMA 7.6. *Let $T > 0$ and $\mu \in \mathcal{M}_1(\mathcal{S})$ be such that $q_\mu^{(T)} < 1$. Then*

$$\sup_{N\in\mathbb{N}} \mathbb{P}^{(N)}_{\mathrm{Poi}_{N\mu}}\big(\mathcal{V}_N^{(T)} \notin \mathcal{A}_{f,\beta}\big) \le \frac{C}{\beta}, \quad \beta \in (0,\infty), \tag{7.11}$$

where $C = \int M_\mu^{(T)}(\mathrm{d}\xi)|\xi_0|^2 < \infty$.

PROOF. By Markov inequality it is enough to show that the expectation of $\int \mathcal{V}_{N,0}^{(T)}(\mathrm{d}k)|k|^2$ under $\mathbb{E}^{(N)}_{\mathrm{Poi}_{N\mu}}$ is bounded in N. Abbreviate $\mathtt{P}_N = \mathtt{P}_{NM_\mu^{(T)}}$. Applying Theorem 2.1 with kernel $\frac{K}{N}$, the density change from Corollary 6.1 and choosing f as the constant function $\nu \mapsto 1$, gives us

$$1 = \mathtt{E}_N\big[\mathrm{e}^{-\frac{1}{2N}\sum_{i,j\colon i\neq j} R^{(T)}(\Xi_i,\Xi_j)}\mathrm{e}^{-\frac{1}{N}\int \varphi_{\xi_0}(\xi)Y_N(\mathrm{d}\xi)}\big]\mathrm{e}^{N(|M_\mu^{(T)}|-1)}.$$

Notice that, since $q_\mu^{(T)} < 1$, we have that $|M_\mu^{(T)}| < \infty$. Then we apply Theorem 2.1 and Corollary 6.1 a second time to the function $\nu \mapsto \int \nu_0(\mathrm{d}k)|k|^2$ and combine the formulas to get

$$\begin{aligned}
&\mathbb{E}^{(N)}_{\mathrm{Poi}_{N\mu}}\Big[\int \mathcal{V}_{N,0}^{(T)}(\mathrm{d}k)|k|^2\Big]\\
&\quad= \frac{\mathtt{E}_N[\int \frac{1}{N}Y_{N,0}(\mathrm{d}k)|k|^2\mathrm{e}^{-\frac{1}{2N}\sum_{i,j\colon i\neq j} R^{(T)}(\Xi_i,\Xi_j)}\mathrm{e}^{-\frac{1}{N}\int \varphi_{\xi_0}(\xi)Y_N(\mathrm{d}\xi)}]}{\mathtt{E}_N[\mathrm{e}^{-\frac{1}{2N}\sum_{i,j\colon i\neq j} R^{(T)}(\Xi_i,\Xi_j)}\mathrm{e}^{-\frac{1}{N}\int \varphi_{\xi_0}(\xi)Y_N(\mathrm{d}\xi)}]}\\
&\quad= \frac{\mathtt{E}_N[f_N(Y_N)g_N(Y_N)]}{\mathtt{E}_N[g_N(Y_N)]},
\end{aligned}\tag{7.12}$$

where, for $\nu = \sum_i \delta_{\xi_i} \in \mathcal{M}_{\mathbb{N}_0}(\Gamma_T^{(1)})$, we defined

$$f_N(\nu) = \int \frac{1}{N}\nu_0(\mathrm{d}k)|k|^2, \qquad g_N(\nu) = \mathrm{e}^{-\frac{1}{2N}\sum_{i,j\colon i\neq j} R^{(T)}(\xi_i,\xi_j)}\mathrm{e}^{-\frac{1}{N}\int \varphi_{\xi_0}(\xi)\nu(\mathrm{d}\xi)}.$$

Observe that f_N and $-g_N$ are increasing on $\mathcal{M}_{\mathbb{N}_0}(\Gamma_T^{(1)})$ under the addition of points. Thus, we can apply the Harris–FKG inequality (see Theorem 20.4 in [32]) to bound the right-hand side of (7.12) by

$$\mathtt{E}_N\big[f_N(Y_N)\big] = \mathtt{E}_N\Big[\int \frac{1}{N}Y_{N,0}(\mathrm{d}k)|k|^2\Big] = \int_{\Gamma_T^{(1)}} M_\mu^{(T)}(\mathrm{d}\xi)|\xi_0|^2, \tag{7.13}$$

where we used Campbell's formula (see Proposition 2.7 in [32]). Note that the right-hand side is finite under the assumption $q_\mu^{(T)} < 1$ due to Lemma 7.1. Hence, we have a established a bound that is uniform in $N \in \mathbb{N}$. This finishes the proof. □

It is standard (see Corollary 2.4) that, given the LDP of Theorem 2.3, accumulation points of $(\mathcal{V}_N^{(T)})_{N\in\mathbb{N}}$ exist and are concentrated on the set of minimisers of the rate function. However, this holds a priori only under conditioning on $\mathcal{V}_N^{(T)} \in \mathcal{A}_{f,\beta}$. However, we now derive that under $q_\mu^{(T)} < 1$ this holds under the unconditioned measure $\mathbb{P}^{(N)}_{\mathrm{Poi}_{N\mu}}$ as well.

LEMMA 7.7 (Law of large numbers). *Fix $T > 0$, and assume that $q_\mu^{(T)} < 1$. Let $\mathbb{P}_{\mathcal{V}_N}$ denote the distribution of $\mathcal{V}_N^{(T)}$ under $\mathbb{P}^{(N)}_{\mathrm{Poi}_{N\mu}}$. Then the sequence of measures $(\mathbb{P}_{\mathcal{V}_N})_{N\in\mathbb{N}}$ is tight (and thus relatively compact), and each limit point $\mathbb{P}$ is concentrated on the set of minimisers of $I_\mu^{(T)}$, that is,*

$$\operatorname{supp}(\mathbb{P}) \subset D_0 := \{\nu \in \mathcal{M}(\Gamma_T^{(1)}) \colon I_\mu^{(T)}(\nu) = \inf I_\mu^{(T)}\}. \tag{7.14}$$

PROOF. Recall that we are working with $f(r) = r^2$. Abbreviate the distribution of $\mathcal{V}_N^{(T)}$ under $\mathbb{P}^{(N)}_{\mathrm{Poi}_{N\mu}}(\cdot|\mathcal{V}_N^{(T)} \in \mathcal{A}_{f,\beta})$ by $\mathbb{P}_{\mathcal{V}_N,\beta}$. According to Theorem 2.3, $(\mathbb{P}_{\mathcal{V}_N,\beta})_{N\in\mathbb{N}}$ satisfies an LDP on $\mathcal{A}_{f,\beta}$ with good rate function $I_{\beta,\mu}^{(T)}$ given by $I_{\beta,\mu}^{(T)}(\nu) = I_\mu^{(T)}(\nu) - \chi_\beta$ for $\nu \in \mathcal{A}_{f,\beta}$. According to Lemma 7.1, because of $q_\mu^{(T)} < 1$, $I_\mu^{(T)}$ possesses at least one minimiser on $\mathcal{M}(\Gamma_T^{(1)})$. For sufficiently large β, every minimiser ν of $I_\mu^{(T)}$ lies in $\mathcal{A}_{f,\beta}$, since it satisfies the EL-equation in (7.6) and satisfies, therefore, the estimate in (7.8) for the second moment, which does not depend on ν. Hence, $\{I_{\beta,\mu}^{(T)} = 0\} = D_0$ for all sufficiently large β.

As a consequence of Corollary 2.4, $(\mathbb{P}_{\mathcal{V}_N,\beta})_{N\in\mathbb{N}}$ is tight, and any accumulation point $\mathbb{P}_\beta$ is concentrated on $\{I_{\beta,\mu}^{(T)} = 0\}$, that is, on D_0. We now show that $(\mathbb{P}_{\mathcal{V}_N,\beta})_{N\in\mathbb{N}}$ and $(\mathbb{P}_{\mathcal{V}_N})_{N\in\mathbb{N}}$ have the same limiting behaviour, that is, they are tight and every accumulation point is concentrated on D_0. Indeed, for any open neighbourhood U of D_0, we have

$$\mathbb{P}_{\mathcal{V}_N}(U^{\mathrm{c}}) \le \mathbb{P}_{\mathcal{V}_N}(U^{\mathrm{c}} \cap \mathcal{A}_{f,\beta}) + \mathbb{P}_{\mathcal{V}_N}(\mathcal{A}_{f,\beta}^{\mathrm{c}}) \le \mathbb{P}_{\mathcal{V}_N,\beta}(U^{\mathrm{c}}) + \frac{C}{\beta},$$

according to Lemma 7.6, where C/β can be taken arbitrarily small, uniformly in N. Using the above, we see that $\mathbb{P}_{\mathcal{V}_N}(U^{\mathrm{c}})$ vanishes as $N \to \infty$. Hence, any accumulation point $\mathbb{P}$ of $(\mathbb{P}_{\mathcal{V}_N})_{N\in\mathbb{N}}$ is concentrated on U and hence on D_0. □

COROLLARY 7.8 (Law of large numbers of the total mass). *Fix $T > 0$, and assume that $q_\mu^{(T)} < 1$. Let $\mathbb{P}_{\mathcal{V}_N}$ denote the distribution of $\mathcal{V}_N^{(T)}$ under $\mathbb{P}^{(N)}_{\mathrm{Poi}_{N\mu}}$, and recall that $(\mathbb{P}_{\mathcal{V}_N})_{N\in\mathbb{N}}$ is tight by Lemma 7.7. Take a subsequence, also denoted $(\mathbb{P}_{\mathcal{V}_N})_{N\in\mathbb{N}}$, with limit point $\mathbb{P}$, and let $\mathcal{V}$ be a random variable with distribution $\mathbb{P}$. Recalling the definition of c_{ν_0} from (2.16) and that $|c_{\nu_0}| = c_{\nu_0}(\mathcal{S})$, for $\nu \in \mathcal{M}(\Gamma_T^{(1)})$, we have that*

$$|c_{\mathcal{V}_{N,0}^{(T)}}| \to |c_{\mathcal{V}_0}| \quad \textit{in distribution, as } N \to \infty \tag{7.15}$$

Further, we have that $|c_{\mathcal{V}_{N,0}^{(T)}}| \to 1$ in probability, as $N \to \infty$. Hence, $|c_{\mathcal{V}_0}| = 1$ $\mathbb{P}$-almost surely.

PROOF. We start with proving (7.15). In the proof of Lemma 7.7, we have seen that $\mathbb{P}$ is concentrated on $D_0 \subset \mathcal{A}_{f,\beta}$, for β large enough. Since the map $\nu \mapsto |c_{\nu_0}|$ is continuous on $\mathcal{A}_{f,\beta}$, the claim follows by Lemma 7.7 and the continuous mapping theorem.

It remains to argue the last statement. Recall that $Nc_{\mathcal{V}_{N,0}^{(T)}}$ is equal to the number of atoms $n(0)$ of the coagulation process, which is Poi_N-distributed under $\mathbb{P}^{(N)}_{\mathrm{Poi}_{N\mu}}$. By the law of large numbers, $c_{\mathcal{V}_{N,0}^{(T)}} \to 1$ in probability. We combine this with the result (7.15), which implies that $\mathbb{P}(||c_{\mathcal{V}_0}| - 1| > \varepsilon) \le \liminf_{N\to\infty} \mathbb{P}^{(N)}_{\mathrm{Poi}_{N\mu}}(||c_{(\mathcal{V}_N^{(T)})_0}| - 1| > \varepsilon) = 0$ for any $\varepsilon > 0$. Hence, $|c_{\mathcal{V}_0}| = 1$ $\mathbb{P}$-almost surely. □

Recall the definition of $\mathrm{NG}_T^{(\mu)}$ from (2.14), and observe that via (2.17) we have that

$$\mathrm{NG}_T^{(\mu)} = \lim_{L\to\infty} \limsup_{N\to\infty} \mathbb{E}^{(N)}_{\mathrm{Poi}_{N\mu}}[|c^{(\le L)}_{(\mathcal{V}_N^{(T)})_0}|].$$

COROLLARY 7.9 (No gelation if $q_\mu^{(T)} < 1$). *Fix $T \in (0,\infty)$ and $\mu \in \mathcal{M}_1(\mathcal{S})$, and assume that $q_\mu^{(T)} < 1$. Then* $\mathrm{NG}_T^{(\mu)} = 1$.

PROOF. Since $|c^{(\le L)}_{\mathcal{V}^{(T)}_{N,0}}| \le |c_{\mathcal{V}^{(T)}_{N,0}}|$ holds for any $L, N \in \mathbb{N}$, it is already clear that $\mathrm{NG}_T^{(\mu)} \le 1$. Now we argue that $\mathrm{NG}_T^{(\mu)} \ge 1$. Fix any limit point $\mathbb{P}$ of $(\mathbb{P}_{\mathcal{V}_N})_{N\in\mathbb{N}}$. Then

$$\lim_{L\to\infty}\liminf_{N\to\infty}\mathbb{E}_{\mathcal{V}_N}\big[|c^{(\le L)}_{\mathcal{V}_{N,0}}|\big] \ge \lim_{L\to\infty}\mathbb{E}_{\mathcal{V}}\big[|c^{(\le L)}_{\mathcal{V}_0}|\big] = \mathbb{E}_{\mathcal{V}}\big[|c_{\mathcal{V}_0}|\big] = 1,$$

where we used monotone convergence in L for the first equality, Corollary 7.8 for the second one. □

This finishes the proof of part 1 of Theorem 2.8.

PROOF OF PROPOSITION 2.10, 1. By Lemma A.5 we know that $TH < 1/\mathrm{e}^2$ implies $q_\mu^{(T)} < 1$, which gives us assertion (i). If $TH < \frac{1}{\mathrm{e}^2}\frac{\pi}{1+\pi}$, then the EL equation (7.6) has a unique solution $\nu^{(T)}$ according to Lemma 7.5, which gives the first assertion of (ii). Then, by Lemma 7.7, every limit point $\mathbb{P}$ of $\mathbb{P}^{(N)}_{\mathrm{Poi}_{N\mu}}(\mathcal{V}_N^{(T)} \in \cdot)$ is a probability measure concentrated on $\{\nu^{(T)}\}$; that is, the only possible limit is $\mathbb{P} = \delta_{\nu^{(T)}}$. Consequently, $\mathbb{P}^{(N)}_{\mathrm{Poi}_{N\mu}}(\mathcal{V}_N^{(T)} \in \cdot) \Longrightarrow \delta_{\nu^{(T)}}$ as $N \to \infty$. Now Lemma 2.5 and (1.19) imply the last statement of (ii). □

7.4. *Supercritical phase*: *Loss of mass and gelation*. In this section we assume that the upper bound on the kernel K in (1.2) holds as well as the lower bound (1.3) and show that gelation occurs under the assumption that $\kappa := \inf_{\nu\in\mathcal{M}(\Gamma_T^{(1)})} I_\mu^{(T)}(\nu) > 0$. That is, for the process $\mathcal{V}_N^{(T)}$ we can observe mass loss in the sense that $\mathrm{NG}_T^{(\mu)} < 1$; see (2.14), which is the assertion in Theorem 2.8, 2.

Fix $\mu \in \mathcal{M}_1(\mathcal{S})$. Pick any $b \in (0,1]$ such that the total mass of $M_{b\mu}^{(T)}$ (defined in (A.9)) is finite. As in the proof of the lower bound in Theorem 2.3, we make use of a cut-off argument for a fixed L and of the results from Section A.5. Therefore, we refer to the proof of Theorem 2.3 and to Section A.5 for notation and definitions. Recall the equality about the $L \to \infty$ limit from (6.14), and note that this also holds if the infimum is taken over $\nu \in \mathcal{M}(\Gamma_T^{(1)})$. This implies that L can be picked so large that $\kappa^{(\le L)} = \inf_{\nu\in\mathcal{M}(\Gamma^{(1)}_{T,\le L})}(H(\nu|M_{b\mu}^{(T,\le L)}) - \phi_b(\nu)) \ge \frac{\kappa}{2}$, where we recall the definition of ϕ_b from (6.5) and the fact that $H(\nu|M_{b\mu}^{(T)}) - \phi_b(\nu) = I_\mu^{(T)}(\nu)$.

Note that $|c^{(\le L)}_{\mathcal{V}^{(T,N)}_{N,0}}| + |c^{(>L)}_{\mathcal{V}^{(T)}_{N,0}}|$ (where the first term is defined after (2.16) and the second one is its obvious analogue) is equal to $\frac{1}{N}$ times a Poi_N-distributed variable under $\mathbb{E}^{(N)}_{\mathrm{Poi}_{N\mu}}$. Hence, for any $\varepsilon \in (0,1)$, on the event $\{|c^{(>L)}_{\mathcal{V}^{(T)}_{N,0}}| > \varepsilon\}$, the expectation of $|c^{(\le L)}_{\mathcal{V}^{(T)}_{N,0}}|$ is not larger than $1 - \varepsilon + o(1)$. Hence, it suffices to show that, for some $\varepsilon \in (0,1)$,

$$\limsup_{N\to\infty}\mathbb{E}^{(N)}_{\mathrm{Poi}_{N\mu}}\big[|c^{(\le L)}_{\mathcal{V}^{(T)}_N}|\mathbb{1}\{|c^{(>L)}_{\mathcal{V}^{(T)}_{N,0}}| \le \varepsilon\}\big] = 0. \tag{7.16}$$

Indeed, we will show that this expectation decays even exponentially fast. To show this, we apply Theorem 2.1 and Corollary 6.1 to obtain

$$\mathbb{E}^{(N)}_{\mathrm{Poi}_{N\mu}}\big[|c^{(\le L)}_{\mathcal{V}^{(T)}_N}|\mathbb{1}\{|c^{(>L)}_{\mathcal{V}^{(T)}_{N,0}}| \le \varepsilon\}\big]$$

$$= \mathrm{E}_{NM^{(T)}_{b\mu}}\big[\mathrm{e}^{-\frac{1}{2N}\sum_{i\neq j} R^{(T)}(\Xi_i,\Xi_j)} b^{-ND_N}\mathrm{e}^{(b-1)|Y_{N,0}|}$$
$$\times\, \mathrm{e}^{-\int \varphi_{\xi_0}(\xi)\frac{1}{N}Y_N(\mathrm{d}\xi)} D_N^{(\le L)} \mathbb{1}\{D_N^{(>L)} \le \varepsilon\}\big]\mathrm{e}^{N(|M^{(T)}_{b\mu}|-1)},$$

where we introduced the abbreviation $D_N^{(\le L)} = \frac{1}{N}\int Y_{N,0}(\mathrm{d}k)|k|\mathbb{1}\{|k| \le L\}$, respectively, defining $D_N^{(>L)}$ and $D_N = D_N^{(\le L)} + D_N^{(>L)}$.

We now decompose the PPP into $Y_N = Y_N^{(\le L)} + Y_N^{(>L)}$, where $Y_N^{(\le L)}$ and $Y_N^{(>L)}$ are the restrictions of Y_N to $\Gamma^{(1)}_{T,\le L}$ (defined in (6.10)) and to $\Gamma^{(1)}_T \setminus \Gamma^{(1)}_{T,\le L}$, respectively. Note that they are independent PPPs with intensity measures $NM^{(T,\le L)}_{b\mu}$ and $NM^{(T,>L)}_{b\mu}$, respectively.

We drop, in the first term in the exponent, the sum involving all Ξ_i's with $|\Xi_{i,0}| > L$, and obtain that the right-hand side of the last display is not larger than

$$\begin{aligned}&\mathrm{E}_{NM^{(T,\le L)}_{b\mu}}\big[\mathrm{e}^{N\phi_b(\frac{1}{N}Y_N^{(\le L)})}\mathrm{e}^{-\int \varphi_{\xi_0}(\xi)\frac{1}{N}Y_N^{(\le L)}(\mathrm{d}\xi)} D_N^{(\le L)}\big]\\ &\quad\times \mathrm{E}_{NM^{(T,>L)}_{b\mu}}\big[b^{-ND_N^{(>L)}}\mathrm{e}^{(b-1)|Y_{N,0}^{(>L)}|}\mathbb{1}\{D_N^{(>L)} \le \varepsilon\}\big],\end{aligned} \tag{7.17}$$

where ϕ_b is defined in (6.5). It is clear that the last line is $\le \mathrm{e}^{NO(\varepsilon)}$, since we can assume $b < 1$ without loss of generality. Define

$$\widehat{\mathbb{P}}_N^{(\le L)}(\mathrm{d}\nu) = \big(\widehat{Z}_N^{(\le L)}\big)^{-1}\mathrm{e}^{N\phi_b(\nu)}\mathrm{e}^{-\int \varphi_{\xi_0}(\xi)\nu^{(\le L)}(\mathrm{d}\xi)}\mathrm{P}_{NM^{(T,\le L)}_{b\mu}}\Big(\frac{1}{N}Y_N^{(\le L)} \in \mathrm{d}\nu\Big), \tag{7.18}$$

$$\widehat{Z}_N^{(\le L)} = \mathrm{E}_{NM^{(T,\le L)}_{b\mu}}\big[\mathrm{e}^{N\phi_b(\frac{1}{N}Y_N^{(\le L)})}\mathrm{e}^{-\int \varphi_{\xi_0}(\xi)\frac{1}{N}Y_N^{(\le L)}(\mathrm{d}\xi)}\big]. \tag{7.19}$$

Then (7.17) is not larger than $\mathrm{e}^{NO(\varepsilon)}\widehat{Z}_N^{(\le L)}\widehat{\mathbb{E}}_N^{(\le L)}(D_N^{(\le L)})$. Now, in the same as was in the LDP from Theorem A.7, we can derive the LDP for $\frac{1}{N}Y_N^{(\le L)}$ under $\widehat{\mathbb{P}}_N^{(\le L)}$, using Lemma A.4 and Varadhan's lemma). From this we first get that $\widehat{Z}_N^{(\le L)} \le \mathrm{e}^{-N(\kappa^{(\le L)}+o(1))} \le \mathrm{e}^{-N\kappa/3}$ for all sufficiently large N. Furthermore, analogously to Corollary A.8, the sequence of distributions of $\frac{1}{N}Y_N^{(\le L)}$ under $\widehat{\mathbb{P}}_N^{(\le L)}$ has accumulation points, and each one lies in the set of minimisers of $I_\mu^{(T,\le L)}$. Lemma A.9 says that all these accumulation points $\nu^{(T,\le L)}$ satisfy the Euler–Lagrange equation in (A.17), and Corollary A.10(ii) says that they all satisfy $\int \nu^{(T,\le L)}(\mathrm{d}k)|k| \le C_T$ for some constant C_T that depends only on T and the constants H from (1.2) and h from (1.3). $D^{(\le L)}$, being a continuous function of $\mathcal{V}_N^{(T)}$, has accumulations points too under $\widehat{\mathbb{P}}_N^{(\le L)}$, and their expectations are bounded by C_T as well.

Hence, we have shown that, for all sufficiently large N,

$$\begin{aligned}\mathbb{E}^{(N)}_{\mathrm{Poi}_{N\mu}}\big[|c^{(\le L)}_{\mathcal{V}_N^{(T)}}|\mathbb{1}\{|c^{(>L)}_{\mathcal{V}_{N,0}^{(T)}}| \le \varepsilon\}\big] &\le (7.17)\\ &\le \mathrm{e}^{NO(\varepsilon)}\widehat{Z}_N^{(\le L)}\widehat{\mathbb{E}}_N^{(\le L)}(D_N^{(\le L)})\\ &\le \mathrm{e}^{N[O(\varepsilon)-\kappa/3]}\big(C_T + o(1)\big).\end{aligned}$$

Now we pick $\varepsilon > 0$ so small that the exponent on the right-hand side is strictly negative. This implies, for any sufficiently large N, that

$$\begin{aligned}\mathbb{E}^{(N)}_{\mathrm{Poi}_{N\mu}}\big[|c^{(\le L)}_{\mathcal{V}_N^{(T)}}|\big] &\le 1-\varepsilon+o(1)+\mathbb{E}^{(N)}_{\mathrm{Poi}_{N\mu}}\big[|c^{(\le L)}_{\mathcal{V}_N^{(T)}}|\mathbb{1}\{|c^{(>L)}_{\mathcal{V}_{N,0}^{(T)}}| \le \varepsilon\}\big]\\ &= 1-\varepsilon+o(1).\end{aligned} \tag{7.20}$$

This implies that $\mathrm{NG}_T^{(\mu)} < 1$ and shows that gelation holds. This finishes the proof of Theorem 2.8, 2.

7.5. *The Smoluchowski equation.* Now we prove Lemma 2.12. We pick a limit point $\nu^{(T)}$ of $(\mathcal{V}_N^{(T)})_{N\in\mathbb{N}}$ under $\mathbb{P}^{(N)}_{\mathrm{Poi}_{N\mu}}$. Note that, according to Proposition 2.10, 1(ii) $\nu^{(T)}$ is uniquely determined as a solution to the Euler–Lagrange equation. We consider $\rho_t^{(T)}(\nu^{(T)}) = \int \nu^{(T)}(\mathrm{d}\xi)\xi_t$ for $t\in[0,T]$, where we wrote $\rho^{(T)}$ for the map ρ defined in (2.11). First, let us argue that $\rho_t^{(T)}(\nu^{(T)})$ does not depend on T, as long as $t\le T<\frac{1}{H}\frac{1}{\mathrm{e}^2}\frac{\pi}{\pi+1}$. Indeed, by (1.19) we have, for any $N\in\mathbb{N}$,

$$\rho_t^{(T_1)}(\mathcal{V}_N^{(T_1)}) = \frac{1}{N}\Xi_t = \rho_t^{(T)}(\mathcal{V}_N^{(T)}), \quad 0\le t\le T_1 < T < \frac{1}{H}\frac{1}{\mathrm{e}^2}\frac{\pi}{\pi+1}. \tag{7.21}$$

We want to pass to the limit $N\to\infty$ and use the continuities of the maps $\nu\mapsto\rho^{(T)}(\nu)$ and $\nu\mapsto\rho^{(T_1)}(\nu)$ on their respective domains (and the fact that also the marginal map $\nu\mapsto\rho_t^{(T)}(\nu)$ is continuous), we obtain that $\rho_t^{(T)}(\nu^{(T)}) = \rho_t^{(T_1)}(\nu^{(T_1)})$. To justify the continuity, recall that the EL-equation (2.19) for $\nu^{(T)}$ implies that $\int \nu^{(T)}(\mathrm{d}\xi)|\xi_0|^2 \le \int M_\mu^{(T)}(\mathrm{d}\xi)|\xi_0|^2$, which is finite by Lemma A.5 and our assumption $TH<\frac{1}{e^2}\frac{\pi}{1+\pi}$. Hence, $\nu^{(T)}\in\mathcal{A}_{f,\beta}$ for $f(r)=r^2$ and some $\beta\in(0,\infty)$. The EL-equations imply that $H(\nu^{(T)}|M_\mu^{(T)})$ is finite. Thus, Lemma 2.5 applies, and $\rho^{(T)}$ is continuous in $\nu^{(T)}$. Since $T_1\le T$, the latter statements are also true for $\nu^{(T_1)}$, the (unique) limit point of $\mathcal{V}_N^{(T_1)}$ under $\mathbb{P}^{(N)}_{\mathrm{Poi}_{N\mu}}$, and hence $\rho^{(T_1)}$ is also continuous in $\nu^{(T_1)}$. Now equation (7.21) implies that $\rho_t^{(T)}(\nu^{(T)}) = \rho_t^{(T_1)}(\nu^{(T_1)})$ for any $t\le T_1<T$. This shows that $\rho_t^{(T)}(\nu^{(T)}) = \int \nu^{(T)}(\mathrm{d}\xi)\xi_t$ does not depend on T, as long as $t\le T<\frac{1}{H}\frac{1}{\mathrm{e}^2}\frac{\pi}{\pi+1}$. Therefore, we write from now $\rho_t=\rho_t^{(T)}(\nu^{(T)})$ (in a small abuse of notation).

Our task is to show that, for any $m^*\in\mathbb{N}$ and any bounded continuous test function $g\colon\mathcal{S}\to\mathbb{R}$,

$$\begin{aligned}
&\frac{\mathrm{d}}{\mathrm{d}t}\int_{\mathcal{S}}\rho_t(\mathrm{d}x^*,m^*)g(x^*)\\
&\quad = -\int_{\mathcal{S}}\rho_t(\mathrm{d}x^*,m^*)K\rho_t(x^*,m^*)g(x^*)\\
&\qquad + \frac{1}{2}\sum_{\substack{m,m'\in\mathbb{N}:\\ m+m'=m^*}}\int_{\mathcal{S}}\int_{\mathcal{S}}\int_{\mathcal{S}}\rho_t(\mathrm{d}x,m)\rho_t(\mathrm{d}x',m')\mathbf{K}\big((x,m),(x',m'),\mathrm{d}x^*\big)g(x^*),\\
&\quad m^*\in\mathbb{N}.
\end{aligned}\tag{7.22}$$

The base of this is the fact that $\nu^{(T)}$ satisfies the EL-equation in (2.19). We need to rewrite that equation a bit. Recall that we introduced $\mathbb{Q}_k^{(T)}$ in (5.10) as

$$\mathbb{Q}_k^{(T)}(\mathrm{d}\xi) = \mathbb{P}_k\big(\Xi\in\mathrm{d}\xi, |\Xi_T|=1\big)\mathrm{e}^{\varphi_k(\xi)},$$

where we rephrased the event that Ξ lies in $\Gamma_T^{(1)}$ as the event $\{|\Xi_T|=1\}$ and recall that the density φ_k was defined in (5.11) as

$$\varphi_k(\xi) = \int_0^T \Phi(\xi_t)\,\mathrm{d}t, \quad \text{where } \Phi(\phi) = \frac{1}{2}\big[\langle\phi,K\phi\rangle - \langle\phi,K^{(\mathrm{diag})}\rangle\big], \phi\in\mathcal{M}_{\mathbb{N}_0}(\mathcal{S}\times\mathbb{N}).$$

From (2.8) and (2.2) and Fubini's theorem, we see that

$$\mathfrak{R}^{(T)}(\nu^{(T)})(\xi) = \int_0^T \mathrm{d}s\Big\langle\int_{\Gamma_T^{(1)}}\nu^{(T)}(\mathrm{d}\xi')\xi'_s, K\xi_s\Big\rangle = \int_0^T \mathrm{d}s\langle\rho_s,K\xi_s\rangle.$$

Hence, we derive from (2.19) that

$$\rho_t = \mathrm{e}\mathbb{E}_{\mathrm{Poi}_\mu}\big[\Xi_t\mathbb{1}\{|\Xi_T|=1\}\mathrm{e}^{\int_0^T[\Phi(\Xi_s)-\langle\rho_s,K\Xi_s\rangle]\,\mathrm{d}s}\big], \quad t\in[0,T]. \tag{7.23}$$

Recall that ρ_t does not depend on T, as long as $t \le T$. Hence, the right-hand side of (7.23) does not depend on T, and we may put T equal to t. We define the function

$$\mathcal{M}(\mathcal{S}\times\mathbb{N}) \ni \phi \mapsto f(\phi) = \int_{\mathcal{S}} \phi(\mathrm{d}x^*, m^*)\mathbb{1}\{|\phi| = 1\} g(x^*),$$

then

$$\int_{\mathcal{S}} \rho_t(\mathrm{d}x^*, m^*) g(x^*) = \mathrm{e}\mathbb{E}_{\mathrm{Poi}_\mu}\big[f(\Xi_t)\mathrm{e}^{\int_0^t [\Phi(\Xi_s) - \langle \rho_s, K\,\Xi_s\rangle]\,\mathrm{d}s}\big], \quad t \in [0, T]. \tag{7.24}$$

We are going to identify the t-derivative of both sides. The expectation on the right-hand side is with respect to a Markov chain in continuous time with only finitely many possible Markovian steps, together with an additional execution of a certain expectation (namely, the one with respect to Υ) at every elapsure of one of the holding times, and this does not depend on the time. Hence, the right-hand side is differentiable with respect to t, as follows from general theory of Markov chains in continuous time on a discrete space, plus the said execution of another expectation that does not depend on time. Furthermore, the derivate may be identified in terms of the generator G of the Marcus–Lushnikov process, using a kind of product differentiation rule, as

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}t}\int_{\mathcal{S}} \rho_t(\mathrm{d}x^*, m^*) g(x^*) &= \mathrm{e}\mathbb{E}_{\mathrm{Poi}_\mu}\big[\big((Gf)(\Xi_t) + f(\Xi_t)\big(\Phi(\Xi_t) - \langle \rho_t, K\,\Xi_t\rangle\big)\big) \\ &\qquad \times \mathrm{e}^{\int_0^t [\Phi(\Xi_s) - \langle \rho_s, K\,\Xi_s\rangle]\,\mathrm{d}s}\big].\end{aligned} \tag{7.25}$$

One easily checks that $f(\Xi_t)\Phi(\Xi_t) = 0$.

To derive the Smoluchowski equation, we will prove the following two equations, the first one dealing with the gain of particles of type $(\mathrm{d}x^*, m^*)$:

$$\begin{aligned}&\mathrm{e}\mathbb{E}_{\mathrm{Poi}_\mu}\big[(Gf)(\Xi_t)\mathrm{e}^{\int_0^t [\Phi(\Xi_s) - \langle \rho_s, K\,\Xi_s\rangle]\,\mathrm{d}s}\big] \\ &\quad = \frac{1}{2}\sum_{\substack{m,m':\\ m+m'=m^*}} \int_{\mathcal{S}}\int_{\mathcal{S}}\int_{\mathcal{S}} \rho_t(\mathrm{d}x, m)\rho_t(\mathrm{d}x', m')\mathbf{K}\big((x,m),(x',m'),\mathrm{d}x^*\big) g(x^*),\end{aligned} \tag{7.26}$$

and the second one dealing with the loss of particles of type $(\mathrm{d}x^*, m^*)$

$$\mathrm{e}\mathbb{E}_{\mathrm{Poi}_\mu}\big[f(\Xi_t)\langle \rho_t, K\,\Xi_t\rangle \mathrm{e}^{\int_0^t [\Phi(\Xi_s) - \langle \rho_s, K\,\Xi_s\rangle]\,\mathrm{d}s}\big] = \int_{\mathcal{S}} \rho_t(\mathrm{d}x^*, m^*) K\rho_t(x^*, m^*) g(x^*). \tag{7.27}$$

The second one is an immediate consequence of the fact that

$$f(\Xi_t)\langle \rho_t, K\,\Xi_t\rangle = \int_{\mathcal{S}} g(x^*)\,\Xi_t(\mathrm{d}x^*, m^*)\mathbb{1}\{|\Xi_t| = 1\}\langle \Xi_t, K\rho_t\rangle$$

Then, equation (7.27) follows from (7.24) by interchanging the integration of x^* and the expectation $\mathbb{E}_{\mathrm{Poi}_\mu}$.

It remains to show equation (7.26). From Section 1.2 we see that the generator of the Marcus–Lushnikov process may be written as

$$\begin{aligned}G(f)(\phi) &= \sum_{\{(x,m),(x',m')\}} \int_{\mathcal{S}} \mathbf{K}_\phi\big((x,m),(x',m'),\mathrm{d}z\big) \\ &\qquad \times \big[f(\phi - \delta_{(x,m)} - \delta_{(x',m')} + \delta_{(z,m+m')}) - f(\phi)\big], \quad \phi \in \mathcal{M}_{\mathbb{N}_0}(\mathcal{S}\times\mathbb{N}),\end{aligned} \tag{7.28}$$

where we sum over the possible (unordered) pairs $(x,m), (x',m') \in \mathrm{supp}(\phi)$. Observe that $\mathbf{K}_\phi f(\phi) = 0$, since $\mathbf{K}_\phi = 0$, if $|\phi| = 1$, and $f(\phi) = 0$, if $|\phi| \neq 1$, and hence

$$(Gf)(\phi) = \sum_{\{(x,m),(x',m')\}} \int_{\mathcal{S}} \mathbf{K}_\phi\big((x,m),(x',m'),\mathrm{d}x^*\big) f(\phi - \delta_{(x,m)} - \delta_{(x',m')} + \delta_{(x^*,m+m')}).$$

In fact, $(Gf)(\phi)$ is only nontrivial if $\phi = \delta_{(x,m)} + \delta_{(x',m')}$ for some $x, x' \in \mathcal{S}$ and $m, m' \in \mathbb{N}$ with $m + m' = m^*$, and in that case

$$(7.29)\quad (Gf)(\phi) = \int_{\mathcal{S}} \mathbf{K}_\phi\big((x,m),(x',m'),\mathrm{d}x^*\big)g(x^*) = \int_{\mathcal{S}} \mathbf{K}\big((x,m),(x',m'),\mathrm{d}x^*\big)g(x^*),$$

where the second equality can be checked by distinguishing the two cases in the definition (1.7) of $\mathbf{K}_\phi$.

For now, we only spell out the condition induced by the fixed m^*, which gives that

$$\begin{aligned}(7.30)\quad & \mathrm{e}\mathbb{E}_{\mathrm{Poi}_\mu}\big[(Gf)(\Xi_t)\mathrm{e}^{\int_0^t[\Phi(\Xi_s)-\langle\rho_s,K\Xi_s\rangle]\,\mathrm{d}s}\big] \\ &= \frac{1}{m^*!}\int \mu^{\otimes m^*}(\mathrm{d}\mathbf{x})\mathbb{E}_{\mathbf{x}}\big[(Gf)(\Xi_t)\mathrm{e}^{\int_0^t[\Phi(\Xi_s)-\langle\rho_s,K\Xi_s\rangle]\,\mathrm{d}s}\big] \\ &= \frac{1}{m^*!}\int \mu^{\otimes m^*}(\mathrm{d}\mathbf{x}) \\ &\quad \times \frac{1}{2}\sum_{\substack{A,B:\ A\dot\cup B=[m^*]:\\ A,B\neq\varnothing}} \mathbb{E}_{\mathbf{x}}\big[\mathbb{1}\{A \nleftrightarrow B\}(Gf)(\Xi_t)\mathrm{e}^{\int_0^t[\Phi(\Xi_s)-\langle\rho_s,K\Xi_s\rangle]\,\mathrm{d}s}\big],\end{aligned}$$

where the factor $1/2$ cancels out the double counting in the double sum over A, B. Recall that Ξ is a function of the labelled coagulation process Z (recalling (1.12)). Further, recall that the event $\{A \nleftrightarrow B\}$ means that during $[0,t]$ no coagulations take place between any particle pair C, C' where $C \subset A$ and $C' \subset B$. Under this event we can decompose the process as $\Xi = \Xi^{(A)} + \Xi^{(B)}$, where we used the short-hand notation $\Xi^{(A)} = \Xi^{(t,A)}$, which is the empirical measure belonging to the subprocess of Z that only deals with particles C with $C \subset A$ (and analogously for B). By basic calculations one gets that

$$(7.31)\quad \int_0^t \Phi\big(\Xi_s^{(A)} + \Xi_s^{(B)}\big)\,\mathrm{d}s = \int_0^t \Phi\big(\Xi_s^{(A)}\big)\,\mathrm{d}s + \int_0^t \Phi\big(\Xi_s^{(B)}\big)\,\mathrm{d}s + R^{(t)}\big(\Xi^{(A)}, \Xi^{(B)}\big).$$

For a fixed pair of nonempty sets A, B with $A\dot\cup B = [m^*]$, we abbreviate $m = |A|$ and $m' = |B|$. According to (7.29), we know that $(Gf)(\Xi_t^{(A)} + \Xi_t^{(B)})$ is only nontrivial if $|\Xi_t^{(A)}| = 1$ and $|\Xi_t^{(B)}| = 1$, and in that case we can also (somewhat artificially) write

$$\begin{aligned}(7.32)\quad & (Gf)\big(\Xi_t^{(A)} + \Xi_t^{(B)}\big) \\ &= \int_{\mathcal{S}}\int_{\mathcal{S}}\int_{\mathcal{S}} \Xi_t^{(A)}(\mathrm{d}x,m)\,\Xi_t^{(B)}(\mathrm{d}x',m')\mathbf{K}\big((x,m),(x',m'),\mathrm{d}x^*\big)g(x^*).\end{aligned}$$

Now we get the following by first inserting equation (7.31), then applying Lemma 4.2, and finally using (7.32):

$$\begin{aligned}& \mathbb{E}_{\mathbf{x}}\big[\mathbb{1}\{A \nleftrightarrow B\}(Gf)(\Xi_t)\mathrm{e}^{\int_0^t[\Phi(\Xi_s)-\langle\rho_s,K\Xi_s\rangle]\,\mathrm{d}s}\big] \\ &= \mathbb{E}_{\mathbf{x}}\big[\mathbb{1}\{A \nleftrightarrow B\}(Gf)\big(\Xi_t^{(A)} + \Xi_t^{(B)}\big)\mathrm{e}^{R^{(t)}(\Xi^{(A)},\Xi^{(B)})} \\ &\quad \times \mathrm{e}^{\int_0^t[\Phi(\Xi_s^{(A)})-\langle\rho_s,K\Xi_s^{(A)}\rangle]\,\mathrm{d}s}\mathrm{e}^{\int_0^t[\Phi(\Xi_s^{(B)})-\langle\rho_s,K\Xi_s^{(B)}\rangle]\,\mathrm{d}s}\big] \\ &= \mathbb{E}_{\mathbf{x}^{(A)}} \otimes \mathbb{E}_{\mathbf{x}^{(B)}}\big[\mathbb{1}\{|\Xi_t^{(A)}|, |\Xi_t^{(B)}| = 1\}(Gf)\big(\Xi_t^{(A)} + \Xi_t^{(B)}\big) \\ &\quad \times \mathrm{e}^{\int_0^t[\Phi(\Xi_s^{(A)})-\langle\rho_s,K\Xi_s^{(A)}\rangle]\,\mathrm{d}s}\mathrm{e}^{\int_0^t[\Phi(\Xi_s^{(B)})-\langle\rho_s,K\Xi_s^{(B)}\rangle]\,\mathrm{d}s}\big] \\ &= \int_{\mathcal{S}}\int_{\mathcal{S}}\int_{\mathcal{S}} \mathbf{K}\big((x,m),(x',m'),\mathrm{d}x^*\big)g(x^*) \\ &\quad \times \mathbb{E}_{\mathbf{x}^{(A)}}\big[\Xi_t^{(A)}(\mathrm{d}x,m)\mathbb{1}\{|\Xi_t^{(A)}| = 1\}\mathrm{e}^{\int_0^t[\Phi(\Xi_s^{(A)})-\langle\rho_s,K\Xi_s^{(A)}\rangle]\,\mathrm{d}s}\big]\end{aligned}$$

$$\times\, \mathbb{E}_{\mathbf{x}^{(B)}}\big[\Xi_t^{(B)}(\mathrm{d}x',m')\mathbb{1}\{|\Xi_t^{(B)}|=1\}\mathrm{e}^{\int_0^t[\Phi(\Xi_s^{(B)})-\langle\rho_s,K\,\Xi_s^{(B)}\rangle]\,\mathrm{d}s}\big],$$

where we have written $\mathbf{x}^{(A)}=(x_i)_{i\in A}$, $\mathbf{x}^{(B)}=(x_i)_{i\in B}$. Now, still for fixed sets A, B (with $m=|A|, m'=|B|$), we can split the term $\int\mu^{\otimes m^*}(\mathrm{d}\mathbf{x})$ from our previous representation (7.30) into $\int\mu^{\otimes m}(\mathrm{d}\mathbf{x}^{(A)})\int\mu^{\otimes m'}(\mathrm{d}\mathbf{x}^{(B)})$. The last thing we need is that

$$\begin{aligned}&\frac{1}{m!}\int\mu^{\otimes m}(\mathrm{d}(\mathbf{x}^{(A)}))\mathbb{E}_{\mathbf{x}^{(A)}}\big[\Xi_t(\mathrm{d}x,m)\mathbb{1}\{|\Xi_t^{(A)}|=1\}\mathrm{e}^{\int_0^t[\Phi(\Xi_s^{(A)})-\langle\rho_s,K\,\Xi_s^{(A)}\rangle]\,\mathrm{d}s}\big]\\&\quad=\mathrm{e}\mathbb{E}_{\mathrm{Poi}_\mu}\big[\Xi_t(\mathrm{d}x,m)\mathbb{1}\{|\Xi_t^{(A)}|=1\}\mathrm{e}^{\int_0^t[\Phi(\Xi_s)-\langle\rho_s,K\,\Xi_s\rangle]\,\mathrm{d}s}\big]=\rho_t(\mathrm{d}x,m),\end{aligned}$$

where the right-hand side does not depend on A anymore, but only on its cardinality m. The same holds for the terms derived from the set B. Hence, in (7.30) we can interchange the summation over nonempty sets A, B with $A\dot\cup B=[m^*]$ by a summation over $m,m'\in\mathbb{N}$ with $m+m'=m^*$, adding a correction term $\frac{m^*!}{m!m'!}$. This finally gives us (7.26).

APPENDIX: MISCELLANEOUS TECHNICALITIES

A.1. Topologies and metrics. We need to make some comments on the topologies used. We need to do this on various levels. Recall that we assumed $\mathcal{S}$ to be a Polish space. Let d be a metric on $\mathcal{S}\times\mathbb{N}$ such that $(\mathcal{S}\times\mathbb{N},d)$ is a complete space.

On the space $\mathcal{M}(\mathcal{S}\times\mathbb{N})$, we consider the weak topology, that is, $\phi^{(n)}\to\phi$ as $n\to\infty$ if $\langle\phi^{(n)},f\rangle\to\langle\phi,f\rangle$ as $n\to\infty$ for any continuous and bounded test functional $f\in C_{\mathrm{b}}(\mathcal{S}\times\mathbb{N})$. Note that $\mathcal{M}(\mathcal{S}\times\mathbb{N})$ is Polish. Weak convergence on $\mathcal{M}(\mathcal{S}\times\mathbb{N})$ is induced by the Lévy–Prohorov metric $\mathfrak{d}$, which is defined as follows; see [14]. For all $\phi,\phi'\in\mathcal{M}(\mathcal{S}\times\mathbb{N})$,

$$\begin{aligned}\mathfrak{d}(\phi,\phi')=\inf\{&\varepsilon>0\colon\phi(A)\le\phi'(A^\varepsilon)+\varepsilon,\phi'(A)\le\phi(A^\varepsilon)+\varepsilon,\\&\text{for any mb. }A\subset\mathcal{S}\times\mathbb{N}\},\end{aligned}\tag{A.1}$$

where $A^\varepsilon=\{x\in\mathcal{S}\times\mathbb{N}\colon d(x,y)<\varepsilon\text{ for some }y\in A\}$ denotes the open ε-neighbourhood of $A\subset\mathcal{S}\times\mathbb{N}$.

Eventually, we want to consider paths of measures. Denote by $\mathcal{M}$ either $\mathcal{M}(\mathcal{S}\times\mathbb{N})$ or $\mathcal{M}_{\mathbb{N}_0}(\mathcal{S}\times\mathbb{N})$ (the space of point measure on $\mathcal{S}\times\mathbb{N}$) and equip it with the metric $\mathfrak{d}$. Then $\mathbb{D}_T=\mathbb{D}_T(\mathcal{M})$ denotes the space of càdlàg functions $[0,T]\to\mathcal{M}$. We endow $\mathbb{D}_T$ with the Skorohod J_1-topology, which is induced by a certain metric d_T such that the space $(\mathbb{D}_T,\mathrm{d}_T)$ is separable and complete (see [30], Theorem. A2.2); that is, it is a Polish space. Convergence in $(\mathbb{D}_T,\mathrm{d}_T)$ can be characterised via time-changes on $[0,T]$, which are strictly increasing bijections $\lambda\colon[0,T]\to[0,T]$ (which are necessarily continuous with $\lambda(0)=0$ and $\lambda(T)=T$) as follows. For $\xi,\xi^{(1)},\xi^{(2)},\dots\in\mathbb{D}_T$, it holds that $\xi^{(n)}\to\xi$ as $n\to\infty$ if and only if

$$\sup_{t\in[0,T]}|\lambda_n(t)-t|+\sup_{t\in[0,T]}\mathfrak{d}(\xi^{(n)}_{\lambda_n(t)},\xi_t)\to0\quad\text{as }n\to\infty\tag{A.2}$$

for some time-changes λ_n on $[0,T]$.

Recall that the set of history trees is

$$\begin{aligned}\Gamma_T^{(1)}=\{\xi=(\xi_t)_{t\in[0,T]}\in\mathbb{D}_T(\mathcal{M}_{\mathbb{N}_0}(\mathcal{S}\times\mathbb{N}))\colon&\ \xi_0\text{ is concentrated on }\mathcal{S}\times\{1\},\\ t\mapsto\xi_t\text{ is piecewise constant and makes steps as in (1.6) and }&\xi_T(\mathcal{S}\times\mathbb{N})=1\}.\end{aligned}$$

Since $\mathbb{D}_T(\mathcal{M}_{\mathbb{N}_0}(\mathcal{S}\times\mathbb{N}))$ is separable, the same is true for $(\Gamma_T^{(1)},\mathrm{d}_T)$. In Lemma A.1 we will show that $(\Gamma_T^{(1)},\mathrm{d}_T)$ is also closed and hence complete as a closed subset of the complete space $\mathbb{D}_T(\mathcal{M}_{\mathbb{N}_0}(\mathcal{S}\times\mathbb{N}))$. With other words, it is itself a Polish space.

Recall that the state space of our process $\mathcal{V}_N^{(T)}$ is equal to the set $\mathcal{M}(\Gamma_T^{(1)})$ of positive measures on $\Gamma_T^{(1)}$. We equip $\mathcal{M}(\Gamma_T^{(1)})$ with the weak topology, that is, $\nu^{(n)} \to \nu$ as $n \to \infty$ if $\langle \nu^{(n)}, f\rangle \to \langle \nu, f\rangle$ as $n \to \infty$ for any continuous and bounded test functional $f \in C_{\mathrm{b}}(\Gamma_T^{(1)})$.

Now we show that the history tree set $\Gamma_T^{(1)}$ defined in (1.9) and (1.15) is closed.

LEMMA A.1. *The set* $\Gamma_T^{(1)}$ *is a closed subset of* $\mathbb{D}_T(\mathcal{M}_{\mathbb{N}_0}(\mathcal{S} \times \mathbb{N}))$. *Consequently,* $(\Gamma_T^{(1)}, \mathrm{d}_T)$ *is a Polish space.*

PROOF. We abbreviate $\mathbb{D}_T$ for $\mathbb{D}_T(\mathcal{M}_{\mathbb{N}_0}(\mathcal{S} \times \mathbb{N}))$. Let $(\xi^{(n)})_{n\in\mathbb{N}}$ be a sequence in $\Gamma_T^{(1)}$ that converges to $\xi \in \mathbb{D}_T$ as $n \to \infty$. We will argue that $\xi \in \Gamma_T^{(1)}$. There exist time changes λ_n, $n \in \mathbb{N}$, such that (A.2) holds. First, we need to show that ξ_0 is concentrated on $\mathcal{S} \times \{1\}$ and that $\xi_T(\mathcal{S} \times \mathbb{N}) = 1$. But this is easy. Indeed, since $\lambda_n(0) = 0$ and $\lambda_n(T) = T$, we have that $\xi_0 = \lim_{n\to\infty} \xi_0^{(n)}$ is concentrated on $\mathcal{S} \times \{1\}$ and $\xi_T(\mathcal{S} \times \mathbb{N}) = \lim_{n\to\infty} \xi_T^{(n)}(\mathcal{S} \times \mathbb{N}) = 1$.

Now we need to show that ξ is piecewise constant and makes only steps as in (1.6). For this we need a uniform lower bound for the holding times of $\xi^{(n)}$, $n \in \mathbb{N}$. Otherwise, one might have a succession of jumps that happen increasingly fast and add up in the limit to a larger jump that is no longer of the form (1.6). For any $\tilde{\xi} \in \mathbb{D}_T$, let $\tilde{t}_1 < \tilde{t}_2 < \cdots$ be the discontinuity points of $\tilde{\xi}$, and define

$$s_*(\tilde{\xi}) = \inf_{j\geq 1} (\tilde{t}_{j+1} - \tilde{t}_j) \wedge T.$$

If $\tilde{\xi}$ is piecewise constant (which is the case for $\tilde{\xi} \in \Gamma_T^{(1)}$), then $s_*(\tilde{\xi})$ is the infimum over all holding times of $\tilde{\xi}$. We also define the modulus of continuity for $\tilde{\xi} \in \mathbb{D}_T$ with spacing $h > 0$ as

$$w(\tilde{\xi}, h) = \inf_{(I_i)_i} \max_i \sup_{u,v\in I_i} \eth(\tilde{\xi}_u, \tilde{\xi}_v), \tag{A.3}$$

where the infimum extends over all partitions of the interval $[0, T]$ into subintervals $I_i = [s_i, t_i)$ such that $|I_i| := t_i - s_i > h$ for all i. Note that if $\tilde{\xi}$ is piecewise constant, then $w(\tilde{\xi}, h) = 0$ for all $h < s_*(\tilde{\xi})$.

By Theorem A2.2 in [30], the convergence of $\xi^{(n)}$ toward ξ implies that

$$\lim_{h\to 0} \sup_{n\in\mathbb{N}} w(\xi^{(n)}, h) = 0.$$

We want to show that $\inf_{n\in\mathbb{N}} s_*(\xi^{(n)}) > 0$. We argue via contraposition, that is, we show that $\inf_{n\in\mathbb{N}} s_*(\xi^{(n)}) = 0$ implies that $\lim_{h\to 0} \sup_{n\in\mathbb{N}} w(\xi^{(n)}, h) > 0$. Fix any $h > 0$, and let n be such that $s_*(\xi^{(n)}) < h$. Then any partition $(I_i)_i$ of $[0, T]$ satisfying $|I_i| > h$ for all i contains an interval I_j that contains a discontinuity point t_n of $\xi^{(n)}$, and thus

$$\sup_{u,v\in I_j} \eth(\xi_u^{(n)}, \xi_v^{(n)}) \geq \xi_{t_n-}^{(n)}(\mathcal{S} \times \mathbb{N}) - \xi_{t_n}^{(n)}(\mathcal{S} \times \mathbb{N}) = 1.$$

This implies that $\lim_{h\to 0} \sup_{n\in\mathbb{N}} w(\xi^{(n)}, h) \geq 1 > 0$. Hence, we have shown that

$$\inf_{n\in\mathbb{N}} s_*(\xi^{(n)}) > 0.$$

Now we argue that ξ is piecewise constant. Fix $h > 0$ with $h < \inf_{n\in\mathbb{N}} s_*(\xi^{(n)})$, and note that $\sup_n w(\xi^{(n)}, h) = 0$. Using the characterization of convergence from (A.2), one can show that for any $\varepsilon \in (0, h)$ one has that $w(\xi, h) \leq w(\xi^{(n)}, h - \varepsilon) + \varepsilon = \varepsilon$, if n is large enough. Hence, $w(\xi, h) = 0$, which implies that ξ is piecewise constant.

Now we show that ξ makes steps as in (1.6). Fix a discontinuity point t of ξ. Take $\varepsilon > 0$ small enough such that $3\varepsilon < \inf_{n\in\mathbb{N}} s_*(\xi^{(n)}) \wedge t$. We have that $\xi_t \neq \xi_{t-\varepsilon}$ and

$$\xi_t - \xi_{t-\varepsilon} = \lim_{n\to\infty} \big(\xi^{(n)}_{\lambda_n(t)} - \xi^{(n)}_{\lambda_n(t-\varepsilon)}\big).$$

We choose n large enough such that $\mathfrak{d}(\xi^{(n)}_{\lambda_n(t-\varepsilon)}, \xi^{(n)}_{\lambda_n(t)}) > 0$ and $\sup_u |\lambda_n(u) - u| \le \varepsilon$. Then $\mathfrak{d}(\xi^{(n)}_{\lambda_n(t-\varepsilon)}, \xi^{(n)}_{\lambda_n(t)}) > 0$ implies that the interval $[\lambda_n(t-\varepsilon), \lambda_n(t))$ contains at least one discontinuity point of $\xi^{(n)}$. On the other hand, we have that $\lambda_n(t) - \lambda_n(t-\varepsilon) \le 3\varepsilon < \inf_{n\in\mathbb{N}} s_*(\xi^{(n)})$ and this implies that the interval $[\lambda_n(t-\varepsilon), \lambda_n(t))$ contains at most one discontinuity point of $\xi^{(n)}$. This gives us that

$$\xi_t - \xi_{t-\varepsilon} = \lim_{n\to\infty} \big(\xi^{(n)}_{\lambda_n(t)} - \xi^{(n)}_{\lambda_n(t-\varepsilon)}\big) = \lim_{n\to\infty} \big(-\delta_{(x_n,m_n)} - \delta_{(x'_n,m'_n)} + \delta_{(z_n,m_n+m'_n)}\big)$$

for some $(x_n, m_n), (x'_n, m'_n) \in \mathcal{S} \times \mathbb{N}$ and $z_n \in \mathcal{S}$. It is not hard to argue that the left-hand side does not depend on ε and that the convergence of the measures on the right-hand side implies the convergence of the atoms, and hence the right-hand side is equal to $-\delta_{(x,m)} - \delta_{(x',m')} + \delta_{(z,m+m')}$, which finishes the proof. □

A.2. The rates of noncoagulation: Properties. In this section we study properties of the operator $\mathfrak{R}^{(T)}$ defined in (2.8) with kernel equal to the noncoagulation probability, $R^{(T)}$, given in (2.2).

LEMMA A.2 (Properties of $\mathfrak{R}^{(T)}$). *Assume that $K\colon (\mathcal{S} \times \mathbb{N})^2 \to [0, \infty)$ is continuous and symmetric, and fix $T \in (0, \infty)$. Then the following hold:*

(i) *The mapping $(\xi, \xi') \mapsto R^{(T)}(\xi, \xi')$, is continuous on $\Gamma^{(1)}_T \times \Gamma^{(1)}_T$.*

(ii) *The map $\nu \mapsto \langle \nu, \mathfrak{R}^{(T)}(\nu)\rangle$ is lower semicontinuous with respect to the weak topology.*

(iii) *Assume that K satisfies (1.2), and fix a majorizing function f satisfying $f(r)/r \to \infty$ as $r \to \infty$ and $f(r) \ge r$ for any $r \in \mathbb{N}$. Then $\nu \mapsto \langle \nu, \mathfrak{R}^{(T)}(\nu)\rangle$ is bounded and continuous on $\mathcal{A}_{f,\beta}$ for any $\beta \in (0, \infty)$.*

PROOF. We start by showing (i), Recall the Skorohod J_1-topology introduced at the beginning of Section A.1. Fix $\xi' \in \Gamma^{(1)}_T$, and let $\xi^{(n)}, \xi \in \Gamma^{(1)}_T$ be such that $\xi^{(n)} \to \xi$, and let λ_n be time-changes such that (A.2) holds. We can assume that $|\xi^{(n)}_0| = |\xi_0|$ for all n. For $i = 1, \dots, |\xi'_0| - 1$, let ϕ'_i be the value of the path ξ' on $[t_{i-1}, t_i)$. Continuity of K implies that the mappings $(x, m) \mapsto K\phi'_i(x, m)$ are continuous. They are also bounded on $\mathcal{S} \times \{1, \dots, |\xi_0|\}$ and hence bounded on the support of $\bigcup_{n,t} \xi^{(n)}_{\lambda_n(t)}$. Hence, $\sup_{t\in[0,T]} \mathfrak{d}(\xi^{(n)}_{\lambda_n(t)}, \xi_t) \to 0$, as $n \to \infty$, implies that

$$\lim_{n\to\infty} R^{(T)}(\xi^{(n)}_{\lambda_n}, \xi') = \lim_{n\to\infty} \sum_{i=1}^{|\xi'_0|-1} \int_{t_{i-1}}^{t_i} \langle \xi^{(n)}_{\lambda_n(t)}, K\phi'_i\rangle \,\mathrm{d}t = \sum_{i=1}^{|\xi'_0|-1} \int_{t_{i-1}}^{t_i} \langle \xi_t, K\phi'_i\rangle \,\mathrm{d}t = R^{(T)}(\xi, \xi').$$

(The fact that we can control the continuity of the mappings $K\xi'_t$ uniformly for $t \in [0, T]$ is implied more generally by the fact that $\lim_{h\to 0} w(\xi', h) = 0$, recalling (A.3)). Also, using that the jumps are of the form (1.6), one has that

$$\begin{aligned} &\big|R^{(T)}(\xi^{(n)}, \xi') - R^{(T)}(\xi^{(n)}_{\lambda_n}, \xi')\big| \\ &\quad \le 3T \sup_{i,x,m\colon m\le|\xi_0|} K\phi'_i(x, m) \sup_{t\in[0,T]} \big|\lambda_n(t) - t\big| \to 0 \quad \text{as } n \to \infty. \end{aligned}$$

Altogether, we proved that $R^{(T)}(\xi^{(n)}, \xi') \to R^{(T)}(\xi, \xi')$, as $n \to \infty$. Continuity of the mapping $(\xi, \xi') \mapsto R^{(T)}(\xi, \xi')$ is then implied by symmetry.

We continue with a proof for (ii). If $\nu_n \to \nu$ as $n \to \infty$ weakly, then also $\nu_n \otimes \nu_n \to \nu \otimes \nu$. Using the continuity of $R^{(T)}$ that we established in point (i), Lemma 4.1, and Fatou's lemma, one easily sees that then $\liminf_{n\to\infty}\langle \nu_n, \mathfrak{R}^{(T)}(\nu_n)\rangle \geq \langle \nu, \mathfrak{R}^{(T)}(\nu)\rangle$, which shows lower semicontinuity.

Let us now show (iii). We show only the upper semicontinuity. Because of (1.2), we have for any $\xi, \xi' \in \Gamma_T^{(1)}$ the upper bound

$$\begin{aligned} R^{(T)}(\xi,\xi') &= \int_0^T \mathrm{d}t \langle \xi_t, K\xi_t'\rangle \\ &\leq H\int_0^T \mathrm{d}t \|\xi_t\|_1 \|\xi_t'\|_1 = HT\|\xi_0\|_1\|\xi_0'\|_1 = HT|\xi_0||\xi_0'|, \end{aligned} \tag{A.4}$$

where we used that $t \mapsto \|\xi_t\|_1 = \sum_{x,m} \xi_t(x,m)m$ is constant for any $\xi \in \Gamma_T^{(1)}$. For $t = 0$, $\|\xi_0\|_1 = |\xi_0|$ is the total mass of ξ_0.

Hence, we may split, for any $L > 0$,

$$\begin{aligned} \langle \nu, \mathfrak{R}^{(T)}(\nu)\rangle &= \int_{|\xi_0|\leq L} \nu(\mathrm{d}\xi) \int_{|\xi_0'|\leq L} \nu(\mathrm{d}\xi') R^{(T)}(\xi,\xi') \\ &\quad + \int \nu(\mathrm{d}\xi) \int \nu(\mathrm{d}\xi') R^{(T)}(\xi,\xi') \mathbb{1}\{|\xi_0| > L \text{ or } |\xi_0'| > L\} \\ &\leq \int_{|\xi_0|\leq L} \nu(\mathrm{d}\xi) \int_{|\xi_0'|\leq L} \nu(\mathrm{d}\xi') R^{(T)}(\xi,\xi') + 2HT|c_{\nu_0}| \int_{|\xi_0|>L} \nu(\mathrm{d}\xi)|\xi_0|, \end{aligned} \tag{A.5}$$

where we recall the definition of c_λ from (2.16) and note that $|c_{\nu_0}| = \int \nu(\mathrm{d}\xi)|\xi_0|$ is bounded by β for all $\nu \in \mathcal{A}_{f,\beta}$. Consider the last term on the right-hand side of (A.5). Note that, for $r > L$, we can estimate $r \leq f(r)\varepsilon_L$ with $\varepsilon_L = \sup_{r>L} r/f(r)$, which vanishes as $L \to \infty$. Hence,

$$\int_{|\xi_0|>L} \nu(\mathrm{d}\xi)|\xi_0| \leq \varepsilon_L \int_{|\xi_0|>L} \nu(\mathrm{d}\xi) f(|\xi_0|) \leq \beta\varepsilon_L.$$

So the last term on the right-hand side of (A.5) vanishes, as $L \to \infty$. Consider the first term on the right-hand side of (A.5). Observe that, due to the point (i), the function $(\xi,\xi') \mapsto R(\xi,\xi')\mathbb{1}\{|\xi_0| \leq L\}\mathbb{1}\{|\xi_0'| \leq L\}$ is dominated by a bounded and continuous function. Together with the fact that $\nu \mapsto \nu \otimes \nu$ is continuous in the weak topology, this implies that $\nu \mapsto \int_{|\xi_0|\leq L} \nu(\mathrm{d}\xi) \int_{|\xi_0'|\leq L} \nu(\mathrm{d}\xi') R(\xi,\xi')$ is continuous with respect to the weak topology. □

Now, for the proof of the upper bound in the LDP in Section 6, we need the following lemma that provides an upper bound for the self-interaction cost of a collection of history trees, which we call the diagonal term. Let us fix a majorizing function $f : (0,\infty) \to (0,\infty)$ satisfying $f(r)/r \to \infty$ as $r \to \infty$ and $f(r) \geq r$ for any $r \in \mathbb{N}$ as well as a constant $\beta \in (0,\infty)$.

LEMMA A.3 (Upper bound for the diagonal term). *Fix a point process $Y_N = \sum_i \delta_{\Xi_i}$ on $\Gamma_T^{(1)}$, and assume the upper bound on K from (1.2). Then for any $L, N \in \mathbb{N}$,*

$$\frac{1}{N}\sum_i R(\Xi_i, \Xi_i) \leq HTL\beta + HTN\beta^2\varepsilon_L^2, \quad \text{with } \varepsilon_L = \sup_{r>L} \frac{r}{f(r)}, \tag{A.6}$$

holds on the event $\{\frac{1}{N} Y_N \in \mathcal{A}_{f,\beta}\}$.

Note that the lemma implies that the diagonal term is $o(N)$ on the specified event.

PROOF. We use the estimate (A.4) that holds under (1.2) and the fact that $\|\xi_0\|_1 = |\xi_0|$ for $\xi \in \Gamma_T^{(1)}$ to get

$$\begin{aligned}\frac{1}{N}\sum_i R^{(T)}(\Xi_i, \Xi_i) &= \int_{\Gamma_T^{(1)}} \frac{1}{N} Y_N(\mathrm{d}\xi) R^{(T)}(\xi, \xi)\\ &\le HT \int \frac{1}{N} Y_N(\mathrm{d}\xi)|\xi_0|^2 \qquad \text{(A.7)}\\ &\le HTL \int_{|\xi_0|\le L} \frac{1}{N} Y_N(\mathrm{d}\xi)|\xi_0| + HT \int_{|\xi_0|>L} \frac{1}{N} Y_N(\mathrm{d}\xi)|\xi_0|^2.\end{aligned}$$

Now, if $\{\frac{1}{N} Y_N \in \mathcal{A}_{f,\beta}\}$, we have that the first integral is smaller than β, and for the second integral, we have that

$$\int_{|\xi_0|>L} \frac{1}{N} Y_N(\mathrm{d}\xi)|\xi_0|^2 \le N\left(\int_{|\xi_0|>L} \frac{1}{N} Y_N(\mathrm{d}\xi)|\xi_0|\right)^2 \le N\beta^2 \varepsilon_L^2. \qquad \square$$

A.3. LDP for the PPP Y_N. We state and prove an LDP for the random variable Y_N introduced in Theorem 2.1. We feel that this result is not new, but we did not find a reference; hence, we give an outline of a proof. Recall that for two finite measures m, p on a Polish space $\mathcal{X}$ the relative entropy of m with respect to p is denoted by

$$H(m|p) = \int_{\mathcal{X}} m(\mathrm{d}x) \log \frac{\mathrm{d}m}{\mathrm{d}p}(x) + p(\mathcal{X}) - m(\mathcal{X}), \qquad \text{(A.8)}$$

if $m \ll p$, and $H(m|p) = \infty$ otherwise. Note that $H(\cdot|p)$ is convex and nonnegative with the only zero p. Furthermore, all its sublevel sets $\{m : H(m|p) \le C\}$ for $C \in \mathbb{R}$ are compact in the weak topology (a short proof of which is given below).

LEMMA A.4 (LDP for Y_N under PPP($N\mathfrak{m}$)). *Assume that $\mathcal{X}$ is a Polish space, pick a finite and positive measure $\mathfrak{m}$ on $\mathcal{X}$, and assume that Y_N is a Poisson point process with intensity measure $N\mathfrak{m}$. Then $(\frac{1}{N} Y_N)_{N\in\mathbb{N}}$ satisfies an LDP on $\mathcal{M}(\mathcal{X})$ with rate function $\nu \mapsto H(\nu|\mathfrak{m})$. All the level sets $\{\nu : H(\nu|\mathfrak{m}) \le C\}$ with $C \in \mathbb{R}$ are compact.*

PROOF. The abstract version of Cramér's theorem gives immediately an LDP for $\frac{1}{N} Y_N$. Indeed, note that we have in distribution that $Y_N = Z_1 + \cdots + Z_N$, where $Z_1, \ldots, Z_N$ are independent PPPs with intensity measure $\mathfrak{m}$. Then $\frac{1}{N} Y_N$, as the average of i.i.d. random objects, satisfies the LDP with rate function equal to the Legendre transform of the logarithm of the moment generating function of Z_1, which reads

$$\begin{aligned}\mathcal{M}^*(\mathcal{X}) \ni \nu \mapsto &\sup_{f\in\mathcal{C}_0(\mathcal{X})} \left(\langle \nu, f\rangle - \log \mathbb{E}\left[\mathrm{e}^{\langle f, Z_1\rangle}\right]\right)\\ &= \sup_{f\in\mathcal{C}_0(\mathcal{X})} \left(\langle \nu, f\rangle - \langle \mathrm{e}^f - 1, \mathfrak{m}\rangle\right),\end{aligned}$$

where we used a well-known formula for exponential Poisson moments, and $\mathcal{C}_0(\mathcal{X})$ is the closure in the uniform norm of the set of all continuous, compactly supported functions $f : \mathcal{X} \to \mathbb{R}$, the dual of which is set $\mathcal{M}_\pm(\mathcal{X})$ of all signed measures on $\mathcal{X}$. Now that we have formulated our problem in a Banach space setting, it is special case of the Gärtner–Ellis theorem. In order to finish the proof of the LDP that we need (i.e., to restrict from $\mathcal{M}_\pm(\mathcal{X})$ to $\mathcal{M}(\mathcal{X})$), we need to check that the restriction of the above rate function to $\mathcal{M}_\pm(\mathcal{X}) \setminus \mathcal{M}(\mathcal{X})$ is constantly equal to $+\infty$, which we leave to the reader.

Now we identify the rate function as $H(\nu|\mathfrak{m})$ by standard means. Indeed, if $\nu \ll \mathfrak{m}$, then we may insert (continuous bounded approximations of) $f = \log \frac{\mathrm{d}\nu}{\mathrm{d}\mathfrak{m}}$ and obtain that the rate function is $\geq H(\nu|\mathfrak{m})$, and the opposite inequality is seen by

$$H(\nu|\mathfrak{m}) = H(\nu|\mathrm{e}^f\,\mathrm{d}\mathfrak{m}) + \langle \nu, f\rangle - \langle \mathrm{e}^f, \mathfrak{m}\rangle + \mathfrak{m}(\mathcal{X}) \geq \langle \nu, f\rangle - \langle \mathrm{e}^f - 1, \mathfrak{m}\rangle,$$

since the entropy is nonnegative. In the case that ν is not absolutely continuous with respect to $\mathfrak{m}$, we take f as a continuous and bounded approximation of $M\mathbb{1}_A$ with a large M and a measurable set A that satisfies $\mathfrak{m}(A) = 0 < \nu(A)$; see [20], Lemma 3.2.13 for details.

The sets $\{\frac{\mathrm{d}\nu}{\mathrm{d}\mathfrak{m}} \colon H(\nu|\mathfrak{m}) \leq C\}$ are weakly compact in $L^1(\mathfrak{m})$ by uniform integrability since we can write $H(\nu|\mathfrak{m}) = \int (\frac{\mathrm{d}\nu}{\mathrm{d}\mathfrak{m}} \log \frac{\mathrm{d}\nu}{\mathrm{d}\mathfrak{m}} - \frac{\mathrm{d}\nu}{\mathrm{d}\mathfrak{m}} + 1)\,\mathrm{d}\mathfrak{m}$. From this we see that the sets $\{\nu \colon H(\nu|\mathfrak{m}) \leq C\}$ are compact in $\mathcal{M}(\mathcal{X})$ with respect to its (functional analytic) weak topology and, therefore, also in the topology generated by testing against bounded measurable functions (see [16], Theorem 4.7.25), which certainly include all continuous bounded functions. □

A.4. Alternate characterisation of the LDP rate function. We introduce here another representation of the rate function $I_\mu^{(T)}$ appearing in our LDP in Theorem 2.3. This will involve the well-known notion of a relative entropy, which allows us to deduce crucial properties of $I_\mu^{(T)}$ like the compactness of the sublevel sets. Since the measure $M_\mu^{(T)}$, defined in (2.5), does not always have finite total mass, we need to slightly extend its definition. Take any finite measure $\nu \in \mathcal{M}(\mathcal{S})$, then we define

$$M_\nu^{(T)} = \mathrm{e}\,\mathrm{Poi}_\nu \otimes \mathbb{Q}^{(T)} \in \mathcal{M}(\Gamma_T^{(1)}). \tag{A.9}$$

Now we are able to work with $M_{b\mu}^{(T)}$ for any $b \in (0,\infty)$. Notice that $M_\mu^{(T)}$ might not have a finite total mass, but we show in the next lemma that, under the assumption in (1.2), $|M_{b\mu}^{(T)}| < \infty$ for sufficiently small $b \in (0,\infty)$.

LEMMA A.5 (Moments of $M_{b\mu}^{(T)}$ under (1.2)). *Assume that* (1.2) *holds, and fix any $T, b \in (0,\infty)$. Then for any $n \in \mathbb{N}_0$,*

$$M_{b\mu}^{(T)}(\{\xi \in \Gamma_T^{(1)} \colon |\xi_0| = n\}) \leq \frac{\mathrm{e}^{1-b}}{2\pi T H}(bTH\mathrm{e}^2)^n n^{-2}. \tag{A.10}$$

Consequently, if $b < 1/H\mathrm{e}^2 T$, then $\int M_{b\mu}^{(T)}(\mathrm{d}\xi)|\xi_0|^\alpha < \infty$ for any $\alpha \in [0,\infty)$. Further, if $TH\mathrm{e}^2 < 1$, then $q_\mu^{(T)} < 1$.

PROOF. With the help of the estimate (5.17) for $\mathbb{Q}_k^{(T)}(\Gamma_{T,k}^{(1)})$ derived in Lemma 5.4 under assumption (1.2) and also using that $\mathbb{Q}_0^{(T)}(\Gamma_{T,0}^{(1)}) = 0$, we obtain that for any $n \in \mathbb{N}_0$

$$\begin{aligned} M_{b\mu}^{(T)}(\{\xi \in \Gamma_T^{(1)} \colon |\xi_0| = n\}) &= \mathrm{e}\int \mathrm{Poi}_\mu(\mathrm{d}k)\mathrm{e}^{1-b} b^{|k|}\mathbb{Q}_k^{(T)}(\Gamma_{T,k}^{(1)})\mathbb{1}\{|k| = n\} \\ &\leq \mathrm{e}^{1-b}\frac{b^n}{n!}\frac{(TH)^{n-1}}{(n-1)!}n^{2(n-1)} \\ &= \frac{\mathrm{e}^{1-b}}{TH}\frac{(bTH)^n}{(n!)^2}n^{2n-1} \leq \frac{\mathrm{e}^{1-b}}{2\pi TH}(bTH\mathrm{e}^2)^n n^{-2}, \end{aligned} \tag{A.11}$$

where we applied the Stirling bound $n! \geq n^n \mathrm{e}^{-n}\sqrt{2\pi n}$. Hence, (A.10) holds. For any $\alpha \in [0,\infty)$, we get that

$$\int M_{b\mu}^{(T)}(\mathrm{d}\xi)|\xi_0|^\alpha \leq \frac{\mathrm{e}^{1-b}}{TH}\sum_{n=0}^\infty \frac{(bTH)^n}{(n!)^2}n^{2n-1} \leq \frac{\mathrm{e}^{1-b}}{2\pi TH}\sum_{n=0}^\infty (bTH\mathrm{e}^2)^n n^{\alpha-2} < \infty$$

if $bTH\mathrm{e}^2 < 1$, since the geometric series with that parameter converges. Choosing $b = 1$, we get that $q_\mu^{(T)} \le TH\mathrm{e}^2$, which implies the last claim. □

Now we are ready to give an alternative formulation of the rate function $I_\mu^{(T)}(\cdot)$. Recall the definition of relative entropy from (A.8). Recall from (2.2) the noncoagulation functional $R^{(T)}$, and recall the operator with kernel $R^{(T)}$ from (2.8); furthermore, fix $\mu \in \mathcal{M}_1(\mathcal{S})$, and recall the definition of $I_\mu^{(T)} : \mathcal{M}(\Gamma_T^{(1)}) \to [0,\infty]$ from (2.9).

As mentioned above, even if $M_\mu^{(T)}$ might not have a finite total mass, we know from Lemma A.5 that, under the assumption in (1.2), $|M_{b\mu}^{(T)}| < \infty$ for sufficiently small $b \in (0,\infty)$, since $M_{b\mu}^{(T)}(\mathrm{d}\xi) = M_\mu^{(T)}(\mathrm{d}\xi)\mathrm{e}^{1-b}b^{|\xi_0|}$. As a consequence, the following corollary holds.

COROLLARY A.6. *For any $b \in (0,\infty)$ such that $|M_{b\mu}^{(T)}| < \infty$, if the density $\frac{\mathrm{d}\nu}{\mathrm{d}M_\mu^{(T)}}$ exists (which exists if and only $\frac{\mathrm{d}\nu}{\mathrm{d}M_{b\mu}^{(T)}}$ exists), we have the alternative representation*

$$
\begin{aligned}
I_\mu^{(T)}(\nu) = 1 - \big|M_{b\mu}^{(T)}\big| + H\big(\nu|M_{b\mu}^{(T)}\big) + \frac{1}{2}\langle \nu, \mathfrak{R}^{(T)}(\nu)\rangle \\
+ \int \nu_0(\mathrm{d}k)|k|\log b + (1-b)|\nu|.
\end{aligned}
\tag{A.12}
$$

(*Recall that we write $\nu_0 = \nu \circ \pi_0^{-1} \in \mathcal{M}(\mathcal{M}_{\mathbb{N}_0}(\mathcal{S}))$ for the projection of a measure $\nu \in \mathcal{M}(\Gamma_T^{(1)})$.*)

Now we see that the rate function $I_\mu^{(T)}$ has compact level sets. Indeed, as is seen from (A.12), $I_\mu^{(T)}(\nu)$ is a sum of terms, each of which is a lower semicontinuous function of ν on the set $\mathcal{A}_{f,\beta}$, defined in (2.7). Indeed, the quadratic term is lower semicontinuous by Lemma A.2(ii); the maps $\nu \mapsto \int \nu_0(\mathrm{d}k)|k| = |c_{\nu_0}|$ and $\nu \mapsto |\nu|$ are continuous on $\mathcal{A}_{f,\beta}$, as is seen using the arguments from the proof of Lemma A.2(iii). Finally, by Lemma A.4 the entropy $H(\cdot|M_{b\mu}^{(T)})$ has even compact sublevel sets (in $\mathcal{M}(\Gamma_T^{(1)})$ and hence in $\mathcal{A}_{f,\beta}$).

A.5. LDP conditioned on bounded particle sizes. In this section we present a variant of the LDP of Theorem 2.3 with conditioning on having only particles of sizes $\le L$ rather than conditioning on $\mathcal{A}_{f,\beta}$. This (more precisely, a variant) is used in Section 7.4 when we prove sufficient gelation criteria.

Let us define the measure $M_\mu^{(T,\le L)}$ for some fixed $L \in \mathbb{N}$, as the restriction of $M_\mu^{(T)}$ to the set

$$
\Gamma_{T,\le L}^{(1)} = \big\{\xi \in \Gamma_T^{(1)} : \xi_T \text{ is concentrated on } \mathcal{S} \times [L]\big\}, \quad \text{where } [L] = \{1,\dots,L\},
\tag{A.13}
$$

of particle trajectories with particles of sizes $\le L$. It is an easy consequence from Lemma 5.4 that $|M_\mu^{(T,\le L)}| < \infty$ under the assumption (1.2). In the following we often identify measures ν on $\Gamma_T^{(1)}$ that satisfy $\nu(\Gamma_T^{(1)} \setminus \Gamma_{T,\le L}^{(1)}) = 0$ with measures on $\Gamma_{T,\le L}^{(1)}$.

THEOREM A.7 (LDP for $\mathcal{V}_N^{(T)}$ under $\mathbb{P}_{\mathrm{Poi}_\mu}(\cdot|\mathcal{V}_N^{(T)}(\Gamma_T^{(1)} \setminus \Gamma_{T,\le L}^{(1)}) = 0)$). *Assume that the kernel K is continuous and satisfies the upper bound in (1.2). Replace the kernel K by $\frac{1}{N}K$. Pick $T \in (0,\infty)$ and $\mu \in \mathcal{M}_1(\mathcal{S})$. For any $L \in \mathbb{N}$, the distribution of $\mathcal{V}_N^{(T)}$ under $\mathbb{P}_{\mathrm{Poi}_\mu}(\cdot|\mathcal{V}_N^{(T)}(\Gamma_T^{(1)} \setminus \Gamma_{T,\le L}^{(1)}) = 0)$ satisfies an LDP on $\mathcal{M}(\Gamma_T^{(1)})$ with speed N and rate function*

$$
\mathcal{M}(\Gamma_T^{(1)}) \to [0,\infty], \qquad \nu \mapsto \begin{cases} I_\mu^{(T,\le L)} - \inf\limits_{\mathcal{M}(\Gamma_{T,\le L}^{(1)})} I_\mu^{(T,\le L)} & \text{if } \nu \in \mathcal{M}(\Gamma_{T,\le L}^{(1)}), \\ \infty & \text{otherwise}, \end{cases}
\tag{A.14}
$$

where

$$I_\mu^{(T,\le L)}(\nu) = H\big(\nu|M_\mu^{(T,\le L)}\big) + 1 - |M_\mu^{(T,\le L)}| + \frac{1}{2}\langle \nu, \mathfrak{R}^{(T)}(\nu)\rangle, \quad \nu \in \mathcal{M}\big(\Gamma_{T,\le L}^{(1)}\big). \tag{A.15}$$

The sublevel sets of this rate function are compact.

The proof of Theorem A.7 is a straightforward extension of the proof of Theorem 2.3. Actually, the proof of the lower bound in Theorem 2.3 gives already many details of the proof of the lower bound in Theorem A.7. The proof of the upper bound in Theorem 2.3 needs more than the proof of the upper bound in Theorem A.7. Finally, the compactness of the sublevel sets of the rate function in (A.15) is clear, since $|M_\mu^{(T,\le L)}|$ is finite (hence, one can take $b = 1$; compare to (7.1)) and $\nu \mapsto \langle \nu, \mathfrak{R}^{(T)}(\nu)\rangle$ is lower semicontinuous.

Equivalently, we have the following lemmas and corollaries.

COROLLARY A.8 (Accumulation points). *In the situation of Theorem A.7, the distribution of $\mathcal{V}_N^{(T)}$ under $\mathbb{P}_{\mathrm{Poi}_\mu}^{(N)}(\cdot|\mathcal{V}_N^{(T)}(\Gamma_T^{(1)} \setminus \Gamma_{T,\le L}^{(1)}) = 0)$ is tight in N, and each limit point of this distribution along any subsequence is concentrated on the set of minimisers of $I_\mu^{(T,\le L)}$.*

LEMMA A.9. *For any $T \in (0,\infty)$ and any $L \in \mathbb{N}$, any minimiser $\nu^{(T,\le L)}$ of $I_\mu^{(T,\le L)}$ defined in (A.15) on the set $\mathcal{M}(\Gamma_{T,\le L}^{(1)})$ satisfies the Euler–Lagrange equation*

$$\nu^{(T)}(\mathrm{d}\xi) = M_\mu^{(T,\le L)}(\mathrm{d}\xi)\mathrm{e}^{-\mathfrak{R}^{(T)}(\nu^{(T)})(\xi)}, \quad \xi \in \Gamma_{T,\le L}^{(1)}. \tag{A.16}$$

The proof of the above lemma is analogous to that of Lemma 7.3.

COROLLARY A.10 (Bounds on $\nu^{(T)}$). *Assume that $\nu^{(T)}$ is a minimiser of $I_\mu^{(T,\le L)}$ on $\mathcal{M}(\Gamma_{T,\le L}^{(1)})$ for some $L \in \mathbb{N}$:*

(i) *Under the assumption in (1.2) and if $T < 1/\mathrm{e}^2 H$,*

$$\int \nu^{(T)}(\mathrm{d}\xi)|\xi_0|^2 \le \frac{\mathrm{e}^2}{2\pi(1-\mathrm{e}^2 TH)}.$$

(ii) *Under the assumptions in (1.2) and (1.3), for any $T > 0$,*

$$\int \nu^{(T)}(\mathrm{d}\xi)|\xi_0| \le \max\left\{\frac{1}{hT}\log\big(2TH\mathrm{e}^2\big), \frac{1}{2\pi HT}\right\}.$$

Since $M_\mu^{(T,\le L)} \le M_\mu^{(T)}$, the above corollary is a direct consequence of Lemma 7.4.

COROLLARY A.11 (Minimizers of $I_\mu^{(T,\le L)}$). *Assume that the kernel K is continuous and satisfies (1.2) and (1.3). Then the following hold:*

(i) *For any $L \in \mathbb{N}$, every minimiser $\nu^{(T,\le L)}$ of $I_\mu^{(T,\le L)}$ satisfies the Euler–Lagrange equation on $\mathcal{M}(\Gamma_{T,\le L}^{(1)})$,*

$$\nu(\mathrm{d}\xi) = M_\mu^{(T)}(\mathrm{d}\xi)\mathrm{e}^{-\mathfrak{R}(\nu)(\xi)}, \quad \xi \in \Gamma_{T,\le L}^{(1)}. \tag{A.17}$$

(ii) *It holds that $|c_{\nu_0^{(T)}}^{(\le L)}| \le \frac{2\log T}{hT}$ for any $L \in \mathbb{N}$ and every minimiser $\nu^{(T,\le L)}$ of $I_\mu^{(T,\le L)}$.*

Acknowledgments. The authors would like to thank the anonymous referees for their valuable comments.

Funding. This research has been funded by the Deutsche Forschungsgemeinschaft (DFG) through grant CRC 1114 "Scaling Cascades in Complex Systems," Project C08, by grant SPP 2265 "Random Geometric Systems," Project P01, and by grant SPP 2256, Project 11.

LA and WK acknowledge partial support from GNAMPA, Indam, via the program "Professori Visitatori 2022."

LA acknowledges partial financial support by the Italian Ministry of University and Research (MUR) via PRIN 2022—ConStRAINeD-CUP-2022XRWY7W.

REFERENCES

[1] ADAMS, S., COLLEVECCHIO, A. and KÖNIG, W. (2011). A variational formula for the free energy of an interacting many-particle system. *Ann. Probab.* **39** 683–728. MR2789510 https://doi.org/10.1214/10-AOP565

[2] AGAZZI, A., ANDREIS, L., PATTERSON, R. I. A. and RENGER, D. R. M. (2022). Large deviations for Markov jump processes with uniformly diminishing rates. *Stochastic Process. Appl.* **152** 533–559. MR4459039 https://doi.org/10.1016/j.spa.2022.06.017

[3] AGAZZI, A., DEMBO, A. and ECKMANN, J.-P. (2018). Large deviations theory for Markov jump models of chemical reaction networks. *Ann. Appl. Probab.* **28** 1821–1855. MR3809478 https://doi.org/10.1214/17-AAP1344

[4] ALDOUS, D. J. (1999). Deterministic and stochastic models for coalescence (aggregation and coagulation): A review of the mean-field theory for probabilists. *Bernoulli* **5** 3–48. MR1673235 https://doi.org/10.2307/3318611

[5] ANDREIS, L., IYER, T. and MAGNANINI, E. (2023). Gelation in cluster coagulation processes. *Ann. Inst. Henri Poincaré*. To appear. Available at arXiv:2308.10232. https://doi.org/10.1214/25-AIHP1581

[6] ANDREIS, L., IYER, T. and MAGNANINI, E. (2024). Convergence of cluster coagulation dynamics. Available at arXiv:2406.12401.

[7] ANDREIS, L., KÖNIG, W., LANGHAMMER, H. and PATTERSON, R. I. A. (2023). A large-deviations principle for all the components in a sparse inhomogeneous random graph. *Probab. Theory Related Fields* **186** 521–620. MR4586227 https://doi.org/10.1007/s00440-022-01180-7

[8] ANDREIS, L., KÖNIG, W. and PATTERSON, R. I. A. (2021). A large-deviations principle for all the cluster sizes of a sparse Erdős-Rényi graph. *Random Structures Algorithms* **59** 522–553. MR4323309 https://doi.org/10.1002/rsa.21007

[9] BALL, J. M. and CARR, J. (1990). The discrete coagulation-fragmentation equations: Existence, uniqueness, and density conservation. *J. Stat. Phys.* **61** 203–234. MR1084278 https://doi.org/10.1007/BF01013961

[10] BASILE, G., BENEDETTO, D., BERTINI, L. and CAGLIOTI, E. (2024). Asymptotic probability of energy increasing solutions to the homogeneous Boltzmann equation. *Ann. Appl. Probab.* **34** 3995–4021. MR4783035 https://doi.org/10.1214/24-aap2057

[11] BASILE, G., BENEDETTO, D., BERTINI, L. and ORRIERI, C. (2021). Large deviations for Kac-like walks. *J. Stat. Phys.* **184** 10. MR4281234 https://doi.org/10.1007/s10955-021-02794-2

[12] BASILE, G., BENEDETTO, D., CAGLIOTI, E. and BERTINI, L. (2023). Large deviations for a binary collision model: Energy evaporation. *Math. Eng.* **5** 001. MR4370323 https://doi.org/10.1007/bf02828297

[13] BELAVKIN, V. P. and KOLOKOL'TSOV, V. N. (2003). On a general kinetic equation for many-particle systems with interaction, fragmentation and coagulation. *Proc. R. Soc. Lond. Ser. A Math. Phys. Eng. Sci.* **459** 727–748. MR1996729 https://doi.org/10.1098/rspa.2002.1026

[14] BILLINGSLEY, P. (2013). *Convergence of Probability Measures*. Wiley, New York.

[15] BODINEAU, T., GALLAGHER, I., SAINT-RAYMOND, L. and SIMONELLA, S. (2023). Statistical dynamics of a hard sphere gas: Fluctuating Boltzmann equation and large deviations. *Ann. of Math.* (2) **198** 1047–1201. MR4660136 https://doi.org/10.4007/annals.2023.198.3.3

[16] BOGACHEV, V. I. (2007). *Measure Theory. Vol. I, II*. Springer, Berlin. MR2267655 https://doi.org/10.1007/978-3-540-34514-5

[17] BONACINI, M., NIETHAMMER, B. and VELÁZQUEZ, J. J. L. (2019). Self-similar gelling solutions for the coagulation equation with diagonal kernel. *Ann. Inst. H. Poincaré Anal. Non Linéaire* **36** 705–744. MR3926520 https://doi.org/10.1016/j.anihpc.2018.09.001

[18] COLLIN, O., JAHNEL, B. and KÖNIG, W. (2023). A micro-macro variational formula for the free energy of a many-body system with unbounded marks. *Electron. J. Probab.* **28** 118. MR4659836 https://doi.org/10.1214/23-ejp1014

[19] Dembo, A. and Zeitouni, O. (2010). *Large Deviations Techniques and Applications*. *Stochastic Modelling and Applied Probability* **38**. Springer, Berlin. MR2571413 https://doi.org/10.1007/978-3-642-03311-7

[20] Deuschel, J.-D. and Stroock, D. W. (1989). *Large Deviations*. *Pure and Applied Mathematics* **137**. Academic Press, Boston, MA. MR0997938

[21] Dubovskiĭ, P. B. and Stewart, I. W. (1996). Existence, uniqueness and mass conservation for the coagulation-fragmentation equation. *Math*. *Methods Appl*. *Sci*. **19** 571–591. MR1385155 https://doi.org/10.1002/(SICI)1099-1476(19960510)19:7<571::AID-MMA790>3.0.CO;2-Q

[22] Ferreira, M. A., Lukkarinen, J., Nota, A. and Velázquez, J. J. L. (2021). Localization in stationary non-equilibrium solutions for multicomponent coagulation systems. *Comm*. *Math*. *Phys*. **388** 479–506. MR4328062 https://doi.org/10.1007/s00220-021-04201-z

[23] Flory, P. J. (1941). Molecular size distribution in three dimensional polymers. I. Gelation. *J*. *Amer*. *Chem*. *Soc*. **63** 3083–3090. https://doi.org/10.1021/ja01856a061

[24] Fournier, N. and Giet, J.-S. (2004). Convergence of the Marcus-Lushnikov process. *Methodol*. *Comput*. *Appl*. *Probab*. **6** 219–231. MR2035293 https://doi.org/10.1023/B:MCAP.0000017714.56667.67

[25] Gillespie, D. T. (1972). The stochastic coalescence model for cloud droplet growth. *J*. *Atmos*. *Sci*. **29** 1496–1510. https://doi.org/10.1175/1520-0469(1972)029<1496:TSCMFC>2.0.CO; 2

[26] Guiaş, F. (2001). Convergence properties of a stochastic model for coagulation-fragmentation processes with diffusion. *Stoch*. *Anal*. *Appl*. **19** 245–278. MR1841439 https://doi.org/10.1081/SAP-100001188

[27] Hammond, A. and Rezakhanlou, F. (2007). The kinetic limit of a system of coagulating Brownian particles. *Arch*. *Ration*. *Mech*. *Anal*. **185** 1–67. MR2308858 https://doi.org/10.1007/s00205-006-0033-5

[28] Heydecker, D. (2023). Large deviations of Kac's conservative particle system and energy nonconserving solutions to the Boltzmann equation: A counterexample to the predicted rate function. *Ann*. *Appl*. *Probab*. **33** 1758–1826. MR4583658 https://doi.org/10.1214/22-aap1852

[29] Jeon, I. (1998). Existence of gelling solutions for coagulation-fragmentation equations. *Comm*. *Math*. *Phys*. **194** 541–567. MR1631473 https://doi.org/10.1007/s002200050368

[30] Kallenberg, O. (1997). *Foundations of Modern Probability*. *Probability and Its Applications* (*New York*). Springer, New York. MR1464694

[31] Lang, R. and Nguyen, X.-X. (1980). Smoluchowski's theory of coagulation in colloids holds rigorously in the Boltzmann-Grad-limit. *Z*. *Wahrsch*. *Verw*. *Gebiete* **54** 227–280. MR0602510 https://doi.org/10.1007/BF00534345

[32] Last, G. and Penrose, M. (2018). *Lectures on the Poisson Process*. *Institute of Mathematical Statistics Textbooks* **7**. Cambridge Univ. Press, Cambridge. MR3791470

[33] Léonard, C. (1995). On large deviations for particle systems associated with spatially homogeneous Boltzmann type equations. *Probab*. *Theory Related Fields* **101** 1–44. MR1314173 https://doi.org/10.1007/BF01192194

[34] Lushnikov, A. A. (1978). Coagulation in finite systems. *J*. *Colloid Interface Sci*. **65** 276–285. https://doi.org/10.1016/0021-9797(78)90158-3

[35] Marcus, A. H. (1968). Stochastic coalescence. *Technometrics* **10** 133–143. MR0223151 https://doi.org/10.2307/1266230

[36] Normand, R. and Zambotti, L. (2011). Uniqueness of post-gelation solutions of a class of coagulation equations. *Ann*. *Inst*. *H*. *Poincaré Anal*. *Non Linéaire* **28** 189–215. MR2784069 https://doi.org/10.1016/j.anihpc.2010.10.005

[37] Norris, J. R. (1999). Smoluchowski's coagulation equation: Uniqueness, nonuniqueness and a hydrodynamic limit for the stochastic coalescent. *Ann*. *Appl*. *Probab*. **9** 78–109. MR1682596 https://doi.org/10.1214/aoap/1029962598

[38] Norris, J. R. (2000). Cluster coagulation. *Comm*. *Math*. *Phys*. **209** 407–435. MR1737990 https://doi.org/10.1007/s002200050026

[39] Rezakhanlou, F. (2013). Gelation for Marcus-Lushnikov process. *Ann*. *Probab*. **41** 1806–1830. MR3098059 https://doi.org/10.1214/11-AOP691

[40] Shwartz, A. and Weiss, A. (2005). Large deviations with diminishing rates. *Math*. *Oper*. *Res*. **30** 281–310. MR2142034 https://doi.org/10.1287/moor.1040.0121

[41] Siegmund-Schultze, R. and Wagner, W. (2006). Induced gelation in a two-site spatial coagulation model. *Ann*. *Appl*. *Probab*. **16** 370–402. MR2209346 https://doi.org/10.1214/105051605000000755

[42] Sun, W. (2023). A conditional compound Poisson process approach to the sparse Erdős–Rényi random graphs: Moderate deviations. Available at arXiv:2310.06348.

[43] Sun, W. (2024). Pathwise large deviations for the pure jump k-nary interacting particle systems. *Ann*. *Appl*. *Probab*. **34** 743–794. MR4696290 https://doi.org/10.1214/23-aap1977

[44] von Smoluchowski, M. (1916). Drei Vorträge über Diffusion, Brownsche Molekularbewegung und Koagulation von Kolloidteilchen. *Phys*. *Z*. **XVII** 585–599.

[45] WAGNER, W. (2006). Post-gelation behavior of a spatial coagulation model. *Electron*. *J*. *Probab*. **11** 35. MR2261057 https://doi.org/10.1214/EJP.v11-359

[46] WELLS, C. G. (2006). A stochastic approximation scheme and convergence theorem for particle interactions with perfectly reflecting boundary conditions. *Monte Carlo Methods Appl*. **12** 291–342. MR2274698 https://doi.org/10.1163/156939606778705182