Abstract: We investigate a probabilistic model for routing of messages in relay-augmented multihop ad-hoc networks, where each transmitter sends one message to the origin. Given the (random) transmitter locations, we weight the family of random, uniformly distributed message trajectories by an exponential probability weight, favouring trajectories with low interference (measured in terms of signal-to-interference ratio) and trajectory families with little congestion (measured in terms of the number of pairs of hops using the same relay). Under the resulting Gibbs measure, the system targets the best compromise between entropy, interference and congestion for a common welfare, instead of an optimization of the individual trajectories.

We describe the joint routing strategy in terms of the empirical measure of all message trajectories. In the limit of high spatial density of users, employing large-deviation theory, we derive the limiting free energy and analyze the optimal strategy, given as the minimizer(s) of a characteristic variational formula. Interestingly, expressing the congestion term requires introducing an additional empirical measure. Our results remain valid under replacing the two penalization terms by more general functionals of these two empirical measures.

MSC 2010. 60F10, 60G55, 60K30; 82B21; 90B15

Keywords and phrases. Gibbs distribution of trajectories, high-density limit, large deviations, empirical measure, number of incoming messages, weak topology, variational formula, minimizer, multihop ad-hoc network, message trajectories, signal-to-interference ratio, congestion

1. Introduction and main results

1.1. Background and main goals. In random networks, one of the prominent problems is the question how to conduct a message through the system in an optimal way. Optimality is often measured in terms of determining the shortest path from the transmitter to the recipient, or, if interference is considered, determining the path that yields the least interference. If many messages are considered at the same time, an additional aspect of optimality may be to achieve a minimal amount of congestion.

Many investigations concern the question just for one single transmitter/recipient pair, which is a question that every single participant faces. However, a strategy found in such a setting may lead to a selfish routing, and it is quite likely that the totality of all these routings for all the individuals is by far not optimal for the community of all the users [CCS16 Section 1]. Instead, the entire system...
may work even better if an optimal *compromise* is realized, by which we mean a joint strategy that leads to an optimum for the entire system, though possibly not for every participant.

In this paper, we present a probabilistic ansatz for describing a jointly optimal routing for an unbounded number of transmitter/recipient pairs, which takes into account the following three crucial properties of the family of message trajectories: entropy, interference and congestion. That is, we consider a situation in which all the messages are directed through the system in a random way, such that each hop prefers a low interference, and such that the total amount of congestion is preferred to be low. Parameters control the strengths of influence of the three effects.

Let us describe our model in words. Let the users be located randomly as the sites of a Poisson point process, which we fix. Each user sends out precisely one message, which arrives at the (unique) base station, which is located at the origin. We consider the entire collection of possible trajectories of the messages through the system. We employ an ad-hoc relaying system with multiple hops, such that all the users act as relays for the handoffs of the messages. The maximal number of hops is \(k_{\text{max}} \in \mathbb{N}\) for each message. Each \(k\)-step message trajectory (with \(k \in \{1, \ldots, k_{\text{max}}\}\) itself random) is random and *a priori* uniformly distributed. The family of all trajectories is *a priori* independent.

Now, the probability distribution of this family is given in terms of a Gibbs ansatz by introducing two exponential weight terms. (That is, we introduce a quenched measure on trajectories given the location of the users.) The first one weights the total amount of interference, measured in terms of the signal-to-interference ratio for each hop. The second one weights the total congestion, i.e., the number of times that any two trajectories use the same relay. Under the arising measure, there is a competition between all the three decisive effects of the trajectory family: entropy, interference and congestion. Furthermore, the users form a random environment for the family, which not only determines the starting sites of all the trajectories, but also has a decisive effect on interference and congestion. While the latter has a smoothing effect on the fine details of the spatial distribution of all the trajectories, the effect of the former is not so clear to predict, as the superposition of signals have a very non-local influence.

Our main interest is in understanding the spatial distribution of the totality of all the message trajectories under the Gibbs distribution. The measure under consideration is a highly complex object, as it depends on all the user locations and on many detailed properties and quantities. However, we make a substantial step towards a thorough understanding by deriving an asymptotic formula for the logarithmic behaviour of the normalization constant in the limit of a high spatial density of the users. The limiting situation is then described in terms of a large-deviation rate function and a variational formula, whose minimizers describe the optimal joint choices of the trajectories. This formula is deterministic and depends only on general spatial considerations, not on the individual users. These are our main results in this paper.

The main objects in terms of which we achieve this description are the empirical measures of the trajectories of the messages sent out by the users, disintegrated with respect to the lengths and rescaled to finite asymptotic size. These measures turn out to converge in the weak topology in the high-density limit that we consider in this paper. The counting complexity of the statistics of the message trajectories can be written in terms of multinomial expressions and afterwards, in the limit of finer and finer decompositions of the space, approximated in terms of relative entropies, using Stirling’s formula. The interference term can also be handled in a standard way [HJKP15], since it is a continuous function of the collection of empirical measures of message trajectories.

However, a key finding of our work is that the congestion term is a highly discontinuous function of these empirical measures. Indeed, one cannot express its limiting behaviour in terms of these measures. Instead, one needs to substantially enlarge the probability space of trajectories and to introduce another collection of empirical measures, the ones of the locations of users (relays) who receive given numbers of incoming messages. The congestion expression then turns out to be a lower
A GIBBSIAN MODEL FOR MESSAGE ROUTING

A semi-continuous function of these empirical measures, and hence the limiting congestion term is still expressible in terms of the weak limits of these measures. Again, using explicit combinatorics and Stirling’s formula, we arrive at explicit entropic terms describing the statistics of these measures. The two families of these crucial empirical measures together enable us to describe all the properties of the message trajectories that we are interested in. We establish a full large-deviation principle for the tuple of all these measures with an explicit rate function and obtain in particular their convergence towards the minimizer(s) of a characteristic variational formula. We also derive their positivity properties and characterize them in terms of Euler-Lagrange equations. Unfortunately, due to the complexity of the congestion term, we are not able to decide about the uniqueness of the minimizer.

However, in the special case when congestion is not penalized, the minimizer turns out to be unique, and we obtain an explicit expression that is amenable to further investigation. In certain limiting regimes, we can derive a good understanding of decisive quantities of the system, like the typical number of hops, the typical length of a hop and the typical shape of a trajectory as a function of the distance between the transmission site and the origin. We expect that such properties of the system are similar if also congestion is penalized, as the effect of congestion is not spatial. We decided to analyze such questions in a separate work [KT17b], as they have a strongly analytic, rather than probabilistic, nature.

The main purpose of the present paper is to provide the mathematical framework of a large-deviation theory for the quenched trajectory distribution, given the user locations, in the high-density limit. A discussion about the relevance of the model for traffic and telecommunication theory is deferred to [KT17b]. Therefore, we will formulate the model in a more general, slightly abstract, way in order to bring its mathematical essence to the surface. That is, we consider a random complete geometric graph in a compact subset of \( \mathbb{R}^d \) (in the special case, the set of users), and a distribution of trajectories that has an interaction (the interference) with all the locations of the nodes and suppresses local clumping (the congestion).

1.1.1. Related literature. Apart from the potential value for understanding a new type of message routing models in telecommunication, the present paper provides also some interesting mathematical research on topological fine properties of random paths in random environment in a high-density setting, a subject that received a lot of interest for various types of such processes over the decades. We remind the reader on a number of investigations of the intersection properties of random walks and Brownian motions (both self-intersections and mutual intersections) in highly dense settings, see the monograph [Ch09] and some particular investigations in [KM02, KM13]; in all these works, one is interested in large-deviation properties of suitable empirical measures, and the lack of continuity of the path properties is the main difficulty. Let us mention that the main aspect of the approach in [KM02] is the same as in the present paper: an approximation of combinatorics in finer and finer decompositions of the space by entropic terms. Another line of research in which similar questions arise is a mean-field variant of a spatial version of Bose-Einstein statistics, like in [AK08], where the statistics of the empirical measures of a diverging number of Brownian bridges with symmetrized initial-terminal condition is analyzed in terms of a large-deviation principle in the weak topology. While [AK08] works with the same method as we in the present paper (spatial discretization with limiting fineness), [T08] showed that a method based entirely on the notion of entropy is able to derive such results in a more general framework.

1.1.2. Organization of the remainder of this paper. We introduce the model and necessary notation in Section 1.2, present our main results in Sections 1.3 (the limiting free energy of the model), 1.4 (the description of the minimizer(s)), 1.5 (the large deviation principle and the convergence of the empirical measures), and 1.6 (results in case congestion is not penalized), and in Section 1.7 we discuss and comment our findings. The remaining sections are devoted to the proofs: in Section 2 we prepare for
the proofs by introducing our methods and deriving formulas for the probability terms, in Section 3.5 we put all this together to a proof of the limiting free energy, the large deviation principle and the convergence of the empirical measures, in Section 4 we analyze the minimizer(s) of the characteristic variational formula, and in Section 5.2 we extend the proofs to the case when congestion is not penalized.

1.2. The Gibbsian model. We introduce now the mathematical setting. For any \( n \in \mathbb{N} \) and for any measurable subset \( V \) of \( \mathbb{R}^n \), let \( \mathcal{M}(V) \) denote the set of all finite nonnegative Borel measures on \( V \), which we equip with the weak topology. We are working in \( \mathbb{R}^d \) with some fixed \( d \in \mathbb{N} \). Our model is defined as follows. Let \( W \subset \mathbb{R}^d \) be compact, the territory of our telecommunication system, containing the origin \( o \) of \( \mathbb{R}^d \).

1.2.1. Users. Let \( \mu \in \mathcal{M}(W) \) be an absolutely continuous measure on \( W \) with \( \mu(W) > 0 \). Note that we do not require that \( \text{supp}(\mu) = W \). For \( \lambda > 0 \), we denote by \( X^\lambda \) a Poisson point process in \( W \) with intensity measure \( \lambda \mu \). They points \( X_i \in X^\lambda \) are interpreted as the locations of the users in the system, while the origin \( o \) of \( \mathbb{R}^d \) is the single base station. We assume that \( X^\lambda = \{X_i: i \in I^\lambda\} \) with \( I^\lambda = \{1, \ldots, N(\lambda)\} \) and \( (N(\lambda))_{\lambda>0} \) a standard Poisson process on \( \mathbb{N} \) and \((X_i)_{i\in \mathbb{N}} \) is an i.i.d. sequence of \( W \)-distributed random variables with distribution \( \mu(\cdot)/\mu(W) \) defined on one probability space \((\Omega, \mathcal{F}, \mathbb{P})\). Since \( \mu \) has a density, all points \( X_i \) are mutually different with probability one. Furthermore, \( X^\lambda \) is increasing in \( \lambda \), and its empirical measure, normalized by \( 1/\lambda \),

\[
L^\lambda = \frac{1}{\lambda} \sum_{i \in I^\lambda} \delta_{X_i},
\]

converges towards \( \mu \) almost surely as \( \lambda \to \infty \).

These assumptions on the users can be relaxed, see Section 1.7.4.

1.2.2. Message trajectories. We now introduce the collection of trajectories sent out from the users to \( o \), i.e., for uplink communication. For any \( i \in I^\lambda \), we call a vector of the form

\[
S^i = (S^i_{-1} = K_i, S^i_0 = X_i, S^i_1 \in X^\lambda, \ldots, S^i_{K_i-1} \in X^\lambda, S^i_{K_i} = o) \in \mathbb{N} \times (\bigcup_{k \in \mathbb{N}} W^k) \times \{o\},
\]

a message trajectory from \( X_i \) to \( o \) with \( K_i \) hops. That is, \( S^i \) starts from \( X_i \) and ends in \( o \) after \( K_i \) hops from user to user in \( X^\lambda \). Hence, each user sends exactly one message to \( o \) and the users receive the function of a relay. We fix a number \( k_{\text{max}} \in \mathbb{N} \) and write \( S^i_{k_{\text{max}}} (X^\lambda) \) for the set of all possible realizations of the random variable \( S^i \) with \( K_i \leq k_{\text{max}} \), i.e., with no more than \( k_{\text{max}} \) hops. Hence, elements \( s^i = (s^i_{-1}, s^i_0, s^i_1, \ldots, s^i_{K_i-1}, s^i_{K_i}) \) of \( S^i_{k_{\text{max}}} (X^\lambda) \) satisfy \( s^i_{-1} \in \{1, \ldots, k_{\text{max}}\} \), \( s^i_0 = X_i \) and \( s^i_{K_i} = o \). We write \( S_{k_{\text{max}}} (X^\lambda) = \prod_{i \in I^\lambda} S^i_{k_{\text{max}}} (X^\lambda) \) for the set of all possible realizations of the families \( S^\lambda = (S^i)_{i \in I^\lambda} \). We use the notation \([k] = \{1, \ldots, k\}\) for \( k \in \mathbb{N} \).

Given \( i \in I^\lambda \), we consider each trajectory \( S^i \) in (1.2) as an \( S^i_{k_{\text{max}}} (X^\lambda) \)-valued random variable. Its a priori measure is given by the formula

\[
s^i \mapsto \frac{1}{N(\lambda)^{s^i_{-1}-1}}, \quad s^i \in S^i_{k_{\text{max}}} (X^\lambda).
\]

That is, its restriction to \( \{s^i \in S^i_{k_{\text{max}}} (X^\lambda): s^i_{-1} = k\} \) is the uniform distribution for any \( k \in [k_{\text{max}}] \), and its total mass is equal to \( k_{\text{max}} \). Recall that it fixes the starting point \( X_i \) and the terminal point \( o \).

Under our joint reference measure, all the trajectories are independent; indeed it gives the value

\[
s = (s^i)_{i \in I^\lambda} \mapsto \prod_{i \in I^\lambda} \frac{1}{N(\lambda)^{s^i_{-1}-1}}
\]

to the configuration \( s \in S_{k_{\text{max}}} (X^\lambda) \). Thus, it gives a total mass of \( k_{\text{max}}^{N(\lambda)} \) to \( S_{k_{\text{max}}} (X^\lambda) \).
1.2.3. Gibbsian trajectory distribution. In this section, we define the central object of this study: a Gibbs distribution on the set of collections of trajectories. After providing the abstract definitions related to this, in Section 1.2.4 we sketch the key example that is relevant for application in telecommunications. The general conditions on the ingredients of the Gibbs distribution in this section arise naturally from the properties of this example.

We introduce the following notation. For \( k \in \mathbb{N} \), elements of the product space \( W^k = \{0,1,\ldots,k-1\} \) are denoted as \((x_0,\ldots,x_{k-1})\). For \( i = 0,\ldots,k-1 \), the \( i \)-th marginal of a measure \( \nu_k \in \mathcal{M}(W^k) \) is denoted by \( \pi_i \nu_k \in \mathcal{M}(W) \), i.e., \( \pi_i \nu_k(A) = \nu_k(W^i \times A \times W^{i+1,\ldots,k-1}) \) for any Borel set \( A \subseteq W \).

For fixed \( k \in [k_{\max}] \) and for a collection of trajectories \( s \in S_{k_{\max}}(X^\lambda) \), we define

\[
R_{\lambda,k}(s) = \frac{1}{\lambda} \sum_{i \in I^\lambda: s^i_{k-1} = k} \delta_{(s^0_i,\ldots,s^i_{k-1})},
\]

the empirical measures of all the \( k \)-hop trajectories, which is an element of \( \mathcal{M}(W^k) \). By the assumption that each user sends out exactly one message, we have

\[
\sum_{k=1}^{k_{\max}} \pi_0 R_{\lambda,k}(s) = L_\lambda.
\]

For \( k \in [k_{\max}] \), we choose a continuous function \( f_k: \mathcal{M}(W) \times W^k \to \mathbb{R} \) that is bounded from below. Using (1.6), we put

\[
\mathfrak{S}(s) = \lambda \sum_{k=1}^{k_{\max}} \left\langle R_{\lambda,k}(\cdot), f_k(L_\lambda, \cdot) \right\rangle,
\]

where we write \( \left\langle \nu, f \right\rangle \) for the integral of the function \( f \) against the measure \( \nu \). Moreover, we define

\[
m_i(s) = \sum_{j \in I^\lambda} \sum_{l=1}^{s^i_{l-1}-1} \mathbb{I}\{s^i_{j} = s^i_{0}\}, \quad i \in I^\lambda,
\]

as the number of incoming hops into the user (relay) \( s^i_{0} = X_i \) of any of the trajectories.

We pick a function \( \eta: \mathbb{N}_0 \to \mathbb{R} \), bounded from below such that \( \lim_{n \to \infty} \eta(n)/n = \infty \). Then we put

\[
\mathfrak{M}(s) = \sum_{i \in I^\lambda} \eta(m_i(s)).
\]

Now, we define

\[
P_{\gamma,\lambda,X^\lambda}^{\gamma,\beta}(s) := \frac{1}{Z_{\lambda}^{\gamma,\beta}(X^\lambda)} \left( \prod_{i \in I^\lambda} \frac{1}{N(\lambda)^{s^i_{l-1}-1}} \right) \exp \left\{ -\gamma \mathfrak{S}(s) - \beta \mathfrak{M}(s) \right\},
\]

where \( \gamma > 0 \) and \( \beta > 0 \) are parameters. This is the Gibbs distribution with independent reference measure given in (1.4), subject to two exponential weights with the terms (1.7) and (1.9). Here

\[
Z_{\lambda}^{\gamma,\beta}(X^\lambda) = \sum_{r \in S_{k_{\max}}(X^\lambda)} \left( \prod_{i \in I^\lambda} \frac{1}{N(\lambda)^{r^i_{l-1}-1}} \right) \exp \left\{ -\gamma \mathfrak{S}(r) - \beta \mathfrak{M}(r) \right\}
\]

is the normalizing constant, which we will refer to as partition function. Note that \( P_{\gamma,\lambda,X^\lambda}^{\gamma,\beta}(\cdot) \) is random conditional on \( X^\lambda \), and it is a probability measure on \( S_{k_{\max}}(X^\lambda) \). In the jargon of statistical mechanics, it is a quenched measure, which we will consider almost surely with respect to the process \((X^\lambda)_{\lambda>0}\). In the annealed setting, one would average out over \((X^\lambda)_{\lambda>0}\), see Section 1.7.5.
1.2.4. The key example: penalization of interference and congestion. In this section, we sketch the most important example for the exponents $\mathcal{G}$ and $\mathfrak{M}$ in (1.10), where $\mathcal{G}$ registers interference and $\mathfrak{M}$ expresses congestion in a telecommunication network. Analyzing the qualitative properties of the network with this choices of $\mathfrak{M}$ and $\mathcal{G}$ will be the main topic of our work [KT17b].

Now we introduce interference. We choose a path-loss function, which describes the propagation of signal strength over distance. This is a monotone decreasing, continuous function $\ell: [0, \infty) \rightarrow (0, \infty)$. An example used in practice is $\ell(r) = \min\{1, r^{-\alpha}\}$, for some $\alpha > 0$, see e.g. [GT08 Section II.], for further examples see [BB09 Section 2.3.1]. The signal-to-interference ratio (SIR) [HJKP15] of a transmission from $X_i \in X^\lambda$ to $x \in W$ in the presence of the users in $X^\lambda$ is given as
\[
\text{SIR}(X_i, x, L_\lambda) = \frac{\ell(|X_i - x|)}{\sum_{o \in I} \ell(|X_j - x|)}.
\]

We call the denominator of the r.h.s of (1.12) the interference. Now, given a trajectory configuration $s = (s^i)^{i \in I^\lambda} \in \mathcal{S}_{k_{\text{max}}}^\lambda$, we put
\[
\mathcal{G}(s) = \sum_{i \in I^\lambda} \sum_{l=1}^{s_i} \text{SIR}(s^i_{l-1}, s^i_l, L_\lambda)^{-1} = \lambda \sum_{k=1}^{k_{\text{max}}} \int_{W^k} R_{\lambda,k}(dx_0, \ldots, dx_{k-1}) \int_W \ell(|y-x_i|) L_\lambda(dy) \frac{\ell(|y-x_i|) L_\lambda(dy)}{\ell(|x_{l-1} - x_l|)}, \quad x_k = o.
\]

Then (1.13) is a special case of (1.7) with
\[
f_k(\nu, x_0, \ldots, x_{k-1}) = \int_{W^k} \ell(|y-x_i|) \ell(dy) \frac{\ell(|y-x_i|) \nu(dy)}{\ell(|x_{l-1} - x_l|)}, \quad x_k = o, k \in [k_{\text{max}}].
\]

Next, we introduce congestion. We define $\eta(n) = n(n-1)$, and, as in (1.9), we put
\[
\mathfrak{M}(s) = \sum_{i \in I^\lambda} \eta(m_i(s)) = \sum_{i \in I^\lambda} m_i(s)(m_i(s) - 1), \quad s \in \mathcal{S}_{k_{\text{max}}}^\lambda.
\]

Note that $\eta(s) = m_i(s)(m_i(s) - 1)$ is the number of ordered pairs of hops arriving at the relay $X_i = s^i_0$.

In the downlink scenario, instead of users sending messages to the base station, the base station sends exactly one message to each of the users, using the same relaying rules. One can define a Gibbsian model analogously, now for trajectories from $o$ to $X_i$ instead of the other way around. For this, the interference term and the congestion term have to be redefined in an obvious way. We are certain that analogues of all our results are true and can be proved in the same way, hence we abstained from spelling them out.

1.3. The limiting free energy. The main goal of this paper is the description of this model in the limit $\lambda \rightarrow \infty$ in the quenched setting. Our first result describes the limiting free energy, i.e., the exponential behaviour of the partition function $Z_{\lambda}^{\beta,\beta}(X^\lambda)$. One expects that this is entirely governed by the large-deviation behaviour of the empirical measures $\{(R_{\lambda,k}(S))_{k \in [k_{\text{max}}]}\}_{\lambda > 0}$. This expectation is supported by the fact that, for $i \in I^\lambda$, we can express $m_i(s)$ defined in (1.8) in terms of $(R_{\lambda,k}(s))_{k \in [k_{\text{max}}]}$ as follows
\[
m_i(s) = \lambda \sum_{k=1}^{k_{\text{max}}} \sum_{l=1}^{s_i} \pi_l R_{\lambda,k}(\{s^i_l\}).
\]

Surprisingly, it turns out that the limiting free energy cannot be described entirely in terms of these measures. The reason is that the function in (1.16) that maps them onto $m_i(s)$ is highly discontinuous in the limit $\lambda \rightarrow \infty$; even a proper formulation of such continuity would be awkward since both $i$ and $s$ depend on $\lambda$. 

One therefore needs to substantially extend the probability space and to choose an additional family of empirical measures such that \((m_i(s))_{i \in I^\lambda}\) can be written as a (lower semi-)continuous function of these measures in the limit \(\lambda \to \infty\). A natural choice of such a family is the one of the measures

\[
P_{\lambda,m}(s) = \frac{1}{\lambda} \sum_{i \in I^\lambda: m_i(s) = m} \delta_{s^i_0}, \quad m \in \mathbb{N}_0. \tag{1.17}\]

Then for \(m \in \mathbb{N}_0\), \(P_{\lambda,m}(s) \in \mathcal{M}(W)\) is the empirical measure of the users \(s^i_0\) whose number of incoming messages equals \(m\). For any \(s \in S_{k_{\text{max}}}^\lambda (X^\lambda)\) the following hold

\[
(i) \quad \sum_{k=1}^{k_{\text{max}}} \pi_0 R_{\lambda,k}(s) = L_\lambda, \quad (ii) \quad \sum_{m=0}^\infty P_{\lambda,m}(s) = L_\lambda, \quad (iii) \quad \sum_{m=0}^\infty m P_{\lambda,m}(s) = \sum_{k=1}^{k_{\text{max}}} \sum_{l=1}^{k-1} \pi_l R_{\lambda,k}(s). \tag{1.18}\]

Condition (i) expresses our assumption that each user transmits precisely one message, (ii) says that the relays can be calculated in two ways: according to the number of incoming hops and according to the index of the hop of a trajectory that uses it. Moreover, we can write (1.9) in terms of \((P_{\lambda,m})_{m \in \mathbb{N}_0}\) as follows

\[
\mathfrak{M}(s) = \sum_{i \in I^\lambda} \eta(m_i(s)) = \lambda \sum_{m=0}^\infty \eta(m) P_{\lambda,m}(s)(W).
\]

We note that the function \(\mathcal{M}(W)^{\mathbb{N}_0} \to \mathbb{R} \cup \{\infty\}, (\xi_k)_{m \in \mathbb{N}_0} \mapsto \sum_{m=0}^\infty \eta(m) \xi_k(W)\) is lower semicontinuous, and even continuous on \(\{(\xi_k)_{m \in \mathbb{N}_0} : \sum_{m=0}^\infty \eta(m) \xi_k(W) \leq \alpha\}\) for any \(\alpha \in \mathbb{R}\).

The limiting free energy will be described in terms of the following kind of families of measures, which will turn out to be subsequential limits of the families \((\mathcal{R}_{\lambda,k}(S))_{k=1}^{k_{\text{max}}}, (P_{\lambda,m}(S))_{m=0}^\infty\) in the quenched limit \(\lambda \to \infty\).

**Definition 1.1.** An admissible trajectory setting is a collection of measures \(\Psi = ((\nu_k)_{k=1}^{k_{\text{max}}}, (\mu_m)_{m=0}^\infty)\) with \(\nu_k \in \mathcal{M}(W^k)\) for all \(k\) and \(\mu_m \in \mathcal{M}(W)\) for all \(m\), satisfying the following properties.

\[
(i) \quad \sum_{k=1}^{k_{\text{max}}} \pi_0 \nu_k = \mu, \quad (ii) \quad \sum_{m=0}^\infty \mu_m = \mu, \quad (iii) \quad M := \sum_{m=0}^\infty m \mu_m = \sum_{k=1}^{k_{\text{max}}} \sum_{l=1}^{k-1} \pi_l \nu_k. \tag{1.19}\]

The measure \(\nu_k\) is the measure of the \(k\)-step trajectories and \(\mu_m\) the measure of the users that receive precisely \(m\) incoming hops; note that there is no reason that they be normalized (like for \(\mu\)). Observe that in (1.18), \(L_\lambda, R_{\lambda,k}(s)\) and \(P_{\lambda,m}(s)\) play the role of \(\nu_k\) and \(\mu_m\), respectively. In particular, after replacing \(\mu\) by \(L_\lambda, ((R_{\lambda,k})_{k \in [k_{\text{max}}]}, (P_{\lambda,m})_{m \in \mathbb{N}_0})\) satisfies the definition of an admissible trajectory setting. See Section [1.7] for more explanations and interpretations, moreover for a modified version of our model where the assumption (i) is relaxed. By

\[
\mathcal{H}_V(\nu \mid \overline{\nu}) = \begin{cases} 
\int_V \nu \log \frac{d\nu}{d\overline{\nu}} - \nu(V) + \overline{\nu}(V), & \text{if the density } \frac{d\nu}{d\overline{\nu}} \text{ exists}, \\
+\infty, & \text{otherwise},
\end{cases} \tag{1.20}
\]

we denote the relative entropy [GZ93, Section 2.3] of a Borel measure \(\nu\) with respect to another Borel measure \(\overline{\nu}\) on a measurable set \(V\).

For an admissible trajectory setting \(\Psi = ((\nu_k)_{k=1}^{k_{\text{max}}}, (\mu_m)_{m=0}^\infty)\) we define

\[
S(\Psi) = \sum_{k=1}^{k_{\text{max}}} \int_{W^k} d\nu_k \tilde{f}_k, \quad \text{where } \tilde{f}_k(x_0, \ldots, x_{k-1}) = f_k(\mu, x_0, \ldots, x_{k-1}), \tag{1.21}
\]

and

\[
M(\Psi) = \sum_{m=0}^\infty \eta(m) \mu_m(W) \tag{1.22}
\]
and
\[ I(\Psi) = \sum_{k=1}^{k_{\max}} \mathcal{H}_W(\nu_k | \mu \otimes M^\otimes(k-1)) + \sum_{m=0}^{\infty} \mathcal{H}_W(\mu_m | \mu c_m) + \mu(W) \left(1 - \sum_{k=1}^{k_{\max}} M(W)^{k-1}\right) - \frac{1}{e}, \quad (1.23) \]

where we recall \( M = \sum_{m \in \mathbb{N}_0} m \mu_m \) from Definition 1.1(iii), \( \eta \) defined before (1.9) and \( c_m = \exp(-1/(e\mu(W))) \) are the weights of the Poisson distribution with parameter \( 1/(e\mu(W)) \). Note that according to (i) and (iii) in (1.9), we have \( M(W) \leq (k_{\max} - 1) \mu(W) \). From the representation in (1.31) below, one easily sees that \( I(\Psi) \) is well-defined as an element of \((-\infty, \infty)\) and \( \Psi \mapsto I(\Psi) \) is a lower semicontinuous function that is bounded from below. A tedious but elementary calculation shows that \( I \) is convex. In Section 1.5, \( I \) will turn out to govern the large deviations of the trajectory configuration.

We fix all the parameters \( W, \mu, f_k, \eta, k_{\max}, \gamma \) and \( \beta \) of the model. Our first main result is the following.

**Theorem 1.2** (Quenched exponential rate of the partition function). For \( \mathbb{P} \)-almost all \( \omega \in \Omega \),
\[ \lim_{\lambda \to \infty} \frac{1}{\lambda} \log Z_\lambda^\gamma(\lambda(\omega)) = - \inf_{\Psi \text{ admissible trajectory setting}} \left(I(\Psi) + \gamma S(\Psi) + \beta M(\Psi)\right). \tag{1.24} \]

See Section 1.7 for a discussion and Section 3.4 for the proof.

### 1.4. Description of the minimizers

From the variational formula in (1.24), descriptive information about the typical behaviour of the network can be deduced, especially in the case of the specific choice of \( \mathfrak{M} \) and \( \mathcal{S} \) from Section 1.2.4, see Sections 1.5 and 1.7. Hence, it is important to derive the Euler-Lagrange equations and to characterize the minimizers in most explicit terms. Our main results in this respect are the following. Note that the case \( k_{\max} = 1 \) is trivial.

**Proposition 1.3** (Characterization of the minimizer(s)). Let \( k_{\max} > 1 \). The infimum in the variational formula in (1.24) is attained, and every minimizer \( \Psi = ((\nu_k)_{k=1}^{k_{\max}}, (\mu_m)_{m=0}^{\infty}) \) has the following form.
\[ \nu_k(dx_0, \ldots, dx_{k-1}) = \mu(dx_0) A(x_0) \prod_{l=1}^{k-1} (C(x_l)M(dx_l)) e^{-\gamma \tilde{f}_k(x_0, \ldots, x_{k-1})}, \quad k \in [k_{\max}], \tag{1.25} \]
\[ \mu_m(dx) = \mu(dx) B(x) \left(\frac{(C(x)\mu(W))}{m!}\right)^m e^{-\beta \eta(m)}, \quad m \in \mathbb{N}_0, \tag{1.26} \]

where \( A, B, C : W \to [0, \infty) \) are functions such that the conditions in (1.19) are satisfied.

The proof of Proposition 1.3 is in Section 4.

While explicit formulas for the functions \( A \) and \( B \) can, given the function \( C \), easily be derived from (i) and (ii) in (1.19) (see (4.10)), the condition for \( C \) coming from (iii) is deeply involved and cannot be easily solved intrinsically; see (4.12) – (4.14). We have no argument for its existence to offer other than via proving the existence of a minimizer \( \Psi \) and deriving the Euler-Lagrange equations. By convexity of \( I, S \) and \( M \), every solution \( \Psi \) to these equations is a minimizer. Even the uniqueness of \( C \) is unknown to us. The equations (1.25)–(1.26) become more explicit in case \( \beta = 0 \), and in this case, uniqueness of the minimizer holds, see Section 1.6.

In case \( k_{\max} = 1 \), the only admissible trajectory setting is \( \Psi = (\nu_1, (\mu_m)_{m \in \mathbb{N}_0}) \) with \( \mu_0 = \nu_1 = \mu \) and \( \mu_m = 0 \) otherwise, therefore this \( \Psi \) minimizes (1.24). Thus, the limiting free energy has value \( -\gamma \int_W \mu(dx_0) \tilde{f}_1(x_0) \).
1.5. Large deviations for the empirical trajectory measure. Actually, the minimizers of the variational formula in (1.24) receive a rigorous interpretation in terms of important objects that describe the network. As we have already mentioned, the family of empirical measures satisfies the definition of an admissible trajectory setting, apart from the fact that in Definition 1.5 \( \mu \) has to replaced by \( L_\lambda \) everywhere, where we recall that \( L_\lambda \) converges to \( \mu \) almost surely as \( \lambda \to \infty \). According to our remarks after Definition 1.1 \( R_{\lambda,k}(s) \) and \( P_{\lambda,m}(s) \) play the roles of \( \nu_k \) and \( \mu_m \), respectively, in an admissible trajectory setting, which explains this term. Furthermore, for \( s \in S_{k_{\max}}(X^\lambda) \), we can express the term \( \mathfrak{M}(s) \) as

\[
\mathfrak{M}(s) = \lambda M(\Psi_\lambda(s)).
\]

Moreover, for the continuous penalization term we have

\[
\mathcal{G}(s) \approx \lambda S(\Psi_\lambda(s)),
\]

where we typically do not have an equality, because we have \( \mathcal{G}(s) = \lambda \sum_{k=1}^{k_{\max}} \int_{W^k} dR_{\lambda,k}(s)f_k(L_\lambda,\cdot) \), which is usually not equal to \( \lambda S(\Psi_\lambda(s)) = \lambda \sum_{k=1}^{k_{\max}} \int_{W^k} dR_{\lambda,k}(s)f_k(\mu,\cdot) \). However, since \( L_\lambda \to \mu \) almost surely, this difference vanishes in the limit, see Proposition 3.2.

We consider now the distribution of \( \Psi_{\lambda}(s) \) with \( S \) distributed under the product reference measure introduced in (1.4), normalized to a probability measure, \( \mathbb{P}^0_{\lambda,X^\lambda} \); note that the normalization \( Z^0_{\lambda}(X^\lambda) \) is equal to \( k_{\max}^{N(\lambda)} \). Our next main result, Theorem 1.4, is a large-deviation principle (LDP; see (1.30)–(1.31)) and the convergence towards the minimizers of the variational formula.

**Theorem 1.4** (LDP and convergence for the empirical measures). The following statements hold almost surely with respect to \( (X^\lambda)_{\lambda>0} \).

(i) The distribution of \( \Psi_{\lambda}(S) \) under \( \mathbb{P}^0_{\lambda,X^\lambda} \) satisfies an LDP as \( \lambda \to \infty \) with scale \( \lambda \) on the set

\[
\mathcal{A} = \left( \prod_{k=1}^{k_{\max}} \mathcal{M}(W^k) \right) \times \mathcal{M}(W)^{\mathbb{N}_0}
\]

with rate function given by \( \mathcal{A} \ni \Psi \mapsto I(\Psi) + \mu(W)\log k_{\max} \), which we define as \( \infty \) if \( \Psi \) is not an admissible trajectory setting.

(ii) For any \( \gamma, \beta \in (0, \infty) \), the distribution of \( \Psi_{\lambda}(S) \) under \( \mathbb{P}^\gamma_{\lambda,X^\lambda} \) converges towards the set of minimizers of the variational formula in (1.24).

For the reader’s convenience, we recall that the LDP states that the rate function \( I + \mu(W)\log k_{\max} \) is lower semicontinuous and

\[
\limsup_{\lambda \to \infty} \frac{1}{\lambda} \log \mathbb{P}^0_{\lambda,X^\lambda}(\Psi_{\lambda}(S) \in F) \leq - \inf_F (I + \mu(W)\log k_{\max}),
\]

\[
\liminf_{\lambda \to \infty} \frac{1}{\lambda} \log \mathbb{P}^0_{\lambda,X^\lambda}(\Psi_{\lambda}(S) \in G) \geq - \inf_G (I + \mu(W)\log k_{\max}),
\]

for any closed set \( F \) and any open set \( G \) in \( \mathcal{A} \). See [DZ98] for more on large deviation theory. On \( \mathcal{A} \), we consider the product topology that is induced by weak convergence in each factor; this is equal to coordinatewise weak convergence, see Section 3.5 for more details. Convergence of a distribution towards a set is defined by requiring that for any neighbourhood of the set, the probability of not being in the neighbourhood vanishes.

The proof of Theorem 1.4(i) is carried out in Section 3.5 using Lemma 4.1. Assertion (ii) is a simple consequence of (i), since the functionals \( S \) and \( M \) are bounded and continuous on the set \( B_C = \{ \Psi \in \mathcal{A} : M(\Psi) \leq C \} \) for any \( C \), and \( B_C \) is compact in \( \mathcal{A} \) (see Lemma 4.1). Denoting the level
sets of the rate function $I + \mu \log k_{\text{max}}$ by $\Phi_\alpha = \{ \Psi \in A : I(\Psi) + \mu(W) \log k_{\text{max}} \leq \alpha \}$ for $\alpha \in \mathbb{R}$, we see that $\Phi_\alpha \cap B_C$ is compact for any $\alpha$ and $C$. Thus, Varadhan’s lemma can be applied to prove Assertion (ii).

1.6. Dropping the congestion term. Proposition 1.3 yields a rather implicit description of the minimizers of (1.24) in the case $\beta, \gamma > 0$. The cardinality of the set of minimizers is also unclear. However, in the special case $\beta = 0$, where the congestion functional $\mathfrak{M}$ (1.9) is absent, the situation is much better. Indeed, it turns out that the minimizer is unique and is explicitly given in terms of the parameters of the model. Especially for the specific choice of Section 1.2.4 where $\mathfrak{S}$ penalizes interference (1.13), on base of this knowledge, we will be able in [KT17b] to derive a number of relevant qualitative properties of the trajectories.

In what follows, we call $\Sigma = (\nu_k)_{k \in k_{\text{max}}}$ with $\nu_k \in \mathcal{M}(W^k)$ for all $k \in [k_{\text{max}}]$ an asymptotic routing strategy if we have $\sum_{k=1}^{k_{\text{max}}} \pi_0 \nu_k = \mu$. In (1.19) we see that the first coordinate, $\Sigma$, of an admissible trajectory setting $\Psi$ is an asymptotic routing strategy, and in (1.21) we see that $S(\Psi)$ depends only on $\Sigma$. We will therefore write $S(\Sigma)$ for $S(\Psi)$. Further, we write $M = \sum_{k=1}^{k_{\text{max}}} \sum_{j=1}^{k-1} \pi_j \nu_k$, in accordance with (1.19) but with no regard to the measures $(\mu_m)_{m \in \mathbb{R}_0}$. We define an entropic term $J$ for asymptotic routing strategies as follows.

$$J(\Sigma) = \sum_{k=1}^{k_{\text{max}}} \mathcal{H}_{W^k}(\nu_k | \mu^\otimes k) - \sum_{k=2}^{k_{\text{max}}} \mu(W)^k + M(W) \log \mu(W).$$

(1.32)

Similarly to $I$ in (1.23), $J$ describes counting complexity in the high-density limit, but without reference to the measures $(\mu_m)_m$.

The following proposition summarizes the analogues of Theorem 1.2, Proposition 1.3 and Theorem 1.4 in case $\beta = 0$, after dropping all the measures $\mu_m$.

**Proposition 1.5.** The following statements hold almost surely with respect to $(X^\lambda)_{\lambda > 0}$.

1. The distribution of

$$\Sigma_\lambda(S) = (R_{\lambda,k}(S))_{k \in [k_{\text{max}}]}$$

under $P_{\lambda,X,\lambda}^{0,0}$ satisfies an LDP as $\lambda \to \infty$ with scale $\lambda$ on the set $\mathcal{A}_0 = \prod_{k=1}^{k_{\text{max}}} \mathcal{M}(W^k)$ with rate function given by $\mathcal{A}_0 \ni \Psi \mapsto J(\Sigma) + \mu(W) \log k_{\text{max}}$, which we define as $\infty$ if $\Sigma$ is not an asymptotic routing strategy. Further, the rate function has compact level sets.

2. For any $\gamma \in (0, \infty)$,

$$\lim_{\lambda \to \infty} \frac{1}{\lambda} \log Z_{\lambda,0}^{\gamma,0}(X^\lambda) = - \inf_{\Sigma \text{ asymptotic routing strategy}} (J(\Sigma) + \gamma S(\Sigma))$$

(1.34)

3. Let $\gamma > 0$ and $k_{\text{max}} > 1$. The variational formula on the r.h.s. of (1.34) exhibits a unique minimizer $\Sigma = (\nu_k)_{k \in [k_{\text{max}}]}$ given as

$$\nu_k(dx_0, \ldots, dx_{k-1}) = \mu(dx_0)A(x_0) \prod_{l=1}^{k-1} \frac{\mu(dx_l)}{\mu(W)} e^{-\tilde{\gamma}_k(x_0, \ldots, x_{k-1})}, \quad k \in [k_{\text{max}}],$$

(1.35)

where

$$\frac{1}{A(x_0)} = \sum_{k=1}^{k_{\text{max}}} \int_{W^{k-1}} \prod_{l=1}^{k-1} \frac{\mu(dx_l)}{\mu(W)} e^{-\tilde{\gamma}_k(x_0, \ldots, x_{k-1})}, \quad x_0 \in W.$$

(1.36)

4. For any $\gamma \in (0, \infty)$, the distribution of $\Sigma_\lambda(S)$ under $P_{\lambda,X,\lambda}^{\gamma,0}$ converges to the minimizer of the variational formula in (1.34).
Proposition 1.5 is proved in Section 6.

Let us explain in what way Proposition 1.5 is the special case of the aforementioned results for \( \beta = 0 \) and in what way it differs. It is true that the LDP in Assertion (1) directly follows from the LDP in Theorem 1.2 (1) via the contraction principle [DZ98, Theorem 4.2.1] for the projection map \((\Sigma, (\mu_m)_m) \mapsto \Sigma\), however, with rate function given by

\[
J(\Sigma) = \inf_{(\mu_m)_m \in \mathbb{N}_0 : \Psi = (\Sigma, (\mu_m)_m) \text{ admissible trajectory setting}} I(\Psi).
\]

It is an elementary but tedious task to identify this as in (1.32) by identifying

\[
\mu_m(dx) = \mu(dx) \left( \frac{M(dx)}{\mu(dx)} \right)^m \frac{1}{m!} e^{-M(dx)/\mu(dx)}, \quad m \in \mathbb{N}_0,
\]

as the unique minimizer on the right-hand side of (1.37), given \( M = \sum_{k=1}^{k_{\max}} \sum_{i=1}^{k-1} \pi_i \nu_k \). However, we chose an alternative route for proving the LDP with explicit identification of \( J \), which is a variant of the proof of Theorem 1.2 (1). From (1.37) it is clear that the variational formula in (1.34) is indeed the special case of (1.24) for \( \beta = 0 \), i.e.,

\[
\inf_{\Sigma \text{ asymptotic routing strategy}} (J(\Sigma) + \gamma S(\Sigma)) = \inf_{\Psi \text{ admissible trajectory setting}} (I(\Psi) + \gamma S(\Psi)).
\]

Note that also the minimizer \( \Psi \) is unique. This raises the additional question whether or not the measures \((P_{\lambda,m}(S))_{m \in \mathbb{N}_0}\) converge to the minimizer \((\mu_m)_{m \in \mathbb{N}_0}\) in (1.38) under \( P_{\lambda,X}^{\gamma,0} \) for \( M \) corresponding to the minimal \( \nu_k \)'s of (1.35). Since the congestion term, which gave rise to a strong compactness argument, is now absent, this question cannot immediately be decided using large-deviation arguments, but we nevertheless believe it is true. Moreover, this compactness property was also used in the proof of Theorem 1.2, which is another reason that we had to redo the proofs of Proposition 1.5 (2) and (3), given our proof of Assertion (1).

1.7. Discussion.

1.7.1. Interpretation of the entropy. We give an alternative representation of the entropy term \( I \), throwing some light on the probabilistic mechanisms that are present in our model.

The interpretation of an admissible trajectory setting \( \Psi = (\nu_k)_{k=1}^{k_{\max}}, (\mu_m)_{m=0}^{\infty} \) is given after Definition 1.3. These measures play the role of the empirical measures introduced in Section 1.3. The entropic term \( I \) in (1.24) describes the entropy of the choices of the indices \( i \) of the users \( X_i \) and the indices \( j = 1, \ldots, K_i - 1 \) of the relays \( S_j^i \) of the trajectories \( S^i \); it can be understood as the exponential rate of counting complexity.

Let us give a rewrite of \( I \). For a measurable set \( V \) and for \( \nu, \tilde{\nu} \in \mathcal{M}(V), \) let us write

\[
H_V(\nu | \tilde{\nu}) = \int_V d\nu \log \frac{d\nu}{d\tilde{\nu}}, \quad \text{if } \nu \ll \tilde{\nu} \text{ and } \infty \text{ otherwise}.
\]

Note that \( H_V(\nu | \tilde{\nu}) = H_V(\nu | \nu) \) if \( \nu(V) = \tilde{\nu}(V) \). Thus, we have

\[
I(\Psi) = \mu(W) H_{[k_{\max}]} \left( \left( \frac{\nu_k(W^k)}{\mu(W)} \right)_{k \in [k_{\max}]} \right) + \mu(W) \mathcal{H}_{\mathbb{N}_0} \left( \left( \frac{\mu_m(W)}{\mu(W)} \right)_{m \in \mathbb{N}_0} | \text{Po}_1/(\epsilon \mu(W)) \right)
\]

\[
- M(W) \log \frac{M(W)}{\mu(W)} - \frac{1}{e} + \sum_{k \in [k_{\max}]} \nu_k(W^k) \mathcal{H}_{W^k} \left( \nu_k | \mu \otimes \mathcal{N}^{0(k-1)} \right) + \sum_{m \in \mathbb{N}_0} \mu(m) \mathcal{H}_W \left( \mu_m | \mu \right).
\]

where we wrote \( \mathcal{N} = N/N(V) \) for the normalized version of a measure \( N \) on a set \( V \), \( \text{Po}_\alpha \) for the Poisson distribution on \( \mathbb{N}_0 \) with parameter \( \alpha \) and \( \epsilon \) for the counting measure on \( [k_{\max}] \). The terms on
the r.h.s. in the first line are entropies for the trajectory length and the number of incoming messages per relay with respect to natural reference measures. The terms in the last line are entropies for the distribution of the trajectories and of the locations of the relays that receive a given number of incoming messages. From \(1.41\), it is easy to see that \(I\) is bounded from below, using Jensen’s inequality and the finiteness of the counting measure on \([k_{\max}]\). (From the LDP in Theorem \(1.24\), one obtains that \(\inf I = -\mu(W) \log k_{\max}\); this equals minus the logarithmic rate of the total mass of the \(a\ priori\) measure \(1.41\).)

1.7.2. Interpretation of the minimizer(s). In case \(\beta, \gamma > 0\), Proposition \(1.3\) tells us some information about the limiting trajectory distribution and the limiting spatial distribution of users with a given number of incoming messages under the measure \(P_{\lambda,\mu}^{\gamma,\beta}\). Indeed, both have densities that are \(\mu\)-almost everywhere positive. It is remarkable that the \(k\)-step trajectories follow a distribution that comes from choosing independently all the \(k\) sites with measures that do not depend on \(k\) (the starting point according to \(A(x) \mu(dx)\) and all the other \(k - 1\) sites each according to \(C(x) M(dx)\)), exponentially weighted with the term \(\gamma f_k\). Furthermore, all the measures of the users receiving \(m\) messages superpose each other on the full set \(\text{supp}(\mu)\), and at each space point \(x\), this number \(m\) is distributed according to some Poisson distribution, exponentially weighted with the term \(\beta \eta(m)\).

As for the case \(\beta = 0, \gamma > 0\), we have a unique minimizer, which exhibits all the properties enumerated for \(\beta, \gamma > 0\). In the \(k\)-hop trajectories, the starting point is chosen according to \(A(x) \mu(dx)\) and all the other \(k - 1\) sites according to the measure \(\mu(dx)/\mu(W)\), weighted with \(\gamma f_k\). Moreover, the number of incoming messages at a given relay at the site \(x \in W\) is Poisson distributed with parameter equal to \(\mu(W)/\mu(dx)\).

1.7.3. Rotation symmetry. Let us assume that \(\beta = 0\) (cf. Section \(1.6\)), and let us consider the special setting of Section \(1.2.1\) where \(S\) penalizes interference and \(R\) penalizes congestion. If \(W = B_r(o) \subset \mathbb{R}^d\) is a closed ball and \(\mu\) is invariant under rotations, then the measures \((\nu_k)_k\) in \(1.35\) are also invariant under rotations of the entire trajectory, i.e., for any orthogonal \(d \times d\)-matrix \(O\), we have that \(\nu_k(dx_0, \ldots, dx_{k-1}) = \nu_k^O(dx_0, \ldots, dx_{k-1}) = \nu_k(d(Ox_0), \ldots, d(Ox_{k-1}))\) for any \(k \in [k_{\max}]\). This is easily seen by an inspection of the formulas for the entropy \(J\) in \(1.32\) and for the interference term \(S\) in \(1.21\), as the function \((x, y) \mapsto \int_W \ell(||z - y||) \mu(dz)/\ell(||x - y||)\) is invariant under rotation of both arguments with the same matrix.

1.7.4. Non-Poissonian users. In fact, the main results of this paper hold for any collection of (random or non-random) point processes \((X_i)_{i=1,\ldots,N(\lambda)}\) on \(W\) for which \(L_\lambda = \frac{1}{\lambda} \sum_{i=1}^{N(\lambda)} \delta_{X_i}\) converges weakly (almost surely, if random) to \(\mu\) as \(\lambda \to \infty\). Neither the independence or monotonicity in \(\lambda\), nor the Poissonity of \((N(\lambda))_{\lambda > 0}\) is used for the proofs. For example, our results remain also true for the deterministic set \(X^\lambda = W \cap (\frac{1}{\lambda} \mathbb{Z}^d)\) and \(\mu\) the Lebesgue measure on \(W\).

1.7.5. The annealed setting. Of mathematical interest might also be the annealed setting, where we average also over the locations of the users. In order to get an interesting result, we have to assume that \(L_\lambda\) satisfies a large deviation principle on the set \(M(W)\) with some good rate function \(H\). (In the case of a Poisson point process with intensity measure \(\lambda\mu\), \(H\) would be \(\text{IIJP}16\), Proposition 3.6) the relative entropy with respect to \(\mu\), see \(1.20\).) Then the large-\(\lambda\) exponential rate of the annealed free energy should be equal to the negative infimum over \(\mu_0 \in M(W)\) of \(H(\mu_0)\) plus the quenched rate function terms from the right-hand side of \(1.21\) with \(\mu\) replaced by \(\mu_0\) everywhere. Also our other results on the LDP and the form of the minimizer(s) should have some analogue, which we do not spell out.
2. The distribution of the empirical measures

Having seen in Section 1.5 that the Gibbsian model can be entirely described in terms of the trajectory setting \( \Psi_\lambda(s) \), i.e., of the crucial empirical measures \( R_{\lambda,k}(s) \) and \( P_{\lambda,m}(s) \) defined in (1.5)–(1.17), we now consider the question how to describe their distributions. We have to quantify the number of message trajectory families \( s \) that give the same family of empirical measures. The plain and short (but wrong) answer is

\[
\sum_{s \in \mathcal{S}_{\lambda} \times \mathcal{L}(\lambda^\lambda)} R_{\lambda,k}(s) = \nu_k \forall k, \quad P_{\lambda,m}(s) = \mu_m \forall m
\]

\[
\prod_{i \in \mathcal{I}_\lambda} \frac{1}{N(\lambda)^{d-1}} \approx e^{-\lambda I(\Psi)},
\]

where we recall \( I(\Psi) \) from (1.23) and recall that \( \Psi = ((\nu_k)_{k \in [k_{\max}]} , (\mu_m)_{m \in \mathbb{N}_0}) \). From such an assertion, it is indeed not far to conclude Theorem 1.2, but the problem is that this statement is not true like this. Actually, there are very many \( \Psi \)'s such that the left-hand side is equal to zero, for example if any of the \( \nu_k \)'s or \( \mu_m \)'s has values outside \( \frac{1}{2} \mathbb{N}_0 \). However, if we do not consider single \( \Psi \)'s, but open sets of \( \Psi \)'s, then the idea behind (2.1) is sustainable. Therefore, we proceed in a standard way by decomposing the area \( W \) into finitely many subsets and count the message trajectories only according to the discretization sets that they visit. In Section 2.2 we introduce necessary notation for carrying out this strategy, and in Section 2.2 we derive explicit formulas for the distribution of the empirical measures in this discretization.

For the purpose of the present paper, where we consider the high-density limit \( \lambda \to \infty \), we later need to take this limit and afterwards the limit as the fineness parameter \( \delta \) of the decomposition of \( W \) goes to zero. The outcome of these parts of the procedure is formulated in Proposition 3.1. In Proposition 3.2 the consequences for the interference term and for the congestion term are formulated.

2.1. Our discretization procedure. Let us now head towards the formulation of the discretization procedure. We proceed by triadic spatial discretization of the Poisson point process \( \lambda \times X_\lambda \), similarly to the approach of [HJKP15]. To be more precise, we perform the following discretization argument. Note that we may assume that our communication area \( W \) is taken as \( W = [-r,r]^d \), by accordingly extending \( \mu \) trivially. We write \( \mathbb{B} = \{3^{-n} | n \in \mathbb{N}_0 \} \). For \( \delta \in \mathbb{B} \), we define the set

\[
W_\delta = \{(x - r\delta, x + r\delta)^d : x \in (2r\delta\mathbb{Z})^d \cap W\}
\]

of congruent sub-cubes of \( W \) of side length \( 2r\delta \) and centers in \( (2r\delta\mathbb{Z})^d \). Note that \( W_\delta \) is a finite set, \( o \) is a center of an element of \( W_\delta \) and any intersection of two distinct elements of \( W_\delta \) has zero Lebesgue measure. Elements of \( W_\delta \) will be called \( \delta \)-subcubes. We will assume that for all \( \delta \in \mathbb{B} \), the \( \delta \)-subcubes are canonically numbered as \( W_\delta^1, \ldots, W_\delta^m \), which can be done e.g. according to the increasing lexicographic order of the midpoints of the subcubes. For \( j = 1, \ldots, \delta^{-d} \), let \( C(W_\delta^j) \) denote the centre of the \( \delta \)-subcube \( W_\delta^j \). Now, for Lebesgue-almost every \( x \in W \), for all \( \delta \in \mathbb{B} \) there exists a unique \( W_\delta^j \) that contains \( x \); let us denote this \( W_\delta^j \) by \( W_\delta^x \), and the set of all \( x \in W \) for which \( W_\delta^x \) is well-defined by \( W_\delta \). For such \( x \), the \( \delta \)-discretization operator is defined as \( \varrho_\delta : x \mapsto C(W_\delta^x) \). We will often use the simplified notation \( x^\delta = \varrho_\delta(x) \).

Now, if \( \nu \in \mathcal{M}(W) \), then for any \( \delta \in \mathbb{B} \), \( \nu^\delta = \nu \circ \varrho_\delta^{-1} \) is an element of \( \mathcal{M}(W_\delta) \) with the property \( \nu^\delta(W_\delta^j) = \nu(W_\delta^j) \), \( \forall j = 1, \ldots, \delta^{-d} \). Note that \( \mathcal{M}(W_\delta) = [0, \infty)^{W_\delta} \), which can be embedded in \( \mathbb{R}^{W_\delta} \). Thus, weak convergence in \( \mathcal{M}(W_\delta) \) is equivalent to norm convergence. On the other hand, if \( \nu \in \mathcal{M}(W_\delta) \) for some \( \delta \in \mathbb{B} \), then \( \nu \) defines an atomic measure on \( W \) that has the same weights on each \( W_\delta^j \) as \( \nu \) and no mass anywhere else. Throughout the rest of this paper, we will denote this measure on \( W \) the same way as \( \nu \), for simplicity. We proceed analogously for \( W^k, k \in [k_{\max}] \) instead of \( W \).

Now we are able to define what a standard setting is, the interpretation of which will be given right after the definition. For any set \( X \), let \( \mathcal{P}(X) \) denote the power set of \( X \).
Definition 2.1. A standard setting is a collection of measures
\[
\Psi = \left( \left( (\nu_k)_{k=1}^{k_{\max}}, \left( \left( (\delta_{\lambda}^k)_{k=1}^{k_{\max}} \right)_{\delta \in B}, \left( \left( \nu_{\lambda}^k \right)_{k=1}^{k_{\max}} \right)_{\delta \in B} \right)_{\lambda > 0}, \left( (\mu_m)_{m=0}^{\infty}, \left( (\mu_{\lambda}^m)_{m=0}^{\infty} \right)_{\delta \in B} \right)_{\lambda > 0}, \left( (\mu_{\lambda}^m)_{m=0}^{\infty} \right)_{\delta \in B} \right) \right)
\] (2.2)

with the following properties: For any \( \delta, \delta' \in B, \lambda > 0, k \in [k_{\max}], m \in \mathbb{N}_0 \) and \( s, s_0, \ldots, s_{k-1} = 1, \ldots, \delta^{-d}, \) respectively,
\begin{enumerate}
\item \( \mu_{\lambda}^\delta \in \mathcal{M}(W), \) with the property that the event \( \{ L_\lambda^\delta = \mu_{\lambda}^\delta \} \) has positive probability,
\item \( \delta' \leq \delta \implies \mu_{\lambda}^{\delta',\lambda} = \left[ \mathcal{P}(W_\lambda) \right] = \mu_{\lambda}^{\delta,\lambda}, \)
\item \( \mu_{\lambda}^{\delta,\lambda} \to \infty \implies \mu_{\lambda}^\delta, \)
\item \( \mu_{\lambda}^\delta = \mu \circ \varrho_{\delta}^{-1}. \) In particular, \( \mu_{\lambda}^\delta \to \mu, \)
\item \( \nu_{\lambda}^k \in \mathcal{M}(W^k). \) Further, we have \( \sum_{k=1}^{k_{\max}} \pi_0 \nu_{\lambda}^k = \mu_{\lambda}^\delta, \) moreover \( \lambda \nu_{\lambda}^k \left( W_{s_0}^{\delta} \times \cdots \times W_{s_{k-1}}^{\delta} \right) \in \mathbb{N}_0. \)
\item \( \delta' \leq \delta \implies \nu_{\lambda}^{\delta',\lambda} \left[ \mathcal{P}(W_\lambda^k) \right] = \nu_{\lambda}^{\delta,\lambda}, \)
\item \( \nu_{\lambda}^{\delta,\lambda} \to \infty \implies \nu_{\lambda}^\delta, \)
\item \( \nu_{\lambda}^k = \nu_k \circ \left( \varrho_{\delta} \times \cdots \times \varrho_{\delta} \right)^{-1}. \) In particular, \( \nu_{\lambda}^k \to \nu_k, \)
\item \( \nu_{\lambda}^m \in \mathcal{M}(W) \) with the property that \( \sum_{m=0}^{\infty} \mu_{\lambda}^m = \mu_{\lambda}^\delta, \) moreover \( \lambda \mu_{\lambda}^m (W_{s}^{\delta}) \in \mathbb{N}_0. \)
\item \( \sum_{m=0}^{\infty} m \mu_{\lambda}^m = \sum_{k=1}^{k_{\max}} \sum_{l=0}^{k_{\max}} \pi_l \mu_{\lambda}^k, \)
\item \( \delta' \leq \delta \implies \mu_{\lambda}^{\delta',\lambda} \left[ \mathcal{P}(W_\lambda) \right] = \mu_{\lambda}^{\delta,\lambda}, \)
\item \( \mu_{\lambda}^m \to \infty \implies \mu_{\lambda}^m, \)
\item \( \mu_{\lambda}^m = \mu \circ \varrho_{\delta}^{-1}. \) In particular, \( \mu_{\lambda}^m \to \mu_m. \)
\end{enumerate}

Let us introduce also the empirical measure
\[
P_{\lambda}(s) = \sum_{m \in \mathbb{N}_0} P_{\lambda,m}(s) = \frac{1}{\lambda} \sum_{i \in I^\lambda} \delta_{s_{i}}, \quad s \in S_{k_{\max}}(X^\lambda).
\] (2.3)

The interpretation of a standard setting \( \Psi \) is the following:
\begin{enumerate}
\item For \( \lambda > 0 \) and \( \delta \in B, \) \( \mu_{\lambda}^\delta \) is the \( \delta \)-discretized version \( P_{\lambda}^\delta(s) \) of the empirical measure \( P_{\lambda}(s) \) of any configuration \( s \in S_{k_{\max}}(X^\lambda); \) recall that this coincides with the empirical measure \( L_\lambda \) of the Poisson point process \( X^\lambda \) of users defined in [11] by means of our assumption that each user is picked precisely once in such a configuration. The consistency criterion [2] ensures that \( \mu_{\lambda}^\delta = P_{\lambda}^\delta(s) \) for the same \( s. \) For any \( \delta \in B, \) \( \mu_{\lambda}^\delta \) converges to the \( \delta \)-discretized version \( \mu_{\lambda}^\delta \) of \( \mu. \)
\item If \( \mu_{\lambda}^\delta \) corresponds to the discretized version of the rescaled empirical measure of the transmitters, then \( \nu_{\lambda}^k \) equals the \( \delta \)-discretized version \( R_{\lambda,k}^\delta(s) \) of the rescaled empirical measure \( R_{\lambda,k} \) of the \( k \)-hop trajectories, related to \( L_\lambda \) via the constraint \( \sum_{k=1}^{k_{\max}} \pi_0 \nu_{\lambda}^k = \mu_{\lambda}^\delta \) in [5], which means that each user sends out exactly one message. Again, we have a consistency relation [5], which ensures that for any \( \lambda > 0 \) and \( k \in [k_{\max}], \nu_{\lambda}^k = R_{\lambda,k}^\delta(s) \) for the same \( s \) for all \( \delta \in B. \)
\item Finally, for any \( m \in \mathbb{N}_0, \lambda > 0 \) and \( \delta \in B, \) \( \mu_{\lambda}^m \) equals the \( \delta \)-discretized version \( (L_\lambda^m)^\delta \) of the rescaled empirical measure
\[
L_\lambda^m = \sum_{i \in I^\lambda; m_i = m} \delta X_i
\]
of the spatial locations of users receiving exactly \( m \) incoming messages. The constraint \( \sum_{m=0}^{\infty} \mu_{\lambda}^m = \mu_{\lambda}^\delta \) in [4] means that each index \( i \in I^\lambda \) belongs to exactly one of the sets
\{i \in I^\delta : m_i(s) = m\}, while the constraint \(\sum_{m=0}^{\infty} m\mu_m^\delta = \sum_{k=1}^{k_{\text{max}}} \sum_{l=1}^{k-1} \pi\nu_k^\delta\) means that the total number of relaying hops taken by all users equals the total number of incoming messages received by each relay, on any subset of \(W_\delta\). The consistency relation (1) ensures that for any \(\lambda > 0\) and \(n \in \mathbb{N}\), \(\mu_m^\delta = P_{\lambda,m}^\delta(s)\) for the same \(s\) for all \(\delta \in \mathbb{B}\). For fixed \(\delta \in \mathbb{B}\) and \(m \in \mathbb{N}_0\), \(\mu_m^\delta\) converges to \(\mu_m^\delta\), and the \(\mu_m^\delta\)'s are the corresponding \(\delta\)-discretized versions of a limiting (continuous) measure \(\mu_m\) describing the asymptotic spatial distribution of \(m\)-hop trajectories.

Note that the condition (1) in Definition 2.1 in particular implies that for any \(\lambda' > \lambda > 0\) and \(\delta \in \mathbb{B}\) we have

\[\lambda'\mu_{\lambda',\delta}(A) \geq \lambda\mu_{\lambda,\delta}(A), \quad \forall A \subset W_\delta.\]

as a direct consequence of the fact that almost surely, \((X_\lambda)_{\lambda>0}\) is increasing.

Since in the definition of an admissible trajectory setting it is not required that \(\mu_m(W) > 0\) holds only for finitely many \(m\), we will often need the following notion of controlled standard setting in order to perform our large deviation analysis.

**Definition 2.2.** A controlled standard setting is a standard setting \(\Psi\) as in (2.2) with the following extra property:

\[
\lim_{\lambda \to \infty} \sum_{m=0}^{\infty} \eta(m)\mu_m^\delta(W_\delta) = \sum_{m=0}^{\infty} \eta(m)\mu_m^\delta(W_\delta) < \infty, \quad \text{for all} \ \delta \in \mathbb{B}. 
\] (2.4)

Note that by part (8) of Definition 2.1 we have \(\sum_{k=1}^{k_{\text{max}}} k\nu_k^\delta(W_\delta^k) = \sum_{k=1}^{k_{\text{max}}} k\nu_k(W_\delta^k)\) for any standard setting. Using this, we have the following lemma.

**Lemma 2.3.** Let \(\Psi\) be a controlled standard setting as in (2.2). Then \(\Psi = ((\nu_k^\delta)_{k=1}^{k_{\text{max}}}, (\mu_m^\delta)_{m=0}^{\infty})\) is an admissible trajectory setting.

**Proof.** Part (5) of Definition 2.1 claims that for all \(\delta \in \mathbb{B}\) and \(\lambda > 0\) we have \(\sum_{k=1}^{k_{\text{max}}} \pi_l\nu_k^\delta = \mu_l^\delta\). By parts (4) and (6) of Definition 2.1 we have \(\lim_{\lambda \to \infty} \sum_{m=0}^{\infty} \eta(m)\mu_m^\delta(W_\delta) = \sum_{m=0}^{\infty} \eta(m)\mu_m^\delta(W_\delta)\) for any standard setting. In order to see that (ii) holds for \((\mu_m^\delta)_{m=0}^{\infty}\), one can use part (9) of Definition 2.1 together with (2.4) and dominated convergence. Finally, by part (10) of Definition 2.1 in Definition 2.2 the fact that \(\lim_{n \to \infty} \eta(n)/n = \infty\) and dominated convergence, we see that for any controlled setting \(\Psi\), we also have

\[
\sum_{m=0}^{\infty} m\mu_m^\delta = \lim_{\delta \to 0} \sum_{m=0}^{\infty} m\mu_m^\delta = \lim_{\delta \to 0} \lim_{\lambda \to \infty} \sum_{m=0}^{\infty} m\mu_m^\delta = \lim_{\delta \to 0} \lim_{\lambda \to \infty} \sum_{k=1}^{k_{\text{max}}} k\pi_l\nu_k^\delta = \sum_{k=1}^{k_{\text{max}}} k\pi_l\nu_k \tag{2.5}
\]

in the weak topology of \(M(W)\). This implies (iii) in (1.19) for \(\Psi\). Hence, \(\Psi\) is an admissible trajectory setting.

**2.2. The distribution of the empirical measures.** In this section, we describe the combinatorics of the system. For a standard setting \(\Psi\) as in Definition 2.1 let us introduce the configuration set

\[
J_{\delta,\lambda}(\Psi) = \left\{ s \in S_{k_{\text{max}}}(X^\lambda) \mid P_{\lambda,k}^\delta(s) = \nu_k^\delta \ \forall k, \ P_{\lambda,m}^\delta(s) = \mu_m^\delta \ \forall m \right\} 
\] (2.6)

for fixed \(\delta \in \mathbb{B}\) and \(\lambda > 0\). In words, \(J_{\delta,\lambda}(\Psi)\) is the set of families of trajectories such that the \(\delta\)-coarsenings of the empirical measures of the trajectories and the hop numbers are given by the respective measures in the setting \(\Psi\). Note that \(J_{\delta,\lambda}(\Psi)\) depends only on the \(\delta-\lambda\) depending measures in the collection \(\Psi\).
In case $\mu^{\delta,\lambda}(W) > 0$, we will refer to the entity $s_0^i$, $i = 1, \ldots, \lambda \mu^{\delta,\lambda}(W_0)$ as the $i$th user or $i$th transmitter, the entity $s_l^i$, $i = 1, \ldots, \lambda \mu^{\delta,\lambda}(W_0)$ as the trajectory of the $i$th user, $s_{l-1}^i$ as the length (number of hops) of $s^i$, $s_l^i$ as the $l$th relay of $s^i$ (for $l = 1, \ldots, s_{l-1}^i - 1$), finally $m_i(s)$ as the number of incoming messages at the relay $s_0^i$.

The combinatorics of computing $\#J^{\delta,\lambda}(\Psi)$ is given as follows.

**Lemma 2.4 (Cardinality of $J^{\delta,\lambda}(\Psi)$).** For any $\delta, \lambda > 0$, and for any standard setting $\Psi$,

$$\#J^{\delta,\lambda}(\Psi) = N_{\delta,\lambda}^1(\Psi) \times N_{\delta,\lambda}^2(\Psi) \times N_{\delta,\lambda}^3(\Psi),$$

where

$$N_{\delta,\lambda}^1(\Psi) = \prod_{i=1}^{\delta-d} \left( (\lambda \mu_k^\delta(W_0^i) \times W_0^i)_{(i_1, \ldots, i_{k-1})} \right)^{\delta-d} \left( \prod_{i=1}^{k_{\text{max}}} \mu_k^\delta(W_0^i) \right),$$

$$N_{\delta,\lambda}^2(\Psi) = \prod_{i=1}^{\delta-d} \left( (\lambda \mu_k^\delta(W_0^i) \times W_0^i)_{m \in \mathbb{N}_0} \right),$$

$$N_{\delta,\lambda}^3(\Psi) = \prod_{i=1}^{\delta-d} \left( \prod_{m=0}^{\infty} \frac{m! \mu_m(W_0^i)}{m^n \cdot \prod_{m=0}^{\infty} \mu_m(W_0^i)} \right).$$

**Proof.** We proceed in three steps by counting first the trajectories, registering only the partition sets $W_0^i$ that they travel through, second, for each $m \in \mathbb{N}_0$, the sets of relays in each partition set that receive precisely $m$ incoming hops and finally the choices of the relays for each hop in each partition set. Since every choice in the three steps can be freely combined with the other ones, the product of the three cardinalities is equal to the number of all trajectory configurations with the requested coarsened empirical measures.

(A) **Number of the transmitters of trajectories passing through given sequences of $\delta$-subcubes.** For each configuration $s \in J^{\delta,\lambda}(\Psi)$ defined in (2.6), in each $\delta$-subcube $W_0^i$, $i = 1, \ldots, \delta-d$, there are $\lambda \mu_k^\delta(W_0^i)$ users. Out of them exactly $\lambda \mu_k^\delta(W_0^i)_{(i_1, \ldots, i_{k-1})}$ take exactly $k$ hops, having their first relay in $W_0^i$, their second in $W_0^i$, etc. and their $(k-1)$st relay in $W_0^i$, for any $k \in [k_{\text{max}}]$ and $i_1, \ldots, i_{k-1} = 1, \ldots, \delta-d$. Such choices in different sub-cubes $W_0^i$ corresponding to the transmitters are independent. Thus, the total number of such choices equals the number $N_{\delta,\lambda}^1(\Psi)$ defined in (2.8). Note that for $i = 1, \ldots, \delta-d$,

$$\sum_{k=1}^{k_{\text{max}}} \sum_{i_1, \ldots, i_{k-1}=1}^{\delta-d} \mu_k^\delta(W_0^i)_{(i_1, \ldots, i_{k-1})} = \sum_{k=1}^{k_{\text{max}}} \prod_{i=1}^{\delta-d} \pi_0 \mu_k^\delta(W_0^i) = \lambda \mu_k^\delta(W_0^i),$$

where we used part [A] of Definition 2.1 hence the multinomial expressions in (2.8) are well-defined.

(B) **Number of incoming messages.** In this step, for any $\delta$-subcube $W_0^i$, we count all the possible ways to distribute the incoming messages among the relays (= users) $X_j \in W_0^i$, under the two constraints that in $W_0^i$ there are $\lambda \mu_k^\delta(W_0^i)$ potential relays, and for any $m \in \mathbb{N}_0$, exactly $\lambda \mu_m^\delta(W_0^i)$ receive precisely $m$ incoming messages. Such choices are clearly independent of each other for different $\delta$-subcubes. Hence, the total number of such choices equals the number $N_{\delta,\lambda}^2(\Psi)$ defined in (2.9). Again, the constraint [A] from Definition 2.1 implies that the multinomial expression (2.9) is well-defined. Clearly, all choices in this part are independent of the choices in part [A].

(C) **Number of assignments of the hops to the relays.** Assume that we have chosen one possible choice in part [A] and one possible choice in part [B]. We now derive the number of possible ways
of distributing, for any \( i \), all the incoming hops in \( W^\delta_i \) among the users in \( W^\delta_i \). Let us call this number \( M_i \), then we know from part (A) that
\[
M_i = \lambda \sum_{k=1}^{\lambda \mu_m} \sum_{l=1}^{k-1} \pi_{\nu_k(W^\delta_i)} \quad \text{since each such hop is the} \ l \text{-th of some of the trajectories for some} \ l.
\]
The cardinality of the set of relays in \( W^\delta_i \) is equal to \( \lambda \sum_{m=0}^{\infty} \mu_m(W^\delta_i) \), and in part (B) we decomposed it into sets, indexed by \( m \), in which each relay receives precisely \( m \) ingoing hops. Let us call such a relay an \( m \)-relay. Think of each such relay as being replaced by precisely \( m \) copies (in particular those with \( m = 0 \) are discarded), then we have \( \lambda \sum_{m=0}^{\infty} m \mu_m(W^\delta_i) \) virtual relays in \( W^\delta_i \). (Note that this is equal to \( M_i \) by one of our constraints.) Now, if all these \( m \) copies of the \( m \)-relays were distinguishable, then the number of ways to distribute the \( M_i \) ingoing hops to the relays would be simply equal to \( M_i! \). However, since these \( m \) copies are identical, we overcount by a factor of \( m! \) for any \( m \)-relay. This means that the number of hops into \( W^\delta_i \) is equal to \( M_i! / \prod_{m=0}^{\infty} (m!)^{\mu_m(W^\delta_i)} \). Since all these cardinalities can freely be combined with each other, we have deduced that the number of possible choices is equal to the number \( N_{\delta,\lambda}(\Psi) \) defined in (2.10).

We also see that all the choices in the three parts are independent of each other, i.e., can be freely combined with each other and yield different combinations. Hence, we arrived at the assertion. \( \square \)

3. The limiting free energy: proofs of Theorems 1.2 and 1.4

In this section, we prove Theorem 1.2 that is, we derive the variational formula in (1.24) for the high-density (i.e., \( \lambda \to \infty \)) exponential rate of the partition function. Our first step is to derive the large-\( \lambda \) exponential rate of the combinatorial formulas for the empirical measures of Lemma 2.4 in Section 3.1. Furthermore, in Section 3.2 we formulate and prove how the interference term and the congestion term behave in the limits \( \lambda \to \infty \), followed by \( \delta \downarrow 0 \). In Section 3.3, given an admissible trajectory setting, we construct a standard setting containing it. Using all these, in Section 3.4 we prove Theorem 1.2.

For the rest of this section, we fix the set \( \Omega_1 \subset \Omega \) of full \( \mathbb{P} \)-measure on which we do our quenched investigations:
\[
\Omega_1 = \left\{ \omega \in \Omega : X_i(\omega) \in W^\delta_i \quad \forall i \in \mathbb{N}, \right\}
\]
\[
\lim_{\lambda \to \infty} \frac{\#\{i \in I^\lambda(\omega) : X_i(\omega) \in W^\delta_i \}}{\lambda} = \mu(W^\delta_j), \quad \forall j = 1, \ldots, \delta^{-d}, \forall \delta \in \mathbb{B}\right\}. \tag{3.1}
\]
That \( \mathbb{P}(\Omega_1) = 1 \) holds follows immediately from the Restriction Theorem [K93 Section 2.2] combined with the Poisson Law of Large Numbers [K93 Section 4.2] and the fact that \( \mu \) is absolutely continuous.

3.1. The asymptotics of the combinatorics. Let us fix a controlled standard setting \( \Psi \) as in (2.2). Fix any \( \omega \in \Omega_1 \), and let the quantities \( I^\lambda \) and \( X^\lambda \) refer to this \( \omega \). Denote
\[
N_{\delta, \lambda}^\delta(\Psi) = \prod_{i=1}^{\delta^{-d}} \prod_{k=1}^{\lambda \mu_m} \prod_{l=1}^{k-1} N(\lambda)_{\lambda \nu_k(W^\delta_i)}. \tag{3.2}
\]
Recall the notation \( H_V(\cdot \, | \cdot) \) from (1.40) and \( c_m = \exp(-1/(e \mu(W))(-e \mu(W))^{-m}/m! \) from (1.23). Note that the rate function \( I \) defined in (1.23) has also the representation
\[
I(\Psi) = \sum_{k=1}^{\lambda \mu_m} H_{W^\delta}((\nu_k \, | \, \mu^{\otimes k}) - H_W(\sum_{m=0}^{\infty} m \mu_m | \mu) + \sum_{m=0}^{\infty} H_W(\mu_m | \mu c_m) - 1/e, \tag{3.3}
\]
which we are going to use here. We now identify the large-\( \lambda \) exponential rate of the cardinality of \( J_{\delta, \lambda}(\Psi) \) both on the scale \( \lambda \log \lambda \) and \( \lambda \):
Proposition 3.1 (Exponential rates of counting terms). Let $\Psi$ be a controlled standard setting. Let us write $\Psi = ((\nu_k)_{k=1}^{\infty}, (\mu_m)_{m=0}^{\infty})$. We have
\[
\lim_{\delta \downarrow 0} \lim_{\lambda \to \infty} \frac{1}{\lambda} \log \frac{\# J^\delta,\lambda(\Psi)}{N^0_{\delta,\lambda}(\Psi)} = -I(\Psi),
\]
as an equality in $[0, \infty)$. Moreover if $I(\Psi) < \infty$, then
\[
\lim_{\delta \downarrow 0} \lim_{\lambda \to \infty} \frac{1}{\lambda \log \lambda} \log \# J^\delta,\lambda(\Psi) = \sum_{k=1}^{k_{\max}} (k-1)\nu_k(W^k) = \sum_{m=0}^{\infty} m\mu_m(W) < \infty,
\]
amost surely.

Proof. Recall that $\Psi$ is an admissible trajectory setting, according to Lemma 2.3. In particular, $I(\Psi) \in (-\infty, \infty]$ is well-defined.

We use Stirling’s formula $\lambda! = (\lambda/e)^\lambda e^{o(\lambda)}$ in the limit $\lambda \to \infty$, which leads to
\[
\lim_{\lambda \to \infty} \frac{1}{\lambda} \log \left( a^{(\lambda)}_1, \ldots, a^{(\lambda)}_n \right) = - \sum_{i=1}^{n} a_i \log \frac{a_i}{a}, \tag{3.4}
\]
for any integers $a^{(\lambda)}_1, \ldots, a^{(\lambda)}_n$ that sum up to $a^{(\lambda)}$ and satisfy $\frac{1}{\lambda} a^{(\lambda)}_i \xrightarrow{\lambda \to \infty} a_i$ for $i = 1, \ldots, n$ with positive numbers $a_1, \ldots, a_n$ satisfying $\sum_{i=1}^{n} a_i = a$.

From (2.3) we obtain that
\[
I^1_\delta(\Psi) = - \lim_{\lambda \to \infty} \frac{1}{\lambda} \log N^1_{\delta,\lambda}(\Psi)
\]
\[
= \sum_{i=1}^{\delta-d} \sum_{k=1}^{k_{\max}} \sum_{i_1, \ldots, i_{k-1} = 1}^{\delta-d} \nu^\delta_k(W^\delta_{i_1} \times W^\delta_{i_2} \times \ldots \times W^\delta_{i_{k-1}}) \log \frac{\nu^\delta_k(W^\delta_{i_1} \times W^\delta_{i_2} \times \ldots \times W^\delta_{i_{k-1}})}{\mu^\delta(W^\delta_{i_1})},
\]
where we also used that all the measures $\nu^\delta_k$ and $\mu^\delta$ converge as $\lambda \to \infty$ to $\nu^\delta_k$ and $\mu^\delta$, respectively.

Now we add the term $\prod_{l=1}^{k-1} \mu^\delta(W^\delta_{i_l})$ both in the numerator and the denominator under the logarithm and separate these two terms. In the former, we write its logarithm as $\sum_{i=1}^{k-1} \log \mu^\delta(W^\delta_{i_l})$, interchange this sum on $l$ with all the other sums on the $i_0, \ldots, i_{k-1}$ and write the sums over $i_0, \ldots, i_{l-1}, i_{l+1}, \ldots, i_{k-1}$ in terms of the $l$-th marginal measure of $\nu^\delta_k$. This gives
\[
I^1_\delta(\Psi) = \sum_{i=1}^{\delta-d} \sum_{k=1}^{k_{\max}} \sum_{i_1, \ldots, i_{k-1} = 1}^{\delta-d} \nu^\delta_k(W^\delta_{i_1} \times W^\delta_{i_2} \times \ldots \times W^\delta_{i_{k-1}}) \log \frac{\nu^\delta_k(W^\delta_{i_1} \times W^\delta_{i_2} \times \ldots \times W^\delta_{i_{k-1}})}{\mu^\delta(W^\delta_{i_1}) \prod_{l=1}^{k-1} \mu^\delta(W^\delta_{i_l})} + \sum_{i=1}^{\delta-d} \sum_{k=1}^{k_{\max}} \sum_{l=1}^{k-1} \pi_l \nu^\delta_k(W^\delta_{i_l}) \log \mu^\delta(W^\delta_{i_l}). \tag{3.5}
\]
In the same way as for $I^1_\delta$, we obtain
\[
I^2_\delta(\Psi) = - \lim_{\lambda \to \infty} \frac{1}{\lambda} \log N^2_{\delta,\lambda}(\Psi) = \sum_{i=1}^{\delta-d} \sum_{m=0}^{\infty} \mu^\delta_m(W^\delta_{i_l}) \log \frac{\mu^\delta_m(W^\delta_{i_l})}{\mu^\delta(W^\delta_{i_l})}. \tag{3.6}
\]
Using (3.1), on $\Omega_1$ we have that the asymptotic behaviour of (3.2) is the following
\[
N^0_{\delta,\lambda}(\Psi) = N(\lambda)^\delta \sum_{i=1}^{\delta-d} \sum_{k=1}^{k_{\max}} \sum_{l=1}^{k-1} \pi_l \nu^\delta_k(W^\delta_{i_l}) = (\lambda \mu(W))^{\lambda(1+o(1))} \sum_{i=1}^{\delta-d} \sum_{k=1}^{k_{\max}} \sum_{l=1}^{k-1} \pi_l \nu^\delta_k(W^\delta_{i_l}).
Let us begin with the first line on the right-hand side of (3.8). For counting terms in (2.10) and (3.2) as follows:

\[
I_\delta^{3,0}(\Psi) = - \lim_{\lambda \to \infty} \frac{1}{\lambda} \log \frac{N_{\delta,\lambda}^3(\Psi)}{N_{\delta,\lambda}(\Psi)}
\]

\[
= - \lim_{\lambda \to \infty} \frac{1}{\lambda} \log \prod_{i=1}^{\delta-d} k_{\text{max}}^{-1} \frac{\pi_{\delta k}(W^\delta_i)}{\pi_{\delta k}(W^\delta_i)} \frac{\alpha_{\delta k}(W^\delta_i)}{\alpha_{\delta k}(W^\delta_i)} \frac{\beta_{\delta k}(W^\delta_i)}{\beta_{\delta k}(W^\delta_i)} \prod_{m=0}^{\infty} m! \mu_m(W^\delta_i)
\]

\[
= - \sum_{i=1}^{\delta-d} k_{\text{max}}^{-1} \sum_{k=1}^{\delta-d} \sum_{l=1}^{\delta-d} \pi_{\delta k}(W^\delta_i) \left( \log \sum_{k=1}^{\delta-d} \sum_{l=1}^{\delta-d} \pi_{\delta k}(W^\delta_i) - (1 + \log \mu(W)) \right)
\]

\[
+ \sum_{i=1}^{\delta-d} k_{\text{max}}^{-1} \sum_{m=0}^{\infty} \mu_m(W^\delta_i) \log(m!)
\]

where for the last term we used the fact that \( \Psi \) is controlled (see also Lemma 2.3), together with dominated convergence. We can summarize the sum of the terms in (3.5), (3.6) and (3.7) as

\[
- \lim_{\lambda \to \infty} \frac{1}{\lambda} \log \frac{\# I_{\delta,\lambda}^3(\Psi)}{N_{\delta,\lambda}^0(\Psi)} = I_\delta^1(\Psi) + I_\delta^2(\Psi) + I_\delta^{3,0}(\Psi)
\]

\[
= \sum_{k=1}^{\delta-d} \sum_{i_0,\ldots,i_{k-1}=1} \nu_{\delta k}(W^\delta_{i_0} \times \ldots \times W^\delta_{i_{k-1}}) \log \frac{\nu_{\delta k}(W^\delta_{i_0} \times \ldots \times W^\delta_{i_{k-1}})}{\prod_{i=0}^{k-1} \mu^\delta(W^\delta_{i})}
\]

\[
+ \sum_{i=1}^{\delta-d} \sum_{m=0}^{\infty} \mu_m(W^\delta_i) \log \mu_m(W^\delta_i)
\]

\[
- \sum_{i=1}^{\delta-d} \sum_{k=1}^{\delta-d} \sum_{l=1}^{\delta-d} \pi_{\delta k}(W^\delta_i) \left( \log \sum_{k=1}^{\delta-d} \sum_{l=1}^{\delta-d} \pi_{\delta k}(W^\delta_i) - (1 + \log \mu(W)) \right)
\]

\[
+ \sum_{m=0}^{\infty} \mu_m(W^\delta)[m(1 + \log \mu(W)) + \log(m!)]
\]

where in the first line on the right-hand side we changed the summing index \( i \) into \( i_0 \). Since we have

\[
\sum_{m=0}^{\infty} \mu_m(W^\delta) = \sum_{m=0}^{\infty} \mu_m(W) = \mu(W),
\]

and thus

\[
\sum_{i=1}^{\delta-d} \sum_{m=0}^{\infty} \mu_m(W^\delta) \log \mu_m(W^\delta) + (m(1 + \log \mu(W)) + \log(m!)) = \sum_{m=0}^{\infty} \mathcal{H}_{W^\delta}(\mu_m | \mu^\delta c_m) - \frac{1}{e},
\]

we obviously arrived at the discrete version of the entropy terms in (3.3).

Now we argue that taking the limit as \( \delta \downarrow 0 \) through \( \delta \in \mathbb{B} \), yields the desired entropy terms in (3.3). Let us begin with the first line on the right-hand side of (3.8). For \( \delta \in \mathbb{B} \), let us define \((\nu^\delta_k)_{k=1}^{\delta-d}\) with \(\nu^\delta_k \in \mathcal{M}(W^k)\) as follows,

\[
\nu^\delta_k = \mu_k \otimes \ldots \otimes \mu_k \mathbb{1}_{W^\delta_{i_0} \times \ldots \times W^\delta_{i_{k-1}}},
\]

On the other hand, also by Stirling’s formula, we can identify the large-\( \lambda \) rate of the quotient of the counting terms in (2.10) and (3.2) as follows:
so that for all $k$,

$$H_W(\nu_k \delta \mid \mu \otimes^k) = \sum_{i_0, \ldots, i_{k-1}=1}^{\delta - d} \nu_k(W_i^{\delta} \times \ldots \times W_{i_{k-1}}^{\delta}) \log \frac{\nu_k(W_i^{\delta} \times \ldots \times W_{i_{k-1}}^{\delta})}{\prod_{l=0}^{k-1} \mu^{\delta}(W_{i l}^{\delta})}.$$ 

Now, $\nu_k^{\delta}$ also converges to $\nu_k$ in the weak topology of $M(W^k)$, for all $k$. Therefore, by lower semicontinuity of the relative entropy (cf. [DZ98 Lemma 6.2.12 and Theorem D.12])

$$\lim \inf_{\delta, k \to 0} \sum_{i=1}^{\delta - d} \left( \sum_{k=1}^{k_{\max}} \sum_{l=1}^{l_{\max}} \pi_l \nu_k^\delta(W_i^\delta) \right) \log \frac{\sum_{k=1}^{k_{\max}} \sum_{l=1}^{l_{\max}} \pi_l \nu_k^\delta(W_i^\delta)}{\mu^\delta(W_i^\delta)} = \lim \inf_{\delta, k \to 0} \sum_{i=1}^{\delta - d} \left( \sum_{k=1}^{k_{\max}} \sum_{l=1}^{l_{\max}} \pi_l \nu_k^\delta(W_i^\delta) \right) \log \frac{\sum_{k=1}^{k_{\max}} \sum_{l=1}^{l_{\max}} \pi_l \nu_k^\delta(W_i^\delta)}{\mu^\delta(W_i^\delta)} = H_W(\sum_{m=0}^{\infty} \mu_m \mid \mu).$$

Finally, we have

$$\lim \inf_{\delta, k \to 0} \sum_{i=1}^{\delta - d} \left( \sum_{k=1}^{k_{\max}} \sum_{l=1}^{l_{\max}} \pi_l \nu_k^\delta(W_i^\delta) \right) \log \frac{\sum_{k=1}^{k_{\max}} \sum_{l=1}^{l_{\max}} \pi_l \nu_k^\delta(W_i^\delta)}{\mu^\delta(W_i^\delta)} = \lim \inf_{\delta, k \to 0} \sum_{i=1}^{\delta - d} \left( \sum_{k=1}^{k_{\max}} \sum_{l=1}^{l_{\max}} \pi_l \nu_k^\delta(W_i^\delta) \right) \log \frac{\sum_{k=1}^{k_{\max}} \sum_{l=1}^{l_{\max}} \pi_l \nu_k^\delta(W_i^\delta)}{\mu^\delta(W_i^\delta)} = \lim \inf_{\delta, k \to 0} \sum_{i=1}^{\delta - d} \left( \sum_{k=1}^{k_{\max}} \sum_{l=1}^{l_{\max}} \pi_l \nu_k^\delta(W_i^\delta) \right) \log \frac{\sum_{k=1}^{k_{\max}} \sum_{l=1}^{l_{\max}} \pi_l \nu_k^\delta(W_i^\delta)}{\mu^\delta(W_i^\delta)} = \lim \inf_{\delta, k \to 0} \sum_{i=1}^{\delta - d} \left( \sum_{k=1}^{k_{\max}} \sum_{l=1}^{l_{\max}} \pi_l \nu_k^\delta(W_i^\delta) \right) \log \frac{\sum_{k=1}^{k_{\max}} \sum_{l=1}^{l_{\max}} \pi_l \nu_k^\delta(W_i^\delta)}{\mu^\delta(W_i^\delta)} = \lim \inf_{\delta, k \to 0} \sum_{i=1}^{\delta - d} \left( \sum_{k=1}^{k_{\max}} \sum_{l=1}^{l_{\max}} \pi_l \nu_k^\delta(W_i^\delta) \right) \log \frac{\sum_{k=1}^{k_{\max}} \sum_{l=1}^{l_{\max}} \pi_l \nu_k^\delta(W_i^\delta)}{\mu^\delta(W_i^\delta)} = \lim \inf_{\delta, k \to 0} \sum_{i=1}^{\delta - d} \left( \sum_{k=1}^{k_{\max}} \sum_{l=1}^{l_{\max}} \pi_l \nu_k^\delta(W_i^\delta) \right) \log \frac{\sum_{k=1}^{k_{\max}} \sum_{l=1}^{l_{\max}} \pi_l \nu_k^\delta(W_i^\delta)}{\mu^\delta(W_i^\delta)}$$

The first part of Proposition 3.1 follows.

Moreover, if $I(\Psi) < \infty$, then we have by continuity

$$\lim \inf_{\delta, k \to 0} \sum_{i=1}^{\delta - d} \left( \sum_{k=1}^{k_{\max}} \sum_{l=1}^{l_{\max}} \pi_l \nu_k^\delta(W_i^\delta) \right) \log \frac{\sum_{k=1}^{k_{\max}} \sum_{l=1}^{l_{\max}} \pi_l \nu_k^\delta(W_i^\delta)}{\mu^\delta(W_i^\delta)} = \lim \inf_{\delta, k \to 0} \sum_{i=1}^{\delta - d} \left( \sum_{k=1}^{k_{\max}} \sum_{l=1}^{l_{\max}} \pi_l \nu_k^\delta(W_i^\delta) \right) \log \frac{\sum_{k=1}^{k_{\max}} \sum_{l=1}^{l_{\max}} \pi_l \nu_k^\delta(W_i^\delta)}{\mu^\delta(W_i^\delta)}$$

where in the last equality we used that by Fubini’s theorem, $\pi_0 \nu_k(W) = \nu_k(W^k)$ for all $k$. Hence, the second part of Proposition 3.1 follows.

3.2. Approximations for the penalization terms. The limiting relations between the penalization terms depending on the numbers of incoming messages in (1.9) and (1.22), and between the continuous penalization terms in (1.7) and (1.21) are given as follows.

**Proposition 3.2.** Let $\Psi$ be a controlled standard setting. Let us write $\Psi = (\nu_k)_{k=1}^{k_{\max}}, (\mu_m)_{m=0}^{\infty}$ for the admissible trajectory setting contained in $\Psi$. Then, almost surely

$$\lim \inf_{\delta, k \to 0} \sum_{s \in J_i^k, \lambda(\Psi)} \frac{1}{\lambda} M(s) - M(\Psi) = 0,$$

and

$$\lim \inf_{\delta, k \to 0} \sum_{s \in J_i^k, \lambda(\Psi)} \frac{1}{\lambda} S(s) - S(\Psi) = 0.$$
Proof. Throughout the proof, we perform our analysis under $\Omega_1$. First, we show (3.11). Consider some $s \in J^{v_0}(\Psi)$ for $\lambda > 0$ and $\delta \in \mathbb{R}$. Additionally assume that $s_i^l \in W_\delta$ for all $i \in I^\lambda$ and $l = 0, \ldots, k$ (which is always the case for $s = (s_i^l)_{i \in I^\lambda}$ on $\Omega_1$).

Then $P^\delta_m(s) = \mu^\delta_m$ and $P^\delta_{\lambda,m}(s) = \mu^\delta_m$ for all $m \in \mathbb{N}_0$, see the definition (2.6) of $J^{v_0}(\Psi)$ and (2.2). Recall that $m_i(s)$ is the number of ingoing messages at relay $X_i$ for the trajectory configuration $s$. Hence we have

$$M(s) = \sum_{i \in I^\lambda} \eta(m_i(s)) = \sum_{m=0}^{\infty} \eta(m)\#\{i \in I^\lambda \mid m_i(s) = m\} = \sum_{m=0}^{\infty} \eta(m)P^\delta_{\lambda,m}(W)$$

$$= \sum_{m=0}^{\infty} \eta(m)P^\delta_{\lambda,m}(W) = \lambda \sum_{m=0}^{\infty} \eta(m)\mu^\delta_m(W),$$

for all such $s$. Note that by parts (12) and (13) of Definition 2.4, we obtain that $\mu^\delta_m$ tends to $\mu_m$ as $\lambda \to \infty$ followed by $\delta \downarrow 0$. Now, (2.4) in Definition 2.2 together with the fact that the total mass of $\mu^\delta_m$ equals the one of $\mu_m$ for any $m$, implies the assertion in (3.11).

We continue with verifying (3.12). Let us fix an arbitrary controlled standard setting $\Psi$. Our goal is to prove that (3.12) holds for this $\Psi$. Let us recall from Section 2.1 that we can conceive all measures on $W^\delta$ as measures on $W^\mu$. Using this, and that, for an admissible trajectory setting $\Psi = (\nu_k)_{k=1}^{k_{\infty}}$, we have for any $\lambda > 0$, $\delta \in \mathbb{R}$, $s \in J^{v_0}(\Psi)$ and $k \in [k_{\infty}]$

$$\frac{1}{\lambda} \mathcal{S}(s) - S(\Psi) = \langle R_{\lambda,k}(s), f_k(L_{\lambda,\cdot}) \rangle - \langle \nu_k, f_k(\cdot, \cdot) \rangle.$$
therefore it is bounded. Now, since eventually \( L_\lambda \in \mathcal{M}_{\leq 2\mu(W)}(W) \), the first term on the right-hand side of (3.13) tends to 0.

As for the second term, note that for any \( \delta \in \mathbb{B} \), \( L_\lambda - L_\lambda^\delta \) tends to \( \mu - \mu^\delta \) as \( \lambda \to \infty \), which tends to 0 as \( \delta \downarrow 0 \). Thus, by the fact that \( f_k \) is continuous and bounded on \( \mathcal{M}_{\leq 2\mu(W)}(W) \times W^k \) and eventually \( L_\lambda, L_\lambda^\delta, \nu_k^\delta, \mu^\delta \in \mathcal{M}_{\leq 2\mu(W)}(W) \) for all \( \delta \in \mathbb{B} \), the second term also tends to 0 as first \( \lambda \to \infty \) and afterwards \( \delta \downarrow 0 \). An analogous argument applies for the third term, using that \( L_\lambda^\delta \) converges to \( \mu \) as first \( \lambda \to \infty \) and then \( \delta \downarrow 0 \) by part (3) of Definition 2.1 and the definition of \( \mu^\delta, \delta \in \mathbb{B} \). The fourth term tends to zero since it easily follows from parts (7) and (8) that \( \nu_k^\delta, \lambda \in \mathbb{B} \) converges weakly to \( \nu_k \), and \( f_k \) is continuous as bounded on \( \mathcal{M}_{\leq 2\mu(W)}(W) \times W^k \). We conclude (3.12) and Proposition 3.2. \( \square \)

3.3. Existence of standard settings. Recall that we equip \( \mathcal{A} \) defined in (1.29) with the product topology of the weak topologies of the factors \( \mathcal{M}(W^k) \) and that this is the topology of coordinatewise weak convergence. For \( k \in \mathbb{N} \), let \( d_k(\cdot, \cdot) \) be the Lipschitz-bounded metric (3.13) on \( \mathcal{M}(W^k) \), which generates the weak topology on this space. Then,

\[
d_0(\Psi^1, \Psi^2) = \sum_{k=1}^{k_{\text{max}}} d_k(\nu_k^1, \nu_k^2) + \sum_{m=0}^{\infty} 2^{-m} d_1(\mu_m^1, \mu_m^2), \quad \Psi^1, \Psi^2 \in \mathcal{A}
\]

(3.15)
is a metric on \( \mathcal{A} \) that generates the product topology. For \( \varrho > 0 \) and \( \Psi \in \mathcal{A} \), let us write \( B_\mathcal{A}(\Psi) = \{ \Psi' \in \mathcal{A} : d_0(\Psi', \Psi) < \varrho \} \) for the open \( \varrho \)-ball around \( \Psi \).

We have the following.

**Proposition 3.3.** On \( \Omega_1 \), for any admissible trajectory setting (see Definition 1.1), \( \Psi = ((\nu_k)_k, (\mu_m)_m) \), there exists a standard setting \( \mathcal{O} \) containing it. If \( \sum_{m} \eta_{m}(\mu_{m}(W)) < \infty \), then \( \mathcal{O} \) can be chosen to be a controlled standard setting.

**Proof.** We fix an admissible trajectory setting \( \Psi \) and construct \( \Psi \) as follows. As is required in Definition 2.3, the measures \( \mu^\delta, \nu^\delta_k \) for \( k \in [k_{\text{max}}] \) and \( \mu^\delta_m \) for \( m \in \mathbb{N}_0 \) are the \( \delta \)-coarsenings of the measures \( \nu_k \) and \( \mu_m \), respectively, and \( \mu^\delta \in \mathcal{M}(W) \). Now for \( \delta \in \mathbb{B} \) and \( \lambda > 0 \), pick some measures \( \nu^\delta_k \) and \( \mu^\delta_m \) with values in \( \mathbb{L}^\infty_\mathcal{A} = \sum_{m=0}^{\infty} \sum_{k=1}^{k_{\text{max}}} \mu^\delta_m \mu^\delta_k \) and \( \sum_{m=0}^{\infty} m \mu^\delta_m = \sum_{k=1}^{k_{\text{max}}} \sum_{m=1}^{k} \sum \nu^\delta_k \mu^\delta_m \) of Definition 2.3 are met, such that \( \nu^\delta_k \Rightarrow \nu_k^\delta \) and \( \mu^\delta_m \Rightarrow \mu_m^\delta \) as \( \lambda \to \infty \) and such that the collection \( \mathcal{O} \) of all these measures is a standard setting containing \( \Psi \), which is controlled if \( \sum_{m} \eta_{m}(\mu_{m}(W)) < \infty \).

We claim that this can be done by taking suitable up- and downroundings of the numbers

\[
\nu^\delta_k(W_{s_0}^\delta \times \ldots \times W_{s_{k-1}}^\delta) = \nu^\delta_k(W_{s_0}^\delta \times \ldots \times W_{s_{k-1}}^\delta) \frac{L_\lambda^\delta(W_{s_0}^\delta)}{\mu^\delta(W_{s_0}^\delta)} \mathbb{1}\{\mu^\delta(W_{s_0}^\delta) > 0\}, \quad k \in [k_{\text{max}}],
\]

(3.16)
for all \( s_0, \ldots, s_{k-1} = 1, \ldots, \delta^{-d} \), and dividing by \( \lambda \), analogously for the \( \mu_m^\delta \)'s. Now, using the \( d \)-metric defined in (3.15), we prove that the convergences required in Definition 2.3 hold for such \( \Psi \).

First, we prove the convergence of the \( \delta \)-coarsenings \( \Psi^\delta = ((\nu_k^\delta)_k, (\mu^\delta_m)_m) \) to \( \Psi \) in the \( d_0 \)-metric. We claim that for any \( \varrho > 0 \), there exists \( \delta_0 \in \mathbb{B} \) such that \( \Psi^\delta \in B_\mathcal{A}(\Psi) \) for all \( \mathbb{B} \ni \delta \leq \delta_0 \). Indeed, for \( k \in [k_{\text{max}}], \nu_k \in \mathcal{M}(W^k) \) and \( \delta \in \mathbb{B} \) we see that the distance between \( \nu_k \) and its \( \delta \)-coarsening is of order \( \delta \) w.r.t. the Lipschitz-bounded metric:

\[
d_k(\nu_k, \nu^\delta_k) = \sup_{f \in \text{Lip}_1(W^k)} \sum_{j_0, \ldots, j_{k-1}=1}^{\delta-d} \int_{W_{j_0}^\delta \times \ldots \times W_{j_{k-1}}^\delta} |f(x) - f(C(W_{j_0}^\delta \times \ldots \times W_{j_{k-1}}^\delta))| \nu_k(dx)
\]

\[
\leq \sum_{j_0, \ldots, j_{k-1}=1}^{\delta-d} \int_{W_{j_0}^\delta \times \ldots \times W_{j_{k-1}}^\delta} |x - C(W_{j_0}^\delta \times \ldots \times W_{j_{k-1}}^\delta)| \nu_k(dx) \leq \nu_k(W^k) \frac{\sqrt{d} \delta}{2},
\]
where we wrote $x = (x_0, \ldots, x_{k-1})$; and analogously for $\mu_m$. Thus, we have

$$d_0(\Psi, \Psi^\delta) \leq \delta \sqrt{\frac{d}{2}} \left[ \sum_{k=1}^{k_{\max}} \nu_k(W^k) \sqrt{k} + \sum_{m=0}^{\infty} \mu_m(W) 2^{-m} \right].$$

Since $\sum_{m=0}^{\infty} \mu_m(W) < \infty$ by (ii) in (3.19), there exists a constant $C$, only depending on $\Psi$, such that $\Psi^\delta \in B_\delta(\Psi)$ for any $\delta \leq C \varrho$.

Second, we ignore the up- or downroundings in the construction of $\Psi$ and prove the following. For $\delta \in \mathbb{B}$ and $\lambda > 0$, let $\Psi^{\delta, \lambda}$ be the collection of the measures introduced in (3.16). We claim that on $\Omega_1$, we have

$$\limsup_{\lambda \to \infty} d_0(\Psi^\delta, \Psi^{\delta, \lambda}) = 0.$$ 

Indeed, for any $k \in [k_{\max}]$ and $s_0, \ldots, s_{k-1} = 1, \ldots, \delta^{-d}$, $d_k(\nu_k, \nu_k^{\delta, \lambda})$ is bounded from above by

$$\sup_{f \in L_{\delta, 1}(W^k)} \left| \sum_{s_k = 1}^{\delta^{-d}} \nu_k(W^\delta_{s_k} \times \ldots \times W_{s_0}) - \frac{L_A(\nu_k^{\delta, \lambda})}{\mu_k(\nu_k^{\delta, \lambda})} - \frac{L_A(\nu_k^{\delta, \lambda})}{\mu_k(\nu_k^{\delta, \lambda})} \right| \leq \delta^{-d} \max_{s_k=1}^{\delta^{-d}} \frac{L_A(\nu_k^{\delta, \lambda})}{\mu_k(\nu_k^{\delta, \lambda})} - 1. \quad (3.17)$$

Thus,

$$d_0(\Psi^\delta, \Psi^{\delta, \lambda}) \leq \left( \sum_{k=1}^{k_{\max}} \nu_k(W^\delta_k) + \sum_{m=0}^{\infty} 2^{-m} \mu_m(W^\delta_k) \right) \max_{s_k=1}^{\delta^{-d}} \frac{L_A(\nu_k^{\delta, \lambda})}{\mu_k(\nu_k^{\delta, \lambda})} - 1,$$

which tends to 0 on $\Omega_1$ as $\lambda \to \infty$, according to (3.1).

Now, if we add the suitable up- and downroundings, we only change distances in the $d$-metric by an error term of order $1/\lambda$, which vanishes as $\lambda \to \infty$. This implies that $\Psi$ is a standard setting. It also follows easily that if $\sum_m \eta(m) \mu_m(W) < \infty$, then $\Psi$ is controlled.

3.4. Proof of Theorem 1.2

Abbreviate

$$\mathcal{Y}(r) = \left( \prod_{i \in I^\lambda} N(\lambda)^{(r^i - r^{i-1} - 1)} \right) \exp \left\{ -\gamma S(r) - \beta \mathcal{M}(r) \right\}, \quad \lambda > 0, r \in S_{k_{\max}}(X^\lambda),$$

and note that the partition function is given as

$$Z_{\lambda}^{\gamma, \beta}(X^\lambda) = \sum_{r \in S_{k_{\max}}(X^\lambda)} \mathcal{Y}(r). \quad (3.18)$$

Then Theorem 1.2 says that its large-$\lambda$ negative exponential rate is given as the infimum of $I(\Psi) + \gamma S(\Psi) + \beta \mathcal{M}(\Psi)$, taken over all admissible trajectory settings $\Psi$. Throughout the proof, we assume that the configuration $X^\lambda = X^\lambda(\omega)$ comes from some $\omega \in \Omega_1$ defined in (3.1).

Having proved Proposions 3.3 3.2 and 3.3, our strategy to prove Theorem 1.2 is the following. First, Proposition 3.3 gives a standard way how to construct from an admissible trajectory setting $\Psi$ a standard setting $\Psi$ that contains $\Psi$. Then the lower bound for the partition function is easily given in terms of the objects that are contained in any such $\Psi$ and using the logarithmic asymptotics for their combinatorics from Propositions 3.1 and 3.2 and finally taking the infimum over all such $\Psi$, respectively $\Psi$. The upper bound needs more care, since the entire sum over $r$ has to be handled. First of all, we show that the sum can be restricted for all $\lambda > 0$, modulo some error term that is negligible on the exponential scale, to the sum of those configurations whose congestion exponent is at most $CA$ for some appropriate large constant $C > 0$. This sum can be decomposed, for any $\delta \in \mathbb{B}$, to sums on configurations coming from a particular choice of empirical measures on the $\delta$-partitions of $W$. The number of these empirical measures and the sum on the partitions is negligible in the limit $\lambda \to \infty$, and the asymptotics of the sums on $r$ in these partitions can be evaluated with the help of
our spatial discretization procedure, using arguments of the proofs of Propositions 3.1 and 3.2 in the limit \( \lambda \to \infty \), followed by \( \delta \downarrow 0 \). Using these, we arrive at the said formula.

Let us give the details. We start with the proof of the lower bound. For any admissible trajectory setting \( \Psi \), we pick \( \Psi \) as in Proposition 3.3 and recall the configuration class \( J_{\delta,\lambda}^k(\Psi) \) from (2.6). Then, for any \( \lambda > 0 \) and \( \delta \in \mathbb{B} \),

\[
Z_{\lambda}^{\gamma,\beta}(X^\lambda) \geq \sum_{r \in J_{\delta,\lambda}(\Psi)} \mathcal{Y}(r) \geq \frac{\#J_{\delta,\lambda}^k(\Psi)}{\sup_{r \in J_{\delta,\lambda}^k(\Psi)} \prod_{i \in I^\lambda} N(\lambda)^{-1/(r_i-1)}} \exp \left\{ -\sup_{r \in J_{\delta,\lambda}^k(\Psi)} \left( \gamma \mathcal{S}(r) + \beta \mathcal{M}(r) \right) \right\}.
\]

Hence,

\[
\liminf_{\lambda \to \infty} \frac{1}{\lambda} \log Z_{\lambda}^{\gamma,\beta}(X^\lambda) \geq \liminf_{\delta \downarrow 0} \liminf_{\lambda \to \infty} \frac{1}{\lambda} \log \frac{\#J_{\delta,\lambda}^k(\Psi)}{\sup_{r \in J_{\delta,\lambda}^k(\Psi)} \prod_{i \in I^\lambda} N(\lambda)^{-1/(r_i-1)}} - \gamma \limsup_{\delta \downarrow 0} \sup_{r \in J_{\delta,\lambda}^k(\Psi)} \frac{1}{\lambda} \mathcal{S}(r) - \beta \limsup_{\lambda \to \infty} \sup_{r \in J_{\delta,\lambda}^k(\Psi)} \frac{1}{\lambda} \mathcal{M}(r) = -I(\Psi) - \gamma S(\Psi) - \beta M(\Psi).
\]

In the last step we also used Propositions 3.1 and 3.2 together with the fact that \( \Psi \) is controlled. Now take the supremum over all such \( \Psi \) on the r.h.s. of (3.20) to conclude that the lower bound in (1.24) holds.

The upper bound of Theorem 1.2 requires some additional work. We start from (3.18). For \( C > 0 \) we have

\[
Z_{\lambda}^{\gamma,\beta}(X^\lambda) = \sum_{r \in \mathcal{S}_{k_{\max}}(X^\lambda) : \mathcal{M}(r) \leq \lambda C} \mathcal{Y}(r) + \sum_{r \in \mathcal{S}_{k_{\max}}(X^\lambda) : \mathcal{M}(r) > \lambda C} \mathcal{Y}(r).
\]

Since the total mass of our \textit{a priori} measure has a bounded large-\( \lambda \) exponential rate (see Section 1.2.2) and \( \mathcal{S}, \mathcal{M} \) are bounded from below, we see that

\[
\limsup_{C \to \infty} \limsup_{\lambda \to \infty} \frac{1}{\lambda} \log \sum_{r \in \mathcal{S}_{k_{\max}}(X^\lambda) : \mathcal{M}(r) > \lambda C} \mathcal{Y}(r) = -\infty.
\]

Thus, for \( C \) sufficiently large, the exponential rate of \( Z_{\lambda}^{\gamma,\beta}(X^\lambda) \) is equal to the one of the first term on the right-hand side of (3.21). We additionally require \( \overline{C} \) so large that

\[
\inf_{\text{adm. traj. setting}, \ M(\Psi) \leq \overline{C}} (I(\Psi) + \gamma S(\Psi) + \beta M(\Psi)) = \inf_{\text{adm. traj. setting}} (I(\Psi) + \gamma S(\Psi) + \beta M(\Psi)).
\]

Let us write \( \mathcal{S}_{k_{\max}}(X^\lambda) = \{ r \in \mathcal{S}_{k_{\max}}(X^\lambda) : \mathcal{M}(r) \leq \lambda C \} \) and \( Z_{\lambda}^{\gamma,\beta}(X^\lambda) = \sum_{r \in \mathcal{S}_{k_{\max}}(X^\lambda)} \mathcal{Y}(r) \).

The upper bound of Theorem 1.2 follows as soon as we show that

\[
\limsup_{\lambda \to \infty} \frac{1}{\lambda} \log Z_{\lambda}^{\gamma,\beta}(X^\lambda) \leq -\sup_{\text{adm. trajectory setting}, \ M(\Psi) \leq \overline{C}} (I(\Psi) + \gamma S(\Psi) + \beta M(\Psi)).
\]

For fixed \( \lambda > 0 \) and \( \delta \in \mathbb{B} \), let us say that a collection of measures \( \Psi^{\delta,\lambda} = (\mu_{k,\lambda}^{\delta}, \mu_{m,\lambda}^{\delta})_{m=0}^{\infty} \) lies in \( G(\delta,\lambda)(X^\lambda) \) if all these measures take values in \( \frac{1}{\lambda} \mathbb{N}_0 \) only and satisfy the constraints \( \sum_{k=1}^{k_{\max}} \pi_{0,k}^{\delta,\lambda} = L_{\lambda}, \sum_{m=0}^{\infty} \mu_{m}^{\delta,\lambda} = L_{\lambda}^{\delta}, \sum_{k=1}^{k_{\max}} \sum_{l=1}^{k_{\max}} \pi_{l,k}^{\delta,\lambda} = \sum_{m=0}^{\infty} m \mu_{m}^{\delta,\lambda} \). We will write \( J_{\delta,\lambda}^k(\Psi^{\delta,\lambda}) \) for the set \( J_{\delta,\lambda}^k(\Psi) \) defined in (2.6). Then the union of \( J_{\delta,\lambda}^k(\Psi^{\delta,\lambda}) \) over all \( \Psi^{\delta,\lambda} \) with

\[
\sum_{m=0}^{\infty} \eta(m) \mu_{m}^{\delta,\lambda}(W_{\delta}) \leq \overline{C} \quad \text{is equal to}
\]

\[
\{ (R_{\lambda,k}^{\delta}(r))_{k=1}^{k_{\max}}, (P_{\lambda,m}^{\delta}(r))_{m=0}^{\infty} : r \in \mathcal{S}_{k_{\max},C}(X^\lambda) \},
\]

since these three equations characterize the tuple of the measures \( (R_{\lambda,k}^{\delta}(S))_{k=1}^{k_{\max}} \) and \( (P_{\lambda,m}^{\delta}(S))_{m=0}^{\infty} \) if \((S^i)_{i \in I^\lambda} \in \mathcal{S}_{k_{\max}}(X^\lambda)\).
Using this, we can estimate, for any \( \delta \in \mathbb{B} \),
\[
Z_\lambda^{\gamma, \beta, C}(X^\lambda) = \sum_{\Psi^{\delta, \lambda} \in G(\delta, \lambda): M(\Psi^{\delta, \lambda}) \leq C} \sum_{r \in J^{\delta, \lambda}(\Psi^{\delta, \lambda})} \Psi(r) \leq \#G(\delta, \lambda) \sup_{\Psi^{\delta, \lambda} \in G(\delta, \lambda): M(\Psi^{\delta, \lambda}) \leq C} \sum_{r \in J^{\delta, \lambda}(\Psi^{\delta, \lambda})} \Psi(r).
\]
Hence,
\[
\lim_{\lambda \to \infty} \frac{1}{\lambda} \log Z_\lambda^{\gamma, \beta, C}(X^\lambda) \leq \lim_{\delta \downarrow 0} \lim_{\lambda \to \infty} \frac{1}{\lambda} \log \#G(\delta, \lambda) + \lim_{\delta \downarrow 0} \lim_{\lambda \to \infty} \frac{1}{\lambda} \log \#J^{\delta, \lambda}(\Psi^{\delta, \lambda}) \sup_{r \in J^{\delta, \lambda}(\Psi^{\delta, \lambda})} \inf_{\lambda \to \infty} \frac{1}{\lambda} \log \#(G(r)) - \beta \lim_{\delta \downarrow 0} \lim_{\lambda \to \infty} \frac{1}{\lambda} \log \inf_{r \in J^{\delta, \lambda}(\Psi^{\delta, \lambda})} \frac{1}{\lambda} \mathcal{Y}(r).
\]
(3.25)

According to Lemma 3.3 below, the first term on the right-hand side is equal to zero. Now pick a sequence \( (\delta_n)_n \) and for each \( n \) a sequence \((\lambda_{n,j})_j\) along which the superior limits as \( n \to \infty \), respectively \( j \to \infty \), are realized. Now pick, for any \( n \) and \( j \), a maximizer \( \tilde{\Psi}_{\delta_n, \lambda_{n,j}} \). Pick \( \lambda_0 \) so large that \( N(\lambda) \leq 2\mu(W) \lambda \) for all \( \lambda \geq \lambda_0 \). Hence,
\[
\bigcup_{\lambda > \lambda_0, \delta \in \mathbb{B}} G(\delta, \lambda) \subseteq \left( \prod_{k=1}^{k_{\text{max}}} \mathcal{M}_{\leq 2\mu(W)}(W^k) \right) \times \mathcal{M}_{\leq 2\mu(W)}(W)^{\mathbb{N}_0},
\]
where we recall that \( \mathcal{M}_{\leq \alpha}(V) \) is the set of measures on a space \( V \) with total mass \( \leq \alpha \). (We recall from Section 2.1 that we conceive all measures on \( W^k_\delta \) as measures on \( W^k \).) Note that \( \mathcal{M}_{\leq 2\mu(W)}(W^k) \) is compact in the weak topology of \( \mathcal{M}(W^k) \) for any \( k \), according to Prohorov’s theorem.

Without loss of generality (using two diagonal sequence arguments), we can assume that for all \( n \in \mathbb{N} \), \( \tilde{\Psi}_{\delta_n, \lambda_{n,j}} \) converges coordinatewise weakly to a collection of measures \( \tilde{\Psi}_{\delta_n} = (\bigotimes_{k=1}^{k_{\text{max}}} (\tilde{\mu}_{m})_{m=0}^{\infty} \) as \( j \to \infty \), and \( \tilde{\Psi}_{\delta_n} \) converges coordinatewise weakly to a collection of measures \( \tilde{\Psi} \) as \( n \to \infty \). Then, it is clear that \( \tilde{\Psi} \) satisfies (i) from (1.19), and also that
\[
\lim_{n \to \infty} \lim_{j \to \infty} \sum_{k=1}^{k_{\text{max}}} \sum_{l=1}^{k-1} \pi_l \mu_{\delta_n, \lambda_{n,j}} = \sum_{k=1}^{k_{\text{max}}} \sum_{l=1}^{k-1} \pi_l \tilde{\mu}_k.
\]
In order to see that (iii) holds for \( \tilde{\Psi} \), it remains to show that \( \lim_{n \to \infty} \lim_{j \to \infty} \sum_{m=0}^{\infty} m \mu_{\delta_n, \lambda_{n,j}} = \sum_{m=0}^{\infty} m \tilde{\mu}_m \). For \( N \in \mathbb{N} \) and for any continuous function \( f: W \to \mathbb{R} \), we estimate
\[
\left| \left( \sum_{m=0}^{N} m \left( \mu_{m, \lambda_{n,j}} - \tilde{\mu}_m \right), f \right) \right| \leq \sum_{m=0}^{N} m \left| \left( \mu_{m, \lambda_{n,j}} - \tilde{\mu}_m, f \right) \right| + \sum_{m=N+1}^{\infty} \left| f \right| \left( \left( \mu_{m, \lambda_{n,j}} - \tilde{\mu}_m, f \right) \right).
\]
The first term on the r.h.s. clearly tends to 0 as \( j \to \infty \), followed by \( n \to \infty \), for any fixed \( N \). The second term can further be estimated from above as follows
\[
\left| f \right| \left( \sum_{m=0}^{N} \eta(m) \left( \sup_{\tilde{m} > N} \frac{\tilde{m}}{\eta(m)} \left( \mu_{m, \lambda_{n,j}}(W) + \tilde{\mu}_m(W) \right) \right) \right) \leq \left| f \right| \left( \sum_{m=0}^{N} \frac{m}{\eta(m)} C \right).
\]
By the assumption that \( \left( \eta(N)/N \right) \to \infty \) as \( N \to \infty \), the right-hand side tends to 0. One can analogously show that \( \sum_{m=0}^{\infty} \mu_{m, \lambda_{n,j}} \) tends to \( \sum_{m=0}^{\infty} \tilde{\mu}_m \) as \( j \to \infty \) followed by \( n \to \infty \), and hence condition (ii) from (1.19) holds. Also we have \( \sum_{m=0}^{\infty} \eta(m) \tilde{\mu}_m(W) \leq C \). Altogether, \( \tilde{\Psi} \) is an admissible trajectory setting.
Now, using the arguments of the proofs of Propositions 3.1 and 3.2 (which also involve the coarsened limits $\Psi^{\delta_n}$ for fixed $n \in \mathbb{N}$) for the subsequential limits $j \to \infty$ followed by $n \to \infty$, we conclude that

$$
\lim_{n \to \infty} \lim_{j \to \infty} \inf_{\tilde{r} \in J^{\delta_n,\lambda_n,j}(\tilde{\Psi}^{\delta_n,\lambda_n,j})} \prod_{i \in I^{\lambda_n,j}} N(\lambda_n,j)^{-(r_{i-1}-1)} = I(\tilde{\Psi})
$$

and, using the boundedness and continuity of each $f_k$ on $\mathcal{M}_{\leq 2\mu(W)}(W) \times W^k$,

$$
\lim_{n \to \infty} \lim_{j \to \infty} \inf_{\tilde{r} \in J^{\delta_n,\lambda_n,j}(\tilde{\Psi}^{\delta_n,\lambda_n,j})} \frac{1}{\lambda_n,j} \mathcal{E}(r) = S(\tilde{\Psi}).
$$

Furthermore, the lower semicontinuity of $\mathcal{M}(W)^{\mathbb{N}} \mapsto (-\infty, \infty]$, $(\nu_m)_{m \in \mathbb{N}_0} \mapsto \sum_{m \in \mathbb{N}_0} \eta(m)\nu_m(W)$ together with Fatou’s lemma implies that

$$
-\beta \lim_{n \to \infty} \inf_{j \to \infty} \inf_{\tilde{r} \in J^{\delta_n,\lambda_n,j}(\tilde{\Psi}^{\delta_n,\lambda_n,j})} \frac{1}{\lambda_n,j} \mathcal{M}(r) \leq -\beta \mathcal{M}(\tilde{\Psi}). \tag{3.27}
$$

Thus, we conclude that (3.23) (and therefore the upper bound in Theorem 1.2) holds, as soon as Lemma 3.4 is formulated and verified. This we do now.

**Lemma 3.4.** For any $\delta \in \mathbb{B}$, almost surely,

$$
\limsup_{\lambda \to \infty} \frac{1}{\lambda} \log \#G(\delta, \lambda) = 0.
$$

**Proof.** For $\lambda > 0$, let $G_1(\delta, \lambda)$ denote the set of $(\nu^{\delta,\lambda}_{\max})_{k=1}^{\kappa_{\max}}$ satisfying part [5] from Definition 2.1. It is easily seen that its cardinality increases only polynomially in $\lambda$. Now, given $(\nu^{\delta,\lambda}_{\max})_{k=1}^{\kappa_{\max}} \in G_1(\delta, \lambda)$, we will give an upper bound for the number of $(\mu^{\delta,\lambda}_{\max})_{m=0}^{\kappa_{\max}-1}$ such that the pair of these tuples is in $G(\delta, \lambda)$. This is much more demanding, since there is a priori no upper bound for $m$. We will provide a $\lambda$-dependent one.

For any $\lambda > 0$, $\Psi^{\delta,\lambda} \in G(\delta, \lambda)$ and $j = 1, \ldots, \delta - d$ we have that

$$
\sum_{m=0}^{\kappa_{\max}} m \mu^{\delta,\lambda}_{\max}(W^{\delta}_j) = \sum_{k=1}^{\kappa_{\max}} \sum_{i=1}^{k} \pi^{\delta,\lambda}_{i\mu^{\delta,\lambda}_{\max}}(W^{\delta}_j) \leq (\kappa_{\max} - 1)N(\lambda),
$$
in particular $\mu^{\delta,\lambda}_{\max}(W^{\delta}_j) = 0$ for $m > (\kappa_{\max} - 1)N(\lambda)$. We also have that the numbers $\mu^{\delta,\lambda}_{0}(W^{\delta}_j), \ldots, \mu^{\delta,\lambda}_{(\kappa_{\max} - 1)N(\lambda)}(W^{\delta}_j)$, are $\frac{1}{\lambda}$ times nonnegative integers.

Let $\varepsilon > 0$ be fixed. We claim that for all sufficiently large $\lambda > 0$, there are not more than $\varepsilon N(\lambda) \sim \varepsilon \lambda \mu(W)$ nonzero ones out of these quantities. Indeed, if there were at least $\lceil \varepsilon N(\lambda) \rceil$ nonzero ones, denoted $\mu^{\delta,\lambda}_{\max}(W^{\delta}_j), \ldots, \mu^{\delta,\lambda}_{\lceil \varepsilon N(\lambda) \rceil - 1}(W^{\delta}_j)$ with $0 < m_0 < m_1 < \ldots < m_{\lceil \varepsilon N(\lambda) \rceil - 1} \leq (\kappa_{\max} - 1)N(\lambda)$, then we could estimate

$$
(\kappa_{\max} - 1)N(\lambda) \geq \sum_{m=0}^{\lceil \varepsilon N(\lambda) \rceil - 1} \lambda m \mu^{\delta,\lambda}_{\max}(W^{\delta}_j) \geq \sum_{m=0}^{\lceil \varepsilon N(\lambda) \rceil - 1} \lambda m \mu^{\delta,\lambda}_{m}(W^{\delta}_j) \mathbb{I}\{\mu^{\delta,\lambda}_{m}(W^{\delta}_j) > 0\}
$$

$$
= \sum_{m=0}^{\lceil \varepsilon N(\lambda) \rceil - 1} \lambda m \mu^{\delta,\lambda}_{m}(W^{\delta}_j) \mathbb{I}\{\mu^{\delta,\lambda}_{m}(W^{\delta}_j) \geq \frac{1}{\lambda}\} \geq \sum_{m=0}^{\lceil \varepsilon N(\lambda) \rceil - 1} \lambda m \mathbb{I}\{\mu^{\delta,\lambda}_{m}(W^{\delta}_j) \geq \frac{1}{\lambda}\} \geq \sum_{m=0}^{\lceil \varepsilon N(\lambda) \rceil - 1} \lambda m \sim \frac{1}{2}(\varepsilon N(\lambda))(\varepsilon N(\lambda) - 1),
$$

which is a contradiction for all $\lambda > 0$ sufficiently large.

Now, $\#G(\delta, \lambda)$ can be estimated as follows. Let us first fix $(\nu^{\delta,\lambda}_{k})_{k=1}^{\kappa_{\max}} \in G_1(\delta, \lambda)$, i.e., satisfying part [5] from Definition 2.1, and let us count the number of $(\mu^{\delta,\lambda}_{m})_{m=0}^{\kappa_{\max} - 1}$ such that $(\nu^{\delta,\lambda}_{k})_{k=1}^{\kappa_{\max}} \in G(\delta, \lambda)$. Out of the $\kappa_{\max}\delta^{-d}N(\lambda)$ quantities $\mu^{\delta,\lambda}_{0}(W^{\delta}_j), \ldots, \mu^{\delta,\lambda}_{(\kappa_{\max} - 1)N(\lambda)}(W^{\delta}_j)$, at most $\lceil \varepsilon N(\lambda) \rceil\delta^{-d}$ are nonzero. The number of
ways to choose them equals \(\binom{k_{\text{max}}N(\lambda)\delta^{-d}}{\lceil \varepsilon N(\lambda)\rceil\delta^{-d}}\). Having chosen \(\lceil \varepsilon N(\lambda)\rceil\delta^{-d}\) potentially nonzero ones so that the remaining \(k_{\text{max}}\delta^{-d}N(\lambda) - \lceil \varepsilon N(\lambda)\rceil\delta^{-d}\) ones are equal to zero, according to part (ii) of Definition 2.1 we note that the potentially nonzero ones sum up to \(N(\lambda)\), and each one has a value in \(\frac{1}{\lambda}N_0\). For this, there are at most \(\binom{N(\lambda)+\lceil \varepsilon N(\lambda)\rceil\delta^{-d}-1}{\lceil \varepsilon N(\lambda)\rceil\delta^{-d}-1}\) combinations, for any choice of the set of the potentially nonzero ones. Therefore, using Stirling’s formula as in (3.31), for any sufficiently large \(\lambda\), we have the following estimate

\[
\# G(\delta, \lambda) \leq \# G_1(\delta, \lambda) \left(\frac{k_{\text{max}}N(\lambda)\delta^{-d}}{\lceil \varepsilon N(\lambda)\rceil\delta^{-d}}\right) \left(\frac{N(\lambda) + \lceil \varepsilon N(\lambda)\rceil\delta^{-d} - 1}{\lceil \varepsilon N(\lambda)\rceil\delta^{-d} - 1}\right) = e^{o(\lambda)} \exp \left(-\lambda \mu(W)\left(\frac{1}{\delta}\right)^{\delta^{-d}} \log \left(\frac{k_{\text{max}} - \varepsilon}{k_{\text{max}}\delta^{-d}}\right) + \varepsilon \delta^{-d} \log \left(\frac{\varepsilon \delta^{-d}}{1 + \varepsilon \delta^{-d}}\right)\right) \times \exp \left(-\lambda \mu(W)\left(\varepsilon \delta^{-d} \log \left(1 + \varepsilon \delta^{-d}\right) + \log \frac{1}{1 + \varepsilon \delta^{-d}}\right)\right).
\]

Making \(\varepsilon \downarrow 0\), we conclude that \(\limsup_{\lambda \to \infty} \frac{1}{\lambda} \log \# G(\delta, \lambda) = 0\).

3.5. The large deviation principle: proof of Theorem 1.4(i). In this section, we prove Theorem 1.4(i). The combinatorial essence of this theorem has already been proven in Proposition 3.1 including the relations with \(\delta\)-coarsenings. What remains to be done is to relate this to the coordinatewise weak convergence on \(A\). We will be able to use some of the arguments of Section 3.4.

The lower semicontinuity of \(I + \mu(W)\log k_{\text{max}}\) was already discussed in Section 1.3. the nonnegativity in Section 1.5. These together mean that \(I + \mu(W)\log k_{\text{max}}\) is a rate function.

We proceed with the proof of the lower bound. Let \(G \subseteq A\) be open. If \(\inf_G I = \infty\), then there is nothing to show, therefore let us assume that there exists \(\Psi \in G\) with \(I(\Psi) < \infty\). According to Proposition 3.3 there is a standard setting \(\Psi\) containing \(\Psi\). Since \(G\) is open, there exists \(\rho > 0\) such that \(B_\rho(\Psi) \subseteq G\). Let us choose \(\delta_0 \in \mathbb{B}\) and, for any \(\mathbb{B} \ni \delta \leq \delta_0\), some \(\lambda_0 = \lambda_0(\delta) > 0\) such that \(\Psi^\delta, \Psi^\delta, \Psi^\delta, \lambda \in B_\rho(\Psi)\) for any \(\lambda > \lambda_0\). Now we can estimate, for these \(\delta\) and \(\lambda\),

\[
P_{\lambda, X, \lambda}^{0, 0}(\Psi_X(S) \in G) \geq P_{\lambda, X, \lambda}^{0, 0}(\Psi_X(S) \in B_\rho(\Psi)) \geq P_{\lambda, X, \lambda}^{0, 0}(\Psi_X(S)^\delta = \Psi^\delta, \lambda) = \frac{1}{Z_{\lambda, X, \lambda}^{0, 0}(X^\lambda)} \sum_{r \in J^\delta, \lambda(\Psi^\delta, \lambda)} \prod_{i \in I^\delta} N(\lambda)^{r_i - 1} \geq \frac{\# J^\delta, \lambda(\Psi^\delta, \lambda)}{\lambda_{\text{max}} \sup_{r \in J^\delta, \lambda(\Psi^\delta, \lambda)} \prod_{i \in I^\delta} N(\lambda)^{r_i - 1}}.
\]

Now, using Proposition 3.1 and the fact that \(N(\lambda) / \lambda \to \mu(W)\), we obtain

\[
\liminf_{\lambda \to \infty} \frac{1}{\lambda} \log P_{\lambda, X, \lambda}^{0, 0}(\Psi_X(S) \in G) \geq -\mu(W) \log k_{\text{max}} - I(\Psi).
\]

Note that \(\Psi\) is not necessarily controlled because \(M(\Psi) < \infty\) is not guaranteed. However, since for all \(\delta \in \mathbb{B}, s = 1, \ldots, \delta^{-d}, \lambda > 0, \mu^\delta_m(W_\delta^\delta) / \mu_m^\delta(W_\delta^\delta)\) does not depend on \(m\), we easily see that Proposition 3.1 holds for this \(\Psi\) as well. Now, take the supremum over \(\Psi \in G \cap \{I < \infty\}\) to conclude that the lower bound holds.

We continue with the upper bound. Let \(F \subseteq A\) be closed. Let us choose an increasing sequence \((\lambda_n)_{n \in \mathbb{N}}\) of positive numbers along which the limit superior in (1.30) is realized. For \(\lambda > 0\), let us put

\[
O(\lambda) = \{\Psi \in A: P_{\lambda, X, \lambda}^{0, 0}(\Psi_X(S) = \Psi) > 0\}.
\]

If for all but finitely many \(n \in \mathbb{N}\) we have \(F \cap O(\lambda_n) = \emptyset\), then

\[
\limsup_{\lambda \to \infty} \frac{1}{\lambda} \log P_{\lambda, X, \lambda}^{0, 0}(\Psi_X(S) \in F) = -\infty.
\]
Therefore, without loss of generality, we can assume that $O(\lambda_n) \cap F$ is non-empty for all $n \in \mathbb{N}$. For $\delta \in \mathbb{B}$ and $A \subset \mathcal{A}$, let us write $A^\delta = \{ \Psi^\delta : \Psi \in A \}$, where $\Psi^\delta$ is the coordinatewise $\delta$-coarsened version of $\Psi$. Then we have

$$
P_{\lambda_n, X, \lambda_n}^{0,0} (\Psi_{\lambda_n} (S) \in F) = P_{\lambda_n, X, \lambda_n}^{0,0} (\Psi_{\lambda_n} (S) \in F \cap O(\lambda_n)) = P_{\lambda_n, X, \lambda_n}^{0,0} ((\Psi_{\lambda_n} (S))^\delta \in (F \cap O(\lambda_n))^\delta)
\leq \#(F \cap O(\lambda_n))^\delta \sup_{\Psi \in F \cap O(\lambda_n)} \frac{\#J_{\delta, \lambda_n}^\delta (\Psi^\delta)}{k_{\max}^N(\lambda_n) \inf_{\nu \in J_{\delta, \lambda_n}^\delta (\Psi^\delta)} \prod_{i \in I_{\lambda_n}} N(\lambda_n)^r_{i-1-1}}.
\tag{3.29}
$$

It is clear that $(F \cap O(\lambda_n))^\delta \subseteq G(\delta, \lambda_n) = (O(\lambda_n))^\delta$ for all $n \in \mathbb{N}$ and $\delta \in \mathbb{B}$, where $G(\delta, \lambda_n)$ was defined in Section 3.4. Hence, by Lemma 3.3

$$
\lim_{\delta, n \to \infty} \frac{1}{\lambda_n} \log \#(F \cap O(\lambda_n))^\delta = 0.
$$

It remains to show that

$$
\lim_{\delta, n \to \infty} \frac{1}{\lambda_n} \log \left[ \sup_{\Psi \in F \cap O(\lambda_n)} \frac{\#J_{\delta, \lambda_n}^\delta (\Psi^\delta)}{\inf_{\nu \in J_{\delta, \lambda_n}^\delta (\Psi^\delta)} \prod_{i \in I_{\lambda_n}} N(\lambda_n)^r_{i-1-1}} \right] \leq - \inf_{\Psi \in F} I(\Psi). \tag{3.30}
$$

One can do this analogously to the proof of the upper bound of Theorem 1.2 starting from (3.29). Indeed, using Prohorov’s theorem together with a diagonal sequence argument, we find $\Psi^* \in \mathcal{A}$ that the maximizer in (3.29) converges to along a subsequence of $\delta$’s and $\lambda_n$’s. The limit lies in $F$ because $F$ is closed. Using the lower semicontinuity of $I$ together with Fatou’s lemma, we conclude that the left-hand side of (3.30) is not larger than $-I(\Psi^*)$, which itself is not larger than $- \inf_F I$. This finishes the proof of the upper bound in Theorem 1.4 (i).

4. Analysis of the minimizers

This section is devoted to the proof of Proposition 1.3. In particular, in Section 4.1, we show that the infimum in (1.24) is attained and, for any minimizer $\Psi = ((\nu_k)_{k=1}^{k_{\max}}, (\mu_m)_{m=0}^{\infty})$, for any $k \in [k_{\max}]$, $\mu \otimes k$ is absolutely continuous with respect to $\nu_k$ and $\mu$ is absolutely continuous with respect to each $\mu_m$. This is a prerequisite for perturbing the minimizer in many admissible directions. In Section 4.2, we finish the proof of Proposition 1.3 by deriving the Euler–Lagrange equations. For the rest of the section, we fix all parameters $W, \mu, \gamma, \beta$ and $k_{\max}$. Moreover, we use the following representation of $I$ from (1.23).

$$
I(\Psi) = \sum_{k=1}^{k_{\max}} H_{W_k}(\mu \otimes M^{\otimes (k-1)}) + \sum_{m=0}^{\infty} H_{W}(\mu_m | \mu) - \mu_m(W) \log \left( \frac{e\mu(W)}{m!} \right).
$$

4.1. Existence and positivity of the minimizers. We start with the following lemma, which follows almost immediately from the arguments of the proof of the upper bound of Theorem 1.2 in Section 3.4.

**Lemma 4.1.** The set of minimizers for the variational formula in (1.24) is non-empty, compact and convex.

**Proof.** Recall that the three functionals $I, S, M$ are lower semicontinuous and convex. Furthermore, it is clear that we can restrict the infimum in (1.24) to those $\Psi$ that satisfy also $M(\Psi) \leq C$ for any sufficiently large $C$. But, as we have seen in Section 3.4, this set of $\Psi$’s is compact. From this, all our assertions easily follow. \qed
Now we prove that, for each minimizer $\Psi$, $\mu^{\otimes k}$ is absolutely continuous with respect to $\nu_k$ and $\mu$ is absolutely continuous with respect to each $\mu_m$. (Note that the opposite absolute continuities are true by finiteness of the entropies.) We need to show this only for $k_{\max} > 1$, as we explained after Proposition [1.3]

**Lemma 4.2.** If $k_{\max} > 1$ and $\Psi = ((\nu_k)_{k=1}^{k_{\max}}, (\mu_m)_{m=0}^{\infty})$ is a minimizer of (1.24), then $\mu^{\otimes k} \ll \nu_k$ for any $k \in [k_{\max}]$, and $\mu \ll \mu_m$ for any $m \in \mathbb{N}_0$.

**Proof.** The essence of the proof is the following. The functionals $M(\cdot)$ and $S(\cdot)$ are linear in each $\mu_m$ respectively $\nu_k$, as well as the third term in $I(\cdot)$ in (1.23) in each $\nu_k$. On the other hand, the function $x \mapsto x \log x$ has the slope $-\infty$ at $x \downarrow 0$. We show the following assertions about the minimizer $\Psi$ step by step as follows. Recall that $M = \sum_{m \in \mathbb{N}_0} m \mu_m = \sum_{k \in [k_{\max}]} \sum_{l=1}^{k-1} \pi_l \nu_k$. We write $\geq$ and $>$, respectively, between measures in $M(W^k)$ if their difference lies in $M(W^k)$, respectively in $M(W^k) \setminus \{0\}$.

Fix a measurable set $A \subset W$ such that $\mu(A) > 0$. Then we have:

1. $M(A) > 0$.
2. for any $m_1 < m_0 < m_2$ such that $\mu_{m_1}(A) > 0$ and $\mu_{m_2}(A) > 0$, also $\mu_{m_0}(A) > 0$.
3. $\mu_0(A) > 0$.
4. $\mu_m(A) > 0$ for any $m \geq k_{\max}$.
5. $\nu_k(A^k) > 0$ for any $k \in [k_{\max}]$.

Indeed, these steps are verified respectively as follows. In each of the steps, for $\varepsilon \in (0, 1)$, we construct an admissible trajectory setting $\Psi^\varepsilon = ((\nu_k)_{k=1}^{k_{\max}}, (\mu_m)_{m=0}^{\infty})$ such that $I(\Psi^\varepsilon) + \gamma S(\Psi^\varepsilon) + \beta M(\Psi^\varepsilon) < I(\Psi) + \gamma S(\Psi) + \beta M(\Psi)$ for sufficiently small $\varepsilon > 0$, and therefore $\Psi$ is not a minimizer of (1.24).

1. If $M(A) = 0$, then in particular $\mu_0(A) = \nu_1(A) = \mu(A)$ and $\mu_m(A) = 0$ for all $m > 0$. Also, $\pi_1 \nu_2(A) = \nu_2(W \times A) = 0$, according to the definition of $M$.

Let us define $\Psi^\varepsilon$ as follows: $\nu_2^\varepsilon = (1 - \varepsilon) \nu_2 + \varepsilon (\mu^{\otimes 2})/\mu(W)$, $\nu_k^\varepsilon = (1 - \varepsilon) \nu_k$ for $k \neq 2$, $\mu_1^\varepsilon = (1 - \varepsilon) \mu_1 + \varepsilon \mu$ and $\mu_m^\varepsilon = (1 - \varepsilon) \mu_m$ for $m \neq 1$. Then we compute and estimate the three terms of the entropy $I(\Psi)$ as follows.

\[
\sum_{k=1}^{k_{\max}} H_{W^k} (\nu_k^\varepsilon \mid \mu^{\otimes (M^\varepsilon)} \otimes (k-1))
\leq \sum_{k=1}^{k_{\max}} H_{W \times (W \setminus A)^{k-1}} ((1 - \varepsilon) \nu_k \mid \mu^{\otimes (M^\varepsilon)} \otimes (k-1)) + H_{W \times A} (\varepsilon \mu^{\otimes 2} / \mu(W) \mid \varepsilon \mu^{\otimes 2}) + O(\varepsilon)
\leq \sum_{k=1}^{k_{\max}} H_{W^k} (\mu \otimes M^{\otimes (k-1)}) + O(\varepsilon),
\]

furthermore

\[
\sum_{m=0}^{\infty} H_{W} (\mu_m^\varepsilon \mid \mu) - \mu_m^\varepsilon(W) \log (e \mu(W))^{-m} m! 
\leq H_{W}((1 - \varepsilon) \mu_m \mid \mu) - \mu_m(W) \log (e \mu(W))^{-m} m! + \mu(A) \varepsilon \log \varepsilon + O(\varepsilon).
\]

For the second term we used the convexity of the relative entropy in the form

\[
H_{W}((1 - \varepsilon) \nu_1 + \varepsilon \mu \mid \mu) \leq (1 - \varepsilon) H_{W}(\nu_1 \mid \mu) \leq H_{W}(\nu_1 \mid \mu) + O(\varepsilon).
\]
This in turn follows from [HJKP15] Lemmas 3.10, 3.11, which implies that, for any \( k \in \mathbb{N}, \xi, \eta \in \mathcal{M}(W^k) \) with \( \eta \neq 0 \) and \( \xi \ll \eta \),

\[
|H_{W^k}(\xi \mid \eta) - H_{W^k}((1 - \varepsilon)\xi \mid \eta)| \asymp \varepsilon.
\]

It follows that, as \( \varepsilon \downarrow 0 \),

\[
I(\Psi^\varepsilon) + \gamma S(\Psi^\varepsilon) + \beta M(\Psi^\varepsilon) - [I(\Psi) + \gamma S(\Psi) + \beta M(\Psi)] \leq O(\varepsilon) + \mu(A)\varepsilon \log \varepsilon,
\]

which is negative for all sufficiently small \( \varepsilon > 0 \). Thus, \( \Psi \) is not a minimizer.

(2) If \( M(A) > 0 \) but \( \mu_{m_1}(A) > 0 \), \( \mu_{m_2}(A) > 0 \) and \( \mu_{m_0}(A) = 0 \) for some \( m_1 < m_0 < m_2 \), then let \( \nu_k^\varepsilon = \nu_k \) for all \( k \in [k_{\text{max}}] \) and let \( \mu_{\text{max}}^{m_0} = (1 - \varepsilon)\mu_{m_0} + \varepsilon(\alpha_1\mu_{m_1} + \alpha_2\mu_{m_2}), \mu_{m_1}^{m_0} = (1 - \alpha_1\varepsilon)\mu_{m_1}, \mu_{m_2}^{m_0} = (1 - \varepsilon\alpha_2)\mu_{m_2} \), where \( \alpha_1, \alpha_2 \in (0,1) \) are such that \( \alpha_1 + \alpha_2 = 1 \) and \( m_1\alpha_1 + m_2\alpha_2 = m_0 \). Then, \( \Psi^\varepsilon \) is an admissible trajectory setting with \( M^\varepsilon = M \). It follows similarly to the previous computation that \( I(\Psi^\varepsilon) + \gamma S(\Psi^\varepsilon) + \beta M(\Psi^\varepsilon) < I(\Psi) + \gamma S(\Psi) + \beta M(\Psi) \) for all sufficiently small \( \varepsilon > 0 \). However, instead of (4.1), we have

\[
\sum_{m=0}^{\infty} H_W(\mu^{m^\varepsilon}_m \mid \mu) - \mu^{m^\varepsilon}(W) \log \frac{(e\mu(W))^{-m}}{m!} \leq \sum_{m=0}^{\infty} H_W(\mu_m \mid \mu) - \mu_m(W) \log \frac{(e\mu(W))^{-m}}{m!} + (\alpha_1\mu_{m_1}(A) + \alpha_2\mu_{m_2}(A))\varepsilon \log \varepsilon + O(\varepsilon),
\]

as \( \varepsilon \downarrow 0 \).

(3) If \( M(A) > 0 \) but \( \mu_0(A) = 0 \), let \( \nu_k^\varepsilon = (1 - \varepsilon)\nu_k \) for all \( 1 < k \leq k_{\text{max}}, \mu_k^{\varepsilon} = (1 - \varepsilon)\mu_k \) for all \( m > 0, \mu_0^{\varepsilon} = \varepsilon\mu + (1 - \varepsilon)\mu_0 \) and \( \mu^\varepsilon = (1 - \varepsilon)\mu_1 + \varepsilon\mu \). It is again sufficient to consider the entropy terms in \( I \). The summands on \( k > 1 \) can be estimated as follows.

\[
\sum_{k=2}^{k_{\text{max}}} H_W(\nu_k^\varepsilon \mid \mu \otimes (M^\varepsilon)^{(k-1)}) = \sum_{k=2}^{k_{\text{max}}} H_W((1 - \varepsilon)\nu_k \mid (1 - \varepsilon)^{k-1}\mu \otimes M^{k-1}) \leq \sum_{k=2}^{k_{\text{max}}} H_W(\nu_k \mid \mu \otimes M^{(k-1)}) + O(\varepsilon).
\]

The summand for \( k = 1 \) can be estimated with the help of (4.2). For the summand for \( m = 0 \), we have

\[
H_W(\mu_0^{\varepsilon} \mid \mu) = H_W\backslash A((1 - \varepsilon)\mu_0 + \varepsilon\mu \mid \mu) + \mu(A)\varepsilon \log \varepsilon
\]

\[
\leq H_W\backslash A((1 - \varepsilon)\mu_0 \mid \mu) + \mu(A)\varepsilon \log \varepsilon + O(\varepsilon) = H_W(\mu_0 \mid \mu) + \mu(A)\varepsilon \log \varepsilon + O(\varepsilon).
\]

while the remaining sum is handled as follows.

\[
\sum_{m=1}^{\infty} H_W(\mu^{m^\varepsilon}_m \mid \mu) - \mu^{m^\varepsilon}(W) \log \frac{(e\mu(W))^{-m}}{m!} = \sum_{m=1}^{\infty} H_W((1 - \varepsilon)\mu_m \mid \mu) - \mu_m(W) \log \frac{(e\mu(W))^{-m}}{m!} + O(\varepsilon).
\]

Thus, (4.3) holds also here, which implies the claim.

(4) If \( M(A) > 0 \) but \( \mu_{m_0}(A) = 0 \) for some \( m_0 \geq k_{\text{max}}, \) let \( \mu_{m_0}^{\varepsilon} = (1 - \varepsilon)\mu_{m_0} + \varepsilon M/m_0, \mu_m^{\varepsilon} = (1 - \varepsilon)\mu_m \) for \( m \notin \{0,m_0\} \) and, moreover \( \nu_k^{\varepsilon} = \nu_k \) for all \( k \in [k_{\text{max}}] \).

\[
\sum_{m=1}^{\infty} m\mu^{m^\varepsilon}_m = (1 - \varepsilon) \sum_{m=1}^{\infty} m\mu_m + \frac{\varepsilon m_0}{m_0} \sum_{k=1}^{k_{\text{max}}} \sum_{l=1}^{k-1} \pi_l \nu_k = \sum_{k=1}^{k_{\text{max}}} \sum_{l=1}^{k-1} \pi_l \nu_k,
\]

as required.
On the other hand, we have

$$
\mu - \sum_{m=1}^{\infty} \mu_m^e \geq \mu - (1 - \varepsilon) \sum_{m=1}^{\infty} \mu_m - \varepsilon (k_{\max} - 1) \mu_0 \geq (1 - \varepsilon) \mu - (1 - \varepsilon) \sum_{m=1}^{\infty} \mu_m = (1 - \varepsilon) \mu_0.
$$

Therefore, if we put $\mu_0 = \mu - \sum_{m=1}^{\infty} \mu_m^e$, then $\mu_0 \geq (1 - \varepsilon) \mu_0$ and $\Psi^e$ is an admissible trajectory setting. Now we can proceed analogously to (3) to conclude that $I(\Psi^e) + \gamma S(\Psi^e) + \beta M(\Psi^e) < I(\Psi) + \gamma S(\Psi) + \beta M(\Psi)$ for sufficiently small $\varepsilon > 0$.

The proof of (3) is very similar to the ones of (2), (3) and (4), therefore we leave it to the reader.

\[\square\]

4.2. Deriving the Euler–Lagrange equations. In this section, we finish the proof of Proposition 4.3. According to the results of Section 4.1, now we see that (4.24) exhibits at least one minimizer, and all minimizers have almost everywhere positive Lebesgue density on the corresponding powers of $\operatorname{supp} \mu$. Knowing this, we now carry out the perturbation analysis for the minimizer(s) of the optimization problem in (1.24) and derive the shape of the minimizers in most explicit terms.

We use the method of Lagrange multipliers in the framework of a perturbation argument. Let $\Psi = ((\nu_k)_{k=1}^{k_{\max}}, (\mu_m)_{m=0})$ minimize (1.24). Fix any collection of signed measures $\Phi = ((\tau_k)_{k=1}^{k_{\max}}, (\sigma_m)_{m=0})$ such that only finitely many $\sigma_m$’s are different from zero, each $\tau_k$ and each $\sigma_m$ has a simple function as a Lebesgue density and they satisfy the following constraints:

\[
\begin{align*}
(i) & \quad \sum_{k=1}^{k_{\max}} \pi_0 \tau_k = 0, \\
(ii) & \quad \sum_{m=0}^{\infty} \sigma_m = 0, \\
(iii) & \quad \sum_{m=0}^{\infty} m \sigma_m = \sum_{k=1}^{k_{\max}} \sum_{l=1}^{k_{\max}} \pi_l \tau_k.
\end{align*}
\]

Then it follows from Lemma 4.2 that, for any $\varepsilon \in \mathbb{R}$ with sufficiently small $|\varepsilon|$, $\Psi + \varepsilon \Phi = ((\nu_k + \varepsilon \tau_k)_{k=1}^{k_{\max}}, (\mu_m + \varepsilon \sigma_m)_{m=0})$ is a collection of (non-negative!) measures that satisfies (1.19) and is therefore admissible in the variational formula in (1.24). That (1.19) is satisfied follows easily from (1.14). Furthermore, the non-negativity follows from the fact that each $\tau_k$ and each $\sigma_m$ is a finite linear combination of measures of the form $\mathbb{I}_A \mathcal{L}_{\nu}$ with $A \subset W$. Since only finitely many such summands are involved, there is a constant $C > 0$ such that $|\tau_k| \leq C \mu_k$ and $|\sigma_m| \leq C \mu_m$ for any $k \in [k_{\max}]$ and $m \in \mathbb{N}_0$, and therefore it suffices to take $|\varepsilon| < 1/C$.

From minimality, we deduce that

\[
0 = \frac{\partial}{\partial \varepsilon} \bigg|_{\varepsilon=0} \left( I(\Psi + \varepsilon \Phi) + \gamma S(\Psi + \varepsilon \Phi) + \beta M(\Psi + \varepsilon \Phi) \right).
\]

We calculate the latter two terms as

\[
\frac{\partial}{\partial \varepsilon} \bigg|_{\varepsilon=0} \left( \gamma S(\Psi + \varepsilon \Phi) + \beta M(\Psi + \varepsilon \Phi) \right) = \gamma \sum_{k \in [k_{\max}]} \langle \tau_k, f_k \rangle + \beta \sum_{m \in \mathbb{N}_0} \eta(m) \sigma_m(W),
\]

where, as before, we used the notation $\langle \nu, f \rangle$ for the integral of a function $f$ with respect to a measure $\nu$. Abbreviating $M = \sum_{k \in [k_{\max}]} \sum_{l=1}^{k_{\max}} \pi_l \nu_k$ and $M_T = \sum_{k \in [k_{\max}]} \sum_{l=1}^{k_{\max}} \pi_l \tau_k$, we see that

\[
\frac{\partial}{\partial \varepsilon} \bigg|_{\varepsilon=0} I(\Psi + \varepsilon \Phi) = \sum_{k \in [k_{\max}]} \langle \tau_k, 1 + \log \frac{d \nu_k}{d \mu_{\varepsilon k}} \rangle + \sum_{m \in \mathbb{N}_0} \langle \sigma_m, 1 + \log \frac{d \mu_m}{d \mu} \rangle - \langle M_T, 1 + \log \frac{d M}{d \mu} \rangle.
\]

Summarizing, we obtain from (4.5) that

\[
0 = \langle \Phi, ((h_k)_{k \in [k_{\max}]}, (g_m)_{m \in \mathbb{N}_0}) \rangle,
\]
where
\[ h_k = \gamma f_k + 2 - k + \log \frac{d\nu_k}{d(\mu \otimes M^{\otimes(k-1)})} \quad \text{and} \quad g_m = \beta \eta(m) + 1 + \log \frac{d\mu_m}{d\mu} - \log \frac{(e\mu(W))^{-m}}{m!}. \]

We conceive \( \Phi \) as an element of the vector space
\[ A = \prod_{k \in [k_{\max}]} \mathcal{M}_{\pm}(W^k) \times \mathcal{M}_{\pm}(W)^{\mathbb{N}_0} \]
where \( \mathcal{M}_{\pm} \) is the set of signed measures, and \( ((h_k)_{k \in [k_{\max}]}, (g_m)_{m \in \mathbb{N}_0}) \) as a function on \( \prod_{k \in [k_{\max}]} W^k \times W^{\mathbb{N}_0} \). The condition in (4.4) means that \( \Phi \) is perpendicular to any function in
\[ \mathcal{F} = \left\{ (\varphi_k)_{k \in [k_{\max}]} : \psi_m : W \to \mathbb{R}, \psi_m : W \to \mathbb{R} \text{ bounded and measurable for any } k, m, \right\}, \]

where \( \exists \tilde{A}, \tilde{B}, \tilde{C} : W \to \mathbb{R} : \varphi_k(x_0, \ldots, x_{k-1}) = \tilde{A}(x_0) + \sum_{l=1}^{k-1} \tilde{C}(x_l), \quad \text{and } \psi_m(x) = \tilde{B}(x) - m \tilde{C}(x) \text{ for } x, x_0, \ldots, x_{k-1} \in W. \}

We have shown that, if \( \Phi \) is perpendicular to any simple function in \( \mathcal{F} \), then it is also perpendicular to \( ((h_k)_{k \in [k_{\max}]}, (g_m)_{m \in \mathbb{N}_0}) \). Since \( \mathcal{F} \) is a closed linear subspace of \( A \), it follows that \( \Phi \) contains this element. That is, there are three functions \( \tilde{A}, \tilde{B}, \tilde{C} \) on \( W \) such that, for any \( k \) respectively \( m \),
\[ h_k(x_0, \ldots, x_{k-1}) = \tilde{A}(x_0) + \sum_{l=1}^{k-1} \tilde{C}(x_l) \quad \text{and} \quad g_m(x) = \tilde{B}(x) - m \tilde{C}(x), \quad x, x_0, \ldots, x_{k-1} \in W. \]

Using an obvious substitution, this is equivalent to the existence of three positive functions \( A, B, C \) such that
\[ v_k(dx_0, \ldots, dx_{k-1}) = \mu(dx_0) A(x_0) \prod_{l=1}^{k-1} (C(x_l)M(dx_l)) e^{-\gamma \tilde{f}_k(x_0, \ldots, x_{k-1})}, \quad k \in [k_{\max}], \quad (4.8) \]
\[ \mu_m(dx) = \mu(dx) B(x) (C(x)\mu(W))^{-m} e^{-\beta \eta(m)} \quad m \in \mathbb{N}_0. \quad (4.9) \]

From (i) and (ii) in (1.19), we can identify \( A \) and \( B \) as
\[ \frac{1}{A(x_0)} = \sum_{k \in [k_{\max}]} \int_{W^{k-1}} \prod_{l=1}^{k-1} (C(x_l)M(dx_l)) e^{-\gamma \tilde{f}_k(x_0, \ldots, x_{k-1})}, \quad (4.10) \]
\[ \frac{1}{B(x)} = \sum_{m \in \mathbb{N}_0} (C(x)\mu(W))^{-m} e^{-\beta \eta(m)}. \quad (4.11) \]

Furthermore, condition (iii) says that
\[ \frac{1}{C(x)} = \frac{1}{C(x)M(dx)} \frac{1}{C(x)\mu(W)} = \Gamma(C \, \text{d}M, x), \quad x \in W, \quad (4.12) \]

where \( \varphi(\alpha) = \sum_{m \in \mathbb{N}_0} \frac{\alpha^m}{m!} e^{-\beta \eta(m)} / \sum_{m \in \mathbb{N}_0} \frac{\alpha^m}{m!} e^{-\beta \eta(m)} \) for \( \alpha \in [0, \infty) \) and
\[ \Gamma(d\tilde{M}, x) = \int_W \mu(dx_0) \sum_{k \in [k_{\max}]} \int_{W^{k-1}} \prod_{l=1}^{k-2} (\tilde{M}(dx_l) F_k(x_0, x_1, \ldots, x_{k-2}, x_l) \sum_{k \in [k_{\max}]} \int_{W^{k-1}} \prod_{l=1}^{k-1} (\tilde{M}(dx_l) \, e^{-\gamma \tilde{f}_k(x_0, \ldots, x_{k-1})}), \quad (4.13) \]

where
\[ F_k(x_0, x_1, \ldots, x_{k-2}, x) = \sum_{l=1}^{k-1} e^{-\gamma \tilde{f}_k(x_0, y_l)}, \quad (4.14) \]
$y^j$ is the vector of length $k - 1$, consisting of $x_1, \ldots, x_{k-2}$; augmented by $x$ at the $l$-th place, and $\tilde{M}(dx) = C(x)M(dx)$. This ends our derivation of the Euler–Lagrange equations for any minimizer $\Psi$ of (1.24).

This description of $C$ and $M$ is rather implicit and involved, therefore we cannot offer any simple criterion for the uniqueness of the minimizers of (1.24). Also, the question of continuity of the tilting functions $A$, $B$ and $C$ is open.

Since $I + \gamma S + \beta M$ is convex, it follows that any admissible trajectory setting $\Psi$ satisfying (1.8)–(1.11) is a minimizer of (1.24).

5. Proof of Proposition 1.5

We proceed analogously to Sections 2 and 3, and thus we start with part (2), i.e., with verifying (1.37). We use the discretization argument from Section 2.1 again. We now provide the definition of receiving given numbers of incoming messages.

Definition 5.1. A transmission setting is a collection of measures

$$\Sigma = \left( \Sigma = \{ (\nu_k)_{k=1}^{k_{\text{max}}}, (\nu_\delta^k)_{k=1}^{k_{\text{max}}}, (\tau_\delta^k)_{k=1}^{k_{\text{max}}}, \theta \in \mathbb{B}, (\tau_\delta^k)_{k=1}^{k_{\text{max}}}, (\tau_\delta^k)_{k=1}^{k_{\text{max}}}, (\mu^\delta)_{\delta \in \mathbb{B}, \lambda > 0}, (\mu^\delta)_{\delta \in \mathbb{B}, \lambda > 0} \right)$$

(5.1)

such that for any $\delta, \delta' \in \mathbb{B}$, $\lambda > 0$, $k \in [k_{\text{max}}]$ and $s, s_0, \ldots, s_{k-1} = 1, \ldots, \delta^{-d}$, respectively, parts (1), (2), (3), (4), (5), (6), (7) and (8) of Definition 2.1 hold.

The following lemma describes the combinatorics of the choices of message trajectories in the system. We recall the empirical measures $(R_{\lambda,k}(s))_{k \in [k_{\text{max}}]}$ from (1.5).

Lemma 5.2. Let $\Sigma$ be a transmission setting. For $\delta \in \mathbb{B}$ and $\lambda > 0$ let

$$K^\delta(\Sigma) = \{ s \in S_{k_{\text{max}}}(X^\lambda) : R_{\lambda,k}(s) = \delta^k \forall k = 1, \ldots, k_{\text{max}} \}.$$ Then we have $\#K^\delta(\Sigma) = N_1(\Sigma) \times N_2(\Sigma)$, where $N_1(\Sigma)$ equals $N_1(\Psi)$ from (2.8) for any standard setting $\Psi$ containing $\Sigma$, and

$$N_2(\Sigma) = \prod_{j=1}^{\delta-d} (\lambda^k \delta^k (W_j^\delta)) \lambda^{k_{\text{max}}} \sum_{k=1}^{k_{\text{max}}} \sum_{l=1}^{k-1} \pi^k \delta^k \mu^k (W_j^\delta).$$

Proof. We proceed in two steps by counting first the trajectories, registering only the partition sets $W_i^\delta$ that they travel through, second, the choices of the relays for each hop in each partition set. Since every choice in the two steps can be freely combined with the other one, the product of the two cardinalities is equal to the number of all trajectory configurations with the prescribed coarsened empirical measures.

(A) Number of the transmitters of trajectories passing through given sequences of $\delta$-subcubes. This is equal to the corresponding quantity in the proof of Lemma 2.14 hence it equals $N_1(\Sigma)$.

(B) Number of assignments of the hops to the relays. For each $i = 1, \ldots, \delta-d$, there are $\lambda \sum_{k=1}^{k_{\text{max}}} \sum_{l=1}^{k-1} \pi^k \nu_k (W_i^\delta)$ incoming messages arriving to the relays in $W_i^\delta$ in total. Each incoming message arriving at $W_i^\delta$ can choose any of the $\lambda^k \delta^k \mu^k (W_j^\delta)$ users as relay. Such choices between different hops in $W_i^\delta$ are independent, moreover all the choices in $W_i^\delta$ are independent from all the choices in $W_j^\delta$ for $j \neq i$. It follows that the number of assignments equals $N_2(\Sigma)$.

We also see that all the choices in the two parts are independent of each other, i.e., they can be freely combined with each other and yield different combinations. Hence, we arrived at the assertion. □

Using the arguments of the proof of Proposition 5.1 the next lemma immediately follows.
Lemma 5.3. Let \( \Sigma \) be a transmission setting. Then
\[
\lim_{\delta \downarrow 0} \lim_{\lambda \to \infty} \frac{1}{\lambda} \log \frac{\# K_{\delta,\lambda}(\Sigma)}{N_{\delta,\lambda}^{0}(\Sigma)} = -J(\Sigma) \in (-\infty, \infty],
\]
(5.2)
where \( N_{\delta,\lambda}^{0}(\Sigma) \) equals \( N_{\delta,\lambda}^{0}(\Psi) \) from (3.2) for any standard setting \( \Psi \) containing \( \Sigma \). Moreover, if the r.h.s. of (5.2) is finite, then
\[
\lim_{\delta \downarrow 0} \lim_{\lambda \to \infty} \frac{1}{\lambda} \log \# K_{\delta,\lambda}(\Sigma) = M(W) = \sum_{k=1}^{k_{\text{max}}} (k - 1) \nu_{k}(W).
\]

Now, the equality in (1.34) follows from the proof of Theorem 1.2, using transmission settings instead of standard settings and replacing Proposition 3.1 by our Lemma 5.3. There is one more major change in the proof. Indeed, instead of the compactness of \( \{ \Psi: \Psi \text{ adm. trajectory setting}, M(\Psi) \leq y \} \) for all \( y \geq 0 \) in Section 3.4 and the fact that any level set of \( I + \gamma S + \beta M \) is contained in a larger level set of \( M \), one shall use the following argument. Using that \( S \) is continuous on the set of asymptotic routing strategies, and that \( J \) is lower semicontinuous, bounded from below and it has compact level sets [DZ98, Section 6.2], it follows that each level set of \( J + \gamma S \) is included in a larger level set of \( S \).

From this, parts (1) and (4) of Proposition 1.5 can be derived analogously to how Theorem 1.4 was derived from Theorem 1.2 in Section 3.5. The additional fact that the rate function \( J + \mu(W) \log k_{\text{max}} \) has compact level sets holds because relative entropies w.r.t. fixed reference measures have compact level sets [DZ98, Section 6.2].

Lastly, we verify (3), i.e., we prove that (1.35) is the unique minimizer of (1.34). The fact that the set of minimizers of the variational formula on the right-hand side of (1.34) is non-empty, compact and convex follows similarly to Lemma 4.1 again by Prohorov’s theorem and the compactness of the sets \( \{ \Sigma: \Sigma \text{ asymptotic routing strategy, } S(\Sigma) \leq y \} \), \( y \geq 0 \). Further, an argument analogous to Lemma 4.2 shows that for all minimizers \( \Sigma = (\nu_{k})_{k \in [k_{\text{max}}]} \), we have that \( \nu_{k} \ll \mu^{\otimes k} \ll \nu_{k} \) for all \( k \in [k_{\text{max}}] \).

Deriving the Euler–Lagrange equations similarly to Section 4.2, it follows that (1.35)–(1.36) hold for any minimizer \( \Sigma = (\nu_{k})_{k \in [k_{\text{max}}]} \) of (1.34). This also implies that the minimizer \( \Sigma \) is unique. Thus, we conclude Proposition 1.5.

Acknowledgements. The authors thank B. Jahnel, C. Hirsch, M. Klimm and M. Renger for interesting discussions and comments. The second author acknowledges support by the Berlin Mathematical School (BMS) and the DFG RTG 1845 ’Stochastic Analysis with Applications in Biology, Finance and Physics’.

References

[AK08] S. Adams and W. König, Large deviations for many Brownian bridges with symmetrised initial-terminal condition. Probab. Theory Relat. Fields 142:1, 79–124 (2008).

[BB09] F. Baccelli and B. Blaszczyszyn, Stochastic Geometry and Wireless Networks: Volume I: Theory. Now Publishers Inc. (2009).

[CCS16] R. Colini-Baldeschi, Roberto Cominetti and Marco Scarsini, On the Price of Anarchy of Highly Congested Nonatomic Network Games, [arXiv:1605.05087] (2016).

[Ch09] X. Chen, Random Walk Intersections: Large Deviations and Related Topics. Mathematical Surveys and Monographs, AMS, Vol. 157, Providence, RI. (2009).

[DZ98] A. Dembo and O. Zeitouni, Large Deviations Techniques and Applications, 2nd edition, Springer, Berlin (1998).
A GIBBSIAN MODEL FOR MESSAGE ROUTING

[GT08] A. J. Ganesh and G. L. Torrisi, Large Deviations of the Interference in a Wireless Communication Model, *IEEE Trans. Inform. Theory* **54**:8, (2008).

[GZ93] H-O. Georgii and H. Zessin, Large deviations and the maximum entropy principle for marked point random fields, *Probab. Theory Relat. Fields* **96**:2, 177–204 (1993).

[HJP16] C. Hirsch, B. Jahnel and R. Patterson, Space-time large deviations in capacity-constrained relay networks, *Latin American Journal of Probability and Statistics* (2017), to appear. [arXiv:1609.06856]

[HJKP15] C. Hirsch, B. Jahnel, P. Keeler and R. Patterson, Large deviations in relay-augmented wireless networks, *Queueing Systems* (2017), to appear. [arXiv:1510.04146]

[K93] J. F. C. Kingman, *Poisson Processes*, Oxford University Press, New York (1993).

[KM02] W. König and P. Mörters, Brownian intersection local times: upper tail asymptotics and thick points. *Ann. Probab.* **30**, 1605–1656 (2002).

[KM13] W. König and C. Mukherjee, Large deviations for Brownian intersection measures, *Commun. Pure Appl. Math.* **66**:2, 263-306 (2013).

[KT17b] W. König and A. Tóbiás, Asymptotic routing properties in a Gibbsian model for highly dense multihop networks, *in preparation* (2017).

[T08] J. Trashorras, Large Deviations for Symmetrised Empirical Measures. *J. Theoret. Probab.* **21**, 397–412 (2008).