

A LARGE-DEVIATIONS PRINCIPLE FOR ALL THE CLUSTER SIZES OF A SPARSE ERDŐS–RÉNYI GRAPH

LUISA ANDREIS, WOLFGANG KÖNIG, AND ROBERT I. A. PATTERSON

(19 August, 2020)

ABSTRACT. Let $\mathcal{G}(N, \frac{1}{N}t_N)$ be the Erdős–Rényi graph with connection probability $\frac{1}{N}t_N \sim t/N$ as $N \rightarrow \infty$ for a fixed $t \in (0, \infty)$. We derive a large-deviations principle for the empirical measure of the sizes of all the connected components of $\mathcal{G}(N, \frac{1}{N}t_N)$, registered according to microscopic sizes (i.e., of finite order), macroscopic ones (i.e., of order N), and mesoscopic ones (everything in between). The rate function explicitly describes the microscopic and macroscopic components and the fraction of vertices in components of mesoscopic sizes. Moreover, it clearly captures the well known phase transition at $t = 1$ as part of a comprehensive picture. The proofs rely on elementary combinatorics and on known estimates and asymptotics for the probability that subgraphs are connected. We also draw conclusions for the strongly related model of the multiplicative coalescent, the Marcus–Lushnikov coagulation model with monodisperse initial condition, and its gelation phase transition.

MSC 2020: 05C80, 60F10, 60K35, 82B26.

Keywords and phrases. Erdős–Rényi random graph, component sizes, large deviations, empirical measure, phase transition, sizes, multiplicative coalescent, gelation.

1. INTRODUCTION

In this paper, we study the Erdős–Rényi random graph $\mathcal{G}(N, \frac{1}{N}t_N)$, that is, the random graph on the vertex set $[N] = \{1, \dots, N\}$, where each two distinct vertices are independently connected with probability $\frac{1}{N}t_N$. We will be working in the sparse regime, i.e., we assume that $\lim_{N \rightarrow \infty} t_N = t$ for some fixed $t \in (0, \infty)$. This is the regime in which the famous phase transition of the emergence of a giant cluster at $t = 1$ occurs, which was detected and characterised for the first time in the seminal paper [ER60]. For an extensive overview on the model see the classical reference [Bol01].

Our new contribution in this paper is a comprehensive study of the family of the sizes of all the connected components of $\mathcal{G}(N, \frac{1}{N}t_N)$, registered according to the asymptotic order of the size in the limit as $N \rightarrow \infty$. We distinguish here *microscopic components* (i.e., with size of order one), *macroscopic components* (i.e., size of order N , usually referred to as *giant components*) and *mesoscopic* ones (everything in between). We summarize all this information in terms of two empirical measures and derive a large-deviation principle (LDP) for them. Our rate function is rather explicit.

Such a principle gives information about the exponential decay rate of all sorts of events, e.g, the emergence of more than one giant cluster or the presence of a non-trivial proportion of

vertices in mesoscopic components. Moreover, the minimizers of the rate function represent the most likely configurations of the graph, which is expressed in terms of a law of large numbers for the objects that satisfy the LDP. In this way, we recover the mentioned phase transition and collect detailed information about the statistics of the sizes of all the components, both in the subcritical regime (where no giant component occurs) and the supercritical one.

Many investigations of the Erdős–Rényi graph and other random graphs in the sparse regime rely on approximations of subgraphs with certain Galton–Watson trees and other branching processes. We would like to stress that our approach does not use such arguments and is therefore an alternate ansatz.

In Section 1.1, we introduce our approach, in Section 1.2, we formulate our main results about large deviations and in Section 1.3 their consequences for the phase transition, and in Section 1.4 we give a literature survey.

Our original interest in this study was triggered by a desire to understand *random particle processes with coagulation*, in particular its simplest variant, the *Marcus–Lushnikov model* with multiplicative coagulation kernel. We introduce this process and its connections with our work on the LDP for the Erdős–Rényi graph in Section 1.5.

Another highly interesting connection appears with a LDP-proof for the well-known *Bose–Einstein condensation* phase transition that appears in the free (i.e., non-interacting) Bose gas; we will explain the similarities and the differences in Section 1.6.

1.1. Micro- and macroscopic empirical measures. Let us introduce the main objects that we study in this paper. For the remainder of the paper, we fix $t \in (0, \infty)$ and will be working with the graph $\mathcal{G}(N, \frac{1}{N}t_N)$, where $t_N = t + o(1)$ as $N \rightarrow \infty$. By

$$S_1^{(N)} \geq S_2^{(N)} \geq \dots \geq S_n^{(N)} \geq 1, \quad \sum_{i=1}^n S_i^{(N)} = N, \quad (1.1)$$

we denote the sizes of all the connected components of $\mathcal{G}(N, \frac{1}{N}t_N)$, ordered in a decreasing way ($n \in \{1, \dots, N\}$ is the number of components). We want to describe the entire family $(S_i^{(N)})_{i \in \{1, \dots, n\}}$ in the limit $N \rightarrow \infty$. This is a comprehensive object, which contains several scales. In order to adequately describe the most important two scales, it will be convenient to work with two *empirical measures* of the component sizes in the microscopic and macroscopic size ranges:

$$\text{Mi}^{(N)} = \frac{1}{N} \sum_{i=1}^n \delta_{S_i^{(N)}} \quad \text{and} \quad \text{Ma}^{(N)} = \sum_{i=1}^n \delta_{\frac{1}{N}S_i^{(N)}}. \quad (1.2)$$

Intuitively, while $\text{Mi}^{(N)}$ registers the proportion of components of ‘microscopic’ sizes $1, 2, 3, \dots$ on the scale N , $\text{Ma}^{(N)}$ registers the components of ‘macroscopic’ sizes, i.e. order N . Note that each of the two measures admits a one-to-one map onto the vector $(S_i^{(N)})_{i \in \{1, \dots, n\}}$ for fixed $N \in \mathbb{N}$ and therefore contains all the information contained in the vector. However, in the limit $N \rightarrow \infty$, they will be able to describe only the statistics of the microscopic, respectively macroscopic, part of the particle configuration. We would like to stress here that this issue lies at the heart of the phase transition of the emergence of a giant component, i.e., a macroscopic size.

Here is a non-technical, intuitive explanation: in the limit as $N \rightarrow \infty$, all sizes $S_i^{(N)}$ that somehow diverge, will vanish from the support of $\text{Mi}^{(N)}$ ‘at infinity’, and all sizes $S_i^{(N)}$ that

are $\ll N$ will vanish from the support of $\text{Ma}^{(N)}$ ‘at zero’. Hence, $\text{Mi}^{(N)}$ may leak out mass at infinity, and $\text{Ma}^{(N)}$ at zero. It is by no means automatic that all the mass that leaks out from the microscopic part at infinity enters the macroscopic part at zero. In order to control that, we also need to take care of the *mesoscopic* mass, coming from particle masses $1 \ll S_i^{(N)} \ll N$. Since here a lot of scales are contained (indeed, a continuum of scales), we will not be able to say anything about these sizes, but only about the total proportion of vertices belonging to such components.

The famous phase transition (proved first in [ER60]) says that, for $t \leq 1$, there is no loss of mass from $\text{Mi}^{(N)}$ (i.e., the first moment stays equal to one in the limit), and $\text{Ma}^{(N)}$ converges to the zero measure (i.e., loses all its mass), while for $t > 1$, the total mass of $\text{Mi}^{(N)}$ loses a positive amount equal to that retained by $\text{Ma}^{(N)}$ in the limit, and this results in a single Dirac measure. In both cases, the mesoscopic part vanishes, even though mesoscopic components are present in the graph with high probability, but their proportion is negligible.

These are assertions of the type of laws of large numbers. However, in the setting of a large-deviation principle as we are working here, we will obtain significant results also about the probabilities of several very unlikely events, like the emergence of non-trivial mesoscopic total mass, of more than one giant cluster and of different statistics of microscopic component sizes. Note that we decided to work on probabilities that are on an exponential scale N and for connection probabilities of the form $\frac{1}{N}(t + o(1))$ for general $t \in (0, \infty)$. This excludes for example all (highly interesting) phenomena that occur with respect to cluster sizes of order $N^{2/3}$ when considering more specified connection probabilities of the size $\frac{1}{N}(1 + cN^{-1/3})$; see [Ald97].

Now let us give a more technical explanation of the issue about possible losses of masses, which will also set the frame for the mathematical treatment. We will conceive the discrete measure $\text{Mi}^{(N)} = (\text{Mi}_k^{(N)})_{k \in \mathbb{N}}$ as a random element of the sequence set $\mathcal{N} = \bigcup_{c \in [0,1]} \mathcal{N}(c)$, where

$$\mathcal{N}(c) = \left\{ \Lambda = (\lambda_k)_{k \in \mathbb{N}} \in [0, \infty)^{\mathbb{N}} : \sum_{k \in \mathbb{N}} k \lambda_k = c \right\}, \quad c > 0. \quad (1.3)$$

We equip $\mathcal{N} = \{ \Lambda : \sum_k k \lambda_k \leq 1 \}$ with the topology of coordinate-wise convergence, which makes it compact by the Bolzano–Weierstrass theorem combined with Fatou’s lemma.

The point measure $\text{Ma}^{(N)}$ is a random element of the set $\mathcal{M} := \bigcup_{c \in [0,1]} \mathcal{M}_{\mathbb{N}_0}((0, 1]; c)$, where

$$\mathcal{M}_{\mathbb{N}_0}((0, 1]; c) = \left\{ \alpha \in \mathcal{M}_{\mathbb{N}_0}((0, 1]) : \int_{(0,1]} x \alpha(dx) = c \right\}, \quad (1.4)$$

and $\mathcal{M}_{\mathbb{N}_0}((0, 1])$ is the set of all measures on $(0, 1]$ with values in $\mathbb{N}_0 = \{0\} \cup \mathbb{N}$. We equip \mathcal{M} with the topology that is induced by functionals of the form $\mu \mapsto \int_{(0,1]} f(x) \mu(dx)$ where $f : (0, 1] \rightarrow \mathbb{R}$ is continuous and compactly supported. We sometimes write the elements of \mathcal{M} as $\alpha = \sum_j \delta_{\alpha_j}$ with $1 \geq \alpha_1 \geq \alpha_2 \geq \dots > 0$ and $\sum_j \alpha_j \leq 1$, where j extends over a finite subset of \mathbb{N} or over \mathbb{N} . Then convergence is equivalent with the pointwise convergence of each of the atoms. By similar arguments as for \mathcal{N} , also \mathcal{M} is compact. We equip the product of \mathcal{N} and \mathcal{M} with the product topology, so that it is also compact.

Important quantities are the expectations of the sub-probability distributions $\Lambda \in \mathcal{N}$ respectively $\alpha \in \mathcal{M}$, i.e., the maps

$$\Lambda \mapsto c_\Lambda := \sum_{k \in \mathbb{N}} k \lambda_k \quad \text{and} \quad \alpha \mapsto c_\alpha := \int_{(0,1]} x \alpha(dx).$$

Note that they are not continuous in the respective topologies, but only lower semicontinuous, according to Fatou's lemma. Indeed, even though the microscopic and macroscopic expectations $c_{\text{Mi}^{(N)}} = \sum_k k \text{Mi}_k^{(N)}$ and $c_{\text{Ma}^{(N)}} = \int_{(0,1]} x \text{Ma}^{(N)}(dx)$ are each equal to one for any N , they may (and will) lose mass in the limit $N \rightarrow \infty$. We sometimes call c_Λ and c_α the *total masses* of the microscopic, respectively macroscopic, configuration Λ and α , since they stand for the total number of particles, after scaling.

The mathematical treatment of the mesoscopic part of the component sizes is more technical, as it requires the introduction of two cutting parameters $R \in \mathbb{N}$ and $\varepsilon \in (0, 1)$. Indeed, a size $S_i^{(N)}$ is called (R, ε) -mesoscopic if $R < S_i^{(N)} < \varepsilon N$, and the definition of mesoscopic sizes requires making the limit $N \rightarrow \infty$, followed by $R \rightarrow \infty$ and $\varepsilon \downarrow 0$. There are several scales (indeed, a continuum of scales) contained in this part and in this regime it does not seem reasonable to consider an empirical measure for this part; therefore we will consider only the total proportion of mesoscopic vertices.

Let us remark that our choice of considering exclusively the size of each component, disregarding its bond structure, comes from the interest in coagulation processes, where only the sizes matter; see Section 1.5. An extension of our work to empirical measures of the components seems to require only moderate additional work, at least as it concerns the microscopic part. See Section 1.4 for earlier LDP-investigation of the components as subgraphs.

1.2. Our results: large-deviations principles. In this section, we present all our results on the LDP satisfied by the empirical measure of statistics of component sizes of the Erdős–Rényi graph $\mathcal{G}(N, \frac{1}{N}t_N)$, the random graph on $[N] = \{1, \dots, N\}$ with connection probability $\frac{1}{N}t_N$, and we assume that $t_N = t + o(1)$ with fixed $t \in (0, \infty)$. In Section 1.3 we will draw conclusions about the phase transition from that. Our main result is the following description of the two empirical measures $\text{Mi}^{(N)}$ and $\text{Ma}^{(N)}$ in terms of a joint large-deviations principle (LDP).

Theorem 1.1 (LDP for the empirical measures). *As $N \rightarrow \infty$, the pair $(\text{Mi}^{(N)}, \text{Ma}^{(N)})$ satisfies a large-deviations principle with speed N and rate function*

$$(\Lambda, \alpha) \mapsto I(\Lambda, \alpha; t) = \begin{cases} I_{\text{Mi}}(\Lambda; t) + I_{\text{Ma}}(\alpha; t) + (1 - c_\Lambda - c_\alpha) \left(\frac{t}{2} - \log t \right), & \text{if } c_\Lambda + c_\alpha \leq 1, \\ \infty & \text{otherwise,} \end{cases}$$

where we write $\Lambda = (\lambda_k)_{k \in \mathbb{N}}$ and

$$I_{\text{Mi}}(\Lambda; t) = \sum_{k=1}^{\infty} \lambda_k \log \frac{k! t \lambda_k}{e k^{k-2}} + c_\Lambda \left(1 + \frac{t}{2} - \log t \right), \quad c_\Lambda = \sum_{k=1}^{\infty} k \lambda_k, \quad (1.5)$$

$$I_{\text{Ma}}(\alpha; t) = \int_0^1 \left[x \log \frac{x}{1 - e^{-tx}} + \frac{t}{2} x(1 - x) \right] \alpha(dx), \quad c_\alpha = \int_{(0,1]} x \alpha(dx). \quad (1.6)$$

The proof of this theorem is in Section 3; it is based on an explicit combinatorial formula for the joint distribution of all the component sizes, followed by analysis of the arising exponential rates. We organised the three terms of I in the way in which they were derived from the

influences of the three parts (micro, macro and meso) in the course of the proof, even though this leads to a cancellation of terms involving c_Λ and c_α . This implies also that separate conclusions about the microscopic and the macroscopic parts can conveniently be made (see Corollaries 1.2 and 1.3). Informally, in (1.5) the terms involving λ_k , e and $k!$ in the logarithm derive from the combinatorial number of possibilities to decompose $[N]$ into the requested configuration of subsets, the term k^{k-2} and the t in the logarithm stem from the probability that these subsets are connected, and all the other terms from the probability that any of these subsets is not connected with the remainder. This interpretation is not immediate, since a number of asymptotic manipulations have been made during the proof. Similar remarks apply to (1.6). Interestingly, on the right-hand side of (1.5) we see, up to normalization, a relative entropy of $(k\lambda_k)_{k \in \mathbb{N}}$ with respect to the *Borel distribution* $\text{Bo}_\mu(k) = e^{-\mu k}(\mu k)^{k-1}/k!$ for a particular choice of μ ; a fact that will be crucial in the analysis of minimizers of the rate function, see the proofs of Corollaries 1.2 and 1.3 and of Theorem 1.5.

Let us recall the notion of an LDP: Theorem 1.1 says that, for any open set $G \subset \mathcal{N} \times \mathcal{M}$ respectively closed set $F \subset \mathcal{N} \times \mathcal{M}$,

$$\begin{aligned} \liminf_{N \rightarrow \infty} \frac{1}{N} \log \mathbb{P}_N((\text{Mi}^{(N)}, \text{Ma}^{(N)}) \in G) &\geq -\inf_G I(\cdot; t), \\ \limsup_{N \rightarrow \infty} \frac{1}{N} \log \mathbb{P}_N((\text{Mi}^{(N)}, \text{Ma}^{(N)}) \in F) &\leq -\inf_F I(\cdot; t), \end{aligned}$$

where we wrote \mathbb{P}_N for the probability measure for $\mathcal{G}(N, \frac{1}{N}t_N)$. For a comprehensive presentation of the theory of large-deviations, see e.g. [DZ10]. It is not difficult to see that since the rate function $I(\cdot, \cdot; t)$ is lower semicontinuous and $\mathcal{N} \times \mathcal{M}$ is compact, it is even a good rate function, i.e., its level sets $\{(\Lambda, \alpha) : I(\Lambda, \alpha; t) \leq r\}$ are compact for any r .

It is well-known in the theory of large deviations (and easy to deduce from the LDP) that for many interesting sets $A \subset \mathcal{N} \times \mathcal{M}$ one also has that $\mathbb{P}_N((\text{Mi}^{(N)}, \text{Ma}^{(N)}) \in A) = e^{-N \inf_A I(\cdot; t)(1+o(1))}$, for example for sets A that are equal to the closure of their open kernel. There are choices of such sets that give the precise exponential rates of interesting events, for instance the event that there are a given number of components larger than Na , for some $a > 0$, or that a given component size appears with a certain least density, or that a given positive percentage of vertices are contained in components of a given range of sizes (e.g., in $\{1, \dots, R\}$ or in $\{R, \dots, \varepsilon N\}$ or in $\{\varepsilon N, \dots, N\}$), and certainly all kinds of combinations of such events.

From our main result, the LDP in Theorem 1.1, a number of other LDPs follow via the contraction principle, according which if a random variable satisfies an LDP, so does its image under a continuous transformation; see [DZ10]. Let us begin with the component size distribution of the microscopic part.

Corollary 1.2 (LDP for microscopic component size statistics). *As $N \rightarrow \infty$, $\text{Mi}^{(N)}$ satisfies an LDP with speed N and rate function $\mathcal{I}_{\text{Mi}}(\cdot; t) : \mathcal{N} \rightarrow [0, \infty]$, given by*

$$\mathcal{I}_{\text{Mi}}(\Lambda; t) = \inf_{\alpha \in \mathcal{M}} I(\Lambda, \alpha; t) = I_{\text{Mi}}(\Lambda; t) - (1 - c_\Lambda) \left(\log \frac{1 - e^{(c_\Lambda - 1)t}}{1 - c_\Lambda} - \frac{c_\Lambda t}{2} \right). \quad (1.7)$$

The first equality comes from the application of the contraction principle; while the second equality is purely analytical and it is checked in Lemma 4.1. There it is seen that, given any $\Lambda \in \mathcal{N}$, it is always optimal to have all the remaining mass $1 - c_\Lambda$ in one single macroscopic component.

In the same way one can investigate the macroscopic part of the system.

Corollary 1.3 (LDP for macroscopic particles). *As $N \rightarrow \infty$, $\text{Ma}^{(N)}$ satisfies an LDP with speed N and rate function $\mathcal{I}_{\text{Ma}}(\cdot; t): \mathcal{M} \rightarrow [0, \infty]$, given by*

$$\begin{aligned} \mathcal{I}_{\text{Ma}}(\alpha; t) &= \inf_{\Lambda \in \mathcal{N}} I(\Lambda, \alpha; t) \\ &= I_{\text{Ma}}(\alpha; t) + (1 - c_\alpha) \left(\frac{t}{2} - \log t \right) + C_{\alpha, t} \left(\log(tC_{\alpha, t}) - \frac{t}{2} C_{\alpha, t} \right), \end{aligned} \quad (1.8)$$

where $C_{\alpha, t} = (1 - c_\alpha) \wedge \frac{1}{t}$ (recall $c_\alpha = \int_0^1 x \alpha(dx)$).

Again, only the second equality has to be checked; this is done in Lemma 4.2. In contrast with the result above, here the optimal configuration Λ^* depends on $\alpha \in \mathcal{M}$, most heavily it depends on whether $1 - c_\alpha \leq \frac{1}{t}$ or not. Indeed, if $1 - c_\alpha \leq \frac{1}{t}$, then $c_{\Lambda^*} = 1 - c_\alpha$ (and no mesoscopic part arises). However, if $1 - c_\alpha > \frac{1}{t}$, then $c_{\Lambda^*} = \frac{1}{t}$, and a non-trivial mesoscopic mass arises in the minimization; see Theorem 1.5. This peculiarity shows already a key difference between the cases $t \leq 1$ and $t > 1$. Indeed, if $t \leq 1$ one cannot have any macroscopic mass distribution α such that $1 - c_\alpha > \frac{1}{t}$ and no difference in the minimizing strategy of the system can be seen. This is a first way to see the phase transition at $t = 1$ from analytic properties of the rate function.

Now we come to the mesoscopic part of the particle configuration. This part comprises particle sizes on all the scales between finite and $O(N)$ and it seems unreasonable to consider an empirical measure for it. Instead, we consider only the total mass of this mesoscopic part. Let $\varepsilon > 0$ and $R \in \mathbb{N}$ be two auxiliary parameters, then we define the (R, ε) -mesoscopic total mass as

$$\overline{\text{Me}}_{R, \varepsilon}^{(N)} = \frac{1}{N} \sum_{i: R < S_i^{(N)} < \varepsilon N} S_i^{(N)}. \quad (1.9)$$

This is the number of vertices that are contained in components with a size between R and εN . The mesoscopic total mass in a strict sense arises after taking the limits $N \rightarrow \infty$, followed by $\varepsilon \downarrow 0$ and $R \rightarrow \infty$, but this does not define a proper random variable. However, it is possible to formulate an LDP in the $N \rightarrow \infty$ limit and then to study the rate function, $\mathcal{J}_{\text{Me}}^{(R, \varepsilon)}$, as $\varepsilon \downarrow 0$ and $R \rightarrow \infty$. Additionally, the proof of Theorem 1.1 shows that it is possible to define a coupled mesoscopic total mass $\overline{\text{Me}}_{R_N, \varepsilon_N}^{(N)}$, for any diverging sequence R_N and vanishing sequence ε_N . This is a well-defined random variable, it satisfies an LDP with speed N and the rate function is the limit of $\mathcal{J}_{\text{Me}}^{(R, \varepsilon)}$ when $\varepsilon \searrow 0$ and $R \nearrow \infty$.

Corollary 1.4 (LDP for mesoscopic mass). *(1) For any $R \in \mathbb{N}$ and $\varepsilon \in (0, 1)$, as $N \rightarrow \infty$, $\overline{\text{Me}}_{R, \varepsilon}^{(N)}$ satisfies an LDP with speed N and rate function $c \mapsto \mathcal{J}_{\text{Me}}^{(R, \varepsilon)}(c; t)$, where*

$$\mathcal{J}_{\text{Me}}^{(R, \varepsilon)}(c; t) = \inf \left\{ I(\Lambda, \alpha; t) : \sum_{k=1}^R k \lambda_k + \int_\varepsilon^1 x \alpha(dx) = 1 - c \right\}.$$

(2) For any $R_N \in \mathbb{N}$ and $\varepsilon_N \in (0, 1)$ such that $1 \ll R_N < \varepsilon_N N \ll N$, the coupled mesoscopic total mass $\overline{\text{Me}}_{R_N, \varepsilon_N}^{(N)}$ satisfies an LDP with speed N and rate function

$$\begin{aligned} \mathcal{J}_{\text{Me}}(c; t) &= \lim_{R \rightarrow \infty, \varepsilon \downarrow 0} \mathcal{J}_{\text{Me}}^{(R, \varepsilon)}(c; t) \\ &= (1 - c) \left(\log(1 - c)t - \frac{(1 - c)t}{2} \right) + \frac{t}{2} - \log t. \end{aligned} \quad (1.10)$$

The function $\mathcal{J}_{\text{Me}}(c; t)$ is strictly increasing in c , its minimum over $[0, 1]$ is $\mathcal{J}_{\text{Me}}(0; t) = 0$.

Corollary 1.4 part (1) is a simple consequence of the contraction principle, as the maps $\Lambda \mapsto \sum_{k=1}^R k \lambda_k$ and $\alpha \mapsto \int_{\varepsilon}^1 x \alpha(dx)$ are continuous. Assertion (2) follows as a byproduct of our proof of Theorem 1.1 in Section 3.

Hence, $\mathcal{J}_{\text{Me}}(\cdot; t)$ can rightfully be called the rate function for the mesoscopic total mass. Since it is positive everywhere outside 0, we have the immediate consequence that the probability that any positive percentage of the vertices lies in mesoscopic components decays exponentially towards zero. This implies the convergence in probability of $\overline{\text{Me}}_{R_N, \varepsilon_N}^{(N)}$ towards zero with exponential decay of the probability of a decay by any positive amount. Interestingly, taking $R_N + 1 = \varepsilon_N N \in \mathbb{N}$, we see that already just one mesoscopic particle alone satisfies the same LDP as the entire (R, ε) -mesoscopic total mass in the limit $R \rightarrow \infty$, $\varepsilon \downarrow 0$.

1.3. Our results: the phase transition in the light of the LDP. We now proceed with the study of the main phenomenon in the sparse Erdős–Rényi random graph: the phase transition of the emergence of a giant component. We will deduce it from our large-deviations rate functions from Section 1.2. The LDPs and the identification of their strict minimiser(s) lead to laws of large numbers for a number of random quantities. Indeed, it is a standard and simple fact from large-deviations theory that a random variable that satisfies an LDP with a rate function that contains precisely one minimizer converges in probability to that minimizer. We will exploit this fact to deduce laws of large numbers. As before, the parameter $t \in (0, \infty)$ will play the decisive role; recall that the connection probability $\frac{1}{N} t_N$ of the graph $\mathcal{G}(N, \frac{1}{N} t_N)$ was picked as $t_N = t + o(1)$ as $N \rightarrow \infty$.

Consider the following functions of the total masses of the microscopic and macroscopic particles respectively:

$$\mathcal{J}_{\text{Mi}}(c; t) = \inf_{\Lambda \in \mathcal{N}(c)} \mathcal{I}_{\text{Mi}}(\Lambda; t) \quad \text{and} \quad \mathcal{J}_{\text{Ma}}(c; t) = \inf_{\alpha \in \mathcal{M}_{\mathbb{N}_0}((0,1];c)} \mathcal{I}_{\text{Ma}}(\alpha; t), \quad (1.11)$$

where $c \in [0, 1]$. These two functions are not entirely analogous to $\mathcal{J}_{\text{Me}}(c; t)$ as rate functions for the total masses of the micro and the macro part, because the total masses both of $\text{Mi}^{(N)}$ and $\text{Ma}^{(N)}$ are equal to one. This is consistent with the fact that the contraction principle cannot be applied to total masses, as they are not continuous functions of the measures. However, they contain rather interesting information about the phase transition.

Theorem 1.5 (Microscopic total mass phase transition). *(1) For any $c \in [0, 1]$,*

$$\mathcal{J}_{\text{Mi}}(c; t) = tc + (1 - c) \log \frac{1 - c}{1 - e^{t(c-1)}} + \begin{cases} c \log c - tc^2 & \text{for } c < \frac{1}{t}, \\ -\frac{1}{2t} - \frac{t}{2}c^2 - c \log t & \text{for } c \geq \frac{1}{t}. \end{cases} \quad (1.12)$$

Moreover, $\mathcal{J}_{\text{Mi}}(c; t) = \mathcal{J}_{\text{Ma}}(1 - c; t)$.

(2) For $c \in (0, 1]$, the minimum of $\mathcal{N}(c) \ni \Lambda \mapsto \mathcal{I}_{\text{Mi}}(\Lambda; t)$ is attained precisely at $\Lambda^(c; t) \in \mathcal{N}(c)$ given by*

$$\lambda_k^*(c; t) = \frac{k^{k-2} c^k t^{k-1} e^{-ctk}}{k!}, \quad k \in \mathbb{N}, \quad (1.13)$$

and the minimum of the function $c \mapsto \mathcal{J}_{\text{Mi}}(c; t)$ is attained precisely at $c = 1$ with value $\mathcal{J}_{\text{Mi}}(1; t) = 0$. Therefore the infimum

$$\inf_{(\Lambda, \alpha) \in \mathcal{N} \times \mathcal{M}} I(\Lambda, \alpha; t) \quad (1.14)$$

is attained at $(\Lambda, \alpha) = (\Lambda^*(1; t), \mathbf{0})$, where $\mathbf{0} = (0, 0, \dots)$.

(3) For $t \in (1, \infty)$, the minimum of the function $c \mapsto \mathcal{J}_{\text{Mi}}(c; t)$ is attained at $c = \beta_t$ where $\beta_t \in (0, t)$ is the smallest positive solution to

$$\log \beta_t = t\beta_t - t. \quad (1.15)$$

The infimum in (1.14) is attained precisely at $(\Lambda, \alpha) = (\Lambda^*(\beta_t; t), (1 - \beta_t, 0, 0, \dots))$.

The proof is found in Section 4.2.

The two different cases in (1.12) refer to the cases that the first minimum in (1.11) is attained or not. Indeed, for $c \leq \frac{1}{t}$, the function $\mathcal{I}_{\text{Mi}}(\cdot; t)$ is minimized in an optimal Λ^* with $c_{\Lambda^*} = c$. However, for $c > \frac{1}{t}$ this is not possible, but only minimizing sequences can be found that achieve a total mass of $\frac{1}{t}$ in the microscopic measure and displace the remaining mass $c - \frac{1}{t}$ to the mesoscopic part. This shows that the phase transition originates from the impossibility of picking an optimal microscopic configuration Λ^* if its total mass c_{Λ^*} is required to be too large; the threshold being $1/t$. If this is exceeded, then a minimization can be done only with the help of some non-trivial mesoscopic part. As we mentioned above, the macroscopic configuration is always minimized in one single giant component.

The same effect is seen in $\mathcal{I}_{\text{Ma}}(c; t)$, where first an optimization over Λ with $c_{\Lambda} \leq 1 - c$ is performed, and such a balance between microscopic and mesoscopic mass can pop out if $1 - c$ is large enough. Subsequently, optimizing over $\mathcal{M}_{\mathbb{N}_0}((0, 1]; c)$ is straightforward. The equality $\mathcal{J}_{\text{Mi}}(c; t) = \mathcal{J}_{\text{Ma}}(1 - c; t)$ follows from this.

In Theorem 1.5(2) and (3) we see the different behaviour for subcritical, respectively supercritical t in terms of the microscopic configuration. Note that this configuration is actually given by

$$k\lambda_k^*(c; t) = c\text{Bo}_{ct}(k),$$

where Bo_{μ} is the Borel distribution with parameter $\mu \in [0, 1]$. We see that such an optimal $\Lambda^*(c; t)$ cannot be found if $c > \frac{1}{t}$, and this is an admissible total mass only when $t > 1$, marking the threshold between subcritical and supercritical regime (otherwise, there is no relevant case distinction as to the value of c). In this way, the Borel distribution appears as the natural minimizer of the microscopic part of the rate function.

In earlier work (see [Pit90]), the appearance of the Borel distribution in this context came from the observation that Bo_{μ} is the distribution of the total progeny of a Galton–Watson tree with offspring that is Poisson-distributed with parameter μ . The characterisation of the emerging cluster-size distribution Λ^* was based on an approximation of the connected subgraphs by such trees and counting the total number of trees of a given size in the graph. This approximation argument was extended to a large-deviation setting in [BC15], see Section 1.4.

Theorem 1.5 characterises the well-known phase transition of the emergence of a giant component at $t = 1$ in terms of a natural notion that is familiar to statistical physics: as a non-analyticity of the limiting free energy for the total mass of the microscopic configuration, which is equal to the infimum of $\mathcal{J}_{\text{Mi}}(\cdot; t)$. Indeed, this function is zero in $[0, 1]$, but positive in $(1, \infty)$.

Another characterisation of this phase transition is in terms of a law of large numbers. Indeed, combining Theorem 1.5 with the LDP in Theorem 1.1 one has

$$(\text{Mi}^{(N)}, \text{Ma}^{(N)}) \xrightarrow{N \rightarrow \infty} \begin{cases} (\Lambda^*(1; t), \mathbf{0}) & \text{if } t \leq 1, \\ (\Lambda^*(\beta_t; t), (1 - \beta_t, 0, \dots)) & \text{if } t \geq 1. \end{cases}$$

In words, this means that, for any $k \in \mathbb{N}$, $\frac{1}{N}$ times the number of components of size k converges to $\lambda_k^*(1; t)$ in the sub-critical regime and to $\lambda_k^*(\beta_t, t)$ in the supercritical regime, while there is no macroscopic component in the first regime and there is precisely one macroscopic cluster of cardinality $\sim N(1 - \beta_t)$ in the second. All these statements are in the sense of convergence in probability, and the probability of a deviation by any positive amount decays even exponentially in N .

One also sees that the cut-off versions of the total masses, $\sum_{k=1}^R k \text{Mi}_k^{(N)}$ and $\int_{[\varepsilon, 1]} x \text{Ma}^{(N)}(dx)$, converge towards the respective cut-off versions of the limits, and their limits as $R \rightarrow \infty$ and $\varepsilon \downarrow 0$ are $(1, 0)$ for $t \leq 1$ and $(\beta_t, 1 - \beta_t)$ for $t \geq 1$.

1.4. Related works on LDPs for Erdős-Rényi graphs. Despite the extensive literature on the Erdős-Rényi graph, there are not many results about large deviations in the sparse regime. Here we summarize, to the best of our knowledge, the existing results and how they relate to our work.

Two LDPs for the size of the largest component and for the number of isolated vertices have been derived in [O'C98]. These are two quantities that are obviously functionals of our measures $\text{Ma}^{(N)}$, respectively of $\text{Mi}^{(N)}$. Indeed, the largest component is the mass of the largest atom of $\text{Ma}^{(N)}$, and the number of isolated vertices is equal to N times $\text{Mi}_1^{(N)}$. Both these two functionals are continuous, such that the contraction principle applies. The approach of [O'C98] is a simplified version of our comprehensive approach for the joint distribution of all the component sizes, and consequently it leads to formulas for the rate functions that are contractions of our rate function, which is straightforward to see. This explains also the remark made in [O'C98] about the lack of convexity of the rate function, since contracted rate functions are rarely convex. Hence, our work includes the results of [O'C98].

A route that is inspired by statistical physics is taken in [EMH04], where the distribution of the random graph is tilted with a parameter $q > 1$ raised to the power of the number of components, properly normalized. The analysis of the free energy of the corresponding partition sum is carried out there. Via the well-known Laplace dualism, the results is essentially equivalent to an LDP for the number of components. This functional is equal to the continuous functional $\text{Mi}^{(N)}(\mathbb{N})$ in our setting. The peculiarity of [EMH04] is that this model is put into relation with the q -state Potts model in the limit $q \downarrow 1$ via diagrammatic expansion techniques. In particular, they derive limiting formulas for the size of the giant component, the degree distributions inside and outside the giant component, and the distribution of small component sizes.

We already mentioned that registering each component as a *subgraph* (rather than only as its size) would give *a priori* a much more detailed description, at least for any fixed N . However, in the limit as $N \rightarrow \infty$, in the LDP regime that we consider, only few subgraph configurations survive: the microscopic components survive only as spanning trees, and only those macroscopic components survive that have an excess of edges of order $\Theta(N)$. The first has been carried out in [BC15], the second in [Puh05].

Indeed, the macroscopic part of our LDP is covered in [Puh05]. The author gives an LDP for the joint distribution of the total number of components, the sequence of the sizes of macroscopic ones, and the sequence of corresponding numbers of the excess edges appended with zeros. Therefore the contraction of this LDP to the total number of components and the macroscopic sizes (see [Puh05, Corollary 2.1]) is equal to the contraction of our LDP from Theorem 1.1 to the LDP for $(\sum_k \text{Mi}_k^{(N)}, \text{Ma}^{(N)})$. The same is of course true, when considering exclusively macroscopic sizes (compare [Puh05, Corollary 2.2] with Corollary 1.3). The approach in [Puh05] goes along a very different route, involving recursive formulas for the graphs $\mathcal{G}(N, \frac{1}{N}t_N)$ if N increases, and consequently the form of the rate function derived there is pretty different from ours; it involves an additional minimization procedure. It would require some work to analytically check that it is identical to ours.

In [BC15], an LDP for the empirical measure of all the components rooted at the vertices, is derived with an explicit rate function. The topology used there comes from a distance that looks only at intersections of graphs with bounded sets, so it can detect only microscopic components. In this way it is contained in our results. However, [BC15] considers the components as graphs, not only as sizes, and gets therefore a much more detailed picture. Moreover, the object described in [BC15] is a size-biased version of our microscopic measure, since we are counting components of a certain size, while they consider the component containing each vertex and therefore counting a certain component proportionally to the number of vertices it contains. Hence, the LDP of [BC15] contains the microscopic part of our LDP (Corollary 1.2) via the contraction principle, but a certain normalization has to be performed to actually compare the two objects. However, let us notice that [BC15, Theorem 1.8] shows that the rate function takes the form of a sort of relative entropy with respect to a Galton Watson tree with Poisson offspring distribution (plus additional constants). This form is shown also in our contracted rate function from Corollary 1.2, which in (4.2) we rewrite in terms of a relative entropy with respect to a distribution related to the Borel distribution (which is the distribution of the total progeny of precisely a Galton Watson tree with Poisson offspring). Also in this case, one sees that the Borel distribution appears as a size-biased version of our reference distribution.

Under assumptions that imply that the connection probability of the Erdős–Rényi graph $\mathcal{G}(N, p)$ satisfies $p \gg N^{-\frac{1}{2}}$, recent progress has been made on the upper tails of sub-graph counts [CD16, Aug18, CD18].

In the case of dense graphs, that is, for $\mathcal{G}(N, p)$ with fixed $p \in (0, 1)$, there is a complete treatment thanks to Chatterjee and Varadhan [CV11], see [Cha16] for an overview. This regime is rather different from the sparse regime, since a proper formulation of the relevant limiting objects requires a abstract setting evolving around the notion of a graphon.

1.5. Application to coagulation models. Our interest in this research came from the desire to understand dynamical particle systems with coagulation in the large-system limit. It turned out that one of the most prominent (and most simple) models, the *Marcus–Lushnikov model of coagulation*, see [Mar68, Gil72, Lus78], admits a one-to-one correspondence to the component sizes of the Erdős–Rényi random graph that we study in this paper. This coagulation process is a continuous-time Markov process of vectors of particle masses $(S_i^{(N)}(t))_{i \in \{1, \dots, n_t\}} \in (\mathbb{N}_0)^{n_t}$ at time $t \in [0, \infty)$, arranged in descending order, precisely as in (1.1), where n_t is the number of particles at time t . This process is specified by the initial configuration, which we take in the monodisperse case, i.e., $S_i^{(N)}(0) = 1$ for all $i = 1, \dots, N = n_0$, and by the transition mechanism,

which is given in terms of a symmetric, non-negative *coagulation kernel* $K_N: \mathbb{N} \times \mathbb{N} \rightarrow [0, \infty)$. That is, we start with N particles of unit mass at time 0, and in the course of the process, each (unordered) pair of particles with respective masses $m, \tilde{m} \in \mathbb{N}$ coagulate to a particle of mass $m + \tilde{m}$ with rate $K_N(m, \tilde{m})$, independently of all the other pairs of particles.

The important special case of the *multiplicative kernel*, $K_N(m, \tilde{m}) = m\tilde{m}/N$ has the two interesting features: (1) it can be mapped onto the Erdős-Rényi random graph that we study in this paper, and (2) it exhibits a-n interesting *gelation phase transition* in the limit $N \rightarrow \infty$ at time $t = 1$, because a *gel*, i.e., a particle of macroscopic size, appears. Indeed, for this process, it turned out in the review [Ald99] that, for fixed $t \in [0, \infty)$, the distribution of the family $(S_i^{(N)})_{i \in \{1, \dots, n\}}$ of the component sizes of $\mathcal{G}(N, \frac{1}{N}t_N)$ defined in (1.1) with $\frac{1}{N}t_N = 1 - e^{-t/N}$ is identical to the family $(S_i^{(N)}(t))_{i \in \{1, \dots, n_t\}}$ of particle masses in the multiplicative coalescent at time t . This correspondence was not mentioned in [BP90], but was discussed one year later in [BP91], which highlights the connection between gelation in the coagulation process and the phase transition given by the formation of a giant connected component in the Erdős-Rényi random graph [ER60].

Hence, the results of this paper also recover the gelation phenomenon in rather explicit terms through a LDP in terms of the microscopic, mesoscopic and macroscopic parts, in the same way as we explained in the above sections for the Erdős-Rényi random graph.

Smoluchowski introduced a (deterministic) ODE model for the concentrations of coagulating particles in the course of his work on Brownian motion [vS16]. Indeed, it is reasonable to assume that $l_k(t) = \lim_{N \rightarrow \infty} \frac{1}{N} \#\{\text{particles of size } k \text{ at time } t\}$ exists under suitable conditions, see [Nor99, LM04, MN14]. These limits satisfy

$$\frac{d}{dt} l_k(t) = \frac{1}{2} \sum_{\substack{m, \tilde{m}: \\ m + \tilde{m} = k}} l_m(t) l_{\tilde{m}}(t) K(m, \tilde{m}) - l_k(t) \sum_m l_m(t) K(k, m), \quad k \in \mathbb{N}, \quad (1.16)$$

where $K(m, \tilde{m}) = \lim_{N \rightarrow \infty} N K_N(m, \tilde{m})$ is the limiting coagulation kernel (in our case, $K(m, \tilde{m}) = m\tilde{m}$). This is the famous *Smoluchowski equation*. Intuitively, the positive terms on the right-hand side of (1.16) take into account that the fraction of particles of mass k increases if a particle of mass m and one of mass \tilde{m} (with $m + \tilde{m} = k$) merge and this happens with rate $K(m, \tilde{m})$. On the other hand, the negative term describes that a particle of mass k can coagulate with particles of any size m with rate $K(k, m)$ and this is why it involves an infinite sum (over all $m \in \mathbb{N}$). One can check for $t \leq 1$ that $\Lambda^*(1; t)$ appearing in Theorem 1.5 is the exact solution of (1.16), the Smoluchowski equation, also given in [Ald99, Table 2]. As a consequence, the above mentioned gelation phase transition as well as the solution of the Smoluchowski ODE are also clear from our results in Sections 1.2 and 1.3 and receive therefore a new interpretation in terms of combinatorial structures.

In the light of the process character of the Marcus-Lushnikov coagulation model, it will be desirable to derive a pathwise version of the LDP of Theorem 1.1. This will require a version of that theorem with start from an arbitrary configuration rather than from $S_i^{(N)}(0) = 1$. This may also be interesting for the time-dependent version of the Erdős-Rényi graph $(\mathcal{G}(N, 1 - e^{t/N}))_{t \in [0, \infty)}$, but not as natural as for the Marcus-Lushnikov model. Another aspect that makes it particular interesting for coagulation models is the availability of alternative methods in the spirit of Wentzell-Freidlin theory to derive pathwise LDPs for coagulation models, see [MPPR17]. Let us also mention that, in the renowned paper [Ald97], time is

expanded around the critical value $t = 1$, and the mesoscopic components of the graph are compared to a stochastic process known as the *multiplicative coalescent*. Although we allow for fluctuations around t (our LDP holds for any sequence $t_N \sim t$), we cannot capture this regime around $t = 1$, since we expect a LDP for mesoscopic particles to hold on a different scale. Another natural direction of our future research is an extension to the case of an *inhomogeneous* Erdős–Rényi graph as introduced in [BJR07]. We will defer future work to these questions.

1.6. Comparison to Bose–Einstein condensation without interaction. Our large-deviations approach to the Marcus–Lushnikov models shows remarkable similarities to another well-known phase transition in a non-spatial model, the non-interacting Bose gas. Here the situation is similar in that the gas can be conceived as a joint distribution of N particles that are randomly grouped into smaller units, called *cycles*, which can become arbitrarily large. The natural question is then, under what circumstances do macroscopic cycles arise. An explicit answer in terms of a large-deviations analysis has been given in [Ada08], where the transition, the famous *Bose–Einstein Condensation (BEC)* in dimensions $d \geq 3$, is derived from the minimization of the rate function, in a way analogous to that in our Theorem 1.5. The two phase transitions differ in that the BEC transition is of *saturation type*, while the gelation transition is not.

For the non-interacting Bose gas in the thermodynamic limit at temperature $1/\beta \in (0, \infty)$ with particle density $\rho \in (0, \infty)$ the partition function is given by

$$Z_{\Lambda_N}^{(\beta)} = \sum_{(\ell_k)_{k \in \mathbb{N}} \in \mathbb{N}_0^{\mathbb{N}}: \sum_k k \ell_k = N} \prod_k \frac{N^{\ell_k}}{\ell_k! k^{\ell_k}} [\rho(4\pi\beta k)^{\frac{d}{2}}]^{-\ell_k},$$

where Λ_N is the centred box in \mathbb{R}^d with volume N/ρ . The free energy per particle is then

$$f(\beta, \rho) = \lim_{N \rightarrow \infty} \frac{1}{N} \log Z_{\Lambda_N}^{(\beta)} = - \inf_{\Lambda \in \mathcal{N}(\rho)} I(\Lambda), \quad \text{where} \quad I(\Lambda) = \sum_k \lambda_k \log \frac{\lambda_k k}{(4\pi\beta k)^{\frac{d}{2}}}.$$

For the Erdős–Rényi graph $\mathcal{G}(N, \frac{1}{N}t_N)$, the equivalent quantity is the rate function \mathcal{I}_{Mi} from (1.7). The key difference between the rate functions is that only \mathcal{I}_{Mi} contains terms in the total mass of microscopic components, c_Λ . This reflects the fact that the giant component makes a significant contribution to the rate function in the graph model, but the condensate in the non-interacting Bose gas does not.

The respective minimisers of \mathcal{I}_{Mi} and I are

$$k\lambda_k^{(\text{ML})}(c; t) = \frac{1}{t} \frac{(cte^{-ct})^k}{k^{1-k} k!} \sim \frac{1}{\sqrt{2\pi t}} \frac{(cte^{-ct+1})^k}{k^{3/2}} \quad \text{and} \quad k\lambda_k^{(\text{BEC})}(\alpha; \beta) = \frac{1}{\rho(4\pi\beta)^{\frac{d}{2}}} \frac{e^{-\alpha k}}{k^{\frac{d}{2}}},$$

where c and α control the values of $\sum_k k\lambda_k$.

The crucial parameters are t for the graph model and the inverse temperature β for the Bose gas. Both models have a trivial upper bound for the total microscopic mass, $\sum_k k\lambda_k$, namely one. One additional upper bound arises in each model from the optimisation of the rate function with respect to the λ_k , but these are not relevant, until t respectively β rises to its critical value. For the graph model this bound is $1/t$, because $\sum_k \frac{(cte^{-ct})^k}{k^{1-k} k!} \leq 1$ for all $ct \in (0, \infty)$, and the summands take their maxima at $ct = 1$, when they correspond to the Borel probability distribution with parameter 1. For $\Lambda^{(\text{BEC})}$ this bound is $\rho^{-1}(4\pi\beta)^{-d/2} \sum_k k^{-\frac{d}{2}}$. At this point we see a difference between the two models, because the total microscopic mass in the Bose

gas remains on this bound as β rises further, while for the graph model it immediately drops strictly below the bound. This explains why BEC is known as a saturation phase transition, but this description cannot be applied to gelation.

2. PREPARATIONS FOR THE PROOF OF THE LDP

We consider the Erdős–Rényi graph $\mathcal{G} = \mathcal{G}(N, p)$ under the corresponding probability measure $\mathbb{P}_{N,p}$. In Section 2.1, we derive, an explicit formula for the distribution of the empirical measure of the component sizes $S_i^{(N)}$ in terms of connectivity probabilities for (smaller) Erdős–Rényi random graphs. Furthermore, we prepare in Section 2.2 for the asymptotic analysis for $p = \frac{1}{N}t_N$ with $t_N = t + o(1)$ by recalling from [Ste70] some estimates and asymptotics for this connectivity probability.

2.1. The joint distribution of the component sizes. An important quantity is

$$\mu_k(p) = \mathbb{P}_{k,p}(\mathcal{G} \text{ is connected}), \quad k \in \mathbb{N}, p \in [0, 1]. \quad (2.1)$$

We will be concerned with this quantity for fixed k , but with connection probability $p = \frac{1}{N}t_N$, in the limit $N \rightarrow \infty$.

We define the state space of the collection of component sizes as

$$E_N = \left\{ (s_i)_{i \in \{1, \dots, n\}} \in \mathbb{N}_0^n : n \in \mathbb{N}, \sum_{i=1}^n s_i = N \right\}. \quad (2.2)$$

To each element $(s_i)_i$ of the state space E_N , we associate a unique element of the space

$$\mathcal{N}_N = \left\{ \ell = (\ell_k)_k \in \mathbb{N}_0^{\mathbb{N}} : \sum_k k \ell_k = N \right\}, \quad (2.3)$$

where for each k , ℓ_k is the number of indices i such that $s_i = k$. The map $(s_i)_i \mapsto \ell$ is a bijection and in the following we refer to configurations equally in terms of $(s_i)_i$ or ℓ .

We denote by \mathcal{P}_N the set of all partitions of $[N] = \{1, \dots, N\}$. We write $B_k(\pi)$ for the number of sets in $\pi \in \mathcal{P}_N$ with cardinality k . Then we can describe the joint distribution of all the component sizes of $\mathcal{G}(N, p)$ as follows.

Lemma 2.1. *For any $p \in [0, 1]$, $N \in \mathbb{N}$ and every $(s_i)_i \in E_N$,*

$$\mathbb{P}_{N,p}((S_i^{(N)})_i = (s_i)_i) = \#\{\pi \in \mathcal{P}_N : B_k(\pi) = \ell_k \forall k\} \times \left(\prod_i \mu_{s_i}(p) \right) \times \left(\prod_{i \neq j} (1-p)^{\frac{1}{2}m_i m_j} \right). \quad (2.4)$$

Proof. A set $A \subset \{1, \dots, N\}$ of vertices is a connected component in the graph $\mathcal{G}(N, p)$ if and only if (1) no bond between any vertex in A and any vertex outside has been put, and (2) the subgraph formed out of the vertices in A and all the bonds between any two vertices in A is connected. This has probability $(1-p)^{|A||A^c|} \times \mu_{|A|}(p)$. Applying this reasoning to A^c and describing the next component, and iterating this argument, shows that the product of the two products on the right-hand side of (2.4) is equal to the probability, for a given partition π with ℓ_k sets of size k for any k , that the components of $\mathcal{G}(N, p)$ are precisely the sets of π . Since this probability depends only on the cardinalities, the counting term completes the formula. \square

Now we rewrite the right-hand side of (2.4) in terms of the empirical measure of $(s_i)_i$, i.e., of the numbers ℓ_k of indices i such that $s_i = k$. Introduce the event

$$A_N(\ell) = \bigcap_{k \in \mathbb{N}} \{\#\{i: S_i^{(N)} = k\} = \ell_k\}, \quad \ell = (\ell_k)_{k \in \mathbb{N}} \in \mathbb{N}_0^{\mathbb{N}}. \quad (2.5)$$

Corollary 2.2. *For any $p \in [0, 1]$, N and any $\ell = (\ell_k)_k \in \mathbb{N}_0^{\mathbb{N}}$ satisfying $\sum_k k \ell_k = N$,*

$$\mathbb{P}_{N,p}(A_N(\ell)) = N! \prod_k \frac{\mu_k(p)^{\ell_k} (1-p)^{\frac{1}{2}k(N-k)\ell_k}}{k!^{\ell_k} \ell_k!}. \quad (2.6)$$

Proof. Note that the last product on the right-hand side of (2.4) can also be written as $\prod_i (1-p)^{\frac{N}{2}m_i(N-m_i)}$. Hence, if ℓ_k is equal to the number of i such that $s_i = k$ for any k , then the product of the last two products can be written as

$$\prod_k \left(\mu_k(p) \right)^{\ell_k} (1-p)^{\frac{1}{2}k(N-k)\ell_k}.$$

The counting term is easily identified as

$$\#\{\pi \in \mathcal{P}_N: \#\{A \in \pi: |A| = k\} = \ell_k \forall k\} = \frac{N!}{\prod_k (k!^{\ell_k} \ell_k!)}.$$

Substituting ends the proof. \square

(To avoid confusion, we note that there is a typographical error in Section 4.5 of [Ald99], where the factor of $\frac{1}{2}$ is missing in the exponent of (2.6).)

2.2. The probability of being connected. Our analysis of (2.6) will depend crucially on an analysis of $\mu_k(\frac{1}{N}t_N)$. The next two lemmas collect results from [Ste70, Lemma 1&2, Theorem 1].

Lemma 2.3 (Bounds and asymptotics for $\mu_k(\frac{1}{N}t_N)$, [Ste70]). *For any $p \in [0, 1]$, $N \in \mathbb{N}$ and any $k \leq N$,*

$$(1-p)^{\frac{1}{2}(k-1)(k-2)} \leq \frac{\mu_k(p)}{k^{k-2} p^{k-1}} \leq 1. \quad (2.7)$$

In particular, if $p = \frac{1}{N}t_N$ with $t_N = t + o(1)$ and $k = o(\sqrt{N})$,

$$\mu_k(\frac{1}{N}t_N) = k^{k-2} \left(\frac{t}{N} \right)^{k-1} (1 + o(1)), \quad N \rightarrow \infty.$$

The expression for the upper bound in (2.7) appears to be present (using somewhat applied chemical language) in [Flo41, equation (5)].

The following is an alternative upper bound for $\mu_k(\frac{1}{N}t_N)$, which will be required for macroscopic components, together with an asymptotic result for the connection probability in the so-called sparse case, where the bond probability is proportional to the inverse of the size of the graph.

Lemma 2.4 ([Ste70]). *For all $p \in [0, 1]$ and $k \in \mathbb{N}$*

$$\mu_k(p) \leq (1 - e^{-kp})^{k-1},$$

with $q = \log(1 - p)$. Moreover, for $\alpha \in (0, 1)$ and $t \in (0, \infty)$ and a sequence $t_N = t + o(1)$, as $N \rightarrow \infty$,

$$\mu_{\lfloor \alpha N \rfloor}(\frac{1}{N}t_N) = \left(1 - \frac{\alpha t}{e^{\alpha t} - 1}\right) (1 - e^{-t\alpha})^{\alpha N} (1 + o(1)). \quad (2.8)$$

3. PROOF OF THE LDP

In this section we prove the main result of this paper, the large-deviations principle in Theorem 1.1. Again, we fix the parameter $t \in (0, \infty)$ and a sequence $t_N = t + o(1)$ and consider the Erdős–Rényi graph $\mathcal{G}(N, \frac{1}{N}t_N)$ with probability measure $\mathbb{P}_{N, \frac{1}{N}t_N}$.

Recall the topological remarks on the two state spaces \mathcal{N} and \mathcal{M} from Section 1.1. The metrics d on \mathcal{N} and D on \mathcal{M} , defined by

$$d(\Lambda, \tilde{\Lambda}) = \sum_{k=1}^{\infty} 2^{-k} |\lambda_k - \tilde{\lambda}_k| \quad \text{and} \quad D(\alpha, \tilde{\alpha}) = \sum_{i=1}^{\infty} 2^{-i} |\alpha_i - \tilde{\alpha}_i|, \quad (3.1)$$

induce the respective topologies of pointwise and vague convergence. We write $B_\delta(\Lambda)$, respectively $B_\rho(\alpha)$, for the closed δ -ball around Λ , respectively for the closed ρ -ball around α . Since the rate function $I(\cdot; t)$ is lower semicontinuous in $\mathcal{N} \times \mathcal{M}$ and the space is compact, we know that it is a good rate function (i.e., its level sets are not only closed but also compact). Therefore, a weak LDP implies our main result, the LDP in Theorem 1.1, and it will be sufficient to prove the following.

Proposition 3.1. *For any $(\Lambda, \alpha) \in \mathcal{N} \times \mathcal{M}$,*

$$\lim_{\delta, \rho \downarrow 0} \lim_{N \rightarrow \infty} \frac{1}{N} \log \mathbb{P}_{N, \frac{1}{N}t_N}(\text{Mi}^{(N)} \in B_\delta(\Lambda), \text{Ma}^{(N)} \in B_\rho(\alpha)) = -I(\Lambda, \alpha; t). \quad (3.2)$$

We split the proof of Proposition 3.1 in several lemmas and finish it at the end of Section 3. We start by bounding the cardinality of \mathcal{N}_N .

Lemma 3.2. *Let \mathcal{N}_N be as defined in (2.3), then*

$$|\mathcal{N}_N| = e^{o(N)}, \quad N \rightarrow \infty.$$

Proof. The following is an argument in [Ada08]. For any $\ell \in \mathcal{N}_N$, the set $H(\ell) = \{k \in \mathbb{N} : \ell_k > 0\}$ has no more than $2\sqrt{N}$ elements, since

$$N = \sum_{k \in H(\ell)} k \ell_k \geq \sum_{k \in H(\ell)} k \geq \sum_{k=1}^{|H(\ell)|} k = |H(\ell)| \frac{1}{2} (|H(\ell)| + 1).$$

Hence,

$$\begin{aligned} |\mathcal{N}_N| &= \left| \left\{ (\ell_k)_k \in \mathbb{N}_0^{\mathbb{N}} : \sum_k k \ell_k = N, |H(\ell)| \leq 2\sqrt{N} \right\} \right| \\ &\leq \sum_{H \subset [N] : |H| \leq 2\sqrt{N}} \left| \left\{ (\ell_k)_{k \in H} \in \mathbb{N}^{|H|} : \sum_{k \in H} k \ell_k = N \right\} \right| \\ &\leq \sum_{H \subset [N] : |H| \leq 2\sqrt{N}} \left| \left\{ (L_k)_{k \in H} \in \mathbb{N}^{|H|} : \sum_{k \in H} L_k = N \right\} \right| \leq \sum_{h=1}^{\lfloor 2\sqrt{N} \rfloor} \binom{N}{h} \binom{N+h}{h} \\ &= e^{o(N)}. \end{aligned}$$

□

Thanks to Lemma 3.2, it will be sufficient to get estimates on $\mathbb{P}_{N, \frac{1}{N}t_N}(A_N(\ell))$, for any $\ell \in \mathcal{N}_N$ close enough to a fixed $(\Lambda, \alpha) \in \mathcal{N} \times \mathcal{M}$. The strategy is to divide the terms in the product representation from Corollary 2.2 into three groups, which we refer to as micro-, meso- and macroscopic, because they take into account the contribution of, respectively, micro-, meso- and macro- components. We fix two increasing sequences R_N and $\varepsilon_N N$ in \mathbb{N} such that $R_N \nearrow \infty$, $\varepsilon_N \downarrow 0$ and $R_N < \varepsilon_N N$. We write

$$\mathbb{P}_{N, \frac{1}{N}t_N}(A_N(\ell)) = N! \times F_{\text{Mi}}(\ell) \times F_{\text{Me}}(\ell) \times F_{\text{Ma}}(\ell), \quad (3.3)$$

where

$$F_{\text{Mi}}(\ell) = \prod_{k=1}^{R_N} z_k(\ell), \quad F_{\text{Me}}(\ell) = \prod_{R_N < k \leq \varepsilon_N N} z_k(\ell), \quad F_{\text{Ma}}(\ell) = \prod_{\varepsilon_N N < k \leq N} z_k(\ell),$$

and

$$z_k(\ell) = \frac{\mu_k \left(\frac{1}{N}t_N\right)^{\ell_k} \left(1 - \frac{t_N}{N}\right)^{\frac{1}{2}k(N-k)\ell_k}}{k!^{\ell_k} \ell_k!}.$$

Let us set

$$c_{\text{Mi}}(\ell/N) = \frac{1}{N} \sum_{k=1}^{R_N} k \ell_k, \quad c_{\text{Me}}(\ell/N) = \frac{1}{N} \sum_{R_N < k \leq \varepsilon_N N} k \ell_k, \quad c_{\text{Ma}}(\ell/N) = \frac{1}{N} \sum_{\varepsilon_N N < k \leq N} k \ell_k. \quad (3.4)$$

Note that the sum of these three terms is equal to one. For the factor $N!$, we use Stirling's formula $N! = \left(\frac{N}{e}\right)^N e^{o(N)}$ so that uniformly in $\ell \in \mathcal{N}_N$

$$N! = \left(\frac{N}{e}\right)^{N c_{\text{Mi}}(\ell/N)} \left(\frac{N}{e}\right)^{N c_{\text{Me}}(\ell/N)} \left(\frac{N}{e}\right)^{N c_{\text{Ma}}(\ell/N)} e^{o(N)}, \quad N \rightarrow \infty. \quad (3.5)$$

We will consider a cut-off version of the distances d and D introduced in (3.1), as follows:

$$d_R(\Lambda, \tilde{\Lambda}) = \sum_{k=1}^R 2^{-k} |\lambda_k - \tilde{\lambda}_k| \quad \text{and} \quad D_\varepsilon(\alpha, \tilde{\alpha}) = \sum_{i=1}^{\infty} 2^{-i} |\alpha_i - \tilde{\alpha}_i| \mathbf{1}(\alpha_i \vee \tilde{\alpha}_i \geq \varepsilon N),$$

such that for a given $\ell \in \mathcal{N}_N$ we can measure simultaneously the distance of its microscopic part from Λ and its macroscopic part from α . We also introduce some new notation, for any $\ell \in \mathcal{N}_N$ we will use the notation $\frac{1}{N}\ell$ to denote the sequence $\left(\frac{\ell_k}{N}\right)_{k \in \mathbb{N}}$, which clearly denotes an element of \mathcal{N} . On the other hand, with $\ell_{[\cdot, N]}$ we denote the point measure on $(0, 1]$ with weight ℓ_k at the point $\frac{k}{N}$ for any $k = 1, \dots, N$ and zero everywhere else. This integer valued measure clearly belongs to \mathcal{M} .

Lemma 3.3. *Fix $(\Lambda, \alpha) \in \mathcal{N} \times \mathcal{M}$ such that $c_\Lambda + c_\alpha \leq 1$. Fix $\delta, \rho > 0$ and pick sequences $\ell \in \mathcal{N}_N$ and $R_N \rightarrow \infty$ and $\varepsilon_N \downarrow 0$ such that $d_{R_N}(\frac{1}{N}\ell, \Lambda) \leq \delta$ and $D_{\varepsilon_N}(\ell_{[\cdot, N]}, \alpha) \leq \rho$ for all N . Further assume that $|\frac{1}{\varepsilon_N} \log \varepsilon_N| \leq o(N)$. Then, for any $R \in \mathbb{N}$ and $\varepsilon > 0$,*

$$\begin{aligned} \limsup_{N \rightarrow \infty} \frac{1}{N} \log \mathbb{P}_{N, \frac{1}{N}t_N}(A_N(\ell)) &\leq -I_{\text{Mi}}(\Lambda; t) - I_{\text{Ma}}(\alpha; t) - \left(\frac{t}{2} - \log t\right)(1 - c_\Lambda - c_\alpha) \\ &\quad + K_{R, \varepsilon}(\delta, \rho), \end{aligned}$$

where $K_{R, \varepsilon}(\delta, \rho)$ vanishes as $\delta \downarrow 0$ and $\rho \downarrow 0$, followed by $R \rightarrow \infty$ and $\varepsilon \downarrow 0$.

Proof. We start by looking at the term $F_{\text{Mi}}(\ell)$, i.e., the microscopic term. For any fix $k \leq R_N$, we use the upper bound in (2.7), the fact that $1 - x \leq e^{-x}$ and Stirling's lower bound for $\ell_k!$ (notice that for k small we expect ℓ_k to be large, $\Theta(N)$) to obtain

$$z_k(\ell) \leq \frac{k^{(k-2)\ell_k} t_N^{(k-1)\ell_k} e^{-\frac{t_N}{2N} k(N-k)\ell_k}}{k!^{\ell_k} N^{(k-1)\ell_k} \left(\frac{1}{e}\ell_k\right)^{\ell_k}}.$$

Now we combine $F_{\text{Mi}}(\ell)$ with the first term in (3.5) and obtain, uniformly for $\ell \in \mathcal{N}_N$, also using that $\sum_{k=1}^{R_N} \frac{t}{2N} k^2 \ell_k \leq \frac{t}{2} R_N c_{\text{Mi}}(\ell/N)$,

$$\begin{aligned} \left(\frac{N}{e}\right)^{N c_{\text{Mi}}(\ell/N)} F_{\text{Mi}}(\ell) &\leq \exp\left(-N \sum_{k=1}^{R_N} \frac{1}{N} \ell_k \log \frac{k! e^k \frac{1}{N} \ell_k}{k^{k-2} t^{k-1} e^{1-\frac{t}{2}k}}\right) e^{o(N)} \\ &= \exp\left(-N I_{\text{Mi}}^{(R_N)}\left(\frac{1}{N}\ell; t\right)\right) e^{o(N)}, \end{aligned}$$

where

$$I_{\text{Mi}}^{(R_N)}(\tilde{\Lambda}; t) = f^{(R_N)}(\tilde{\Lambda}; t) + \sum_{k=1}^{R_N} k \tilde{\lambda}_k \left(\frac{t}{2} - \log t\right) \quad \text{with } f^{(R_N)}(\tilde{\Lambda}; t) := \sum_{k=1}^{R_N} \tilde{\lambda}_k \log \frac{k! t e^{k-1} \tilde{\lambda}_k}{k^{k-2}}, \quad (3.6)$$

is the cut-off version of the rate function defined in (1.5). Recall that $d_{R_N}(\frac{1}{N}\ell, \Lambda) < \delta$ and that $c_\Lambda = \sum_{k \in \mathbb{N}} k \lambda_k \in [0, 1]$ and observe that $\lim_{R \rightarrow \infty} I_{\text{Mi}}^{(R)}(\Lambda; t) = I_{\text{Mi}}(\Lambda; t)$. Therefore we claim that, for any fixed $R \in \mathbb{N}$, as $N \rightarrow \infty$,

$$\left(\frac{N}{e}\right)^{N c_{\text{Mi}}(\ell/N)} F_{\text{Mi}}(\ell) \leq \exp(-N I_{\text{Mi}}(\Lambda; t)) e^{N(C_R(\delta) + \gamma_R) + o(N)} e^{-N(\frac{t}{2} - \log t)(c_{\text{Mi}}(\ell/N) - c_\Lambda)}, \quad (3.7)$$

where $\lim_{R \rightarrow \infty} \gamma_R = 0$ and $\lim_{\delta \downarrow 0} C_R(\delta) = 0$. To prove (3.7), we notice that $f^{(R)}(\cdot; t)$ is continuous, it is clear that $\sup_{\ell: d_R(\frac{1}{N}\ell, \Lambda) < \delta} |f^{(R)}(\frac{1}{N}\ell; t) - f^{(R)}(\Lambda; t)|$ vanishes as $\delta \downarrow 0$ and can therefore be

estimated against such a $C_R(\delta)$. Moreover, we estimate (substituting $\frac{1}{N}\ell$ by $\tilde{\Lambda}$), for any N such that $R_N > R$, with the help of the Stirling bound $k! e^k k^{-k} \geq 1$ and Jensen's inequality for $\varphi(x) = x \log x$, as follows:

$$\begin{aligned} f^{(R_N)}(\tilde{\Lambda}; t) - f^{(R)}(\tilde{\Lambda}; t) &= \sum_{k=R+1}^{R_N} \tilde{\lambda}_k \log \frac{k! t e^{k-1} \tilde{\lambda}_k}{k^{k-2}} \geq \sum_{k=R+1}^{R_N} \tilde{\lambda}_k \log \frac{k^2 t \tilde{\lambda}_k}{e} \\ &\geq \sum_{k=R+1}^{R_N} \frac{e}{t k^2} \varphi\left(\frac{\sum_{k=R+1}^{R_N} \tilde{\lambda}_k}{\sum_{k=R+1}^{R_N} \frac{e}{t k^2}}\right) \\ &= \sum_{k=R+1}^{R_N} \tilde{\lambda}_k \log \left(\frac{\sum_{k=R+1}^{R_N} \tilde{\lambda}_k}{\sum_{k=R+1}^{R_N} \frac{e}{t k^2}}\right) \geq \sum_{k=R+1}^{R_N} \tilde{\lambda}_k \log \left(cR \sum_{k=R+1}^{R_N} \tilde{\lambda}_k\right) \\ &\geq -\gamma_R, \end{aligned} \quad (3.8)$$

for some $c > 0$, where we used that the remainder sum $\sum_{k > R} \frac{1}{k^2}$ is of order $1/R$ as $R \rightarrow \infty$ and that $\sum_{k=R+1}^{R_N} \tilde{\lambda}_k \leq 1/R$ since $\sum_k k \tilde{\lambda}_k \leq 1$ and that the map $x \mapsto x \log(cRx)$ is decreasing in $(0, 1/eRc)$, introducing some $-\gamma_R$ that vanishes as $R \rightarrow \infty$. This proves the claim (3.7). Notice that the last term on the right-hand side of (3.7) cannot be further estimated with the

help of continuity (since $\Lambda \mapsto c_\Lambda$ is not continuous), but will be jointly handled together with the correspondent macroscopic and mesoscopic terms.

Next we focus on the term $F_{\text{Ma}}(\ell)$, the macroscopic term. We proceed analogously, but use the upper bound in Lemma 2.4 and Stirling's lower bound for $k!$ (in this case we know that k is large and we expect ℓ_k small). We obtain, for $k \in \{\varepsilon_N N, \dots, N\}$,

$$z_k(\ell) \leq \frac{(1 - e^{kq_N})^{(k-1)\ell_k} e^{-\frac{t}{2N}k(N-k)\ell_k}}{k^{k\ell_k} e^{-k\ell_k} \ell_k!},$$

where $q_N = \log(1 - \frac{t}{N})$. We pair $F_{\text{Ma}}(\ell)$ with the second term in (3.5), and we obtain, uniformly for $\ell \in \mathcal{N}_N$,

$$\begin{aligned} \left(\frac{N}{e}\right)^{Nc_{\text{Ma}}(\ell/N)} F_{\text{Ma}}(\ell) &\leq \prod_{\varepsilon_N N \leq k \leq N} \left[\left(\frac{N}{e}\right)^{k\ell_k} \frac{(1 - e^{kq_N})^{(k-1)\ell_k} e^{-\frac{t}{2}k\ell_k} e^{\frac{t}{2}k^2\ell_k/N}}{k^{k\ell_k} e^{-k\ell_k} \ell_k!} \right] \\ &\leq \prod_{\varepsilon_N N \leq k \leq N} \left[\left(\frac{N}{k}\right)^{k\ell_k} (1 - e^{kq_N})^{(k-1)\ell_k} e^{-\frac{t}{2}k\ell_k} e^{\frac{t}{2}k^2\ell_k/N} \right] \\ &= \exp\left(-NI_{\text{Ma}}^{(\varepsilon_N)}(\ell_{[\cdot, N]}; t) + o(N)\right) \end{aligned} \quad (3.9)$$

where

$$I_{\text{Ma}}^{(\varepsilon_N)}(\tilde{\alpha}; t) = g^{(\varepsilon_N)}(\tilde{\alpha}; t) + \int_{[\varepsilon_N, 1]} x \tilde{\alpha}(dx) \left(\frac{t}{2} - \log t\right) \quad (3.10)$$

with

$$g^{(\varepsilon_N)}(\tilde{\alpha}; t) = \int_{[\varepsilon_N, 1]} \left[x \log \frac{tx}{1 - e^{-tx}} - \frac{t}{2}x^2 \right] \tilde{\alpha}(dx),$$

denotes the cut-off version of the rate function I_{Ma} defined in (1.5). Indeed, for proving the last line of (3.9) we do the following. In the product, we add the factor $(1 - e^{-kt/N})^{k\ell_k}$ and its reciprocal, substitute $\exp \circ \log$ and turn the sum on k into an integral over x . Then most of the terms are easily asymptotically identified with the corresponding terms in (3.10), with possible exception of the term

$$\prod_{\varepsilon_N N \leq k \leq N} \left[(1 - e^{kq_N})^{(k-1)\ell_k} (1 - e^{-kt/N})^{-k\ell_k} \right], \quad (3.11)$$

of which we now show that it is not larger than $e^{o(N)}$. Now write ℓ in terms of $s = (s_i)_{i \in \{1, \dots, n\}} \in E_N$ defined in (2.2), such that $\sum_i s_i = N$, and we pick i^* minimal such that $s_{i^*+1} < \varepsilon_N N$. Then, also using the inequalities $\log(1 + y) \leq y$ and $1 - e^{-x} \leq x$, we see that

$$\begin{aligned} \sum_{\varepsilon_N N \leq k \leq N} k\ell_k \log \frac{1 - e^{kq_N}}{1 - e^{-k\frac{t}{N}}} &= \sum_{i=1}^{i^*} s_i \log \frac{1 - e^{s_i q_N}}{1 - e^{-s_i t/N}} \leq \sum_{i=1}^{i^*} s_i \frac{e^{-s_i t/N} - e^{s_i q_N}}{1 - e^{-s_i t/N}} \\ &= \sum_{i=1}^{i^*} s_i \frac{e^{-s_i t/N}}{1 - e^{-s_i t/N}} \left(1 - e^{s_i(q_N + t/N)}\right) \leq - \sum_{i=1}^{i^*} s_i^2 \frac{e^{-s_i t/N}}{1 - e^{-s_i t/N}} (q_N + t/N). \end{aligned}$$

Recall the definition of q_N to see that $q_N + t/N \leq o(1/N)$. Use Jensen's inequality and $\sum_i s_i \leq N$ to see that the entire last term is not larger than $o(N)$. To handle the last missing term in (3.11), notice (because of $\sum_k k\ell_k = N$) that $\sum_{k \geq \varepsilon_N N} \ell_k \leq 1/\varepsilon_N$ and $1 - e^{kq_N} \geq$

$1 - e^{\varepsilon_N N q_N} \geq \varepsilon_N N q_N \sim \varepsilon_N t$ and therefore

$$- \sum_{k \geq \lfloor \varepsilon_N N \rfloor} \ell_k \log(1 - e^{k q_N}) \leq -\frac{1}{\varepsilon_N} \log(\varepsilon_N t) \leq o(N),$$

where we recall that we assumed that $|\frac{1}{\varepsilon_N} \log \varepsilon_N| \leq o(N)$. Hence, the term in (3.11) is not larger than $e^{o(N)}$.

Now, recall that $D_{\varepsilon_N}(\ell_{[\cdot, N]}, \alpha) \leq \rho$ and $c_\alpha = \int_{(0,1]} x \alpha(dx) \leq 1$ and observe that $\lim_{\varepsilon \downarrow 0} I_{\text{Ma}}^{(\varepsilon)}(\alpha; t) = I_{\text{Ma}}(\alpha; t)$, then we claim that, for any $\varepsilon > 0$,

$$\left(\frac{N}{e}\right)^{N c_{\text{Ma}}(\ell/N)} F_{\text{Ma}}(\ell) \leq \exp(-N I_{\text{Ma}}(\alpha; t)) e^{N(C_\varepsilon(\rho) + \gamma_\varepsilon + \frac{t}{2}\varepsilon) + o(N)} e^{-N(\frac{t}{2} - \log t)(c_{\text{Ma}}(\ell/N) - c_\alpha)},$$

$$N \rightarrow \infty, \quad (3.12)$$

for some $C_\varepsilon(\rho)$ and γ_ε that satisfy $\lim_{\varepsilon \downarrow 0} \gamma_\varepsilon = 0$ and $\lim_{\rho \downarrow 0} C_\varepsilon(\rho) = 0$. To prove (3.12), we first observe that $g^{(\varepsilon)}(\cdot; t)$ is continuous and hence $|g^{(\varepsilon)}(\ell_{[\cdot, N]}; t) - g^{(\varepsilon)}(\alpha; t)|$ can be estimated against such a $C_\varepsilon(\rho)$, uniformly in $N \in \mathbb{N}$ and ℓ such that $D_{\varepsilon_N}(\ell_{[\cdot, N]}, \alpha) \leq \rho$. Furthermore, for any $\varepsilon > 0$ and any $N \in \mathbb{N}$ such that $\varepsilon_N < \varepsilon$,

$$g^{(\varepsilon_N)}(\ell_{[\cdot, N]}; t) - g^{(\varepsilon)}(\ell_{[\cdot, N]}; t) = \sum_{k=\varepsilon_N N}^{\varepsilon N} \ell_k \frac{k}{N} \left(\log \frac{\frac{k}{N} t}{1 - e^{-\frac{k}{N} t}} - \frac{t}{2} \frac{k}{N} \right) \geq -\frac{t}{2} \varepsilon,$$

since $\log \frac{x}{1-e^{-x}} \geq 0$ for all $x > 0$. Hence, we arrived at the bound in (3.12). Notice that again we refrain from estimating the term $e^{-N(\frac{t}{2} - \log t)(c_{\text{Ma}}(\ell/N) - c_\alpha)}$, which needs to be coupled with the microscopic and the mesoscopic part.

Then we are left to handle the middle term in (3.3), for which we use again the upper bound in (2.7) and Stirling's formula, to see that

$$\begin{aligned} \left(\frac{N}{e}\right)^{N c_{\text{Me}}(\ell/N)} F_{\text{Me}}(\ell) &\leq \prod_{k=R_N+1}^{\lfloor \varepsilon_N N \rfloor} \left[\left(\frac{N}{e}\right)^{k \ell_k} \frac{k^{(k-2)\ell_k} (1 - e^{k q_N})^{(k-1)\ell_k} e^{-\frac{t}{2N} k(N-k)\ell_k}}{k!^{\ell_k} \ell_k!} \right] \\ &\leq \left(\prod_{k=R_N+1}^{\lfloor \varepsilon_N N \rfloor} \left[\frac{N e}{k^2 \ell_k t} \right]^{\ell_k} \right) \left(\prod_{k=R_N+1}^{\lfloor \varepsilon_N N \rfloor} e^{\frac{t}{2N} k^2 \ell_k} \right) (te^{-t/2})^{N c_{\text{Me}}(\ell/N)}. \end{aligned}$$

We claim that the right-hand side is equal to $(te^{-t/2})^{N c_{\text{Me}}(\ell/N)} e^{N L_N(\ell)}$ for some $L_N(\ell)$ that vanishes, uniformly in ℓ , as $N \rightarrow \infty$. First note that the next-to-last term is such a term, since $\frac{t}{2N} \sum_{k=R_N+1}^{\lfloor \varepsilon_N N \rfloor} k^2 \ell_k \leq \frac{t}{2} \varepsilon_N N c_{\text{Me}}(\ell/N)$. Furthermore, $\sum_{k=R_N+1}^{\lfloor \varepsilon_N N \rfloor} \ell_k \leq N/R_N$, which shows that the terms containing t and e in the first product are also so small. With the same approach as in (3.8), we see the lower bound

$$\liminf_{N \rightarrow \infty} \sum_{k=R_N+1}^{\lfloor \varepsilon_N N \rfloor} \frac{\ell_k}{N} \log \frac{k^2 \ell_k}{N} \geq 0.$$

Therefore, uniformly in ℓ such that $D_{\varepsilon_N}(\ell_{[\cdot, N]}, \alpha) \leq \rho$, we have arrived at the estimate

$$\left(\frac{N}{e}\right)^{N c_{\text{Me}}(\ell/N)} F_{\text{Me}}(\ell) \leq (te^{-t/2})^{N c_{\text{Me}}(\ell/N) + o(N)} = \exp\left(-N\left(\frac{t}{2} - \log t\right) c_{\text{Me}}(\ell/N)\right) e^{o(N)}, \quad N \rightarrow \infty. \quad (3.13)$$

Now we collect (3.7), (3.12) and (3.13) and substitute them in (3.3), also using (3.5), then we obtain, uniformly in ℓ such that $d(\frac{1}{N}\ell, \Lambda) < \delta$ and $D(\ell_{[\cdot, N]}, \alpha) < \rho$, for any $R \in \mathbb{N}$ and any $\varepsilon > 0$, as $N \rightarrow \infty$,

$$\begin{aligned} \frac{1}{N} \log \mathbb{P}_{N, \frac{1}{N}t_N}(A_{N,t}(\ell)) &\leq -I_{\text{Mi}}(\Lambda; t) - I_{\text{Ma}}(\alpha; t) + C_R(\delta) + \gamma_R + C_\varepsilon(\rho) + \gamma_\varepsilon + \frac{t}{2}\varepsilon \\ &\quad - \left(\frac{t}{2} - \log t\right)(1 - c_\Lambda - c_\alpha) + o(1) \\ &= -I(\Lambda, \alpha; t) + K_{R,\varepsilon}(\delta, \rho) + o(1), \end{aligned}$$

where $K_{R,\varepsilon}(\delta, \rho)$ vanishes as $\delta \downarrow 0$ and $\rho \downarrow 0$, followed by $R \rightarrow \infty$ and $\varepsilon \downarrow 0$, and we recall that $c_{\text{Me}}(\ell/N) = 1 - c_{\text{Mi}}(\ell/N) - c_{\text{Ma}}(\ell/N)$. This implies the upper bound in (3.2) in the case where $c_\Lambda + c_\alpha \leq 1$. □

In the following lemma, we implicitly use the lower semicontinuity of the maps $\Lambda \mapsto c_\Lambda$ and $\alpha \mapsto c_\alpha$ to show that when $c_\Lambda + c_\alpha > 1$, then the event $A_{N,t}(\ell)$ is empty for any ℓ such that $d_{R_N}(\frac{1}{N}\ell, \Lambda) \leq \delta$ and $D_{\varepsilon_N}(\ell_{[\cdot, N]}, \alpha) \leq \rho$, if δ and ρ are small enough. This will give the right super-exponential upper bound for $\mathbb{P}_{N, \frac{1}{N}t_N}(A_{N,t}(\ell))$, since $I(\Lambda, \alpha; t) = \infty$.

Lemma 3.4. *Let $(\Lambda, \alpha) \in \mathcal{N} \times \mathcal{M}$ such that $c_\Lambda + c_\alpha > 1$, then there exists $R \in \mathbb{N}$, $\varepsilon, \delta, \rho > 0$ and N_0 large enough such that for all $N > N_0$:*

$$\left\{ \ell \in \mathcal{N}_N : d_R(\frac{1}{N}\ell, \Lambda) \leq \delta, \quad D_\varepsilon(\ell_{[\cdot, N]}, \alpha) \leq \rho \right\} = \emptyset.$$

Proof. We pick $R \in \mathbb{N}$ so large and $\varepsilon \in (0, 1)$ so small that $\sum_{k=1}^R k\lambda_k + \int_{[\varepsilon, 1]} x\alpha(dx)$ are larger than one, say equal to $1 + \eta$ for some $\eta > 0$. Then choose δ and ρ in $(0, 1)$ so small that, for any ℓ such that $d_R(\frac{1}{N}\ell, \Lambda) \leq \delta$ and $D_\varepsilon(\ell_{[\cdot, N]}, \alpha) \leq \rho$, we have $\frac{1}{N} \sum_{k=1}^R k\ell_k - \sum_{k=1}^R k\lambda_k \geq -\frac{\eta}{3}$ and $\frac{1}{N} \sum_{k=R+1}^N k\ell_k - \int_{[\varepsilon, 1]} x\alpha(dx) \geq -\frac{\eta}{3}$. Therefore we see that

$$\begin{aligned} 1 &= \frac{1}{N} \sum_{k=1}^N k\ell_k \geq \frac{1}{N} \sum_{k=1}^N k\ell_k - \sum_{k=1}^R k\lambda_k - \int_{[\varepsilon, 1]} x\alpha(dx) + \sum_{k=1}^R k\lambda_k + \int_{[\varepsilon, 1]} x\alpha(dx) \\ &\geq -\frac{\eta}{3} - \frac{\eta}{3} + \sum_{k=1}^R k\lambda_k + \int_{[\varepsilon, 1]} x\alpha(dx) > 1 + \frac{\eta}{3}, \end{aligned}$$

which yields a contradiction. □

The last lemma of this section deals with the construction of an optimal sequence $(\ell^{(N)})_{N \in \mathbb{N}}$, which will give a lower bound on the probability that matches the upper bound from Lemma 3.3.

Lemma 3.5. *For all $(\Lambda, \alpha) \in \mathcal{N} \times \mathcal{M}$ such that $I(\Lambda, \alpha; t) < \infty$, there exists a sequence $(\ell^{(N)})_{N \in \mathbb{N}}$ such that $\ell^{(N)} \in \mathcal{N}_N$ and*

$$\lim_{N \rightarrow \infty} d(\frac{1}{N}\ell^{(N)}, \Lambda) = 0, \tag{3.14}$$

$$\lim_{N \rightarrow \infty} D(\ell_{[\cdot, N]}^{(N)}, \alpha) = 0, \tag{3.15}$$

$$\liminf_{N \rightarrow \infty} \frac{1}{N} \log \mathbb{P}_{N, \frac{1}{N}t_N}(A_{N,t}(\ell^{(N)})) \geq -I(\Lambda, \alpha; t). \tag{3.16}$$

Proof. For N large enough define $\ell^{(N)} \in \mathcal{N}_N$ by

$$\ell_k^{(N)} = \begin{cases} \lfloor \lambda_k N \rfloor & \text{for } k = 2, \dots, R_N - 1; \\ \lfloor \frac{(1-c_\lambda - c_\alpha)N}{R_N} \rfloor & \text{for } k = R_N; \\ \alpha\left(\frac{k-1}{N}, \frac{k}{N}\right) & \text{for } k = R_N + 1, \dots, N; \\ N - \sum_{j \geq 2}^N j \ell_j^{(N)} & \text{for } k = 1, \end{cases} \quad (3.17)$$

where R_N is an arbitrary diverging sequence in \mathbb{N} such that $R_N \ll N$.

We notice that our sequence $(\ell^{(N)})_{N \in \mathbb{N}}$ is such that the so-called mesoscopic mass is simply concentrated in components all of the same size (namely R_N) and surprisingly no specific requirement is imposed on R_N , except that it diverges. We will underline this in the steps of our proof. It is clear by construction that (3.14) and (3.15) hold. We now prove (3.16), starting from the formulation in (3.3). By abuse of notation, we will drop the index (N) from ℓ .

We use the lower bound in (2.7) from Lemma 2.3 to perform a calculation similar to that for the upper bound (using now Stirling's upper bound on $\ell_k!$):

$$\begin{aligned} \left(\frac{N}{e}\right)^{N c_{\text{Mi}}(\ell/N)} F_{\text{Mi}}(\ell) &\geq \prod_{k=1}^{R_N} \left[\left(\frac{N}{e}\right)^{k \ell_k} \frac{k^{(k-2)\ell_k} t_N^{(k-1)\ell_k}}{k! \ell_k N^{(k-1)\ell_k} (\ell_k)^{\ell_k} e^{-\ell_k} \sqrt{2\pi \ell_k}} \left(1 - \frac{t_N}{N}\right)^{\frac{1}{2} k \ell_k N - \frac{3}{2} k \ell_k + \ell_k} \right] \\ &= \exp\left(-N \sum_{k=1}^{R_N} \frac{1}{N} \ell_k \log \frac{k! e^k \frac{1}{N} \ell_k}{k^{k-2} t^{k-1} e^{1-\frac{t}{2}k}}\right) e^{o(N)} \\ &= \exp\left(-N I_{\text{Mi}}^{(R_N)}(\Lambda; t)\right) e^{o(N)}, \end{aligned} \quad (3.18)$$

where $I_{\text{Mi}}^{(R_N)}(\Lambda; t)$ is defined in (3.6). Notice that no restriction is required on R_N in order for the above lower bound to coincide with the upper bound in the proof of Lemma 3.3 (except that R_N diverges). Indeed, although the lower bound in (2.7) differs from the upper bound for a factor $(1 - \frac{t_N}{N})^{\frac{(k-1)(k-2)}{2}}$ the probability $\mu_{t_N}(k)$ is paired with $(1 - \frac{t_N}{N})^{\frac{k(N-k)}{2}}$ (the probability of a component to be separated by any other), which is an exact term and it balances the error coming from (2.7).

In the same way, one checks that for $k = R_N$

$$\begin{aligned} \left(\frac{N}{e}\right)^{N(1-c_\lambda - c_\alpha)} z_{R_N}(\ell_{R_N}) &\geq \left(\frac{N}{e}\right)^{N(1-c_\lambda - c_\alpha)} \frac{R_N^{(R_N-2)\ell_{R_N}} \left(\frac{t_N}{N}\right)^{(R_N-1)\ell_{R_N}} \left(1 - \frac{t_N}{N}\right)^{\frac{1}{2} R_N \ell_{R_N} N - \frac{3}{2} R_N \ell_{R_N} + \ell_{R_N}}}{R_N^{R_N \ell_{R_N}} e^{-R_N \ell_{R_N}} (\sqrt{2\pi R_N})^{\ell_{R_N}} \ell_{R_N}^{\ell_{R_N}} e^{-\ell_{R_N}} \sqrt{2\pi \ell_{R_N}}} \\ &= \exp\left(- (1 - c_\lambda - c_\alpha) \left(\frac{t}{2} - \log t\right)\right) e^{o(N)}. \end{aligned} \quad (3.19)$$

The above lower bound relies on the lower bound (2.7) in Lemma 2.3, on Stirling's upper bound for $R_N!$ and $\ell_{R_N}!$ (which are both large) and on how we defined ℓ_{R_N} in (3.17). We notice that the above term, in order to give the desired lower bound matching the upper bound from Lemma 3.3, does not add any condition on the sequence R_N , which is just supposed to diverge.

Now, fix $\delta \in (0, 1)$. We use the lower bound (2.7) from Lemma 2.3 and Stirling's upper bound on $k!$ and we write:

$$\begin{aligned} \prod_{k=R_N+1}^{\lfloor \delta N \rfloor} \left(\frac{N}{e}\right)^{k\ell_k} z_k(\ell_k) &\geq \prod_{k=R_N+1}^{\lfloor \delta N \rfloor} \left[\left(\frac{N}{e}\right)^{k\ell_k} \frac{k^{(k-2)\ell_k} t_N^{(k-1)\ell_k} \left(1 - \frac{t_N}{N}\right)^{\frac{kN}{2}\ell_k - \frac{3}{2}k\ell_k + \frac{\ell_k}{2}}}{\ell_k! N^{(k-1)\ell_k} k^{k\ell_k} e^{-k\ell_k} (2\pi k)^{\frac{\ell_k}{2}}} \right] \\ &\geq \exp\left(-N\left(\frac{t}{2} - \log t\right) \sum_{k=R_N+1}^{\delta N} \frac{k}{N} \alpha\left(\frac{k-1}{N}, \frac{k}{N}\right] + \mathcal{R}_{R_N, \delta}\right), \end{aligned}$$

where the remainder is defined as

$$\mathcal{R}_{R_N, \delta} = \sum_{k=R_N+1}^{\delta N} \ell_k \left(\log N - \log t + \frac{t}{2N} - 2 \log k - \frac{1}{2} \log(2\pi k) - \frac{3t}{2N} k \right) - \sum_{k=R_N+1}^{\delta N} \log \ell_k!.$$

By defining $\ell_k := \alpha\left(\frac{k-1}{N}, \frac{k}{N}\right]$ and using that $\log k \leq k$, we see that

$$\begin{aligned} \frac{1}{\delta} \int_{\frac{R_N+1}{N}}^{\delta} x \alpha(dx) &\leq \sum_{k=R_N+1}^{\delta N} \ell_k \leq \frac{N}{R_N} \int_{\frac{R_N+1}{N}}^{\delta} x \alpha(dx), \\ \sum_{k=R_N+1}^{\delta N} \ell_k \log k &\leq N \int_{\frac{R_N+1}{N}}^{\delta} x \alpha(dx). \end{aligned}$$

Finally, we use that $\sum_{h=1}^{\ell_k} \log h \leq \int_1^{\ell_k} \log(y) dy$ and Jensen's inequality (since $x \log x$ is concave) to give the bound

$$\sum_{k=R_N+1}^{\delta N} \log \ell_k! \leq \sum_{k=R_N+1}^{\delta N} (\ell_k \log \ell_k - \ell_k + 1) \leq N\delta.$$

Exploiting the bounds above, we can say that

$$\prod_{k=R_N+1}^{\lfloor \delta N \rfloor} \left(\frac{N}{e}\right)^{k\ell_k} z_k(\ell_k) \geq \exp\left(-N\left(\frac{t}{2} - \log t\right) \sum_{k=R_N+1}^{\delta N} \frac{k}{N} \alpha\left(\frac{k-1}{N}, \frac{k}{N}\right] - r_\delta\right) + o(N), \quad (3.20)$$

where r_δ depends on $\int_0^\delta x \alpha(dx)$ and it goes to zero when $\delta \searrow 0$. Notice that we could not handle directly the terms from $R_N + 1$ to N in one go, so we decided to fix a fictitious threshold δN , where δ can be chosen arbitrarily close to zero. This however does not affect the choice of R_N and there is no need to add some condition on R_N to handle such small, but macroscopic, components.

The remaining terms (the purely macroscopic ones) in (3.3) are treated as follows. We use Stirling's upper bound for $k!$ and we see that

$$\begin{aligned} \prod_{k=\lfloor \delta N \rfloor}^N \left(\frac{N}{e}\right)^{k\ell_k} z_k(\ell_k) &\geq \prod_{k=\lfloor \varepsilon_N N \rfloor}^{\lfloor \delta N \rfloor} \left[\left(\frac{N}{e}\right)^{k\ell_k} \frac{\mu_k\left(\frac{1}{N}t_N\right)^{\ell_k} \left(1 - \frac{t_N}{N}\right)^{\frac{kN}{2}\ell_k - \frac{k^2}{2}\ell_k}}{\ell_k! N^{k\ell_k} \left(\frac{k}{N}\right)^{k\ell_k} e^{-k\ell_k} (2\pi k)^{\frac{\ell_k}{2}}} \right] \\ &\geq \exp\left(\sum_{k=\lfloor \delta N \rfloor+1}^N k \ell_k^{(N)} \left[\log\left(\frac{\mu_k\left(\frac{1}{N}t_N\right)^{\frac{1}{k}}}{k/N}\right) - \frac{t}{2N}(N-k)\right]\right) e^{o(N)} \\ &= \exp\left(N \int_{(\delta, 1]} x \left[\log\left(\frac{\mu_{\lfloor Nx \rfloor}\left(\frac{1}{N}t_N\right)^{\frac{1}{\lfloor Nx \rfloor}}}{x}\right) - \frac{t}{2}(1-x)\right] \alpha(dx)\right) e^{o(N)}. \end{aligned}$$

Let us focus on the integral $\int_{(\delta,1]} x \left[\log \left(\frac{\mu_{\lfloor Nx \rfloor}(\frac{1}{N}t_N)^{\lfloor \frac{1}{Nx} \rfloor}}{x} \right) - \frac{t}{2}(1-x) \right] \alpha(dx)$. Now by (2.8) in Lemma 2.4, we know that the integrand converges pointwise to

$$x \left[\log \left(\frac{1 - e^{-tx}}{x} \right) - \frac{t}{2}(1-x) \right].$$

Since, for N large enough,

$$x \log \frac{x}{\mu_{\lfloor Nx \rfloor}(\frac{1}{N}t_N)^{\lfloor \frac{1}{Nx} \rfloor}} \leq x \left[\log \left(\frac{x}{1 - e^{-tx}} \right) + 1 \right],$$

which is clearly integrable over $x \in (\delta, 1]$ with respect to α , we can apply the dominated convergence theorem and we get

$$\prod_{k=\lfloor \delta N \rfloor}^N \left(\frac{N}{e} \right)^{k\ell_k} z_k(\ell_k) \geq \exp \left(-NI_{\text{Ma}}^{(\delta)}(\alpha; t) + o(N) \right), \quad (3.21)$$

where $I_{\text{Ma}}^{(\delta)}(\alpha; t)$ is defined in (3.10).

Finally, we combine the lower bounds in (3.18), (3.19), (3.20) and (3.21), to get that there exists a_δ with $\lim_{\delta \searrow 0} a_\delta = 0$ and

$$\liminf_{N \rightarrow \infty} \frac{1}{N} \log \mathbb{P}_{N, \frac{1}{N}t_N}(A_{N,t}(\ell^{(N)})) \geq -I(\Lambda, \alpha; t) - a_\delta \quad (3.22)$$

so (3.16) follows on taking the limit $\delta \searrow 0$. □

Now we are ready to prove Proposition 3.1 by combining the above lemmas.

Proof of Proposition 3.1. Fix $\delta, \rho > 0$ and $N \in \mathbb{N}$ and recall the definition of $A_{N,t}(\ell)$ in (2.5), then we see that

$$\begin{aligned} & \mathbb{P}_{N, \frac{1}{N}t_N}(\text{Mi}^{(N)} \in B_\delta(\Lambda), \text{Ma}^{(N)} \in B_\rho(\alpha)) \\ &= \sum_{\ell \in \mathcal{N}_N} \mathbb{1}\{d(\frac{1}{N}\ell, \Lambda) \leq \delta\} \mathbb{1}\{D(\ell_{\lfloor \cdot \rfloor}, \alpha) \leq \rho\} \mathbb{P}_{N, \frac{1}{N}t_N}(A_N(\ell)). \end{aligned} \quad (3.23)$$

Because of Lemma 3.2, we only have to give asymptotic estimates on the single summands on the right-hand side of (3.23). Let $(\Lambda, \alpha) \in \mathcal{N} \times \mathcal{M}$ such that $c_\Lambda + c_\alpha \leq 1$. We fix any diverging sequence R_N and vanishing sequence ε_N with $R_N < \lfloor \varepsilon_N N \rfloor$ and $|\frac{1}{\varepsilon_N} \log \varepsilon_N| \leq o(N)$, then we see that

$$\begin{aligned} & \mathbb{P}_{N, \frac{1}{N}t_N}(\text{Mi}^{(N)} \in B_\delta(\Lambda), \text{Ma}^{(N)} \in B_\rho(\alpha)) \\ & \leq \sum_{\ell \in \mathcal{N}_N} \mathbb{1}\{d_{R_N}(\frac{1}{N}\ell, \Lambda) \leq \delta\} \mathbb{1}\{D_{\varepsilon_N}(\ell_{\lfloor \cdot \rfloor}, \alpha) \leq \rho\} \mathbb{P}_{N, \frac{1}{N}t_N}(A_N(\ell)) \\ & \leq \exp \left(-NI(\Lambda, \alpha; t) + NK_{R,\varepsilon}(\delta, \rho) + o(N) \right), \end{aligned}$$

as a consequence of Lemma 3.3. Taking first δ and ρ to zero and then $R \nearrow \infty$ and $\varepsilon \searrow 0$, we get the desired upper bound.

Given any fixed δ and ρ , we can construct the sequence $(\ell^{(N)})_{N \in \mathbb{N}}$ from Lemma 3.5 and see that, for such a sequence

$$\mathbb{P}_{N, \frac{1}{N}t_N}(\text{Mi}^{(N)} \in B_\delta(\Lambda), \text{Ma}^{(N)} \in B_\rho(\alpha)) \geq \mathbb{P}_{N, \frac{1}{N}t_N}(A_N(\ell^{(N)})) \geq \exp(-NI(\Lambda, \alpha; t) - Na),$$

for an a arbitrarily small.

Finally, if $(\Lambda, \alpha) \in \mathcal{N} \times \mathcal{M}$ are such that $c_\Lambda + c_\alpha > 1$, Lemma 3.3 gives us that

$$\limsup_{N \rightarrow \infty} \frac{1}{N} \log \mathbb{P}_{N, \frac{1}{N}t_N}(\text{Mi}^{(N)} \in B_\delta(\Lambda), \text{Ma}^{(N)} \in B_\rho(\alpha)) \leq -\infty.$$

□

4. COROLLARIES AND STUDY OF THE RATE FUNCTIONS

In this section we analyse, for fixed $t \in [0, \infty)$, the minima of the rate function, $I(\Lambda, \alpha; t)$, over the configurations Λ respectively α , and afterwards the minimima of the rate functions for the total masses, \mathcal{J}_{Mi} , \mathcal{J}_{Me} and \mathcal{J}_{Ma} . In particular, we will see in Lemma 4.2 that there is a drastic difference when minimizing $I(\Lambda, \alpha; t)$ over Λ if $c_\Lambda \leq \frac{1}{t}$ or not. Clearly, no such difference can be spotted when $t \leq 1$, since we do not allow for $c_\Lambda > 1$. However, when $t > 1$ we see that $c_\Lambda > \frac{1}{t}$ is perfectly admissible and we interpret this as an analytic sign of the phase transition in $t = 1$.

4.1. Rate functions for the microscopic part. We start by minimizing $I(\Lambda, \alpha; t)$, for a fixed $\Lambda \in \mathcal{N}$ over all compatible $\alpha \in \mathcal{M}$. We will obtain the rate function for the microscopic part, and we will see that this minimum is attained for α of the form $\alpha = \delta_{c_\alpha}$. Informally speaking, the following in particular implies that, with probability tending to one, there is at most one macroscopic particle.

Lemma 4.1 (Analysis of the microscopic rate function). *Fix $\Lambda \in \mathcal{N}$ and recall that $c_\Lambda = \sum_{k \in \mathbb{N}} k \lambda_k \in [0, 1]$, then*

$$\inf_{\alpha \in \mathcal{M}_{\mathbb{N}}} I(\Lambda, \alpha; t) = I_{\text{Mi}}(\Lambda; t) + I_{\text{Ma}}(\delta_{1-c_\Lambda}; t).$$

where $I(\Lambda, \alpha; t)$, $I_{\text{Mi}}(\Lambda; t)$ and $I_{\text{Ma}}(\alpha; t)$ are defined in the statement of Theorem 1.1.

Proof. Clearly

$$\begin{aligned} \inf_{\alpha \in \mathcal{M}_{\mathbb{N}}} I(\Lambda, \alpha; t) &= \inf_{c \in [0, 1-c_\Lambda]} \inf_{\alpha \in \mathcal{M}_{\mathbb{N}_0}(c)} I(\Lambda, \alpha; t) \\ &= I_{\text{Mi}}(\Lambda; t) + \inf_{c \in [0, 1-c_\Lambda]} \left(\inf_{\alpha \in \mathcal{M}_{\mathbb{N}_0}(c)} I_{\text{Ma}}(\alpha; t) + (1 - c_\Lambda - c) \left(\frac{t}{2} - \log t \right) \right). \end{aligned}$$

Fix $c \in [0, 1]$ and $\alpha \in \mathcal{M}_{\mathbb{N}_0}(c)$. Note that $\alpha((c, 1]) = 0$ since α is a point measure with $\int_{(0,1]} x \alpha(dx) = c$. We have, denoting $f_t(x) = \log \frac{x}{1-e^{-tx}} + \frac{t}{2}(1-x)$,

$$I_{\text{Ma}}(\alpha; t) = \int_{(0,c]} x f_t(x) \alpha(dx) \geq \int x f_t(c) \alpha(dx) = c f_t(c) = I_{\text{Ma}}(\delta_c; t), \quad (4.1)$$

since f_t is strictly decreasing in $[0, \infty)$. Indeed,

$$f_t'(x) = \frac{1}{x} - \frac{te^{-tx}}{1-e^{-tx}} - \frac{t}{2} = \frac{t(1+y)}{2y(1-e^{-2y})} \left[\frac{1-y}{1+y} - e^{-2y} \right], \quad y = \frac{tx}{2}.$$

We want to prove that $f'(x) < 0$ for $x \in [0, \infty)$. For $y \geq 1$, this is obvious from above, and for $y \in [0, 1)$, this is easily seen as follows.

$$e^{2y} = 1 + \sum_{k=1}^{\infty} \frac{(2y)^k}{k!} < 1 + \sum_{k=1}^{\infty} 2y^k = (1+y) \sum_{k=0}^{\infty} y^k = \frac{1+y}{1-y},$$

since $\frac{2^k}{k!} < 2$ for all $k \geq 3$. Hence, we see that $f'_t(x) \leq 0$ for $x \in [0, \infty)$, and (4.1) follows.

Furthermore, when we study $g_t(c) := c \log \frac{c}{1-e^{-ct}} + \frac{t}{2}c(1-c) + (1-c_\Lambda - c)(\frac{t}{2} - \log t)$, we see that its derivative is

$$g'_t(c) = 1 - \frac{cte^{-ct}}{(1-e^{-ct})} + \log \frac{cte^{-ct}}{(1-e^{-ct})},$$

which is strictly negative if $\frac{cte^{-ct}}{(1-e^{-ct})} \neq 1$. Since $\frac{cte^{-ct}}{(1-e^{-ct})} < 1$ if $ct > 0$, we see that $g_t(c)$ is strictly decreasing in c , and hence the optimal value of c is $c = 1 - c_\Lambda$. \square

Now the proof of Corollary 1.2 directly follows from Theorem 1.1, Lemma 4.1 and the contraction principle since the projection $(\Lambda, \alpha) \mapsto \Lambda$ is continuous in the product topology. Let us mention that, since $c_\Lambda = \sum_{k \in \mathbb{N}} k \lambda_k$, we can rewrite

$$\mathcal{I}_{\text{Mi}}(\Lambda; t) = \widehat{I}(\Lambda) - (1 - c_\Lambda) \left(\log \frac{1 - e^{(c_\Lambda - 1)t}}{(1 - c_\Lambda)} - \frac{c_\Lambda t}{2} \right) + c_\Lambda \left(\frac{t}{2} - \log t \right) - \frac{1}{2t}, \quad (4.2)$$

with

$$\widehat{I}(\Lambda) = \sum_{k=1}^{\infty} (\lambda_k \log \frac{\lambda_k}{p_k} + p_k - \lambda_k), \quad (4.3)$$

with $p_k = \frac{1}{t} \frac{k^{k-2} e^{-k}}{k!}$ for all $k \in \mathbb{N}$. Hence, the term $\widehat{I}(\Lambda)$ is the relative entropy of two non-normalized measures Λ and $p := (p_k)_{k \in \mathbb{N}}$. Notice that the reference measure p is such that

$$p_k = \frac{1}{tk} \text{Bo}_1(k),$$

where

$$\text{Bo}_\mu(k) = \frac{e^{-\mu k} (\mu k)^{k-1}}{k!}, \quad k \in \mathbb{N}, \quad (4.4)$$

are the probabilities of the Borel distribution with parameter $\mu \in [0, 1]$. The total mass of p is therefore given by

$$\sum_{k=1}^{\infty} p_k = \frac{1}{t} \mathbb{E} \left[\frac{1}{X} \right],$$

where X is Borel distributed with parameter 1. Now this expectation [AP98, §4.5] is precisely $\frac{1}{2}$, which explains why we added and subtracted the term $\frac{1}{2t}$ to $\mathcal{I}_{\text{Mi}}(\Lambda; t)$ in order to obtain the formulation (4.2). As already mentioned in Section 1.4, the above entropy form for the rate function (4.2) strictly relates to the rate function obtained in [BC15, Theorem 1.8] which also takes the form of an entropy with respect to a standard Galton Watson tree, whose total progeny is precisely Borel distributed.

Let us analyse the minimising statistics of the macroscopic part.

Lemma 4.2 (Analysis of the macroscopic rate function). *Fix $\alpha \in \mathcal{M}_{\mathbb{N}_0}$ and recall that $c_\alpha = \int_{(0,1]} x \alpha(dx) \in [0, 1]$, then*

$$\inf_{\Lambda \in \mathcal{N}} I(\Lambda, \alpha; t) = I_{\text{Ma}}(\alpha; t) + C_{\alpha, t} \left(\log(t C_{\alpha, t}) - \frac{t}{2} C_{\alpha, t} \right) + (1 - c_\alpha) \left(\frac{t}{2} - \log t \right), \quad (4.5)$$

where $C_{\alpha,t} = (1 - c_\alpha) \wedge \frac{1}{t}$.

Furthermore, the unique minimizer is equal to $\Lambda^*(C_{\alpha,t}; t)$, defined in (1.13).

Proof. As in the proof of Lemma 4.1, we see that

$$\begin{aligned} \inf_{\Lambda \in \mathcal{N}} I(\Lambda, \alpha; t) &= \inf_{c \in [0, 1 - c_\alpha]} \inf_{\Lambda \in \mathcal{N}(c)} I(\Lambda, \alpha; t) \\ &= I_{\text{Ma}}(\alpha; t) + \inf_{c \in [0, 1 - c_\alpha]} \left(\inf_{\Lambda \in \mathcal{N}(c)} I_{\text{Mi}}(\Lambda; t) + (1 - c_\alpha - c) \left(\frac{t}{2} - \log t \right) \right) \\ &= I_{\text{Ma}}(\alpha; t) + (1 - c_\alpha) \left(\frac{t}{2} - \log t \right) - \frac{1}{2t} + \inf_{c \in [0, 1 - c_\alpha]} \inf_{\Lambda \in \mathcal{N}(c)} \widehat{I}(\Lambda), \end{aligned} \quad (4.6)$$

with $\widehat{I}(\Lambda)$ defined in (4.3). Fix $c \in [0, 1]$. Since \widehat{I} is strictly convex on the convex set $\mathcal{N}(c)$, we see by evaluating the variational equations that the only candidate for a minimiser in the interior is

$$\lambda_k^*(c; t) = \frac{e^{k(\rho-1)} k^{k-2}}{k! t}, \quad k \in \mathbb{N},$$

with $\rho \in \mathbb{R}$ such that $\sum_{k=1}^{\infty} k \lambda_k^*(c; t) = c$. Interestingly, we can identify $k \lambda_k^*(c; t) = \text{Bo}_{\mu_\rho}(k) \mu_\rho / t$, where μ_ρ is determined by $\mu_\rho - \log \mu_\rho = 1 - \rho$ and Bo_μ is defined in (4.4). Note that $\text{Bo}_\mu(k)$ is not summable for $\mu > 1$. Hence, ρ must be picked such that $c = \mu_\rho / t$. The largest value c that can be realised in this way is $c = 1/t$ by picking $\rho = 0$. Hence, the preceding is possible at most for $c \in [0, 1 \wedge \frac{1}{t}]$. By continuity and strict monotonicity of $\sum_{k=1}^{\infty} k \lambda_k^*(c; t)$ in ρ , indeed, any $c \in [0, 1 \wedge \frac{1}{t}]$ can be uniquely realized, by picking $\rho = -tc + \log tc + 1 \leq 0$ such that $\sum_{k=1}^{\infty} k \lambda_k^*(c; t) = c$. In this case, it is clear that the minimizer of \widehat{I} in the interior of $\mathcal{N}(c)$ is equal to

$$\lambda_k^*(c; t) = \frac{k^{k-2} c^k t^{k-1} e^{-ctk}}{k!}, \quad k \in \mathbb{N},$$

as claimed in (1.13), with value

$$\widehat{I}(\Lambda^*(c; t)) = c(\log ct + 1 - ct) - \sum_{k=1}^{\infty} \lambda_k^*(c; t) + \frac{1}{2t} = c \left(\log tc - \frac{tc}{2} \right) + \frac{1}{2t}, \quad (4.7)$$

where we used that if $X \sim \text{Bo}_{ct}$, then $\sum_{k=1}^{\infty} \lambda_k^*(c; t) = \mathbb{E} \left[\frac{1}{X} \right] = 1 - \frac{ct}{2}$, see [AP98, §4.5]. Now we give an argument why $\Lambda^*(c; t)$ realises the minimum of \widehat{I} over $\mathcal{N}(c)$. We show that any such minimiser must be positive in every component. Indeed, if $\lambda_{k^*} = 0$ for some $k^* \in \mathbb{N}$, then we consider $\widehat{\Lambda} \in \mathcal{N}(c)$, defined by

$$\widehat{\lambda}_k = \begin{cases} \varepsilon, & \text{if } k = k^*, \\ \lambda_{\widehat{k}} - \varepsilon C, & \text{if } k = \widehat{k}, \\ \lambda_k & \text{otherwise,} \end{cases}$$

with $\widehat{k} \in \mathbb{N} \setminus \{k^*\}$ such that $\lambda_{\widehat{k}} > 0$ and $C > 0$ such that $\widehat{\Lambda} \in \mathcal{N}(c)$ for any sufficiently small $\varepsilon > 0$. Now a simple insertion shows that $\widehat{I}(\widehat{\Lambda}) < \widehat{I}(\Lambda)$, if $\varepsilon > 0$ is small enough, since the slope of $\varepsilon \mapsto \varepsilon \log \varepsilon$ at zero is $-\infty$. Hence, Λ cannot be a minimizer. On the other hand, $\Lambda^*(c; t)$ has the property that all directional derivatives of \widehat{I} in all admissible directions with compact support are zero; hence it is the minimizer of \widehat{I} over $\mathcal{N}(c)$ for $c \in [0, \frac{1}{t}]$.

When $c > \frac{1}{t}$, it is possible to pick a sequence of $\Lambda^{(n)} \in \mathcal{N}(c)$ such that $\lim_{n \rightarrow \infty} \widehat{I}(\Lambda^{(n)}) = 0$ (pick $\lambda_k^{(n)}$ as $\lambda_k^*(\frac{1}{t}; t) + \varepsilon_n \delta_n(k)$ for some suitable $\varepsilon_n > 0$). Furthermore, since $\widehat{I}(\Lambda)$ is a relative

entropy, we know that $\inf_{\Lambda \in \mathcal{N}} \widehat{I}(\Lambda) \geq 0$. Hence, the infimum of \widehat{I} over $\Lambda \in \mathcal{N}(c)$ for $c \geq \frac{1}{t}$ is equal to 0. This shows that the infimum over $\Lambda \in \mathcal{N}(c)$ in the last line of (4.6) is equal to $(c \wedge \frac{1}{t})(\log(t(c \wedge \frac{1}{t})) - \frac{t}{2}(c \wedge \frac{1}{t}))$, and (4.5) follows. \square

Then, the proof of Corollary 1.3 directly follows from Theorem 1.1, Lemma 4.2 and the contraction principle, since the projection is continuous.

Finally, let us draw some conclusions regarding the mesoscopic mass. As stated after Corollary 1.4, it is not possible to apply the contraction principle, if we want to derive an LDP for the sequence of random variables $\overline{\text{Me}}_{R_N, \varepsilon_N}^{(N)}(t)$, however we can still identify the rate function by minimizing I over all pairs (Λ, α) such that $c_\Lambda + c_\alpha = 1 - c$. Even if the contraction principle cannot be applied directly, the following lemma proves that the rate function $\mathcal{J}_{\text{Me}}(c; t)$ has exactly the expected form, given by (4.9).

Lemma 4.3. *Fix $t \in [0, \infty)$. Then, for any $c \in [0, 1]$ and any $R_N \in \mathbb{N}$ and $\varepsilon_N \in (0, 1)$ such that $1 \ll R_N < \varepsilon_N N \ll N$,*

$$\lim_{\delta \downarrow 0} \lim_{N \rightarrow \infty} \frac{1}{N} \log \mathbb{P}_{N, \frac{1}{N}t_N} \left(\left| \overline{\text{Me}}_{R_N, \varepsilon_N}^{(N)}(t) - c \right| \leq \delta \right) = -\mathcal{J}_{\text{Me}}(c; t). \quad (4.8)$$

Proof. We first verify that, for a fixed $c \in [0, 1]$,

$$\inf_{\substack{\Lambda \in \mathcal{N} \\ \alpha \in \mathcal{M}_{\mathbb{N}_0} \\ c_\Lambda + c_\alpha = 1-c}} I(\Lambda, \alpha; t) = \mathcal{J}_{\text{Me}}(c; t) = (1-c) \left(\log(1-c)t - \frac{(1-c)t}{2} \right) + \frac{t}{2} - \log t. \quad (4.9)$$

Fix $x \in [0, 1-c]$, then for a fixed $\Lambda \in \mathcal{N}(x)$

$$\begin{aligned} \inf_{\alpha \in \mathcal{M}_{\mathbb{N}_0}(1-c-x)} I(\Lambda, \alpha; t) &= I_{\text{Mi}}(\Lambda; t) \\ &+ c \left(\frac{t}{2} - \log t \right) + \left[(1-c-x) \log \frac{1-c-x}{1-e^{-t(1-c-x)}} + \frac{t}{2}(1-c-x)(c+x) \right], \end{aligned}$$

since the infimum is attained in $\alpha = \delta_{1-c-x}$, as proved in Lemma 4.1. Then, with the same procedure of Lemma 4.2, we see that the infimum over $\Lambda \in \mathcal{N}(x)$ is attained in

$$\lambda_k^*(x \wedge \frac{1}{t}; t) = (x \wedge \frac{1}{t}) \frac{e^{-(x \wedge \frac{1}{t})tk} \left((x \wedge \frac{1}{t}) t \right)^{k-1} k^{k-2}}{k!},$$

giving

$$\begin{aligned} \inf_{\substack{\alpha \in \mathcal{M}_{\mathbb{N}_0}(1-c-x) \\ \Lambda \in \mathcal{N}(x)}} I(\Lambda, \alpha; t) &= \left(x \log tx - \frac{tx}{2} + \frac{1}{2t} \right) \mathbb{1}(x < \frac{1}{t}) + (c+x) \left(\frac{t}{2} - \log t \right) - \frac{1}{2t} \\ &+ (1-c-x) \log \frac{1-c-x}{1-e^{-t(1-c-x)}} + \frac{t}{2}(1-c-x)(c+x). \end{aligned} \quad (4.10)$$

Minimizing then for $x \in [0, 1-c]$, we see that the infimum is attained in x^* , the smallest solution to

$$x^* = (1-c)e^{-t(1-c-x^*)},$$

which is $x^* = 1-c$, for all $t \geq \frac{1}{1-c}$ and $x^* < 1-c$ otherwise. By substituting the optimal x^* in (4.10), we see that (4.9) holds.

Now, notice that the procedure to get the upper bound in the proof of Proposition 3.1 implies in a straightforward way that

$$\lim_{\delta \downarrow 0} \limsup_{N \rightarrow \infty} \mathbb{P}_{N, \frac{1}{N}t_N} \left(\left| \overline{\text{Me}}_{R_N, \varepsilon_N}^{(N)}(t) - c \right| \leq \delta \right) \leq - \inf_{\substack{\Lambda \in \mathcal{N} \\ \alpha \in \mathcal{M}_{\mathbb{N}_0} \\ c_\Lambda + c_\alpha = 1-c}} I(\Lambda, \alpha; t) = -\mathcal{J}_{\text{Me}}(c; t).$$

In the same way, from the proof of Proposition 3.1, we borrow the strategy of constructing a “recovery sequence”, this time using $\Lambda^*(x^*; t)$ and $\alpha^* = \delta_{1-c-x^*}$ to construct $\ell^{(N)}$ as in (3.17). This gives

$$\lim_{\delta \downarrow 0} \liminf_{N \rightarrow \infty} \mathbb{P}_{N, \frac{1}{N}t_N} \left(\left| \overline{\text{Me}}_{R_N, \varepsilon_N}^{(N)}(t) - c \right| \leq \delta \right) \geq -\mathcal{J}_{\text{Me}}(c; t).$$

□

The proof of the second point in Corollary 1.4 follows as a direct consequence of Lemma 4.3.

4.2. Proof of Theorem 1.5. Item (1) follows by Lemma 4.1 and 4.2. Following the approach of those proofs, one can easily see that the order of minimization is not important, in particular:

$$\mathcal{J}_{\text{Mi}}(c; t) = \inf_{\Lambda \in \mathcal{N}(c)} \mathcal{I}_{\text{Mi}}(\Lambda; t) = \inf_{\alpha \in \mathcal{M}_{\mathbb{N}_0}(1-c)} \mathcal{I}_{\text{Ma}}(\alpha; t) = \mathcal{J}_{\text{Ma}}(1-c; t).$$

The minimizer, given a certain microscopic mass $c \in [0, 1]$, is seen to take the form

$$(\Lambda^*(c \wedge \tfrac{1}{t}; t), \delta_{1-c}),$$

where Λ^* is defined in (1.13) and $\alpha = \delta_{1-c}$ is a single macroscopic component. Imposing a certain microscopic (respectively macroscopic) mass influences the optimal configuration. Indeed, although it is optimal for the system to avoid mesoscopic mass (as seen in Corollary 1.4(2)), the impossibility of minimizing the microscopic configuration under a certain constraint on the mass, namely $c_\Lambda > \frac{1}{t}$, forces the system to actually have a mesoscopic mass of size $c_\Lambda - \frac{1}{t}$. The same happens when we impose a macroscopic mass which is too small, namely $c_\alpha < \frac{t-1}{t}$. The form of the function $\mathcal{J}_{\text{Mi}}(c; t)$ in (1.12) comes directly from such minimization procedures.

Let us now prove assertion (2). The form of the minimizing Λ follows from Lemma 4.2. Fix $t \in [0, 1]$. Then $\mathcal{J}_{\text{Mi}}(c; t) = c \log c - tc^2 + tc + (1-c) \log \frac{1-c}{1-e^{t(c-1)}}$ is strictly decreasing in $c \in [0, 1]$. Indeed

$$\frac{d}{dc} \mathcal{J}_{\text{Mi}}(c; t) = \log tc - tc + \frac{t(1-c)e^{-t(1-c)}}{1-e^{-t(1-c)}} - \log \frac{t(1-c)e^{-t(1-c)}}{1-e^{-t(1-c)}} = F\left(\frac{t(1-c)e^{-t(1-c)}}{1-e^{-t(1-c)}}\right) - F(tc),$$

where we introduced the function $F(x) = x - \log x$, which is decreasing in $x \in (0, 1]$. Hence, monotonicity of $\mathcal{J}_{\text{Mi}}(\cdot; t)$ in $[0, 1]$ follows from

$$tc \leq \frac{t(1-c)e^{-t(1-c)}}{1-e^{-t(1-c)}} \leq 1. \quad (4.11)$$

The first inequality follows by observing that the function $\phi_t(c) = e^{-t(1-c)} - c$ is nonnegative for all $c \in [0, 1 \wedge \frac{1}{t}]$, since $\phi_t(0) = e^{-t} > 0$, $\phi_t(1) = 0$, and ϕ_t is strictly decreasing in $[0, 1]$, since $t \leq 1$. The second inequality follows from the fact that $\psi(z) := 1 - e^{-z} - ze^{-z} \geq 0$ for all $z \in [0, 1]$ (substitute $z = t(1-c)$), since $\psi(0) = 0$, $\psi(1) = 1 - 2e^{-1} \geq 0$ and ψ is strictly increasing in $[0, 1]$. Therefore, $\mathcal{J}_{\text{Mi}}(\cdot; t)$ is minimized in $c = 1$, which implies the conclusion.

Now we turn to assertion (3). For $t \in (1, \infty)$, the derivative of $\mathcal{J}_{\text{Mi}}(c; t)$ writes as follows

$$\frac{d}{dc} \mathcal{J}_{\text{Mi}}(c; t) = \frac{t(1-c)e^{-t(1-c)}}{1-e^{-t(1-c)}} - \log \frac{t(1-c)e^{-t(1-c)}}{1-e^{-t(1-c)}} + \begin{cases} \log tc - tc & \text{for } c \leq \frac{1}{t}, \\ -1 & \text{for } c > \frac{1}{t}. \end{cases}$$

It is clear that $\mathcal{J}_{\text{Mi}}(c; t)$ is strictly increasing in $c \in (\frac{1}{t}, 1]$, while for $c \in [0, \frac{1}{t}]$, we need to go back to (4.11). The right inequality there is still true for any $c < \frac{1}{t}$. Since the quotient in (4.11) is strictly increasing in c and since $F(x) = x - \lg x$ is strictly convex in x , the unique zero of $\frac{d}{dc} \mathcal{J}_{\text{Mi}}(c; t)$ is given by the unique solution c of

$$tc = \frac{t(1-c)e^{-t(1-c)}}{1-e^{-t(1-c)}},$$

which is precisely the solution $c = \beta_t$ of (1.15). The remaining assertions follow.

ACKNOWLEDGEMENTS

The authors acknowledge three anonymous referees for their careful reviews and many suggestions for improving the exposition. This research has been funded by the Deutsche Forschungsgemeinschaft (DFG) through grant CRC 1114 “Scaling Cascades in Complex Systems”, Project C08.

REFERENCES

- [Ada08] S. Adams. Large deviations for empirical cycle counts of integer partitions and their relation to systems of Bosons. In *Analysis and stochastics of growth processes and interface models*, pages 148–172. Oxford Univ. Press, Oxford, 2008.
- [Ald97] D. J. Aldous. Brownian excursions, critical random graphs and the multiplicative coalescent. *Ann. Probab.*, 25(2):812–854, 1997.
- [Ald99] D. J. Aldous. Deterministic and stochastic models for coalescence (aggregation and coagulation): a review of the mean-field theory for probabilists. *Bernoulli*, 5(1):3–48, 02 1999.
- [AP98] D. J. Aldous and J. Pitman. Tree-valued Markov chains derived from Galton-Watson processes. *Annales de l’Institut Henri Poincaré (B) Probability and Statistics*, 34(5):637 – 686, 1998.
- [Aug18] F. Augeri. Nonlinear large deviation bounds with applications to traces of Wigner matrices and cycles counts in Erdős-Rényi graphs. *arxiv preprint 1810.01558*, 2018.
- [BC15] C. Bordenave and P. Caputo. Large deviations of empirical neighborhood distribution in sparse random graphs. *Probability Theory and Related Fields*, 163(1-2):149–222, 2015.
- [BJR07] B. Bollobás, S. Janson, and O. Riordan. The phase transition in inhomogeneous random graphs. *Random Structures & Algorithms*, 31(1):3–122, 2007.
- [Bol01] B. Bollobás. *Random Graphs*. Cambridge Studies in Advanced Mathematics. Cambridge University Press, 2 edition, 2001.
- [BP90] E. Buffet and J. V. Pulé. On Lushnikov’s model of gelation. *J. Statist. Phys.*, 58(5-6):1041–1058, 1990.
- [BP91] E. Buffet and J. V. Pulé. Polymers and random graphs. *J. Statist. Phys.*, 64(1-2):87–110, 1991.
- [CD16] S. Chatterjee and A. Dembo. Nonlinear large deviations. *Advances in Mathematics*, 299:396 – 450, 2016.
- [CD18] N. A. Cook and A. Dembo. Large deviations of subgraph counts for sparse Erdős-Rényi graphs. *arXiv preprint arXiv:1809.11148*, 2018.
- [Cha16] S. Chatterjee. An introduction to large deviations for random graphs. *Bull. Amer. Math. Soc. (N.S.)*, 53(4):617–642, 2016.
- [CV11] S. Chatterjee and S.R.S. Varadhan. The large deviation principle for the Erdős-Rényi random graph. *European Journal of Combinatorics*, 32(7):1000 – 1017, 2011. Homomorphisms and Limits.

- [DZ10] A. Dembo and O. Zeitouni. *Large deviations techniques and applications*, volume 38 of *Stochastic Modelling and Applied Probability*. Springer-Verlag, Berlin, 2010. Corrected reprint of the second (1998) edition.
- [EMH04] A. Engel, R. Monasson, and A. K. Hartmann. On large deviation properties of Erdős–Rényi random graphs. *J. Statist. Phys.*, 117(3-4):387–426, 2004.
- [ER60] P. Erdős and A. Rényi. On the evolution of random graphs. *Magyar Tud. Akad. Mat. Kutató Int. Közl.*, 5:17–61, 1960.
- [Flo41] P. J. Flory. Molecular size distribution in three dimensional polymers. III. Tetrafunctional branching units. *Journal of the American Chemical Society*, 63:3096–3100, 1941.
- [Gil72] D. T. Gillespie. The stochastic coalescence model for cloud droplet growth. *J. Atmos. Sci.*, 29:1496–1510, 1972.
- [LM04] P. Laurençot and S. Mischler. On coalescence equations and related models. In *Modeling and computational methods for kinetic equations*, pages 321–356. Springer, 2004.
- [Lus78] A. A. Lushnikov. Coagulation in finite systems. *Journal of Colloid and Interface Science*, 65(2):276–285, 1978.
- [Mar68] A. H. Marcus. Stochastic coalescence. *Technometrics*, 10:133–143, 1968.
- [MN14] M. Merle and R. Normand. Self-organized criticality in a discrete model for smoluchowski’s equation. *arXiv preprint arXiv:1410.8338*, 2014.
- [MPPR17] A. Mielke, R. Patterson, M. Peletier, and D. M. Renger. Non-equilibrium thermodynamical principles for chemical reactions with mass-action kinetics. *SIAM Journal on Applied Mathematics*, 77(4):1562–1585, 2017.
- [Nor99] J. R. Norris. Smoluchowski’s coagulation equation: Uniqueness, nonuniqueness and a hydrodynamic limit for the stochastic coalescent. *Annals of Applied Probability*, pages 78–109, 1999.
- [O’C98] N. O’Connell. Some large deviation results for sparse random graphs. *Probability theory and related fields*, 110(3):277–285, 1998.
- [Pit90] Boris Pittel. On tree census and the giant component in sparse random graphs. *Random Structures & Algorithms*, 1(3):311–342, 1990.
- [Puh05] A. A. Puhalskii. Stochastic processes in random graphs. *The Annals of Probability*, 33(1):337–412, 2005.
- [Ste70] V. E. Stepanov. On the probability of connectedness of a random graph $\mathcal{G}_m(t)$. *Theory of Probability & Its Applications*, 15(1):55–67, 1970.
- [vS16] M. von Smoluchowski. Drei Vorträge über Diffusion, Brownsche Molekularbewegung und Koagulation von Kolloidteilchen. *Physik. Zeitschr.*, XVII:585–599, 1916. The lectures do not have separate titles; coagulation is treated in lecture III, which starts on page 593.

WIAS, MOHRENSTRASSE 39, 10117 BERLIN

Email address: luisa.andreis@wias-berlin.de

TU BERLIN AND WIAS, MOHRENSTRASSE 39, 10117 BERLIN

Email address: wolfgang.koenig@wias-berlin.de

WIAS, MOHRENSTRASSE 39, 10117 BERLIN

Email address: robert.patterson@wias-berlin.de