LARGE DEVIATION PRINCIPLES FOR EMPIRICAL MEASURES OF COLOURED RANDOM GRAPHS

By Kwabena Doku-Amponsah and Peter Mörters[1]

Abstract. For any finite coloured graph we define the *empirical neighbourhood measure*, which counts the number of vertices of a given colour connected to a given number of vertices of each colour, and the *empirical pair measure*, which counts the number of edges connecting each pair of colours. For a class of models of sparse coloured random graphs, we prove large deviation principles for these empirical measures in the weak topology. The rate functions governing our large deviation principles can be expressed explicitly in terms of relative entropies. We derive a large deviation principle for the degree distribution of Erdős-Rényi graphs near criticality.

1. Introduction

In this paper we study a random graph model where each vertex of the graph carries a random symbol, spin or colour. The easiest model of this kind is that of an Erdős-Rényi graph where additionally each vertex is equipped with an independently chosen colour. The more general models of coloured random graphs we consider here allow for a dependence between colour and connectivity of the vertices.

With each coloured graph we associate its *empirical neighbourhood measure*, which records the number of vertices of a given colour with a given number of adjacent vertices of each colour. From this quantity one can derive a host of important characteristics of the coloured graph, like its degree distribution, the number of edges linking two given colours, or the number of isolated vertices of any colour. The aim of this paper is to derive a large deviation principle for the empirical neighbourhood measure.

To be more specific about our model, we consider coloured random graphs constructed as follows: In the first step each of n fixed vertices independently gets a colour, chosen according to some law μ on the finite set $\mathcal{X}$ of colours. In the second step we connect any pair of vertices independently with a probability $p(a,b)$ depending on the colours $a,b \in \mathcal{X}$ of the two vertices. This model, which comprises the simple Erdős-Rényi graph with independent colours as a special case, was introduced by Penman in his thesis [15], see [6] for an exposition, and rediscovered later by Söderberg [16]. It is also a special case of the inhomogeneous random graphs studied in [3, 13].

Our main concern in this paper are asymptotic results when the number n of vertices goes to infinity, while the connection probabilities are of order $1/n$. This leads to an average number of edges of order n, the *near critical* or *sparse* case. Our methods also allow the study of the sub- and supercritical regimes. Some results on these cases are discussed in [10].

[1]Corresponding Author.

AMS Subject Classification: 60F10, 05C80

Keywords: Random graph, Erdős-Rényi graph, random randomly coloured graph, typed graph, spins, joint large deviation principle, empirical pair measure, empirical measure, degree distribution, entropy, relative entropy, Ising model on a random graph, partition function.

Recall that a *rate function* is a non-constant, lower semicontinuous function I from a polish space $\mathcal{M}$ into $[0,\infty]$, it is called *good* if the level sets $\{I(m)\le x\}$ are compact for every $x\in[0,\infty)$. A functional M from the set of finite coloured graphs to $\mathcal{M}$ is said to satisfy a *large deviation principle* with rate function I if, for all Borel sets $B\subset\mathcal{M}$,

$$-\inf_{m\in\operatorname{int}B} I(m)\le\liminf_{n\to\infty}\tfrac{1}{n}\log\mathbb{P}_n\big\{M(X)\in B\big\}$$
$$\le\limsup_{n\to\infty}\tfrac{1}{n}\log\mathbb{P}_n\big\{M(X)\in B\big\}\le-\inf_{m\in\operatorname{cl}B} I(m)\,,$$

where X under $\mathbb{P}_n$ is a coloured random graph with n vertices and int B and cl B refer to the interior, resp. closure, of the set B.

Apart from the empirical neighbourhood measure defined above, we also consider the *empirical pair measure*, which counts the number of edges connecting any given pair of colours, and the *empirical colour measure*, which simply counts the number of vertices of any given colour. The main result of this paper is a joint large deviation principle for the empirical neighbourhood measure and the empirical pair measure of a coloured random graph in the weak topology, see Theorem 2.1. In the course of the proof of this principle, two further interesting large deviation principles are established: A large deviation principle for the empirical neighbourhood measure conditioned to have a given empirical pair and colour measure, see Theorem 2.5, and a joint large deviation principle for the empirical colour measure and the empirical pair measure, see Theorem 2.3 (b). For all these principles we obtain a completely explicit rate function given in terms of relative entropies.

Our motivation for this project is twofold. On the one hand one may consider the coloured random graphs as a very simple random model of *networked data.* The data is described as a text of fixed length, consisting of words chosen from a finite dictionary, together with a random number of unoriented edges or links connecting the words. Large deviation results for the empirical neighbourhood measure permit the calculation of the asymptotic number of bits needed to transmit a large amount of such data with arbitrarily small error probability, see [10] where this idea is followed up.

On the other hand we are working towards understanding simple models of *statistical mechanics* defined on random graphs. Here, typically, the colours of the vertices are interpreted as spins, taken from a finite set of possibilities, and the Hamiltonian of the system is an integral of some function with respect to the empirical neighbourhood measure. As a very simple example we provide the annealed asymptotics of the random partition function for the Ising model on an Erdős-Rényi graph, as the graph size goes to infinity.

As a more substantial example, we consider the Erdős-Rényi graph model on n vertices, where edges are inserted with probability $p_n\in[0,1]$ independently for any pair of vertices. We assume that $np_n\to c\in(0,\infty)$. From our main result we derive a large deviation principle for the degree distribution, see Corollary 2.2. This example seems to be new in this explicit form.

2. Statement of the results

Let V be a fixed set of n vertices, say $V=\{1,\dots,n\}$ and denote by $\mathcal{G}_n$ the set of all (simple) graphs with vertex set $V=\{1,\dots,n\}$ 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 describe simply *unordered* edges. Note that for all n, we have $0\le|E|\le\frac{1}{2}\,n(n-1)$. Let $\mathcal{X}$ be a finite alphabet or colour set $\mathcal{X}$ and denote by $\mathcal{G}_n(\mathcal{X})$ be the set of all coloured graphs with colour set $\mathcal{X}$ and n vertices.

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 *connection probability* $p_n(X(u), X(v))$. We always consider $X = ((X(v) : v \in V), E)$ under the joint law of graph and colour and interpret X as coloured random graph.

We are interested in the properties of the randomly coloured graphs for large n in the *sparse* or *near critical* case, i.e. we assume that the connection probabilities satisfy $np_n(a,b) \to C(a,b)$ for all $a, b \in \mathcal{X}$, where $C\colon \mathcal{X} \times \mathcal{X} \to [0,\infty)$ is a symmetric function, which is not identically equal to zero.

To fix some notation, 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. When applying $\nu \in \tilde{\mathcal{M}}(\mathcal{Y})$ to some function $g\colon \mathcal{Y} \to \mathbb{R}$ we use the scalar-product notation

$$\langle \nu, g \rangle := \sum_{y \in \mathcal{Y}} \nu(y)\, g(y),$$

and denote by $\|\nu\|$ its total mass. Further, if $\mu \in \tilde{\mathcal{M}}(\mathcal{Y})$ and $\nu \ll \mu$ we denote by

$$H(\nu \,\|\, \mu) = \sum_{y \in \mathcal{Y}} \nu(y) \log\left(\tfrac{\nu(y)}{\mu(y)}\right)$$

the *relative entropy* of ν with respect to μ. We set $H(\nu \,\|\, \mu) = \infty$ if $\nu \not\ll \mu$. By $\mathcal{N}(\mathcal{Y})$ we denote the space of counting measures on $\mathcal{Y}$, i.e. those measures taking values in $\mathbb{N} \cup \{0\}$, endowed with the discrete topology. Finally, we denote by $\tilde{\mathcal{M}}_*(\mathcal{Y}\times\mathcal{Y})$ the subspace of symmetric measures in $\tilde{\mathcal{M}}(\mathcal{Y}\times\mathcal{Y})$.

With any coloured graph $X = ((X(v) : v \in V), E)$ with n vertices we associate a probability measure, the *empirical colour measure* $L^1 \in \mathcal{M}(\mathcal{X})$, by

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

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

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

The total mass $\|L^2\|$ of the empirical pair measure is $2|E|/n$. Finally we define a further probability measure, the *empirical neighbourhood measure* $M \in \mathcal{M}(\mathcal{X} \times \mathcal{N}(\mathcal{X}))$, by

$$M(a,\ell) := \frac{1}{n} \sum_{v \in V} \delta_{(X(v), L(v))}(a,\ell), \qquad \text{for } (a,\ell) \in \mathcal{X} \times \mathcal{N}(\mathcal{X}),$$

where $L(v) = (l^v(b),\, b \in \mathcal{X})$ and $l^v(b)$ is the number of vertices of colour b connected to vertex v. For every $\nu \in \mathcal{M}(\mathcal{X} \times \mathcal{N}(\mathcal{X}))$ let ν_1 and ν_2 be the $\mathcal{X}$-marginal and the $\mathcal{N}(\mathcal{X})-$marginal of the measure ν, respectively. Moreover, we define a measure $\langle \nu(\cdot,\ell),\, \ell(\cdot) \rangle \in \tilde{\mathcal{M}}(\mathcal{X} \times \mathcal{X})$ by

$$\langle \nu(\cdot,\ell),\, \ell(\cdot) \rangle(a,b) := \sum_{\ell \in \mathcal{N}(\mathcal{X})} \nu(a,\ell)\ell(b), \qquad \text{for } a, b \in \mathcal{X}.$$

Define the function $\Phi\colon \mathcal{M}(\mathcal{X} \times \mathcal{N}(\mathcal{X})) \to \mathcal{M}(\mathcal{X}) \times \tilde{\mathcal{M}}(\mathcal{X} \times \mathcal{X})$ by $\Phi(\nu) = (\nu_1, \langle \nu(\cdot,\ell),\, \ell(\cdot) \rangle)$, and observe that $\Phi(M) = (L^1, L^2)$, if these quantities are defined as empirical neighbourhood, colour, and pair measures of a coloured graph. Note that while the *first* component of Φ is a continuous function, the *second* component is *discontinuous* in the weak topology.

To formulate the large deviation principle, we call a pair of measures $(\varpi,\nu)\in\tilde{\mathcal{M}}(\mathcal{X}\times\mathcal{X})\times\mathcal{M}(\mathcal{X}\times\mathcal{N}(\mathcal{X}))$ *sub-consistent* if

$$\langle\nu(\cdot,\ell),\,\ell(\cdot)\rangle(a,b)\le\varpi(a,b),\qquad\text{for all } a,b\in\mathcal{X},\tag{2.1}$$

and *consistent* if equality holds in (2.1). Observe that, if ν is the empirical neighbourhood measure and ϖ the empirical pair measure of a coloured graph, (ϖ,ν) is consistent and both sides in (2.1) represent

$$\tfrac{1}{n}\,(1+\mathbb{1}_{\{a=b\}})\,\sharp\big\{\text{edges between vertices of colours } a \text{ and } b\big\}.$$

For a measure $\varpi\in\tilde{\mathcal{M}}_*(\mathcal{X}\times\mathcal{X})$ and a measure $\omega\in\mathcal{M}(\mathcal{X})$, define

$$\mathfrak{H}_C(\varpi\,\|\,\omega):=H\big(\varpi\,\|\,C\omega\otimes\omega\big)+\|C\omega\otimes\omega\|-\|\varpi\|\,,$$

where the measure $C\omega\otimes\omega\in\tilde{\mathcal{M}}(\mathcal{X}\times\mathcal{X})$ is defined by $C\omega\otimes\omega(a,b)=C(a,b)\omega(a)\omega(b)$ for $a,b\in\mathcal{X}$. It is not hard to see that $\mathfrak{H}_C(\varpi\,\|\,\omega)\ge 0$ and equality holds if and only if $\varpi=C\omega\otimes\omega$ (see Lemma 3.2). For every $(\varpi,\nu)\in\tilde{\mathcal{M}}_*(\mathcal{X}\times\mathcal{X})\times\mathcal{M}(\mathcal{X}\times\mathcal{N}(\mathcal{X}))$ define a probability measure $Q=Q[\varpi,\nu]$ on $\mathcal{X}\times\mathcal{N}(\mathcal{X})$ by

$$Q(a\,,\,\ell):=\nu_1(a)\prod_{b\in\mathcal{X}}e^{-\frac{\varpi(a,b)}{\nu_1(a)}}\,\frac{1}{\ell(b)!}\,\Big(\frac{\varpi(a,b)}{\nu_1(a)}\Big)^{\ell(b)},\quad\text{for } a\in\mathcal{X},\,\ell\in\mathcal{N}(\mathcal{X}).\tag{2.2}$$

We have now set the stage to state our principal theorem, the large deviation principle for the empirical pair measure and the empirical neighbourhood measure.

Theorem 2.1. *Suppose that X is a coloured random graph with colour law μ and connection probabilities $p_n\colon\mathcal{X}\times\mathcal{X}\to[0,1]$ satisfying $np_n(a,b)\to C(a,b)$ for some symmetric function $C\colon\mathcal{X}\times\mathcal{X}\to[0,\infty)$ not identical to zero. Then, as $n\to\infty$, the pair $(L^2,\,M)$ satisfies a large deviation principle in $\tilde{\mathcal{M}}_*(\mathcal{X}\times\mathcal{X})\times\mathcal{M}(\mathcal{X}\times\mathcal{N}(\mathcal{X}))$ with good rate function*

$$J(\varpi,\nu)=\begin{cases}H(\nu\,\|\,Q)+H(\nu_1\,\|\,\mu)+\frac{1}{2}\,\mathfrak{H}_C(\varpi\,\|\,\nu_1) & \textit{if } (\varpi,\nu)\ \textit{sub-consistent,}\\ \infty & \textit{otherwise.}\end{cases}$$

Remark 1 The rate function can be interpreted as follows: $J(\varpi\,,\nu)$ represents the cost of obtaining an empirical pair measure ϖ and an empirical neighbourhood measure ν. This cost is divided into three sub-costs:

(i) $H(\nu_1\,\|\,\mu)$ represents the cost of obtaining the empirical colour measure ν_1, this cost is nonnegative and vanishes iff $\nu_1=\mu$,

(ii) $\frac{1}{2}\mathfrak{H}_C(\varpi\,\|\,\nu_1)$ represent the cost of obtaining an empirical pair measure ϖ if the empirical colour measure is ν_1, again this cost is nonnegative and vanishes iff $\varpi=C\,\nu_1\otimes\nu_1$,

(iii) $H(\nu\,\|\,Q)$ represents the cost of obtaining an empirical neighbourhood measure ν if the empirical colour measure is ν_1 and the empirical pair measure is ϖ, this cost is nonnegative and vanishes iff $\nu=Q$.

Consequently, $J(\varpi\,,\nu)$ is nonnegative and vanishes if and only if $\varpi=C\,\mu\otimes\mu$ and

$$\nu(a,\ell)=\mu(a)\prod_{b\in\mathcal{X}}e^{-C(a,b)\mu(b)}\,\frac{(C(a,b)\mu(b))^{\ell(b)}}{\ell(b)!},\qquad\text{for all } (a,\ell)\in\mathcal{X}\times\mathcal{N}(\mathcal{X}).$$

This is the law of a pair (a,ℓ) where a is distributed according to μ and, given the value of a, the random variables $\ell(b)$ are independently Poisson distributed with parameter $C(a,b)\mu(b)$. Hence, as $n\to\infty$, the random measure $M(X)$ converges in probability to this law.

Remark 2 Our large deviation principle implies individual large deviation principles for the measures L^2 and M by contraction, see [9, Theorem 4.2.1]. Note that, by the discontinuity of Φ, the functional relationship $L^2 = \Phi_2(M)$ may break down in the limit, and hence the rate function may be finite on pairs which are not consistent. We have not been able to extend the large deviation principle to a *stronger topology* in which Φ is continuous, as this leads to considerable compactness problems, see [11, 8] for discussions of some of the problems and opportunities arising when extending large deviation principles to stronger topologies.

As usual, the *degree distribution* $D \in \mathcal{M}(\mathbb{N} \cup \{0\})$ of a graph with empirical neighbourhood measure M is defined by

$$D(k) = \sum_{a \in \mathcal{X}} \sum_{\ell \in \mathcal{N}(\mathcal{X})} \delta_k\big(\textstyle\sum_{b \in \mathcal{X}} \ell(b)\big)\, M(a, \ell), \qquad \text{for } k \in \mathbb{N} \cup \{0\},$$

i.e. $D(k)$ is the proportion of vertices in the graph with degree k. As the degree distribution D is a continuous function of M, Theorem 2.1 and the contraction principle imply a large deviation principle for D. For a classical Erdős-Rényi graph the rate function takes on a particularly simple form (see Section 6 for details).

Corollary 2.2. *Suppose D is the degree distribution of an Erdős-Rényi graph with connection probability $p_n \in [0, 1]$ satisfying $np_n \to c \in (0, \infty)$. Then D satisfies a large deviation principle, as $n \to \infty$, in the space $\mathcal{M}(\mathbb{N} \cup \{0\})$ with good rate function*

$$\delta(d) = \begin{cases} \frac{1}{2}\, x \, \log\big(\frac{x}{c}\big) - \frac{1}{2} x + \frac{c}{2} + H(d \,\|\, q_x), & \text{if } \langle d \rangle \le c, \\ \frac{1}{2}\, \langle d \rangle \, \log\big(\frac{\langle d \rangle}{c}\big) - \frac{1}{2}\, \langle d \rangle + \frac{c}{2} + H(d \,\|\, q_{\langle d \rangle}), & \text{if } c < \langle d \rangle < \infty, \\ \infty & \text{if } \langle d \rangle = \infty, \end{cases} \tag{2.3}$$

where, in the case $\langle d \rangle \le c$, the value $x = x(d)$ is the unique solution of

$$x = c e^{-2\left(1 - \frac{\langle d \rangle}{x}\right)},$$

and where q_λ is a Poisson distribution with parameter λ and $\langle d \rangle := \sum_{m=0}^{\infty} m d(m)$.

Remark 3 On probability measures d with mean c the rate simplifies to the relative entropy of d with respect to the Poisson distribution of the same mean. In [5, Theorem 7.1] a large deviation principle for the degree distribution is formulated for this situation, albeit with a rather implicitly defined rate function. Moreover, the proof given there contains a serious gap: The exponential equivalence stated in [5, Lemma 7.2] is not proved there and we conjecture that it does not hold.

Remark 4 O'Connell [7] provides further large deviation principles for sparse Erdős-Rényi graphs. His focus is on the size of the biggest component, and he also studies the number of isolated vertices. A large deviation principle for the latter is also a consequence of our corollary.

We now state the two large deviation results, Theorems 2.3 (b) and 2.5, which are the main ingredients for our proof of Theorem 2.1, but are also of independent interest. The first of these is a joint large deviation principle for the empirical colour measure L^1 and the empirical pair measure L^2, the second a large deviation principle for the empirical neighbourhood measure M given L^1 and L^2.

For any $n \in \mathbb{N}$ we define

$$\mathcal{M}_n(\mathcal{X}) := \big\{\omega \in \mathcal{M}(\mathcal{X}) \,:\, n\omega(a) \in \mathbb{N} \text{ for all } a \in \mathcal{X}\big\},$$
$$\tilde{\mathcal{M}}_{*,n}(\mathcal{X} \times \mathcal{X}) := \big\{\varpi \in \tilde{\mathcal{M}}_*(\mathcal{X} \times \mathcal{X}) \,:\, \tfrac{n}{1 + \mathbb{1}\{a=b\}}\, \varpi(a, b) \in \mathbb{N} \text{ for all } a, b \in \mathcal{X}\big\}\,,$$

Theorem 2.3. *Suppose that X is a coloured random graph with colour law μ and connection probabilities satisfying $np_n(a,b) \to C(a,b)$ for some symmetric function $C\colon \mathcal{X} \times \mathcal{X} \to [0,\infty)$ not identical to zero.*

(a) *Suppose the sequence $\omega_n \in \mathcal{M}_n(\mathcal{X})$ converges to a limit $\omega \in \mathcal{M}(\mathcal{X})$. Then, as $n \to \infty$, conditional on the event $\{L^1 = \omega_n\}$ the empirical pair measure L^2 of X satisfies a large deviation principle on the space $\tilde{\mathcal{M}}_*(\mathcal{X} \times \mathcal{X})$ with good rate function*

$$I_\omega(\varpi) = \tfrac{1}{2}\mathfrak{H}_C(\varpi \,\|\, \omega)\,. \tag{2.4}$$

(b) *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*

$$I(\omega, \varpi) = H(\omega \,\|\, \mu) + \tfrac{1}{2}\mathfrak{H}_C(\varpi \,\|\, \omega)\,. \tag{2.5}$$

Example 1 We look at the Erdős-Rényi graph with connection probabilities p_n satisfying $np_n \to c \in (0,\infty)$ and study the random partition function for the *Ising model* on the graph, which is defined as

$$Z(\beta) := \sum_{\eta \in \{-1,+1\}^V} \exp\Big(\beta \sum_{(u,v)\in E} \eta(u)\eta(v)\Big) \text{ for the inverse temperature } \beta > 0.$$

Denoting by E expectation with respect to the graph, we note that

$$\mathsf{E}Z(\beta) = 2^n \mathbb{E} \exp\Big(n\,\tfrac{\beta}{2} \int xy\, L^2(dx\,dy)\Big)\,,$$

where $\mathbb{E}$ is expectation with respect to the graph randomly coloured using independent colours chosen uniformly from $\mathcal{X} = \{-1,1\}$. Then Varadhan's lemma, see e.g. [14, III.3], Theorem 2.3 (b), gives

$$\begin{aligned} &\lim_{n\to\infty} \tfrac{1}{n} \log \mathsf{E}Z(\beta) \\ &\quad = \log 2 + \sup\Big\{ \tfrac{\beta}{2} \int xy\, \varpi(dx\,dy) - I(\omega,\varpi) \,:\, \omega \in \mathcal{M}(\mathcal{X}),\ \varpi \in \mathcal{M}_*(\{-1,1\}\times\{-1,1\}) \Big\} \\ &\quad = \sup\Big\{ \tfrac{\beta}{2}\big(\varpi(\Delta) - \varpi(\Delta^c)\big) - x\log(x) - (1-x)\log(1-x) - \tfrac{1}{2}\big(H(\varpi \,\|\, \omega_x) + c - \|\varpi\|\big) \Big\}, \end{aligned} \tag{2.6}$$

where Δ is the diagonal in $\{-1,1\}\times\{-1,1\}$, and the supremum is over all $x \in [0,1]$ and $\varpi \in \mathcal{M}_*(\{-1,1\}\times\{-1,1\})$, and the measure $\omega_x \in \tilde{\mathcal{M}}_*(\{-1,1\}\times\{-1,1\})$ is defined by

$$\omega_x(i,j) = c x^{(2+i+j)/2}(1-x)^{(2-i-j)/2} \text{ for } i,j \in \{-1,1\}\,.$$

Note that the last expression in (2.6) is an optimisation problem in only four real variables.

We obtain from Theorem 2.3 (b) the following corollary (see Section 6 for details).

Corollary 2.4. *Suppose that X is a coloured random graph with colour law μ and connection probabilities satisfying $np_n(a,b) \to C(a,b)$ for some symmetric function $C\colon \mathcal{X}\times\mathcal{X} \to [0,\infty)$ not identical to zero. Then, as $n \to \infty$, the number of edges per vertex $|E|/n$ satisfies a large deviation principle in $[0,\infty)$ with good rate function*

$$\zeta(x) = x\log x - x + \inf_{y>0} \big\{\psi(y) - x\log(\tfrac{1}{2}y) + \tfrac{1}{2}y\big\},$$

where $\psi(y) = \inf H(\omega \,\|\, \mu)$ over all probability vectors ω with $\omega^T C \omega = y$.

Remark 5 In the Erdős-Rényi case $C(a,b) = c$ one obtains $\psi(y) = 0$ for $y = c$, and $\psi(y) = \infty$ otherwise. Hence $\zeta(x) = x\log x - x - x\log(\frac{c}{2}) + \frac{c}{2}$, which is the Cramér rate function for the Poisson distribution with parameter $\frac{c}{2}$. In [2] a large deviation principle for $|E|/n^2$ is proved for coloured random graphs with *fixed* connection probabilities.

For a given $\varpi \in \tilde{\mathcal{M}}_*(\mathcal{X}\times\mathcal{X})$ and $\nu \in \mathcal{M}(\mathcal{X}\times\mathcal{N}(\mathcal{X}))$ we recall the definition of the measure $Q \in \mathcal{M}(\mathcal{X}\times\mathcal{N}(\mathcal{X}))$ from (2.2).

Theorem 2.5. *Suppose $(\omega_n, \varpi_n) \in \mathcal{M}_n(\mathcal{X}) \times \tilde{\mathcal{M}}_{*,n}(\mathcal{X}\times\mathcal{X})$ converges to a limit $(\omega,\varpi) \in \mathcal{M}(\mathcal{X}) \times \tilde{\mathcal{M}}_*(\mathcal{X}\times\mathcal{X})$. Let X be a coloured random graph with n vertices conditioned on the event $\{\Phi(M) = (\omega_n, \varpi_n)\}$. Then, as $n\to\infty$, the empirical neighbourhood measure M of X satisfies a large deviation principle in the space $\mathcal{M}(\mathcal{X}\times\mathcal{N}(\mathcal{X}))$ with good rate function*

$$\tilde{J}_{(\omega,\varpi)}(\nu) = \begin{cases} H(\nu \,\|\, Q) & \text{if } (\varpi,\nu) \text{ is sub-consistent and } \nu_1 = \omega, \\ \infty & \text{otherwise.} \end{cases} \tag{2.7}$$

In the remainder of the paper we give the proofs of the results set out so far. Section 3 is devoted to the proof of Theorem 2.3 (a), which uses the Gärtner-Ellis theorem. By contrast, the proof of Theorem 2.5, carried out in Section 4, is based on nontrivial combinatorial arguments combined with a partially randomised approximation procedure. This approximation is the most demanding argument of the paper and requires a fairly sophisticated technique. In Section 5 we first combine Sanov's Theorem [9, Theorem 2.1.10] and Theorem 2.3 (a) to obtain Theorem 2.3 (b), and then Theorem 2.3 (b) and Theorem 2.5 to get Theorem 2.1, using the setup and result of Biggins [1] to 'mix' the large deviation principles. The paper concludes with the proofs of Corollaries 2.2 and 2.4, which are given in Section 6.

3. Proof of Theorem 2.3 (a) by the Gärtner-Ellis theorem

Throughout this section we assume that the sequence $\omega_n \in \mathcal{M}_n(\mathcal{X})$ converges to $\omega \in \mathcal{M}(\mathcal{X})$. Let $\mathbb{P}\{\,\cdot\,|\, L^1 = \omega_n\}$ be the law of coloured random graph X with connection probabilities satisfying $np_n(a,b) \to C(a,b)$ conditioned on the event $\{L^1 = \omega_n\}$. In the next lemma we verify the assumption of the Gärtner-Ellis theorem [9, Theorem 2.3.6]. We denote by $\mathcal{C}_2$ the space of symmetric functions on $\mathcal{X}\times\mathcal{X}$.

Lemma 3.1. *For each $g \in \mathcal{C}_2$, we have*

$$\lim_{n\to\infty} \tfrac{1}{n} \log \mathbb{E}\big[e^{n\langle g,\, L^2\rangle} | L^1 = \omega_n\big] = -\tfrac{1}{2}\,\langle C\omega\otimes\omega,\, (1-e^g)\rangle.$$

Proof. Let $g \in \mathcal{C}_2$. Observe that given the colours $a, b \in \mathcal{X}$ the random variables $nL^2(a,b)$ are binomial with parameters $n^2\omega_n(a)\omega_n(b) - n\omega_n(a)1\!\!1_{\{a=b\}}$ and $p_n(a,b)$, and the variables $nL^2(a,b)$, for $\{a,b\} \subset \mathcal{X}$ are independent. Hence, we have that

$$\mathbb{E}\big[e^{n\langle g,\, L^2\rangle}\big| L^1 = \omega_n\big] = \prod_{\{a,b\}} \Big(1 - p_n(a,b) + p_n(a,b)e^{g(a,b)}\Big)^{n^2\omega_n(a)\omega_n(b) - n\omega_n(a)1\!\!1_{\{a=b\}}}.$$

Now, for any $\varepsilon > 0$ and for large n we have

$$\Big(1 - \tfrac{C(a,b)(1-e^{g(a,b)})+\varepsilon}{n}\Big)^n \le \Big(1 - p_n(a,b) + p_n(a,b)e^{g(a,b)}\Big)^n \le \Big(1 - \tfrac{C(a,b)(1-e^{g(a,b)})-\varepsilon}{n}\Big)^n.$$

Using Euler's formula and taking the product over all $\{a,b\} \subset \mathcal{X}$ we obtain

$$-\tfrac{1}{2}\,\langle C\omega\otimes\omega,\, (1-e^g)\rangle - \varepsilon \le \lim_{n\to\infty} \tfrac{1}{n} \log \mathbb{E}\big[e^{n\langle g,\, L^2\rangle} | L^1 = \omega_n\big] \le -\tfrac{1}{2}\,\langle C\omega\otimes\omega,\, (1-e^g)\rangle + \varepsilon,$$

and the result follows as $\varepsilon > 0$ was arbitrary. ∎

Now, by the Gärtner-Ellis theorem, under $\mathbb{P}\{\cdot \mid L^1 = \omega_n\}$ the measure L^2 obeys a large deviation principle on $\tilde{\mathcal{M}}_*(\mathcal{X}\times\mathcal{X})$ with rate function $\hat{I}_\omega(\varpi) = \frac{1}{2}\sup_{g\in\mathcal{C}_2}\{\langle\varpi, g\rangle + \langle C\omega\otimes\omega,\, (1-e^g)\rangle\}$.

Next, we express the rate function in term of relative entropies and consequently show that it is a good rate function. Recall the definition of the function I_ω from Theorem 2.3 (a).

Lemma 3.2.

(i) $\hat{I}_\omega(\varpi) = I_\omega(\varpi)$, *for any* $\varpi \in \tilde{\mathcal{M}}_*(\mathcal{X}\times\mathcal{X})$,
(ii) $I_\omega(\varpi)$ *is a good rate function and*
(iii) $\mathfrak{H}_C(\varpi \,\|\, \omega) \ge 0$ *with equality if and only if* $\varpi = C\omega\otimes\omega$.

Proof. (i) Suppose that $\varpi \not\ll C\omega\otimes\omega$. Then, there exists $a_0, b_0 \in \mathcal{X}$ with $C\omega\otimes\omega(a_0,b_0) = 0$ and $\varpi(a_0,b_0) > 0$. Define $\hat{g}\colon \mathcal{X}\times\mathcal{X}\to\mathbb{R}$ by

$$\hat{g}(a,b) = \log\big[K(\mathbb{1}_{(a_0,b_0)}(a,b) + \mathbb{1}_{(b_0,a_0)}(a,b)) + 1\big], \text{ for } a,b\in\mathcal{X} \text{ and } K>0.$$

For this choice of $\hat{g}$ we have

$$\tfrac{1}{2}\langle\varpi,\hat{g}\rangle + \tfrac{1}{2}\langle C\,\omega\otimes\omega, 1-e^{-\hat{g}}\rangle \ge \tfrac{1}{2}\log(K+1)\varpi(a_0,b_0) \to \infty, \text{ for } K\uparrow\infty.$$

Now suppose that $\varpi \ll C\omega\otimes\omega$. We have

$$\hat{I}_\omega(\varpi) = \tfrac{1}{2}\|C\,\omega\otimes\omega\| + \tfrac{1}{2}\sup_{g\in\mathcal{C}_2}\big\{\langle\varpi, g\rangle - \langle C\,\omega\otimes\omega, e^g\rangle\big\}.$$

By the substitution $h = e^g\,\frac{C\omega\otimes\omega}{\varpi}$ the supremum equals

$$\begin{aligned}\sup_{\substack{h\in\mathcal{C}_2\\ h\ge 0}}\big\langle\varpi, \log\big(h\frac{\varpi}{C\omega\otimes\omega}\big) - h\big\rangle &= \sup_{\substack{h\in\mathcal{C}_2\\ h\ge 0}}\langle\varpi, \log h - h\rangle + \big\langle\varpi, \log\frac{\varpi}{C\omega\otimes\omega}\big)\big\rangle\\ &= -\|\varpi\| + H(\varpi\,\|\,C\omega\otimes\omega),\end{aligned}$$

where we have used $\sup_{x>0}\log x - x = -1$ in the last step. This yields that $\hat{I}_\omega(\varpi) = I_\omega(\varpi)$.

(ii) Recall that $I_\omega = \hat{I}_\omega$ is a rate function. Moreover, for all $\alpha<\infty$, the level sets $\{\varpi\in\tilde{\mathcal{M}}_*(\mathcal{X}\times\mathcal{X})\colon \frac{1}{2}\,\mathfrak{H}_C(\varpi\,\|\,\omega)\le\alpha\}$ are bounded and therefore compact, so I_ω is a good rate function.

(iii) Consider the nonnegative function $\xi(x) = x\log x - x + 1$, for $x>0$, $\xi(0)=1$, which has its only root at $x=1$. Note that

$$\mathfrak{H}_C(\varpi\,\|\,\omega) = \begin{cases} \int \xi\circ g\; dC\omega\otimes\omega & \text{if } g := \frac{d\varpi}{dC\omega\otimes\omega} \ge 0 \text{ exists,}\\ \infty & \text{otherwise.}\end{cases} \tag{3.1}$$

Hence $\mathfrak{H}_C(\varpi\,\|\,\omega)\ge 0$, and, if $\varpi = C\omega\otimes\omega$, then $\xi(\frac{d\varpi}{dC\omega\otimes\omega}) = \xi(1) = 0$ and so $\mathfrak{H}_C(C\omega\otimes\omega\,\|\,\omega) = 0$. Conversely, if $\mathfrak{H}_C(\varpi\,\|\,\omega) = 0$, then $\varpi(a,b)>0$ implies $C\omega\otimes\omega(a,b)>0$, which then implies $\xi\circ g(a,b) = 0$ and further $g(a,b)=1$. Hence $\varpi = C\omega\otimes\omega$, which completes the proof of (iii). ■

4. Proof of Theorem 2.5 by the method of types

Throughout the proof we may assume that $\omega(a)>0$ for all $a\in\mathcal{X}$. It is easy to see that the law of the randomly coloured graph conditioned to have empirical colour measure ω_n and empirical pair measure ϖ_n,

$$\mathbb{P}_{(\omega_n,\varpi_n)} := \mathbb{P}\{\,\cdot\mid\Phi(M) = (\omega_n,\varpi_n)\},$$

can be described in the following manner:

- Assign colours to the vertices by sampling without replacement from the collection of n colours, which contains any colour $a \in \mathcal{X}$ exactly $n\omega_n(a)$ times;
- for every unordered pair $\{a,b\}$ of colours create exactly $n(a,b)$ edges by sampling without replacement from the pool of possible edges connecting vertices of colour a and b, where

$$n(a,b) := \begin{cases} n\,\varpi_n(a,b) & \text{if } a \neq b, \\ \frac{n}{2}\,\varpi_n(a,b) & \text{if } a = b\,. \end{cases} \tag{4.1}$$

We would like to reduce the calculation of probabilities to the counting of suitable configurations. To this end we introduce a numbering system, which specifies, for each $\{a,b\}$, the order in which edges are drawn in the second step. More precisely, the edge-number k is attached to both vertices connecting the k^{th} edge. Note that the total number of edge-numbers attached to every vertex corresponds to the degree of the vertex in the graph. All permitted numberings are equally probable.

Denote by $Y_j^{\{a,b\}}$ be the j^{th} edge drawn in the process of connecting vertices of colours $\{a,b\}$. Let $\mathcal{A}_n(\omega_n,\varpi_n)$ be the set of all possible configurations

$$\Big(\big(X(v)\,:\,v\in V\big);\,\big(Y_k^{\{a,b\}}\,:\,k=1,\dots,n(a,b);\,\{a,b\}\subset\mathcal{X}\big)\Big),$$

and let $\mathcal{B}_n(\omega_n,\varpi_n)$ be the set of all coloured graphs x with $L^1(x)=\omega_n$ and $L^2(x)=\varpi_n$. Define $\Psi\colon \mathcal{A}_n(\omega_n,\varpi_n)\to\mathcal{B}_n(\omega_n,\varpi_n)$ as the canonical mapping which associates the coloured graph to any configuration, i.e. 'forgets' the numbering of the edges. Finally, define

$$\mathcal{K}^{(n)}(\omega_n,\varpi_n) := \big\{M(x) \text{ for some } x\in\mathcal{B}_n(\omega_n,\varpi_n)\big\}$$

to be the set of all empirical neighbourhood measures $M(x)$ arising from coloured graphs x with n vertices with $\Phi(M(x))=(\omega_n,\varpi_n)$. For any $\nu_n\in\mathcal{K}^{(n)}(\omega_n,\varpi_n)$ we have

$$\mathbb{P}\big\{M=\nu_n\,\big|\,\Phi(M)=(\omega_n,\varpi_n)\big\} = \frac{\sharp\big\{\tilde{x}\in\mathcal{A}_n(\omega_n,\varpi_n)\,:\,M\circ\Psi(\tilde{x})=\nu_n\big\}}{\sharp\big\{\tilde{x}\in\mathcal{A}_n(\omega_n,\varpi_n)\big\}}\,. \tag{4.2}$$

In our proofs we use the following form of Stirling's formula, see [12, page 54]: For all $n\in\mathbb{N}$,

$$n^n\,e^{-n} \le n! \le (2\pi n)^{\frac12}\,n^n\,e^{-n+1/(12n)}\,.$$

4.1 A bound on the number of empirical neighbourhood measures

In this section we provide an upper bound on the number of measures in $\mathcal{K}^{(n)}(\omega_n,\varpi_n)$. We write m for the number of elements in $\mathcal{X}$.

Lemma 4.1. *There exists $\vartheta>0$, depending on m such that, if $\omega_n\in\mathcal{M}_n(\mathcal{X})$ and $\varpi_n\in\tilde{\mathcal{M}}_{*,n}(\mathcal{X}\times\mathcal{X})$, then*

$$\sharp\mathcal{K}^{(n)}(\omega_n,\varpi_n) \le \exp\Big[\vartheta\,(\log n)\,(n\|\varpi_n\|)^{\frac{2m-1}{2m}}\Big].$$

The proof is based on counting integer partitions of vectors. To fix some notation, let $\mathfrak{I}_m=\big(\mathbb{N}\cup\{0\}\big)^m$ be the collection of (nonnegative) integer vectors of length m. For any $\ell\in\mathfrak{I}_m$ we denote by $\|\ell\|$ its magnitude, i.e. the sum of its entries.

We introduce an ordering $\succcurlyeq$ on $\mathfrak{I}_m$ such that, for any vectors

$$\ell_1=(\ell_1^{(1)},\dots,\ell_1^{(m)}) \text{ and } \ell_2=(\ell_2^{(1)},\dots,\ell_2^{(m)}),$$

we write $\ell_1\succcurlyeq\ell_2$ if either

(i) $\|\ell_1\|>\|\ell_2\|$, *or*
(ii) $\|\ell_1\|=\|\ell_2\|$ and there is $j\in\{1,\dots,m\}$ with $\ell_1^{(k)}=\ell_2^{(k)}$, for all $k<j$, and $\ell_1^{(j)}>\ell_2^{(j)}$, *or*
(iii) $\ell_1=\ell_2$.

A collection $(\ell_1, \ldots, \ell_k)$ of elements in $\mathfrak{I}_m$ is an *integer partition* of the vector $\ell \in \mathfrak{I}_m$, if

$$\ell_1 \succcurlyeq \ldots \succcurlyeq \ell_k \neq 0 \qquad \text{and} \qquad \ell_1 + \ldots + \ell_k = \ell.$$

Any integer partition of a vector $\ell \in \mathfrak{I}_m$ induces an integer partition $\|\ell_1\|, \ldots, \|\ell_k\|$ of its magnitude $\|\ell\|$, which we call its *sum-partition.* We denote by $\mathcal{P}_m(\ell)$ the set of integer partitions of ℓ.

Lemma 4.2. *There exists $\vartheta > 0$, which depends on m such that, for any $\ell \in \mathfrak{I}_m$ of magnitude n,*

$$\sharp\mathcal{P}_m(\ell) \leq \exp\Big[\vartheta\,(\log n)\, n^{\frac{2m-1}{2m}}\Big].$$

Proof. Let $\ell \in \mathfrak{I}_m$ be a vector of magnitude n and $(\ell_1, \ldots, \ell_k)$ be an integer partition of ℓ. We relabel the partition as $(\mathfrak{m}_{1,1}, \ldots, \mathfrak{m}_{1,k_1}; \mathfrak{m}_{2,1}, \ldots, \mathfrak{m}_{2,k_2}; \ldots; \mathfrak{m}_{r,1}, \ldots, \mathfrak{m}_{r,k_r})$ such that all vectors in the same block (indicated by the first subscript) have the same magnitude, which we denote $y_1, \ldots, y_r$, and such that $y_1 > \cdots > y_r > 0$. Note that for the block sizes we have $k_1 + \cdots + k_r = k$ and that $k_1 y_1 + \cdots + k_r y_r = n$.

For a moment, look at a fixed block $\mathfrak{m}_{j,1}, \ldots, \mathfrak{m}_{j,k_j}$. It is easy to see that the number of integer vectors of length m and magnitude y_j is given by

$$b(y_j, m) := \binom{y_j + m - 1}{m - 1} \leq c(m)\, y_j^{m-1}\,.$$

Writing $\mathfrak{m}_{j,0}$ for the largest and $\mathfrak{m}_{j,k_j+1}$ for the smallest of these vectors in the ordering $\succcurlyeq$, we note that

$$p\colon \{0, \ldots, k_j + 1\} \to \big\{\mathfrak{m} \in \mathfrak{I}_m \colon \|\mathfrak{m}\| = y_j\big\}, \qquad p(i) = \mathfrak{m}_{j,i},$$

is a non-increasing path of length $k_j + 2$ into an ordered set of size $b(y_j, m)$, which connects the smallest to the largest element. The number of such paths is easily seen to be

$$\binom{b(y_j, m) + k_j}{k_j}\,.$$

Therefore, the number of integer partitions of ℓ with given sum-partition $(y_1, \overset{k_1}{\ldots}, y_1, \ldots, y_r, \overset{k_r}{\ldots}, y_r)$ is

$$\prod_{j=1}^{r} \binom{b(y_j, m) + k_j}{k_j} \leq \max_{\substack{a_1, \ldots, a_r > 0 \\ \sum a_j = n}} \prod_{j=1}^{r} \Big\{ \max_{\substack{y,k \in \mathbb{N} \\ yk = a_j}} \binom{c(m) y^{m-1} + k}{k} \Big\}.$$

To maximize the binomial coefficient over the set $yk = a_j$, we distinguish between the cases when (i) $a_j \leq c(m) y^m$, (ii) $a_j > c(m) y^m$ and observe that

$$\binom{c(m) y^{m-1} + \frac{a_j}{y}}{\frac{a_j}{y}} \leq \begin{cases} \binom{2c(m) y^{m-1}}{\frac{a_j}{y}} & \text{if } a_j \leq c(m) y^m, \\ \binom{2\frac{a_j}{y}}{c(m) y^{m-1}} & \text{if } a_j > c(m) y^m. \end{cases}$$

Case (i): Using the upper bound $\binom{i}{r} \leq \big(\frac{ie}{r}\big)^r$, for $r, i \in \mathbb{N}$ with $r \leq i$ and the inequality $(\frac{a_j}{c(m)})^{1/m} \leq y \leq a_j \leq n$ we obtain, for some constants $C_0 = C_0(m)$, $C_1 = C_1(m)$,

$$\binom{2c(m) y^{m-1}}{a_j/y} \leq \Big(\frac{2c(m) y^{m-1} e}{a_j/y}\Big)^{a_j/y} \leq \exp\big(C_0\,(a_j/y)\,\log n\big) \leq \exp\big((\log n)\, C_1\, a_j^{\frac{m-1}{m}}\big).$$

Case (ii): The same upper bound for binomial coefficients and $1 \le y \le (\frac{a_j}{c})^{1/m} \le a_j \le n$ yield for some constant $C_2 = C_2(m)$,

$$\binom{2a_j/y}{c(m)y^{m-1}} \le \left(\frac{2(a_j/y)e}{c(m)y^{m-1}}\right)^{c(m)y^{m-1}} \le \exp\Big((\log n)\, C_2\, a_j^{\frac{m-1}{m}}\Big).$$

From these cases, we have for some $C = C(m) > 0$, the upper bound

$$\prod_{j=1}^{r} \binom{b(y_j, m) + k_j}{k_j} \le \max_{\substack{a_1,\dots,a_r>0\\ \sum a_j = n}} \prod_{j=1}^{r} \exp((\log n) C\, a_j^{\frac{m-1}{m}}),$$

which is estimated further (using Hölder's inequality) by

$$\exp\Big((\log n)\, C \sum_{j=1}^{r} a_j^{\frac{m-1}{m}}\Big) \le \exp\Big((\log n)\, C\, r^{\frac{1}{m}} \Big(\sum_{j=1}^{r} a_j\Big)^{\frac{m-1}{m}}\Big).$$

We observe that all y_j are different, positive and that their sum is not greater than n, so we have that

$$r^2/2 \le 1 + \ldots + r \le y_1 + \ldots + y_r \le n.$$

Recalling that $a_1 + \cdots + a_r = n$, our upper bound becomes

$$\exp\Big((\log n)\, C\, (2n)^{\frac{1}{2m}}\, n^{\frac{m-1}{m}}\Big) = \exp\Big(\tfrac{\vartheta}{2}\,(\log n)\, n^{\frac{2m-1}{2m}}\Big),$$

for some $\vartheta = \vartheta(m) > 0$. Note that from our argument so far one can easily recover the well-known fact that the number of integer partitions of n is bounded by $e^{(\vartheta/2)\sqrt{n}}$, the full asymptotics being discovered by Hardy and Ramanujan in 1918. Combining this with the upper bound for the number of integer partitions with a given sum-partition, we obtain the claim. ■

Proof of Lemma 4.1 Suppose $\omega_n \in \mathcal{M}_n(\mathcal{X})$ and $\varpi_n \in \tilde{\mathcal{M}}_{*,n}(\mathcal{X}\times\mathcal{X})$. For $a \in \mathcal{X}$, we look at the mappings

$$\Phi_a \colon \mathcal{K}^{(n)}(\omega_n, \varpi_n) \ni \frac{1}{n}\sum_{v\in V} \delta_{(X(v),L(v))} \mapsto \big(L_1^a, \ldots, L_{n\omega(a)}^a\big),$$

where $(L_1^a, \ldots, L_{n\omega(a)}^a)$ is the ordering of the vectors $L(v)$, for all $v \in V$ with $X(v) = a$, and thus constitutes an integer partition of the vector $(n\varpi_n(a,b) \colon b \in \mathcal{X})$, which has magnitude $n\sum_b \varpi_n(a,b)$. The combined mapping $\Phi = (\Phi_a \colon a \in \mathcal{X})$ is injective, and therefore, by Lemma 4.2,

$$\begin{aligned}\sharp\mathcal{K}^{(n)}(\omega_n, \varpi_n) &\le \exp\Big[\vartheta \sum_{a\in\mathcal{X}} \log\big(n\sum_{b\in\mathcal{X}}\varpi_n(a,b)\big)\big(n\sum_{b\in\mathcal{X}}\varpi_n(a,b)\big)^{\frac{2m-1}{2m}}\Big]\\ &\le \exp\Big[\vartheta\, m\, \log\big(n\|\varpi_n\|\big)\big(n\|\varpi_n\|\big)^{\frac{2m-1}{2m}}\Big],\end{aligned}$$

where we have used the fact that $\sum_b \varpi_n(a,b) \le \|\varpi_n\|$ in the last step. ■

4.2 Proof of the upper bound in Theorem 2.5.

We are now ready to prove an upper bound for the large deviation probability in Theorem 2.5.

Lemma 4.3. *For any sequence (ν_n) with $\nu_n \in \mathcal{K}^{(n)}(\omega_n, \varpi_n)$ we have*

$$\mathbb{P}\big\{M = \nu_n \,\big|\, \Phi(M) = (\omega_n, \varpi_n)\big\} \le \exp\big(-nH(\nu_n \,\|\, Q_n) + \varepsilon_1^{(n)}(\nu_n)\big),$$

where

$$Q_n(a\,,\,\ell)=\omega_n(a)\prod_{b\in\mathcal{X}}\frac{e^{-\varpi_n(a,b)/\omega_n(a)}[\varpi_n(a,b)/\omega_n(a)]^{\ell(b)}}{\ell(b)!},\quad \textit{for } \ell\in\mathcal{N}(\mathcal{X}),$$

and

$$\lim_{n\uparrow\infty}\frac{1}{n}\sup_{\nu_n\in\mathcal{K}^{(n)}(\omega_n,\varpi_n)}\varepsilon_1^{(n)}(\nu_n)=0\,.$$

Proof. The proof of this lemma is based on the method of types, see [9, Chapter 2]. Recall from (4.2) that, for any $\nu_n\in\mathcal{K}^{(n)}(\omega_n,\varpi_n)$, we have

$$\mathbb{P}\big\{M=\nu_n\,\big|\,\Phi(M)=(\omega_n,\varpi_n)\big\}=\frac{\sharp\big\{\tilde{x}\in\mathcal{A}_n(\omega_n,\varpi_n)\,:\,M\circ\Psi(\tilde{x})=\nu_n\big\}}{\sharp\big\{\tilde{x}\in\mathcal{A}_n(\omega_n,\varpi_n)\big\}}\,.$$

Now, by elementary counting, the denominator on the right side of (4.2) is

$$\binom{n}{n\omega_n(a),\,a\in\mathcal{X}}\prod_{\{a,b\}}\prod_{k=1}^{n(a,b)}\Big(\frac{n^2\omega_n(a)\omega_n(b)-n\omega_n(a)\mathbb{1}_{\{a=b\}}}{1+\mathbb{1}_{\{a=b\}}}-(k-1)\Big)\,. \tag{4.3}$$

For a given empirical neighbourhood measure $\nu_n\in\mathcal{K}^{(n)}(\omega_n,\varpi_n)$ the numerator is probably too tricky to find explicitly. However, an easy upper bound is

$$\binom{n}{n\nu_n(a,\ell),\,a\in\mathcal{X}\,,\ell\in\mathcal{N}(\mathcal{X})}\,2^{-\frac{n}{2}\varpi_n(\Delta)}\prod_{(a,b)}\binom{n\varpi_n(a,b)}{\ell_a^{(j)}(b),\,j=1,\dots,n\omega_n(a)}\,, \tag{4.4}$$

where $\ell_a^{(j)}(b)$, $j=1,\dots,n\omega_n(a)$ are any enumeration of the family containing each $\ell(b)$ with multiplicity $n\nu_n(a,\ell)$. This can be seen from the following construction: First allocate to each vertex some $(a,\ell)\in\mathcal{X}\times\mathcal{N}(\mathcal{X})$ in such a way that every vector (a,ℓ) is allocated $n\nu_n(a,\ell)$ times. The first binomial coefficient in (4.4) represents the number of possible ways to do this. For any $(a,b)\in\mathcal{X}\times\mathcal{X}$ distribute the numbers in $\{1,\dots,n\varpi_n(a,b)\}$ among the vertices with colour a so that a vertex carrying vector (a,ℓ) gets exactly $\ell(b)$ numbers. Once this is done for both (a,b) and (b,a), each vertex of colour a or b carries a set of numbers; if $a\neq b$ each number in $\{1,\dots n\varpi_n(a,b)\}$ occurs exactly twice in total, if $a=b$ it occurs exactly once. Next, for $k=1,\dots n(a,b)$, if $a\neq b$ draw the kth edge between the two vertices of colour a and b carrying number k, if $a=b$ draw the kth edge between the vertices with number k and $2k$. The remaining factor in (4.4) represents the number of possible ways to do this, with the power of two discounting the fact that for edges connecting vertices of the same colour two numbering schemes lead to the same configuration. By this construction, every element $\tilde{x}\in\mathcal{A}_n(\omega_n,\varpi_n)$ with $M\circ\Psi(\tilde{x})=\nu_n$ has been constructed exactly once, but also some graphs with loops or multiple edges can occur, so that (4.4) is an upper bound for the numerator in (4.2).

Combining (4.2), (4.3), and (4.4) we get

$$\begin{aligned}\mathbb{P}\big\{M=\nu_n\,\big|\,\Phi(M)=(\omega_n,\varpi_n)\big\}&\\ \le\prod_{a\in\mathcal{X}}\binom{n\omega_n(a)}{n\nu_n(a,\ell),\,\ell\in\mathcal{N}(\mathcal{X})}&\prod_{(a,b)}\binom{n\varpi_n(a,b)}{\ell_a^{(j)}(b),\,j=1,\dots,n\omega_n(a)}\\ \times\,2^{-\frac{n}{2}\varpi_n(\Delta)}&\prod_{\{a,b\}}\prod_{k=1}^{n(a,b)}\big(\tfrac{n^2\omega_n(a)\omega_n(b)-n\omega_n(a)\mathbb{1}_{\{a=b\}}}{1+\mathbb{1}_{\{a=b\}}}-(k-1)\big)^{-1}\,.\end{aligned} \tag{4.5}$$

It remains to analyse the asymptotics of this upper bound. Using Stirling's formula, we obtain

$$\prod_{a\in\mathcal{X}}\binom{n\omega_n(a)}{n\nu_n(a,\ell),\,\ell\in\mathcal{N}(\mathcal{X})}\le\exp\Big(n\sum_a\omega_n(a)\log\omega_n(a)-n\sum_{(a,\ell)}\nu_n(a,\ell)\log\nu_n(a,\ell)\Big)$$
$$\times\exp\Big(\tfrac{m}{2}\log(2\pi n)+\tfrac{1}{n}\sum_a\tfrac{1}{12\omega_n(a)}\Big).$$

We observe that

$$\prod_{j=1}^{n\omega_n(a)}\big(\ell_a^{(j)}(b)\big)!=\exp\Big(n\sum_\ell\log\big(\ell(b)!\big)\,\nu_n(a,\,\ell)\Big),$$

and hence

$$\binom{n\varpi_n(a,b)}{\ell_a^{(j)}(b),\,j\le n\omega_n(a)}\le\exp\Big(-n\sum_\ell\log\big(\ell(b)!\big)\,\nu_n(a,\ell)+n\varpi_n(a,b)\log\big(n\varpi_n(a,b)\big)-n\varpi_n(a,b)\Big)$$
$$\times\exp\Big(\tfrac{1}{12n\varpi_n(a,b)}+\tfrac{1}{2}\log(2\pi n)\Big).$$

Next, we obtain,

$$\prod_{k=1}^{n(a,b)}\big(\tfrac{n^2\omega_n(a)\omega_n(b)-n\omega_n(a)\mathbb{1}_{\{a=b\}}}{1+\mathbb{1}_{\{a=b\}}}-(k-1)\big)\ge\exp\Big(n(a,b)\,\log\big(\tfrac{n^2\omega_n(a)\omega_n(b)}{1+\mathbb{1}_{\{a=b\}}}\big)\Big)$$
$$\times\exp\Big(n(a,b)\,\log\big(1-\tfrac{\mathbb{1}\{a=b\}}{2n\omega_n(a)}-\tfrac{2n(a,b)}{n^2\omega_n(a)\omega_n(b)}\big)\Big)\,.$$

Putting everything together and denoting by $H(\omega)=-\sum_{y\in\mathcal{Y}}\omega(y)\log\omega(y)$ the entropy of a measure $\omega\in\mathcal{M}(\mathcal{Y})$, we get

$$\begin{aligned}&\mathbb{P}\big\{M=\nu_n\,\big|\,\Phi(M)=(\omega_n,\varpi_n)\big\}\\&\le\exp\Big(-nH(\omega_n)+nH(\nu_n)-n\sum_{(a,b)}\sum_\ell(\log\ell(b)!)\,\nu_n(a,\ell)+n\sum_{(a,b)}\varpi_n(a,b)\log\varpi_n(a,b)\\&\qquad-n\sum_{(a,b)}\varpi_n(a,b)-\tfrac{n}{2}\sum_{(a,b)}\varpi_n(a,b)\,\log\big(\tfrac{\omega_n(a)\omega_n(b)}{1+\mathbb{1}_{\{a=b\}}}\big)-\tfrac{n}{2}\,\varpi_n(\Delta)\log 2+\varepsilon_1^{(n)}\Big),\end{aligned}$$

for a sequence $\varepsilon_1^{(n)}$ which does not depend on ν_n and satisfies $\lim_{n\uparrow\infty}\frac{1}{n}\,\varepsilon_1^{(n)}=0$. To give the right hand side the form as stated in the theorem, we observe that

$$\begin{aligned}&H(\omega_n)-H(\nu_n)+\sum_{(a,b)}\sum_\ell(\log\ell(b)!)\,\nu_n(a,\ell)-\sum_{(a,b)}\varpi_n(a,b)\log\varpi_n(a,b)+\sum_{(a,b)}\varpi_n(a,b)\\&\qquad+\tfrac{1}{2}\sum_{(a,b)}\varpi_n(a,b)\,\log\big(\omega_n(a)\omega_n(b)\big)\\&=\sum_{(a,\ell)}\nu_n(a,\ell)\Big[\log\nu_n(a,\ell)-\log\omega_n(a)-\sum_b\Big(\log\big(\tfrac{\varpi_n(a,b)}{\omega_n(a)}\big)^{\ell(b)}-\tfrac{\varpi_n(a,b)}{\omega_n(a)}-(\log\ell(b)!)\Big)\Big]\\&=\sum_{(a,\ell)}\nu_n(a,\ell)\Big[\log\nu_n(a,\ell)-\log\big(\omega_n(a)\prod_b\tfrac{(\varpi_n(a,b)/\omega_n(a))^{\ell(b)}\exp(-\varpi_n(a,b)/\omega_n(a))}{\ell(b)!}\big)\Big]\\&=H(\nu_n\,\|\,Q_n),\end{aligned}$$

which completes the proof of Lemma 4.3. ■

We can now complete the proof of the upper bound in Theorem 2.5 by combining Lemma 4.1 and Lemma 4.3. Suppose that $\Gamma \subset \mathcal{M}(\mathcal{X} \times \mathcal{N}(\mathcal{X}))$ is a closed set. Then,

$$\mathbb{P}\Big\{M \in \Gamma \,\big|\, \Phi(M) = (\omega_n, \varpi_n)\Big\} = \sum_{\nu_n \in \Gamma \cap \mathcal{K}^{(n)}(\omega_n, \varpi_n)} \mathbb{P}\big\{M = \nu_n \,\big|\, \Phi(M) = (\omega_n, \varpi_n)\big\}$$
$$\leq \sharp\mathcal{K}^{(n)}(\omega_n, \varpi_n) \, \exp\Big(- n \inf_{\nu_n \in \Gamma \cap \mathcal{K}^{(n)}(\omega_n, \varpi_n)} H(\nu_n \,\|\, Q_n) + \sup_{\nu_n \in \mathcal{K}^{(n)}(\omega_n, \varpi_n)} \varepsilon_1^{(n)}(\nu_n)\Big) .$$

We have already seen that $\frac{1}{n} \sup_{\nu_n} \varepsilon_1^{(n)}(\nu_n)$ and $\frac{1}{n} \log \sharp\mathcal{K}^{(n)}(\omega_n, \varpi_n)$ converge to zero. It remains to check that

$$\lim_{n\to\infty} \sup_{\nu_n \in \mathcal{K}^{(n)}(\omega_n, \varpi_n)} \big|H(\nu_n \,\|\, Q_n) - H(\nu_n \,\|\, Q)\big| = 0 \,. \tag{4.6}$$

To do this, we observe that

$$H(\nu_n \,\|\, Q_n) - H(\nu_n \,\|\, Q) = \sum_{(a,\ell) \in \mathcal{X} \times \mathcal{N}(\mathcal{X})} \nu_n(a, \ell) \log \tfrac{Q(a,\ell)}{Q_n(a,\ell)}$$
$$= -H(\omega_n \,\|\, \omega) - H(\varpi_n \,\|\, \varpi) - \sum_{a,b \in \mathcal{X}} \varpi(a, b) \tfrac{\omega_n(a)}{\omega(a)} + \sum_{a,b \in \mathcal{X}} \varpi(a, b) \log \tfrac{\omega_n(a)}{\omega(a)} + \|\varpi_n\|. \tag{4.7}$$

Note that this expression does not depend on ν_n. As the first, second and fourth term of (4.7) converge to 0, and the third and fifth term converge to $\|\varpi\|$, the expression (4.7) vanishes in the limit, and this completes the proof of the upper bound in Theorem 2.5.

4.3 An upper bound on the support of empirical neighbourhood measures

The cardinality of the support, denoted $\sharp\mathcal{S}(\nu)$, of an empirical neighbourhood measure ν of a graph with n vertices is naturally bounded by n. For the proof of the lower bound in Theorem 2.5 we need a better upper bound. We still use m to denote the cardinality of $\mathcal{X}$, and let

$$C := 2^m \, \frac{\Gamma(m + 2)^{\frac{m}{m+1}}}{\Gamma(m)} \text{ and } D := 2^m \, \frac{(m + 1)^m}{\Gamma(m)},$$

where $\Gamma(\cdot)$ is the Gamma function.

Lemma 4.4. *For every* $(\omega_n, \varpi_n) \in \mathcal{M}_n(\mathcal{X}) \times \tilde{\mathcal{M}}_{*,n}(\mathcal{X} \times \mathcal{X})$ *and* $\nu_n \in \mathcal{M}_n(\mathcal{X} \times \mathcal{N}(\mathcal{X}))$ *with* $\Phi(\nu_n) = (\omega_n, \varpi_n)$*, we have*

$$\sharp\mathcal{S}(\nu_n) \leq C \,\Big[n\|\varpi_n\|\Big]^{\frac{m}{m+1}} + D. \tag{4.8}$$

The following lemma provides a step in the proof of Lemma 4.4.

Lemma 4.5. *Suppose* $j \in \mathbb{N} \cup \{0\}$ *and* $n \in \mathbb{N}$*. Then,*

$$\tfrac{1}{\Gamma(n)}\, j^{n-1} \leq \sharp\Big\{(l_1, \ldots, l_n) \in \big(\mathbb{N} \cup \{0\}\big)^n \colon l_1 + \ldots + l_n = j\Big\} \leq \tfrac{1}{\Gamma(n)}\, (j + n)^{n-1}. \tag{4.9}$$

Proof. The proof is by induction on n. Equation (4.9) holds trivially for all $j \in \mathbb{N} \cup \{0\}$ and $n = 1, 2$, so we assume it holds for all j and $n \geq 2$. By the induction hypothesis, for any j,

$$\frac{1}{\Gamma(n)} \sum_{l=0}^{j} (j - l)^{n-1} \leq \sum_{l=0}^{j} \sharp\Big\{(l_1, \ldots, l_{n-1}) \in \big(\mathbb{N} \cup \{0\}\big)^{n-1} \colon l_1 + \ldots + l_{n-1} = j - l\Big\}$$
$$= \sharp\Big\{(l_1, \ldots, l_n) \in \big(\mathbb{N} \cup \{0\}\big)^n \colon l_1 + \ldots + l_n = j\Big\} \leq \frac{1}{\Gamma(n)} \sum_{l=0}^{j} (j - l + n)^{n-1}.$$

For the first and last term, we obtain the lower and upper bounds

$$\sum_{l=0}^{j}(j-l)^{n-1} \geq \int_0^j y^{n-1}\,dy = \frac{1}{n}\,j^n = \frac{\Gamma(n)}{\Gamma(n+1)}j^n$$

and

$$\sum_{l=0}^{j}(j-l+n)^{n-1} \leq \int_n^{j+n} y^{n-1}dy \leq \int_0^{j+n+1} y^{n-1}dy = \frac{1}{n}\,(j+n+1)^n = \frac{\Gamma(n)}{\Gamma(n+1)}(j+n+1)^n,$$

which yields inequality (4.9) for $n+1$ instead of n, and completes the induction. ■

Proof of Lemma 4.4. Suppose $(\omega_n,\varpi_n)\in\mathcal{M}_n(\mathcal{X})\times\mathcal{M}_n(\mathcal{X}\times\mathcal{X})$. Let

$$\begin{aligned} a_m(j) &:= \sharp\big\{(a,\ell)\in\mathcal{X}\times\mathcal{N}(\mathcal{X})\,:\,\sum_{b\in\mathcal{X}}\ell(b)=j\big\} \\ &= m\times\sharp\Big\{(l_1,\,\ldots,\,l_m)\in\big(\mathbb{N}\cup\{0\}\big)^m : l_1+\ldots+l_m=j\Big\}. \end{aligned}$$

For any positive integer k we write

$$\theta_k=\min\Big\{\theta\in\mathbb{N}:\sum_{j=0}^{\theta}a_m(j)\geq k\Big\}.$$

We observe from Lemma 4.5 that,

$$k\leq\sum_{j=0}^{\theta_k}a_m(j)\leq m\sum_{j=0}^{\theta_k}\tfrac{1}{\Gamma(m)}(j+m)^{m-1}\leq\tfrac{m}{\Gamma(m)}\int_0^{\theta_k+m}y^{m-1}dy=\tfrac{1}{\Gamma(m)}\big(\theta_k+m\big)^m.$$

Thus, we have $\theta_k\geq\big(k\Gamma(m)\big)^{\frac{1}{m}}-m=:\alpha_k$. This yields

$$\sum_{j=0}^{\theta_k}ja_m(j)\geq\tfrac{1}{\Gamma(m)}\sum_{j=0}^{\lceil\alpha_k\rceil}j^m\geq\tfrac{1}{\Gamma(m)}\int_0^{\alpha_k-1}y^m dy\geq\tfrac{1}{\Gamma(m+2)}\big(\alpha_k-1\big)^{m+1}, \tag{4.10}$$

where $\lceil y\rceil$ is the smallest integer greater or equal to y.
Observe that the size of the support of the measure $\nu_n\in\mathcal{K}^{(n)}(\omega_n,\varpi_n)$ satisfies

$$\sharp\mathcal{S}(\nu_n)\leq\max\Big\{k\colon\sum_{j=0}^{\theta_k}ja_m(j)\leq n\|\varpi_n\|\Big\},$$

and hence, using (4.10) and the inequality $(a+b)^m\leq 2^m(a^m+b^m)$ for $a,b\geq 0$,

$$\begin{aligned}\sharp\mathcal{S}(\nu_n)&\leq\max\Big\{k\colon\tfrac{1}{\Gamma(m+2)}\big(\alpha_k-1\big)^{m+1}\leq n\|\varpi_n\|\Big\}\\ &\leq\Gamma(m)^{-1}\Big((n\|\varpi_n\|)^{\frac{1}{m+1}}\Gamma(m+2)^{\frac{1}{m+1}}+m+1\Big)^m\leq C\,(n\|\varpi_n\|)^{\frac{m}{m+1}}+D\,,\end{aligned}$$

where the constants C, D were defined before the formulation of the lemma. ■

4.4 Approximation by empirical neighbourhood measures

Throughout this section we assume that $\omega_n \in \mathcal{M}_n(\mathcal{X})$ with $\omega_n \to \omega$, $\varpi_n \in \tilde{\mathcal{M}}_{*,n}(\mathcal{X}\times\mathcal{X})$ with $\varpi_n \to \varpi$, and that (ϖ,ν) is sub-consistent and $\nu_1 = \omega$. Our aim is to show that ν can be approximated in the weak topology by some $\nu_n \in \mathcal{M}_n(\mathcal{X}\times\mathcal{N}(\mathcal{X}))$ with $\Phi(\nu_n) = (\omega_n, \varpi_n)$ and the additional feature that

$$\sum_{b\in\mathcal{X}} \ell(b) \le n^{1/3} \text{ for } \nu_n\text{-almost every } (a,\ell). \tag{4.11}$$

The approximation will be done in three steps, given as Lemma 4.6, 4.7 and 4.9. We denote by d the metric of total variation, i.e.

$$d(\nu,\tilde{\nu}) = \tfrac{1}{2} \sum_{(a,\ell)\in\mathcal{X}\times\mathcal{N}(\mathcal{X})} |\nu(a,\ell) - \tilde{\nu}(a,\ell)|, \qquad \text{for } \nu,\tilde{\nu} \in \mathcal{M}(\mathcal{X}\times\mathcal{N}(\mathcal{X})).$$

This metric generates the weak topology.

Lemma 4.6 (Approximation Step 1)**.** *For every $\varepsilon > 0$, there exist $\hat{\nu} \in \mathcal{M}(\mathcal{X}\times\mathcal{N}(\mathcal{X}))$ and $\hat{\varpi} \in \tilde{\mathcal{M}}(\mathcal{X}\times\mathcal{X})$ such that $|\varpi(a,b) - \hat{\varpi}(a,b)| \le \varepsilon$ for all $a,b\in\mathcal{X}$, $d(\nu,\hat{\nu}) \le \varepsilon$ and $(\hat{\varpi},\hat{\nu})$ is consistent.*

Proof. By our assumption (ϖ,ν) is sub-consistent. For any $b\in\mathcal{X}$ define $e^{(b)} \in \mathcal{N}(\mathcal{X})$ by $e^{(b)}(a) = 0$ if $a \ne b$, and $e^{(b)}(b) = 1$. For large n define measures $\hat{\nu}_n \in \mathcal{M}(\mathcal{X}\times\mathcal{N}(\mathcal{X}))$ by

$$\hat{\nu}_n(a,\ell) = \nu(a,\ell)\Big(1 - \tfrac{\|\varpi\| - \|\langle \nu(\cdot,\ell),\ell(\cdot)\rangle\|}{n}\Big) + \sum_{b\in\mathcal{X}} \mathbb{1}\{\ell = ne^{(b)}\}\, \tfrac{\varpi(a,b) - \langle \nu(\cdot,\ell),\ell(\cdot)\rangle(a,b)}{n}\,.$$

Note that $\hat{\nu}_n \to \nu$ and that, for all $a,b\in\mathcal{X}$,

$$\begin{aligned}\sum_{\ell\in\mathcal{N}(\mathcal{X})} \hat{\nu}_n(a,\ell)\ell(b) &= \Big(1 - \tfrac{\|\varpi\| - \|\langle \nu(\cdot,\ell),\ell(\cdot)\rangle\|}{n}\Big) \sum_{\ell\in\mathcal{N}(\mathcal{X})} \nu(a,\ell)\ell(b) + \varpi(a,b) - \langle \nu(\cdot,\ell),\ell(\cdot)\rangle(a,b)\\ &= \varpi(a,b) - \tfrac{\|\varpi\| - \|\langle \nu(\cdot,\ell),\ell(\cdot)\rangle\|}{n}\, \langle \nu(\cdot,\ell),\ell(\cdot)\rangle(a,b) \overset{n\uparrow\infty}{\longrightarrow} \varpi(a,b)\,.\end{aligned}$$

Hence, defining $\hat{\varpi}_n$ by $\hat{\varpi}_n(a,b) = \sum \hat{\nu}_n(a,\ell)\ell(b)$, we have a sequence of consistent pairs $(\hat{\varpi}_n,\hat{\nu}_n)$ converging to (ϖ,ν), as required. □

Lemma 4.7 (Approximation Step 2)**.** *For every $\varepsilon > 0$, there exists $n(\varepsilon)$ such that, for all $n \ge n(\varepsilon)$ there exists $\nu_n \in \mathcal{M}_n(\mathcal{X}\times\mathcal{N}(\mathcal{X}))$ with $\Phi(\nu_n) = (\omega_n,\varpi_n)$ such that $d(\nu_n,\nu) \le \varepsilon$.*

The key to the construction of the measure ν_n is the following 'law of large numbers'.

Lemma 4.8. *For every $\delta > 0$, there exists $\hat{\nu} \in \mathcal{M}(\mathcal{X}\times\mathcal{N}(\mathcal{X}))$ with $d(\nu,\hat{\nu}) < \delta$ such that, for i.i.d. $\mathcal{N}(\mathcal{X})$-valued random variables ℓ_j^a, $j = 1,\dots,n\omega_n(a)$ with law $\hat{\nu}(\,\cdot\,|\,a) := \hat{\nu}(\{a\}\times\cdot\,)/\hat{\nu}_1(a)$, almost surely,*

$$\limsup_{n\to\infty} \Big(\frac{1}{n}\sum_{j=1}^{n\omega_n(a)} \ell_j^a(b) - \varpi_n(a,b)\Big) \le \delta, \qquad \text{for all } a,b\in\mathcal{X}. \tag{4.12}$$

Proof. By Lemma 4.6 we can choose a consistent pair $(\hat{\varpi},\hat{\nu})$ such that $d(\nu,\hat{\nu}) < \delta$ and, for all $a,b\in\mathcal{X}$,

$$\tfrac{\nu_1(a)}{\hat{\nu}_1(a)} \le 1 + \tfrac{\delta}{\|\varpi\|+1} \quad \text{and} \quad \hat{\varpi}(a,b)\big(1 + \tfrac{\delta}{\|\varpi\|+1}\big) \le \varpi(a,b)\big(1 + \tfrac{\delta}{\varpi(a,b)}\big)\,.$$

The random variables $\ell_j^a(b)$, $j = 1,\dots,n\omega_n(a)$ are i.i.d. with expectation

$$\mathbb{E}\ell_1^a(b) = \sum_\ell \hat{\nu}(\ell\,|\,a)\ell(b) = \frac{\hat{\varpi}(a,b)}{\hat{\nu}_1(a)}\,.$$

Hence, by the strong law of large numbers, almost surely,

$$\limsup_{n\to\infty}\Big(\frac{1}{n}\sum_{j=1}^{n\omega_n(a)}\ell_j^a(b)-\varpi_n(a,b)\Big)\le\tfrac{\nu_1(a)}{\hat\nu_1(a)}\hat\varpi(a,b)-\varpi(a,b)\le\delta,$$

where we also used that $\omega_n(a)\to\omega(a)=\nu_1(a)$ and $\varpi_n(a,b)\to\varpi(a,b)$. ■

Proof of Lemma 4.7. We use a randomised construction. Given (ϖ,ν) sub-consistent with $\nu_1=\omega$ and $\varepsilon>0$, choose $\hat\nu$ as in Lemma 4.8 with $\delta=\varepsilon/(3m)$, where m is the cardinality of $\mathcal{X}$. For every $a\in\mathcal{X}$, we draw tuples ℓ_j^a, $j=1,\dots,n\omega_n(a)$ independently according to $\hat\nu(\,\cdot\,|\,a)$ and define $e_n(a,b)$ by

$$e_n(a,b):=\frac{1}{n}\sum_{j=1}^{n\omega_n(a)}\ell_j^a(b)-\varpi_n(a,b),\ \text{ for all } a,b\in\mathcal{X}.$$

We modify the tuples $(\ell_j^a\,:\,j=1,\dots,n\omega_n(a))$ as follows:

- If $e_n(a,b)<0$, we add an amount to the last element $\ell^a_{n\omega_n(a)}(b)$ such that the modified tuple satisfies $e_n(a,b)=0$;
- if $e_n(a,b)>0$, by Lemma 4.8, the 'overshoot' $ne_n(a,b)$ cannot exceed $n\delta$. We successively deduct one from the nonzero elements in $\ell_j^a(b)$, $j=1,\dots,n\omega_n(a)$ until the modified tuples satisfy $e_n(a,b)=0$;
- if $e_n(a,b)=0$ we do not modify $\ell_j^a(b)$.

We denote by $(\tilde\ell_j^a\,:\,j=1,\dots,n\omega_n(a))$ the tuples after all modifications.

For each $a\in\mathcal{X}$ define probability measures $\tilde\Delta_n(\,\cdot\,|\,a)$ and $\Delta_n(\,\cdot\,|\,a)$ by

$$\tilde\Delta_n(\ell\,|\,a)=\frac{1}{n\omega_n(a)}\sum_{j=1}^{n\omega_n(a)}\mathbb{1}_{\{\tilde\ell_j^a=\ell\}},\qquad\text{for }\ell\in\mathcal{N}(\mathcal{X}),$$

respectively,

$$\Delta_n(\ell\,|\,a)=\frac{1}{n\omega_n(a)}\sum_{j=1}^{n\omega_n(a)}\mathbb{1}_{\{\ell_j^a=\ell\}},\qquad\text{for }\ell\in\mathcal{N}(\mathcal{X}).$$

We define probability measures $\tilde\nu_n\in\mathcal{M}_n(\mathcal{X}\times\mathcal{N}(\mathcal{X}))$ and $\nu_n\in\mathcal{M}_n(\mathcal{X}\times\mathcal{N}(\mathcal{X}))$ by $\tilde\nu_n(a,\ell)=\omega_n(a)\Delta_n(\ell\,|\,a)$, respectively $\nu_n(a,\ell)=\omega_n(a)\tilde\Delta_n(\ell\,|\,a)$, for $(a,\ell)\in\mathcal{X}\times\mathcal{N}(\mathcal{X})$. Recall from our modification procedure that, in the worst case, we have changed $nm\delta$ of the tuples. Thus,

$$d(\tilde\nu_n,\nu_n)\le m\delta\le\tfrac{1}{3}\,\varepsilon\,.$$

As a result of our modifications we have $\Phi(\nu_n)=(\omega_n,\varpi_n)$. We observe that, for all $(a,\ell)\in\mathcal{X}\times\mathcal{N}(\mathcal{X})$, the random variables

$$\mathbb{1}\{\ell_1^a=\ell\},\dots,\mathbb{1}\{\ell^a_{n\omega_n(a)}=\ell\}$$

are independent Bernoulli random variables with success probability $\hat\nu(\ell\,|\,a)$ and hence, almost surely,

$$\lim_{n\to\infty}\Delta_n(\ell\,|\,a)=\hat\nu(\ell\,|\,a)\,.$$

Therefore, for all $(a,\ell)\in\mathcal{X}\times\mathcal{N}(\mathcal{X})$, we obtain $\lim_{n\to\infty}\tilde\nu_n(a,\ell)=\hat\nu(a,\ell)$, almost surely. Thus, almost surely, for all large n, we have $d(\nu_n,\nu)\le d(\nu_n,\tilde\nu_n)+d(\tilde\nu_n,\hat\nu)+d(\hat\nu,\nu)\le\varepsilon$, as claimed. ■

Lemma 4.9 (Approximation Step 3)**.** *Let $\nu_n \in \mathcal{M}_n(\mathcal{X} \times \mathcal{N}(\mathcal{X}))$ with $\Phi(\nu_n) = (\omega_n, \varpi_n)$. For every $\varepsilon > 0$ there exists $n(\varepsilon)$ such that, for all $n \geq n(\varepsilon)$, we can find $\tilde{\nu}_n \in \mathcal{M}_n(\mathcal{X} \times \mathcal{N}(\mathcal{X}))$ satisfying* (4.11) *and $\Phi(\tilde{\nu}_n) = (\omega_n, \varpi_n)$.*

Proof. As $\nu_n \in \mathcal{M}_n(\mathcal{X} \times \mathcal{N}(\mathcal{X}))$, there is a representation

$$\nu_n = \frac{1}{n} \sum_{k=1}^{n} \delta_{(a_k, \ell_k)}, \qquad \text{for } a_k \in \mathcal{X}, \ell_k \in \mathcal{N}(\mathcal{X}).$$

Fix $\delta > 0$ and $a \in \mathcal{X}$. Look at the sets

- $V^+ = \{1 \leq k \leq n \colon a_k = a, \sum_b \ell_k(b) > n^{1/3}\}$ with cardinality $\sharp V^+ \leq (n \sum_b \varpi_n(a,b))^{2/3}$,
- $V^- = \{1 \leq k \leq n \colon a_k = a, \sum_b \ell_k(b) \leq n^{1/4}\}$ with cardinality $\sharp V^- \geq n - (n \sum_b \varpi_n(a,b))^{3/4}$.

For each $k \in V^+$ we replace ℓ_k by a smaller vector $\tilde{\ell}_k$ such that $\sum_b \tilde{\ell}_k(b) = n^{1/3}$. As

$$\sum_{k \in V_1} \sum_b \ell_k(b) \leq n \sum_{b \in \mathcal{X}} \varpi_n(a,b)$$

we may replace (for large n) no more than δn of the vectors ℓ_k, $k \in V^-$, by larger vectors $\tilde{\ell}_k$ such that

$$\sum_b \tilde{\ell}_k(b) \leq n^{1/3} \text{ and } \sum_{k=1}^{n} \sum_{b \in \mathcal{X}} 1\!\!1\{a_k = a\}\, \tilde{\ell}_k(b) = \sum_{k=1}^{n} \sum_{b \in \mathcal{X}} 1\!\!1\{a_k = a\}\, \ell_k(b)\,,$$

where we use the convention $\tilde{\ell}_k = \ell_k$ if this vector was not changed in the procedure. Performing such an operation for every $a \in \mathcal{X}$ we may define

$$\tilde{\nu}_n = \frac{1}{n} \sum_{k=1}^{n} \delta_{(a_k, \tilde{\ell}_k)},$$

and observe that (4.11) holds and $\Phi(\tilde{\nu}_n) = (\omega_n, \varpi_n)$. Moreover,

$$d(\nu_n, \tilde{\nu}_n) \leq \tfrac{m}{2n} \big((n \sum_b \varpi_n(a,b))^{2/3} + \delta n \big),$$

which is less than $\varepsilon > 0$ for a suitable choice of $\delta > 0$, and all sufficiently large n. ∎

4.5 Proof of the lower bound in Theorem 2.5.

There is a partial analogue to Lemma 4.3 for the lower bounds.

Lemma 4.10. *For any $\nu_n \in \mathcal{M}_n(\mathcal{X} \times \mathcal{N}(\mathcal{X}))$ which satisfies* (4.11) *with $\Phi(\nu_n) = (\omega_n, \varpi_n)$ and any $\varepsilon > 0$, we have*

$$\mathbb{P}\big\{ d(M, \nu_n) < \varepsilon \,\big|\, \Phi(M) = (\omega_n, \varpi_n) \big\} \geq \exp\big(- nH(\nu_n \,\|\, Q_n) - \varepsilon_2^{(n)}(\nu_n) \big)\,,$$

where Q_n is as Lemma 4.3 and

$$\lim_{n \uparrow \infty} \frac{1}{n}\, \varepsilon_2^{(n)}(\nu_n) = 0\,.$$

Proof of Lemma 4.10. We use the notation and some results from the proof of the upper bound, Lemma 4.3. In particular, recall the definition of $n(a,b)$ from (4.1) and, from (4.2), that

$$\mathbb{P}\big\{ d(M, \nu_n) < \varepsilon \,\big|\, \Phi(M) = (\omega_n, \varpi_n) \big\} = \frac{\sharp\big\{ \tilde{x} \in \mathcal{A}_n(\omega_n, \varpi_n) \,:\, d(M \circ \Psi(\tilde{x}), \nu_n) < \varepsilon \big\}}{\sharp\big\{ \tilde{x} \in \mathcal{A}_n(\omega_n, \varpi_n)) \big\}}\,, \tag{4.13}$$

and that the denominator was evaluated in (4.3) as

$$\binom{n}{n\omega_n(a),\ a\in\mathcal{X}}\prod_{\{a,b\}}\prod_{k=1}^{n(a,b)}\Big(\frac{n^2\omega_n(a)\omega_n(b)-n\omega_n(a)\mathbb{1}_{\{a=b\}}}{1+\mathbb{1}_{\{a=b\}}}-(k-1)\Big).$$

We now describe a procedure which yields (for sufficiently large n) a lower bound of

$$\binom{n}{n\nu_n(a,\ell),\ a\in\mathcal{X}\,,\ell\in\mathcal{N}(\mathcal{X})}\prod_{(a,b)}\frac{(n\varpi_n(a,b)-2\lceil n^{2/3}\rceil-2)!}{\prod_{j=1}^{n\omega_n(a)}(\ell_a^{(j)}(b))!}\,2^{-\frac{n}{2}\varpi_n(\Delta)}\,. \tag{4.14}$$

for the numerator, where $\ell_a^{(j)}(b)$, $j=1,\dots,n\omega_n(a)$ are any enumeration of the family containing each $\ell(b)$ with multiplicity $n\nu_n(a,\ell)$.

First, we allocate to each vertex some $(a,\ell)\in\mathcal{X}\times\mathcal{N}(\mathcal{X})$ in such a way that every vector (a,ℓ) is allocated $n\nu_n(a,\ell)$ times. There are

$$\binom{n}{n\nu_n(a,\ell),\ a\in\mathcal{X}\,,\ell\in\mathcal{N}(\mathcal{X})}$$

ways to do this. Next we add edges between the vertices of two *different* colours, $a,b\in\mathcal{X}$. To this end, we distribute the numbers in $\{1,\dots,n(a,b)-\lceil n^{2/3}\rceil-1\}$ among the vertices with colour a so that a vertex carrying vector (a,ℓ) gets at most $\ell(b)$ numbers. A crude lower bound for the number of ways to do this is

$$\frac{(n(a,b)-\lceil n^{2/3}\rceil-1)!}{\prod_{j=1}^{n\omega_n(a)}(\ell_a^{(j)}(b))!}.$$

Now the numbers in $\{1,\dots,n(a,b)-\lceil n^{2/3}\rceil-1\}$ are distributed successively, this time among the vertices of colour b. Again we do this in such a way that a vertex carrying vector (b,ℓ) has a capacity to carry no more than $\ell(a)$ numbers. However, we are more cautious now: When distributing k we look at the vertex of colour a, which already carries k. If this carries numbers from $\{1,\dots,k-1\}$, we do not allow k to be associated with any vertex of colour b which carries one of these numbers. By (4.11) this rules out no more than $n^{1/3}$ vertices, each of which has a capacity no more than $n^{1/3}$, so that the number of ways to do this is at least

$$\frac{(n(a,b)-\lceil n^{2/3}\rceil-1)!}{\prod_{j=1}^{n\omega_n(b)}(\ell_b^{(j)}(a))!}.$$

Next, for $k=1,\dots n(a,b)-\lceil n^{2/3}\rceil-1$ draw the kth edge between the two vertices of colour a and b carrying number k and observe that when allocating the numbers we have been cautious not to cause any multiple edges. Obviously, there is at least one way to establish a further $\lceil n^{2/3}\rceil+1$ edges between vertices of colour a and b without creating multiple edges.

We now add the edges connecting vertices of *the same* colour $a\in\mathcal{X}$. For this purpose, we successively distribute the numbers in $\{1,\dots,n(a,a)-\lceil n^{2/3}\rceil-1\}$ and $\{n(a,a)+1,\dots,2n(a,a)-\lceil n^{2/3}\rceil-1\}$ among the vertices with colour a so that a vertex carrying vector (a,ℓ) gets at most $\ell(b)$ numbers. When distributing $k>n(a,a)$ we look at the vertex of colour a, which already carries $k-n(a,a)$. If this vertex carries numbers $j\in\{1,\dots,k-n(a,a)-1\}$, we do not allow k to be associated with the vertices carrying numbers $j+n(a,a)$. We also do not allow k to be associated with the vertex itself. By (4.11) these restrictions rule out no more than $n^{1/3}$ vertices, each of which has a capacity no more than $n^{1/3}$, so that the number of ways to do this is at least

$$\frac{(2n(a,a)-2\lceil n^{2/3}\rceil-2)!}{\prod_{j=1}^{n\omega_n(a)}(\ell_a^{(j)}(a))!}.$$

Obviously, there is at least one way of allocating the remaining numbers $\{n(a,a)-\lceil n^{2/3}\rceil,\ldots,n(a,a)\}$ and $\{2n(a,a)-\lceil n^{2/3}\rceil,\ldots,2n(a,a)\}$ to vertices so that no single vertex carries a matching pair $j, j+n(a,a)$, and no pair of vertices carry two or more matching pairs between them. Next, for $k=1,\ldots,n(a,b)$ draw the kth edge between the two vertices carrying numbers k and $k+n(a,a)$ and observe that when allocating the numbers we have been cautious not to cause any loops or multiple edges. As, for every $k\in\{1,\ldots,n(a,a)\}$ the numbers k and $k+n(a,a)$ could be interchanged without changing the configuration, the total number of different configurations constructable in this procedure is bounded from below by

$$\binom{n}{n\nu_n(a,\ell),\, a\in\mathcal{X}\,,\ell\in\mathcal{N}(\mathcal{X})}\times\prod_{\substack{(a,b)\\a\neq b}}\frac{(n(a,b)-\lceil n^{2/3}\rceil-1)!}{\prod_{j=1}^{n\omega_n(a)}(\ell_a^{(j)}(b))!}\times\prod_{a\in\mathcal{X}}\frac{(2n(a,a)-2\lceil n^{2/3}\rceil-2)!}{\prod_{j=1}^{n\omega_n(a)}(\ell_a^{(j)}(a))!}\,2^{-n(a,a)}\,,$$

and this is bounded from below by the quantity in (4.14). Every resulting graph satisfies the constraint $\Phi(M)=(\omega_n,\varpi_n)$. To measure the distance between its empirical neighbourhood measure M and ν_n, we say that a vertex $v\in V$ is *successful* if the associated $(X(v),L(v))$ is identical to the (a,ℓ) they were carrying after the initial step. Note that after allocation of the edges with numbers in $\{1,\ldots,n(a,b)-\lceil n^{2/3}\rceil-1\}$ among the vertices of all colours, all but at most $2m^2(\lceil n^{2/3}\rceil+1)$ vertices $v\in V$ were successful. Adding in the further edges in the last step can lead to up to $2m^2(\lceil n^{2/3}\rceil+1)$ further unsuccessful vertices. Hence

$$d(\nu_n,M)\le\frac{1}{2n}\sum_{v\in V}1\!\!1\{v \text{ unsuccessful }\}\le 4m^2\,n^{-1/3}\overset{n\to\infty}{\longrightarrow}0\qquad\text{as } n\uparrow\infty.$$

To complete the proof, we again use Stirling's formula to analyse the combinatorial terms obtained as an estimate for the numerator and denominator in (4.13). For the denominator we get the same main terms as in Lemma 4.3 with slightly different error terms, which however do not depend on ν_n. More interestingly, we have

$$\prod_{a\in\mathcal{X}}\binom{n\omega_n(a)}{n\nu_n(a,\ell),\,\ell\in\mathcal{N}(\mathcal{X})}\ge\exp\Big(n\sum_a\omega_n(a)\log\omega_n(a)-n\sum_{(a,\ell)}\nu_n(a,\ell)\log\nu_n(a,\ell)\Big)$$
$$\times\exp\Big(-\tfrac{|\mathcal{S}(\nu_n)|}{2}\log(2\pi n)-\sum_{\substack{(a,\ell)\\n\nu_n(a,\ell)\ge1}}\tfrac{1}{12n\nu_n(a,\ell)}\Big)\,,$$

where the exponent in the error term is of order $o(n)$, by the bound on the size of the support of ν_n given in Lemma 4.4. Further,

$$\frac{(n\varpi_n(a,b)-2\lceil n^{2/3}\rceil-2)!}{\prod_{j=1}^{n\omega_n(a)}(\ell_a^{(j)}(b))!}\ge\exp\Big(-n\sum_\ell\log\big(\ell(b)!\big)\,\nu_n(a,\ell)+n\varpi_n(a,b)\log\big(n\varpi_n(a,b)\big)-n\varpi_n(a,b)\Big)$$
$$\times\exp\Big(-(2n^{2/3}+2)\,\log\big(n\varpi_n(a,b)\big)+n\varpi_n(a,b)\,\log\big(1-\tfrac{2n^{2/3}+2}{n\varpi_n(a,b)}\big)\Big)\,,$$

and the result follows by combining this with facts discussed in the context of the upper bound. ■

To complete the proof of the lower bound in Theorem 2.5, take an open set $\Gamma\subset\mathcal{M}(\mathcal{X}\times\mathcal{N}(\mathcal{X}))$. Then, for any $\nu\in\Gamma$ with (ϖ,ν) sub-consistent and $\nu_1=\omega$ we may find $\varepsilon>0$ with the ball around ν of radius $2\varepsilon>0$ contained in Γ. By our approximation, see Lemma 4.7 and 4.9, we may find $\nu_n\in\Gamma\cap\mathcal{M}_n(\mathcal{X}\times\mathcal{N}(\mathcal{X}))$ with $\Phi(\nu_n)=(\omega_n,\varpi_n)$ such that (4.11) holds and $d(\nu_n,\nu)\downarrow0$. Hence, for all large $n \ge n(\varepsilon)$,

$$\mathbb{P}\Big\{M \in \Gamma \,\Big|\, \Phi(M) = (\omega_n, \varpi_n)\Big\} \ge \mathbb{P}\big\{d(\nu_n, M) < \varepsilon \,\big|\, \Phi(M) = (\omega_n, \varpi_n)\big\}$$
$$\ge \exp\Big(- nH(\nu_n \,\|\, Q_n) - \varepsilon_2^{(n)}(\nu_n)\Big).$$

We observe that

$$\lim_{n\to\infty} H(\nu_n \,\|\, Q_n) - H(\nu \,\|\, Q) = \lim_{n\to\infty} H(\nu_n \,\|\, Q_n) - H(\nu_n \,\|\, Q) + \lim_{n\to\infty} H(\nu_n \,\|\, Q) - H(\nu \,\|\, Q) = 0,$$

where the last term vanishes by continuity of relative entropy, and the first term was shown to vanish in the proof of Lemma 4.10. This completes the proof of Theorem 2.5.

5. Proof of Theorems 2.1 and 2.3 (b) by mixing

We denote by $\Theta_n := \mathcal{M}_n(\mathcal{X}) \times \tilde{\mathcal{M}}_{*,n}(\mathcal{X} \times \mathcal{X})$ and $\Theta := \mathcal{M}(\mathcal{X}) \times \tilde{\mathcal{M}}_{*}(\mathcal{X} \times \mathcal{X})$. Define

$$\begin{aligned}
P^{(n)}_{(\omega_n,\varpi_n)}(\nu_n) &:= \mathbb{P}\big\{M = \nu_n \,\big|\, \Phi(M) = (\omega_n, \varpi_n)\big\}\,,\\
P^{(n)}(\omega_n, \varpi_n) &:= \mathbb{P}\big\{(L^1, L^2) = (\omega_n, \varpi_n)\big\}\,,\\
P^{(n)}_{\omega_n}(\varpi_n) &:= \mathbb{P}\{L^2 = \varpi_n \,|\, L^1 = \omega_n\}\,,\\
P^{(n)}(\omega_n) &:= \mathbb{P}\{L^1 = \omega_n\}\,.
\end{aligned}$$

The joint distribution of L^1, L^2 and M is the mixture of $P^{(n)}_{(\omega_n,\varpi_n)}$ with $P^{(n)}(\omega_n, \varpi_n)$ defined as

$$d\tilde{P}^n(\omega_n, \varpi_n, \nu_n) := dP^{(n)}_{(\omega_n,\varpi_n)}(\nu_n)\, dP^{(n)}(\omega_n, \varpi_n)\,, \tag{5.1}$$

whilst the joint distribution of L^1 and L^2 is the mixture of $P^{(n)}_{\omega_n}$ with $P^{(n)}$ given by

$$dP^{(n)}(\omega_n,\, \varpi_n) = dP^{(n)}_{\omega_n}(\varpi_n)\, dP^{(n)}(\omega_n). \tag{5.2}$$

Biggins [1, Theorem 5(b)] gives criteria for the validity of large deviation principles for the mixtures and for the goodness of the rate function if individual large deviation principles are known. The following two lemmas ensure validity of these conditions.

Lemma 5.1 (Exponential tightness)**.** *The following families of distributions are exponentially tight.*

(a) $(P^{(n)} \colon n \in \mathbb{N})$ *on* $\mathcal{M}(\mathcal{X}) \times \tilde{\mathcal{M}}_*(\mathcal{X} \times \mathcal{X})$,
(b) $(\tilde{P}^{(n)} \colon n \in \mathbb{N})$ *on* $\mathcal{M}(\mathcal{X}) \times \tilde{\mathcal{M}}_*(\mathcal{X} \times \mathcal{X}) \times \mathcal{M}(\mathcal{X} \times \mathcal{N}(\mathcal{X}))$.

Proof. (a) It suffices to show that, for every $\theta > 0$, there exists $N \in \mathbb{N}$ such that

$$\limsup_{n\to\infty} \tfrac{1}{n} \log \mathbb{P}\Big\{|E| > nN\Big\} \le -\theta.$$

To see this, let $c > \max_{a,b\in\mathcal{X}} C(a,b) > 0$. By a simple coupling argument we can define, for all sufficiently large n, a new coloured random graph $\tilde{X}$ with colour law μ and connection probability $\frac{c}{n}$, such that any edge present in X is also present in $\tilde{X}$. Let $|\tilde{E}|$ be the number of edges of $\tilde{X}$. Using Chebyshev's inequality, the binomial formula, and Euler's formula, we have that

$$\mathbb{P}\Big\{|\tilde{E}| \ge nl\Big\} \le e^{-nl}\mathbb{E}\big[e^{|\tilde{E}|}\big] = e^{-nl} \sum_{k=0}^{\frac{n(n-1)}{2}} e^k \binom{n(n-1)/2}{k} \Big(\frac{c}{n}\Big)^k \Big(1 - \frac{c}{n}\Big)^{n(n-1)/2-k}$$
$$= e^{-nl}\Big(1 - \frac{c}{n} + e\frac{c}{n}\Big)^{n(n-1)/2} \le e^{-nl} e^{nc(e-1+o(1))}.$$

Now given $\theta > 0$ choose $N \in \mathbb{N}$ such that $N > \theta + c(e-1)$ and observe that, for sufficiently large n,

$$\mathbb{P}\big\{|E| \geq nN\big\} \leq \mathbb{P}\big\{|\tilde{E}| \geq nN\big\} \leq e^{-n\theta},$$

which implies the statement.

(b) Given $\theta > 0$, we observe from (a) that there exists $N(\theta) \in \mathbb{N}$ such that, for all sufficiently large n,

$$\mathbb{P}\big\{M(\{\|\ell\| \geq 2\theta N(\theta)\}) \geq \theta^{-1} \text{ or } \|L^2\| \geq 2N(\theta)\big\} \leq \mathbb{P}\big\{|E| \geq nN(\theta)\big\} \leq e^{-\theta n}.$$

We define the set Ξ_θ by

$$\Xi_\theta := \big\{(\varpi, \nu) \in \tilde{\mathcal{M}}_*(\mathcal{X} \times \mathcal{X}) \times \mathcal{M}(\mathcal{X} \times \mathcal{N}(\mathcal{X})) \,:\, \nu\{\|\ell\| > 2lN(l)\} < l^{-1}\,\forall l \geq \theta \text{ and } \|\varpi\| < 2N(\theta)\big\}.$$

As $\{\|\ell\| \leq 2lN(l)\} \subset \mathcal{N}(\mathcal{X})$ is finite, hence compact, the set Ξ_θ is relatively compact in the weak topology, by Prohorov's criterion. Moreover, we have that

$$\tilde{P}^n\big((\Xi_\theta)^c\big) \leq \mathbb{P}\{\|L^2\| \geq 2N(\theta)\} + \sum_{l=\theta}^{\infty} \mathbb{P}\big\{M(\{\|\ell\| > 2lN(l)\}) \geq l^{-1}\big\} \leq C(\theta)\, e^{-n\theta}.$$

Therefore, $\limsup_{n\to\infty} \frac{1}{n} \log \tilde{P}^n((\mathrm{cl}\,\Xi_\theta)^c) \leq -\theta$, which completes the proof, as $\theta > 0$ was arbitrary. ∎

Now, we observe that the function $I(\omega, \varpi) = H(\omega \,\|\, \mu) + \mathcal{H}_C(\varpi\|\omega)$ is a good rate function, by a similar argument as in the proof of Lemma 3.2 (ii). Therefore, applying [1, Theorem 5(b)] to Sanov's theorem [9, Theorem 2.1.10] and Theorem 2.3 (a) we obtain the large deviation principle for $P^{(n)}$ on $\mathcal{M}(\mathcal{X}) \times \tilde{\mathcal{M}}_*(\mathcal{X} \times \mathcal{X})$ with good rate function $I(\omega, \varpi)$, which is Theorem 2.3 (b).

To prove Theorem 2.5 define the function

$$\tilde{J} \colon \Theta \times \mathcal{M}(\mathcal{X} \times \mathcal{N}(\mathcal{X})) \to [0, \infty], \qquad \tilde{J}((\omega, \varpi),\, \nu) = \tilde{J}_{(\omega,\varpi)}(\nu).$$

Lemma 5.2. *$\tilde{J}$ is lower semicontinuous.*

Proof. Suppose $\theta_n := ((\omega_n, \varpi_n),\, \nu_n)$ converges to $\theta := ((\omega, \varpi),\, \nu)$ in $\Theta \times \mathcal{M}(\mathcal{X} \times \mathcal{N}(\mathcal{X}))$. There is nothing to show if $\liminf_{\theta_n \to \theta} \tilde{J}(\theta_n) = \infty$. Otherwise, if (ϖ_n, ν_n) is sub-consistent for infinitely many n, then

$$\varpi(a, b) = \lim_{n\uparrow\infty} \varpi_n(a, b) \geq \liminf_{n\uparrow\infty} \langle \nu_n(\cdot, \ell), \ell(\cdot)\rangle(a, b) \geq \langle \nu(\cdot, \ell), \ell(\cdot)\rangle(a, b)\,,$$

hence (ϖ, ν) is sub-consistent. Similarly, if the first marginal of ν_n is ω_n, we see that the first marginal of ν is ω. We may therefore argue as in (4.6) to obtain

$$\liminf_{\theta_n\to\theta} J(\theta_n) = \liminf_{\theta_n\to\theta} H(\nu_n \,\|Q_n) \geq \lim_{\theta_n\to\theta} H(\nu_n \,\|\, Q_n) - H(\nu_n \,\|\, Q) + \liminf_{\nu_n\to\nu} H(\nu_n \,\|Q) = H(\nu \,\|Q),$$

where the last step uses continuity of relative entropy. This proves the lemma. ∎

Lemma 5.2 and Lemma 5.1 (b) ensure that we can apply [1, Theorem 5(b)] to the large deviation principles established in Theorem 2.3 (b) and 2.5. This yields a large deviation principle for $(\tilde{P}^n \colon n \in \mathbb{N})$ on $\mathcal{M}(\mathcal{X}) \times \tilde{\mathcal{M}}_*(\mathcal{X} \times \mathcal{X}) \times \mathcal{M}(\mathcal{X} \times \mathcal{N}(\mathcal{X}))$ with good rate function

$$\hat{J}(\omega, \varpi, \nu) = \begin{cases} H(\omega \,\|\, \mu) + \frac{1}{2}\,\mathfrak{H}_C(\varpi \,\|\, \omega) + H(\nu \,\|Q)\,, & \text{if } (\varpi, \nu) \text{ sub-consistent and } \nu_1 = \omega, \\ \infty\,, & \text{otherwise.} \end{cases}$$

By projection onto the last two components we obtain the large deviation principle as stated in Theorem 2.1 from the contraction principle, see e.g. [9, Theorem 4.2.1].

6. Proof of Corollaries 2.2 and 2.4

We derive the corollaries from Theorem 2.1 by applying the contraction principle, see e.g. [9, Theorem 4.2.1]. It just remains to simplify the rate functions.

6.1 Proof of Corollary 2.2 In the case of an uncoloured Erdős-Renyi graph, the function C degenerates to a constant c, $L^2 = |E|/n \in [0,\infty)$ and $M = D \in \mathcal{M}(\mathbb{N}\cup\{0\})$. Theorem 2.1 and the contraction principle imply a large deviation principle for D with good rate function

$$\delta(d) = \inf\big\{J(x,d)\colon x\ge 0\big\} = \inf\big\{H(d\,\|\,q_x) + \tfrac12 x\log x - \tfrac12 x\log c + \tfrac12\, c - \tfrac12 x\colon \langle d\rangle \le x\,\big\},$$

which is to be understood as infinity if $\langle d\rangle$ is infinite. We denote by $\delta^x(d)$ the expression inside the infimum and consider the cases (i) $\langle d\rangle \le c$ and (ii) $\infty > \langle d\rangle \ge c$ separately.

Case (i): Under our condition the equation $x = ce^{-2(1-\langle d\rangle/x)}$ has a unique solution, which satisfies $x\ge\langle d\rangle$. Elementary calculus shows that the global minimum of $y\mapsto\delta^y(d)$ on $(0,\infty)$ is attained at the value $y=x$, where x is the solution of our equation.

Case (ii): For any $\varepsilon>0$, we have

$$\delta^{\langle d\rangle+\varepsilon}(d) - \delta^{\langle d\rangle}(d) = \tfrac{\varepsilon}{2} + \tfrac{\langle d\rangle-\varepsilon}{2}\log\tfrac{\langle d\rangle}{\langle d\rangle+\varepsilon} + \tfrac{\varepsilon}{2}\log\tfrac{\langle d\rangle}{c} \ge \tfrac{\varepsilon}{2} + \tfrac{\langle d\rangle-\varepsilon}{2}\left(\tfrac{-\varepsilon}{\langle d\rangle}\right) + \tfrac{\varepsilon}{2}\log\tfrac{\langle d\rangle}{c} > 0,$$

so that the minimum is attained at $x=\langle d\rangle$.

6.2 Proof of Corollary 2.4 We begin the proof by defining the continuous linear map $W\colon\mathcal{M}(\mathcal{X})\times\tilde{\mathcal{M}}_*(\mathcal{X}\times\mathcal{X})\to[0,\infty)$ by $W(\omega,\varpi)=\frac12\|\varpi\|$. We infer from Theorem 2.3 and the contraction principle that $W(L^1,L^2)=|E|/n$ satisfies a large deviation principle in $[0,\infty)$ with the good rate function

$$\zeta(x) = \inf\big\{I(\omega,\varpi)\colon W(\omega,\varpi)=x\big\}.$$

To obtain the form of the rate in the corollary, the infimum is reformulated as unconstrained optimisation problem (by normalising ϖ)

$$\inf_{\substack{\varpi\in\mathcal{M}_*(\mathcal{X}\times\mathcal{X})\\ \omega\in\mathcal{M}(\mathcal{X})}}\Big\{H(\omega\,\|\,\mu) + xH(\varpi\,\|\,C\omega\otimes\omega) + x\log 2x + \tfrac12\,\|C\omega\otimes\omega\| - x\Big\}. \tag{6.1}$$

By Jensen's inequality $H(\varpi\,\|\,C\omega\otimes\omega)\ge-\log\|C\omega\otimes\omega\|$, with equality if $\varpi=\frac{C\omega\otimes\omega}{\|C\omega\otimes\omega\|}$, and hence, by symmetry of C we have

$$\begin{aligned}\min_{\varpi\in\mathcal{M}_*(\mathcal{X}\times\mathcal{X})}&\Big\{H(\omega\,\|\,\mu) + xH(\varpi\,\|\,C\omega\otimes\omega) + x\log 2x + \tfrac12\,\|C\omega\otimes\omega\| - x\Big\}\\ &= H(\omega\,\|\,\mu) - x\log\|C\omega\otimes\omega\| + x\log 2x + \tfrac12\,\|C\omega\otimes\omega\| - x.\end{aligned}$$

The form given in Corollary 2.4 follows by defining $y=\sum_{a,b\in\mathcal{X}} C(a,b)\omega(a)\omega(b)$.

Acknowledgments: We thank Nadia Sidorova for helpful discussions and, in particular, for contributing the idea of the proof of Lemma 4.2, Peter Eichelsbacher for helpful discussions on large deviations principles in strong topolgies, and two referees for numerous suggestions which significantly improved the paper. In particular the proof of Theorem 2.3 was consideably shortened by a referee's suggestion to use Gärtner-Ellis and mixing instead of a direct change-of-measure argument. This paper contains material from the first author's PhD thesis. The work of the second author was supported by an Advanced Research Fellowship of the EPSRC, and the revision of this paper was done during a visit of the first author to Bath, which was supported by the London Mathematical Society.

Bibliography

[1] J.D. Biggins. Large deviations for mixtures. *El. Comm. Probab.* **9** 60–71 (2004).

[2] J.D. Biggins and D.B. Penman. Large deviations in randomly coloured random graphs. *El. Comm. Probab.* **14** 290–301 (2009).

[3] B. Bollobás, S. Janson and O. Riordan. The phase transition in inhomogeneous random graphs. *Random Structures Algorithms* **31** 3–122 (2007).

[4] S. Boucheron, O. Bousquet and G. Lugosi. Concentration inequalities. In: *Advanced lectures in machine learning*, Eds: O. Bousquet, U. v. Luxburg and G. Rötsch. Springer (2004) 208-240.

[5] S. Boucheron, F. Gamboa and C. Leonard. Bins and balls: Large deviations of the empirical occupancy process. *Ann. Appl. Probab.* **12** 607-636 (2002).

[6] C. Cannings and D.B. Penman. Models of random graphs and their applications. In: *Handbook of Statistics 21. Stochastic Processes: Modelling and Simulation.* Eds: D.N. Shanbhag and C.R. Rao. Elsevier (2003) 51-91.

[7] N. O'Connell. Some large deviation results for sparse random graphs. *Probab. Theory Relat. Fields* **110** 277–285 (1998).

[8] A. Dembo, P. Mörters and S. Sheffield. Large deviations of Markov chains indexed by random trees. *Ann. Inst. Henri Poincaré: Probab. et Stat.* **41** 971-996 (2005).

[9] A. Dembo and O. Zeitouni. *Large deviations techniques and applications.* Springer, New York, (1998).

[10] K. Doku-Amponsah. *Large deviations and basic information theory for hierarchical and networked data structures.* PhD Thesis, University of Bath, 2006.

[11] P. Eichelsbacher and U. Schmock. Large deviations of U-empirical measures in strong topologies and applications. *Ann. Inst. Henri Poincaré: Probab. et Stat.* **38** 779-797 (2002).

[12] W. Feller. *An introduction to probability theory and its applications.* Volume I. Wiley, New York. Third edition, (1967).

[13] R. van der Hofstad. Critical behaviour in inhomogeneous random graphs. Preprint arXiv 0902.0216v1 (2009).

[14] F. den Hollander. *Large deviations.* 'AMS Fields Institute Monographs (2000).

[15] D.B. Penman. *Random graphs with correlation structure.* PhD Thesis, University of Sheffield, 1998.

[16] B. Söderberg. General formalism for inhomogeneous random graphs. Phys. Rev. E **66** (2002), no. 6, 066121, 6 pp.

Kwabena Doku-Amponsah, University of Ghana, Department of Statistics, P.O. Box LG 115, Legon-Accra, Ghana. E-mail: `kdoku@ug.edu.gh`.

Peter Mörters, University of Bath, Department of Mathematical Sciences, Claverton Down, Bath BA2 7AY, United Kingdom. E-mail: `maspm@bath.ac.uk`.