

# Universality of the mean-field for the Potts model

Anirban Basak<sup>1</sup> · Sumit Mukherjee<sup>2</sup>

Received: 17 August 2015 / Revised: 4 May 2016  
© Springer-Verlag Berlin Heidelberg 2016

**Abstract** We consider the Potts model with  $q$  colors on a sequence of weighted graphs with adjacency matrices  $A_n$ , allowing for both positive and negative weights. Under a mild regularity condition on  $A_n$  we show that the mean-field prediction for the log partition function is asymptotically correct, whenever  $\text{tr}(A_n^2) = o(n)$ . In particular, our results are applicable for the Ising and the Potts models on any sequence of graphs with average degree going to  $+\infty$ . Using this, we establish the universality of the limiting log partition function of the ferromagnetic Potts model for a sequence of asymptotically regular graphs, and that of the Ising model for bi-regular bipartite graphs in both ferromagnetic and anti-ferromagnetic domain. We also derive a large deviation principle for the empirical measure of the colors for the Potts model on asymptotically regular graphs.

**Keywords** Ising measure · Potts model · Log partition function · Mean-field · Large deviation

**Mathematics Subject Classification** 60K35 · 82B20 · 82B44

## 1 Introduction

One of the fundamental models in statistical physics is the nearest neighbor  $q$ -state Potts model. For a finite undirected graph  $\mathbf{G} := (V, E)$ , with vertex set  $V$ , and edge set  $E$ , the Potts model is a probability measure on  $[q]^{|V|}$  with  $[q] := \{1, 2, \dots, q\}$ ,

---

✉ Anirban Basak  
anirbanb.math@gmail.com

<sup>1</sup> Department of Mathematics, Duke University, Durham, NC 27705, USA

<sup>2</sup> Department of Statistics, Columbia University, Manhattan, NY 10027, USA

where  $|\cdot|$  denotes the cardinality of a set. The probability mass function for the Potts model at  $\mathbf{y} := \{y_i, i \in V\}$  is given by

$$\mu^{\beta, B}(\mathbf{y}) := \frac{1}{Z_G(\beta, B)} \exp \left\{ \beta \sum_{(i,j) \in E} \delta(y_i, y_j) + B \sum_{i \in V} \delta(y_i, 1) \right\}. \tag{1.1}$$

Here  $\delta(y, y') = 1_{y=y'}$ , and  $Z_G(\beta, B)$  is the normalizing constant, which is commonly termed as the *partition function*. The parameters  $\beta$  and  $B$  are known as *inverse temperature* parameter and *external magnetic field* parameters respectively, with  $\beta \geq 0$  is said to be the *ferromagnetic* regime, and  $\beta < 0$  is the *anti-ferromagnetic* regime. When  $q = 2$ , the measure  $\mu^{\beta, B}(\cdot)$  is the well known *Ising measure*.

Although Ising and Potts models originated from statistical physics [34,41], due to its wide applications it has received a lot of recent interest from varied areas, including statistics (cf. [1,5,15,42] and references therein), computer science (cf. [4,12,31,44] and references therein), combinatorics, finance, social networks, computer vision, biology, and signal processing. Potts models on graphs also have connections with many graph properties, such as the number of proper colorings, max cut, min cut, min bisection (cf. [2,10,11,22] and references therein), which are of interest in classical graph theory. One of the main difficulties in the study of the Ising and the Potts model is the intractability of its partition function. If the partition function were available in closed form, one could analyze it to compute moments and limiting distributions, carry on inference in a statistical framework using maximum likelihood, or compute thermodynamic limits of these models which are of interest in statistical physics. As the partition function involves summing the unnormalized mass function over exponentially many terms, computing the partition function numerically or otherwise is challenging in general. Since exact computations are infeasible, they are broadly two approaches to tackle this problem. A branch of research is directed towards devising efficient algorithms to approximate the log partition function (cf. [35,46], and the references therein). Whereas, probabilists are interested in studying the asymptotics of the log partition function for sequence of graphs  $G_n$  for large  $n$  (cf. [23,24,28,29] and references therein), in an attempt to understand these measures. More precisely, considering a sequence of graphs  $G_n := ([n], E_n)$ , with growing size, the goal is to compute the asymptotic *limiting log partition function*  $\Phi(\beta, B)$ , where

$$\Phi(\beta, B) := \lim_{n \rightarrow \infty} \frac{1}{n} \log Z_{G_n}(\beta, B),$$

and  $\Phi_n(\beta, B) := \log Z_{G_n}(\beta, B)$ . To get a non-trivial value of  $\Phi(\beta, B)$  one must scale  $\beta$  appropriately depending on  $|E_n|$ . In particular, the inverse temperature parameter in (1.1) should be replaced by  $\beta_n := (n/2|E_n|)\beta$  for the Potts model on  $G_n$ . This scaling ensures that  $\Phi(\beta, B)$  is not a constant function for all choices of  $\beta$ , and  $B$ . By a slight abuse of notation we denote this measure by  $\mu_n^{\beta, B}(\cdot)$ .

One common scheme of approximating  $\Phi_n(\beta, B)$  is via the *naive mean-field* method. Mean-field method has been in the statistical physics literature for a long time (see [14,38]). Below we describe the mean-field method in our context in detail:

### 1.1 Mean-field method

Let  $\mathcal{P}([q]^n)$  denote the space of probability measures on  $[q]^n$ . For any two measures  $\mu, \nu \in \mathcal{P}([q]^n)$  define the Kullback–Leibler divergence between  $\mu$  and  $\nu$  by

$$D(\mu||\nu) := \sum_{y \in [q]^n} \mu(y) \log \mu(y) - \sum_{y \in [q]^n} \mu(y) \log \nu(y),$$

where  $0 \log 0 = 0$  and  $\log 0 = -\infty$  by convention.

Then, for any  $q \in \mathcal{P}([q]^n)$  an easy computation gives

$$D(q||\mu_n^{\beta,B}) = \Phi_n(\beta, B) + \sum_{y \in [q]^n} q(y) \log q(y) - \sum_{y \in [q]^n} q(y) \mathbf{H}_n^{\beta,B}(y),$$

where

$$\mathbf{H}_n^{\beta,B}(y) := \beta_n \sum_{(i,j) \in E_n} \delta(y_i, y_j) + B \sum_{i \in [n]} \delta(y_i, 1).$$

Since  $D(q||\mu_n^{\beta,B}) \geq 0$ , with equality iff  $q = \mu_n^{\beta,B}$ , we get

$$\Phi_n(\beta, B) = \sup_{q \in \mathcal{P}([q]^n)} \left\{ \sum_{y \in [q]^n} q(y) \mathbf{H}_n^{\beta,B}(y) - \sum_{y \in [q]^n} q(y) \log q(y) \right\}. \tag{1.2}$$

In literature (1.2) is known as the *variational formula* for the log partition function  $\Phi_n(\beta, B)$ . From (1.2) one can obtain a lower bound on  $\Phi_n(\beta, B)$  by restricting the supremum in (1.2) to product measures, i.e.  $q = \prod_{i \in [n]} q_i \in \mathcal{P}([q]^n)$ . Therefore

$$\Phi_n(\beta, B) \geq \sup_{q \in \mathcal{P}([q]^n)} \mathbf{M}_n^{\beta,B}(q), \tag{1.3}$$

where

$$\mathbf{M}_n^{\beta,B}(q) := \left\{ \beta_n \sum_{(i,j) \in E_n} \sum_{r \in [q]} q_i(r) q_j(r) + B \sum_{i \in [n]} q_i(1) - \sum_{i \in [n], r \in [q]} q_i(r) \log q_i(r) \right\}. \tag{1.4}$$

The RHS of (1.3) is referred as the *mean-field approximation* for the log-partition function  $\Phi_n(\beta, B)$ .

Since the supremum in (1.3) is much more tractable than the one in (1.2), it is therefore naturally interesting to find graph sequences for which (1.3) is asymptotically tight. For the complete graph it has been long known that the mean-field prediction is indeed tight for both Ising and Potts measure (see [28–30]). However, for *locally*

*tree-like graphs* (see [23, Definition 1.1]) this is not the case. Indeed, in [20] it is shown that the *Bethe prediction* is the correct answer for Ising measures on such graphs when the limiting tree is a Galton–Watson tree whose off-spring distribution have a finite variance. In [26] it was extended for power law distribution, and finally in [23] it was extended to full generality. Moreover the same was shown be true for the Potts model on regular graphs in [23,24].

For the complete graph on  $n$  vertices one has  $\Theta(n^2)$  edges, whereas locally tree-like graphs has only  $O(n)$  edges (see Definition 1.2 for  $O(\cdot)$ , and  $\Theta(\cdot)$ ). Therefore, it is natural to ask for graph sequences such that  $n \ll |E_n| \ll n^2$ , if one of the two predictions is correct for the limiting log partition function. Very few results are known about the asymptotics of the log partition function in this regime. See however [9, Theorem 2.10] which in particular shows that if a sequence of graphs converges in  $L^p$  cut metric, then corresponding log partition functions converge. Also, it follows from [13, Theorem 1] that the mean field approximation is correct for the limiting log-partition function of Potts models on a sequence of growing graphs in  $\mathbb{Z}^d$ , when  $d$  goes to  $\infty$  as well. We re-derive both these results to demonstrate flexibility of our approach (see Theorem 2.4 and Example 1.3.1(d) respectively).

In this paper, we consider Ising and Potts measures (we consider a slightly generalized version of standard Potts model, see Definition 1.1) on graphs with growing sizes such that  $|E_n|/n \rightarrow \infty$ , as  $n \rightarrow \infty$ , and show that the asymptotic log partition function can be expressed as a *variational problem* (see Theorem 1.1). Building on Theorem 1.1, and focusing on *asymptotically regular graphs*, we prove the universality of the limiting log partition function in the ferromagnetic domain, and confirm that it matches with the one obtained from the complete graph (see Theorem 2.1). We further derive asymptotic log partition function for bi-regular bipartite graphs (see Theorem 2.3). Recently, in [9] the asymptotic log partition function was derived for graph sequences converging in *cut metric*. As a byproduct of Theorem 1.1 we give an alternate proof of the same (see Sect. 2.3). For an outline of the proof techniques of the results we refer the reader to Sect. 1.4.

### 1.2 Statement of main theorem

We will work with the following slightly general version of the Potts model.

**Definition 1.1** For  $q \geq 2$ , let  $\mathbf{J}, \mathbf{h}$  be a symmetric  $q \times q$  matrix, and a vector of length  $q$  respectively. Also let  $A_n$  be a real symmetric  $n \times n$  matrix. We define a hamiltonian  $H_n^{\mathbf{J},\mathbf{h}}(\cdot)$  on  $[q]^n$  by setting

$$H_n^{\mathbf{J},\mathbf{h}}(\mathbf{y}) := \frac{1}{2} \sum_{i,j=1}^n A_n(i, j) \sum_{r,s=1}^q J_{rs} \delta(y_i, r) \delta(y_j, s) + \sum_{i=1}^n \sum_{r=1}^q h_r \delta(y_i, r), \tag{1.5}$$

where  $\mathbf{y} := (y_1, \dots, y_n)$ . Using  $H_n^{\mathbf{J},\mathbf{h}}(\cdot)$  we now define the following probability measure on  $[q]^n$ :

$$\mu_n^{\mathbf{J},\mathbf{h}}(\mathbf{y}) := \frac{1}{Z_n(\mathbf{J}, \mathbf{h})} \exp(H_n^{\mathbf{J},\mathbf{h}}(\mathbf{y})), \tag{1.6}$$

where

$$Z_n(\mathbf{J}, \mathbf{h}) := \sum_{y \in [q]^n} e^{\mathbf{H}_n^{\mathbf{J}, \mathbf{h}}(y)}.$$

Considering  $\mathbf{J}$  to be the identity matrix  $I_q$ ,  $\mathbf{h} = B(1, 0, 0, \dots, 0)$ , and  $A_n$  to be the adjacency matrix of  $\mathbf{G}_n$  divided by  $2|E_n|/n$ , we see that the probability measure  $\mu_n^{\mathbf{J}, \mathbf{h}}$  in (1.6) is a generalized version of the standard Potts measure  $\mu_n^{\beta, B}$ . Throughout most of the article, we will fix a choice of  $\{A_n\}_{n \in \mathbb{N}}$ ,  $\mathbf{J}$ , and  $\mathbf{h}$ . Therefore, to lighten the notation we will often write  $\mu_n(\cdot)$  instead of  $\mu_n^{\mathbf{J}, \mathbf{h}}(\cdot)$ .

Now similarly as before we define the log partition function

$$\Phi_n(\mathbf{J}, \mathbf{h}) := \log Z_n(\mathbf{J}, \mathbf{h}).$$

Arguing same as before we also obtain that

$$\Phi_n(\mathbf{J}, \mathbf{h}) = \sup_{q \in \mathcal{P}([q]^n)} \left\{ \sum_{y \in [q]^n} q(y) \mathbf{H}_n^{\mathbf{J}, \mathbf{h}}(y) - \sum_{y \in [q]^n} q(y) \log q(y) \right\}, \tag{1.7}$$

and

$$\Phi_n(\mathbf{J}, \mathbf{h}) \geq \sup_{q \in \mathcal{P}([q]^n)} M_n^{\mathbf{J}, \mathbf{h}}(q), \tag{1.8}$$

where

$$M_n^{\mathbf{J}, \mathbf{h}}(q) := \left\{ \frac{1}{2} \sum_{i, j=1}^n A_n(i, j) \sum_{r, s=1}^q q_i(r) q_j(s) J_{rs} + \sum_{i=1}^n \sum_{r=1}^q h_r q_i(r) - \sum_{i=1}^n \sum_{r=1}^q q_i(r) \log q_i(r) \right\}. \tag{1.9}$$

In Theorem 1.1 below we show that under a fairly general condition (1.8) is actually tight as  $n \rightarrow \infty$ . Before going to the statement of Theorem 1.1, for convenience of writing, first let us introduce the following notation:

**Definition 1.2** Let  $a_n$  and  $b_n$  be two non-negative sequences of real numbers. We write  $a_n = o(b_n)$  if  $\lim_{n \rightarrow \infty} \frac{a_n}{b_n} = 0$ , whereas  $a_n = O(b_n)$  implies  $\limsup_{n \rightarrow \infty} \frac{a_n}{b_n} < \infty$ . Note that  $a_n = O(b_n)$  includes the possibility of  $a_n = o(b_n)$ . Next we use the notation  $a_n = \Theta(b_n)$ , if  $a_n = O(b_n)$  and  $b_n = O(a_n)$ .

Note that for both Ising and Potts model we must assume some conditions on  $A_n$  to ensure that the resulting log partition is  $O(n)$ , or equivalently the limiting log partition function to be non-trivial. In this paper we work with the following condition:

$$\sup_{\mathbf{x} \in [0, 1]^n} \sum_{i \in [n]} \left| \sum_{j \in [n]} A_n(i, j) x_j \right| = O(n). \tag{1.10}$$

Now let us denote  $\|J\|_\infty := \max_{r,s \in [q]} |J_{r,s}|$  and  $\|h\|_\infty := \max_{r \in [q]} |h_r|$ . Since

$$\begin{aligned} |H_n^{J,h}(y)| &\leq \frac{\|J\|_\infty}{2} \sum_{i \in [n], r,s \in [q]} \left| \sum_{j \in [n]} A_n(i, j) \delta(y_j, s) \right| + \|h\|_\infty \sum_{i \in [n]} \sum_{r \in [q]} \delta(y_i, r) \\ &\leq \frac{\|J\|_\infty}{2} \sum_{r,s \in [q]} \sup_{x \in \{0,1\}^n} \sum_{i \in [n]} \left| \sum_{j \in [n]} A_n(i, j) x_j \right| + n \|h\|_\infty, \end{aligned}$$

it follows by (1.10) that  $|\sup_{y \in [q]^n} H_n^{J,h}(y)| = O(n)$ , which implies  $\Phi_n(J, h) = O(n)$  as well.

When all entries of  $A_n$  have the same sign, condition (1.10) is equivalent to

$$\|A_n\|_1 := \sum_{i,j \in [n]} |A_n(i, j)| = O(n).$$

If (1.10) does not hold then there exists  $J, h$  such that the resulting log partition function  $\Phi_n(J, h)$  scales super linearly. For example, if all entries of  $A_n$  are positive,  $J = \beta I_q$ , then for any  $\beta > 0$  an application of the mean-field lower bound gives  $\lim_{n \rightarrow \infty} \frac{1}{n} \Phi_n(\beta, B) = +\infty$ , thus proving that (1.10) is necessary for the log partition function to be  $O(n)$  in general. If  $A_n$  has both positive and negative entries, (1.10) continues to hold for many well-known models with both positive and negative entries, such as the Sherrington–Kirkpatrick model and Hopfield model (see Sect. 1.3).

Of course we do not expect the mean-field approximation to hold for all matrices  $A_n$  satisfying (1.10). For example, it is known that the mean-field approximation is not correct for the Sherrington–Kirkpatrick model [45], or Ising models on sparse graphs [21]. With this in mind we introduce the following definition.

**Definition 1.3** Suppose  $A_n$  is a sequence of symmetric  $n \times n$  matrices satisfying (1.10). We say that  $A_n$  satisfies the *mean-field* assumption if  $\text{tr}(A_n^2) = o(n)$ .

Now we are ready to state our first result.

**Theorem 1.1** *If  $A_n$  satisfies the mean-field assumption, then*

$$\lim_{n \rightarrow \infty} \frac{1}{n} \left[ \Phi_n(J, h) - \sup_{q \in \mathcal{P}(\{q\})^n} M_n^{J,h}(q) \right] = 0.$$

Theorem 1.1 essentially says that if  $A_n$  is a sequence of matrices which satisfies the mean-field assumption then the mean-field approximation gives the right answer for the log partition function upto an error which is  $o(n)$ .

As an application of Theorem 1.1, one immediately obtains the following corollary. This corollary will be used in all of our applications involving graphs.

**Corollary 1.2** *Suppose  $G_n$  is a sequence of simple graphs, and  $A_n$  is the adjacency matrix of  $G_n := ([n], E_n)$  multiplied by  $n/(2|E_n|)$ , where  $|E_n|$  is the number of edges. Then the conclusion of Theorem 1.1 holds if  $n = o(|E_n|)$ .*

*Proof* Since

$$\sup_{x \in [0,1]} \sum_{i \in [n]} \left| \sum_{j \in [n]} A_n(i, j)x_j \right| = \sum_{i, j \in [n]} A_n(i, j) = n,$$

(1.10) holds. Also we have

$$\frac{1}{n} \sum_{i, j \in [n]} A_n(i, j)^2 = \frac{n}{2|E_n|^2} |E_n| = O\left(\frac{n}{|E_n|}\right) = o(1),$$

and so  $A_n$  satisfies the mean-field assumption. The conclusion then follows by Theorem 1.1. □

Below we consider few different choices of  $A_n$ , and verify for which of those the mean-field assumption is satisfied.

### 1.3 Examples

This is broadly divided into two categories.

#### 1.3.1 Matrices $A_n$ which are scaled adjacency graphs

- (a) Let  $\mathbf{G}_n$  be any sequence of simple dense labeled graphs on  $n$  vertices, i.e. it has  $\Theta(n^2)$  edges. Let  $A_n$  be adjacency matrix of  $\mathbf{G}_n$  scaled by  $n$ , i.e.  $A_n(i, j) := \frac{1}{n} 1_{(i,j) \in E_n}$ . Since this scaling is equivalent to the scaling proposed in Corollary 1.2, it suffices to check that  $n = o(|E_n|)$ . But this is immediate as  $|E_n| = \Theta(n^2)$ .
- (b) Let  $\mathbf{G}_n$  be a  $d_n$  regular graph, and  $A_n(i, j) := \frac{1}{d_n} 1_{(i,j) \in E_n}$ . In this case again the scaling is the same one as that of Corollary 1.2, and so it suffices to check that  $n = o(|E_n|)$ . Since  $2|E_n| = nd_n$ , Corollary 1.2 holds iff  $d_n \rightarrow \infty$ .
- (c) Let  $\mathbf{G}_n$  be an Erdős-Rényi random graph with parameter  $p_n$ . Setting  $A_n(i, j) := \frac{1}{np_n} 1_{(i,j) \in E_n}$  it again suffices to check by Corollary 1.2 that  $n = o(|E_n|)$ , in probability. Since  $|E_n|$  has  $\text{Bin}\binom{n}{2}, p_n$  distribution,

$$\frac{2|E_n|}{n^2 p_n} \rightarrow 1, \text{ in probability,}$$

as soon as  $n^2 p_n \rightarrow \infty$ , the mean-field assumption holds in probability iff  $np_n \rightarrow \infty$ . In particular the mean-field condition does not hold if  $p_n = \frac{\lambda}{n}$  for some  $\lambda < \infty$ .

- (d) Let  $\mathbf{G}_n^{(d)}$  be the  $[-n^{1/d}, n^{1/d}]^d$  box of the  $d$ -dimensional integer lattice  $\mathbb{Z}_d$ . Physicists have long been interested in studying Ising and Potts models on lattices (see [40,47], and the references therein). For any finite  $d$ , setting  $A_n^{(d)}(i, j) := \frac{1}{d} 1_{\{(i, j) \in E_n\}}$  we note that  $\frac{1}{n} \text{tr}((A_n^{(d)})^2) = O(\frac{1}{d})$ , and thus the sequence does

not satisfy the mean-field assumption. So our results are not applicable on  $\mathbb{Z}^d$  for finite  $d$ . However, if we allow  $d$  to go to infinity (at any rate) along with  $n$ , then Corollary 1.2 is applicable. One can check that this also implies that if we let  $d \rightarrow \infty$  after letting  $n \rightarrow \infty$ , the same conclusion continues to hold. Behavior of limiting log-partition function for the Potts model on  $\mathbb{Z}^d$  for large  $d$  has been studied in [6, 13]. We recover their results as an application of Corollary 1.2

### 1.3.2 Matrices with both positive and negative entries

A general sufficient condition for (1.10) to hold is  $\|A_n\| := \sup_{\mathbf{x}: \|\mathbf{x}\|_2=1} \|A_n \mathbf{x}\|_2 = O(1)$ . To see this note that an application of Cauchy–Schwarz inequality gives

$$\sup_{\mathbf{x} \in [0, 1]^n} \sum_{i \in [n]} \left| \sum_{j \in [n]} A_n(i, j)x_j \right| \leq \sqrt{n} \sup_{\mathbf{x} \in [0, 1]^n} \|A_n \mathbf{x}\|_2 \leq \sqrt{n} \|A_n\| \sup_{\mathbf{x} \in [0, 1]^n} \|\mathbf{x}\|_2 = O(n).$$

- (a) Let  $A_n$  be a symmetric matrix with 0 on the diagonal, and  $A_n(i, j) = \frac{1}{\sqrt{n}}Z(i, j)$  with

$$\{Z(i, j)\}_{1 \leq i < j \leq \infty} \stackrel{i.i.d.}{\sim} N(0, 1).$$

This is the celebrated Sherrington–Kirkpatrick model of statistical physics introduced in [43]. Since  $\|A_n\| = O(1)$ , in probability, in this case (see [3, Theorem 2.12]), (1.10) holds. However  $A_n$  does not satisfy the mean-field assumption, as

$$\frac{1}{n} \sum_{i, j \in [n]} A_n(i, j)^2 = \frac{1}{n^2} \sum_{i, j \in [n]} Z(i, j)^2 \rightarrow 1, \text{ in probability.}$$

This is expected, as the log partition function in this case is given by the Parisi formula, and not by the mean-field approximation.

- (b) Let  $\eta$  be an  $n \times m$  matrix of i.i.d. random variables with  $\mathbb{P}(\eta_{ik} = \pm 1) = \frac{1}{2}$ , and let

$$A_n(i, j) = \frac{1}{n} \sum_{k \in [m]} \eta_{ik} \eta_{jk}.$$

This is the Hopfield model of neural networks, first introduced in [33]. In this case also one has  $\|A_n\| = O(1)$ , in probability, when  $m = \Theta(n)$  (see [3, Section 2.2.2]), and therefore (1.10) holds. Proceeding to check the mean-field condition one has

$$\begin{aligned} \frac{1}{n} \mathbb{E} \sum_{i, j \in [n]} A_n(i, j)^2 &= \frac{1}{n^3} \sum_{i, j \in [n], k, l \in [m]} [\delta(i, j) + \delta(k, l) - \delta(i, j)\delta(k, l)] \\ &= \frac{nm^2 + n^2m - mn}{n^3}, \end{aligned}$$

and so the mean-field condition does not hold for  $m = \Theta(n)$ .



### 1.4 Proof technique

Establishing the conclusion of Theorem 1.1 for graphs whose adjacency matrix has a single dominant eigenvalue is much easier, since in that case the behavior of the log partition function is governed by that eigenvalue. This is indeed the case for Erdős-Rényi random graphs on  $n$  vertices with parameter  $p_n$  such that  $np_n \gg \log n$ . For example, in this regime the largest eigenvalue equals  $np_n(1 + o(1))$  (see [36, Section 1]), whereas the second largest eigenvalue is  $o(np_n)$  (see [32, Theorem 1.1]), providing a spectral gap. Similarly for random  $d_n$ -regular graphs on  $n$  vertices, one also has a spectral gap, as long as  $d_n \geq (\log n)^\gamma$  for some  $\gamma$  positive (see [17, 19]). More generally, any *expander graph* has a spectral gap, and therefore for such graphs one can show that the mean-field approximation is asymptotically tight. However, there are many graphs which are not expanders, such as the  $d$ -dimensional hypercube  $\{0, 1\}^d$  with  $d \rightarrow \infty$ . In this case the number of vertices in the graph is  $n = 2^d$ , and it is well known that the set of eigenvalues are  $\{d - 2i, 0 \leq i \leq d\}$  with multiplicity of  $d - 2i$  being  $\binom{d}{i}$ . Thus the two largest eigenvalues are  $d$  and  $d - 2$  whose ratio converges to 1 as  $d$  becomes large, and consequently there is no dominant eigenvalue.

Even though there is no spectral gap in the hypercube, it is still the case that the number of *big* eigenvalues is small. For example, the largest eigenvalue is  $d$ , and the proportion of eigenvalues that lie outside the interval  $[-d\delta, d\delta]$ , for any  $\delta > 0$ , equals

$$\frac{1}{n} \sum_{i=1}^n 1\{|d - 2i| > d\delta\} = \mathbb{P} \left( \left| \frac{1}{d} \sum_{i \in [d]} B_i - \frac{1}{2} \right| > \delta \right),$$

where  $\{B_i\}_{i \in [d]}$  are i.i.d. Bernoulli random variables with  $\mathbb{P}(B_i = 0) = \mathbb{P}(B_i = 1) = .5$ . By weak law of large numbers the RHS above is  $o(1)$ , as  $d \rightarrow \infty$ , and so the proportion of eigenvalues which are comparable to the leading eigenvalue is  $o(1)$ . Our proof makes this precise proving Theorem 1.1 which covers not just the hypercube, but any sequence of graphs  $\mathbf{G}_n$  satisfying  $n = o(|E_n|)$  (see Corollary 1.2). In fact the main condition of Theorem 1.1. i.e. the condition  $\text{tr}(A_n^2) = o(n)$ , can be rewritten as

$$\frac{1}{n} \sum_{i=1}^n \lambda_i(A_n)^2 = o(1),$$

which says that the (properly scaled) empirical eigenvalue distribution converges to 0 in  $L^2$ . And of course, as already pointed out that the mean-field approximation does not hold in general when  $|E_n| = \Theta(n)$ , thus demonstrating that the conditions of Theorem 1.1, and Corollary 1.2 are tight.

The main tool in the proof of Theorem 1.1 is a modified version of [16, Theorem 1.5]. For readers not familiar with [16], we informally describe the theorem and the ideas behind the proof of [16, Theorem 1.5]. Before proceeding, we define the notion of a *net* of a set.

**Definition 1.4** For any  $S \subset \mathbb{R}^n$  and  $\varepsilon > 0$ , a set  $\tilde{S} \subset \mathbb{R}^n$  is said to be a  $\varepsilon$  net of  $S$ , if given  $s \in S$  there exists (at least one)  $\tilde{s} \in \tilde{S}$  such that  $\|s - \tilde{s}\|_2 \leq \varepsilon$ .

The theorem assumes that  $f : [0, 1]^n \mapsto \mathbb{R}$  is a smooth function such that the set  $\{\nabla f(\mathbf{u}) : \mathbf{u} \in \{0, 1\}^n\}$  has an  $\sqrt{n}\varepsilon$  net  $\mathcal{D}_n(\varepsilon)$  with  $\log |\mathcal{D}_n(\varepsilon)| = o(n)$ , and concludes that

$$\log \sum_{\mathbf{u} \in \{0, 1\}^n} e^{f(\mathbf{u})} = \sup_{\mathbf{u} \in [0, 1]^n} \{f(\mathbf{u}) - I_n(\mathbf{u})\} + o(n),$$

where  $I_n(\mathbf{u}) := \sum_{i=1}^n u_i \log u_i + (1 - u_i) \log(1 - u_i)$  is the binary entropy function, and  $\mathbf{u} := (u_1, \dots, u_n)$ .

For the proof, they introduce a measure  $\nu_n(\cdot)$  on  $\{0, 1\}^n$  given by  $\nu_n(\mathbf{u}) \propto \exp(f(\mathbf{u}))$  for  $\mathbf{u} := (u_1, u_2, \dots, u_n) \in \{0, 1\}^n$ . First it is argued that  $f(\mathbf{u})$  and  $f(\widehat{\mathbf{u}})$  are close on a set with *high probability* under  $\nu_n(\cdot)$ , say  $\mathcal{A}_n$  (see [16, Lemma 3.1]). Here  $\widehat{u}_i$  is conditional expectation of  $u_i$ , conditioned on everything else. Therefore  $\sum_{\mathbf{u} \in \{0, 1\}^n} \exp(f(\mathbf{u}))$  can be well approximated by  $\sum_{\mathbf{u} \in \mathcal{A}_n} \exp(f(\widehat{\mathbf{u}}))$ . Turning to evaluate the latter summation, it is further noted that  $g(\mathbf{u}, \widehat{\mathbf{u}})$ , and  $I_n(\widehat{\mathbf{u}})$  are also close on  $\mathcal{A}_n$  (see [16, Lemma 3.2]), where for  $\mathbf{u} \in [0, 1]^n$ , and  $\mathbf{w} \in (0, 1)^n$ ,

$$g(\mathbf{u}, \mathbf{w}) := \sum_{i \in [n]} u_i \log w_i + (1 - u_i) \log(1 - w_i), \text{ and } I_n(\mathbf{w}) := g(\mathbf{w}, \mathbf{w}).$$

Therefore one only needs to control  $\sum_{\mathbf{u} \in \mathcal{A}_n} \exp(f(\widehat{\mathbf{u}}) + g(\mathbf{u}, \widehat{\mathbf{u}}) - I_n(\widehat{\mathbf{u}}))$ . To control the above, the summation over  $\mathcal{A}_n$  is broken into smaller sets where each sum is over only those  $\mathbf{u}$  for which  $\widehat{\mathbf{u}} \approx \mathbf{p}$ , for some  $\mathbf{p} \in [0, 1]^n$ . Next instead of summing over all choices of  $\mathbf{p} \in [0, 1]^n$ , the sum is restricted on the  $\sqrt{n}\varepsilon$ -net of the image of the map  $\mathbf{u} \mapsto \widehat{\mathbf{u}}$ , using the set  $\mathcal{D}_n(\varepsilon)$ . Thus one obtains

$$\begin{aligned} \log \sum_{\mathbf{u} \in \mathcal{A}_n} \exp(f(\widehat{\mathbf{u}}) + g(\mathbf{u}, \widehat{\mathbf{u}}) - I_n(\widehat{\mathbf{u}})) &\approx \log \sum_{\mathbf{p} \in \mathcal{D}_n(\varepsilon)} \sum_{\mathbf{u} : \widehat{\mathbf{u}} \approx \mathbf{p}} \\ &\exp(f(\mathbf{p}) + g(\mathbf{u}, \mathbf{p}) - I_n(\mathbf{p})). \end{aligned} \tag{1.11}$$

Finally noting that

$$\sum_{\mathbf{u} \in \{0, 1\}^n} e^{g(\mathbf{u}, \mathbf{p})} = 1,$$

the proof follows as the size of  $\mathcal{D}_n(\varepsilon)$  is sub-exponential.

In our proof we follow the same scheme. However, there are several challenges that we had to overcome to apply this idea in our set-up. First, we need to find a net  $\mathcal{D}_n(\varepsilon)$  with appropriate properties. In our set-up, we need to find a  $\sqrt{n}\varepsilon$ -net  $\mathcal{D}_n(\varepsilon)$  of the set  $\{A_n \mathbf{v} : \mathbf{v} \in \{0, 1\}^n\}$ . Since we have very limited assumptions on the structure of  $A_n$ , obtaining a  $\sqrt{n}\varepsilon$ -net is not straightforward. The main difficulty comes from the fact that the eigenvalues of  $A_n$  can be unbounded. To overcome this, we split the range of the eigenvalues into its *level sets*, and then we choose nets of varying size across each of the level sets (for more details see proof of Lemma 3.4).

Equipped with Lemma 3.4, a direct application of [16, Theorem 1.5] proves Theorem 1.1 for graphs  $\mathbb{G}_n$  such that

$$\limsup_{n \rightarrow \infty} n \sum_{i \in [n]} \left( \frac{d_i(\mathbf{G}_n)}{\sum_{j \in [n]} d_j(\mathbf{G}_n)} \right)^2 < \infty, \quad (1.12)$$

where  $\{d_1(\mathbf{G}_n), \dots, d_n(\mathbf{G}_n)\}$  are the degrees of  $\mathbf{G}_n$ . The hypercube does satisfy this condition, as does any regular graph. There are many graphs in literature such that  $n = o(|E_n|)$ , but (1.12) does not hold. For example, let  $\mathbf{G}_n$  denote the complete bipartite graph  $K_{a_n, n-a_n}$ , where  $a_n$  is a sequence of natural numbers going to  $\infty$  such that  $a_n = o(n)$ . In this case the LHS of (1.12) equals

$$\frac{n[a_n(n-a_n)^2 + (n-a_n)a_n^2]}{4a_n^2(n-a_n)^2} = O\left(\frac{n}{a_n}\right),$$

which is not  $O(1)$ , as  $a_n = o(n)$ . Since  $|E_n| = a_n(n-a_n)$  with  $a_n \rightarrow \infty$ , Corollary 1.2 is still applicable for  $K_{a_n, n-a_n}$  but [16, Theorem 1.5] does not apply.

To remove the requirement of (1.12) we modify the proofs of [16, Lemma 3.1], and [16, Lemma 3.2]. In the proof of these two lemmas, at many places, supremum norm bound is used for several functions. The condition (1.12) arises because of that. Instead, we carefully use the assumption (1.10), and the fact that the hamiltonian in our set-up is a quadratic function. This part of the proof has been inspired from [15].

In Sect. 2 we provide several applications of Theorem 1.1. One of which is the computation of the limit for asymptotically regular graphs. To be more precise, we call a sequence of graphs to be asymptotically regular if the empirical distribution of the row sums of the properly scaled adjacency matrix converges to  $\delta_1$ , and if its mean also converges to one. Using a truncation argument we derive the desired result. We also find the limit for bi-regular bipartite graphs, for which we carefully analyze the solutions of some fixed point equations. Lastly, we identify the limit for a sequence of simple graphs converging in cut metric. This follows from a straightforward analysis upon using Theorem 1.1.

## 1.5 Outline

The outline of the rest of the paper is as follows. As applications of Theorem 1.1, in Sect. 2 we derive the asymptotics of the log partition function for ferromagnetic Potts models on asymptotically regular graphs, that of Ising models (both ferromagnetic and anti-ferromagnetic) on bi-regular bipartite graphs, and that of Potts model on a sequence of simple graphs converging in cut metric in the  $L_p$  sense. Section 3 carries out the proof of Theorem 1.1 using three auxiliary lemmas, whose proofs are deferred to Sect. 4. Finally in Sect. 5 we prove the results appearing in Sect. 2.

## 2 Applications of Theorem 1.1

### 2.1 Asymptotically regular graphs

In Theorem 1.1 we saw that the mean-field prediction is asymptotically correct when  $A_n$  satisfies the mean-field condition. However, computing the supremum of  $M_n^{J,h}(q)$

may often be very hard for general matrices  $A_n$ . Restricting ourselves to the case  $\mathbf{J} = \beta I_q$  for  $\beta > 0$ , in Theorem 2.1 below we show that when the matrices  $A_n$  are ‘‘asymptotically regular’’ one can write the  $n$ -dimensional supremum as a one-dimensional supremum, and thereby providing more tractable form of the limit. In particular, setting  $h_r = B\delta(r, 1)$ , for asymptotically regular graphs the limit is same as the one obtained for a Curie–Weiss Potts model.

**Theorem 2.1** (a) *Let  $A_n$  satisfies the mean-field assumption, and each entry of  $A_n$  is non-negative. Also let  $\mathbf{J} = \beta I_q$ , for some  $\beta \geq 0$ . Set  $\mathcal{R}_n(i) := \sum_{j=1}^n A_n(i, j)$ . If*

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n \delta_{\mathcal{R}_n(i)} \rightarrow \delta_1, \text{ in distribution,} \tag{2.1}$$

and

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n \mathcal{R}_n(i) = 1, \tag{2.2}$$

then

$$\lim_{n \rightarrow \infty} \frac{1}{n} \Phi_n(\mathbf{J}, \mathbf{h}) = \sup_{q \in \mathcal{P}([q])} \left[ \frac{\beta}{2} \sum_{r=1}^q q(r)^2 - \sum_{r=1}^q q(r) \log q(r) + \sum_{r=1}^q h_r q(r) \right]. \tag{2.3}$$

- (b) *In particular, the conclusion of part (a) applies in the following two cases:*
- (i)  $\mathbf{G}_n$  is a sequence of  $d_n$  regular graphs with  $d_n \rightarrow \infty$ , and  $A_n = \frac{1}{d_n} 1_{(i,j) \in E_n}$ .
  - (ii)  $\mathbf{G}_n$  is an Erdős–Rényi random graph with parameter  $p_n$  such that  $np_n \rightarrow \infty$ , and  $A_n = \frac{1}{np_n} 1_{(i,j) \in E_n}$ .

As an application of the above theorem, the following theorem derives the large deviation for the empirical measure  $L_n$  on  $\mathcal{P}([q])$  defined by

$$L_n(r) := \frac{1}{n} \sum_{i \in [n]} \delta(y_i, r).$$

Below we recall a few definitions of large deviation theory which are necessary for our paper.

**Definition 2.1** Let  $(\mathcal{X}, \mathcal{B})$  be a measure space equipped with a topology such that every open set is in  $\mathcal{B}$ . A function  $I : \mathcal{X} \mapsto [0, \infty]$  is said to be a rate function if it is lower semi continuous, i.e. for every  $\alpha < \infty$  the set  $\{x \in \mathcal{X} : I(x) \leq \alpha\}$  is closed. The function  $I$  is said to be a good rate function, if further the set  $\{x \in \mathcal{X} : I(x) \leq \alpha\}$  is compact as well. In particular if  $\mathcal{X}$  is compact, any rate function is a good rate function.

A sequence of probability measures  $\mathbb{P}_n$  on  $(\mathcal{X}, \mathcal{B})$  is said to satisfy a large deviation on  $\mathcal{X}$  with respect to a good rate function  $I(\cdot)$ , at speed  $n$ , if for every closed set  $F$ , and open set  $U$ , we have

$$\begin{aligned} \limsup_{n \rightarrow \infty} \frac{1}{n} \log \mathbb{P}_n(F) &\leq - \inf_{x \in F} I(x), \\ \liminf_{n \rightarrow \infty} \frac{1}{n} \log \mathbb{P}_n(U) &\geq - \inf_{x \in U} I(x). \end{aligned}$$

The large deviation reduces the concentration of measure problem to an optimization problem involving the rate function. Next we introduce a few notations which will be needed while solving this optimization problem.

**Definition 2.2** For  $\beta > 0, B \neq 0$  let  $m_{\beta,B}$  denote the unique solution of  $m = \tanh(\beta m + B)$  with the same sign as that of  $B$ . For  $\beta > 1, B = 0$  let  $m_{\beta,0}$  denote the unique positive root of the equation  $m = \tanh(\beta m)$ . The assertions about the roots of the equation  $m = \tanh(\beta m + B)$  can be found in [21, Section 1.1.3].

**Theorem 2.2** (a) *In the setting of Theorem 2.1, the sequence of empirical measures  $L_n$  satisfies a large deviation principle on  $\mathcal{P}([q])$  with speed  $n$  with respect to Euclidean topology, with the good rate function  $\tilde{I}_{\beta,h}(\mu) := I_{\beta,h}(\mu) - \min_{\mu \in \mathcal{P}([q])} I_{\beta,h}(\mu)$ , where*

$$I_{\beta,h}(\mu) := \sum_{r \in [q]} \left( \mu_r \log \mu_r - \frac{\beta \mu_r^2}{2} - h_r \mu_r \right).$$

Consequently letting  $K_{\beta,h} := \arg \min_{\mu \in \mathcal{P}([q])} I_{\beta,h}(\mu)$ , for any  $\delta > 0$  we have

$$\limsup_{n \rightarrow \infty} \frac{1}{n} \log \mu_n \left( \min_{\mu \in K_{\beta,h}} \|L_n - \mu\|_{\infty} \geq \delta \right) < 0. \tag{2.4}$$

(b) *Suppose we are in the setting of Theorem 2.1 with  $q = 2$  (which corresponds to Ising model).*

(i) *If  $h_1 - h_2 = 0$  then*

- *For  $\beta \leq 2$ , for any  $\delta > 0$  there exists  $\varepsilon = \varepsilon(\beta, \delta)$  such that for all large  $n$  we have*

$$\mu_n \left( \frac{1}{n} \sum_{i \in [n]} \{ \delta(y_i, 1) - \delta(y_i, 2) \} \in [-\delta, \delta] \right) \geq 1 - e^{-n\varepsilon}.$$

- *For  $\beta > 2$ , for any  $\delta > 0$  there exists  $\varepsilon = \varepsilon(\beta, \delta)$  such that for all large  $n$  we have*

$$\mu_n \left( \frac{1}{n} \sum_{i \in [n]} \{ \delta(y_i, 1) - \delta(y_i, 2) \} \in [m_{\beta/2,0} - \delta, m_{\beta/2,0} + \delta] \right) \geq \frac{1}{2} - e^{-n\varepsilon},$$

$$\mu_n \left( \frac{1}{n} \sum_{i \in [n]} \{ \delta(y_i, 1) - \delta(y_i, 2) \} \in [-m_{\beta/2,0} - \delta, -m_{\beta/2,0} + \delta] \right) \geq \frac{1}{2} - e^{-n\varepsilon},$$

where  $m_{\beta,0}$  is as in Definition 2.2.

(ii) If  $h_1 - h_2 = B \neq 0$ , for any  $\delta > 0$  there exists  $\varepsilon = \varepsilon(\beta, B, \delta)$  such that for all large  $n$  we have

$$\mu_n \left( \frac{1}{n} \sum_{i \in [n]} \{ \delta(y_i, 1) - \delta(y_i, 2) \} \in [m_{\beta/2,B/2} - \delta, m_{\beta/2,B/2} + \delta] \right) \geq 1 - e^{-n\varepsilon},$$

where  $m_{\beta,B}$  is as in Definition 2.2.

*Remark 2.1* Theorem 2.2(b) gives concentration results for  $\frac{1}{n} \sum_{i \in [n]} \{ \delta(y_i, 1) - \delta(y_i, 2) \}$ , for the Ising model, i.e. for the Potts model of (1.1) for  $q = 2$ . If the Ising model is formulated in such a way that the spins take values in  $\{-1, 1\}$ , then one can easily see that the results of Theorem 2.2(b) are equivalent to the exponential concentration of average spin configuration in that set-up. This gives a complete picture for the ferromagnetic Ising model  $\mu_n^{\beta I_2, \mathbf{h}}$  for all choices of the vector  $\mathbf{h}$ , for asymptotically regular graphs. The optimization of  $I_{\beta, \mathbf{h}}$  for general  $q$  for some specific choices of  $\mathbf{h}$  is well known in the literature (see [7, 18, 27, 29, 30]). Using these results similar concentration results can be derived for the Potts model on asymptotically regular graphs, for those choices of  $\mathbf{h}$ . We omit the details.

## 2.2 Ising model on bipartite graphs

This section focuses on the Ising model ( $q = 2$ ) on bipartite graphs.

**Definition 2.3** Let  $\mathbf{G}_{(a,b),(c,d)}$  denote a bi-regular bipartite graph on  $a + b$  labeled vertices, such that the two partite sets have sizes  $a$  and  $b$ , and the common degree of vertices in those two partite sets are  $c$  and  $d$  respectively. Thus we must have  $ac = bd$ , which equals the number of edges.

In particular  $\mathbf{G}_{(a,b),(b,a)}$  denotes the complete bipartite graph with the two partite sets having sizes  $a$  and  $b$ .

**Definition 2.4** For any  $p \in (0, 1)$  and  $\beta \in \mathbb{R}$  set  $\eta_{\beta,p}(\mathfrak{s}) := \tanh(\beta(1 - p) \tanh(\beta p \mathfrak{s}))$ . By elementary calculus it follows that

- (a) For  $\beta^2 p(1 - p) \leq 1$  the equation  $\mathfrak{s} = \eta_{\beta,p}(\mathfrak{s})$  has the unique root 0.
- (b) For  $\beta^2 p(1 - p) > 1$  the equation  $\mathfrak{s} = \eta_{\beta,p}(\mathfrak{s})$  has a unique positive root, denoted hereafter by  $\mathfrak{s}_{\beta,p}$ . Thus the aforementioned equation has three roots, namely 0,  $\mathfrak{s}_{\beta,p}$ , and  $-\mathfrak{s}_{\beta,p}$ . Applying implicit function theorem, we also note that the function  $(\beta, p) \mapsto \mathfrak{s}_{\beta,p}$  is a continuously differentiable in the open set  $\{(\beta, p) : p(1 - p)\beta^2 > 1\}$ .

**Theorem 2.3** Let  $\mathbf{G}_{(a_n, n-a_n), (c_n, d_n)}$  be a sequence of bipartite graphs on  $n$  labeled vertices, such that

$$\lim_{n \rightarrow \infty} \frac{a_n}{n} = p \in (0, 1), \tag{2.5}$$

and  $c_n + d_n \rightarrow \infty$ , as  $n \rightarrow \infty$ . Thus for  $q = 2$ ,  $\mathbf{J} = \beta \mathbf{I}_2$  for some  $\beta \in \mathbb{R}$ ,  $\mathbf{h} = \mathbf{0}$  in (1.6), setting  $A_n$  to be the adjacency matrix of  $\mathbf{G}_{(a_n, n-a_n), (c_n, d_n)}$  scaled by  $c_n + d_n$  we have

(a) If  $\beta^2 p(1 - p) \leq 1$ , then

$$\lim_{n \rightarrow \infty} \Phi_n(\beta, 0) = \frac{\beta p(1 - p)}{2} + \log 2.$$

(b) If  $\beta^2 p(1 - p) > 1$ , then

$$\begin{aligned} \lim_{n \rightarrow \infty} \Phi_n(\beta, 0) = & \frac{\beta p(1 - p)}{2} + \frac{|\beta| p(1 - p)}{2} \mathfrak{s}_{|\beta|, p \mathfrak{s}_{|\beta|, 1-p}} \\ & + p H(\mathfrak{s}_{|\beta|, p}) + (1 - p) H(\mathfrak{s}_{|\beta|, 1-p}), \end{aligned}$$

where  $\mathfrak{s}_{\beta, p}(\cdot)$  is as in Definition 2.4, and  $H(\mathfrak{s}) := -\frac{1+\mathfrak{s}}{2} \log \frac{1+\mathfrak{s}}{2} - \frac{1-\mathfrak{s}}{2} \log \frac{1-\mathfrak{s}}{2}$  for  $\mathfrak{s} \in [-1, 1]$ .

### 2.3 Potts model on converging sequence of graphs in cut metric

The theory of dense graph limits was developed by Borgs, Chayes, Lovasz, and coauthors [10, 11, 37], and has received phenomenal attention over the last few years.

Recent works of Borgs et al [8, 9] have extended this theory beyond the regime of dense graphs. One of the results in [9] is the asymptotics of the log partition function  $\Phi_n(\mathbf{J}, \mathbf{h})$  of (1.6) of a sequence of graphs converging in the sense of cut metric to functions  $W$  that are unbounded. As a byproduct of Theorem 1.1 we are able to provide a short proof of their result. Before going to the statement of the result, we first need to introduce necessary notations, and concepts. These are taken from [8, 9].

**Definition 2.5** A function  $W : [0, 1]^2 \mapsto \mathbb{R}$  is called a symmetric function if  $W(x, y) = W(y, x)$  for all  $x, y \in [0, 1]$ . Any symmetric measurable function  $W : [0, 1]^2 \mapsto \mathbb{R}$  which is  $L^1$  integrable, i.e.  $\|W\|_1 := \int_{[0, 1]^2} |W(x, y)| dx dy < \infty$  is called a graphon.

Given a symmetric  $n \times n$  matrix  $A_n$ , define a graphon on  $[0, 1]^2$  by dividing  $[0, 1]^2$  into  $n^2$  smaller squares each of length  $1/n$ , and setting  $W_{A_n}(x, y) := A_n(i, j)$  if  $(x, y)$  is in the  $(i, j)$ -th box, i.e.  $\lceil nx \rceil = i$ ,  $\lceil ny \rceil = j$ .

The cut norm of a graphon  $W$  is given by

$$\|W\|_{\square} = \left| \sup_{S, T \subset [0, 1]} \int_{S \times T} W(x, y) dx dy \right|.$$

After identifying graphons with cut distance zero, the set of equivalence classes of graphons equipped with the cut metric is a compact metric space. The cut norm is equivalent to the  $L^\infty \mapsto L^1$  operator norm defined by

$$\|W\|_{\infty \mapsto 1} := \sup_{f, g: \|f\|_\infty, \|g\|_\infty \leq 1} \left| \int_{[0,1]^2} W(x, y) f(x) g(y) dx dy \right|.$$

More precisely, we have  $\|W\|_{\square} \leq \|W\|_{\infty \mapsto 1} \leq 4 \|W\|_{\square}$ .

Next we introduce the notion of fractional partition.

**Definition 2.6** A  $q$  tuple of measurable functions  $\rho := (\rho_1, \dots, \rho_q) : [0, 1]^q \mapsto [0, 1]^q$ , such that

$$\sum_{r \in [q]} \rho_r(x) = 1, \forall x \in [0, 1],$$

will be called a fractional partition of  $[0, 1]$  into  $q$  classes. The set of fractional partitions of  $[0, 1]$  into  $q$  classes will be denoted by  $\mathbf{FP}_q$ .

Now we are ready to state the result about the limiting log partition function for a sequence of graphs converging in cut metric.

**Theorem 2.4** Let  $\mathbf{G}_n$  be a sequence of simple graphs, and let  $A_n$  be the adjacency matrix of  $\mathbf{G}_n$  scaled by  $\frac{2|E_n|}{n}$ . If  $W_{nA_n}$  converges in cut metric to a graphon  $W$ , then we have

$$\lim_{n \rightarrow \infty} \frac{1}{n} \Phi_n(\mathbf{J}, \mathbf{h}) = \sup_{\rho \in \mathbf{FP}_q} F^{\mathbf{J}, \mathbf{h}}(W, \rho),$$

where

$$\begin{aligned} F^{\mathbf{J}, \mathbf{h}}(W, \rho) &:= \frac{1}{2} \sum_{r, s \in [q]} J_{rs} \int_{[0,1]^2} \rho_r(x) \rho_s(y) W(x, y) dx dy \\ &+ \sum_{r \in [q]} h_r \int_{[0,1]} \rho_r(x) dx - \int_{[0,1]} \sum_{r \in [q]} \rho_r(x) \log \rho_r(x) dx. \end{aligned}$$

Theorem 2.4 follows from [9, Theorem 2.10], and [9, Lemma 3.2]. In Sect. 5 we give a shorter proof of the same using Corollary 1.2.

### 3 Proof of Theorem 1.1

We begin with a simple lemma which allows us to assume that the entries of  $A_n$  are  $o(1)$ .



**Lemma 3.1** *Let  $A_n$  be a sequence of matrices that satisfies the mean-field assumption. Then there is a sequence of matrices  $\tilde{A}_n$  with 0 diagonal entries which also satisfies the mean-field assumption such that  $\max_{i,j \in [n]} |\tilde{A}_n(i, j)| = o(1)$ , and*

$$|\Phi_n(\mathbf{J}, \mathbf{h}) - \tilde{\Phi}_n(\mathbf{J}, \mathbf{h})| = o(n), \quad \sup_{\mathbf{q} \in \mathcal{P}([q]^n)} |\mathbf{M}_n^{\mathbf{J}, \mathbf{h}}(\mathbf{q}) - \tilde{\mathbf{M}}_n^{\mathbf{J}, \mathbf{h}}(\mathbf{q})| = o(n),$$

where  $\tilde{\Phi}_n(\mathbf{J}, \mathbf{h})$  and  $\tilde{\mathbf{M}}_n^{\mathbf{J}, \mathbf{h}}(\mathbf{q})$  are obtained by replacing  $A_n$  with  $\tilde{A}_n$  in the corresponding definitions.

*Proof* Since  $A_n$  satisfies the mean-field assumption, setting  $\varepsilon_n := n^{-1/2} \sqrt{\text{tr}(A_n^2)}$ , we see that  $\varepsilon_n \rightarrow 0$ . Now defining an  $n \times n$  symmetric matrix  $\tilde{A}_n$  by

$$\tilde{A}_n(i, i) := 0, \quad \tilde{A}_n(i, j) := A_n(i, j) 1_{|A_n(i, j)| \leq \varepsilon_n},$$

one immediately has  $\max_{i,j \in [n]} |\tilde{A}_n(i, j)| \leq \varepsilon_n \rightarrow 0$ . Extending the definition of  $\mathbf{H}_n^{\mathbf{J}, \mathbf{h}}(\cdot)$  to  $\mathcal{P}([q]^n)$  (see Definition 3.1 below for more details), and defining  $\tilde{\mathbf{H}}_n^{\mathbf{J}, \mathbf{h}}$  analogously one has

$$\begin{aligned} \sup_{\mathbf{q} \in \mathcal{P}([q]^n)} \left| \mathbf{H}_n^{\mathbf{J}, \mathbf{h}}(\mathbf{q}) - \tilde{\mathbf{H}}_n^{\mathbf{J}, \mathbf{h}}(\mathbf{q}) \right| &\leq \frac{q \|\mathbf{J}\|_\infty}{2} \left( \sum_{i,j \in [n]} |A_n(i, j)| 1_{|A_n(i, j)| > \varepsilon_n} + \sum_{i \in [n]} |A_n(i, i)| \right) \\ &\leq \frac{q \|\mathbf{J}\|_\infty}{2\varepsilon_n} \sum_{i,j \in [n]} A_n(i, j)^2 + \frac{q \|\mathbf{J}\|_\infty}{2} \sqrt{n \sum_{i \in [n]} A_n(i, i)^2} \\ &\leq \frac{nq \|\mathbf{J}\|_\infty \varepsilon_n}{2} + \frac{q \|\mathbf{J}\|_\infty}{2} \sqrt{n \text{tr}(A_n^2)} = o(n), \end{aligned} \tag{3.1}$$

which immediately implies  $\sup_{\mathbf{q} \in \mathcal{P}([q]^n)} |\mathbf{M}_n^{\mathbf{J}, \mathbf{h}}(\mathbf{q}) - \tilde{\mathbf{M}}_n^{\mathbf{J}, \mathbf{h}}(\mathbf{q})| = o(n)$ . Also we have

$$\begin{aligned} |\Phi_n(\mathbf{J}, \mathbf{h}) - \tilde{\Phi}_n(\mathbf{J}, \mathbf{h})| &= \left| \log \frac{\sum_{\mathbf{y} \in [q]^n} e^{\mathbf{H}_n^{\mathbf{J}, \mathbf{h}}(\mathbf{y})}}{\sum_{\mathbf{y} \in [q]^n} e^{\tilde{\mathbf{H}}_n^{\mathbf{J}, \mathbf{h}}(\mathbf{y})}} \right| \leq \sup_{\mathbf{y} \in [q]^n} |\mathbf{H}_n^{\mathbf{J}, \mathbf{h}}(\mathbf{y}) - \tilde{\mathbf{H}}_n^{\mathbf{J}, \mathbf{h}}(\mathbf{y})| \\ &\leq \sup_{\mathbf{q} \in \mathcal{P}([q]^n)} |\mathbf{H}_n^{\mathbf{J}, \mathbf{h}}(\mathbf{q}) - \tilde{\mathbf{H}}_n^{\mathbf{J}, \mathbf{h}}(\mathbf{q})|, \end{aligned}$$

where the last inequality follows on noting that for any  $\mathbf{y} \in [q]^n$  setting  $\mathbf{q}_i(r) = \delta(y_i, r)$  one has  $\mathbf{q} \in \mathcal{P}([q]^n)$ . Since the RHS above is  $o(n)$  by (3.1), the proof of the lemma is complete.  $\square$

For the remaining of this section and the next, without loss of generality we will assume that diagonal elements of  $A_n$  are 0, and  $\max_{i,j \in [n]} |A_n(i, j)| = o(1)$ . Next we state three lemmas which are necessary for proving Theorem 1.1. First, for ease of writing we introduce a few notations.

**Definition 3.1** For any  $y \in [q]^n$  define the  $nq \times 1$  vector  $\mathbf{x} := \mathbf{x}(y) \in \mathcal{X}_n$  by setting  $x_{ir} := \delta(y_i, r)$ , where

$$\mathcal{X}_n := \left\{ \mathbf{z} \in \{0, 1\}^{nq} : \sum_{r \in [q]} z_{ir} = 1 \text{ for all } i \in [n] \right\}.$$

Let  $m : [0, 1]^{nq} \mapsto [0, 1]^{nq}$  by

$$m_{ir}(\mathbf{z}) := \sum_{s=1}^q J_{rs} \sum_{j=1}^n A_n(i, j) z_{js}.$$

Note that, since diagonal entries of  $A_n$  are zero,  $m_{ir}(\mathbf{z})$  is free of  $\{z_{is}, s \in [q]\}$ . Next for every  $r \in [q]$ , define a map  $T_r : (-\infty, \infty)^q \mapsto (0, 1)$  by

$$T_r(m_1, m_2, \dots, m_q) := \frac{e^{m_r}}{\sum_{s \in [q]} e^{m_s}}.$$

Define another  $nq \times 1$  vector  $\hat{\mathbf{x}}$  by

$$\begin{aligned} \hat{x}_{ir} &:= \mathbb{P}_{\mu_n}(Y_i = r \mid Y_k = y_k, k \neq i) = T_r(m_{i1} + h_1, \dots, m_{iq} + h_q) \\ &= \frac{\exp(m_{ir}(\mathbf{x}) + h_r)}{\sum_{s=1}^q \exp(m_{is}(\mathbf{x}) + h_s)}, \end{aligned}$$

and note that  $\hat{\mathbf{x}} \in \hat{\mathcal{X}}_n$ , where

$$\hat{\mathcal{X}}_n := \left\{ \mathbf{z} \in (0, 1)^{nq} : \sum_{r \in [q]} z_{ir} = 1 \text{ for all } i \in [n] \right\},$$

When  $\mathbf{Y} := (Y_i)_{i \in [n]} \sim \mu_n$ , let  $\mathbf{X}, \hat{\mathbf{X}}$  denote the corresponding random vectors. Finally by a slight abuse of notation for any  $\mathbf{z} \in [0, 1]^{nq}$  let  $H_n^{J,h}(\mathbf{z})$  stand for  $F_n(\mathbf{z}) + \sum_{i \in [n], r \in [q]} h_r z_{ir}$ , where  $F_n : [0, 1]^{nq} \mapsto \mathbb{R}$  is defined by

$$F_n(\mathbf{z}) := \frac{1}{2} \sum_{r,s \in [q], i,j \in [n]} J_{rs} z_{ir} z_{js} A_n(i, j) = \frac{1}{2} \sum_{i \in [n], r \in [q]} m_{ir}(\mathbf{z}) z_{ir} = \frac{1}{2} \sum_{r,s \in [q]} J_{rs} \mathbf{z}'_r A_n \mathbf{z}_s,$$

$\mathbf{z}_r := (z_{ir})_{1 \leq i \leq n} \in \mathbb{R}^n$ , and  $\mathbf{z}'_r$  denotes the transpose of  $\mathbf{z}_r$ . In this notation  $H_n^{J,h}(\mathbf{x}(y))$  is the Hamiltonian of the Potts model at  $y \in [q]^n$  in (1.6).

With this notation we have

**Lemma 3.2** *If  $A_n$  satisfies the mean-field assumption, then*

$$\mathbb{E}_{\mu_n} \left\{ [F_n(\mathbf{X}) - F_n(\hat{\mathbf{X}})]^2 \right\} = o(n^2).$$

**Lemma 3.3** *If  $A_n$  satisfies the mean-field assumption, then*

$$\mathbb{E}_{\mu_n} \left[ \left( \sum_{i \in [n], r \in [q]} (X_{ir} - \hat{X}_{ir}) m_{ir}(\mathbf{X}) \right)^2 \right] = o(n^2), \tag{3.2}$$

and,

$$\mathbb{E}_{\mu_n} \left[ \sum_{r \in [q]} \left( \sum_{i \in [n]} (X_{ir} - \hat{X}_{ir}) \right)^2 \right] = o(n^2). \tag{3.3}$$

Recalling the definition of net (see Definition 1.4) we now state our next lemma.

**Lemma 3.4** *If  $A_n$  satisfies the mean-field assumption, then given any  $\varepsilon > 0$ , there exists a  $\sqrt{n}\varepsilon$ -net  $\mathcal{D}_n(\varepsilon)$  of the set  $\{A_n \mathbf{v} : \mathbf{v} \in [0, 1]^n\}$ , such that*

$$\lim_{n \rightarrow \infty} \frac{1}{n} \log |\mathcal{D}_n(\varepsilon)| = 0. \tag{3.4}$$

We now complete the proof of Theorem 1.1 using Lemmas 3.2, 3.3, and 3.4, deferring the proof of the lemmas to Sect. 4.

*Proof of Theorem 1.1* For  $\mathbf{z} \in [0, 1]^{nq}$ , and  $\mathbf{w} \in (0, 1)^{nq}$  define

$$g_n(\mathbf{z}, \mathbf{w}) := \sum_{i \in [n]} \sum_{r \in [q]} z_{ir} \log w_{ir}, \quad I_n(\mathbf{z}) := g_n(\mathbf{z}, \mathbf{z}).$$

Note that

$$\begin{aligned} g_n(\mathbf{x}, \hat{\mathbf{x}}) - I_n(\hat{\mathbf{x}}) &= \sum_{i \in [n], r \in [q]} (x_{ir} - \hat{x}_{ir}) \log \hat{x}_{ir} = \sum_{i \in [n], r \in [q]} (x_{ir} - \hat{x}_{ir})(m_{ir}(\mathbf{x}) + h_r) \\ &\quad - \sum_{i \in [n], r \in [q]} (x_{ir} - \hat{x}_{ir}) \log \sigma_i, \end{aligned}$$

where

$$\sigma_i := \sum_{s \in [q]} \exp(m_{is}(\mathbf{x}) + h_s).$$

Since for each  $i \in [n]$ ,

$$\sum_{r \in [q]} x_{ir} = \sum_{r \in [q]} \hat{x}_{ir} = 1,$$

we have that

$$g_n(\mathbf{x}, \hat{\mathbf{x}}) - I_n(\hat{\mathbf{x}}) = \sum_{i \in [n], r \in [q]} (x_{ir} - \hat{x}_{ir})(m_{ir}(\mathbf{x}) + h_r).$$

Therefore, from Lemma 3.3 we deduce that

$$\begin{aligned} \mathbb{E}_{\mu_n} \left[ (g_n(\mathbf{X}, \widehat{\mathbf{X}}) - I_n(\widehat{\mathbf{X}}))^2 \right] &\leq 2\mathbb{E}_{\mu_n} \left[ \left( \sum_{i \in [n], r \in [q]} (X_{ir} - \widehat{X}_{ir}) m_{ir}(\mathbf{X}) \right)^2 \right] \\ &\quad + 2 \|\mathbf{h}\|_\infty^2 q \mathbb{E}_{\mu_n} \sum_{r \in [q]} \left[ \left( \sum_{i \in [n]} X_{ir} - \widehat{X}_{ir} \right)^2 \right] = o(n^2), \end{aligned}$$

where we recall  $\|\mathbf{h}\|_\infty = \max_{r \in [q]} |h_r|$ . Similarly, recalling that  $\mathbf{H}_n^{J, \mathbf{h}}(\mathbf{z}) = F_n(\mathbf{z}) + \sum_{i \in [n], r \in [q]} h_r z_{ir}$ , combining Lemmas 3.2, and 3.3, we get

$$\begin{aligned} \mathbb{E}_{\mu_n} \left[ \left( \mathbf{H}_n^{J, \mathbf{h}}(\mathbf{X}) - \mathbf{H}_n^{J, \mathbf{h}}(\widehat{\mathbf{X}}) \right)^2 \right] &\leq 2\mathbb{E}_{\mu_n} \left[ (F_n(\mathbf{X}) - F_n(\widehat{\mathbf{X}}))^2 \right] \\ &\quad + 2 \|\mathbf{h}\|_\infty^2 q \mathbb{E}_{\mu_n} \left[ \sum_{r \in [q]} \left( \sum_{i \in [n]} X_{ir} - \widehat{X}_{ir} \right)^2 \right] = o(n^2). \end{aligned}$$

Hence, applying Markov’s inequality we see that  $\mathbb{P}_{\mu_n}(\mathbf{X} \in \mathcal{A}_n) \geq 1/2$ , where

$$\mathcal{A}_n := \{ \mathbf{x} \in \mathcal{X}_n : |\mathbf{H}_n(\mathbf{x}) - \mathbf{H}_n(\widehat{\mathbf{x}})|, |g_n(\mathbf{x}, \widehat{\mathbf{x}}) - I_n(\widehat{\mathbf{x}})| \leq \delta_n/2 \},$$

for some  $\delta_n = o(n)$ , and

This implies that

$$\begin{aligned} \Phi_n(\mathbf{J}, \mathbf{h}) &\leq \log 2 + \log \left( \sum_{\mathbf{x} \in \mathcal{A}_n} \exp(\mathbf{H}_n^{J, \mathbf{h}}(\mathbf{x})) \right) \\ &\leq \log 2 + \delta_n + \log \left( \sum_{\mathbf{x} \in \mathcal{A}_n} \exp \left[ \mathbf{H}_n^{J, \mathbf{h}}(\widehat{\mathbf{x}}) - I_n(\widehat{\mathbf{x}}) + g_n(\mathbf{x}, \widehat{\mathbf{x}}) \right] \right). \end{aligned} \tag{3.5}$$

Since  $\delta_n = o(n)$ , it is enough to upper bound the rightmost term in the RHS of (3.5). This will be done by approximating the summation over  $\mathcal{A}_n$ , by a summation over a suitable net of  $\mathcal{A}_n$ .

To this end, using Lemma 3.4 we obtain an  $\sqrt{n}\varepsilon$ -net  $\mathcal{D}_n(\varepsilon)$  having a sub-exponential size, of the set  $\{A_n \mathbf{v}, \mathbf{v} \in [0, 1]^n\}$ . For any  $\bar{\mathbf{v}} := (\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_q)$  such that  $\mathbf{v}_r \in \mathcal{D}_n(\varepsilon)$  for each  $r \in [q]$ , choose (if exists) a  $\mathbf{v}(\bar{\mathbf{v}}) \in \mathcal{A}_n \subset \mathcal{X}_n \subset \{0, 1\}^{nq}$  such that  $\|A_n \mathbf{v}_r(\bar{\mathbf{v}}) - \bar{\mathbf{v}}_r\|_2 \leq \sqrt{n}\varepsilon$  for all  $r \in [q]$ . Here  $\mathbf{v}_r(\bar{\mathbf{v}}) := (\mathbf{v}_{ir}(\bar{\mathbf{v}}))_{i \in [n]}$ . Also for any  $\mathbf{v}(\bar{\mathbf{v}})$  define

$$\mathcal{D}(\mathbf{v}(\bar{\mathbf{v}})) := \{ \mathbf{x} \in \mathcal{A}_n : \|A_n \mathbf{x}_r - A_n \mathbf{v}_r(\bar{\mathbf{v}})\|_2 \leq 2\sqrt{n}\varepsilon, r \in [q] \}.$$

By triangle inequality it is easy to see that

$$\mathcal{A}_n \subset \bigcup_{\mathbf{v}_r \in \mathcal{D}_n(\varepsilon), r \in [q]} \mathcal{D}(\mathbf{v}(\bar{\mathbf{v}})) = \bigcup_{\bar{\mathbf{v}} \in \mathcal{D}_n^q(\varepsilon)} \mathcal{D}(\mathbf{v}(\bar{\mathbf{v}})),$$

and so

$$\begin{aligned} & \sum_{\mathbf{x} \in \mathcal{A}_n} \exp \left[ \mathbf{H}_n^{J, \mathbf{h}}(\hat{\mathbf{x}}) - I_n(\hat{\mathbf{x}}) + g_n(\mathbf{x}, \hat{\mathbf{x}}) \right] \\ & \leq \sum_{\bar{\mathbf{v}} \in \mathcal{D}_n^q(\varepsilon)} \sum_{\mathbf{x} \in \mathcal{D}(\mathbf{v}(\bar{\mathbf{v}}))} \exp \left[ \mathbf{H}_n^{J, \mathbf{h}}(\hat{\mathbf{x}}) - I_n(\hat{\mathbf{x}}) + g_n(\mathbf{x}, \hat{\mathbf{x}}) \right]. \end{aligned} \tag{3.6}$$

We then claim that for any  $\mathbf{x} \in \mathcal{D}(\mathbf{v}(\bar{\mathbf{v}}))$ ,

$$\begin{aligned} & |\mathbf{H}_n^{J, \mathbf{h}}(\hat{\mathbf{x}}) - \mathbf{H}_n^{J, \mathbf{h}}(\widehat{\mathbf{v}(\bar{\mathbf{v}})})| + |g_n(\mathbf{x}, \hat{\mathbf{x}}) - g_n(\mathbf{x}, \widehat{\mathbf{v}(\bar{\mathbf{v}})})| + |I_n(\hat{\mathbf{x}}) - I_n(\widehat{\mathbf{v}(\bar{\mathbf{v}})})| \\ & \leq (q \|\mathbf{h}\|_\infty + 1)\delta_n + 2q^2 \|\mathbf{J}\|_\infty (\|\mathbf{h}\|_\infty + 1)n\varepsilon + 4q^3 \|\mathbf{J}\|_\infty n\varepsilon. \end{aligned} \tag{3.7}$$

Since  $\delta_n = o(n)$  the RHS of (3.7) is bounded by  $C(q)n\varepsilon$  for some finite constant  $C(q)$ , for all large  $n$ . Thus using (3.5)-(3.7) and noting the fact that

$$\sum_{\mathbf{x} \in \mathcal{X}_n} e^{g_n(\mathbf{x}, \mathbf{z})} = 1,$$

for any  $\mathbf{z} \in \hat{\mathcal{X}}_n$ , we deduce that

$$\begin{aligned} \Phi_n(\mathbf{J}, \mathbf{h}) & \leq \log 2 + C(q)n\varepsilon + \log \left( \sum_{\bar{\mathbf{v}} \in \mathcal{D}_n^q(\varepsilon)} \sum_{\mathbf{x} \in \mathcal{D}(\mathbf{v}(\bar{\mathbf{v}}))} \exp \left[ \mathbf{H}_n^{J, \mathbf{h}}(\widehat{\mathbf{v}(\bar{\mathbf{v}})}) - I_n(\widehat{\mathbf{v}(\bar{\mathbf{v}})}) + g_n(\mathbf{x}, \widehat{\mathbf{v}(\bar{\mathbf{v}})}) \right] \right) \\ & \leq \log 2 + C(q)n\varepsilon + \log \left( \sum_{\mathbf{v}(\bar{\mathbf{v}}) \in \mathcal{D}_n^q(\varepsilon)} \exp \left[ \mathbf{H}_n(\widehat{\mathbf{v}(\bar{\mathbf{v}})}) - I_n(\widehat{\mathbf{v}(\bar{\mathbf{v}})}) \right] \right) \\ & \leq \log 2 + C(q)n\varepsilon + q \log |\mathcal{D}_n(\varepsilon)| + \sup_{\mathbf{q} \in \mathcal{P}([q])^n} \mathbf{M}_n^{J, \mathbf{h}}(\mathbf{q}), \end{aligned} \tag{3.8}$$

where the last inequality uses the fact that for any  $\mathbf{x} \in \mathcal{X}_n$  setting  $\mathbf{q}_i(r) = \hat{x}_{ir}$  one has  $\mathbf{q}_i \in \mathcal{P}([q])$ , for each  $i \in [n]$ . Thus using the fact that  $\log |\mathcal{D}_n(\varepsilon)| = o(n)$  we have

$$\limsup_{n \rightarrow \infty} \frac{1}{n} \left[ \Phi_n(\mathbf{J}, \mathbf{h}) - \sup_{\mathbf{q} \in \mathcal{P}([q])^n} \mathbf{M}_n^{J, \mathbf{h}}(\mathbf{q}) \right] \leq C(q)\varepsilon.$$

Since  $\varepsilon > 0$  is arbitrary, letting  $\varepsilon \rightarrow 0$  the proof completes. Therefore it only remains to verify (3.7).

Turning to prove (3.7), we recall that for every  $\mathbf{v}(\bar{\mathbf{v}}) \in \mathcal{D}_n^{(q)}(\varepsilon)$ , and any  $\mathbf{x} \in \mathcal{D}(\mathbf{v}(\bar{\mathbf{v}}))$ , both  $\mathbf{x}$  and  $\mathbf{v}(\bar{\mathbf{v}})$  are in the set  $\mathcal{A}_n$ . Therefore

$$|\mathbf{H}_n^{J,h}(\mathbf{x}) - \mathbf{H}_n^{J,h}(\hat{\mathbf{x}})| \leq \delta_n/2, \quad \text{and} \quad |\mathbf{H}_n^{J,h}(\mathbf{v}(\bar{\mathbf{v}})) - \mathbf{H}_n^{J,h}(\widehat{\mathbf{v}(\bar{\mathbf{v}})})| \leq \delta_n/2.$$

Thus, in order to bound  $|\mathbf{H}_n^{J,h}(\hat{\mathbf{x}}) - \mathbf{H}_n^{J,h}(\widehat{\mathbf{v}(\bar{\mathbf{v}})})|$ , we only need to consider  $|\mathbf{H}_n^{J,h}(\mathbf{x}) - \mathbf{H}_n^{J,h}(\mathbf{v}(\bar{\mathbf{v}}))|$ . Now recall that

$$\mathbf{H}_n^{J,h}(\mathbf{x}) = \frac{1}{2} \sum_{r,s \in [q]} J_{rs} \mathbf{x}'_r A_n \mathbf{x}_s + \sum_{r \in [q]} h_r \mathbf{1}' \mathbf{x}_r.$$

Note that  $\mathbf{x} \in \mathcal{D}(\mathbf{v}(\bar{\mathbf{v}}))$  we have

$$\begin{aligned} |\mathbf{x}'_r A_n \mathbf{x}_s - \mathbf{v}_r(\bar{\mathbf{v}})' A_n \mathbf{v}_s(\bar{\mathbf{v}})| &\leq |\mathbf{x}'_r A_n \mathbf{x}_s - \mathbf{x}'_r A_n \mathbf{v}_s(\bar{\mathbf{v}})| + |\mathbf{x}'_r A_n \mathbf{v}_s(\bar{\mathbf{v}}) - \mathbf{v}_r(\bar{\mathbf{v}})' A_n \mathbf{v}_s(\bar{\mathbf{v}})| \\ &\leq \sqrt{n} \|A_n \mathbf{x}_r - A_n \mathbf{v}_r(\bar{\mathbf{v}})\|_2 + \sqrt{n} \|A_n \mathbf{x}_s - A_n \mathbf{v}_s(\bar{\mathbf{v}})\|_2 \leq 4n\varepsilon. \end{aligned} \tag{3.9}$$

Next we proceed to bound  $|\mathbf{1}' \mathbf{x}_r - \mathbf{1}' \mathbf{v}_r(\bar{\mathbf{v}})|$ . From Lemma 3.3, applying Markov’s inequality it follows that  $|\mathbf{1}' \mathbf{x}_r - \mathbf{1}' \hat{\mathbf{x}}_r| \leq \delta_n/2$ , for every  $\mathbf{x} \in \mathcal{A}_n$ , and  $r \in [q]$ . Hence it remains to find an upper bound on  $\|\hat{\mathbf{x}}_r - \widehat{\mathbf{v}_r(\bar{\mathbf{v}})}\|_2$ . To this end, recalling that  $\hat{x}_{ir} = \mathsf{T}_r(m_{i1}(\mathbf{x}) + h_1, \dots, m_{iq}(\mathbf{x}) + h_q)$  and noting that

$$\left\| \frac{\partial \mathsf{T}_r(m_1, \dots, m_q)}{\partial m_s} \right\|_{\infty} = \|\mathsf{T}_r(m_1, \dots, m_q)\{\delta(r, s) - \mathsf{T}_s(m_1, \dots, m_q)\}\|_{\infty} \leq 1, \tag{3.10}$$

applying a multivariate version of the mean-value theorem we obtain that

$$\begin{aligned} |\hat{x}_{ir} - \widehat{\mathbf{v}(\bar{\mathbf{v}})}_{ir}| &\leq \sum_{s \in [q]} |m_{is}(\mathbf{x}) - m_{is}(\mathbf{v}(\bar{\mathbf{v}}))| = \sum_{s \in [q]} \left| \sum_{s' \in [q]} J_{ss'} \{(A_n \mathbf{x}_{s'})_i - (A_n \mathbf{v}_{s'}(\bar{\mathbf{v}}))_i\} \right| \\ &\leq q \|\mathbf{J}\|_{\infty} \sum_{s' \in [q]} |(A_n \mathbf{x}_{s'})_i - (A_n \mathbf{v}_{s'}(\bar{\mathbf{v}}))_i|. \end{aligned}$$

This further implies that

$$\|\hat{\mathbf{x}}_r - \widehat{\mathbf{v}(\bar{\mathbf{v}})}\|_2^2 \leq q^3 \|\mathbf{J}\|_{\infty}^2 \sum_{i \in [n], s' \in [q]} |(A_n \mathbf{x}_{s'})_i - (A_n \mathbf{v}_{s'}(\bar{\mathbf{v}}))_i|^2 \leq 4q^4 \|\mathbf{J}\|_{\infty}^2 n\varepsilon^2.$$

Therefore,

$$\begin{aligned} |\mathbf{H}_n^{J,h}(\hat{\mathbf{x}}) - \mathbf{H}_n^{J,h}(\widehat{\mathbf{v}(\bar{\mathbf{v}})})| &\leq |\mathbf{H}_n^{J,h}(\mathbf{x}) - \mathbf{H}_n^{J,h}(\mathbf{v}(\bar{\mathbf{v}}))| + \delta_n \\ &\leq \delta_n + \frac{\|\mathbf{J}\|_{\infty}}{2} \sum_{r,s \in [q]} |\mathbf{x}'_r A_n \mathbf{x}_s - \mathbf{v}_r(\bar{\mathbf{v}})' A_n \mathbf{v}_s(\bar{\mathbf{v}})| + \|\mathbf{h}\|_{\infty} \sum_{r \in [q]} |\mathbf{1}' \mathbf{x}_r - \mathbf{1}' \mathbf{v}_r(\bar{\mathbf{v}})| \end{aligned}$$

$$\begin{aligned} &\leq (q \|\mathbf{h}\|_\infty + 1)\delta_n + 2q^2 \|\mathbf{J}\|_\infty n\varepsilon + \sqrt{n} \|\mathbf{h}\|_\infty \sum_{r \in [q]} \left\| \hat{\mathbf{x}}_r - \widehat{\mathbf{v}}_r(\bar{\mathbf{v}}) \right\|_2 \\ &\leq (q \|\mathbf{h}\|_\infty + 1)\delta_n + 2q^2 \|\mathbf{J}\|_\infty (\|\mathbf{h}\|_\infty + 1)n\varepsilon. \end{aligned} \tag{3.11}$$

Next we proceed to bound  $|g_n(\mathbf{x}, \hat{\mathbf{x}}) - g_n(\mathbf{x}, \widehat{\mathbf{v}}(\bar{\mathbf{v}}))|$ . To this end, we have

$$\begin{aligned} &|g_n(\mathbf{x}, \hat{\mathbf{x}}) - g_n(\mathbf{x}, \widehat{\mathbf{v}}(\bar{\mathbf{v}}))| \\ &\leq \sum_{i \in [n], r \in [q]} \left| \log \mathsf{T}_r(\mathbf{m}_{i1}(\mathbf{x}) + h_1, \dots, \mathbf{m}_{iq}(\mathbf{x}) + h_q) \right. \\ &\quad \left. - \log \mathsf{T}_r(\mathbf{m}_{i1}(\widehat{\mathbf{v}}(\bar{\mathbf{v}})) + h_1, \dots, \mathbf{m}_{iq}(\widehat{\mathbf{v}}(\bar{\mathbf{v}})) + h_q) \right| \\ &\leq \sum_{i \in [n], r, s \in [q]} |\mathbf{m}_{is}(\mathbf{x}) - \mathbf{m}_{is}(\widehat{\mathbf{v}}(\bar{\mathbf{v}}))| \left\| \frac{\partial \log \mathsf{T}_r}{\partial m_s} \right\|_\infty \\ &\leq \sum_{i \in [n], s \in [q]} |\mathbf{m}_{is}(\mathbf{x}) - \mathbf{m}_{is}(\widehat{\mathbf{v}}(\bar{\mathbf{v}}))|, \end{aligned}$$

where the last inequality uses (3.10) to conclude that  $\left\| \frac{\partial \log \mathsf{T}_r}{\partial m_s} \right\|_\infty \leq 1$ .

This gives

$$\begin{aligned} |g_n(\mathbf{x}, \hat{\mathbf{x}}) - g_n(\mathbf{x}, \widehat{\mathbf{v}}(\bar{\mathbf{v}}))| &\leq q \sum_{i \in [n], s \in [q]} |\mathbf{m}_{is}(\mathbf{x}) - \mathbf{m}_{is}(\widehat{\mathbf{v}}(\bar{\mathbf{v}}))| \\ &= q \sum_{i \in [n], s \in [q]} \left| \sum_{s' \in [q]} J_{ss'} \{ (A_n \mathbf{x}_{s'})_i - (A_n \mathbf{v}_{s'}(\bar{\mathbf{v}}))_i \} \right| \\ &\leq q^2 \|\mathbf{J}\|_\infty \sum_{i \in [n], s' \in [q]} |(A_n \mathbf{x}_{s'})_i - (A_n \mathbf{v}_{s'}(\bar{\mathbf{v}}))_i| \\ &\leq q^2 \sqrt{n} \|\mathbf{J}\|_\infty \sum_{s' \in [q]} \|A_n \mathbf{x}_{s'} - A_n \mathbf{v}_{s'}(\bar{\mathbf{v}})\|_2 \leq 2q^3 \|\mathbf{J}\|_\infty n\varepsilon, \end{aligned} \tag{3.12}$$

where the penultimate step uses Cauchy–Schwarz inequality, and the last step uses the fact that  $\mathbf{x} \in \mathcal{D}(\widehat{\mathbf{v}}(\bar{\mathbf{v}}))$ .

Now it remains to bound  $|I_n(\hat{\mathbf{x}}) - I_n(\widehat{\mathbf{v}}(\bar{\mathbf{v}}))|$ , for which we follow a similar program. Setting  $\gamma(t) := t \log t$  for  $t \geq 0$ , we have

$$\begin{aligned} &|I_n(\hat{\mathbf{x}}) - I_n(\widehat{\mathbf{v}}(\bar{\mathbf{v}}))| \\ &\leq \sum_{i \in [n], r \in [q]} |\gamma(\mathsf{T}_r(\mathbf{m}_{i1}(\mathbf{x}) + h_1, \dots, \mathbf{m}_{iq}(\mathbf{x}) + h_q)) \end{aligned}$$

$$\begin{aligned}
 & -\gamma \left( \mathbb{T}_r(\mathbf{m}_{i_1}(\mathbf{v}(\bar{\mathbf{v}})) + h_1, \dots, \mathbf{m}_{i_q}(\mathbf{v}(\bar{\mathbf{v}})) + h_q) \right) \\
 \leq & \sum_{i \in [n], r, s \in [q]} |\mathbf{m}_{i_s}(\mathbf{x}) - \mathbf{m}_{i_s}(\mathbf{v}(\bar{\mathbf{v}}))| \left\| \frac{\partial(\gamma \circ \mathbb{T}_r)}{\partial m_s} \right\|_{\infty}
 \end{aligned}$$

Using (3.10) gives

$$\left\| \frac{\partial(\gamma \circ \mathbb{T}_r)}{\partial m_s} \right\|_{\infty} = \|\mathbb{T}_r(1 + \log \mathbb{T}_r)\{\delta(r, s) - \mathbb{T}_s\}\|_{\infty} \leq \sup_{t \in [0, 1]} t|1 + \log t| \leq 1,$$

and therefore

$$|I_n(\hat{\mathbf{x}}) - I_n(\widehat{\mathbf{v}(\bar{\mathbf{v}})})| \leq q \sum_{i \in [n], s \in [q]} |\mathbf{m}_{i_s}(\mathbf{x}) - \mathbf{m}_{i_s}(\widehat{\mathbf{v}(\bar{\mathbf{v}})})| \leq 2q^3 \|\mathbf{J}\|_{\infty} n\varepsilon, \tag{3.13}$$

where the last bound follows by arguments similar to (3.12). Finally combining (3.11)-(3.13) we arrive at (3.7), and this completes the proof.  $\square$

### 4 Proof of auxiliary Lemmas

In this section we prove Lemmas 3.2, 3.3, and 3.4. We start with the proof of Lemma 3.2.

*Proof of Lemma 3.2* To lighten the notation, we drop the subscript  $n$  in  $F_n$ , and write  $F$  through out the proof. Before we begin the proof let us introduce some notation:

$$F_{ir}(\mathbf{x}) := \frac{\partial}{\partial x_{ir}} F(\mathbf{x}), \quad \text{and} \quad F_{ir, js}(\mathbf{x}) := \frac{\partial^2}{\partial x_{ir} \partial x_{js}} F(\mathbf{x}).$$

Equipped with these notation by mean-value theorem we have

$$F(\mathbf{x}) - F(\widehat{\mathbf{x}}) = \int_0^1 \sum_{i \in [n], r \in [q]} (x_{ir} - \hat{x}_{ir}) F_{ir}(t\mathbf{x} + (1-t)\widehat{\mathbf{x}}) dt.$$

Thus denoting  $\Delta(\mathbf{x}) := F(\mathbf{x}) - F(\widehat{\mathbf{x}})$ , and  $u_{ir}(t, \mathbf{x}) := F_{ir}(t\mathbf{x} + (1-t)\widehat{\mathbf{x}})$ , we have

$$\mathbb{E}_{\mu_n} \left\{ [F(\mathbf{X}) - F(\widehat{\mathbf{X}})]^2 \right\} = \int_0^1 \sum_{i \in [n], r \in [q]} \mathbb{E}_{\mu_n} \left( (X_{ir} - \hat{X}_{ir}) u_{ir}(t, \mathbf{X}) \Delta(\mathbf{X}) \right) dt. \tag{4.1}$$

Hence to complete the proof it is enough to find upper bound on the RHS of (4.1) for each value of  $t \in [0, 1]$ . To this end observe that for any  $i \in [n], r \in [q]$  we have



$$\mathbb{E}_{\mu_n} \left( (X_{ir} - \widehat{X}_{ir}) u_{ir}(t, \mathbf{X}^{(ir)}) \Delta(\mathbf{X}^{(ir)}) \right) = 0,$$

where  $\mathbf{X}^{(ir)}$  is obtained by setting  $X_{ir} = 0$  in the random vector  $\mathbf{X}$ . Therefore for each  $i \in [n], r \in [q]$  it suffices to consider the difference

$$\mathbb{E}_{\mu_n} \left( (X_{ir} - \widehat{X}_{ir}) u_{ir}(t, \mathbf{X}) \Delta(\mathbf{X}) - (X_{ir} - \widehat{X}_{ir}) u_{ir}(t, \mathbf{X}^{(ir)}) \Delta(\mathbf{X}^{(ir)}) \right),$$

and show that

$$\sup_{t \in [0,1]} \left| \sum_{i \in [n], r \in [q]} \mathbb{E}_{\mu_n} \left( (X_{ir} - \widehat{X}_{ir}) (u_{ir}(t, \mathbf{X}) - u_{ir}(t, \mathbf{X}^{(ir)})) \Delta(\mathbf{X}^{(ir)}) \right) \right| = o(n^2), \tag{4.2}$$

and

$$\sup_{t \in [0,1]} \left| \sum_{i \in [n], r \in [q]} \mathbb{E}_{\mu_n} \left( (X_{ir} - \widehat{X}_{ir}) u_{ir}(t, \mathbf{X}) (\Delta(\mathbf{X}) - \Delta(\mathbf{X}^{(ir)})) \right) \right| = o(n^2). \tag{4.3}$$

To establish (4.2), we first note that using (1.10) there exists  $C_1 < \infty$  such that

$$\begin{aligned} & \left| \mathbb{E}_{\mu_n} \left( (X_{ir} - \widehat{X}_{ir}) (u_{ir}(t, \mathbf{X}) - u_{ir}(t, \mathbf{X}^{(ir)})) \Delta(\mathbf{X}^{(ir)}) \right) \right| \\ & \leq C_1 n \mathbb{E}_{\mu_n} \left| u_{ir}(t, \mathbf{X}) - u_{ir}(t, \mathbf{X}^{(ir)}) \right|. \end{aligned} \tag{4.4}$$

Since

$$\begin{aligned} u_{ir}(t, X) &= t m_{ir}(\mathbf{x}) + (1-t) m_{ir}(\widehat{\mathbf{x}}) \\ &= t m_{ir}(\mathbf{x}) + (1-t) \sum_{j \in [n], s \in [q]} J_{rs} A_n(i, j) \mathbb{T}_s(m_{j1}(\mathbf{x}) + h_1, \dots, m_{jq}(\mathbf{x}) + h_q), \end{aligned}$$

and  $m_{ir}(\mathbf{x})$  is free of  $\{x_{is}\}_{s \in [q]}$ , by chain rule the RHS of (4.4) can be bounded by

$$\begin{aligned} & C_1 n \left| \sum_{j \in [n], s \in [q]} J_{rs} A_n(i, j) \sum_{s' \in [q]} \left( \frac{\partial \mathbb{T}_s}{\partial m_{s'}}(m_1, \dots, m_q) \Big|_{\{m_{j\ell} = m_{j\ell'}(\mathbf{x}) + h_{\ell}, \ell' \in [q]\}} \right) A_n(i, j) J_{rs'} \right| \\ & \leq C_1 q^2 n \|J\|_{\infty}^2 \sum_{j \in [n]} |A_n(i, j)|^2, \end{aligned}$$

where the last step uses (3.10). This, on summing over  $i \in [n]$ , and  $r \in [q]$  gives (4.2) by the mean-field assumption. Next turning to bound (4.3), we first write

$$\begin{aligned}
 2(\Delta(\mathbf{X}) - \Delta(\mathbf{X}^{(ir)})) &= \sum_{a,b \in [q]} J_{ab} \left[ \mathbf{X}'_a A_n \mathbf{X}_b - \mathbf{X}^{(ir)'}_a A_n \mathbf{X}^{(ir)}_b \right] \\
 &\quad - \sum_{a,b \in [q]} J_{ab} \left[ \widehat{\mathbf{X}}'_a A_n \widehat{\mathbf{X}}_b - \widehat{\mathbf{X}}^{(ir)'}_a A_n \left( \widehat{\mathbf{X}}^{(ir)} \right)_b \right] \\
 &\quad - \sum_{a,b \in [q]} J_{ab} \left[ \widehat{\mathbf{X}}'_a A_n \widehat{\mathbf{X}}^{(ir)}_b - \widehat{\mathbf{X}}^{(ir)'}_a A_n \left( \widehat{\mathbf{X}}^{(ir)} \right)_b \right]. \tag{4.5}
 \end{aligned}$$

Here the notation  $(\widehat{\mathbf{X}}^{(ir)})_b$  means the  $b^{th}$  column of the matrix  $\widehat{\mathbf{X}}^{(ir)}$ . Now note for any  $a \in [q] \setminus \{r\}$ ,  $\mathbf{x}_a = \mathbf{x}_a^{(ir)}$ ,

$$\mathbf{x}'_a A_n \mathbf{x}_r - \mathbf{x}_a^{(ir)'} A_n \mathbf{x}_r^{(ir)} = \mathbf{x}'_a A_n \mathbf{x}_r - \mathbf{x}_a A_n \mathbf{x}_r^{(ir)} = x_{ir} (A_n \mathbf{x}_a)_i,$$

and

$$\mathbf{x}'_r A_n \mathbf{x}_r - \mathbf{x}_r^{(ir)'} A_n \mathbf{x}_r^{(ir)} = x_{ir} (A_n \mathbf{x}_r)_i + x_{ir} \left( A_n \mathbf{x}_r^{(ir)} \right)_i = 2x_{ir} (A_n \mathbf{x}_r)_i,$$

where the last equality follows from the fact that  $A_n(i, i) = 0$ . Thus recalling the definition of  $m_{ir}(\mathbf{x})$ , we have

$$\begin{aligned}
 &\left| \sum_{i \in [n], a,b,r \in [q]} J_{ab} \left( (X_{ir} - \widehat{X}_{ir}) u_{ir}(t, \mathbf{x}) \{ \mathbf{x}'_a A_n \mathbf{x}_b - \mathbf{x}_a^{(ir)'} A_n \mathbf{x}_b^{(ir)} \} \right) \right| \\
 &\leq 2 \sum_{i \in [n], r \in [q]} |m_{ir}(\mathbf{x})| (|m_{ir}(\mathbf{x})| + |m_{ir}(\widehat{\mathbf{x}})|) \\
 &\leq 2 \left\{ \sum_{r \in [q], i \in [n]} m_{ir}(\mathbf{x})^2 + \sum_{r \in [q]} \sqrt{\sum_{i=1}^n m_{ir}(\mathbf{x})^2} \sqrt{\sum_{i=1}^n m_{ir}(\widehat{\mathbf{x}})^2} \right\}, \tag{4.6}
 \end{aligned}$$

where the first step uses the fact that  $|u_{ir}(t, \mathbf{x})| \leq t|m_{ir}(\mathbf{x})| + (1-t)|m_{ir}(\widehat{\mathbf{x}})|$ , and last step follows by an application of Cauchy–Schwarz inequality.

Also the mean-field assumption implies that  $\lambda_{\max}(A_n) = o(\sqrt{n})$ , and therefore we have

$$\sum_{i \in [n]} m_{ir}(\mathbf{x})^2 \leq q \|J\|_\infty^2 \sum_{s \in [q]} \|A_n \mathbf{x}_s\|_2^2 \leq q \|J\|_\infty^2 \lambda_{\max}^2(A_n) \sum_{s \in [q]} \|\mathbf{x}_s\|_2^2 = o(n^2).$$

By similar arguments, from (4.6) we deduce that

$$\sup_{t \in [0,1]} \left| \sum_{i \in [n], r \in [q]} \mathbb{E}_{\mu_n} \left( (X_{ir} - \widehat{X}_{ir}) u_{ir}(t, \mathbf{X}) \sum_{a,b \in [q]} J_{ab} \left[ \mathbf{X}'_a A_n \mathbf{X}_b - \mathbf{X}_a^{(ir)'} A_n \mathbf{X}_b^{(ir)} \right] \right) \right| = o(n^2). \tag{4.7}$$

Next we consider the second term in the RHS of (4.5), where using first order Taylor’s theorem, upon application of chain rule, followed by (3.10), gives

$$\begin{aligned} \widehat{\mathbf{x}}'_a A_n \widehat{\mathbf{x}}_r - \widehat{\mathbf{x}}'_a A_n \left(\widehat{\mathbf{x}}^{(ir)}\right)_b &= \sum_{j,k \in [n]} \widehat{x}_{ja} A_n(j, k) \left\{ \widehat{x}_{kr} - \left(\widehat{x}^{(ir)}\right)_{kb} \right\} \\ &= \sum_{j,k \in [n]} \widehat{x}_{ja} A_n(j, k) A_n(i, k) \xi_{i,k,b,r}, \end{aligned}$$

for some  $\xi_{i,k,b,r}$ , such that  $|\xi_{i,k,b,r}| \leq q \|J\|_\infty$ . Denoting  $\|A_n\|_\infty := \sup_{i,j} |A_n(i, j)|$ , and summing over  $i \in [n], a, b, r \in [q]$  this gives

$$\begin{aligned} &\left| \sum_{a,b,r \in [q], i \in [n]} J_{ab} \left( (x_{ir} - \widehat{x}_{ir}) u_{ir}(t, \mathbf{x}) \left\{ \widehat{\mathbf{x}}'_a A_n \widehat{\mathbf{x}}_b - \widehat{\mathbf{x}}'_a A_n \left(\widehat{\mathbf{x}}^{(ir)}\right)_b \right\} \right) \right| \\ &= \left| \sum_{i,k \in [n], b,r \in [q]} ((x_{ir} - \widehat{x}_{ir}) x_{ir} m_{kb}(\widehat{\mathbf{x}}) A_n(i, k) \xi_{i,k,b,r} u_{ir}(t, \mathbf{x})) \right| \\ &\leq q \|A_n\|_\infty \|J\|_\infty \sum_{i,k \in [n], b,r \in [q]} (|m_{ir}(\mathbf{x})| + |m_{ir}(\widehat{\mathbf{x}})|) |m_{kb}(\widehat{\mathbf{x}})| \\ &= q \|A_n\|_\infty \|J\|_\infty \left[ \left( \sum_{i \in [n], r \in [q]} |m_{ir}(\mathbf{x})| \right) \left( \sum_{k \in [n], b \in [q]} |m_{kb}(\widehat{\mathbf{x}})| \right) \right. \\ &\quad \left. + \left( \sum_{i \in [n], r \in [q]} |m_{ir}(\widehat{\mathbf{x}})| \right) \left( \sum_{k \in [n], b \in [q]} |m_{kb}(\widehat{\mathbf{x}})| \right) \right]. \end{aligned}$$

Now using (1.10) we obtain

$$\sum_{i \in [n], r \in [q]} |m_{ir}(\mathbf{x})|, \quad \sum_{i \in [n], r \in [q]} |m_{ir}(\widehat{\mathbf{x}})| = O(n).$$

This together with the fact that  $\|A_n\|_\infty = o(1)$ , implies that the RHS above is  $o(n^2)$ , thus giving

$$\sup_{t \in [0,1]} \left| \sum_{i \in [n], r \in [q]} \mathbb{E}_{\mu_n} \left( (X_{ir} - \widehat{X}_{ir}) u_{ir}(t, \mathbf{X}) \sum_{a,b \in [q]} J_{ab} \left[ \widehat{\mathbf{X}}'_a A_n \widehat{\mathbf{X}}_b - \widehat{\mathbf{X}}'_a A_n \left(\widehat{\mathbf{X}}^{(ir)}\right)_b \right] \right) \right| = o(n^2). \tag{4.8}$$

Finally, considering the third term in the RHS of (4.5) and using first order Taylor’s theorem again, we also note that

$$\begin{aligned} \widehat{\mathbf{x}}'_b A_n(\widehat{\mathbf{x}}^{(ir)})_a - (\widehat{\mathbf{x}}^{(ir)})'_b A_n(\widehat{\mathbf{x}}^{(ir)})_a &= \sum_{j,k \in [n]} (\widehat{\mathbf{x}}^{(ir)})_{kb} (\widehat{x}_{ja} - (\widehat{\mathbf{x}}^{(ir)})_{ja}) A_n(j, k) \\ &= \sum_{j,k \in [n]} (\widehat{\mathbf{x}}^{(ir)})_{kb} \xi_{i,j,a,r} A_n(j, k) A_n(i, k). \end{aligned}$$

From this, proceeding similarly as in the proof of (4.8) we have

$$\begin{aligned} &\left| \sum_{i,j,k \in [n], a,b,r \in [q]} J_{ab} \left( (x_{ir} - \widehat{x}_{ir}) u_i(t, \mathbf{x}) \left\{ \widehat{\mathbf{x}}'_b A_n \widehat{\mathbf{x}}^{(ir)}_a - (\widehat{\mathbf{x}}^{(ir)})'_b A_n(\widehat{\mathbf{x}}^{(ir)})_a \right\} \right) \right| \\ &= \left| \sum_{i,j \in [n], a,r \in [q]} \left( (x_{ir} - \widehat{x}_{ir}) x_{ir} m_{ja}(\widehat{\mathbf{x}}^{(ir)}) A_n(i, k) \xi_{i,j,a,r} u_{ir}(t, \mathbf{x}) \right) \right| \\ &\leq \|A_n\|_\infty \|J\|_\infty \sum_{i,j \in [n], a,r \in [q]} (|m_{ir}(\mathbf{x})| + |m_{ir}(\widehat{\mathbf{x}})|) (|\widehat{m}_{ja}(\widehat{\mathbf{x}}^{(ir)})|) = o(n^2) \quad (4.9) \end{aligned}$$

as before, and so

$$\begin{aligned} &\left| \sum_{i \in [n], r \in [q]} \mathbb{E}_{\mu_n} \left( (X_{ir} - \widehat{X}^{(ir)}) u_{ir}(t, \mathbf{X}) \sum_{a,b \in [q]} J_{ab} \left[ \widehat{\mathbf{X}}'_a A_n(\widehat{\mathbf{X}}^{(ir)})_b - (\widehat{\mathbf{X}}^{(ir)})'_a A_n(\widehat{\mathbf{X}}^{(ir)})_b \right] \right) \right| \\ &= o(n^2). \quad (4.10) \end{aligned}$$

Finally combining (4.7), (4.8), and (4.10), the proof is complete. □

Now we prove Lemma 3.3.

*Proof of Lemma 3.3* First we prove (3.2). To this end, for any  $\mathbf{x} \in \mathcal{X}_n$  define

$$G(\mathbf{x}) := \sum_{i \in [n], r \in [q]} (x_{ir} - \widehat{x}_{ir}) m_{ir}(\mathbf{x})$$

and note that

$$\mathbb{E}_{\mu_n} \left( (X_{ir} - \widehat{X}_{ir}) m_{ir}(X) G(X^{(ir)}) \right) = 0.$$

Thus we need to show that

$$\sum_{i \in [n], r \in [q]} \mathbb{E}_{\mu_n} \left[ (X_{ir} - \widehat{X}_{ir}) m_{i,r}(X) (G(X) - G(X^{(ir)})) \right] = o(n^2).$$

To this end, we first observe that

$$\begin{aligned}
 G(\mathbf{x}) - G(\mathbf{x}^{(ir)}) &= 2x_{ir}m_{ir}(\mathbf{x}) + \sum_{j \in [n], s \in [q]} \hat{x}_{js} \left( m_{js}(\mathbf{x}^{(ir)}) - m_{js}(\mathbf{x}) \right) \\
 &\quad + \sum_{j \in [n], s \in [q]} \widehat{(x^{(ir)})}_{js} - \hat{x}_{js} m_{js}(\mathbf{x}^{(ir)}). \tag{4.11}
 \end{aligned}$$

For the first term in the RHS of (4.11), proceeding as in (4.7), by a Cauchy–Schwarz argument we have

$$\left| \sum_{i \in [n], r \in [q]} (x_{ir} - \hat{x}_{ir})x_{ir}m_{ir}^2 \right| \leq \sum_{i \in [n], r \in [q]} m_{ir}(\mathbf{x})^2 = o(n^2),$$

giving

$$\left| \mathbb{E}_{\mu_n} \sum_{i \in [n], r \in [q]} \left[ (X_{ir} - \hat{X}_{ir})X_{ir}m_{ir}(\mathbf{X})^2 \right] \right| = o(n^2). \tag{4.12}$$

For controlling the second term in the RHS of (4.11) first note that  $m_{js}(\mathbf{x}) - m_{js}(\mathbf{x}^{(ir)}) = A_n(i, j)J_{rs}x_{ir}$ . Thus proceeding as in (4.6) again we further have

$$\begin{aligned}
 &\left| \sum_{i, j \in [n], r, s \in [q]} (x_{ir} - \hat{x}_{ir})m_{ir}(\mathbf{x})\hat{x}_{js}(m_{js}(\mathbf{x}) - m_{js}(\mathbf{x}^{(ir)})) \right| \\
 &= \left| \sum_{i \in [n], r \in [q]} (x_{ir} - \hat{x}_{ir})x_{ir}m_{ir}(\mathbf{x})m_{ir}(\hat{\mathbf{x}}) \right| \leq \sum_{i \in [n], r \in [q]} |m_{ir}(\mathbf{x})||m_{ir}(\hat{\mathbf{x}})|
 \end{aligned}$$

which is  $o(n^2)$  by a Cauchy–Schwarz argument as in the proof of (4.7). Thus we have

$$\left| \mathbb{E}_{\mu_n} \sum_{i, j \in [n], r, s \in [q]} (X_{ir} - \hat{X}_{ir})m_{i,r}(\mathbf{X})\hat{X}_{js}(m_{js}(\mathbf{X}) - m_{js}(\mathbf{X}^{(ir)})) \right| = o(n^2). \tag{4.13}$$

Finally for controlling the third term in the RHS of (4.11), applying first order Taylor’s theorem yields that  $\hat{x}_{js} - \widehat{(x^{(ir)})}_{js} = A_n(i, j)\xi_{i,j,r,s}$  with  $|\xi_{i,j,r,s}| \leq q \|\mathbf{J}\|_\infty$ . Thus we have

$$\begin{aligned}
 &\left| \sum_{i, j \in [n], r, s \in [q]} (x_{ir} - \hat{x}_{ir})m_{ir}(\mathbf{x})m_{js}(\mathbf{x}^{(ir)}) \left\{ \hat{x}_{js} - \widehat{(x^{(ir)})}_{js} \right\} \right| \\
 &\leq q \|A_n\|_\infty \|\mathbf{J}\|_\infty \left( \sum_{i \in [n], r \in [q]} |m_{ir}(\mathbf{x})| \right)^2 + nq^2 \|A_n\|_\infty^2 \|\mathbf{J}\|_\infty^2 \sum_{i \in [n], r \in [q]} |m_{ir}(\mathbf{x})|,
 \end{aligned}$$

which is  $o(n^2)$  by arguments similar to the proof of (4.8). This gives

$$\left| \mathbb{E}_{\mu_n} \sum_{i,j \in [n], r,s \in [q]} (X_{ir} - \widehat{X}_{ir}) m_{i,r}(\mathbf{X}) m_{js}(\mathbf{X}^{(ir)}) \left\{ \widehat{X}_{js} - \left( \widehat{X}^{(ir)} \right)_{js} \right\} \right| = o(n^2), \tag{4.14}$$

which on combining with (4.12) and (4.13) completes the proof of (3.2).

Next to prove (3.3), we define

$$\widetilde{G}_r(\mathbf{x}) = \sum_{i \in [n]} (x_{ir} - \widehat{x}_{ir}),$$

and therefore

$$\widetilde{G}_r(\mathbf{x}) - \widetilde{G}_r(\mathbf{x}^{(ir)}) = x_{ir} - \sum_{j \in [n]} \left\{ \widehat{x}_{jr} - \left( \widehat{x}^{(ir)} \right)_{jr} \right\}.$$

Thus observing that

$$\mathbb{E}_{\mu_n} \left[ (X_{ir} - \widehat{X}_{ir}) \widetilde{G}_r(\mathbf{X}^{(ir)}) \right] = 0,$$

for any  $i \in [n], r \in [q]$ , we only need to show that

$$\sum_{i \in [n]} \mathbb{E}_{\mu_n} \left[ (X_{ir} - \widehat{X}_{ir}) (\widetilde{G}_r(\mathbf{X}) - \widetilde{G}_r(\mathbf{X}^{(ir)})) \right] = o(n^2).$$

This can be done proceeding similarly as above. We omit the details. □

Now we prove Lemma 3.4. Before going to the proof let us introduce the following notation:

For  $r \in \mathbb{N}$  and  $R > 0$  let  $\mathcal{B}_r(R)$  denote the Euclidean ball of radius  $R$  in dimension  $r$ , i.e.

$$\mathcal{B}_r(R) := \{ \mathbf{v} \in \mathbb{R}^r : \|\mathbf{v}\|_2 \leq R \}.$$

The proof of Lemma 3.4 also requires the following standard estimate on an  $\eta$ -net  $\mathcal{B}_r(R)$ . Its proof is based on simple volumetric argument. We refer the reader to [39, Lemma 2.6] for its proof.

**Lemma 4.1** *For any  $R, \eta \in \mathbb{R}$ , and  $r \in \mathbb{N}$ , there exists an  $\eta$ -net of  $\mathcal{B}_r(R)$  of size at most  $\max\{1, (3R/\eta)^r\}$ .*

*Proof of Lemma 3.4* Let  $\{\lambda_1(A_n), \dots, \lambda_n(A_n)\}$  denote the eigenvalues of  $A_n$ . Fixing  $\varepsilon \in (0, 1)$ , let  $N_n$  denote the number of eigenvalues of  $A_n$  which are greater than  $\varepsilon/2$  in absolute value. Since  $A_n$  satisfies the mean-field assumption, by Chebyshev's inequality we have that

$$0 \leq \lim_{n \rightarrow \infty} \frac{N_n}{n} \leq \lim_{n \rightarrow \infty} \frac{4}{n\varepsilon^2} \sum_{i \in [n]} \lambda_i(A_n)^2 = 0. \quad (4.15)$$

Set  $\ell = \ell_n := \lceil \log_2 \sqrt{n} \rceil$ , and for  $1 \leq k \leq \ell$  let  $I_k := \{1 \leq i \leq n : 2^{k-1} < |\lambda_i(A_n)| \leq 2^k\}$ . Thus with  $I_0 := \{1 \leq i \leq n : \varepsilon/2 < |\lambda_i(A_n)| \leq 1\}$  and  $I := \bigcup_{k=0}^{\ell} I_k$ , and using the fact that  $\text{tr}(A_n^2) = O(n)$ , we have

$$\sum_{k=0}^{\ell} |I_k| = |I| = N_n.$$

For  $0 \leq k, j \leq \ell$ , if  $I_k \neq \emptyset$  we let  $\mathcal{C}_k(j)$  denote an  $\varepsilon 2^{-(k+1)} \sqrt{|I_k|}$ -net of the set  $\mathcal{B}_{|I_k|}(2^j)$ . By Lemma 4.1 we may and will assume that

$$|\mathcal{C}_k(j)| \leq \max \left\{ 1, \left( \frac{6}{\varepsilon} \right)^{|I_k|} \left( \frac{2^{k+j}}{\sqrt{|I_k|}} \right)^{|I_k|} \right\}. \quad (4.16)$$

Setting

$$\mathcal{S}_n(\varepsilon) := \bigcup_{0 \leq j_0, j_1, \dots, j_{\ell} \leq \ell : \sum_{k=0}^{\ell} 2^{2j_k} \leq 5n} \{\mathcal{C}_0(j_0) \times \mathcal{C}_1(j_1) \times \dots \times \mathcal{C}_{\ell}(j_{\ell})\}$$

we first claim that

$$\lim_{n \rightarrow \infty} \frac{1}{n} \log |\mathcal{S}_n(\varepsilon)| = 0. \quad (4.17)$$

Deferring the proof of (4.17) and setting

$$\mathcal{D}_n(\varepsilon) := \left\{ \sum_{i \in I} \lambda_i(A_n) c_i \mathbf{p}_i : \mathbf{c} := (c_i)_{i \in I} \in \mathcal{S}_n(\varepsilon) \right\},$$

where  $\mathbf{p}_1, \mathbf{p}_2, \dots, \mathbf{p}_n$  are the eigenvectors of  $A_n$ , we will now show that  $\mathcal{D}_n(\varepsilon)$  is indeed an  $\sqrt{n}\varepsilon$ -net of  $\{A_n \mathbf{v} : \mathbf{v} \in [0, 1]^n\}$  having a sub-exponential size. Since (3.4) is immediate from (4.17), it only remains to show that  $\mathcal{D}_n(\varepsilon)$  is a  $\sqrt{n}\varepsilon$ -net.

To this end, fix  $\mathbf{v} \in [0, 1]^n$ , and expand  $\mathbf{v}$  in the basis  $\{\mathbf{p}_1, \mathbf{p}_2, \dots, \mathbf{p}_n\}$  as

$$\mathbf{v} = \sum_{i=1}^n \alpha_i \mathbf{p}_i,$$

where  $\alpha_1, \alpha_2, \dots, \alpha_n \in \mathbb{R}$  satisfies

$$\sum_{i=1}^n \alpha_i^2 = \sum_{i=1}^n v_i^2 \leq n, \quad (4.18)$$

For  $0 \leq k \leq \ell$  setting  $s_k := \sqrt{\sum_{i \in I_k} \alpha_i^2}$ , we will now find a vector  $\mathbf{c} \in \mathcal{S}_n(\varepsilon)$  such that

$$\left\| A\mathbf{v} - \sum_{i \in I} \lambda_i(A_n) c_i \mathbf{p}_i \right\|_2 \leq \sqrt{n}\varepsilon.$$

If  $I_k \neq \emptyset$  for some  $k$ , setting  $j_k := \max(0, \lceil \log_2 s_k \rceil)$  we note that  $(\alpha_i, i \in I_k) \in \mathcal{B}_{|I_k|}(2^{j_k})$ , and so there exists  $(c_i, i \in I_k) \in \mathcal{C}_k(j_k)$  such that

$$\sum_{i \in I_k} (\alpha_i - c_i)^2 \leq \frac{|I_k| \varepsilon^2}{2^{2k+2}}. \tag{4.19}$$

By our choice of  $j_k$  we have  $2^{j_k} \leq 2s_k$  if  $s_k \geq 1$ , and  $j_k = 0$  if  $s_k < 1$ . This gives

$$\sum_{k=0}^{\ell} 2^{2j_k} = \sum_{k: j_k=0} 2^{2j_k} + \sum_{k: j_k \geq 1} 2^{2j_k} \leq \ell + 4 \sum_{k=0}^{\ell} s_k^2 \leq \ell + 4 \sum_{i=1}^n \alpha_i^2 \leq 5n,$$

where the last step uses (4.18). Thus we have shown that  $\mathbf{c} = (c_i)_{i \in I} \in \mathcal{S}_n(\varepsilon)$ . Finally, recalling that  $|\lambda_i(A_n)| \leq 2^k$  for any  $i \in I_k$ , we note

$$\begin{aligned} \left\| A\mathbf{v} - \sum_{i \in I} \lambda_i(A_n) c_i \mathbf{p}_i \right\|_2^2 &= \sum_{k=0}^{\ell} \sum_{i \in I_k} \lambda_i(A_n)^2 (\alpha_i - c_i)^2 + \sum_{i \notin I} \lambda_i(A_n)^2 \alpha_i^2 \\ &\leq \frac{\varepsilon^2}{4} \sum_{k=0}^{\ell} 2^{2k} \frac{|I_k|}{2^{2k}} + \frac{\varepsilon^2}{4} \sum_{i=1}^n \alpha_i^2 \\ &\leq (N_n + n) \frac{\varepsilon^2}{4} \leq \frac{n\varepsilon^2}{2}, \end{aligned}$$

where the first two inequalities follow by an use of (4.19) and (4.18), respectively. Thus we have shown that  $\mathcal{D}_n(\varepsilon)$  is indeed an  $\sqrt{n}\varepsilon$ -net.

Therefore to complete the proof it suffices to show (4.17). To this effect, fix any  $j_0, j_1, j_2, \dots, j_{\ell}$  such that  $\sum_{k=0}^{\ell} 2^{2j_k} \leq 5n$ , and set  $\mathcal{K} = \mathcal{K}(j_0, j_1, \dots, j_{\ell}) := \{0 \leq k \leq \ell : 6 \times 2^{k+j_k} \geq \varepsilon \sqrt{|I_k|}\}$ . Thus we have

$$\log |\mathcal{C}_0(j_0) \times \mathcal{C}_1(j_1) \times \dots \times \mathcal{C}_{\ell}(j_{\ell})| \leq \sum_{k \in \mathcal{K}} |I_k| \log \left( \frac{6 \cdot 2^{k+j_k}}{\varepsilon \sqrt{|I_k|}} \right) \tag{4.20}$$

Further denote  $N'_n = N'_n(j_0, j_1, \dots, j_{\ell}) := \sum_{k \in \mathcal{K}} |I_k|$ , and obviously  $N'_n \leq N_n$ . Now using Jensen's inequality, applied for  $\log(\cdot)$ , we have



$$\begin{aligned} \frac{1}{N'_n} \sum_{k \in \mathcal{K}} |I_k| \log \left( \frac{6 \cdot 2^{k+j_k}}{\varepsilon \sqrt{|I_k|}} \right) &\leq \log \left\{ \frac{6}{\varepsilon N'_n} \sum_{k \in \mathcal{K}} 2^{k+j_k} \sqrt{|I_k|} \right\} \\ &\leq \log \left\{ \frac{6}{\varepsilon N'_n} \sqrt{\sum_{k=0}^{\ell} 2^{2k} |I_k|} \sqrt{\sum_{k=0}^{\ell} 2^{2j_k}} \right\}, \end{aligned} \tag{4.21}$$

where the last step follows by Cauchy–Schwarz’s inequality. Since for any  $k \geq 1$ , and any  $i \in I_k$ , we have  $|\lambda_i(A_n)| \geq 2^{k-1}$ , we therefore deduce that

$$\sum_{k=0}^{\ell} 2^{2k} |I_k| \leq |I_0| + 4 \sum_{k=1}^{\ell} \sum_{i \in I_k} |\lambda_i(A_n)|^2 \leq N_n + 4 \sum_{i=1}^n \lambda_i(A_n)^2.$$

Now note that Assumption 1.3 in particular implies that  $\sum_{i=1}^n \lambda_i(A_n)^2 \leq Cn$  for some positive constant  $C$ . Therefore recalling that  $\sum_{k=0}^{\ell} 2^{2j_k} \leq 5n$ , using (4.20), and (4.21), we deduce that

$$\begin{aligned} \log |\mathcal{C}_0(j_0) \times \mathcal{C}_1(j_1) \times \cdots \times \mathcal{C}_{\ell}(j_{\ell})| &\leq N'_n \left( \log \frac{6}{\varepsilon} \right) + N'_n \log \frac{n \sqrt{5(4C+1)}}{N'_n} \\ &\leq N_n \left( \log \frac{6}{\varepsilon} \right) + N_n \log \frac{n \sqrt{5(4C+1)}}{N_n}, \end{aligned} \tag{4.22}$$

where the last step uses the facts that  $\lim_{n \rightarrow \infty} \frac{N_n}{n} = 0$ , and  $x \mapsto x \log(1/x)$  is increasing near 0. Therefore from the definition of the set  $\mathcal{S}_n(\varepsilon)$  it now follows that

$$|\mathcal{S}_n(\varepsilon)| \leq (1 + \ell)^{1+\ell} \exp \left[ N_n \left( \log \frac{6}{\varepsilon} \right) + N_n \log \frac{n \sqrt{5(4C+1)}}{N_n} \right].$$

Now using the fact that  $\lim_{n \rightarrow \infty} \frac{N_n}{n} = 0$  again the proof completes. □

*Remark 4.1* Note that the proof of Lemma 3.4 goes through as long as the following hold:

$$\frac{1}{n} \sum_{i=1}^n \delta_{\lambda_i}(A_n) \xrightarrow{w} \delta_0, \quad \limsup_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n \lambda_i(A_n)^2 < \infty. \tag{4.23}$$

For example, if  $A_n$  is the adjacency matrix of the  $n$ -star graph  $K_{1,n-1}$  then it does not satisfy the mean-field assumption. Indeed, this follows from observing that

$$\frac{1}{n} \sum_{i=1}^n \lambda_i^2(A_n) = \frac{2|E(K_{1,n-1})|}{n} \rightarrow 2.$$

However, all but 2 of the eigenvalues of the adjacency matrix are zero. Therefore (4.23) holds here, and hence proof of Lemma 3.4 goes through unchanged in this case. For the  $n$ -star graph, one can directly check that the mean-field approximation (1.8) is tight. In light of this and similar other examples, we believe the mean-field assumption can be weakened to (4.23), and we conjecture that the conclusion of Theorem 1.1 continues to hold as long as (4.23) holds.

## 5 Proof of applications

### 5.1 Proofs of Theorems 2.1 and 2.2

In this section we compute the limiting log partition function for asymptotically regular graphs. This is followed by the proof of large deviation principle for the empirical measure of the colors for such graphs.

*Proof of Theorem 2.1* (a) Since  $A_n$  satisfies the mean-field assumption, applying Theorem 1.1 we get

$$\lim_{n \rightarrow \infty} \frac{1}{n} \left[ \Phi_n(\mathbf{J}, \mathbf{h}) - \sup_{\mathbf{q} \in \mathcal{P}(\{[q]\}^n)} M_n^{\mathbf{J}, \mathbf{h}}(\mathbf{q}) \right] = 0.$$

Proceeding to estimate  $M_n^{\mathbf{J}, \mathbf{h}}(\mathbf{q})$ , fixing  $\delta > 0$ , we denote

$$A_n^{(\delta)}(i, j) := A_n(i, j) 1_{|\mathcal{R}_n(i)-1| \leq \delta} 1_{|\mathcal{R}_n(j)-1| \leq \delta}.$$

Thus we have

$$\begin{aligned} & \frac{1}{n} \sum_{r=1}^q \sum_{i, j=1}^n A_n(i, j) q_i(r) q_j(r) \\ & \leq \frac{1}{n} \sum_{r=1}^q \sum_{i, j=1}^n A_n^{(\delta)}(i, j) q_i(r) q_j(r) + \frac{1}{n} \sum_{i: |\mathcal{R}_n(i)-1| > \delta} \sum_{r=1}^q \sum_{j=1}^n A_n(i, j) q_i(r) q_j(r) \\ & \quad + \frac{1}{n} \sum_{j: |\mathcal{R}_n(j)-1| > \delta} \sum_{r=1}^q \sum_{i=1}^n A_n(i, j) q_i(r) q_j(r) \\ & = \frac{1}{n} \sum_{r=1}^q \sum_{i, j=1}^n A_n^{(\delta)}(i, j) q_i(r) q_j(r) + \frac{2}{n} \sum_{i: |\mathcal{R}_n(i)-1| > \delta} \sum_{r=1}^q \sum_{j=1}^n A_n(i, j) q_i(r) q_j(r). \end{aligned} \tag{5.1}$$

Note that the second term in the RHS of (5.1) is bounded above by

$$\frac{2q}{n} \sum_{i:|\mathcal{R}_n(i)-1|>\delta} \mathcal{R}_n(i) \leq \frac{2q}{n} \left[ \sum_{i=1}^n \mathcal{R}_n(i) - (1-\delta) \sum_{i=1}^n 1_{|\mathcal{R}_n(i)-1|\leq\delta} \right] = 2qa_n^{(\delta)}, \tag{5.2}$$

where

$$a_n^{(\delta)} := \frac{1}{n} \left[ \sum_{i=1}^n \mathcal{R}_n(i) - (1-\delta) \sum_{i=1}^n 1_{|\mathcal{R}_n(i)-1|\leq\delta} \right].$$

Considering the first term in the RHS of (5.1), and noting that  $A_n^{(\delta)}$  is a symmetric entries with non negative matrix whose row sums are bounded by  $1 + \delta$ , we apply Gershgorin circle theorem to obtain

$$\frac{1}{n} \sum_{r=1}^q \sum_{i,j=1}^n A_n^{(\delta)}(i, j) q_i(r) q_j(r) \leq \frac{1+\delta}{n} \sum_{r=1}^q \sum_{i=1}^n q_i(r)^2. \tag{5.3}$$

Combining (5.2)–(5.3), along with the expression for  $M_n^{J,h}(\mathbf{q})$ , we get

$$\sup_{\mathbf{q} \in \mathcal{P}(\{q\})^n} \frac{1}{n} M_n^{J,h}(\mathbf{q}) \leq \sup_{\mathbf{q} \in \mathcal{P}(\{q\})^n} \left\{ \frac{\beta}{2} \sum_{r=1}^q q(r)^2 + \sum_{r=1}^q h_r q(r) - \sum_{r=1}^q q(r) \log q(r) \right\} + \frac{q\beta}{2} (\delta + 2a_n^{(\delta)}). \tag{5.4}$$

Now note that  $a_n^{(\delta)} \rightarrow \delta$  as  $n \rightarrow \infty$  by (2.1)–(2.2). Thus taking limits as  $n \rightarrow \infty$  on both sides of (5.4), we get

$$\begin{aligned} & \limsup_{n \rightarrow \infty} \sup_{\mathbf{q} \in \mathcal{P}(\{q\})^n} \frac{1}{n} M_n^{J,h}(\mathbf{q}) \\ & \leq \sup_{\mathbf{q} \in \mathcal{P}(\{q\})} \left\{ \frac{\beta}{2} \sum_{r=1}^q q(r)^2 + \sum_{r=1}^q h_r q(r) - \sum_{r=1}^q q(r) \log q(r) \right\} + 2q\beta\delta, \end{aligned}$$

from which the upper bound of (2.3) follows, as  $\delta > 0$  is arbitrary.

For the lower bound, taking a supremum over all  $\mathbf{q} = \prod_{i=1}^n q_i$  such that  $q_i$  is same for all  $i$ , we have

$$\begin{aligned} & \sup_{\mathbf{q} \in \mathcal{P}(\{q\})^n} \frac{1}{n} M_n^{J,h}(\mathbf{q}) \\ & \geq \sup_{\mathbf{q} \in \mathcal{P}(\{q\})} \left\{ \frac{\beta}{2} \sum_{r=1}^q q(r)^2 \frac{1}{n} \sum_{i=1}^n \mathcal{R}_n(i) + \sum_{r=1}^q h_r q(r) - \sum_{r=1}^q q(r) \log q(r) \right\}, \end{aligned}$$

which on dividing by  $n$ , and taking limits using (2.2), gives the lower bound in (2.3). This completes the proof of part (a).

(b) To prove part (i), we note that  $\mathcal{R}_n(i) = 1$  for all  $i \in [n]$ , and thus both (2.1) and (2.2) hold trivially.

Turning to prove part (ii), note that  $\mathcal{R}_n(i) = \frac{d_i(\mathbf{G}_n)}{np_n}$  with  $d_i(\mathbf{G}_n)$  denoting the degree of vertex  $i \in [n]$ . Since the number of edges  $|E_n|$  has a  $\text{Bin}\left(\binom{n}{2}, p_n\right)$  distribution,

$$\frac{1}{n} \sum_{i=1}^n \mathcal{R}_n(i) = \frac{2|E_n|}{n^2 p_n} \rightarrow 1, \text{ in probability.}$$

This verifies (2.2). To check (2.1), fixing  $\delta > 0$ , it suffices to check that  $\lim_{n \rightarrow \infty} \frac{1}{n} \mathbb{E}N_n^{(\delta)} = 0$ , where

$$N_n^{(\delta)} := \sum_{i=1}^n 1_{|d_i(\mathbf{G}_n) - np_n| > np_n \delta}.$$

This follows using Chebyshev’s inequality:

$$\frac{1}{n} \mathbb{E}N_n^{(\delta)} = \frac{1}{n} \sum_{i \in [n]} \mathbb{P}(|d_i(\mathbf{G}_n) - np_n| > np_n \delta) \leq \frac{(n-1)p_n(1-p_n)}{n^2 p_n^2 \delta^2} \rightarrow 0,$$

as  $np_n \rightarrow \infty$ . □

Now as an application of Theorem 2.1, we derive the following large deviation principle. As a byproduct we also get an exponential concentration of the average sample spins in Ising model.

*Proof of Theorem 2.2* (a) The proof of this theorem is based on Baldi’s theorem (cf. [25, Theorem 4.5.20]). To this end, we first need to compute logarithmic moment generating function. Fixing a vector  $\mathbf{t} = (t_1, t_2, \dots, t_q) \in \mathbb{R}^q$ , using Theorem 2.1, one has

$$\begin{aligned} \frac{1}{n} \log \mathbb{E}_{\mu_n} e^{n \sum_{r \in [q]} t_r L_n(r)} &= \frac{1}{n} [\Phi_n(\mathbf{J}, \mathbf{h} + \mathbf{t}) - \Phi_n(\mathbf{J}, \mathbf{h})] \\ &\xrightarrow{n \rightarrow \infty} \sup_{\mathbf{q} \in \mathcal{P}([q])} \left[ \frac{\beta}{2} \sum_{r=1}^q \mathbf{q}(r)^2 - \sum_{r=1}^q \mathbf{q}(r) \log \mathbf{q}(r) + \sum_{r=1}^q (h_r + t_r) \mathbf{q}(r) \right] \\ &\quad - \sup_{\mathbf{q} \in \mathcal{P}([q])} \left[ \frac{\beta}{2} \sum_{r=1}^q \mathbf{q}(r)^2 - \sum_{r=1}^q \mathbf{q}(r) \log \mathbf{q}(r) + \sum_{r=1}^q h_r \mathbf{q}(r) \right]. \end{aligned}$$

Denoting the RHS above by  $\Lambda(\mathbf{t})$  we note that

$$\Lambda(\mathbf{t}) = \sup_{\mu \in \mathcal{P}([q])} \left\{ \sum_{r \in [q]} t_r \mu_r - \tilde{I}_{\beta, \mathbf{h}}(\mu) \right\}.$$

Therefore, applying the duality lemma (see [25, Lemma 4.5.8]), we have

$$\tilde{I}_{\beta, \mathbf{h}}(\mu) = \sup_{\mathbf{t} \in \mathbb{R}^q} \left\{ \sum_{r \in [q]} t_r \mu_r - \Lambda(\mathbf{t}) \right\}.$$

Next note that the set  $\mathcal{P}([q])$  being compact, the law of  $L_n(\cdot)$  is automatically exponentially tight. Thus using the fact that  $\Lambda(\mathbf{t}) < \infty$  for all  $\mathbf{t} \in \mathbb{R}^q$ , applying [25, Theorem 4.5.20(a)] we obtain that for any closed set  $F \subset \mathcal{P}([q])$ ,

$$\limsup_{n \rightarrow \infty} \frac{1}{n} \log \mu_n(L_n \in F) \leq - \inf_{\mu \in F} \tilde{I}_{\beta, \mathbf{h}}(\mu).$$

To derive the lower bound we use part (b) of [25, Theorem 4.5.20]. To this end, we note that it is enough to prove that any  $\mu \in \mathcal{P}([q])$  is an exposed point of  $\tilde{I}_{\beta, \mathbf{h}}(\cdot)$ , i.e. for any  $\nu \in \mathcal{P}([q])$  with  $\mu \neq \nu$  there exists  $\mathbf{t} \in \mathbb{R}^q$  such that

$$\begin{aligned} & \sum_{r \in [q]} \left\{ \frac{\beta}{2} \sum_{r=1}^q \mu_r^2 - \sum_{r=1}^q \mu_r \log \mu_r + \sum_{r=1}^q (h_r + t_r) \mu_r \right\} \\ & > \sum_{r \in [q]} \left\{ \frac{\beta}{2} \sum_{r=1}^q \nu_r^2 - \sum_{r=1}^q \nu_r \log \nu_r + \sum_{r=1}^q (h_r + t_r) \nu_r \right\}. \end{aligned}$$

This follows on noting the existence of  $r \in [q]$  such that  $\mu_r > \nu_r$ , and then choosing  $t_r$  large enough for all  $r$  such that  $\mu_r > \nu_r$ , and  $t_r = 0$  for all  $r$  such that  $\mu_r \leq \nu_r$ .

Now to prove (2.4), we note that the function  $\mu \mapsto I_{\beta, \mathbf{h}}(\mu)$  is a non constant analytic function on a compact set, and thus the infimum is attained on a finite set  $K_{\beta, \mathbf{h}}$ . Thus (2.4) follows from the large deviation principle on noting that the set  $\{v \in \mathcal{P}([q]) : \min_{\mu \in K_{\beta, \mathbf{h}}} \|v - \mu\|_{\infty} \geq \delta\}$  is closed.

(b) By the last conclusion of part (a) it suffices to minimize the function  $I_{\beta, \mathbf{h}}(\mu)$ . To begin introduce the variable  $m = \mu_1 - \mu_2 \in [-1, 1]$  and note that

$$I_{\beta, \mathbf{h}}(\mu) = -\frac{\beta}{4} m^2 - \frac{B}{2} m + H(m) - \left[ \frac{\beta}{4} - \frac{h_1 + h_2}{2} \right].$$

The optimization of this function has been carried out in [21, Section 1.1.3], where it is shown that optimum is at  $m = 0$  for  $\beta \leq 2, B = 0$ , at  $m = \pm m_{\beta/2, 0}$  for  $\beta > 2, B = 0$ , and at  $m = m_{\beta/2, B/2}$  for  $\beta > 0, B \neq 0$ . This, along with the symmetry of the Ising model for  $B = 0$  completes the proof of part (b).  $\square$

### 5.2 Proofs of Theorems 2.3 and 2.4

In this section we prove the convergence of log partition function for bi-regular bipartite graphs, followed by the same for a sequence of graphs converging in cut metric.

*Proof of Theorem 2.3* To begin first note that (2.5), along with  $a_n c_n = (n - a_n) d_n$  implies  $c_n = \Theta(d_n)$ , and therefore we deduce that  $c_n$  and  $d_n$  individually converge to  $\infty$ , as  $n \rightarrow \infty$ . This further implies that  $|E_n| = a_n c_n \gg n$ . Thus Corollary 1.2 is applicable, and it suffices only to consider the asymptotics of  $\sup_{q \in \mathcal{P}(\{q\})^n} M_n^{\beta, B}(q)$ . For computing the supremum in this setting, denoting  $b_n := n - a_n$  we have

$$M_n^{\beta, B}(q) = \beta \sum_{i \in [a_n], j \in [b_n], r \in [2]} q_i^{(1)}(r) q_j^{(2)}(r) A_n(i, j) - \sum_{i \in [a_n], r \in [2]} q_i^{(1)}(r) \log q_i^{(1)}(r) - \sum_{j \in [b_n], r \in [2]} q_j^{(2)}(r) \log q_j^{(2)}(r).$$

Introducing variables  $s_i^{(1)} := q_i^{(1)}(1) - q_i^{(1)}(2)$ , and  $s_j^{(2)} := q_j^{(2)}(1) - q_j^{(2)}(2)$ , and noting that  $\sum_{k \in [2]} q_i^{(1)}(k) = \sum_{k \in [2]} q_j^{(2)}(k) = 1$ , the RHS above becomes

$$\begin{aligned} &= \frac{\beta}{2} \sum_{i \in [a_n], j \in [b_n]} (1 + s_i^{(1)} s_j^{(2)}) A_n(i, j) + \sum_{i \in [a_n]} H(s_i^{(1)}) + \sum_{j \in [b_n]} H(s_j^{(2)}) \\ &= \frac{\beta}{2} \sum_{i \in [a_n], j \in [b_n]} s_i^{(1)} s_j^{(2)} A_n(i, j) + \sum_{i \in [a_n]} H(s_i^{(1)}) + \sum_{j \in [b_n]} H(s_j^{(2)}) + \frac{\beta}{2} \frac{a_n c_n}{c_n + d_n}. \end{aligned} \tag{5.5}$$

Hence, it suffices to maximize (5.5) over the set  $\{s_i^{(1)} \in [-1, 1], i \in [a_n]; s_j^{(2)} \in [-1, 1], j \in [b_n]\}$ .

Fixing  $n$  first note that the optimum occurs at an interior point where  $s_i^{(1)} \in (-1, 1)$ ,  $s_j^{(2)} \in (-1, 1)$ , for any  $i \in [a_n], j \in [b_n]$ . This is due to the facts that for any  $i \in [a_n]$ , we have

$$\frac{\partial}{\partial s_i^{(1)}} M_n^{\beta, B} \Big|_{s_i^{(1)} \rightarrow -1+} = +\infty, \quad \frac{\partial}{\partial s_i^{(1)}} M_n^{\beta, B} \Big|_{s_i^{(1)} \rightarrow 1-} = -\infty,$$

and a similar argument holds for  $s_j^{(2)}$  for  $j \in [b_n]$ , as well. Thus differentiating with respect to  $s_i^{(1)}$ ,  $s_j^{(2)}$  and equating to 0, any optimum satisfies the following equations

$$s_i^{(1)} = \tanh \left( \beta \sum_{j \in [b_n]} A_n(i, j) s_j^{(2)} \right), \tag{5.6}$$

$$s_j^{(2)} = \tanh \left( \beta \sum_{i \in [a_n]} A_n(i, j) s_i^{(1)} \right). \tag{5.7}$$

We now split the proof into four different cases.

**Case 1:**  $\beta > 0$ , and  $\beta^2 p(1 - p) < 1$ .

Since  $\beta > 0$ , and  $H(x) = H(-x)$ , without loss of generality we can assume that for any optimum we have  $\mathfrak{s}_i^{(1)}, \mathfrak{s}_j^{(2)} \geq 0$ . Next combining (5.6), and (5.7), for every  $i \in [a_n]$  we get

$$\mathfrak{s}_i^{(1)} = \tanh \left( \beta \sum_{j \in [b_n]} A_n(i, j) \tanh \left( \beta \sum_{k \in [a_n]} A_n(j, k) \mathfrak{s}_k^{(1)} \right) \right). \tag{5.8}$$

Letting  $\mathfrak{s}_{i_0}^{(1)} := \arg \max_{i \in [a_n]} \mathfrak{s}_i^{(1)}$ , (5.8) further yields

$$\begin{aligned} \mathfrak{s}_{i_0}^{(1)} &= \tanh \left( \beta \sum_{j \in [b_n]} A_n(i, j) \tanh \left( \beta \sum_{k \in [a_n]} A_n(j, k) \mathfrak{s}_k^{(1)} \right) \right) \\ &\leq \tanh \left( \beta \sum_{j \in [b_n]} A_n(i, j) \tanh \left( \beta \sum_{k \in [a_n]} A_n(j, k) \mathfrak{s}_{i_0}^{(1)} \right) \right) \\ &= \tanh \left( \beta \frac{c_n}{c_n + d_n} \tanh \left( \beta \mathfrak{s}_{i_0}^{(1)} \frac{d_n}{c_n + d_n} \right) \right) =: \eta_{\beta, \frac{d_n}{c_n + d_n}}(\mathfrak{s}_{i_0}^{(1)}). \end{aligned} \tag{5.9}$$

It is easy to note that

$$\left\| \frac{d\eta_{\beta, \frac{d_n}{c_n + d_n}}(\mathfrak{s})}{d\mathfrak{s}} \right\|_{\infty} = \beta^2 \frac{c_n d_n}{(c_n + d_n)^2} \xrightarrow{n \rightarrow \infty} \beta^2 p(1 - p) < 1.$$

Thus  $\mathfrak{s} \mapsto \eta_{\beta, \frac{d_n}{c_n + d_n}}(\mathfrak{s})$  is a contraction. This implies that for any  $\mathfrak{s} > 0$

$$\eta_{\beta, \frac{d_n}{c_n + d_n}}(\mathfrak{s}) = \left| \eta_{\beta, \frac{d_n}{c_n + d_n}}(\mathfrak{s}) - \eta_{\beta, \frac{d_n}{c_n + d_n}}(0) \right| < |\mathfrak{s} - 0| = \mathfrak{s},$$

for all large  $n$ . Using (5.9), for large  $n$ , we therefore deduce that  $\mathfrak{s}_{i_0}^{(1)}$  must be equal to zero. This further implies that  $\mathfrak{s}_i^{(1)}$  must be equal to zero for all  $i \in [a_n]$ . Similar arguments hold for  $\mathfrak{s}_j^{(2)}$ , proving  $\mathfrak{s}_j^{(2)} = 0$  for all  $j \in [b_n]$ . Plugging in the values of  $\mathfrak{s}_i^{(1)}$ , and  $\mathfrak{s}_j^{(2)}$  in the RHS of (5.5) we have

$$\sup_{q \in \mathcal{P}([q]^n)} M_n^{\beta, 0}(q) = \frac{\beta}{2} \frac{a_n c_n}{c_n + d_n} + n \log 2,$$

which on dividing by  $n$  and taking limits proves **Case 1**.

**Case 2:**  $\beta < 0$ , and  $\beta^2 p(1 - p) < 1$ .

Note that one can rewrite  $M_n^{\beta,0}(q)$  as

$$\frac{(-\beta)}{2} \sum_{i \in [a_n], j \in [b_n]} \mathfrak{s}_i^{(1)} (-\mathfrak{s}_j^{(2)}) A_n(i, j) + \sum_{i \in [a_n]} H(\mathfrak{s}_i^{(1)}) + \sum_{j \in [b_n]} H(-\mathfrak{s}_j^{(2)}) + \frac{\beta}{2} \frac{a_n c_n}{c_n + d_n}.$$

Since  $\beta < 0$ , one can argue that for any optimum we must have  $\mathfrak{s}_i^{(1)}$  and  $-\mathfrak{s}_j^{(2)}$  non negative for all  $i \in [a_n]$ , and  $j \in [b_n]$ . The rest of the arguments is similar to **Case 1**. We omit the details.

**Case 3:**  $\beta > 0$ , and  $\beta^2 p(1 - p) > 1$ .

We begin by noting that

$$\frac{d\eta_{\beta, \frac{d_n}{c_n+d_n}}}{d\mathfrak{s}}(\mathfrak{s}) = \beta^2 \frac{c_n d_n}{(c_n + d_n)^2} \operatorname{sech}^2 \left( \tanh \left( \beta \mathfrak{s} \frac{d_n}{c_n + d_n} \right) \right) \operatorname{sech}^2 \left( \beta \mathfrak{s} \frac{d_n}{c_n + d_n} \right),$$

which is decreasing in  $\mathfrak{s}$ , and goes to zero as  $\mathfrak{s} \rightarrow \infty$ . Further noting that

$$\left. \frac{d\eta_{\beta, \frac{d_n}{c_n+d_n}}}{d\mathfrak{s}}(\mathfrak{s}) \right|_{\mathfrak{s}=0} = \beta^2 \frac{c_n d_n}{(c_n + d_n)^2} \xrightarrow{n \rightarrow \infty} \beta^2 p(1 - p) > 1,$$

we deduce that there is a unique positive root of the equation  $\mathfrak{s} = \eta_{\beta, \frac{d_n}{c_n+d_n}}(\mathfrak{s})$ , denoted by  $\mathfrak{s}_{\beta, \frac{d_n}{c_n+d_n}}$ , for all  $n$  large enough. Also (5.9) implies

$$\max_{i \in [a_n]} \mathfrak{s}_i^{(1)} = \mathfrak{s}_{i_0}^{(1)} \leq \mathfrak{s}_{\beta, \frac{d_n}{c_n+d_n}}.$$

By a similar argument we also deduce

$$\min_{i \in [a_n]} \mathfrak{s}_i^{(1)} \geq \mathfrak{s}_{\beta, \frac{d_n}{c_n+d_n}},$$

and so  $\mathfrak{s}_i^{(1)} = \mathfrak{s}_{\beta, \frac{d_n}{c_n+d_n}}$  for all  $i \in [a_n]$ . Plugging in this solution in (5.7) gives  $\mathfrak{s}_j^{(2)} = \mathfrak{s}_{\beta, \frac{c_n}{c_n+d_n}}$  for all  $j \in [b_n]$ . Thus the optimum solution is

$$\mathfrak{s}_i^{(1)} = \mathfrak{s}_{\beta, \frac{d_n}{c_n+d_n}}, \text{ for all } i \in [a_n], \quad \mathfrak{s}_j^{(2)} = \mathfrak{s}_{\beta, \frac{c_n}{c_n+d_n}} \text{ for all } j \in [b_n].$$

Plugging in this optimal solution in the RHS of (5.5) gives

$$\begin{aligned} \sup_{q \in \mathcal{P}(\{q\})^n} M_n^{\beta,0}(q) &= \frac{\beta a_n c_n}{2(c_n + d_n)} \left\{ 1 + \mathfrak{s}_{\beta, \frac{d_n}{c_n+d_n}} \mathfrak{s}_{\beta, \frac{c_n}{c_n+d_n}} \right\} \\ &\quad + a_n H \left( \mathfrak{s}_{\beta, \frac{d_n}{c_n+d_n}} \right) + b_n H \left( \mathfrak{s}_{\beta, \frac{d_n}{c_n+d_n}} \right), \end{aligned}$$



from which part (b) follows on dividing by  $n$  and taking limits, on noting that the function  $p \mapsto \mathfrak{s}_{\beta,p}$  is continuous.

**Case 4:**  $\beta < 0$ , and  $\beta^2 p(1 - p) > 1$ .

This can be done by combining the arguments of **Case 2**, and **Case 3**. We omit the details.

Note that the above four cases complete the proof, barring the convergence of  $\Phi_n(\beta, 0)$  at  $\beta = \pm\beta_c(p) = \pm\sqrt{p(1 - p)}$ . To complete the proof first we use the fact that  $|\tanh(x)| < |x|$ , for any  $x \neq 0$ , and deduce that  $\mathfrak{s}_{\beta,p} \rightarrow 0$ , as  $\beta \rightarrow \pm\beta_c$ . This implies that

$$\Phi(\beta, 0) := \frac{\beta p(1 - p)}{2} + \frac{|\beta|p(1 - p)}{2} \mathfrak{s}_{|\beta|,p\mathfrak{s}_{|\beta|,1-p}} + p H(\mathfrak{s}_{|\beta|,p}) + (1 - p) H(\mathfrak{s}_{|\beta|,1-p})$$

is continuous for all  $\beta$ . Since  $\{\Phi_n(\cdot, 0)\}$  are convex functions, and limit of such functions is also a convex function, using the fact that  $\limsup_{n \rightarrow \infty} \frac{1}{n} \Phi_n(\beta, 0) < \infty$  at  $\beta = \pm\beta_c(p)$ , the proof completes by a standard analysis argument.  $\square$

*Remark 5.1* Even though we do not pursue it here, by combining the arguments of Theorems 2.1 and 2.3 one should be able to prove Theorem 2.3 for a sequence asymptotically bi-regular bipartite graphs. We believe a similar universality result for the limiting log partition function for the  $q$  Potts model holds for general  $q$ -partite graphs as well, though proving it will require an analysis of fixed points in  $q$  dimensional equations for  $q > 2$ .

Finally we prove Theorem 2.4.

*Proof of Theorem 2.4* By assumption  $W_{nA_n}$  converges to  $W$  in the cut metric, and therefore by [8, Proposition C5 and Proposition C15], we have  $\lim_{n \rightarrow \infty} \frac{|E_n|}{n} = \infty$ . Thus applying Corollary 1.2, we note that it suffices to show

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sup_{q \in (\mathcal{P}[q])^n} M_n^{J,h}(q) = \sup_{\rho \in \mathbf{FP}_q} F^{J,h}(W, \rho). \tag{5.10}$$

To this end, setting  $\rho_r(x) = q_i(r)$  for  $(\frac{i-1}{n}, \frac{i}{n}]$  for each  $1 \leq r \leq q, 1 \leq i \leq n$ , we note that

$$\frac{1}{n} M_n^{J,h}(q) = F^{J,h}(W_{nA_n}, \rho). \tag{5.11}$$

Since  $nW_{A_n}$  converges to  $W$  in the cut metric we have

$$\sup_{\rho \in \mathbf{FP}_q} |F^{J,h}(W_{nA_n}, \rho) - F^{J,h}(W, \rho)| \rightarrow 0. \tag{5.12}$$

This implies that

$$\limsup_{n \rightarrow \infty} \frac{1}{n} \sup_{q \in (\mathcal{P}[q])^n} M_n^{J,h}(q) \leq \sup_{\rho \in \mathbf{FP}_q} F^{J,h}(W, \rho).$$

Thus to establish (5.10), we need to prove the other side of the inequality. Turning to prove the same, we note that it suffices to show that given any  $\rho \in \mathbf{FP}_q$  there exists  $\rho^{(n)} \in \mathbf{FP}_q$ , with  $\rho_r$  being constant on  $(\frac{i-1}{n}, \frac{i}{n}]$  for  $1 \leq i \leq n$ ,  $1 \leq r \leq q$ , such that

$$\lim_{n \rightarrow \infty} F^{J,h}(W, \rho^{(n)}) = F^{J,h}(W, \rho). \quad (5.13)$$

Indeed, using (5.11), we deduce that, for any  $\rho \in \mathbf{FP}_q$

$$F^{J,h}(W, \rho) \leq \left| F^{J,h}(W, \rho) - F^{J,h}(W_{nA_n}, \rho^{(n)}) \right| + \frac{1}{n} M_n^{J,h}(q).$$

Next taking a supremum over  $q \in \mathcal{P}([q])^n$ , followed by a liminf on the both sides, and using (5.12), and (5.13), we further obtain that

$$F^{J,h}(W, \rho) \leq \liminf_{n \rightarrow \infty} \frac{1}{n} \sup_{q \in \mathcal{P}([q])^n} M_n^{J,h}(q).$$

Next taking another supremum over  $\rho \in \mathbf{FP}_q$ , we complete the proof of (5.10).

Now it only remains to establish (5.13). A standard measure theoretic arguments yields the existence  $\rho^{(n)} \in \mathbf{FP}_q$ , with  $\rho_r$  being constant on  $(\frac{i-1}{n}, \frac{i}{n}]$  for  $1 \leq i \leq n$ ,  $1 \leq r \leq q$ , such that

$$\lim_{n \rightarrow \infty} \max_{r=1}^q |\rho_r^{(n)}(x) - \rho_r(x)| = 0, \text{ Lebesgue almost surely.} \quad (5.14)$$

Therefore, noting  $\|W\|_1 < \infty$ , using dominated convergence theorem, and the fact that the function  $x \mapsto x \log x$  is continuous on  $[0, 1]$  we prove (5.13). This completes the proof of the theorem.  $\square$

**Acknowledgments** We thank Andrea Montanari for suggesting to look at the Ising measure on hypercube, Sourav Chatterjee for pointing out the reference [16], and Amir Dembo for helpful comments on earlier version of the manuscript. We also thank Sourav Chatterjee, Amir Dembo, and Andrea Montanari for many helpful discussions. We further thank Marek Biskup and Aernout Van Enter for pointing out the references [6] and [13] respectively. We are grateful to two anonymous referees for their detailed comments and suggestions which have improved the quality of this paper.

## References

1. Anandkumar, A., Tan, V., Huang, F., Willsky, A.: High-dimensional structure estimation in Ising models: local separation criterion. *Ann. Stat.* **40**(3), 1346–1375 (2012)
2. Bayati, M., Gamarnik, D., Tetali, P.: Combinatorial approach to the interpolation method and scaling limits in sparse random graphs. *Ann. Probab.* **41**(6), 4080–4115 (2013)
3. Bai, Z.D.: Methodologies in spectral analysis of large-dimensional random matrices, a review (with discussion). *Stat. Sin.* **9**(3), 611–662 (1999)
4. Bento, J., Montanari, A.: Which graphical models are difficult to learn? *Neural Inf. Process. Syst.* **22**, 1303–1311 (2009)
5. Bhattacharya, B., Mukherjee, S.: Inference in Ising model (2015, preprint). [arXiv:1507.07055](https://arxiv.org/abs/1507.07055)
6. Biskup, M., Chayes, L.: Rigorous analysis of discontinuous phase transitions via mean-field bounds. *Commun. Math. Phys.* **238**(1), 53–93 (2003)

7. Blanchard, P., Gandolfo, D., Ruiz, J., Wouts, M.: Thermodynamic vs topological phase transition: cusp in the Kertész line. *Europhys. Lett.* **82**(5), 50003 (2008)
8. Borgs, C., Chayes, J., Cohn, H., Zhao, Y.: An  $L_p$  theory of sparse convergence I: limits, sparse random graph models, and power law distributions (preprint). <http://yufeizhao.com/papers/2014-LpLimit1>
9. Borgs, C., Chayes, J., Cohn, H., Zhao, Y.: An  $L_p$  theory of sparse convergence II: LD convergence, quotients, and right convergence (preprint). <http://yufeizhao.com/papers/2014-LpLimit2>
10. Borgs, C., Chayes, J., Lovász, L., Sós, V.T., Vesztergombi, K.: Convergent sequences of dense graphs I: subgraph frequencies, metric properties and testing. *Adv. Math.* **219**, 1801–1851 (2009)
11. Borgs, C., Chayes, J., Lovász, L., Sós, V.T., Vesztergombi, K.: Convergent sequences of dense graphs II. Multiway cuts and statistical physics. *Ann. Math.* **176**, 151–219 (2012)
12. Bresler, G.: Efficiently learning Ising models on arbitrary graphs. *Symposium on Theory of Computing (STOC)*, pp. 771–782 (2015)
13. Cant, A., Pearce, P.A.: Mean-field limits of the quantum Potts model. *Commun. Math. Phys.* **90**(3), 373–383 (1983)
14. Chandler, D.: *Introduction to Modern Statistical Mechanics*. Oxford University Press, Oxford (1987)
15. Chatterjee, S.: Estimation in spin glasses: a first step. *Ann. Stat.* **35**(5), 1931–1946 (2007)
16. Chatterjee, S., Dembo, A.: Nonlinear large deviations. *Adv. Math.* (to appear) [arXiv:1401.3495v5](https://arxiv.org/abs/1401.3495v5)
17. Coja-Oghlan, A., Lanka, A.: The spectral gap of random graphs with given expected degrees. *Electron. J. Combin.* **16**(1), R138 (2009)
18. Costeniuc, M., Ellis, R.S., Touchette, H.: Complete analysis of phase transitions and ensemble equivalence for the Curie–Weiss–Potts model. *J. Math. Phys.* **46**(6), 063301 (2005)
19. Chung, F., Lu, L., Vu, V.: The spectra of random graphs with given expected degrees. *J. Internet Math.* **1**(3), 257–275 (2003)
20. Dembo, A., Montanari, A.: Ising models on locally tree-like graphs. *Ann. Appl. Probab.* **20**(2), 565–592 (2010)
21. Dembo, A., Montanari, A.: Gibbs measures and phase transitions on sparse random graphs. *Br. J. Probab. Stat.* **24**(2), 137–211 (2010)
22. Dembo, A., Montanari, A., Sen, S.: Extremal cuts of sparse random graphs. *Ann. Probab.* (to appear). [arXiv:1503.03923](https://arxiv.org/abs/1503.03923)
23. Dembo, A., Montanari, A., Sun, N.: Factor models on locally tree-like graphs. *Ann. Probab.* **41**(6), 4162–4213 (2013)
24. Dembo, A., Montanari, A., Sly, A., Sun, N.: The replica symmetric solution for Potts models on  $d$ -regular graphs. *Commun. Math. Phys.* **327**(2), 551–575 (2014)
25. Dembo, A., Zeitouni, O.: *Large Deviations Techniques and Applications*, vol. 38. Springer, Berlin (2009)
26. Dommers, S., Giardinà, C., van der Hofstad, R.: Ising models on power-law random graphs. *J. Stat. Phys.* **141**(4), 638–660 (2010)
27. Eichelsbacher, P., Martschink, B.: On rates of convergence in the Curie–Weiss Potts model with an external field. *Annales de l’Institut Henri Poincaré Probabilités et Statistiques* **51**(1), 252–282 (2015)
28. Ellis, R.S., Newman, M.E.: The statistics of Curie–Weiss models. *J. Stat. Phys.* **19**(2), 149–161 (1978)
29. Ellis, R.S., Wang, K.: Limit theorems for the empirical vector of the Curie–Weiss–Potts model. *Stochastic Process. Appl.* **35**(1), 59–79 (1990)
30. Ellis, R.S., Wang, K.: Limit theorems for maximum likelihood estimators in the Curie–Weiss–Potts model. *Stochastic Process. Appl.* **40**(2), 251–288 (1992)
31. Gamarnik, D.: Correlation decay method for decision, optimization, and inference in large-scale networks. *Tutorials in Operations Research, INFORMS* (2013)
32. Feige, U., Ofek, E.: Spectral techniques applied to sparse random graphs. *Random Struct. Algorithms* **27**(2), 251–275 (2005)
33. Hopfield, J.: Neural networks and physical systems with emergent collective computational abilities. *Proc. Natl. Acad. Sci.* **79**(8), 2554–2558 (1982)
34. Ising, E.: Beitrag zur theorie der ferromagnetismus. *Z. Phys.* **31**(1), 253–258 (1925)
35. Jerrum, M., Sinclair, A.: Polynomial-time approximation algorithms for the Ising model. *SIAM J. Comput.* **22**(5), 1087–1116 (1993)
36. Krivelevich, M., Sudakov, B.: The largest eigenvalue of sparse random graphs. *Combin. Probab. Comput.* **12**(1), 61–72 (2003)
37. Lovász, L.: *Large Networks and Graph Limits*, vol. 60. AMS, Providence (2012)
38. Parisi, G.: *Statistical Field Theory*. Addison-Wesley, New York (1988)

39. Milman, V.D., Schechtman, G.: Asymptotic theory of finite-dimensional normed spaces. Lecture Notes in Mathematics, vol. 1200. Springer, Berlin (1986)
40. Niss, M.: History of the Lenz-Ising model 1920–1950: from ferromagnetic to cooperative phenomena. *Archi. Hist. Exact Sci.* **59**(3), 267–318 (2005)
41. Potts, R.: Some generalized order-disorder transformations. *Math. Proc. Camb. Philos. Soc.* **48**(1), 106–109 (1952)
42. Ravikumar, P., Wainwright, M.J., Lafferty, J.: High-dimensional Ising model selection using  $\ell_1$ -regularized logistic regression. *Ann. Stat.* **38**(3), 1287–1319 (2010)
43. Sherrington, D., Kirkpatrick, S.: Solvable model of a spin-glass. *Phys. Rev. Lett.* **35**(26), 1792–1796 (1975)
44. Sly, A., Sun, N.: Counting in two-spin models on  $d$ -regular graphs. *Ann. Probab.* **42**(6), 2383–2416 (2014)
45. Talagrand, M.: The Parisi formula. *Ann. Math.* **163**(2), 221–263 (2006)
46. Wainwright, M.J., Jaakkola, T.S., Willsky, A.S.: A new class of upper bounds on the log partition function. *IEEE Trans. Inf. Theory* **51**(7), 2313–2335 (2005)
47. Wu, F.Y.: The Potts model. *Rev. Modern Phys.* **54**(1), 235 (1982)