ESAIM: PS 16 (2012) 114–138
DOI: 10.1051/ps/2010016

ESAIM: Probability and Statistics
www.esaim-ps.org

ASYMPTOTIC EQUIPARTITION PROPERTIES FOR SIMPLE HIERARCHICAL AND NETWORKED STRUCTURES

Kwabena Doku-Amponsah[1]

Abstract. We prove *asymptotic equipartition properties* for simple hierarchical structures (modelled as *multitype Galton-Watson trees*) and networked structures (modelled as *randomly coloured random graphs*). For example, for large n, a networked data structure consisting of n units connected by an average number of links of order $n/\log n$ can be coded by about $H \times n$ bits, where H is an explicitly defined entropy. The main technique in our proofs are large deviation principles for suitably defined empirical measures.

Mathematics Subject Classification. 4A15, 94A24, 60F10, 05C80.

Received June 19, 2009. Revised March 31, 2010 and June 9, 2010.

1. Introduction

Information is often structured in a nonlinear way. For example, in genetics information often has an implicit hierarchical structure, in computer science data is often organized in the form of a network. To transmit or compress data from these sources, one needs efficient coding schemes and approximate pattern matching algorithms, and the *Shannon-McMillan-Breiman theorem* or *asymptotic equipartition property* (AEP) plays a key role in this regard, for example by providing bounds on the possible performance of algorithms.

Two major sets of research work on the AEP (and its applications) within mathematics and information theory have so far been considered. The first of these has focussed on stationary ergodic processes such as Markov chains, see Cover and Thomas [5] and the references therein. The second has dealt with stationary (ergodic) random fields on $\mathbb{Z}^d$, as well as amenable group actions, see, for example Dembo and Kontoyiannis [6] and the reference therein. Whilst typical examples of applications of the former has concentrated on data from linear source, the latter includes recent advances such as image and video processing, geostatistics, and statistical mechanics.

However, numerous types of data we usually come across in applications (communication studies, demographic studies, biological population studies and the field of physics) are naturally structured like networks or trees. For example, the WWW (consisting of a collection of pages residing on a server with a given name, together with 'hyperlinks' with their direction ignored), data on the spread of some disease in a given population and many more, can be described by networks. Equally, the age structure of a given population is best modelled by genealogical trees.

Keywords and phrases. Asymptotic equipartition property, large deviation principle, relative entropy, random graph, multitype Galton-Watson tree, randomly coloured random graph, typed graph, typed tree.

[1] Statistics Department, University of Ghana, Box LG 115, Legon, Ghana. kdoku@ug.edu.gh

Article published by EDP Sciences

© EDP Sciences, SMAI 2012

https://doi.org/10.1051/ps/2010016 Published online by Cambridge University Press

In this paper we use the large deviation techniques, as provided in the recent paper Dembo, Mörters and Sheffied [8], to study the AEP of structured data consisting of a large number of *units*, chosen from a finite set, together with a number of *links* connecting the units.

As an application of our abstract principles, we consider the following concrete examples from biology.

- **Metabolic network:** this is a graph of interactions forming a part of the energy generation and biosynthesis metabolism of the bacterium E. coli. Here, the units represent *substrates* and *products*, and links represent *interactions*. See Newman [13].
- **Mutation study:** consider mutations in mitochondrial DNA (mtDNA for short) e.g. the mtDNA^{4977} deletion (a mutation which causes a deletion of about one third of the mitochondrial genome). The replication of mtDNA can be described by a tree, where the units are a (normal) and b (mutant) and the links indicate 'mother-child' relations. See Olofsson and Shaw [14] and the references therein.

The core results of the paper are the *Shannon-McMillan-Breiman theorems* for two simple probabilistic models: the *multitype Galton-Watson trees* describing hierarchical data structures, and a class of *randomly coloured random graphs* describing networked data structures, see Theorems 2.2 and 2.1.

Specifically, we consider for the first model typed trees described by the following procedure: the root carries a random type chosen according to the some law on a finite alphabet; given the type of a vertex, the number and types of the children (ordered from left to right) are given independently of everything else, by an offspring law. For the second we look at random graph models constructed as follows: assign vertices colours independently and identically according to some colour law on a finite set of colours; connect any pair of vertices independently according to a probability depending on their colours.This model, with the simple Erdős-Rényi graph with independent colours as a special case, was introduced by Penman in his thesis [15], see Canning and Penman [4] for an exposition.

We also present large deviation principles (LDPs) for empirical colour measure and empirical pair measure of *sub*-and *supercritical* coloured random graphs. Major tool used in the proofs of these LDPs is (exponential) *change of measure*. We remark here that some of our results fit well into the framework of large deviations for mixtures which is utilized in the proofs of the LDPs in Doku and Mörters [10].

1.1. A model for simple hierarchical structures

We review in this subsection, the model for simple hierarchical data structures, *multitype Galton-Watson trees*. To begin, we collect some notation and concepts from the paper Dembo *et al.* [8]. By $\mathcal{T}$ we denote the set of all finite rooted planar trees T, by $V = V(T)$ the set of all vertices and by $E = E(T)$ the set of all edges oriented away from the root, which is always denoted by ρ. We write $|T|$ for the number of vertices in the tree T. Let $\mathcal{X}$ be a finite alphabet and write

$$\mathcal{X}^* = \bigcup_{n=0}^{\infty} \{n\} \times \mathcal{X}^n.$$

We equip $\mathcal{X}^*$ with the finest topology with all subsets as open sets. *i.e.* the discrete topology. We observe that the offspring of any vertex $v \in T$ is characterized by an element of $\mathcal{X}^*$ and that there is an element $(0, \emptyset)$ in $\mathcal{X}^*$ symbolizing absence of offspring.

Let μ be a probability measure (initial distribution) on $\mathcal{X}$ and $\mathbb{Q} : \mathcal{X} \times \mathcal{X}^* \to [0, 1]$ be an offspring transition kernel. The law $\mathbb{P}$ of a tree-indexed process X is defined by the following procedure:

- The root ρ carries a random type $X(\rho)$ chosen according to the probability measure μ on $\mathcal{X}$.
- For every vertex with type $a \in \mathcal{X}$ the offspring number and types are given independently of everything else, by the offspring law $\mathbb{Q}\{\cdot \,|\, a\}$ on $\mathcal{X}^*$. We write

$$\mathbb{Q}\big\{ \cdot \,|\, a\big\} = \mathbb{Q}\big\{(N, X_1, \ldots, X_N) \in \cdot \,|\, a\big\},$$

i.e. we have a random number N of descendants with types $X_1, \ldots, X_N$.

https://doi.org/10.1051/ps/2010016 Published online by Cambridge University Press

We shall consider $X = ((X(v), C(v)), v \in V)$ under the joint law of tree and offspring. We interpret X as *multitype Galton-Watson tree* and $X(v)$ as the type of vertex v. For each typed tree X and each vertex v we denote by $C(v) = (N(v), X_1(v), \ldots, X_{N(v)}(v)) \in \mathcal{X}^*$, the number and types of the children of v, ordered from left to right. We notice that the children of the root (denoted by ρ) are ordered but the root itself is not. We call an offspring distribution $\mathbb{Q}$ bounded if for some $N_0 < \infty$, we have

$$\mathbb{Q}\{N > N_0 \,|\, a\} = 0, \text{ for all } a \in \mathcal{X}.$$

Denote, for every $c = (n(c), a_1(c), \ldots, a_n(c)) \in \mathcal{X}^*$ and $a \in \mathcal{X}$, the *multiplicity* of the symbol a in c by

$$m(a, c) = \sum_{i=1}^{n(c)} 1_{\{a_i = a\}}.$$

Define the matrix A with index set $\mathcal{X} \times \mathcal{X}$ and nonnegative entries by

$$A(a, b) = \sum_{c \in \mathcal{X}^*} \mathbb{Q}\{c \,|\, b\} m(a, c), \text{ for } a, b \in \mathcal{X}.$$

$A(a, b)$ is the expected number of offspring of type a of a vertex of type b. Let $A^*(a, b) = \sum_{k=1}^{\infty} A^k(a, b) \in [0, \infty]$. We say that the matrix A is irreducible if $A^*(a, b) > 0$, for all $a, b \in \mathcal{X}$.

The multitype Galton-Watson tree is called irreducible if the matrix A is irreducible. It is called critical (subcritical, supercritical) if the largest eigenvalue of the matrix A is 1 (less than 1, greater 1 resp.). Let π be the *eigenvector* corresponding to the largest *eigenvalue* of the matrix A (normalized to a probability vector). Then π is *unique*, if the Galton-Watson tree is irreducible.

1.2. **A model for simple networked structures**

In this subsection, we review the model for simple networked structures, the *randomly coloured random graph model*. We begin by fixing the following notations. Let $\mathcal{X}$ be a finite alphabet or colour set $\mathcal{X}$. Let V be a fixed set of n vertices, say $V = \{1, \ldots, n\}$. Denote by $\mathcal{G}$ the set of all (simple) graphs and by $\mathcal{G}_n$ the set of all (simple) graphs with vertex set $V = \{1, \ldots, n\}$ and edge set

$$E \subset \mathcal{E} := \big\{(u, v) \in V \times V \,:\, u < v\big\},$$

where the *formal* ordering of edges is introduced as a means to simply describe *unordered* edges.

Given a symmetric function $p_n \colon \mathcal{X} \times \mathcal{X} \to [0, 1]$ and a probability measure μ on $\mathcal{X}$ we may define the *randomly coloured random graph* or simply *coloured random graph* X with n vertices as follows: assign to each vertex $v \in V$ colour $X(v)$ independently according to the *colour law* μ. Given the colours, we connect any two vertices $u, v \in V$, independently of everything else, with a *connection probability* $p_n(X(u), X(v))$ otherwise keep them disconnected. We always consider $X = \big((X(v) \,:\, v \in V), E\big)$ under the joint law of graph and colour. We interpret X as coloured random graph and consider $X(v)$ as the colour of the vertex v. Denote by $\mathcal{G}_n(\mathcal{X})$ the set of all coloured graphs with colour set $\mathcal{X}$ and n vertices.

We look at the coloured random graph models in three regimes, the *near-critical*, *subcritical* and *supercritical* cases. Thus, we consider the cases when the connection probabilities satisfy $a_n^{-1} p_n(a, b) \to C(a, b)$, for all $a, b \in \mathcal{X}$, where the sequence (a_n) is such that either $na_n \to 1$ or $na_n \to 0$ or $na_n \to \infty$ and $C \colon \mathcal{X} \times \mathcal{X} \to [0, \infty)$.

The rest of the paper is organized in the following way. In Section 2 all our results are stated. We state the Asymptotic Equipartition Properties for both models in Section 2.1, beginning with the case of simple hierarchical structures and then followed by the simple networked structures case. In Section 2.2, we compute the asymptotic number of bits needed to encode large amount of data from the model of the mtDNA[4977]

https://doi.org/10.1051/ps/2010016 Published online by Cambridge University Press

and the metabolic network. Section 3 contains proof of main results. We state and prove some large deviation principles for subcritical and supercritical coloured random graphs. We derived our main results from Theorems 3.1, 3.2, 3.4 and 3.6.

2. Statement of main results

2.1. **Asymptotic equipartition properties**

The underlying question is, how many bits are needed to store or transmit the information contained in a structured data consisting of n units connected by number of links?

Clearly, if no probabilistic structure is imposed, one needs of order n bits to transmit the units and of order n^2 bits to transmit the links of the network data structure. By imposing a probabilistic structure one can often transmit the structure at much cheaper cost with arbitrarily high probability. This is explained by the Shannon-McMillan-Breiman theorems for networked structures modelled as sparse coloured random graphs, and hierarchical structures modelled as multitype Galton-Watson trees.

Suppose q is the distribution of a message Y_n generated by a hierarchical or networked source and let H be entropy of the source. Then, we shall say $-\log_2 q(Y_n) \approx nH$ bits with *high probability* if as $n \to \infty$,

$$-\tfrac{1}{n}\log_2 q(Y_n) \to H \text{ in probability.}$$

We denote by $\mathbb{P}_n$ the law of a multitype Galton-Watson tree conditioned to have n vertices and write

$$P_n(x) := \mathbb{P}_n\{X = x\}, \text{for } x \in \mathcal{T}.$$

We state the asymptotic equipartition property (AEP) for simple hierarchical data structures.

Theorem 2.1. *Suppose $X = (X(v) : v \in V(T))$ is an irreducible, critical multitype Galton-Watson tree with finite type space $\mathcal{X}$ and bounded offspring kernel $\mathbb{Q}$. Then, for every $\varepsilon > 0$,*

$$\lim_{n\to\infty} \mathbb{P}_n\Big\{\Big| -\tfrac{1}{n}\log P_n(X) + \sum_{(a,c)\in\mathcal{X}\times\mathcal{X}^*} \pi(a)\mathbb{Q}\{c\,|\,a\}\log\mathbb{Q}\{c\,|\,a\}\Big| \geq \varepsilon\Big\} = 0.$$

We can extract from Theorem 2.1 the following useful information: to transmit the information contained in a large critical multitype Galton-Watson tree one needs with high probability, about

$$n\Big[-\frac{1}{\log 2}\sum_{(a,c)\in\mathcal{X}\times\mathcal{X}^*} \pi\otimes\mathbb{Q}(a,c)\log\mathbb{Q}\{c\,|\,a\}\Big] \qquad \text{bits,}$$

where n is the number of vertices in the tree. We consider the following example from the field of biology.

Mutations in mitochodrial DNA. Mitochondria are organelles in cells carrying their own DNA. Like nuclear DNA, mtDNA is subject to mutations which may take the form of base substitutions, duplication or deletions. The population mtDNA is modelled by two-type process where the units are a (normal) and b (mutant), and the links are mother-child relations. A normal can give birth to either two normals or, if there is mutation, one normal and one mutant. Suppose the latter happens with probability or mutation rate α. Mutants can only give birth to mutants. A DNA molecule may also die without reproducing.

Let the survival probabilities be $p \in \big[0, \frac{1}{(2-\alpha)}\big]$ and $q \in \big[0, \frac{1}{2}\big]$ for normals and mutants resp. We assume that the population is started from one normal ancestor. Suppose the offspring kernel $\mathbb{Q}$ is given by $\mathbb{Q}\{(0,\emptyset)|a\} = 1-p$, $\mathbb{Q}\{(2,(a,b))|a\} = p\alpha$, $\mathbb{Q}\{(2,(a,a))|a\} = p(1-\alpha)$, $\mathbb{Q}\{(0,\emptyset)|b\} = 1-q$ and $\mathbb{Q}\{(2,(b,b))|b\} = q$. Then, the process X is a multitype Galton-Watson process with matrix A (with index set $\{a,b\}$) given by

$$A = \begin{pmatrix} p(2-\alpha) & 0 \\ p\alpha & 2q \end{pmatrix}.$$

https://doi.org/10.1051/ps/2010016 Published online by Cambridge University Press

We restrict ourselves to the special case when $p = q = \frac{1}{2}$ and $\alpha > 0$. This case corresponds to the model for non-dividing tissue such as the brain. This means that the population of mtDNA is kept constant on average but that mitochondrial DNA keeps reproducing also in non-dividing cells. See, for example Arking [1] or Olofsson and Shaw [14] and the references therein.

We observe that, in this special case X is critical and irreducible, with $\pi(a) = \pi(b) = \frac{1}{2}$. Therefore, by Theorem 2.1 one needs with high probability approximately,

$$n\Big[1 - \tfrac{1}{\log 16}(\alpha \log \alpha + (1-\alpha)\log(1-\alpha))\Big] \text{ bits}, \tag{2.1}$$

in order to store or transmit data from a model of non-dividing tissues. For more examples of data source with tree structure, we refer to Kimmel and Axelrod [11] or Mode [12].

We state the AEP for networked data structure described by random coloured graphs. By $\mathbb{P}_n$ we also denote the (probability) law of a coloured random graph with n vertices. We write

$$P_n(x) = \mathbb{P}_n\{X = x\}, \text{ for } x \in \mathcal{G}_n.$$

Theorem 2.2. *Suppose that X is a coloured random graph with colour law $\mu\colon \mathcal{X} \to (0,1]$ and connection probabilities p_n such that $a_n^{-1}p_n(a,b) \to C(a,b)$ for some sequence (a_n) with $a_n n \log n \to \infty$ and $\log a_n / \log n \to -1$. Then, for every $\varepsilon > 0$,*

$$\lim_{n\to\infty} \mathbb{P}_n\left\{\Big| - \tfrac{1}{a_n n^2 \log n} \log P_n(X) - \tfrac{1}{2} \sum_{a,b\in\mathcal{X}} \mu(a)C(a,b)\mu(b)\Big| \geq \varepsilon\right\} = 0.$$

In other words, in order to transmit a coloured random graph in the given regime one needs with high probability, about

$$\frac{a_n n^2 \log n}{2 \log 2} \sum_{a,b\in\mathcal{X}} \mu(a)C(a,b)\mu(b) \text{ bits}.$$

The most interesting regime is when the cost of transmitting colours and transmitting edges is of comparable order, *i.e.* when $a_n^{-1} = n \log n$. In this case one obtains the following Shannon-McMillan-Breiman theorem.

Theorem 2.3. *Suppose that X is a coloured random graph with colour law $\mu\colon \mathcal{X} \to (0,1]$ and connection probabilities p_n such that $(n \log n)\, p_n(a,b) \to C(a,b)$ for $C\colon \mathcal{X} \times \mathcal{X} \to [0,\infty)$ symmetric. Then, for every $\varepsilon > 0$,*

$$\lim_{n\to\infty} \mathbb{P}_n\left\{\Big| - \tfrac{1}{n} \log P_n(X) - \tfrac{1}{2} \sum_{a,b\in\mathcal{X}} \mu(a)C(a,b)\mu(b) + \sum_{a\in\mathcal{X}} \mu(a) \log \mu(a)\Big| \geq \varepsilon\right\} = 0.$$

Interpretation. From Theorem 2.3 one can deduce that, the number of bits needed in order to code a networked data structure consisting of n units connected by an average number of order $n/\log n$ links with high probability is about nH, where H is the entropy defined by

$$H := \frac{1}{\log 2}\left[\tfrac{1}{2} \sum_{a,b\in\mathcal{X}} \mu(a)C(a,b)\mu(b) - \sum_{a\in\mathcal{X}} \mu(a) \log \mu(a)\right]. \tag{2.2}$$

Metabolic network. We consider a metabolic network of the energy and biosynthesis metabolism of the bacterium E. coli. Here, the units represent substrates and products, and links represent interactions. Suppose half the nodes in the graph are of unit a (substrate) and half are of unit b (product), and link between pair of units (a,b) occur independently with connection probability $\frac{C(a,b)}{n}$, where $C : \{a, b\} \times \{a, b\} \to [0,\infty)$ is nonzero symmetric function and n the size of the graph. We write

$$H := \tfrac{1}{8 \log 2}(2C(a,b) + C(a,a) + C(b,b)).$$

https://doi.org/10.1051/ps/2010016 Published online by Cambridge University Press

Then, by Theorem 2.2 one needs with high probability about $(n \log n)\, H$ bits to transmit the data contained in the metabolic network of the bacterium E. Coli.

3. Proof of main results

3.1. LDP for the empirical offspring measure

We present the recent large deviation principle for empirical offspring measures on random trees, see Dembo *et al.* [8]. We recall from the introductory section that $|T|$ is the number of vertices and $V = V(T)$ is the set of all vertices in tree T. We also recall that $m(a, c)$ is the multiplicity of the symbol a in $c = (n, a_1, \ldots, a_n)$ and that

$$\mathcal{X}^* = \bigcup_{n=0}^{\infty} \{n\} \times \mathcal{X}^n.$$

For every multitype Galton-Watson tree X, the *empirical offspring measure* M_X is defined by

$$M_X(a, c) = \frac{1}{|T|} \sum_{v \in V} \delta_{(X(v), C(v))}(a, c), \quad \text{for } (a, c) \in \mathcal{X} \times \mathcal{X}^*.$$

We call ν *shift-invariant* if $\nu_1(a) = \displaystyle\sum_{(b,c) \in \mathcal{X} \times \mathcal{X}^*} m(a, c)\nu(b, c), \quad \text{for all } a \in \mathcal{X} \times \mathcal{X}^*$.

We denote by $\mathcal{M}(\mathcal{X} \times \mathcal{X}^*)$ the space of probability measures ν on $\mathcal{X} \times \mathcal{X}^*$ with $\int n\, \nu(\mathrm{d}a\,, \mathrm{d}c) < \infty$, using the convention $c = (n, a_1, \ldots, a_n)$. We endow this space with the smallest topology which makes the functionals $\nu \mapsto \int f(b, c)\, \nu(db\,, dc)$ continuous, for $f : \mathcal{X} \times \mathcal{X}^* \to \mathbb{R}$ either bounded, or

$$f(b, c) = m(a, c)\mathbb{1}_{b_0}(b) \text{ for some } a, b_0 \in \mathcal{X}.$$

Theorem 3.1 (Dembo *et al.* [8])**.** *Suppose that X is an irreducible, critical multitype Galton-Watson tree with an offspring law whose exponential moments are all finite, conditioned to have exactly n vertices. Then, for $n \to \infty$, the empirical offspring measure M_X satisfies a large deviation principle in $\mathcal{M}(\mathcal{X} \times \mathcal{X}^*)$ with speed n and the convex, good rate function*

$$J(\nu) = \begin{cases} H(\nu \,\|\, \nu_1 \otimes \mathbb{Q}) & \text{if } \nu \text{ is shift-invariant,} \\ \infty & \text{otherwise.} \end{cases} \tag{3.1}$$

We remark that the critical and noncritical cases give the same tree under conditioning. See Dembo *et al.* [8].

3.2. Large deviation principle for sparse random coloured graphs

For any finite or countable set $\mathcal{Y}$ we denote by $\mathcal{M}(\mathcal{Y})$ the space of probability measures, and by $\tilde{\mathcal{M}}(\mathcal{Y})$ the space of finite measures on $\mathcal{Y}$, both endowed with the weak topology. We denote by $\tilde{\mathcal{M}}_*(\mathcal{Y} \times \mathcal{Y})$ the subspace of symmetric measures in $\tilde{\mathcal{M}}(\mathcal{Y} \times \mathcal{Y})$. We recall that V is fixed set of n vertices and $E \subset \mathcal{E} := \big\{(u, v) \in V \times V : u < v\big\}$ is the edge set.

We associate with any coloured random graph X with n vertices a probability measure, the *empirical colour measure* $L^1 \in \mathcal{M}(\mathcal{X})$, defined by

$$L^1(a) := \frac{1}{n} \sum_{v \in V} \delta_{X(v)}(a), \quad \text{for } a \in \mathcal{X},$$

https://doi.org/10.1051/ps/2010016 Published online by Cambridge University Press

and a symmetric finite measure, the *empirical pair measure* $L^2 \in \tilde{\mathcal{M}}_*(\mathcal{X} \times \mathcal{X})$, defined by

$$L^2(a,b) := \frac{1}{n^2 a_n} \sum_{(u,v)\in E} [\delta_{(X(v),\, X(u))} + \delta_{(X(u),\, X(v))}](a,b), \qquad \text{for } a,b \in \mathcal{X}.$$

The total mass $\|L^2\|$ of L^2 is $2|E|/(n^2 a_n)$. The next theorem is the LDP for the empirical colour measure and the empirical pair measure of a class of sparse coloured random graphs. *i.e.* $a_n = \frac{1}{n}$.

Theorem 3.2 (Doku *et al.* [10])**.** *Suppose that X is a coloured random graph with colour law μ and edge probabilities satisfying $np_n(a,b) \to C(a,b)$ for some symmetric function $C\colon \mathcal{X} \times \mathcal{X} \to [0,\infty)$. Then, as $n \to \infty$, the pair (L^1, L^2) satisfies a large deviation principle in $\mathcal{M}(\mathcal{X}) \times \tilde{\mathcal{M}}_*(\mathcal{X} \times \mathcal{X})$ with good rate function,*

$$I(\omega, \varpi) = H(\omega \,\|\, \mu) + \tfrac{1}{2}\mathfrak{H}_C(\varpi \,\|\, \omega)\,, \tag{3.2}$$

where $\mathfrak{H}_C(\varpi \,\|\, \omega) := H\big(\varpi \,\|\, C\omega \otimes \omega\big) + \|C\omega \otimes \omega\| - \|\varpi\|$ is a non-negative function and $C\omega \otimes \omega(a,b) := C(a,b)\omega(a)\omega(b)$.

Remark 3.3. By exponential equivalence, see Dembo and Zeitouni [7], Theorem 4.2.13, one can obtain from Theorem 3.2 the LDP for (L^1, L^2) of any coloured random graph X with connection probabilities satisfying $a_n^{-1}p_n(a,b) \to C(a,b)$, for some sequence (a_n) with $na_n \to 1$ and $C : \mathcal{X} \times \mathcal{X} \to [0, \infty)$ symmetric.

The proof of Theorem 3.2 uses the Gärtner-Ellis theorem, and the technique of mixing, see Biggins [2].

3.3. Large-deviation principles in the sub- and supercritical cases

We use large deviation techniques to study asymptotic properties of the coloured random graphs for large n in the *subcritical* and *supercritical* cases. In the rest of the paper, we assume that $(a_n) \to 0$ as n approaches infinity.

Theorem 3.4. *Suppose that X is a coloured random graph with colour law $\mu\colon \mathcal{X} \to (0,1]$ and edge probabilities $p_n : \mathcal{X} \times \mathcal{X} \to [0,1]$ satisfying $a_n^{-1}p_n(a,b) \to C(a,b)$, for some sequence (a_n) with $na_n \to \infty$ and $C : \mathcal{X} \times \mathcal{X} \to [0,\infty)$ symmetric. Then, for $n \to \infty$, the pair (L^1, L^2) satisfies a large deviation principle in $\mathcal{M}(\mathcal{X}) \times \tilde{\mathcal{M}}_*(\mathcal{X} \times \mathcal{X})$ with speed*

(i) *$a_n n^2$ and good rate function,*

$$I_1(\omega, \varpi) = \tfrac{1}{2}\mathfrak{H}_C(\varpi \,\|\, \omega). \tag{3.3}$$

(ii) *n and good rate function,*

$$I_2(\omega, \varpi) = \begin{cases} H(\,\omega \,\|\, \mu\,) & \text{if } \varpi = C\omega \otimes \omega, \\ \infty & \text{otherwise.} \end{cases} \tag{3.4}$$

Remark 3.5. Intuitively this means that, on the scale $a_n n^2$ the colour law can be changed 'for free', whereas on the scale n once the colour law is fixed, the edge law has to be the typical one.

Theorem 3.6. *Suppose that X is a coloured random graph with colour law $\mu\colon \mathcal{X} \to (0,1]$ and edge probabilities $p_n : \mathcal{X} \times \mathcal{X} \to [0,1]$ satisfying $a_n^{-1}p_n(a,b) \to C(a,b)$, for some sequence (a_n) with $na_n \to 0$ and $C : \mathcal{X} \times \mathcal{X} \to [0,\infty)$ symmetric. Then, for $n \to \infty$, the pair (L^1, L^2) satisfies a large deviation principle in $\mathcal{M}(\mathcal{X}) \times \tilde{\mathcal{M}}_*(\mathcal{X} \times \mathcal{X})$ with speed*

(i) *$a_n n^2$ and good rate function,*

$$I_3(\omega, \varpi) = \begin{cases} \frac{1}{2}\mathfrak{H}_C(\varpi \,\|\, \omega) & \text{if } \omega = \mu, \\ \infty & \text{otherwise.} \end{cases} \tag{3.5}$$

https://doi.org/10.1051/ps/2010016 Published online by Cambridge University Press

(ii) *n and good rate function,*

$$I_4(\omega, \varpi) = H(\,\omega \,\|\, \mu\,) \tag{3.6}$$

Remark 3.7. Intuitively this means that, on the scale n the edge law can be changed 'for free', whereas on the scale $a_n n^2$ the colour law cannot be changed.

Biggins and Penman [3] have proved large deviation principle for the $2|E|/n(n-1)$ using the technique of mixing, see Biggins [2].

In the rest of the section we give the proofs of the large deviation principles (LDPs) for coloured random graphs in the sub- and supercritical regimes, and use our LDPs and Theorems 3.1 and 3.2 to prove the asymptotic equipartition properties for simple hierarchical and networked structures. We prove our large deviation principles using the technique of (exponential) change of measures. Specifically, we use the technique of change of measure to prove the Upper bounds in Theorems 3.4 and 3.6. We then obtain the proofs of all Lower bounds *except* Theorems 3.4(i) and 3.6(ii) from the Upper bounds. Lower bounds of Theorems 3.4(i) and 3.6(ii) are proved from Lower bounds of Theorems 3.4(ii) and 3.6(i) respectively. All our proofs use two important lemmas, Euler's formula and exponential tightness lemma, which we shall state and prove in Section 3.4.

3.4. **Some useful lemmas**

Lemma 3.8 (Euler's formula)**.** *If $a_n^{-1} p_n(a,b) \to C(a,b)$, for all $a, b \in \mathcal{X}$ and $(a_n) \to 0$, then*

$$\lim_{n\to\infty} \left[1 + \alpha p_n(a,b)\right]^{a_n^{-1}} = \mathrm{e}^{\alpha C(a,b)}, \text{ for all } a, b \in \mathcal{X} \text{ and } \alpha \in \mathbb{R}. \tag{3.7}$$

Proof. Observe that, for any $\varepsilon > 0$ and for large n we have

$$\Big[1 + a_n(\alpha C(a,b) - \varepsilon)\Big]^{a_n^{-1}} \le \Big[1 + \alpha p_n(a,b)\Big]^{a_n^{-1}} \le \Big[1 + a_n(\alpha C(a,b) + \varepsilon)\Big]^{a_n^{-1}}, \tag{3.8}$$

by the pointwise convergence. Hence by the sandwich theorem and Euler's formula we have (3.7). □

Note, $\mathbb{P}$ is used instead of $\mathbb{P}_n$ in all our large deviation analysis for sake of a simple presentation.

Lemma 3.9 (exponential tightness)**.** *For every $\alpha > 0$, there exists $N \in \mathbb{N}$ such that*

$$\limsup_{n\to\infty} \tfrac{1}{a_n n^2} \log \mathbb{P}\Big\{|E| > a_n n^2 N\Big\} \le -\alpha. \tag{3.9}$$

Proof. Let $c > \max_{a,b\in\mathcal{X}} C(a,b) > 0$. Using Chebysheff's inequality and Lemma 3.8, we have (for sufficiently large n)

$$\begin{aligned}
\mathbb{P}\Big\{|E| \ge a_n n^2 l\Big\} &\le \mathrm{e}^{-a_n n^2 l} \mathbb{E}\Big\{\mathrm{e}^{|E|}\Big\} \\
&\le \mathrm{e}^{-a_n n^2 l} \sum_{k=0}^{n(n-1)/2} \mathrm{e}^k \binom{n(n-1)/2}{k} \Big(a_n c\Big)^k \Big(1 - a_n c\Big)^{n(n-1)/2-k} \\
&= \mathrm{e}^{-a_n n^2 l} \Big(1 + (e-1) a_n c\Big)^{a_n^{-1}(a_n n(n-1)/2)} \\
&\le \mathrm{e}^{-a_n n^2 l} \mathrm{e}^{a_n n^2 (c(e-1+o(1)))}.
\end{aligned}$$

https://doi.org/10.1051/ps/2010016 Published online by Cambridge University Press

Now given α choose $N \in \mathbb{N}$ such that $N > \alpha + c(e-1)$ and observe that, for sufficiently large n,

$$\mathbb{P}\Big\{|E| \geq a_n n^2 N\Big\} \leq \mathrm{e}^{-a_n n^2 \alpha},$$

which implies the statement. □

3.5. **Change of measure on the scale** n

We denote by $\mathcal{C}_2$ the space of symmetric functions on $\mathcal{X} \times \mathcal{X}$ and by $\mathcal{C}_1$ the space of functions on $\mathcal{X}$. Given a function $\tilde{f}\colon \mathcal{X} \to \mathbb{R}$ and a symmetric function $\tilde{g}\colon \mathcal{X} \times \mathcal{X} \to \mathbb{R}$, define the constant $U_{\tilde{f}}$ by

$$U_{\tilde{f}} = \log \sum_{a \in \mathcal{X}} \mathrm{e}^{\tilde{f}(a)} \mu(a).$$

For the function $\tilde{g} \in \mathcal{C}_2$ we define the symmetric function $\tilde{h}_n^{(2)} : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ by

$$\tilde{h}_n^{(2)}(a,b) = \log\Big[\Big(1 - p_n(a,b) + p_n(a,b)\mathrm{e}^{\tilde{g}(a,b)/na_n}\Big)^{-n}\Big]. \tag{3.10}$$

Use $\tilde{f},\ \tilde{g}$ to define (for sufficiently large n) a new coloured random graph in the following manner.

- To the vertices $V = \{1, \ldots, n\}$ we assign colours from $\mathcal{X}$ independently and identically according to the colour law $\widetilde{\mu}$ defined by
$$\widetilde{\mu}(a) = \mathrm{e}^{\tilde{f}(a) - U_{\tilde{f}}}\mu(a).$$
- Given any two vertices $u, v \in V$, with u carrying colour a and v carrying colour b connect vertex u to vertex v with probability
$$\tilde{p}_n(a,b) = \tfrac{p_n(a,b)\mathrm{e}^{\tilde{g}(a,b)/na_n}}{1-p_n(a,b)+p_n(a,b)\mathrm{e}^{\tilde{g}(a,b)/na_n}}, \tag{3.11}$$
otherwise keep u and v disconnected.

For this new graph, observe $\widetilde{\mu}$ is a probability measure and further that $\tilde{p}_n(a,\, b) \in [0,\, 1],$ for all $a, b \in \mathcal{X}$. Denote by $\tilde{\mathbb{P}}$ the law of the new coloured random graph construct from $\widetilde{\mu}$ and $\tilde{p}$. We note from the construction of the new graph that $\tilde{\mathbb{P}}$ is absolutely continuous with respect to $\mathbb{P}$, as for a coloured random graph X,

$$\begin{aligned}
\frac{\mathrm{d}\tilde{\mathbb{P}}}{\mathrm{d}\mathbb{P}}(X) &= \prod_{u\in V} \tfrac{\widetilde{\mu}(X(u))}{\mu(X(u))} \prod_{(u,v)\in E} \tfrac{\tilde{p}_n(X(u),X(v))}{p_n(X(u),X(v))} \prod_{(v,u)\notin E} \tfrac{1-\tilde{p}_n(X(u),X(v))}{1-p_n(X(u),X(v))} \\
&= \prod_{u\in V} \tfrac{\widetilde{\mu}(X(u))}{\mu(X(u))} \prod_{(u,v)\in E} \tfrac{\tilde{p}_n(X(u),X(v))}{p_n(X(u),X(v))} \times \tfrac{n-np_n(X(u),X(v))}{n-n\tilde{p}_n(X(u),X(v))} \prod_{(u,v)\in \mathcal{E}} \tfrac{n-n\tilde{p}_n(X(u),X(v))}{n-np_n(X(u),X(v))} \\
&= \prod_{u\in V} \mathrm{e}^{\tilde{f}(X(u)) - U_{\tilde{f}}} \prod_{(u,v)\in E} \mathrm{e}^{\tilde{g}(X(u),X(v))/na_n} \prod_{(u,v)\in \mathcal{E}} \mathrm{e}^{\tilde{h}_n^{(2)}(X(u),X(v))/n} \\
&= \mathrm{e}^{n\,\langle L^1,\, \tilde{f} - U_{\tilde{f}}\rangle + n\,\langle \frac{1}{2}L^2,\, \tilde{g}\rangle + n\,\langle \frac{1}{2}L^1 \otimes L^1,\, \tilde{h}_n^{(2)}\rangle - \langle\, \frac{1}{2}L^1_\Delta,\, \tilde{h}_n^{(2)}\rangle},
\end{aligned} \tag{3.12}$$

where $L^1_\Delta(a,\, a) = \frac{1}{n}\sum_{u\in V} \delta_{(X(u),X(u))}(a,\, a)$, for $a \in \mathcal{X}$ and $\sum_{a\in\mathcal{X}} L^1_\Delta(a,\, a) = 1$.

https://doi.org/10.1051/ps/2010016 Published online by Cambridge University Press

3.6. Change of measure on the scale $a_n n^2$

Define for $\tilde{g} \in \mathcal{C}_2$, $\tilde{h}_n^{(1)} : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ by

$$\tilde{h}_n^{(1)}(a,b) = -\log\left[\left(1 - p_n(a,b) + p_n(a,b)\mathrm{e}^{\tilde{g}(a,b)}\right)^{1/a_n}\right].$$

Define for $\tilde{f} \in \mathcal{C}_1$ and $\tilde{g}$ a new coloured random graph (for sufficiently large n) in the following way:

- Assign to the n vertices in V colours from $\mathcal{X}$ independently and identically according to the colour law $\widetilde{\mu}$ defined by
$$\widetilde{\mu}(a) = \mathrm{e}^{\tilde{f}(a) - U_{\tilde{f}}}\mu(a). \tag{3.13}$$
- Given any two vertices $u, v \in V$, with u carrying colour a and v carrying colour b connect vertex u to vertex v with probability
$$\tilde{p}_n(a,b) = \tfrac{p_n(a,b)\mathrm{e}^{\tilde{g}(a,b)}}{1-p_n(a,b)+p_n(a,b)\mathrm{e}^{\tilde{g}(a,b)}}, \tag{3.14}$$
otherwise keep u and v disconnected.

Note the colour law $\tilde{\mu}$ is a probability measure and the connection probabilities satisfy $\tilde{p}_n(a, b) \in [0, 1]$, for all $a, b \in \mathcal{X}$. We denote by $\tilde{\mathbb{P}}$ the law of the coloured random graph obtained from $\widetilde{\mu}$ and $\tilde{p}_n$. By construction $\tilde{\mathbb{P}}$ is absolutely continuous with respect to $\mathbb{P}$, as for a coloured random graph X,

$$\begin{aligned}
\frac{\mathrm{d}\tilde{\mathbb{P}}}{\mathrm{d}\mathbb{P}}(X) &= \prod_{u\in V} \tfrac{\tilde{\mu}(X(u))}{\mu(X(u))} \prod_{(u,v)\in E} \tfrac{\tilde{p}_n(X(u),X(v))}{p_n(X(u),X(v))} \prod_{(v,u)\notin E} \tfrac{1-\tilde{p}_n(X(u),X(v))}{1-p_n(X(u),X(v))} \\
&= \prod_{u\in V} \mathrm{e}^{\tilde{f}(X(u))-U_{\tilde{f}}} \prod_{(u,v)\in E} \tfrac{\tilde{p}_n(X(u),X(v))}{p_n(X(u),X(v))} \times \tfrac{1-\tilde{p}_n(X(u),X(v))}{1-p_n(X(u),X(v))} \prod_{(u,v)\in\mathcal{E}} \tfrac{1-p_n(X(u),X(v))}{1-\tilde{p}_n(X(u),X(v))} \\
&= \prod_{u\in V} \mathrm{e}^{\tilde{f}(X(u))-U_{\tilde{f}}} \prod_{(u,v)\in E} \mathrm{e}^{\tilde{g}(X(u),X(v))} \prod_{(u,v)\in\mathcal{E}} \mathrm{e}^{a_n \tilde{h}_n^{(1)}(X(u),X(v))} \\
&= \mathrm{e}^{n\langle L^1, \tilde{f}-U_{\tilde{f}}\rangle + a_n n^2 \langle \frac{1}{2}L^2, \tilde{g}\rangle + a_n n^2 \langle \frac{1}{2}L^1\otimes L^1, \tilde{h}_n^{(1)}\rangle - a_n n^2 \langle \frac{1}{2}L^2_\Delta, \tilde{h}_n^{(1)}\rangle},
\end{aligned} \tag{3.15}$$

where $L^2_\Delta(a, a) = \frac{1}{n^2}\sum_{u\in V}\delta_{(X(u),X(u))}(a, a)$, for $a \in \mathcal{X}$ and $\sum_{a\in\mathcal{X}} L^2_\Delta(a, a) = \frac{1}{n}$.

3.7. Upper bound in Theorem 3.4(ii)

Define for $(\omega, \varpi) \in \mathcal{M}(\mathcal{X}) \times \tilde{\mathcal{M}}_*(\mathcal{X} \times \mathcal{X})$, $\hat{I}_2(\omega, \varpi)$ by

$$\hat{I}_2(\omega, \varpi) = \sup_{\substack{\tilde{f}\in\mathcal{C}_1 \\ \tilde{g}\in\mathcal{C}_2}} \left\{ \sum_{a\in\mathcal{X}} \big(\tilde{f}(a) - U_{\tilde{f}}\big)\omega(a) + \sum_{a,b\in\mathcal{X}} \tfrac{1}{2}\tilde{g}(a,b)(\varpi(a,b) - C(a,b)\omega(a)\omega(b)) \right\}. \tag{3.16}$$

Lemma 3.10. *For each closed set $F \subset \tilde{\mathcal{M}}_*(\mathcal{X} \times \mathcal{X})$, we have*

$$\limsup_{n\to\infty} \tfrac{1}{n}\log\mathbb{P}\big\{(L^1, L^2) \in F\big\} \le -\inf_{(\omega,\varpi)\in F} \hat{I}_2(\omega, \varpi).$$

Proof. Fix $\tilde{f} \in \mathcal{C}_1$. For any $\tilde{g} \in \mathcal{C}_2$, we define $\tilde{\beta}\colon \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ by $\tilde{\beta}(a,b) = -\tilde{g}(a,b)C(a,b)$.

We notice from Lemma 3.8 that, $\lim_{n\to\infty} \tilde{h}_n^{(2)}(a,b) = \tilde{\beta}(a,b)$, for all $a, b \in \mathcal{X}$. Hence, for any $\delta > 0$ and for (sufficiently) large n, we have

$$\tilde{h}_n^{(2)}(a,b) \le |\tilde{\beta}(a,b)| + \delta, \quad \text{for all } a, b \in \mathcal{X}. \tag{3.17}$$

https://doi.org/10.1051/ps/2010016 Published online by Cambridge University Press

Using (3.12) and (3.17) we obtain

$$\mathrm{e}^{(\max_{a\in\mathcal{X}}|\tilde{\beta}(a,a)|+\delta)/2} \geq \int \mathrm{e}^{\langle \frac{1}{2}L^1_\Delta, \tilde{h}_n^{(2)}\rangle}\mathrm{d}\tilde{\mathbb{P}} = \mathbb{E}\Big\{\mathrm{e}^{n\langle L^1, \tilde{f}-U_{\tilde{f}}\rangle + n\langle \frac{1}{2}L^2, \tilde{g}\rangle + n\langle \frac{1}{2}L^1\otimes L^1, \tilde{h}_n^{(2)}\rangle}\Big\},$$

for any $\delta > 0$ and for large n. Therefore, we have

$$\limsup_{n\to\infty} \tfrac{1}{n}\log \mathbb{E}\Big\{\mathrm{e}^{n\langle L^1, \tilde{f}-U_{\tilde{f}}\rangle + n\langle \frac{1}{2}L^2, \tilde{g}\rangle + n\langle \frac{1}{2}L^1\otimes L^1, \tilde{h}_n^{(2)}\rangle}\Big\} \leq 0. \tag{3.18}$$

We now fix $\varepsilon > 0$ and write $\hat{I}_2^\varepsilon(\omega, \varpi) := \min\{\hat{I}_2(\omega, \varpi), \varepsilon^{-1}\} - \varepsilon$. Let F be a closed subset of $\tilde{\mathcal{M}}_*(\mathcal{X}\times\mathcal{X})$ and suppose $(\omega, \varpi) \in F$. Choose $\tilde{f} \in \mathcal{C}_1$, $\tilde{g} \in \mathcal{C}_2$ such that

$$\langle \omega, \tilde{f} - U_{\tilde{f}}\rangle + \tfrac{1}{2}\langle \varpi, \tilde{g}\rangle - \tfrac{1}{2}\langle \omega\otimes\omega, C\tilde{g}\rangle \geq \hat{I}_2^\varepsilon(\omega, \varpi).$$

Since $\mathcal{X}$ is finite, we can find open neighbourhoods B^2_ϖ and B^1_ω of ϖ, ω such that

$$\inf_{\tilde{\omega}\in B^1_\omega,\, \tilde{\varpi}\in B^2_\varpi} \big\{\langle \tilde{\omega}, \tilde{f} - U_{\tilde{f}}\rangle + \tfrac{1}{2}\langle \tilde{\varpi}, \tilde{g}\rangle - \tfrac{1}{2}\langle \tilde{\omega}\otimes\tilde{\omega}, C\tilde{g}\rangle\big\} \geq \hat{I}_2^\varepsilon(\omega, \varpi) - \varepsilon.$$

Using Chebysheff's inequality and (3.18), we have that

$$\begin{aligned}
&\limsup_{n\to\infty} \tfrac{1}{n}\log \mathbb{P}\big\{(L^1, L^2) \in B^1_\omega \times B^2_\varpi\big\} \\
&\qquad \leq \limsup_{n\to\infty} \tfrac{1}{n}\log \mathbb{E}\Big\{\mathrm{e}^{n\langle L^1, \tilde{f}-U_{\tilde{f}}\rangle + n\langle \frac{1}{2}L^2, \tilde{g}\rangle + n\langle \frac{1}{2}L^1\otimes L^1, \tilde{h}_n^{(2)}\rangle}\Big\} - \hat{I}_2^\varepsilon(\omega, \varpi) + \varepsilon \\
&\qquad \leq -\hat{I}_2^\varepsilon(\omega, \varpi) + \varepsilon.
\end{aligned} \tag{3.19}$$

Use Lemma 3.9 to choose $N \in \mathbb{N}$ such that

$$\limsup_{n\to\infty} \tfrac{1}{n}\log \mathbb{P}\Big\{|E| > a_n n^2 N\Big\} = -\infty.$$

For this N define the set K_N by

$$K_N = \big\{(\omega, \varpi) \in \mathcal{M}(\mathcal{X})\times\tilde{\mathcal{M}}_*(\mathcal{X}\times\mathcal{X}) : \|\varpi\| \leq 2N\big\}.$$

The set $K_N \cap F$ is compact and therefore may be covered by finitely many sets $B^1_{\omega_r} \times B^2_{\varpi_r}$ with $(\omega_r, \varpi_r) \in F$ for $r = 1, \ldots, m$. Hence, we have

$$\mathbb{P}\big\{(L^1, L^2) \in F\big\} \leq \sum_{r=1}^m \mathbb{P}\big\{(L^1, L^2) \in B^1_{\omega_r} \times B^2_{\varpi_r}\big\} + \mathbb{P}\big\{(L^1, L^2) \notin K_N\big\}.$$

We may now use (3.19) to obtain, for all sufficiently small $\varepsilon > 0$,

$$\begin{aligned}
\limsup_{n\to\infty} \tfrac{1}{n}\log \mathbb{P}\big\{(L^1, L^2) \in F\big\} &\leq \max_{r=1}^m \limsup_{n\to\infty} \tfrac{1}{n}\log \mathbb{P}\big\{(L^1, L^2) \in B^1_{\omega_r} \times B^2_{\varpi_r}\big\} - \infty \\
&\leq -\inf_{(\omega,\,\varpi)\in F} \hat{I}_2^\varepsilon(\omega, \varpi) + \varepsilon.
\end{aligned}$$

Taking $\varepsilon \downarrow 0$ we have the desired statement. □

We solve the variational problem on the right side of equation (3.16).

https://doi.org/10.1051/ps/2010016 Published online by Cambridge University Press

Lemma 3.11. $\hat{I}_2(\omega, \varpi) = H(\omega \,\|\, \mu)$ *if (and only if)* $\varpi = C\omega \otimes \omega$, *and* ∞ *otherwise.*

Proof. Suppose that $\varpi \neq C\omega \otimes \omega$. Then there exists $a_0, b_0 \in \mathcal{X}$ such $\varpi(a_0, b_0) > C(a_0, b_0)\omega(a_0)\omega(b_0)$ or $\varpi(a_0, b_0) < C(a_0, b_0)\omega(a_0)\omega(b_0)$. Define for this $a_0, b_0 \in \mathcal{X}$ the symmetric function $\tilde{g}$ by

$$\tilde{g}(a, b) = K(\mathbb{1}_{(a_0, b_0)}(a, b) + \mathbb{1}_{(b_0, a_0)})(a, b), \text{ for } a, b \in \mathcal{X} \text{ and } K \in \mathbb{R} \tag{3.20}$$

Considering this $\tilde{g}$ in (3.16) we have

$$\sum_{a,b\in\mathcal{X}} \tfrac{1}{2}\tilde{g}(a, b)\varpi(a, b) + \sum_{a,b\in\mathcal{X}} -\tfrac{1}{2}\tilde{g}(a, b)C(a, b)\omega(a)\omega(b) = K(\varpi(a_0, b_0) - C(a_0, b_0)\omega(a_0)\omega(b)) \xrightarrow{|K|\uparrow\infty} \infty, \tag{3.21}$$

where the sign of $|K|$ is chosen such that expression in the right side of (3.25) remain positive.

Suppose that $\varpi = C\omega \otimes \omega$. Then, by the variational characterization of the relative entropy we have $I_2(\omega, \varpi) = H(\omega \,\|\, \mu)$, which ends the proof of the upper bounds. □

3.8. **Lower bound Theorem 3.4(ii)**

Lemma 3.12. *For every open set* $O \subset \mathcal{M}(\mathcal{X}) \times \tilde{\mathcal{M}}_*(\mathcal{X} \times \mathcal{X})$,

$$\liminf_{n\to\infty} \tfrac{1}{n} \log \mathbb{P}\big\{(L^1, L^2) \in O\big\} \geq - \inf_{(\omega, \varpi)\in O} I_2(\omega, \varpi). \tag{3.22}$$

Proof. Suppose $(\omega, \varpi) \in O$ is such that we have $\varpi = C\omega \otimes \omega$. Set $\tilde{g}(a, b) = 0$, for all $a, b \in \mathcal{X}$ and define $\tilde{f}_\omega \colon \mathcal{X} \to \mathbb{R}$ by

$$\tilde{f}_\omega(a) = \begin{cases} \log \frac{\omega(a)}{\mu(a)}, & \text{if } \omega(a) > 0, \\ 0, & \text{otherwise.} \end{cases}$$

We note that this choice of $\tilde{g}$ yields $\tilde{h}^{(2)}_n(a, b) = 0$, for all $a, b \in \mathcal{X}$. Choose B^1_ω, B^2_ϖ open neighbourhoods of ω, ϖ, such that $B^1_\omega \times B^2_\varpi \subset O$ and for all $(\tilde{\omega}, \tilde{\varpi}) \in B^1_\omega \times B^2_\varpi$,

$$\langle \tilde{f}_\omega, \omega\rangle - \varepsilon \leq \langle \tilde{f}_\omega, \tilde{\omega}\rangle.$$

We use the probability measure $\tilde{\mathbb{P}}$ given by $\tilde{g}_\varpi$. We observe that the colour law is ω and the connection probabilities satisfy $a_n^{-1}\tilde{p}_n(a, b) \to \tilde{C}(a, b) := \varpi(a, b)/(\omega(a)\omega(b))$, as n approaches infinity. Therefore, using (3.12) we have that

$$\begin{aligned} \mathbb{P}\Big\{(L^1, L^2) \in O\Big\} \geq \tilde{\mathbb{E}}\Big\{\tfrac{\mathrm{d}\mathbb{P}}{\mathrm{d}\tilde{\mathbb{P}}}(X)\mathbb{1}_{\{(L^1,L^2)\in B^1_\omega \times B^2_\varpi\}}\Big\} &= \tilde{\mathbb{E}}\Big\{\mathrm{e}^{-n\langle L^1, \tilde{f}_\omega\rangle}\mathbb{1}_{\{(L^1,\,L^2)\in B^1_\omega \times B^2_\varpi\}}\Big\} \\ &\geq \mathrm{e}^{-n\langle \omega,\, \tilde{f}_\omega\rangle - n\varepsilon} \times \tilde{\mathbb{P}}\Big\{(L^1, L^2) \in B^1_\omega \times B^2_\varpi\Big\}. \end{aligned}$$

Therefore, we have

$$\liminf_{n\to\infty} \tfrac{1}{n} \log \mathbb{P}\big\{(L^1, L^2) \in O\big\} \geq -\langle \omega,\, \tilde{f}_\omega\rangle - \varepsilon + \liminf_{n\to\infty} \tfrac{1}{n} \log \tilde{\mathbb{P}}\big\{(L^1, L^2) \in B^1_\omega \times B^2_\varpi\big\}.$$

The result follows once we prove that

$$\liminf_{n\to\infty} \tfrac{1}{n} \log \tilde{\mathbb{P}}\big\{(L^1, L^2) \in B^1_\omega \times B^2_\varpi\big\} = 0. \tag{3.23}$$

We use the upper bound (but now with the law $\mathbb{P}$ replaced by $\tilde{\mathbb{P}}$) to prove (3.23).

https://doi.org/10.1051/ps/2010016 Published online by Cambridge University Press

Therefore, we have

$$\limsup_{n\to\infty} \tfrac{1}{n} \log \tilde{\mathbb{P}}\big\{(L^1, L^2) \in (B^1_\omega \times B^2_\varpi)^c\big\} \le - \inf_{(\tilde\omega,\tilde\varpi)\in\tilde F} \tilde I_2(\tilde\omega, \tilde\varpi),$$

$$\tilde I_2(\tilde\omega, \tilde\varpi) = \begin{cases} H(\,\tilde\omega \,\|\, \omega\,) & \text{if } \tilde\varpi = \tilde C\tilde\omega \otimes \tilde\omega, \\ \infty & \text{otherwise,} \end{cases}$$

where $\tilde F = (B^1_\omega \times B^2_\varpi)^c$. It therefore suffices to show that the infimum is positive. Suppose for contradiction that there exists a sequence $(\tilde\omega_n, \tilde\varpi_n) \in \tilde F$ with $\tilde I_2(\tilde\omega_n, \tilde\varpi_n) \downarrow 0$. Then, since $\tilde I_2$ is a good rate function and its level sets are compact, and the mapping $(\tilde\omega, \tilde\varpi) \mapsto \tilde I_2(\tilde\omega, \tilde\varpi)$ is lower semicontinuous, we can construct a limit point $(\tilde\omega, \tilde\varpi) \in \tilde F$ with $\tilde I_2(\tilde\omega, \tilde\varpi) = 0$. By Lemma 3.11 this implies $H(\tilde\omega \,\|\, \omega) = 0$ and $\tilde\varpi = C\tilde\omega \otimes \tilde\omega$, hence $\tilde\omega = \omega$, and $\tilde\varpi = \tilde C\tilde\omega \otimes \tilde\omega = \varpi$. This contradicts $(\tilde\omega, \tilde\varpi) \in \tilde F$. □

3.9. **Upper bound in Theorem 3.4(i)**

We obtain the upper bound in a variational formulation. We define for $(\omega, \varpi) \in \mathcal{M}(\mathcal{X}) \times \tilde{\mathcal{M}}_*(\mathcal{X} \times \mathcal{X})$ the rate function $\hat I_1$ by

$$\hat I_1(\omega, \varpi) = \sup_{\tilde g \in \mathcal{C}_2} \Big\{ \sum_{a,b\in\mathcal{X}} \tfrac{1}{2}\tilde g(a,b)\varpi(a,b) + \sum_{a,b\in\mathcal{X}} \tfrac{1}{2}(1 - \mathrm{e}^{\tilde g(a,b)})C(a,b)\omega(a)\omega(b) \Big\}. \tag{3.24}$$

Lemma 3.13. *For each closed set $F \subset \mathcal{M}(\mathcal{X}) \times \tilde{\mathcal{M}}_*(\mathcal{X} \times \mathcal{X})$, we have*

$$\limsup_{n\to\infty} \tfrac{1}{a_n n^2} \log \mathbb{P}\big\{(L^1,\, L^2) \in F\big\} \le - \inf_{(\omega,\,\varpi)\in F} \hat I_1(\omega,\, \varpi).$$

Proof. For any $\tilde g \in \mathcal{C}_2$ we define $\tilde\beta\colon \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ by

$$\tilde\beta(a,b) = (1 - \mathrm{e}^{\tilde g(a,b)})C(a,b).$$

From Lemma 3.8 we note that $\lim_{n\to\infty} \tilde h^{(1)}_n(a,b) = \tilde\beta(a,b)$, for all $a, b \in \mathcal{X}$. We observe that, for any $\delta > 0$ and for (sufficiently) large n we have

$$\tilde h^{(2)}_n(a,b) \le |\tilde\beta(a,b)| + \delta, \quad \text{for all } a, b \in \mathcal{X}. \tag{3.25}$$

We take $\tilde f(a) = 0$, for all $a \in \mathcal{X}$, and use (3.15) and (3.25) to obtain

$$\mathrm{e}^{na_n(\max_{a\in\mathcal{X}} |\tilde\beta(a,a)|+\delta)/2} \ge \int \mathrm{e}^{a_n n^2 \left\langle \frac{1}{2} L^2_\Delta,\, \tilde h^{(1)}_n \right\rangle} \mathrm{d}\tilde{\mathbb{P}} = \mathbb{E}\left\{ \mathrm{e}^{a_n n^2 \left\langle \frac{1}{2} L^2,\, \tilde g \right\rangle + a_n n^2 \left\langle \frac{1}{2} L^1 \otimes L^1,\, \tilde h^{(1)}_n \right\rangle} \right\},$$

for any $\delta > 0$ and for large n. Therefore, we have that

$$\limsup_{n\to\infty} \tfrac{1}{a_n n^2} \log \mathbb{E}\left\{ \mathrm{e}^{a_n n^2 \left\langle \frac{1}{2} L^2,\, \tilde g \right\rangle + a_n n^2 \left\langle \frac{1}{2} L^1 \otimes L^1,\, \tilde h^{(1)}_n \right\rangle} \right\} \le 0. \tag{3.26}$$

Fix $\varepsilon > 0$ and take $\hat I^\varepsilon_1(\omega,\, \varpi) = \min\{\hat I_1(\omega,\, \varpi), \varepsilon^{-1}\} - \varepsilon$. Suppose $(\omega,\, \varpi) \in F$ and choose $\tilde g \in \mathcal{C}_2$ such that

$$\tfrac{1}{2}\langle \varpi,\, \tilde g\rangle + \tfrac{1}{2}\langle \omega \otimes \omega,\, \tilde\beta\rangle \ge \hat I^\varepsilon_1(\omega, \varpi).$$

https://doi.org/10.1051/ps/2010016 Published online by Cambridge University Press

Using the finiteness of $\mathcal{X}$ we can find open neighbourhoods B^1_ω, B^2_ϖ of ω, ϖ such that

$$\inf_{\tilde\omega\in B^1_\omega,\,\tilde\varpi\in B^2_\varpi}\left\{\tfrac12\langle\tilde\varpi,\,\tilde g\rangle+\tfrac12\langle\tilde\omega\otimes\tilde\omega,\,\tilde\beta\rangle\right\}\geq\hat I^\varepsilon_1(\omega,\,\varpi)-\varepsilon.$$

By Chebysheff's inequality and (3.26), we have that

$$\begin{aligned}\limsup_{n\to\infty}\tfrac{1}{a_nn^2}\log\mathbb{P}\big\{(L^1,\,L^2)\in B^1_\omega\times B^2_\varpi\big\}&\leq\limsup_{n\to\infty}\tfrac{1}{a_nn^2}\log\mathbb{E}\Big\{\mathrm{e}^{a_nn^2\,\langle\frac12L^2,\,\tilde g\rangle+a_nn^2\,\langle\frac12L^1\otimes L^1,\,\tilde h^{(1)}_n\rangle}\Big\}-\hat I^\varepsilon_1(\omega,\,\varpi)+\varepsilon\\&\leq-\hat I^\varepsilon_1(\omega,\,\varpi)+\varepsilon.\end{aligned}\tag{3.27}$$

We use Lemma (3.9) with $\alpha=\varepsilon^{-1}$ to choose $N(\varepsilon)\in\mathbb{N}$ such that

$$\limsup_{n\to\infty}\tfrac{1}{a_nn^2}\log\mathbb{P}\Big\{|E|>a_nn^2N(\varepsilon)\Big\}\leq-\varepsilon^{-1}.$$

Define for this N, the set $K_{N(\varepsilon)}$ by

$$K_{N(\varepsilon)}=\big\{(\omega,\,\varpi)\in\mathcal{M}(\mathcal{X})\times\tilde{\mathcal{M}}_*(\mathcal{X}\times\mathcal{X}):\|\varpi\|\leq2N(\varepsilon)\big\}.$$

Now, observe that $K_{N(\varepsilon)}\cap F$ is compact and therefore may be covered by finitely many sets $B^1_\omega\times B^2_{\varpi_r}$, $r=1,\dots,m$ with $(\omega_r,\,\varpi_r)\in F$ for $r=1,\dots,m$. Hence, we have

$$\mathbb{P}\big\{(L^1,\,L^2)\in F\big\}\leq\sum_{r=1}^m\mathbb{P}\big\{(L^1,\,L^2)\in B^1_{\omega_r}\times B^2_{\varpi_r}\big\}+\mathbb{P}\big\{(L^1,\,L^2)\notin K_{N(\varepsilon)}\big\}.$$

Using (3.27) for small enough $\varepsilon>0$, we obtain

$$\begin{aligned}\limsup_{n\to\infty}\tfrac{1}{a_nn^2}\log\mathbb{P}\big\{(L^1,\,L^2)\in F\big\}&\leq\max_{r=1}^m\limsup_{n\to\infty}\tfrac{1}{a_nn^2}\log\mathbb{P}\big\{(L^1,\,L^2)\in B^1_{\omega_r}\times B^2_{\varpi_r}\big\}-\varepsilon^{-1}\\&\leq-\hat I^\varepsilon_1(\omega,\,\varpi)+\varepsilon.\end{aligned}$$

Taking $\varepsilon\downarrow0$ we have the desired statement. □

We identify the rate function by solving the variational problem in the right side of equation (3.24).

Lemma 3.14. *For any $(\omega,\,\varpi)\in\mathcal{M}(\mathcal{X})\times\tilde{\mathcal{M}}_*(\mathcal{X}\times\mathcal{X})$ we have $\hat I_1(\omega,\,\varpi)=\mathfrak{H}_C(\varpi\,\|\,\omega)$.*

Proof. (i) Suppose $\varpi\not\ll C\omega\otimes\omega$. Then there exists $a_0,\,b_0\in\mathcal{X}$ with $C(a_0,b_0)\omega(a_0)\omega(b_0)=0$ and $\varpi(a_0,b_0)>0$. For this (a_0,b_0) we define the symmetric function $\tilde g$ by

$$\tilde g(a,b)=\log(K(\mathbb{1}_{(a_0,b_0)}(a,b)+\mathbb{1}_{(b_0,a_0)}(a,b))+1),$$

for $a,b\in\mathcal{X}$ and $K>0$. Considering our $\tilde g$ in (3.24) we have

$$\sum_{a,b\in\mathcal{X}}\tfrac12\tilde g(a,b)\varpi(a,b)+\sum_{a,b\in\mathcal{X}}\tfrac12(1-\mathrm{e}^{\tilde g(a,b)})C(a,b)\omega(a)\omega(b)=\log(K+1)(\varpi(a_0,b_0))\xrightarrow{K\uparrow\infty}\infty.$$

Suppose that $\varpi\ll C\omega\otimes\omega$. Then, we have

$$\hat I(\omega,\varpi)\geq\tfrac12\sup_{g\in\mathcal{C}_2}\Big\{\sum_{a,b\in\mathcal{X}}g(a,b)\varpi(a,b)-\sum_{a,b\in\mathcal{X}}\mathrm{e}^{g(a,b)}C(a,b)\omega(a)\omega(b)\Big\}+\tfrac12\sum_{a,b\in\mathcal{X}}C(a,b)\omega(a)\omega(b).\tag{3.28}$$

https://doi.org/10.1051/ps/2010016 Published online by Cambridge University Press

Using the substitution $h = \mathrm{e}^g \frac{C\omega\otimes\omega}{\varpi}$ and $\sup_{x>0} \log x - x = -1$ we obtain the expression

$$\begin{aligned}
&\sup_{g\in\mathcal{C}_2} \Big\{ \sum_{a,b\in\mathcal{X}} g(a,b)\varpi(a,b) - \sum_{a,b\in\mathcal{X}} \mathrm{e}^{g(a,b)} C(a,b)\omega(a)\omega(b) \Big\} \\
&\sup_{\substack{h\in\mathcal{C}_2\\ h\geq 0}} \sum_{a,b\in\mathcal{X}} \Big[\log\Big(h(a,b) \frac{\varpi(a,b)}{C(a,b)\omega(a)\omega(b)} \Big) - h(a,b) \Big] \varpi(a,b) \\
&= \sup_{\substack{h\in\mathcal{C}_2\\ h\geq 0}} \sum_{a,b\in\mathcal{X}} \big(\log h(a,b) - h(a,b) \big) \varpi(a,b) + \sum_{a,b\in\mathcal{X}} \log\Big(\frac{\varpi(a,b)}{C(a,b)\omega(a)\omega(b)} \Big) \varpi(a,b) \\
&= -\|\varpi\| + H(\varpi \,\|\, C\omega\otimes\omega).
\end{aligned} \tag{3.29}$$

This gives $\hat{I}_1(\omega, \varpi) = \mathfrak{H}_C(\varpi \,\|\, \omega)$, which concludes the proof of the lemma. □

Remark 3.15. It is not hard to see that $\mathfrak{H}_C(\cdot \,\|\, \cdot)$ is a good rate function, as for for all $\alpha < \infty$, its level sets are the bounded, closed set $\{(\omega, \varpi) \in \mathcal{M}(\mathcal{X}) \times \tilde{\mathcal{M}}(\mathcal{X}\times\mathcal{X})\colon\ \mathfrak{H}_C(\varpi \,\|\, \omega) \leq \alpha\}$ and therefore, are compact.

3.10. Lower bound in Theorem 3.4(i)

We use the LDP on the scale n (but with the law $\mathbb{P}$ replaced by $\tilde{\mathbb{P}}$) to establish the lower bound for some open set $O \subset \mathcal{M}(\mathcal{X}) \times \tilde{\mathcal{M}}_*(\mathcal{X}\times\mathcal{X})$.

Lemma 3.16. *For every open set $O \subset \mathcal{M}(\mathcal{X}) \times \tilde{\mathcal{M}}_*(\mathcal{X}\times\mathcal{X})$.*

$$\liminf_{n\to\infty} \tfrac{1}{a_n n^2} \log \tilde{\mathbb{P}}\big\{(L^1, L^2) \in O\big\} \geq - \inf_{(\tilde{\omega}, \tilde{\varpi})\in O} I_1(\omega, \varpi). \tag{3.30}$$

Proof. Suppose $(\omega, \varpi) \in O$ with $\varpi \ll C\omega\otimes\omega$. We define the function $\tilde{f}_\omega\colon \mathcal{X} \to \mathbb{R}$ by

$$\tilde{f}_\omega(a) = \begin{cases} \log \frac{\omega(a)}{\mu(a)}, & \text{if } \omega(a) > 0, \\ 0, & \text{otherwise,} \end{cases}$$

and the symmetric function $\tilde{g}_\varpi\colon \mathcal{X}\times\mathcal{X} \to \mathbb{R}$ by

$$\tilde{g}_\varpi(a,b) = \begin{cases} \log \frac{\varpi(a,b)}{C(a,b)\omega(a)\omega(b)}, & \text{if } \varpi(a,b) > 0, \\ 0, & \text{otherwise.} \end{cases}$$

We recall that

$$\tilde{h}_n^{(1)}(a,b) = -\log\Big[1 - p_n(a,b) + p_n(a,b)\mathrm{e}^{\tilde{g}_\varpi(a,b)}\Big]^{1/a_n}, \text{ for } a,b \in \mathcal{X}.$$

Define the symmetric function $\tilde{\beta}_\varpi(a,b)$ by

$$\tilde{\beta}_\varpi(a,b) := \lim_{n\to\infty} \tilde{h}_n^{(1)}(a,b) = C(a,b)(1 - \mathrm{e}^{g_\varpi(a,b)}).$$

Choose B^1_ω, B^2_ϖ open neighbourhoods of ω, ϖ such that $B^1_\omega \times B^2_\varpi \subset O$ and for all $(\tilde{\omega}, \tilde{\varpi}) \in B^1_\omega \times B^2_\varpi$,

$$\langle \varpi, \tilde{g}_\varpi\rangle + \langle \omega\otimes\omega, \tilde{\beta}_\varpi\rangle - \varepsilon \leq \langle \tilde{\varpi}, \tilde{g}_\varpi\rangle + \langle \tilde{\omega}\otimes\tilde{\omega}, \tilde{\beta}_\varpi\rangle.$$

We note that, the coloured random graph obtained from the function $\tilde{g}_\varpi$ has colour law ω and connection probabilities $\tilde{p}_n(a,b) \in [0,1]$ satisfying

$$a_n^{-1}\tilde{p}_n(a,b) \to \tilde{C}(a,b) := \varpi(a,b)/(\omega(a)\omega(b)), \text{ as } n\to\infty.$$

https://doi.org/10.1051/ps/2010016 Published online by Cambridge University Press

Write $m := 0 \wedge \min_{a\in\mathcal{X}} \tilde{\beta}_{\varpi}(a,a)$, and $l := 0 \wedge \max_{a\in\mathcal{X}} \tilde{f}(a)$. Now, using (3.15) we have that

$$\begin{aligned}\mathbb{P}\Big\{(L^1,L^2)\in O\Big\} &\geq \tilde{\mathbb{E}}\Big\{\tfrac{\mathrm{d}\mathbb{P}}{\mathrm{d}\tilde{\mathbb{P}}}(X)\mathbb{1}_{\{(L^1,L^2)\in B^1_\omega\times B^2_\varpi\}}\Big\}\\ &= \tilde{\mathbb{E}}\Big\{\mathrm{e}^{-n\langle L^1,\tilde{f}_\omega\rangle - a_n n^2\left\langle \frac{1}{2}L^2,\tilde{g}_\varpi\right\rangle - a_n n^2\left\langle \frac{1}{2}L^1\otimes L^1,\tilde{h}^{(1)}_n\right\rangle + n^2\left\langle \frac{1}{2}L^2_\Delta,\tilde{h}^{(1)}_n\right\rangle}\times\mathbb{1}_{\{(L^1,L^2)\in B^1_\omega\times B^2_\varpi\}}\Big\}\\ &\geq \mathrm{e}^{-nl-a_nn^2\langle\varpi,\tilde{g}_\varpi\rangle/2 - a_nn^2\langle\ \omega\otimes\omega,\tilde{\beta}\rangle/2 + a_nm/4 - a_nn^2\varepsilon/2}\times\tilde{\mathbb{P}}\Big\{(L^1,L^2)\in B^1_\omega\times B^2_\varpi\Big\}.\end{aligned}$$

Therefore, we have

$$\begin{aligned}\liminf_{n\to\infty}\tfrac{1}{a_nn^2}\log\mathbb{P}\big\{(L^1,L^2)\in O\big\} \geq &-\tfrac{1}{2}\langle\tilde{g},\varpi\rangle - \tfrac{1}{2}\langle\tilde{\beta},\omega\otimes\omega\rangle - \varepsilon\\ &+\liminf_{n\to\infty}\tfrac{1}{a_nn^2}\log\tilde{\mathbb{P}}\big\{(L^1,L^2)\in B^1_\omega\times B^2_\varpi\big\}.\end{aligned}$$

The result follows once we prove that

$$\liminf_{n\to\infty}\tfrac{1}{a_nn^2}\log\tilde{\mathbb{P}}\big\{(L^1,L^2)\in B^1_\omega\times B^2_\varpi\big\} = 0. \tag{3.31}$$

To conclude the proof, we use the lower bound of the LDP on the scale n (but with the law $\mathbb{P}$ replaced by $\tilde{\mathbb{P}}$), to prove (3.31). We notice from Theorem 3.4(ii) that, for any $\delta>0$ and for large n we have

$$\tilde{\mathbb{P}}\big\{(L^1,L^2)\in B^1_\omega\times B^2_\varpi\big\}\geq \mathrm{e}^{-n\alpha_2(\omega,\varpi)-n\delta}$$

where $\alpha_2(\omega,\varpi) = \inf\big\{\tilde{I}_2(\tilde{\omega},\tilde{\varpi}) : (\tilde{\omega},\tilde{\varpi})\in B^1_\omega\times B^2_\varpi\big\}$ and

$$\tilde{I}_2(\tilde{\omega},\tilde{\varpi}) = \begin{cases} H(\tilde{\omega}\,\|\,\omega) & \text{if } \tilde{\varpi} = \tilde{C}\tilde{\omega}\otimes\tilde{\omega},\\ \infty & \text{otherwise.}\end{cases}$$

Therefore, we have

$$\liminf_{n\to\infty}\tfrac{1}{a_nn^2}\log\tilde{\mathbb{P}}\big\{(L^1,L^2)\in B^1_\omega\times B^2_\varpi\big\}\geq\liminf_{n\to\infty}\tfrac{1}{a_nn}\{-\alpha_2(\omega,\varpi)-\delta\} = 0,$$

since $a_nn\to\infty$ as $n\to\infty$. This concludes the proof of the Lemma. □

3.11. **Upper bound in theorem 3.6(i)**

To begin we obtain the upper bound in a variational formulation. We recall that $na_n\to 0$ for subcritical coloured graphs and write

$$Z_n(f) := \tfrac{1}{na_n}U_{na_nf}.$$

Notice $Z(f) := \lim_{n\to\infty} Z_n(f) < \infty$. Define for $(\omega,\varpi)\in\mathcal{M}(\mathcal{X})\times\tilde{\mathcal{M}}_*(\mathcal{X}\times\mathcal{X})$ the rate $\hat{I}_3$ by

$$\hat{I}_3(\omega,\varpi) = \sup_{\substack{f\in\mathcal{C}_1\\ \tilde{g}\in\mathcal{C}_2}}\left\{\sum_{a\in\mathcal{X}}(f(a)-Z(f))\omega(a) + \sum_{a,b\in\mathcal{X}}\tfrac{1}{2}\tilde{g}(a,b)\varpi(a,b) + \sum_{a,b\in\mathcal{X}}\tfrac{1}{2}(1-\mathrm{e}^{\tilde{g}(a,b)})C(a,b)\omega(a)\omega(b)\right\}. \tag{3.32}$$

Lemma 3.17. *For each closed set $F\subset\mathcal{M}(\mathcal{X})\times\tilde{\mathcal{M}}_*(\mathcal{X}\times\mathcal{X})$,*

$$\limsup_{n\to\infty}\tfrac{1}{a_nn^2}\log\mathbb{P}\big\{(L^1,L^2)\in F\big\}\leq -\inf_{(\omega,\varpi)\in F}\hat{I}_3(\omega,\varpi).$$

https://doi.org/10.1051/ps/2010016 Published online by Cambridge University Press

Proof. Fix $f \in \mathcal{C}_1$. For any $\tilde{g} \in \mathcal{C}_2$ we define $\tilde{\beta}\colon \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ by $\tilde{\beta}(a,b) = (1 - \mathrm{e}^{\tilde{g}(a,b)})C(a,b)$. Lemma 3.8 gives $\lim_{n\to\infty} \tilde{h}_n^{(1)}(a,b) = \tilde{\beta}(a,b)$, for all $a, b \in \mathcal{X}$. We note that, for any $\delta > 0$ and for large n, we have

$$\tilde{h}_n^{(1)}(a,b) \leq |\tilde{\beta}(a,b)| + \delta, \quad \text{for all } a, b \in \mathcal{X} \tag{3.33}$$

Taking $\tilde{f}(a) = na_n f(a)$, for all $a \in \mathcal{X}$, and using (3.15) and (3.33) we have

$$\mathrm{e}^{na_n(\max_{a\in\mathcal{X}}|\tilde{\beta}(a,a)|+\delta)/2} \geq \int \mathrm{e}^{a_n n^2\langle \frac{1}{2}L^2_\Delta, \tilde{h}_n^{(2)}\rangle}\mathrm{d}\tilde{\mathbb{P}} = \mathbb{E}\left\{\mathrm{e}^{a_n n^2\langle L^1, f - Z_n(f)\rangle + a_n n^2\left\langle \frac{1}{2}L^2, \tilde{g}\right\rangle + a_n n^2\left\langle \frac{1}{2}L^1\otimes L^1, \tilde{h}_n^{(1)}\right\rangle}\right\},$$

for any $\delta > 0$ and for large n. Therefore, we have

$$\limsup_{n\to\infty} \tfrac{1}{a_n n^2} \log \mathbb{E}\left\{\mathrm{e}^{a_n n^2\langle L^1, f - Z_n(f)\rangle + a_n n^2\left\langle \frac{1}{2}L^2, \tilde{g}\right\rangle + a_n n^2\left\langle \frac{1}{2}L^1\otimes L^1, \tilde{h}_n^{(1)}\right\rangle}\right\} \leq 0. \tag{3.34}$$

Fix $\varepsilon > 0$ and take $\hat{I}_3^\varepsilon(\omega, \varpi) = \min\{\hat{I}_3(\omega, \varpi), \varepsilon^{-1}\} - \varepsilon$. Suppose $(\omega, \varpi) \in F$ and choose $f \in \mathcal{C}_1$, $\tilde{g} \in \mathcal{C}_2$ such that

$$\langle \omega, f - Z(f)\rangle + \tfrac{1}{2}\langle \varpi, \tilde{g}\rangle + \tfrac{1}{2}\langle \omega \otimes \omega, \tilde{\beta}\rangle \geq \hat{I}_3^\varepsilon(\omega, \varpi).$$

By finiteness of $\mathcal{X}$, we can find open neighbourhoods B^1_ω, B^2_ϖ of ω, ϖ such that

$$\inf_{\tilde{\omega}\in B^1_\omega, \tilde{\varpi}\in B^2_\varpi} \left\{\langle \tilde{\omega}, f - Z(f)\rangle + \langle \tfrac{1}{2}\tilde{g}, \tilde{\varpi}\rangle + \langle \tfrac{1}{2}\tilde{\omega}\otimes\tilde{\omega}, \tilde{\beta}\rangle\right\} \geq \hat{I}_3^\varepsilon(\omega, \varpi) - \varepsilon.$$

By Chebysheff's inequality and (3.34), we have that

$$\begin{aligned}
&\limsup_{n\to\infty} \tfrac{1}{a_n n^2} \log \mathbb{P}\left\{(L^1, L^2) \in B^1_\omega \times B^2_\varpi\right\} \leq \\
&\qquad \limsup_{n\to\infty} \tfrac{1}{a_n n^2} \log \mathbb{E}\left\{\mathrm{e}^{a_n n^2\langle L^1, \tilde{f} - Z_n(f)\rangle + a_n n^2\langle \frac{1}{2}L^2, \tilde{g}\rangle + a_n n^2\langle \frac{1}{2}L^1\otimes L^1, \tilde{h}_n^{(1)}\rangle}\right\} - \hat{I}_3^\varepsilon(\omega, \varpi) + \varepsilon \\
&\qquad \leq -\hat{I}_3^\varepsilon(\omega, \varpi) + \varepsilon.
\end{aligned} \tag{3.35}$$

By Lemma (3.9) we choose $N(\varepsilon) \in \mathbb{N}$ (with $\alpha = \varepsilon^{-1}$) such that

$$\limsup_{n\to\infty} \tfrac{1}{a_n n^2} \log \mathbb{P}\left\{|E| > a_n n^2 N(\varepsilon)\right\} \leq -\varepsilon^{-1}.$$

Define for this N, the set $K_{N(\varepsilon)}$ by

$$K_{N(\varepsilon)} = \left\{(\omega, \varpi) \in \mathcal{M}(\mathcal{X}) \times \tilde{\mathcal{M}}_*(\mathcal{X}\times\mathcal{X}) : \|\varpi\| \leq 2N(\varepsilon)\right\}.$$

Note $K_{N(\varepsilon)} \cap F$ is compact and therefore may be covered by finitely many sets $B^1_\omega \times B^2_{\varpi_r}$, $r = 1, \ldots, m$ with $(\omega_r, \varpi_r) \in F$ for $r = 1, \ldots, m$. Hence, we have

$$\mathbb{P}\left\{(L^1, L^2) \in F\right\} \leq \sum_{r=1}^m \mathbb{P}\left\{(L^1, L^2) \in B^1_{\omega_r} \times B^2_{\varpi_r}\right\} + \mathbb{P}\left\{(L^1, L^2) \notin K_{N(\varepsilon)}\right\}.$$

https://doi.org/10.1051/ps/2010016 Published online by Cambridge University Press

Using (3.35) for small enough $\varepsilon > 0$, we obtain

$$\limsup_{n\to\infty} \tfrac{1}{a_n n^2} \log \mathbb{P}\big\{(L^1, L^2) \in F\big\} \le \max_{r=1}^{m} \limsup_{n\to\infty} \tfrac{1}{a_n n^2} \log \mathbb{P}\big\{(L^1, L^2) \in B^1_{\omega_r} \times B^2_{\varpi_r}\big\} - \varepsilon^{-1}$$
$$\le -\hat{I}_3^{\varepsilon}(\omega, \varpi) + \varepsilon.$$

Taking $\varepsilon \downarrow 0$ we have the required statement. □

We identify the rate function by solving the variational problem in the right side of equation (3.32).

Lemma 3.18. *For any $(\omega, \varpi) \in \mathcal{M}(\mathcal{X}) \times \tilde{\mathcal{M}}_*(\mathcal{X} \times \mathcal{X})$, we have $\hat{I}_3(\omega, \varpi) = I_3(\omega, \varpi)$.*

Proof. (i) Suppose $\omega \in \mathcal{M}(\mathcal{X})$ is not equal μ. Define the function f by

$$f(a) = K \log(|\omega(a) - \mu(a)| + 1), \quad \text{for } a \in \mathcal{X} \text{ and } K \in \mathbb{R}.$$

Set $\tilde{g}(a, b) = 0$ for all $a, b \in \mathcal{X}$ in (3.32) and note that by the choice of f we have

$$\sum_{a\in\mathcal{X}} (f(a) - Z(f))\omega(a) + \sum_{a,b\in\mathcal{X}} \tfrac{1}{2}\tilde{g}(a, b)\varpi(a, b) \sum_{a,b\in\mathcal{X}} \tfrac{1}{2}(1 - \mathrm{e}^{\tilde{g}(a,b)})C(a, b)\omega(a)\omega(b)$$
$$\ge K \left(\sum_{a\in\mathcal{X}} \log(|\omega(a) - \mu(a)| + 1)\omega(a) - \max_a |\omega(a) - \mu(a)| - 1 \right) \xrightarrow{|K|\uparrow\infty} \infty,$$

where the sign of $|K|$ is such that last expression always stays positive. Suppose $\varpi \not\ll C\omega \otimes \omega$. Then there exists a_0, $b_0 \in \mathcal{X}$ with $C(a_0, b_0)\omega(a_0)\omega(b_0) = 0$ and $\varpi(a_0, b_0) > 0$. For this (a_0, b_0) we define the function $\tilde{g}$ by

$$\tilde{g}(a, b) = \log(K(\mathbb{1}_{(a_0,b_0)}(a, b) + \mathbb{1}_{(b_0,a_0)}(a, b)) + 1),$$

for $a, b \in \mathcal{X}$ and $K > 0$. Considering our $\tilde{g}$ in (3.24) we have

$$\sum_{a,b\in\mathcal{X}} \tfrac{1}{2}\tilde{g}(a, b)\varpi(a, b) + \sum_{a,b\in\mathcal{X}} \tfrac{1}{2}(1 - \mathrm{e}^{\tilde{g}(a,b)}C(a, b)\omega(a)\omega(b) = \log(K + 1)(\varpi(a_0, b_0)) \xrightarrow{K\uparrow\infty} \infty.$$

Suppose that $\varpi \ll C\omega \otimes \omega$. Then, we have

$$\hat{I}(\omega, \varpi) \ge \tfrac{1}{2} \sup_{g\in\mathcal{C}_2} \left\{ \sum_{a,b\in\mathcal{X}} g(a, b)\varpi(a, b) - \sum_{a,b\in\mathcal{X}} \mathrm{e}^{g(a,b)}C(a, b)\omega(a)\omega(b) \right\} + \tfrac{1}{2} \sum_{a,b\in\mathcal{X}} C(a, b)\omega(a)\omega(b).$$

By the substitution $h = \mathrm{e}^g \frac{C\omega\otimes\omega}{\varpi}$ and $\sup_{x>0} \log x - x = -1$ we have (3.29), which yields

$$\hat{I}_3(\omega, \varpi) = I_3(\omega, \varpi).$$ □

3.12. **Lower bound in Theorem 3.6(i)**

Lemma 3.19. *For every open set $O \subset \mathcal{M}(\mathcal{X}) \times \tilde{\mathcal{M}}_*(\mathcal{X} \times \mathcal{X})$.*

$$\liminf_{n\to\infty} \tfrac{1}{a_n n^2} \log \mathbb{P}\Big\{(L^1, L^2) \in O\Big\} \ge - \inf_{(\tilde{\omega}, \tilde{\varpi})\in O} I_3(\omega, \varpi). \tag{3.36}$$

https://doi.org/10.1051/ps/2010016 Published online by Cambridge University Press

Proof. Suppose $(\omega, \varpi) \in O$ with $\varpi \ll C\omega \otimes \omega$ and $\omega = \mu$. Take $\tilde{f}(a) = 0$, for all $a \in \mathcal{X}$. Define the symmetric function $\tilde{g}_\varpi \colon \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ by

$$\tilde{g}_\varpi(a,b) = \begin{cases} \log \frac{\varpi(a,b)}{C(a,b)\omega(a)\omega(b)}, & \text{if } \varpi(a,b) > 0, \\ 0, & \text{otherwise.} \end{cases}$$

Recall that

$$\tilde{h}_n^{(1)}(a,b) = -\log\left[1 - p_n(a,b) + p_n(a,b)\mathrm{e}^{\tilde{g}_\varpi(a,b)}\right]^{1/a_n}, \text{ for } a,b \in \mathcal{X}.$$

Define the function $\tilde{\beta}_\varpi(a,b)$ by

$$\tilde{\beta}_\varpi(a,b) := \lim_{n\to\infty} \tilde{h}_n^{(1)}(a,b) = C(a,b)(1 - \mathrm{e}^{g_\varpi(a,b)}).$$

Choose B^1_ω, B^2_ϖ open neighbourhoods of ω, ϖ such that $B^1_\omega \times B^2_\varpi \subset O$ and for all $(\tilde{\omega}, \tilde{\varpi}) \in B^1_\omega \times B^2_\varpi$,

$$\langle \varpi, \tilde{g}_\varpi \rangle + \langle \omega \otimes \omega, \tilde{\beta}_\varpi \rangle - \varepsilon \le \langle \tilde{\varpi}, \tilde{g}_\varpi \rangle + \langle \tilde{\omega} \otimes \tilde{\omega}, \tilde{\beta}_\varpi \rangle.$$

We note that, the random coloured graph obtained from the function $\tilde{g}_\varpi$ has colour law ω and connection probabilities satisfying

$$a_n^{-1}\tilde{p}_n(a,b) \to \tilde{C}(a,b) := \varpi(a,b)/(\omega(a)\omega(b)), \text{ as } n \to \infty.$$

Using (3.12) we have

$$\begin{aligned} \mathbb{P}\Big\{(L^1, L^2) \in O\Big\} &\ge \tilde{\mathbb{E}}\Big\{ \tfrac{\mathrm{d}\mathbb{P}}{\mathrm{d}\tilde{\mathbb{P}}}(X) \mathbb{1}_{\{(L^1,L^2)\in B^1_\omega \times B^2_\varpi\}} \Big\} \\ &= \tilde{\mathbb{E}}\left\{ \mathrm{e}^{-a_n n^2 \left\langle \frac{1}{2} L^2, \tilde{g}_\varpi \right\rangle - a_n n^2 \left\langle \frac{1}{2} L^1 \otimes L^1, \tilde{h}_n^{(1)} \right\rangle + a_n n^2 \left\langle L^2_\Delta, \tilde{h}_n^{(1)} \right\rangle} \times \mathbb{1}_{\{(L^1,L^2)\in B^1_\omega \times B^2_\varpi\}} \right\} \\ &\ge \mathrm{e}^{-a_n n^2 \left\langle \frac{1}{2}\varpi, \tilde{g}_\varpi \right\rangle - a_n n^2 \left\langle \frac{1}{2}\omega\otimes\omega, \tilde{\beta} \right\rangle + a_n m/4 - \frac{1}{2} a_n n^2 \varepsilon} \times \tilde{\mathbb{P}}\Big\{(L^1, L^2) \in B^1_\omega \times B^2_\varpi\Big\}, \end{aligned}$$

where $m := 0 \wedge \min_{a\in\mathcal{X}} \tilde{\beta}_\varpi(a,a)$. Therefore, we have

$$\liminf_{n\to\infty} \tfrac{1}{a_n n^2} \log \mathbb{P}\big\{(L^1, L^2) \in O\big\} \ge -\tfrac{1}{2}\langle \varpi, \tilde{g}\rangle - \tfrac{1}{2}\langle \omega\otimes\omega, \tilde{\beta}\rangle - \varepsilon + \liminf_{n\to\infty} \tfrac{1}{a_n n^2} \log \tilde{\mathbb{P}}\big\{(L^1, L^2) \in B^1_\omega \times B^2_\varpi\big\}.$$

The result follows once we prove that

$$\liminf_{n\to\infty} \tfrac{1}{a_n n^2} \log \tilde{\mathbb{P}}\big\{(L^1, L^2) \in B^1_\omega \times B^2_\varpi\big\} = 0. \tag{3.37}$$

We use the upper bound (but now with the law $\mathbb{P}$ replaced by $\tilde{\mathbb{P}}$) to prove (3.37).

Therefore, we have

$$\limsup_{n\to\infty} \tfrac{1}{a_n n^2} \log \tilde{\mathbb{P}}\big\{(L^1, L^2) \in (B^1_\omega \times B^2_\varpi)^c\big\} \le -\inf_{(\tilde{\omega},\tilde{\varpi})\in\hat{F}} \tilde{I}_3(\tilde{\omega}, \tilde{\varpi}),$$

$$\tilde{I}_3(\tilde{\omega}, \tilde{\varpi}) = \begin{cases} \frac{1}{2}\mathfrak{H}_{\tilde{C}}(\tilde{\varpi} \,\|\, \omega) & \text{if } \tilde{\omega} = \omega, \\ \infty & \text{otherwise.} \end{cases}$$

https://doi.org/10.1051/ps/2010016 Published online by Cambridge University Press

where $\hat{F} = (B^1_\omega \times B^2_\varpi)^c$ and $(B^1_\omega \times B^2_\varpi)^c$ is the complement of the set $B^1_\omega \times B^2_\varpi$. It remain for us to show that the infimum is positive. Suppose by contradiction there exists the sequence $(\omega_n, \varpi_n) \in \hat{F}$ such that $\tilde{I}_3(\tilde{\omega}, \tilde{\varpi}) \downarrow 0$. Then, because $\tilde{I}_3$ is good rate function with all its level sets compact, and by lower semicontinuity of the mapping $(\tilde{\omega}, \tilde{\varpi}) \to \tilde{I}_3(\tilde{\omega}, \tilde{\varpi})$, we can construct a limit point $(\tilde{\varpi}, \tilde{\omega}) \in \hat{F}$ with $\tilde{I}_3(\tilde{\omega}, \tilde{\varpi}) = 0$. This means $\tilde{\omega} = \omega$ and $\tilde{\varpi} = \tilde{C}\tilde{\omega} \otimes \tilde{\omega} = \varpi$, and hence, contradicting $(\tilde{\omega}, \tilde{\varpi}) \in \hat{F}$. □

3.13. **Upper bound in Theorems 3.6(ii)**

We define for $(\omega, \varpi) \in \mathcal{M}(\mathcal{X}) \times \tilde{\mathcal{M}}_*(\mathcal{X} \times \mathcal{X})$, the function $\hat{I}_4(\omega, \varpi)$ by

$$\hat{I}_4(\omega, \varpi) = \sup_{\tilde{f} \in \mathcal{C}_1} \Big\{ \sum_{a \in \mathcal{X}} \big(\tilde{f}(a) - U_{\tilde{f}}\big)\omega(a) \Big\} \tag{3.38}$$

Lemma 3.20. *For each closed set $F \subset \tilde{\mathcal{M}}_*(\mathcal{X} \times \mathcal{X})$, we have*

$$\limsup_{n \to \infty} \tfrac{1}{n} \log \mathbb{P}\big\{(L^1, L^2) \in F\big\} \le - \inf_{(\omega, \varpi) \in F} \hat{I}_4(\omega, \varpi).$$

Proof. Fix $\tilde{f} \in \mathcal{C}_1$ and take $\tilde{g}(a, b) = 0$, for all $a, b \in \mathcal{X}$. We observe that by this choice of $\tilde{g}$

$$\tilde{h}^{(2)}_n(a, b) = 0, \text{ for all } a, b \in \mathcal{X}.$$

Using (3.12) we obtain $\mathbb{E}\big\{\mathrm{e}^{n\langle L^1, \tilde{f} - U_{\tilde{f}}\rangle}\big\} = \int \mathrm{d}\tilde{\mathbb{P}} \le 1$ and therefore, we have

$$\limsup_{n \to \infty} \tfrac{1}{n} \log \mathbb{E}\Big\{\mathrm{e}^{n\langle L^1, \tilde{f} - U_{\tilde{f}}\rangle}\Big\} \le 0. \tag{3.39}$$

Now fix $\varepsilon > 0$ and write $\hat{I}^\varepsilon_4(\omega, \varpi) := \min\{\hat{I}_4(\omega, \varpi), \varepsilon^{-1}\} - \varepsilon$. We suppose $(\omega, \varpi) \in F$ and choose $\tilde{f} \in \mathcal{C}_1$ such that

$$\langle \tilde{f} - U_{\tilde{f}}, \omega\rangle \ge \hat{I}^\varepsilon_4(\omega, \varpi).$$

By finiteness of $\mathcal{X}$, we can find open neighbourhoods B^2_ϖ and B^1_ω of ϖ, ω such that

$$\inf_{\tilde{\omega} \in B^1_\omega} \big\{\langle \tilde{\omega}, \tilde{f} - U_{\tilde{f}}\rangle\big\} \ge \hat{I}^\varepsilon_4(\omega, \varpi) - \varepsilon.$$

Using Chebysheff's inequality and (3.39), we have that

$$\begin{aligned} \limsup_{n \to \infty} \tfrac{1}{n} \log \mathbb{P}\big\{(L^1, L^2) \in B^1_\omega \times B^2_\varpi\big\} &\le \limsup_{n \to \infty} \tfrac{1}{n} \log \mathbb{E}\Big\{\mathrm{e}^{n\langle L^1, \tilde{f} - U_{\tilde{f}}\rangle}\Big\} - \hat{I}^\varepsilon_4(\omega, \varpi) + \varepsilon \\ &\le -\hat{I}^\varepsilon_4(\omega, \varpi) + \varepsilon. \end{aligned} \tag{3.40}$$

By Lemma 3.9 with $\alpha = \varepsilon^{-1}$, we choose $N(\varepsilon) \in \mathbb{N}$ such that

$$\limsup_{n \to \infty} \tfrac{1}{n} \log \mathbb{P}\Big\{|E| > a_n n^2 N(\varepsilon)\Big\} \le -\varepsilon^{-1}.$$

We define for this N the set $K_{N(\varepsilon)}$ by

$$K_{N(\varepsilon)} = \big\{(\omega, \varpi) \in \mathcal{M}(\mathcal{X}) \times \tilde{\mathcal{M}}_*(\mathcal{X} \times \mathcal{X}) : \|\varpi\| \le 2N(\varepsilon)\big\}.$$

The set $K_{N(\varepsilon)} \cap F$ is compact and therefore may be covered by finitely many sets

https://doi.org/10.1051/ps/2010016 Published online by Cambridge University Press

$B^1_{\omega_r} \times B^2_{\varpi_r}$, $r = 1, \dots, m$ with $(\omega_r, \varpi_r) \in F$ for $r = 1, \dots, m$. Hence, we have

$$\mathbb{P}\big\{(L^1,\, L^2) \in F\big\} \le \sum_{r=1}^{m} \mathbb{P}\big\{(L^1, L^2) \in B^1_{\omega_r} \times B^2_{\varpi_r}\big\} + \mathbb{P}\big\{(L^1, L^2) \notin K_{N(\varepsilon)}\big\}.$$

Now we use (3.40) to obtain, for all sufficiently small $\varepsilon > 0$,

$$\begin{aligned}\limsup_{n\to\infty} \tfrac{1}{n} \log \mathbb{P}\big\{(L^1, L^2) \in F\big\} &\le \max_{r=1}^{m} \limsup_{n\to\infty} \tfrac{1}{n} \log \mathbb{P}\big\{(L^1, L^2) \in B^1_{\omega_r} \times B^2_{\varpi_r}\big\} - \varepsilon^{-1} \\ &\le - \inf_{(\omega,\varpi)\in F} \hat{I}^{\varepsilon}_4(\omega, \varpi) + \varepsilon.\end{aligned}$$

Taking $\varepsilon \downarrow 0$ we have the desired statement. □

By the variational characterization of the relative entropy we have $I_4(\omega,\, \varpi) = H(\omega \,\|\, \mu)$, which ends the proof of the upper bound.

3.14. **Lower bound in Theorems 3.6(ii)**

Lemma 3.21. *For every open set $O \subset \mathcal{M}(\mathcal{X}) \times \tilde{\mathcal{M}}_*(\mathcal{X} \times \mathcal{X})$,*

$$\liminf_{n\to\infty} \tfrac{1}{n} \log \mathbb{P}\Big\{(L^1, L^2) \in O\Big\} \ge - \inf_{(\omega,\, \varpi)\in O} I_4(\omega,\, \varpi). \tag{3.41}$$

Proof. Suppose $(\omega,\, \varpi) \in O$ with $\varpi \ll C\omega \otimes \omega$. We define the function $\tilde{f}_\omega \colon \mathcal{X} \to \mathbb{R}$ by

$$\tilde{f}_\omega(a) = \begin{cases} \log \frac{\omega(a)}{\mu(a)}, & \text{if } \omega(a) > 0, \\ 0, & \text{otherwise,} \end{cases}$$

and the symmetric function $g_\varpi \colon \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ by

$$g_\varpi(a, b) = \begin{cases} \log \frac{\varpi(a,b)}{C(a,b)\omega(a)\omega(b)}, & \text{if } \varpi(a, b) > 0, \\ 0, & \text{otherwise.} \end{cases}$$

Set $\tilde{g}_{\tilde{\varpi}}(a, b) = n a_n g_\varpi(a, b)$, for all $a, b \in \mathcal{X}$ and note by the choice of $\tilde{g}$ we have

$$\lim_{n\to\infty} \tilde{h}^{(2)}_n(a, b) = 0, \ \text{ for all } a, b \in \mathcal{X}.$$

Choose B^1_ω, B^2_ϖ open neighbourhoods of ω, ϖ such that $B^1_\omega \times B^2_\varpi \subset O$ and for all $(\tilde{\omega}, \tilde{\varpi}) \in B^1_\omega \times B^2_\varpi$,

$$\langle \omega,\, \tilde{f}_\omega \rangle + \tfrac{1}{2} \|\varpi\| \varepsilon - \varepsilon \le \langle \tilde{\omega},\, \tilde{f}_\omega \rangle + \tfrac{1}{2} \|\tilde{\varpi}\| \varepsilon.$$

We use the probability measure $\tilde{\mathbb{P}}$ given by $\tilde{g}_\varpi$. We observe that the colour law is ω and the connection probabilities satisfy $a_n^{-1} \tilde{p}_n(a, b) \to \tilde{C}(a, b) := \varpi(a, b)/(\omega(a)\omega(b))$, as n approaches infinity. Therefore, using (3.12) we have that

$$\begin{aligned}\mathbb{P}\Big\{(L^1, L^2) \in O\Big\} &\ge \tilde{\mathbb{E}}\Big\{ \tfrac{\mathrm{d}\mathbb{P}}{\mathrm{d}\tilde{\mathbb{P}}}(X) \mathbb{1}_{\{(L^1,L^2)\in B^1_\omega \times B^2_\varpi\}} \Big\} \\ &= \tilde{\mathbb{E}}\left\{ \mathrm{e}^{-n\langle L^1, \tilde{f}_\omega\rangle - n\left\langle \frac{1}{2} L^2, \tilde{g} \right\rangle - n\left\langle \frac{1}{2} L^1 \otimes L^1, \tilde{g} \right\rangle + n\left\langle L^1_\Delta, \tilde{h}^{(2)}_n \right\rangle} \mathbb{1}_{\{(L^1,\, L^2)\in B^1_\omega \times B^2_\varpi\}} \right\} \\ &\ge \mathrm{e}^{-n\,\langle \omega,\, \tilde{f}_\omega\rangle - n\,\|\varpi\|\varepsilon/2 - n\,\varepsilon - a_n n^2 l/2 + o(1)} \times \tilde{\mathbb{P}}\Big\{(L^1, L^2) \in B^1_\omega \times B^2_\varpi\Big\},\end{aligned}$$

https://doi.org/10.1051/ps/2010016 Published online by Cambridge University Press

where $l = 0 \wedge \max_{a,b\in\mathcal{X}} \tilde{g}_{\tilde{\varpi}}(a,b)$. Therefore, we have that

$$\liminf_{n\to\infty} \tfrac{1}{n} \log \mathbb{P}\big\{(L^1,L^2)\in O\big\} \geq -\langle \omega,\, \tilde{f}_\omega\rangle - \tfrac{1}{2}\,\|\varpi\|\varepsilon - \varepsilon + \liminf_{n\to\infty} \tfrac{1}{n} \log \tilde{\mathbb{P}}\big\{(L^1,L^2)\in B^1_\omega \times B^2_\varpi\big\}.$$

The result follows once we prove that

$$\liminf_{n\to\infty} \tfrac{1}{n} \log \tilde{\mathbb{P}}\big\{(L^1,L^2)\in B^1_\omega \times B^2_\varpi\big\} = 0. \tag{3.42}$$

We use the lower bound of the LDP on the scale $a_n n^2$ (but now with the law $\mathbb{P}$ replaced by $\tilde{\mathbb{P}}$) to prove (3.42). We observe from the lower bound of Theorem 3.6(i) that, for any $\delta > 0$ and for large n we have

$$\tilde{\mathbb{P}}\big\{(L^1,\, L^2)\in B^1_\omega \times B^2_\varpi\big\} \geq \mathrm{e}^{-a_n n^2(\alpha_3(\omega,\,\varpi)+\delta)},$$

where $\alpha_3(\omega,\,\varpi) = \inf\big\{\tilde{I}_3(\tilde{\omega},\,\tilde{\varpi}) : (\tilde{\omega},\,\tilde{\varpi})\in B^1_\omega \times B^2_\varpi\big\}$ and

$$\tilde{I}_3(\tilde{\omega},\,\tilde{\varpi}) = \begin{cases} \frac{1}{2}\mathfrak{H}_{\tilde{C}}(\tilde{\varpi}\,\|\,\omega) & \text{if } \tilde{\omega} = \omega, \\ \infty & \text{otherwise.} \end{cases}$$

Therefore, we have that

$$\liminf_{n\to\infty} \tfrac{1}{n} \log \tilde{\mathbb{P}}\big\{(L^1,\, L^2)\in B^1_\omega \times B^2_\varpi\big\} \geq \liminf_{n\to\infty} a_n n\{-\alpha_3(\omega,\,\varpi) - \delta\} = 0,$$

since $a_n n \to 0$ as n approaches infinity. This concludes the proof of the Lemma. □

3.15. **Proof of Theorem 2.1**

Recall N_0 from the boundedness of $\mathbb{Q}$ and write

$$\mathcal{X}_0^* = \bigcup_{n=0}^{N_0} \{n\} \times \mathcal{X}^n.$$

We equip it with the discrete topology. We recall also that π is the unique eigenvector (normailzed to a probability vector) of the matrix A corresponding to the eigenvalue 1. We recall that $\mathbb{P}_n$ is the law of a multitype Galton-Watson tree conditioned to have n vertices, and derive from Theorem 3.1 the following weak law of large numbers.

Lemma 3.22. *Suppose that X is an irreducible, critical multitype Galton-Watson tree with bounded offspring law $\mathbb{Q}$, conditioned to have exactly n vertices. Then, for any $\varepsilon > 0$*

$$\lim_{n\to\infty} \mathbb{P}_n\Big\{ \max_{(a,c)\in\mathcal{X}\times\mathcal{X}_0^*} \big|M_X(a,c) - \pi\otimes\mathbb{Q}(a,c)\big| \geq \varepsilon \Big\} = 0. \tag{3.43}$$

Proof. Define the closed set

$$F = \big\{\, \nu\in\mathcal{M}(\mathcal{X}\times\mathcal{X}_0) : \max_{(a,c)\in\mathcal{X}\times\mathcal{X}_0^*} \big|\nu(a,c) - \pi\otimes\mathbb{Q}(a,c)\big| \geq \varepsilon \,\big\}.$$

We observe from Theorem 3.1 that,

$$\limsup_{n\to\infty} \tfrac{1}{n} \log \mathbb{P}_n\big\{\, M_X \in F \,\big\} \leq -\inf_{\nu\in F} J(\nu). \tag{3.44}$$

We show by contradiction that the right hand side of (3.44) is negative.

https://doi.org/10.1051/ps/2010016 Published online by Cambridge University Press

To do this, we suppose that there exists sequence ν_n in F such that $J(\nu_n) \downarrow 0$. Then, because J is a good rate function and its level sets are compact, and by lower semicontinuity of the mapping $\nu \mapsto J(\nu)$, there is a limit $\nu \in F$ with $J(\nu) = 0$. Hence, we have that ν is shift-invariant and $H(\nu \,\|\, \nu_1 \otimes \mathbb{Q}) = 0$. This implies $\nu(a,c) = \nu_1 \otimes \mathbb{Q}(a,c)$, for every $(a,c) \in \mathcal{X} \times \mathcal{X}_0^*$. Using shift-invariance of ν, for any $b \in \mathcal{X}$, we have

$$\sum_{a\in\mathcal{X}} A(b,a)\nu_1(a) = \sum_{(a,c)\in\mathcal{X}\times\mathcal{X}_0^*} \mathbb{Q}\{c \,|\, a\} m(b,c)\nu_1(a) = \sum_{(a,c)\in\mathcal{X}\times\mathcal{X}_0^*} \nu(a,c)m(b,c) = \nu_1(b).$$

This means that ν_1 is a nonnegative eigenvector of A. By uniqueness of the Perron-Frobenius eigenvector, see, for example Dembo *et al.* [7], Theorem 3.1.1(d), we infer that $\nu_1 = \pi$. This contradicts $\pi \otimes \mathbb{Q} \notin F$. □

We recall that $\mathcal{T}$ is set of all finite rooted planar trees T, $V = V(T)$ is set of all vertices and $|T|$ is the number of vertices in the tree T. We now compute the probability weight $P_n(x)$ of $x \in \mathcal{T}$ as

$$P_n(x) = \tfrac{\mu(x(\rho))}{\mathbb{P}\{|T|=n\}} \prod_{v\in V(T),|T|=n} \mathbb{Q}\big\{C(v) = c(v) \,|\, X(v) = x(v)\big\},$$

$(x(v), c(v))$ is the type, and the configuration of children of vertex v of $x \in \mathcal{T}$. Therefore, we have that

$$-\tfrac{1}{n}\log P_n(x) = -\tfrac{1}{n}\log \mu(x(\rho)) + \tfrac{1}{n}\log \mathbb{P}\{|T| = n\} + \langle M_x, -\log \mathbb{Q}\rangle.$$

Now the term $\frac{1}{n}\log\mu(x(\rho))$ converges to zero, while the term $\frac{1}{n}\log\mathbb{P}\{|T| = n\}$ converges to zero because $\mathbb{Q}$ is bounded. See Dembo *et al.* [8], Lemma 3.1. We observe that $-\log\mathbb{Q}$ is almost surely bounded on the support of M_X and therefore, by Lemma 3.22 we have $\langle M_x, -\log\mathbb{Q}\rangle \to \langle \pi\otimes\mathbb{Q}, -\log\mathbb{Q}\rangle$, which concludes the proof of Theorem 2.1.

3.16. **Derivation of Theorems 2.2 and 2.3**

We recall that $\mathbb{P}_n$ is the law of a coloured random graph with n vertices, and derive from our large deviation principles for coloured random graphs the following weak law of large numbers.

Lemma 3.23. *Suppose that X is a coloured random graph with colour law $\mu\colon \mathcal{X} \to (0,1]$ and connection probabilities $p_n : \mathcal{X}\times\mathcal{X} \to [0,1]$ such that $a_n^{-1}p_n(a,b) \to C(a,b)$ for some sequence (a_n) with $a_n n \to 0$ or $a_n n \to 1$ or $a_n n \to \infty$ and $C : \mathcal{X}\times\mathcal{X} \to [0,\infty)$ nonzero. Then, for any $\varepsilon > 0$ we have*

$$\lim_{n\to\infty} \mathbb{P}_n\big\{\sup_{a\in\mathcal{X}} |L^1(a) - \mu(a)| \geq \varepsilon\big\} = 0$$

and

$$\lim_{n\to\infty} \mathbb{P}_n\big\{\sup_{a,b\in\mathcal{X}} |L^2(a,b) - \mu(a)C(a,b)\mu(b)| \geq \varepsilon\big\} = 0.$$

From Theorem 3.2, Theorem 3.4 and Theorem 3.6 we prove this lemma.

To begin, we define a closed set

$$F_1 = \big\{(\omega,\varpi) \in \mathcal{M}(\mathcal{X}) \times \tilde{\mathcal{M}}_*(\mathcal{X}\times\mathcal{X})\colon \sup_{a,b\in\mathcal{X}} |\varpi(a,b) - \mu(a)C(a,b)\mu(b)| \geq \varepsilon\big\}.$$

We observe that in the sparse case (when $na_n \to 1$), by Theorem 3.2,

$$\limsup_{n\to\infty} \tfrac{1}{n}\log \mathbb{P}_n\Big\{(L^1,L^2) \in F_1\Big\} \leq -\inf_{(\omega,\varpi)\in F_1} I(\omega,\varpi). \tag{3.45}$$

We show by contradiction that the right handside of (3.45) is negative. For this purpose suppose that there exists sequence (ω_n,ϖ_n) in F_1 such that $I(\omega_n,\varpi_n) \downarrow 0$. Then, because I is a good rate function and its level sets

https://doi.org/10.1051/ps/2010016 Published online by Cambridge University Press

are compact, and by lower semicontinuity of the mapping $(\omega,\varpi)\mapsto I(\omega,\varpi)$, there is a limit point $(\omega,\varpi)\in F_1$ with $I(\omega,\varpi)=0$. By Doku *et al.* [10], Lemma 3.4, we have $H(\omega\,\|\,\mu)=0$ and $\mathfrak{H}_C(\varpi\,\|\,\omega)=0$. This implies $\omega(a)=\mu(a)$, and $\varpi(a,b)=C(a,b)\omega(a)\omega(b)$, for $a,b\in\mathcal{X}$ which contradicts $(\omega,\varpi)\in F_1$. Hence as required.

For the subcritical case we can argue similarly with the LDP on the scale $a_n n^2$ with rate function I_3. The first statement of Lemma 3.23 follows similarly using the set

$$F_2=\big\{(\omega,\varpi)\in\mathcal{M}(\mathcal{X})\times\tilde{\mathcal{M}}_*(\mathcal{X}\times\mathcal{X})\colon\ \sup_{a\in\mathcal{X}}|\omega(a)-\mu(a)|\geq\varepsilon\big\}$$

and the LDP of Theorem 3.2 in the sparse case, and the LDP on the scale n with rate function I_4 in the subcritical case. Finally, in the supercritical case, an analogous argument can be carried out using $F=F_1\cup F_2$ and the LDP on the scale n with rate function I_2.

We recall that V is a fixed set of n vertices, say $V=\{1,\dots,n\}$, $\mathcal{G}_n$ is the set of all (simple) graphs with vertex set $V=\{1,\dots,n\}$ and $E\subset\mathcal{E}:=\big\{(u,v)\in V\times V\,:\,u<v\big\}$ the edge set.

We now compute the probability weight $P_n(x)$ of $x\in\mathcal{G}_n$ as

$$\begin{aligned}P_n(x)&=\prod_{u\in V}\mu(x(u))\prod_{(u,v)\in E}p_n(x(u),x(v))\prod_{(u,v)\notin E}\big(1-p_n(x(u),x(v))\big)\\&=\prod_{u\in V}\mu(x(u))\prod_{(u,v)\in E}\tfrac{p_n(x(u),x(v))}{1-p_n(x(u),x(v))}\prod_{(u,v)\in\mathcal{E}}\big(1-p_n(x(u),x(v))\big).\end{aligned}$$

Therefore, we have in the case of Theorem 2.2

$$\begin{aligned}-\tfrac{1}{a_n n^2\log n}\log P_n(x)=\langle L^1,-\tfrac{\log\mu}{a_n\,n\,\log n}\rangle+\tfrac12\langle L^2,-\tfrac{\log(p_n/(1-p_n))}{\log n}\rangle\\+\tfrac12\langle L^1\otimes L^1,-\tfrac{\log(1-p_n)}{a_n\log n}\rangle+\tfrac12\langle L^1_\Delta,-\tfrac{\log(1-p_n)}{a_n\,n\,\log n}\rangle.\end{aligned}$$

In the case of Theorem 2.3 we have

$$\begin{aligned}-\tfrac1n\log P_n(x)=\langle L^1,-\log\mu\rangle+\tfrac12\langle L^2,-\tfrac{\log(p_n/(1-p_n))}{\log n}\rangle\\+\tfrac12\langle L^1\otimes L^1,-n\log(1-p_n)\rangle+\tfrac12\langle L^1_\Delta,-\log(1-p_n)\rangle.\end{aligned}$$

Now in the first case the integrands $\frac{-\log\mu}{a_n\,n\,\log n}$, $\frac{-\log(1-p_n)}{a_n\log n}$ and $\frac{-\log(1-p_n)}{a_n\,n\,\log n}$ all converge to zero, while $\frac{-\log(p_n/(1-p_n))}{\log n}\to 1$. Hence Theorem 2.2 follows from Theorem 3.23.

In the second case both integrand $-\log(1-p_n)$ and $-n\,\log(1-p_n)$ converges to zero. Therefore, Theorem 2.3 follows from Lemma 3.23.

Acknowledgements. This paper contains material from my Ph.D. thesis (Bath). I would like to thank LMS for supporting my recent visit to Bath.

References

[1] R. Arking, *Biology of Aging*, 2nd edition. Sinauer, Sunderland, MA (1998).
[2] J.D. Biggins, Large deviations for mixtures. *Electron. Commun. Probab.* **9** (2004) 60–71.
[3] J.D. Biggins and D.B. Penman, Large deviations in randomly coloured random graphs. *Electron. Commun. Probab.* **14** (2009) 290–301.
[4] C. Cannings and D.B. Penman, Models of random graphs and their applications, Handbook of Statistics. *Stochastic Processes: Modeling and Simulation,* edited by D.N. Shanbhag and C.R. Rao. Elsevier **21** (2003) 51–91.
[5] T.M. Cover and J.A. Thomas, Elements of Information Theory, in *Wiley Series in Telecommunications* (1991).
[6] A. Dembo and I. Kontoyiannis, Source Coding, Large deviations and Approximate Pattern. *IEEE Trans. Inf. Theory* **48** (2002) 1590–1615.
[7] A. Dembo and O. Zeitouni, *Large deviations techniques and applications*. Springer, New York (1998).

https://doi.org/10.1051/ps/2010016 Published online by Cambridge University Press

[8] A. Dembo, P. Mörters and S. Sheffield, Large deviations of Markov chains indexed by random trees. *Ann. Inst. Henri Poincaré: Probab. Stat.* **41** (2005) 971–996.
[9] K. Doku-Amponsah, *Large deviations and basic information theory for hierarchical and networked data structures*. Ph.D. thesis, Bath (2006).
[10] K. Doku-Amponsah and P. Mörters, Large deviation principle for empirical measures of coloured random graphs. *Ann. Appl. Prob.* **20** (2010) 1989–2021.
[11] M. Kimmel and D.E. Axelrod, *Branching Processes with Biology*. Springer, New York (2002).
[12] C.J. Mode, *Multitype Branching Processes Theory and Applications*. American Elsevier, New York (1971).
[13] M.E. Newman, *Random graphs as models of networks*. `http://arxiv.org/abs/cond-mat/0202208`
[14] P. Olofsson and C.A. Shaw, Exact sampling formulas for multitype Galton-Watson processes. *J. Math. Biol.* **45** (2002) 279–293.
[15] D.B. Penman, *Random graphs with correlation structure*. Ph.D. thesis, Sheffield (1998).

https://doi.org/10.1051/ps/2010016 Published online by Cambridge University Press