

Joint Estimation in Potts Model

Somabha Mukherjee,¹ Sumit Mukherjee² and Sayar Karmakar³

¹*Department of Statistics and Data Science, National University of Singapore, Singapore*
e-mail: somabha@nus.edu.sg

²*Department of Statistics, Columbia University, USA* e-mail: sm3949@columbia.edu

³*Department of Statistics, University of Florida, USA* e-mail: sayarkarmakar@ufl.edu

Abstract: In this paper, we study estimation of parameters in a two-parameter Potts model with q colors and coupling matrix \mathbf{A}_N . We characterize concrete sufficient conditions for existence of the pseudo-likelihood estimator of the Potts model, in terms of the local magnetic fields, and give sufficient conditions for the validity of the above characterization. We then provide sufficient criteria for estimation of both parameters at the optimal rate \sqrt{N} . In particular, if \mathbf{A}_N is the scaled adjacency matrix of a graph G_N , then we show that joint estimation is possible if either G_N has bounded degree or is irregular. In contrast, we give an example of a graph sequence G_N which is approximately regular and dense, where no consistent estimator exists. We also show that one-parameter estimation at the optimal rate \sqrt{N} holds under much milder conditions when the other parameter is known. Along the way, we develop a concentration result for mean-field Potts models using the framework of nonlinear large deviations. Compared to the Ising case, our results for the Potts case require a novel analysis across multiple colors.

Keywords and phrases: Potts model, pseudo-likelihood, random graphs, phase transitions.

1. Introduction

The *Potts model*, whose origin can be traced back to the 1900s (see [Ashkin and Teller \(1943\)](#)), is a statistical physics model for capturing dependence in complex stochastic systems. What began as a generalization of the *Ising model* (see [Ising \(1925\)](#)) in order to accommodate spins with more than two values (see [Potts \(1952\)](#); [Wu \(1982\)](#)) has, over the past several decades, found widespread applications in a number of diverse fields including biomedical problems ([Boas et al., 2018](#); [Moltchanova, Pitkaniemi and Haapala, 2005](#)), image processing and computer vision ([Celeux, Forbes and Peyrard, 2002](#); [Levada, Mascarenhas and Tannús, 2009](#)), spatial statistics ([Zukovic, 2008](#)), social sciences ([Bosconti et al., 2015](#)), finance ([Takaishi, 2005](#); [Bornholdt, 2021](#)) and automata theory ([Graner and Glazier, 1992](#)), among others.

The q -state Potts model, for any positive integer q can be described as a discrete probability distribution supported on the set $[q]^N$. Here and henceforth, the notation $[m]$, for any $m \in \mathbb{N}$, denotes the set $\{1, 2, \dots, m\}$. The positive integer $N \in \mathbb{N}$ indicates the size of the system (number of interacting particles in the system) under consideration. This distribution is given by the probability

mass function

$$\mathbb{P}_{\beta, \mathbf{B}}(\mathbf{x}) := \frac{1}{Z_N(\beta, \mathbf{B})} \exp \left(\frac{\beta}{2} \sum_{1 \leq i, j \leq N} a_{ij} \mathbb{1}_{x_i = x_j} + \sum_{i=1}^N \sum_{r=1}^q B_r \mathbb{1}_{x_i = r} \right) \text{ for } \mathbf{x} \in [q]^N, \quad (1)$$

where $\beta > 0$ represents the *inverse temperature*, $\mathbf{B} := (B_1, \dots, B_{q-1}) \in \mathbb{R}^{q-1}$ represents the *magnetic field* vector, and $\mathbf{A}_N := ((a_{ij}))_{1 \leq i, j \leq N}$ is a symmetric matrix, with zeros on the diagonal. We will refer to \mathbf{A}_N as the *coupling/interaction matrix*. Note here that we did not include a non-zero magnetic field parameter B_q for identifiability reasons, since otherwise, the model remains unchanged if the same constant is added to all the magnetic fields. Throughout the paper, we will use by convention the notation $B_q := 0$. Some of the commonest examples of coupling matrices are suitably scaled adjacency matrices of graphs, defined via

$$a_{ij} := \frac{N}{2|E(G_N)|} \mathbb{1}(i \text{ and } j \text{ form an edge in } G_N), \text{ for all } i, j \in [N], \quad (2)$$

where G_N is any graph on vertex set $[N]$ and edge set $E(G_N)$.

Equation (1) takes the form of a discrete exponential family with natural parameters $\beta > 0$ and $\mathbf{B} \in \mathbb{R}^{q-1}$. The problem we address in our paper is the estimation of these parameters given a *single* sample $\mathbf{X} := (X_1, \dots, X_N)$ from this model. The unavailability of multiple, mutually independent random vectors sampled from the same distribution (as is typical in epidemics, elections, or criminal activity, where the underlying network is typically observed only once, and replications are spatio-temporally dependent) is what poses the primary challenge in this problem. Throughout this paper, we assume that all entries of the matrix \mathbf{A}_N are non-negative and completely known.

The choice $q = 2$ corresponds to the Ising model, and in this special case, changing the domain of \mathbf{x} from $[2]^N = \{1, 2\}^N$ to $\{\pm 1\}^N$, one can write

$$\sum_{i,j=1}^n a_{ij} \mathbb{1}\{x_i = x_j\} = \frac{1}{2} \sum_{i,j=1}^N a_{ij} (1 + x_i x_j), \quad \sum_{i=1}^N \mathbb{1}\{x_i = 1\} = \frac{N}{2} (1 + \bar{x}).$$

Plugging these in (1), the pmf of the Ising model on $\{\pm 1\}^N$ can thus be written as

$$\mathbb{P}_{\beta, B}(\mathbf{X} = \mathbf{x}) \propto \exp \left\{ \frac{\beta}{4} \mathbf{x}' \mathbf{A}_N \mathbf{x} + \frac{B_1}{2} \sum_{i=1}^N x_i \right\}, \text{ for } \mathbf{x} = (x_1, \dots, x_N) \in \{\pm 1\}^N, \quad (3)$$

Statistical inference for general Ising and Potts models traces back to the seminal work Chatterjee (2007a), which analyzed the Ising model (3) in the absence of an external field ($B_1 = 0$). Allowing the coupling matrix \mathbf{A}_N to have both positive and negative entries, under bare minimal conditions Chatterjee (2007a) establishes \sqrt{N} -consistency of the *maximum pseudo-likelihood estimator* of the natural parameter β of this one-parameter exponential family. In particular,

the results of this paper apply to the well known spin glass models such as the celebrated Sherrington-Kirkpatrick model, and the Hopfield model of neural networks. One of the many open questions raised in Chatterjee (2007a) speculates whether the methods developed in Chatterjee (2007a) can be adapted for estimation in multi-parameter models. As an answer to this, Ghosal et al. (2020) considers the two-parameter Ising model in (3) when the coupling matrix \mathbf{A}_N has non-negative entries, and studies the joint estimation of the inverse temperature parameter and the magnetization parameter, i.e. the pair (β, B_1) . In this paper, we will study the analogous question of joint estimation of (β, \mathbf{B}) for the more general Potts model (1).

1.1. Literature Review

The problem of statistical inference in statistical physics models has a body of growing literature, and here we cite some of the relevant literature close to our work. Some of the earliest rigorous studies for the Curie-Weiss Ising model were done in Ellis, Newman and Rosen (1980); Ellis (1985), which established the CLT for the magnetization for the Curie-Weiss Ising model ((3) with $\mathbf{A}_N(i, j) = N^{-1} \mathbb{1}_{i \neq j}$). Subsequently, Comets and Gidas (1991) studied asymptotics of the MLE in the Curie-Weiss Ising model, and showed that one-parameter estimation is possible if the other parameter is known. Going beyond the Curie-Weiss Ising model in a significant way, Chatterjee (2007a) studies the performance of the pseudo-likelihood estimator for the one-parameter Ising model with the temperature parameter $\beta > 0$ unknown for a general matrix \mathbf{A}_N , with the magnetization parameter $B_1 = 0$. In this paper the author allows the coupling matrix \mathbf{A}_N to have both positive and negative values, and gives a sufficient criterion for estimation of β at the optimal rate \sqrt{N} , which covers both graphical models as well as spin glass models. In a follow up work, in Bhattacharya and Mukherjee (2018) the authors extend this to show that the estimation rate for β depends on the order of the log normalizing constant $\log Z_N(\cdot, \mathbf{B})$ in a local neighborhood of the truth β . Using this, they demonstrate phase transitions in the rate of the pseudo-likelihood estimator, which is typically dictated by the critical temperature of the Ising model. The problem of estimating both the parameters (β, B_1) in the Ising model was first studied in Ghosal et al. (2020), where the authors assume that the coupling matrix \mathbf{A}_N is non-negative entry-wise. Under this assumption, Ghosal et al. (2020) show that joint estimation is possible at the optimal rate \sqrt{N} if either the coupling matrix is irregular (see (14)) or non-mean field (see (13)). In contrast, if the coupling matrix is both regular and mean-field, they give an example to show that joint consistent estimation may be impossible. In a more recent paper Chen, Sen and Wu (2024), the authors show that joint estimation is possible for spin glass models, where the coupling matrix \mathbf{A}_N can take both positive and negative values.

Prior to this work, a number of studies have explored statistical inference in Potts models and more general Markov random fields (see, for example, Ali,

Farag and Gimel'farb (2008); Gimenez, Frery and Flesia (2013); Descombes et al. (1999); Okabayashi, Johnson and Geyer (2011); McGrory et al. (2009); Song et al. (2016); Pereyra et al. (2013, 2014); Rosu et al. (2015); Levada, Mascarenhas and Tannús (2008); Levada et al. (2008)). While these contributions provide valuable insights and methodological developments, a fully rigorous treatment of consistency for joint parameter estimators in general Potts models with $q > 2$ colors has not yet been established. To the best of our knowledge, the present work is the first to address this question. In fact, even in the single-parameter setting, rigorous results on consistent estimation in the Potts framework are largely absent, with the notable exception of the Curie–Weiss Potts model (see Ellis and Wang (1992); Bhowal and Mukherjee (2025a,b)). As indicated above, studying general Potts models requires the development of new analytical tools tailored to Potts models, which we expect will also be useful for future investigations.

A natural motivation for our work arises from the extensive literature on *exponential random graph models* (ERGMs), which can be thought of as analogues of the Ising model with higher dimensional tensors. Sampling from ERGMs plays a central role in both parameter estimation and hypothesis testing, and Glauber dynamics provide a standard and widely used approach for this purpose. The mixing properties of Glauber dynamics in ERGMs have been studied in several key works, including Bhamidi, Bresler and Sly (2011) and DeMuse, Easlick and Yin (2019), which demonstrate interesting phase transition properties in the mixing rate. In fact, even in the specialized Curie–Weiss Ising model, mixing rates can be either polynomial or exponential depending on the parameter regime. For details, we refer the interested reader to Levin, Luczak and Peres (2010); Ding, Lubetzky and Peres (2009); Samanta, Mukherjee and Zhang (2024) and references therein. For the Curie–Weiss Potts model, He and Lok (2025) have studied mixing rate for Glauber dynamics, whereas Eichelsbacher and Martschink (2015), Ellis and Wang (1990) and Gandolfo, Ruiz and Marc (2010) study CLT for the magnetization. On the inferential side, the problem of parameter estimation in ERGMs has also received significant attention which demonstrates challenges of their own; see, for instance, Chatterjee and Diaconis (2013); Mukherjee and Xu (2023); Stivala, Robins and Lomi (2020). Similar to the Potts case, the most well-studied tensor for higher order binary models is the complete tensor case (p-spin Curie Weiss model), which has been studied recently in Mukherjee, Son and Bhattacharya (2022, 2021, 2025); Mukherjee et al. (2024) using the perfect symmetry of the complete tensor.

Going in a different direction, another question of interest is the problem of structure learning, i.e. to recover the whole graph/matrix \mathbf{A}_N , which is a high-dimensional parameter estimation problem. Indeed, in this case one Ising/Potts sample will not suffice, and one needs access to i.i.d. samples. In this setting, Anandkumar et al. (2012); Ravikumar, Wainwright and Lafferty (2010); Bresler (2015); Lohov et al. (2018); Vuffray et al. (2016) study graph recovery and support recovery, and establish tight sample complexity bounds, for Ising models. Other questions of interest for Ising-type models include community detection on SBM (Berthet, Rigollet and Srivastava, 2019), property testing (Neykov and

Liu, 2019), and structure detection (Cao, Neykov and Liu, 2022).

1.2. Our contributions

In this paper we study bivariate estimation of parameters in a Potts model with q colors, using the pseudo-likelihood method of Besag (1974, 1975). Prior to our work, the existing literature focuses exclusively on the Ising case ($q = 2$), or on the Curie–Weiss Potts case. Going from the Ising to general Potts case requires us to investigate conditions under which the pseudo-likelihood estimator exists (see (48)). The exact characterization is delicate for $q > 2$ colors, more so because the characterization for $q = 2$ in (Ghosal et al., 2020, Theorem 1.2 (a)) is not entirely correct. The correct characterization in the Ising case was established recently in Chen, Sen and Wu (2024), and in this work we establish the corresponding result for the Potts model. In particular, we require that there exist two colors for which the corresponding local fields are well separated (see Theorem 1.1 for details). Another challenge is the characterization of the subset of the parameter space for the Curie–Weiss Potts model where the local magnetization vector has \sqrt{N} fluctuation, in terms of the Hessian of the variational objective $H_{\beta, \mathbf{B}}(\cdot)$ (see (18)). This is carried out in Lemma H.4, utilizing tools from linear algebra, coupled with a careful application of the inverse function theorem. This lemma is crucially used to show non-existence of consistent estimators for Potts models on dense Erdős–Rényi graphs. Showing that the estimation is possible at the optimal rate of \sqrt{N} in the irregular case (Theorem 1.4) is more delicate for the Potts case with $q > 2$ colors. A fine analysis is needed to show that the RHS of (53) is strictly positive, which translates into a variation bound for the gradient of the free energy function ψ_N (see (48)) from its average. But perhaps most significantly, utilizing the non-linear large deviations framework developed in Chatterjee and Dembo (2016) and Basak and Mukherjee (2017), in this paper we develop a concentration result for mean-field Potts models (see Lemma I.6). This result shows that the local fields for all colors are close to the optimizers of the variational problem resulting from the non-linear large deviations. This is of possible independent interest, particularly if one wants to go beyond the law of large numbers, and study a CLT under Potts models.

1.3. Main Results

In this section, we state the main results of this paper. As mentioned above, our main goal is to derive a consistent estimator of the parameter (β, \mathbf{B}) , when a single vector \mathbf{X} is observed from the model (1). The classical method of maximum likelihood (ML) estimation is not practical in this framework, because of the presence of the intractable normalizing constant $Z_N(\beta, \mathbf{B})$, which is hard to compute and difficult to approximate using MCMC techniques; see Bhamidi, Bresler and Sly (2011). A computationally efficient alternative in the literature Besag (1974, 1975); Chatterjee (2007a); Bhattacharya and Mukherjee (2018);

Ghosal et al. (2020); Daskalakis, Dikkala and Panageas (2020) is to consider the maximum pseudo-likelihood (MPL) estimator, given by:

$$(\hat{\beta}_N, \hat{\mathbf{B}}_N) := \operatorname{argmax}_{(\beta, \mathbf{B}) \in \mathbb{R}^q} L_N(\beta, \mathbf{B}) := \operatorname{argmax}_{(\beta, \mathbf{B}) \in \mathbb{R}^q} \prod_{i=1}^N \mathbb{P}_{\beta, \mathbf{B}}(X_i | (X_j)_{j \neq i})$$

provided the pseudo-likelihood function L_N has a unique maximizer. Indeed, the conditional distribution of X_i given $(X_j)_{j \neq i}$ is easy to compute, and is given by:

$$\mathbb{P}_{\beta, \mathbf{B}}(X_i = r | (X_j)_{j \neq i}) = \frac{\exp\{\beta m_{i,r}(\mathbf{X}) + B_r\}}{\sum_{s=1}^q \exp\{\beta m_{i,s}(\mathbf{X}) + B_s\}} =: \theta_{i,r}(\mathbf{X}) \quad (4)$$

where $m_{i,r}(\mathbf{x}) := \sum_{j=1}^N a_{ij} \mathbb{1}_{x_j=r}$ for $\mathbf{x} \in [q]^N$. We will often drop \mathbf{X} from the notation $\theta_{i,r}$ for simplicity. The pseudo-likelihood function L_N is thus given by:

$$L_N(\beta, \mathbf{B}) := \frac{\exp\left\{\beta \sum_{i=1}^N \sum_{r=1}^q m_{i,r}(\mathbf{X}) \mathbb{1}_{X_i=r} + \sum_{i=1}^N \sum_{r=1}^q B_r \mathbb{1}_{X_i=r}\right\}}{\prod_{i=1}^N \sum_{r=1}^q \exp\{\beta m_{i,r}(\mathbf{X}) + B_r\}}$$

and hence, the log pseudo-likelihood function is given by:

$$\begin{aligned} \ell_N(\beta, \mathbf{B}) := & \beta \sum_{i=1}^N \sum_{r=1}^q m_{i,r}(\mathbf{X}) \mathbb{1}_{X_i=r} + \sum_{i=1}^N \sum_{r=1}^q B_r \mathbb{1}_{X_i=r} - \sum_{i=1}^N \log \left(\sum_{r=1}^q \exp\{\beta m_{i,r}(\mathbf{X}) + B_r\} \right). \end{aligned} \quad (5)$$

The MPL estimator can be obtained by setting the partial derivatives of ℓ_N to 0, which in turn requires the exact expressions of these partial derivatives. Towards this, we have:

$$\frac{\partial \ell_N(\beta, \mathbf{B})}{\partial \beta} = \sum_{i=1}^N \sum_{r=1}^q m_{i,r}(\mathbf{X}) \mathbb{1}_{X_i=r} - \sum_{i=1}^N \frac{\sum_{r=1}^q m_{i,r}(\mathbf{X}) \exp\{\beta m_{i,r}(\mathbf{X}) + B_r\}}{\sum_{r=1}^q \exp\{\beta m_{i,r}(\mathbf{X}) + B_r\}}, \quad (6)$$

$$\frac{\partial \ell_N(\beta, \mathbf{B})}{\partial B_s} = \sum_{i=1}^N \mathbb{1}_{X_i=s} - \sum_{i=1}^N \frac{\exp\{\beta m_{i,s}(\mathbf{X}) + B_s\}}{\sum_{r=1}^q \exp\{\beta m_{i,r}(\mathbf{X}) + B_r\}} \quad (1 \leq s \leq q-1). \quad (7)$$

Henceforth, we will call the equation $\nabla L_N(\beta, \mathbf{B}) = \mathbf{0}$, the pseudo-likelihood equation. Of course, if the MPL estimator $(\hat{\beta}_N, \hat{\mathbf{B}}_N)$ exists, then it is a solution of the pseudo-likelihood equation.

Before stating our first main result about the behavior of the MPL estimator, we introduce two assumptions on the coupling matrix \mathbf{A}_N that we will assume throughout the rest of the paper:

$$\sup_{N \geq 1} \|\mathbf{A}_N\|_1 = \sup_{N \geq 1} \max_{i \in [N]} \sum_{j=1}^N a_{ij} =: \gamma < \infty, \quad (8)$$

$$\liminf_{N \rightarrow \infty} \frac{\mathbf{1}' \mathbf{A}_N \mathbf{1}}{N} = \liminf_{N \rightarrow \infty} \frac{1}{N} \sum_{1 \leq i, j \leq N} a_{ij} > 0. \quad (9)$$

Here $\|\cdot\|_1$ denotes the ℓ_1 operator norm of a matrix, and $\mathbf{1}$ is the constant vector of size N with all entries 1. These conditions are standard in the literature for inference in Ising models, which corresponds to the case $q = 2$ (see Eq. (1.2) and (1.3) in Ghosal et al. (2020); also see Deb et al. (2024); Mukherjee, Mukherjee and Yuan (2018)). Note that when \mathbf{A}_N is the scaled adjacency of a graph (2), condition (8) becomes equivalent to the maximum degree $d_{\max}(G_N)$ of G_N being of the same order as its average degree $\bar{d}(G_N) := \frac{1}{N} \sum_{i=1}^N d_i(G_N)$, where $d_i(G_N)$ denotes the degree of the vertex i in G_N (i.e. $d_{\max}(G_N) = O(\bar{d}(G_N))$). What this essentially says, is that there is no vertex in the graph with *atypically high* degree. Condition (9) is always true in this case, and in fact, one has

$$\frac{1}{N} \sum_{1 \leq i, j \leq N} a_{ij} = 1.$$

We are now ready to state the first main result of this paper, which gives an upper bound to the estimation error in terms of the quantity $T_N(\mathbf{x})$ defined as:

$$T_N(\mathbf{x}) := \sum_{1 \leq r < s \leq q} \left(\frac{1}{N} \sum_{i=1}^N (m_{i,r}(\mathbf{x}) - m_{i,s}(\mathbf{x}))^2 - (\bar{m}_r(\mathbf{x}) - \bar{m}_s(\mathbf{x}))^2 \right). \quad (10)$$

where $\bar{m}_r(\mathbf{x}) := N^{-1} \sum_{i=1}^N m_{i,r}(\mathbf{x})$.

Theorem 1.1. *Suppose \mathbf{X} is a sample from the Potts model (1), where the coupling matrix \mathbf{A}_N has non-negative entries, and satisfies conditions (8) and (9). If $(\beta, \mathbf{B}) \in \Theta := (0, \infty) \times \mathbb{R}^{q-1}$, then the following conclusions hold:*

(a) *The MPL estimator $(\hat{\beta}_N, \hat{\mathbf{B}}_N)$ exists if $\mathbf{X} \in \Omega_N \cap \Lambda_N$, where*

$$\begin{aligned} \Lambda_N &:= \{\mathbf{y} \in [q]^N : \text{for every } r \in [q] \text{ there exists } i \in [N], \text{ such that } y_i = r\}, \\ \Omega_N &:= \{\mathbf{y} \in [q]^N : \text{there exist } 1 \leq r < s \leq q \text{ and } 1 \leq i, j, k, \ell \text{ all distinct, such that} \\ &\quad \{y_i, y_j\} = \{r, s\} = \{y_k, y_\ell\}, \{\tilde{m}_i^{r,s}(\mathbf{y}), \tilde{m}_j^{r,s}(\mathbf{y})\} < \{\tilde{m}_k^{r,s}(\mathbf{y}), \tilde{m}_\ell^{r,s}(\mathbf{y})\}\} \end{aligned}$$

$$\text{where } \tilde{m}_u^{r,s}(\mathbf{y}) := m_{u,r}(\mathbf{y}) - m_{u,s}(\mathbf{y}).$$

(b) *If $T_N(\mathbf{X})^{-1} = o_{\mathbb{P}}(\sqrt{N})$ and the MPL estimator exists, then*

$$\|(\hat{\beta}_N - \beta, \hat{\mathbf{B}}_N - \mathbf{B})\|_2 = O_{\mathbb{P}} \left(\frac{1}{\sqrt{N} T_N(\mathbf{X})} \right).$$

(c) *In particular, if $T_N(\mathbf{X})^{-1} = O_{\mathbb{P}}(1)$, then*

$$\mathbb{P}_{\beta, \mathbf{B}} \left(\mathbf{X} \in \Omega_N \cap \Lambda_N \right) \rightarrow 1 \text{ as } N \rightarrow \infty. \quad (11)$$

Consequently, the MPL estimator $(\hat{\beta}_N, \hat{\mathbf{B}}_N)$ exists with probability tending to 1, and satisfies

$$\|(\hat{\beta}_N - \beta, \hat{\mathbf{B}}_N - \mathbf{B})\|_2 = O_{\mathbb{P}}\left(\frac{1}{\sqrt{N}}\right). \quad (12)$$

Remark 1.2. Note that we are able to prove the existence of the joint MPL estimator $(\hat{\beta}_N, \hat{\mathbf{B}}_N)$ (with high probability) only in the regime $T_N(\mathbf{X})^{-1} = O_{\mathbb{P}}(1)$, but not in the entire regime $T_N(\mathbf{X})^{-1} = o_{\mathbb{P}}(\sqrt{N})$. This suffices to guarantee the \sqrt{N} -consistency of the MPL estimator whenever $T_N(\mathbf{X})^{-1} = O_{\mathbb{P}}(1)$ (part (b) of Theorem 1.1). In particular, this setting covers the cases where \mathbf{A}_N is the adjacency matrix of a sequence of bounded-degree graphs (see Section 1.3.1) or asymptotically irregular graphs (see Section 1.3.2). Part (c) extends the result to the full regime $T_N(\mathbf{X})^{-1} = o_{\mathbb{P}}(\sqrt{N})$, though only under the additional assumption that the MPL estimator exists in this regime.

The proof of Theorem 1.1 is given in Section 2. We now study the two most general types of interaction structures to which the joint consistency result, Theorem 1.1, applies. In fact, in both these cases, $T_N(\mathbf{X})^{-1} = O_{\mathbb{P}}(1)$, and hence, the joint MPL estimator is \sqrt{N} -consistent.

1.3.1. Non mean-field interactions

Throughout this subsection, we will assume that:

$$\liminf_{N \rightarrow \infty} \frac{1}{N} \sum_{1 \leq i, j \leq N} a_{ij}^2 > 0. \quad (13)$$

Condition (13) is often referred to as the *non mean-field condition*. Note that if the coupling matrix is the scaled adjacency of a graph, then condition (13) simply means that the average degree of the graph is bounded. This, coupled with condition (8) implies that the maximum degree of the graph is bounded. The following theorem shows that the joint MPL estimator is \sqrt{N} -consistent for the Potts model with interaction matrix \mathbf{A}_N satisfying (13).

Theorem 1.3. Suppose \mathbf{X} is an observation from the Potts model (1) where the interaction matrix \mathbf{A}_N satisfies the conditions (8), (9) and (13). Then,

$$\|(\hat{\beta}_N - \beta, \hat{\mathbf{B}}_N - \mathbf{B})\|_2 = O_{\mathbb{P}}\left(\frac{1}{\sqrt{N}}\right).$$

The proof of Theorem 1.3 is given in Section A of the appendix. As mentioned above, if the underlying interaction structure is the adjacency matrix of a deterministic graph scaled appropriately (2), then Theorem 1.3 applies as long as the graph is of bounded degree and $\liminf_{N \rightarrow \infty} \bar{d}(G_N) > 0$. This covers as special cases, the classical Ising models on lattices, that have finite-range interactions, and d -regular graphs with d fixed.

1.3.2. Irregular interactions

Throughout this subsection, we will assume that:

$$\liminf_{N \rightarrow \infty} \frac{1}{N} \sum_{i=1}^N (R_i - \bar{R})^2 > 0 \quad (14)$$

where $R_i := \sum_{j=1}^N a_{ij}$ and $\bar{R} := \frac{1}{N} \sum_{i=1}^N R_i$. Note that if the coupling matrix is the scaled adjacency of a graph (2), then Condition (14) says that the graph is asymptotically irregular. The following theorem shows that the joint MPL estimator is \sqrt{N} -consistent for the Potts model with interaction matrix \mathbf{A}_N satisfying (14).

Theorem 1.4. *Suppose that $\mathbf{B} \neq \mathbf{0}$, and \mathbf{X} is an observation from the Potts model (1) where the interaction matrix \mathbf{A}_N satisfies the conditions (8), (9) and (14). Then,*

$$\|(\hat{\beta}_N - \beta, \hat{\mathbf{B}}_N - \mathbf{B})\|_2 = O_{\mathbb{P}}\left(\frac{1}{\sqrt{N}}\right).$$

The proof of Theorem 1.4 is given in Section B of the appendix. Common examples of interaction structures satisfying all the necessary assumptions of Theorem 1.4 are the scaled adjacencies of the complete bipartite graph $K_{m,n}$ and a disjoint union of the cliques K_m and K_n where $N = m + n$ and $\frac{m}{N} \rightarrow \alpha \in (0, 1) \setminus \{\frac{1}{2}\}$ as $N \rightarrow \infty$. In general, it follows from the theory of graphons (see Lovász (2012) for a survey on graph limit theory and the literature of graphons) that if G_N is a sequence of dense graphs converging to a graphon W such that the function $x \mapsto \int_0^1 W(x, y) dy$ is not constant Lebesgue almost everywhere, then all the necessary assumptions of Theorem 1.4 are satisfied. This includes the above two examples as special cases, as well as dense stochastic block models on N nodes with two communities C_1 and C_2 of sizes m and n respectively, where $m/N \rightarrow \alpha$, between-group connection probability q and within-group connection probabilities p_1 (within community C_1) and p_2 (within community C_2), satisfying:

$$\alpha(p_1 - q) \neq (1 - \alpha)(p_2 - q). \quad (15)$$

In this case, the limiting graphon is given by:

$$W(x, y) = \begin{cases} p_1, & (x, y) \in [0, \alpha] \times [0, \alpha], \\ p_2, & (x, y) \in (\alpha, 1] \times (\alpha, 1], \\ q, & (x, y) \in [0, \alpha] \times (\alpha, 1] \cup (\alpha, 1] \times [0, \alpha]. \end{cases}$$

If the block sizes are asymptotically equal (i.e. $\alpha = 1/2$), then condition (15) reduces to unequal within-group connection probabilities (i.e. $p_1 \neq p_2$). On the other hand, if the within-group connection probabilities are the same (i.e. $p_1 = p_2$), and unequal to the between-group connection probability q , then

condition (15) amounts to asymptotically unequal block sizes (i.e. $\alpha \neq 1/2$). See Section 1.2 in Ghosal et al. (2020) for a detailed discussion on such examples.

Having shown that joint consistent estimation at rate $N^{-1/2}$ is possible for non mean-field and irregular interactions, we now go to the opposite extreme, where consistent joint estimation is impossible. This happens in the Curie-Weiss Potts model where the coupling matrix is the adjacency of the complete graph (scaled by N) (see (40)) and more generally, in the Erdős-Rényi Potts model, where the coupling matrix is given by:

$$a_{ij} := \frac{g_{ij}}{Np} \quad (16)$$

with $G := ((g_{ij}))_{1 \leq i, j \leq N}$ being the adjacency of an Erdős-Rényi random graph $\mathcal{G}(N, p)$ with $p > 0$ fixed. Note that in the latter model, the coupling matrix is random, so we will consider the problem of estimation under the joint law $\mathbb{P}_{\beta, \mathbf{B}}^{\text{ER}}$ of \mathbf{X} and G on $[q]^N \times \{0, 1\}^{\binom{N}{2}}$. Throughout the rest of the paper, we will use the notation $\mathcal{P}([q])$ to denote the set of all probability measures on $[q]$, i.e.

$$\mathcal{P}([q]) := \left\{ \mathbf{v} \in [0, 1]^q : \sum_{r=1}^q v_r = 1 \right\}. \quad (17)$$

Theorem 1.5. *For each $\mathbf{m} \in \mathcal{P}([q])$, let $\Theta_{\mathbf{m}}$ be the set of all $(\beta, \mathbf{B}) \in (0, \infty) \times \mathbb{R}^{q-1}$ such that the function*

$$H_{\beta, \mathbf{B}}(\mathbf{t}) := \frac{\beta}{2} \sum_{r=1}^q t_r^2 + \sum_{r=1}^q B_r t_r - \sum_{r=1}^q t_r \log t_r \quad (18)$$

has the unique global maximizer \mathbf{m} on the set $\mathcal{P}([q])$, and

$$\mathbf{u}^\top \nabla^2 H_{\beta, \mathbf{B}}(\mathbf{m}) \mathbf{u} < 0 \text{ for all } \mathbf{u} \in T := \left\{ \mathbf{u} \in \mathbb{R}^q \setminus \{\mathbf{0}\} : \sum_{r=1}^q u_r = 0 \right\}.$$

Then the product measure $\nu := \mathbf{m}^N \times \mathcal{G}(N, p)$ is contiguous to the measure $\mathbb{P}_{\beta, \mathbf{B}}^{\text{ER}}$ for every $(\beta, \mathbf{B}) \in \Theta_{\mathbf{m}}$. Consequently, whenever $|\Theta_{\mathbf{m}}| \geq 2$, under $\mathbb{P}_{\beta, \mathbf{B}}^{\text{ER}}$ there does not exist any sequence of estimators (functions of (\mathbf{X}, G)) which is consistent for (β, \mathbf{B}) in $\Theta_{\mathbf{m}}$.

The proof of Theorem 1.5 is based on a contiguity argument that is presented in Section C of the appendix.

Remark 1.6. *The ambient Hessian of the function $H_{\beta, \mathbf{B}}$ is given by:*

$$\nabla^2 H_{\beta, \mathbf{B}}(\mathbf{t}) := \text{diag}((\beta - t_r^{-1})_{1 \leq r \leq q}).$$

This implies that for $\beta \leq 1$, $H_{\beta, \mathbf{B}}$ is negative definite for all \mathbf{t} in $(0, 1)^q$, and hence, the function $H_{\beta, \mathbf{B}}$ is strictly concave in this case. Therefore, for $\beta \leq 1$, any stationary point of $H_{\beta, \mathbf{B}}$ must be its unique maximizer. By a Lagrangian

argument (see the proof of Lemma H.4), an interior stationary point $\mathbf{m} \in \mathcal{P}([q])$ of $H_{\beta, \mathbf{B}}$ is characterized by the system of equations:

$$\beta(m_r - m_q) + B_r = \log \frac{m_r}{m_q} \quad \text{for } r \in [q-1].$$

Therefore, for any $\mathbf{m} \in \mathcal{P}([q])$, we have:

$$\left\{ \left(\beta, \log \frac{m_1}{m_q} + \beta(m_q - m_1), \dots, \log \frac{m_{q-1}}{m_q} + \beta(m_q - m_{q-1}) \right) : 0 < \beta \leq 1 \right\} \subseteq \Theta_{\mathbf{m}} \quad (19)$$

and note that the LHS of (19) is a non-empty affine (straight) line segment in \mathbb{R}^q , containing a continuum of points.

Finally, we establish consistency results for the partial MPL estimators of β and \mathbf{B} , treating each parameter as known while estimating the other. This setting is comparatively simpler and requires weaker assumptions than those needed for joint consistent estimation. For every fixed \mathbf{B} , the partial MPL estimator of β is defined as the unique maximizer of the function $\beta \rightarrow \ell_N(\beta, \mathbf{B})$, and for every fixed β , the partial MPL estimator of \mathbf{B} is defined as the unique maximizer of the function $\mathbf{B} \rightarrow \ell_N(\beta, \mathbf{B})$, if they exist. The following theorem gives consistency rates of the partial MPL estimator of β in terms of the quantity

$$U_N(\mathbf{x}) = \frac{1}{N} \sum_{1 \leq r < s \leq q} \sum_{i=1}^N (m_{i,r}(\mathbf{x}) - m_{i,s}(\mathbf{x}))^2, \quad (20)$$

and also that of \mathbf{B} .

Theorem 1.7. *Suppose \mathbf{X} is a sample from the Potts model (1), where the coupling matrix \mathbf{A}_N satisfies conditions (8) and (9), and $(\beta, \mathbf{B}) \in \Theta := (0, \infty) \times \mathbb{R}^{q-1}$. Then, the following are true:*

- (a) *The partial MPL estimator $\hat{\mathbf{B}}_N$ exists with probability $1 - o(1)$ for all $(\beta, \mathbf{B}) \in (0, \infty) \times \mathbb{R}^{q-1}$, and the partial MPL estimator $\hat{\beta}_N$ exists with probability $1 - o(1)$ whenever*

$$U_N(\mathbf{X})^{-1} = o_{\mathbb{P}}(\sqrt{N}). \quad (21)$$

- (b) *The partial MPL estimator $\hat{\mathbf{B}}_N$ satisfies:*

$$\|\hat{\mathbf{B}}_N - \mathbf{B}\|_2 = O_{\mathbb{P}}\left(\frac{1}{\sqrt{N}}\right).$$

- (c) *If $U_N(\mathbf{X})^{-1} = o_{\mathbb{P}}(\sqrt{N})$, then the partial MPL estimator $\hat{\beta}_N$ satisfies:*

$$|\hat{\beta}_N - \beta| = O_{\mathbb{P}}\left(\frac{1}{\sqrt{N}U_N(\mathbf{X})}\right).$$

(d) Define

$$\beta_c := \begin{cases} q & \text{if } q \leq 2 \\ \frac{2(q-1)}{q-2} \log(q-1) & \text{otherwise.} \end{cases}$$

If $\mathbf{B} \neq \mathbf{0}$, or if $\mathbf{B} = \mathbf{0}$ and $\beta \liminf_{N \rightarrow \infty} \frac{\mathbf{1}' \mathbf{A}_N \mathbf{1}}{N} > \beta_c$, then $U_N(\mathbf{X})^{-1} = O_{\mathbb{P}}(1)$. In these cases, $\hat{\beta}_N$ is \sqrt{N} -consistent for β .

(e) Finally, for the model $\mathbb{P}_{\beta, \mathbf{B}}^{\text{ER}}$ described above (see (16)), no consistent sequence of estimators exists for $\beta < \beta_c$ (here $\lim_{N \rightarrow \infty} \frac{\mathbf{1}' \mathbf{A}_N \mathbf{1}}{N} = 1$ almost surely) when $\mathbf{B} = \mathbf{0}$.

The proof of Theorem 1.7 is given in Section D of the appendix.

1.4. Future directions

As a possible future direction, a first question is to relax the assumption of non-negativity of the coupling matrix \mathbf{A}_N , and extend our results to the case of spin glass Potts models (similar to what was done for the special case of $q = 2$ in Chen, Sen and Wu (2024)). Another interesting question is to go beyond concentration results, and develop central limit theorems for the magnetization vector, similar to what was done in Deb and Mukherjee (2023) for $q = 2$. This will ultimately lead to the construction of asymptotically valid confidence intervals, a very useful inferential task. Possibly a more challenging direction is to go beyond quadratic interaction models, and study general Gibbs measures with higher order tensors, such as cubics, quartics, and so on. The exponential random graph models (ERGMs) fall under this class of higher order tensor models, and have proved notoriously hard for inference purposes. In particular, the phenomenon of “degeneracy” for ERGMs has a body of growing literature, both in empirical and rigorous work (see Snijders et al. (2006); Handcock et al. (2003); Chatterjee and Diaconis (2013); Mukherjee (2020) and references therein). Other related and more general models in which one may seek to establish analogous results on the joint consistency of parameter estimators include the XY model (Kenna, 2005), the Ashkin–Teller model (Ashkin and Teller, 1943; Aoun, Dober and Glazman, 2024), and the $O(N)$ model (Kirkpatrick and Nawaz, 2016).

1.5. Outline of the paper

The rest of the paper is organized as follows. The proof of our main result (Theorem 1.1) is given in Section 2. In Section 3 we illustrate our theoretical results with a simulation study. In Appendices A and B, we prove Theorems 1.3 and 1.4, respectively. Appendices C and D are dedicated to the proofs of the remaining main results of the paper, namely Theorems 1.5 and 1.7, respectively. In Appendix E, we prove a general result on convergence of Z -estimators, whereas in Appendix F, we develop necessary tools for bounding the derivatives of the log pseudolikelihood, both of which are crucial in establishing consistency and rates

of convergence for our MPL estimators (Theorems 1.1 and 1.7). In Appendix G we prove Lemma A.1, which is a crucial step towards proving Theorem 1.3. In Appendix H, we prove some results necessary for verifying Theorem 1.5. Finally, Appendix I contains additional technical lemmas necessary for proving some of the main results of the paper.

2. Proof of Theorem 1.1

This section is dedicated to proving the main result of this paper (Theorem 1.1). Further technical lemmas necessary for proving these results are given in the appendix.

(a) Suppose that $\mathbf{X} \in \Omega_N \cap \Lambda_N$. Then, there exist $1 \leq a \neq b \leq q$ and $1 \leq i, j, k, l \leq N$ all distinct, such that $\{X_i, X_j\} = \{a, b\}$, $\{X_k, X_l\} = \{a, b\}$ and $\{\tilde{m}_i^{a,b}(\mathbf{X}), \tilde{m}_j^{a,b}(\mathbf{X})\} < \{\tilde{m}_k^{a,b}(\mathbf{X}), \tilde{m}_l^{a,b}(\mathbf{X})\}$ (recall that $\tilde{m}_u^{r,s}(\mathbf{X}) := m_{u,r}(\mathbf{X}) - m_{u,s}(\mathbf{X})$). Since the function ℓ_N is concave (see Lemma F.2), in order to show the existence of the MPL estimator, it suffices to show that:

$$\lim_{\|(\beta, \mathbf{B})\|_\infty \rightarrow \infty} \ell_N(\beta, \mathbf{B}) = -\infty.$$

Without loss of generality, assume that $X_i = X_k = a$ and $X_j = X_l = b$. Since

$$\begin{aligned} & h_w(\beta, \mathbf{B}) \\ := & \beta \sum_{r=1}^q m_{w,r}(\mathbf{X}) \mathbb{1}_{X_w=r} + \sum_{r=1}^q B_r \mathbb{1}_{X_w=r} - \log \left(\sum_{r=1}^q \exp \{ \beta m_{w,r}(\mathbf{X}) + B_r \} \right) \leq 0 \end{aligned}$$

for all $1 \leq w \leq N$, it suffices to show that at least one of $h_i(\beta, \mathbf{B})$, $h_j(\beta, \mathbf{B})$, $h_k(\beta, \mathbf{B})$ and $h_l(\beta, \mathbf{B})$ goes to $-\infty$ as $\|(\beta, \mathbf{B})\|_\infty \rightarrow \infty$. Now, note that:

$$\exp(h_w(\beta, \mathbf{B})) = \frac{1}{1 + \sum_{r \neq X_w} \exp(\beta(m_{w,r}(\mathbf{X}) - m_{w,X_w}(\mathbf{X})) + (B_r - B_{X_w}))}, \quad (22)$$

Setting $w = i, j, k, l$ in (22), it suffices to show that at least one of the following four quantities:

1. $\beta(m_{i,r}(\mathbf{X}) - m_{i,a}(\mathbf{X})) + (B_r - B_a)$,
2. $\beta(m_{k,r}(\mathbf{X}) - m_{k,a}(\mathbf{X})) + (B_r - B_a)$,
3. $\beta(m_{j,r}(\mathbf{X}) - m_{j,b}(\mathbf{X})) + (B_r - B_b)$, and
4. $\beta(m_{l,r}(\mathbf{X}) - m_{l,b}(\mathbf{X})) + (B_r - B_b)$

goes to $+\infty$ as $\|(\beta, \mathbf{B})\|_\infty \rightarrow \infty$, for at least one $r \in [q]$. So, let us assume that none of them goes to $+\infty$, i.e. there exists a constant $K \in (0, \infty)$, such that all of (1), (2), (3) and (4) are bounded above by K along the sequence (β, \mathbf{B}) whose norm goes to $+\infty$, for all $r \in [q]$. In this case, putting $r = b$ in (1) and $r = a$ in (4), we get:

$$-\beta \tilde{m}_i^{a,b}(\mathbf{X}) - (B_a - B_b) \leq K \quad \text{and} \quad \beta \tilde{m}_l^{a,b}(\mathbf{X}) + (B_a - B_b) \leq K$$

$$\implies \beta \leq \frac{2K}{\tilde{m}_\ell^{a,b}(\mathbf{X}) - \tilde{m}_i^{a,b}(\mathbf{X})}.$$

Similarly, putting $r = b$ in (2) and $r = a$ in (3), we get:

$$\begin{aligned} -\beta \tilde{m}_k^{a,b}(\mathbf{X}) - (B_a - B_b) &\leq K \quad \text{and} \quad \beta \tilde{m}_j^{a,b}(\mathbf{X}) + (B_a - B_b) \leq K \\ \implies \beta &\geq -\frac{2K}{\tilde{m}_k^{a,b}(\mathbf{X}) - \tilde{m}_j^{a,b}(\mathbf{X})}. \end{aligned}$$

Thus,

$$|\beta| \leq K_0 := \max \left\{ \frac{2K}{\tilde{m}_\ell^{a,b}(\mathbf{X}) - \tilde{m}_i^{a,b}(\mathbf{X})}, \frac{2K}{\tilde{m}_k^{a,b}(\mathbf{X}) - \tilde{m}_j^{a,b}(\mathbf{X})} \right\}.$$

Now, choose any two distinct colors $s, t \in [q]$. Since $\mathbf{X} \in \Lambda_N$, we can choose $1 \leq u \neq v \leq N$ such that $X_u = s$ and $X_v = t$. If either $h_u(\beta, \mathbf{B})$ or $h_v(\beta, \mathbf{B}) \rightarrow -\infty$, then once again we are done. So, assume otherwise, i.e. both are bounded below by some constant, which implies that there exists $\varepsilon > 0$ such that:

$$\min \{ \exp(h_u(\beta, \mathbf{B})), \exp(h_v(\beta, \mathbf{B})) \} > \varepsilon,$$

which, in view of (22), implies that:

$$\beta \tilde{m}_u^{r,s}(\mathbf{X}) + B_r - B_s < -\log \varepsilon \quad \text{for all } r \neq s \quad (23)$$

and

$$\beta \tilde{m}_v^{r,t}(\mathbf{X}) + B_r - B_t < -\log \varepsilon \quad \text{for all } r \neq t. \quad (24)$$

Now, putting $r = t$ in (23) and $r = s$ in (24), we get:

$$B_t - B_s < -\log \varepsilon - \beta \tilde{m}_u^{t,s}(\mathbf{X}) \quad \text{and} \quad B_s - B_t < -\log \varepsilon - \beta \tilde{m}_v^{s,t}(\mathbf{X}).$$

By Assumption 8, there exists $\gamma \in (0, \infty)$ such that $\sup_{N \geq 1, i \in [N]} \sum_{j=1}^N a_{ij} < \gamma$, which implies that $|\tilde{m}_u^{t,s}(\mathbf{X})| \leq 2\gamma$ and $|\tilde{m}_v^{s,t}(\mathbf{X})| \leq 2\gamma$. This now implies that $|B_t - B_s| \leq -\log \varepsilon + 2K_0\gamma$. Setting $s = q$, we thus have $|B_t| \leq -\log \varepsilon + 2K_0\gamma$, i.e. B_t is bounded. Hence, (β, \mathbf{B}) is a bounded sequence, a contradiction. This proves (a).

(b) It follows from Lemma F.1 (taking $\lambda = 0$) that $\|\nabla \ell_N(\beta, \mathbf{B})\|_2 = O_{\mathbb{P}}(\sqrt{N})$. Part (b) now follows from Proposition E.1 (on taking $w_N(\beta, \mathbf{B}) = \nabla \ell_N(\beta, \mathbf{B})$, $a_N = \sqrt{N}$ and $h_N(\mathbf{X}) = NT_N(\mathbf{X})$) and Lemma F.2. \square

(c) It suffices to check (11), as the other conclusions follow from parts (a) and (b). To this effect, define

$$E_N(\delta) := \{ \mathbf{x} \in [q]^N : T_N(\mathbf{x}) < \delta \}.$$

It follows from the tightness of $T_N(\mathbf{X})^{-1}$, that

$$\limsup_{\delta \rightarrow 0} \sup_{N \geq 1} \mathbb{P}(\mathbf{X} \in E_N(\delta)) = 0. \quad (25)$$

Suppose that $\mathbf{x} \in E_N(\delta)^c$ for some fixed $\delta > 0$. Fixing \mathbf{x} , for notational convenience, we will abbreviate $\tilde{m}_u^{r,s}(\mathbf{x})$ by $\tilde{m}_u^{r,s}$. Then, we have:

$$T_N(\mathbf{x}) = \frac{1}{2N^2} \sum_{r < s} \sum_{i,j} (\tilde{m}_i^{r,s} - \tilde{m}_j^{r,s})^2 \geq \delta,$$

Hence, there exist colors $a < b$, such that

$$\frac{1}{N^2} \sum_{i,j} (\tilde{m}_i^{a,b} - \tilde{m}_j^{a,b})^2 \geq \frac{4\delta}{q(q-1)}. \quad (26)$$

It follows from (8) that for any $r, s \in [q]$ and $i \in [N]$, $\tilde{m}_i^{r,s} \in [-\gamma, \gamma]$, where γ is defined in (8). Fixing a positive integer $R > \frac{9q\gamma}{4\sqrt{\delta}}$, and setting $t_p := \frac{p\gamma}{R}$ for $p \in \mathbb{Z}$, note that

$$[-\gamma, \gamma] = \bigcup_{-R \leq p \leq R-1} [t_p, t_{p+1}].$$

Let us define

$$S_p := \left\{ i \in [N] : \tilde{m}_i^{a,b} \geq \frac{p\gamma}{R} \right\}$$

and

$$p_0 = p_0(\varepsilon) := \max \{ p \in \mathbb{Z} \cap [-R, R+1] : |S_p| \geq \varepsilon N \} - 1.$$

We now claim that whenever $\varepsilon \in (0, \delta/(9q^2\gamma^2))$, $|S_{p_0+1}| \geq \varepsilon N$ and $|S_{p_0-1}^c| \geq \varepsilon N$.

Proof of Claim: Note that $|S_{-R}| = N$, $|S_{R+1}| = 0$ and $|S_p|$ is decreasing in p , so such a p_0 exists. Also, it follows directly from the definition of p_0 , that:

$$|S_{p_0+1}| \geq \varepsilon N \quad \text{and} \quad |S_{p_0+2}| < \varepsilon N.$$

We claim that $|S_{p_0-1}^c| \geq \varepsilon N$. If this is not true, then there must exist at least $(1 - 2\varepsilon)N$ many i 's for which $\tilde{m}_i^{a,b} \in I := [(p_0 - 1)\gamma/R, (p_0 + 2)\gamma/R)$. Hence, we have the following:

$$\begin{aligned} & \frac{1}{N^2} \sum_{i,j} (\tilde{m}_i^{a,b} - \tilde{m}_j^{a,b})^2 \\ & \leq \frac{2}{N^2} \sum_{i \in S_{p_0+2}, 1 \leq j \leq N} (\tilde{m}_i^{a,b} - \tilde{m}_j^{a,b})^2 + \frac{2}{N^2} \sum_{i \in S_{p_0-1}^c, 1 \leq j \leq N} (\tilde{m}_i^{a,b} - \tilde{m}_j^{a,b})^2 \\ & + \frac{1}{N^2} \sum_{i,j: \tilde{m}_i^{a,b}, \tilde{m}_j^{a,b} \in I} (\tilde{m}_i^{a,b} - \tilde{m}_j^{a,b})^2 \\ & \leq \frac{2|S_{p_0+2}|}{N} (4\gamma^2) + \frac{2|S_{p_0-1}^c|}{N} (4\gamma^2) + \frac{9\gamma^2}{R^2} \\ & \leq 16\varepsilon\gamma^2 + \frac{9\gamma^2}{R^2} < \frac{4\delta}{q(q-1)}. \end{aligned}$$

where the last inequality follows from the choice of ε and R above. But this contradicts (26), thus verifying the claim.

Using the claim, the colors a and b satisfy $\tilde{m}_i^{a,b} < (p_0 - 1)\gamma/R$ for at least εN many i and $\tilde{m}_i^{a,b} \geq p_0\gamma/R$ for at least εN many i . Let us now define the following four events:

$$\mathcal{E}_{1,a} := \{\mathbf{x} \in [q]^N : x_i \neq a \forall i \in S_{p_0}\}, \quad \mathcal{E}_{1,b} := \{\mathbf{x} \in [q]^N : x_i \neq b \forall i \in S_{p_0}\}$$

$$\mathcal{E}_{2,a} := \{\mathbf{x} \in [q]^N : x_i \neq a \forall i \in S_{p_0}^c\}, \quad \mathcal{E}_{2,b} := \{\mathbf{x} \in [q]^N : x_i \neq b \forall i \in S_{p_0}^c\}.$$

We claim that each of the above four sets, intersected with $E_N(\delta)^c$, has probability $o(1)$ of containing \mathbf{X} as $N \rightarrow \infty$.

Proof of Claim: Define a function $g : \mathbb{R} \mapsto [0, \infty)$ as:

$$g(x) := \begin{cases} 0 & \text{if } x \leq 0 \\ x^2 & \text{if } x \in (0, 0.5) \\ x - 0.25 & \text{if } x > 0.5. \end{cases}$$

Then it is straightforward to check that g is differentiable on \mathbb{R} with derivative bounded by 1. Also, g is non-decreasing, strictly positive on the positive axis, and bounded on compact intervals. Let us now define:

$$h_1(\mathbf{x}) := \max_{c \in F} \max_{r, s \in [q]} \sum_{i=1}^N (\mathbb{1}_{x_i \neq r} - \mathbb{P}(X_i \neq r | X_j = x_j, j \neq i)) g(\tilde{m}_i^{r,s}(\mathbf{x}) - c),$$

$$h_2(\mathbf{x}) := \max_{c \in F} \max_{r, s \in [q]} \sum_{i=1}^N (\mathbb{1}_{x_i \neq r} - \mathbb{P}(X_i \neq r | X_j = x_j, j \neq i)) g(c - \tilde{m}_i^{r,s}(\mathbf{x})),$$

where

$$F := \left\{ \frac{p\gamma}{R} : p \in \mathbb{Z} \cap [-R, R+1] \right\}.$$

Whenever $\mathbf{x} \in \mathcal{E}_{1,a} \cap E_N(\delta)^c$, we have:

$$\begin{aligned} h_1(\mathbf{x}) &\geq \sum_{i=1}^N (\mathbb{1}_{x_i \neq a} - \mathbb{P}(X_i \neq a | X_j = x_j, j \neq i)) g\left(\tilde{m}_i^{a,b}(\mathbf{x}) - \frac{p_0\gamma}{R}\right) \\ &\geq \alpha \sum_{i \in S_{p_0}} g\left(\tilde{m}_i^{a,b}(\mathbf{x}) - \frac{p_0\gamma}{R}\right) \geq \alpha \sum_{i \in S_{p_0+1}} g\left(\tilde{m}_i^{a,b}(\mathbf{x}) - \frac{p_0\gamma}{R}\right) \geq \alpha \varepsilon N g\left(\frac{\gamma}{R}\right), \end{aligned}$$

where

$$\alpha := q^{-1} \exp\{-\beta\gamma - 2\|\mathbf{B}\|_\infty\} > 0. \quad (27)$$

Here the first inequality in the second line uses the fact that $x_i \neq a$ for $i \in S_{p_0}$, by definition of $\mathcal{E}_{1,a}$, and the last inequality uses the fact that $|S_{p_0+1}| \geq \varepsilon N$. Applying Lemma F.1 with $b_{itrs} := \mathbb{1}_{t \neq r}$, $\lambda = 1$, $t = N[\alpha \varepsilon g(\frac{\gamma}{R})]^2$ and a union bound we get

$$\mathbb{P}(\mathbf{X} \in \mathcal{E}_{1,a} \cap E_N(\delta)^c) \leq \mathbb{P}\left(|h_1(\mathbf{X})| > \alpha \varepsilon N g\left(\frac{\gamma}{R}\right)\right)$$

$$\leq 2(2R+2)q^2 \exp\left(-CN\left[\alpha\varepsilon g\left(\frac{\gamma}{R}\right)\right]^2\right) \rightarrow 0.$$

One can similarly show that $\mathbb{P}(\mathbf{X} \in \mathcal{E}_{1,b} \cap E_N(\delta)^c) = o(1)$. Also, for $\mathbf{x} \in \mathcal{E}_{2,a} \cap E_N(\delta)^c$, a similar calculation gives

$$\begin{aligned} h_2(\mathbf{x}) &\geq \sum_{i=1}^N (\mathbb{1}_{x_i \neq a} - \mathbb{P}(X_i \neq a | X_j = x_j, j \neq i)) g\left(\frac{p_0\gamma}{R} - \tilde{m}_i^{a,b}(\mathbf{x})\right) \\ &\geq \alpha \sum_{i \in S_{p_0}^c} g\left(\frac{p_0\gamma}{R} - \tilde{m}_i^{a,b}(\mathbf{x})\right) \alpha \sum_{i \in S_{p_0-1}^c} g\left(\frac{p_0\gamma}{R} - \tilde{m}_i^{a,b}(\mathbf{x})\right) \geq \alpha\varepsilon N g\left(\frac{\gamma}{R}\right) \end{aligned} \quad (28)$$

Once again, by Lemma F.1, the probability that \mathbf{X} satisfies (28) is $o(1)$ as $N \rightarrow \infty$, which proves that $\mathbb{P}(\mathbf{X} \in \mathcal{E}_{2,a} \cap E_N(\delta)^c) = o(1)$. One can similarly show that $\mathbb{P}(\mathbf{X} \in \mathcal{E}_{2,b} \cap E_N(\delta)^c) = o(1)$, which completes the proof of the claim.

When \mathbf{x} belongs to $\mathcal{E}_{1,a}^c \cap \mathcal{E}_{1,b}^c \cap \mathcal{E}_{2,a}^c \cap \mathcal{E}_{2,b}^c$, there exist $i, j, k, \ell \in [N]$, such that: $x_i = a, x_j = b, x_k = a, x_\ell = b$, and

$$\max\{\tilde{m}_i^{a,b}(\mathbf{x}), \tilde{m}_j^{a,b}(\mathbf{x})\} < \frac{(p_0)\gamma}{R} \leq \min\{\tilde{m}_k^{a,b}(\mathbf{x}), \tilde{m}_\ell^{a,b}(\mathbf{x})\},$$

and so $\mathbf{x} \in \Omega_N$ (as introduced in the statement of Theorem 1.1 (a)). Hence, for all $\delta > 0$, we have:

$$\begin{aligned} \mathbb{P}(\mathbf{X} \in \Omega_N^c) &\leq \mathbb{P}(\mathbf{X} \in E_N(\delta)) + \mathbb{P}(\mathbf{X} \in \Omega_N^c \cap E_N(\delta)^c) \\ &\leq \mathbb{P}(\mathbf{X} \in E_N(\delta)) + \mathbb{P}\left(\mathbf{X} \in \left(\mathcal{E}_{1,a} \cup \mathcal{E}_{1,b} \cup \mathcal{E}_{2,a} \cup \mathcal{E}_{2,b}\right) \cap E_N(\delta)^c\right). \end{aligned}$$

The second term in the RHS converges to 0 on letting $N \rightarrow \infty$ as shown above, and the first term converges to 0 on taking a supremum over N followed by $N \rightarrow \infty$ using (25), thus showing that $\mathbb{P}(\mathbf{X} \in \Omega_N^c) = o(1)$.

Next, we show that $\mathbb{P}(\mathbf{X} \notin \Lambda_N) = o(1)$. For $\mathbf{x} \in \Lambda_N^c$, there exists $r \in [q]$ such that $x_i \neq r$ for all i . This implies that for all $i \in [N]$, we have

$$\sum_{s \neq r} \mathbb{1}_{x_i=s} = 1 \Rightarrow \sum_{s \neq r} (\mathbb{1}_{x_i=s} - \theta_{i,s}(\mathbf{x})) = 1 - \sum_{s \neq r} \theta_{i,s}(\mathbf{x}),$$

where $\theta_{i,s}(\mathbf{x}) = \mathbb{P}(X_i = s | X_j = x_j, j \neq i)$ is as in (4). Now, we have $1 - \sum_{s \neq r} \theta_{i,s}(\mathbf{x}) \geq \alpha$ (see (27) for the definition of α), which implies:

$$\sum_{s \neq r} \sum_{i=1}^N (\mathbb{1}_{x_i=s} - \theta_{i,s}(\mathbf{x})) \geq \frac{N}{q \exp(\beta\gamma + 2 \max_{r \in [q]} |B_r|)}.$$

Now, by Lemma F.1 (taking $b_{itrs} := \mathbb{1}_{t \neq r}$ and $g \equiv 1$), we have:

$$\sum_{s \neq r} \sum_{i=1}^N (\mathbb{1}_{X_i=s} - \theta_{i,s}(\mathbf{x})) = O_{\mathbb{P}}(\sqrt{N}).$$

Therefore, $\mathbb{P}(\mathbf{X} \in \Lambda_N^c) = o(1)$, completing the proof of (11). The proof of (12) will follow from part (b) above.

3. Numerical Study

To illustrate our theoretical results, we simulate observations from the Potts model on the Erdős-Rényi graph $\mathcal{G}(N, p)$ with $N = 100$ and 200 , and $p = 0.025$ and 0.25 . The $p = 0.025$ case illustrates the sparse regime where joint estimation is possible, and the $p = 0.25$ case illustrates the dense regime where joint estimation is not possible on the level-sets of the map $(\beta, \mathbf{B}) \mapsto \arg \max_{\mathbf{t}} H_{\beta, \mathbf{B}}(\mathbf{t})$ provided this maximizer is unique. We do this for $q = 3$ for the ease of representation of the numerical results through 3D graphs. One can easily derive using Lagrange multipliers, that for each such unique maximizer \mathbf{m} , the inestimability curve $\Theta_{\mathbf{m}}$ is a straight line that has equation:

$$B_1 = \log \left(\frac{m_1 \exp(\beta m_3)}{m_3 \exp(\beta m_1)} \right) \quad \text{and} \quad B_2 = \log \left(\frac{m_2 \exp(\beta m_3)}{m_3 \exp(\beta m_2)} \right)$$

We call this straight line the *line of inestimability*, which is plotted in blue in Figure 1. We take $\mathbf{m} = (0.2, 0.5, 0.3)$, and for each value of β within the range 0 to 2 in increments of 0.01, compute the MPL estimates based on samples generated from the Erdős-Rényi Potts model with the corresponding true parameters lying on the line of inestimability. We use Gibbs sampling for the simulation.

For both $N = 100$ and $N = 200$, the green points, representing the MPL estimates for the sparse case, lie very close to the line of inestimability, thereby supporting our result that joint estimation is always possible in the sparse (bounded-degree) case. The closeness increases from $N = 100$ to $N = 200$, as indicated by the $N^{-1/2}$ rate of convergence. On the other hand, the red points, representing the MPL estimates for the dense case, seem to scatter away from the blue line, a phenomenon which is best observed from the significantly high number of red points that have very large values of $\hat{\beta}$ in comparison to the height of the blue line. The green points, on the other hand, have $\hat{\beta}$ values lying more or less within the height limits of the line of inestimability. This demonstrates the inestimability phenomenon in the dense case.

4. Acknowledgement

The first author gratefully acknowledges support from the National University of Singapore Start-Up Grant A-0008523-00-00 and the AcRF Tier 1 grants A-8001449-00-00 and A-8002932-00-00. The second author gratefully acknowledges NSF support (DMS-2515519, DMS-2113414) during this research. The third author gratefully acknowledges NSF support (DMS-2124222) during this research. We also thank Moumanti Podder for her valuable assistance with an earlier version of this manuscript.

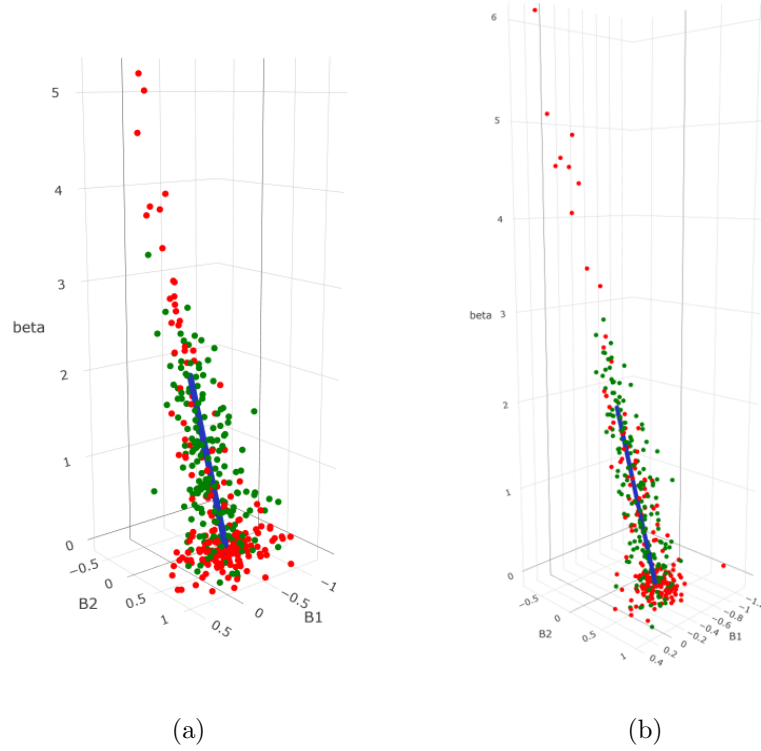


Fig 1: Plot of the MPL estimate $(\hat{\beta}, \hat{B}_1, \hat{B}_2)$ for the Potts model on $\mathcal{G}(N, p)$ with (a) $N = 100$ and (b) $N = 200$. Blue line denotes the line of inestimability, green points denote the MPL estimates for $p = 0.025$ (the sparse case) and red points denote the MPL estimates for $p = 0.25$ (the dense case).

References

- ALI, A. M., FARAG, A. A. and GIMEL'FARB, G. (2008). Analytical method for MGRF Potts model parameter estimation. In *2008 19th International Conference on Pattern Recognition* 1–4. IEEE.
- ANANDKUMAR, A., TAN, V. Y., HUANG, F. and WILLSKY, A. S. (2012). High-dimensional structure estimation in Ising models: Local separation criterion. *The Annals of Statistics* 1346–1375.
- AOUN, Y., DOBER, M. and GLAZMAN, A. (2024). Phase Diagram of the Ashkin–Teller Model. *Communications in Mathematical Physics* **405** 37.
- ASHKIN, J. and TELLER, E. (1943). Statistics of Two-Dimensional Lattices with Four Components. *Physical Review* **64** 178–184.
- BASAK, A. and MUKHERJEE, S. (2017). Universality of the mean-field for the Potts model. *Probability theory and related fields* **168** 557–600.
- BERTHET, Q., RIGOLLET, P. and SRIVASTAVA, P. (2019). Exact recovery in the Ising blockmodel. *Annals of Statistics* **47** 1805–1834.
- BESAG, J. (1974). Spatial Interaction and the Statistical Analysis of Lattice Systems. *Journal of the Royal Statistical Society. Series B (Methodological)* **36** 192–236.
- BESAG, J. (1975). Statistical Analysis of Non-Lattice Data. *The Statistician* **24** 179–195.
- BHAMIDI, S., BRESLER, G. and SLY, A. (2011). Mixing Time of Exponential Random Graphs. *The Annals of Applied Probability* **21** 2146–2170.
- BHATTACHARYA, B. B. and MUKHERJEE, S. (2018). Inference in Ising models. *Bernoulli* **24** 493–525.
- BHOWAL, S. and MUKHERJEE, S. (2023). Supplement to “Limit theorems and phase transitions in the tensor Curie–Weiss Potts model”. *Information and Inference: A Journal of the IMA* 1–43.
- BHOWAL, S. and MUKHERJEE, S. (2025a). Limit theorems and phase transitions in the tensor Curie–Weiss Potts model. *Information and Inference: A Journal of the IMA* **14** iaaf014.
- BHOWAL, S. and MUKHERJEE, S. (2025b). Rates of Convergence of the Magnetization in the Tensor Curie–Weiss Potts Model. *Journal of Statistical Physics* **192** 2.
- BOAS, S. E. M., JIANG, Y., MERKS, R. M. H., PROKOPIOU, S. A. and RENS, E. G. (2018). Cellular Potts Model: Applications to Vasculogenesis and Angiogenesis. *Probabilistic Cellular Automata* **27** 279–310.
- BORNHOLDT, S. (2021). A q-spin Potts model of markets: Gain–loss asymmetry in stock indices as an emergent phenomenon. *arXiv preprint arXiv:2112.06290*.
- BOSCONTI, C., CORALLO, A., FORTUNATO, L., GENTILE, A. A., MASSAFRA, A. and PELLÈ, P. (2015). Reconstruction of a Real World Social Network using the Potts Model and Loopy Belief Propagation. *Frontiers in Psychology* **6**.
- BRESLER, G. (2015). Efficiently learning Ising models on arbitrary graphs. In *Proceedings of the forty-seventh annual ACM symposium on Theory of com-*

- puting 771–782.
- CAO, Y., NEYKOV, M. and LIU, H. (2022). High-temperature structure detection in ferromagnets. *Information and Inference: A Journal of the IMA* **11** 55–102.
- CELEUX, G., FORBES, F. and PEYRARD, N. (2002). EM-based image segmentation using Potts models with external field Technical Report No. RR-4456, INRIA.
- CHATTERJEE, S. (2007a). Estimation in spin glasses: A first step. *The Annals of Statistics* **35** 1931–1946.
- CHATTERJEE, S. (2007b). Stein’s method for concentration inequalities. *Probability theory and related fields* **138** 305–321.
- CHATTERJEE, S. and DEMBO, A. (2016). Nonlinear large deviations. *Advances in Mathematics* **299** 396–450.
- CHATTERJEE, S. and DIACONIS, P. (2013). Estimating and understanding exponential random graph models. *The Annals of Statistics* **41** 2428–2461.
- CHEN, W.-K., SEN, A. and WU, Q. (2024). Joint Parameter Estimations for Spin Glasses. *arXiv preprint arXiv:2406.10760*.
- COMETS, F. and GIDAS, B. (1991). Asymptotics of Maximum Likelihood Estimators for the Curie–Weiss Model. *The Annals of Statistics* **19** 557–578.
- DASKALAKIS, C., DIKKALA, N. and PANAGEAS, I. (2020). Logistic regression with peer-group effects via inference in higher-order Ising models. In *International Conference on Artificial Intelligence and Statistics* 3653–3663. PMLR.
- DEB, N. and MUKHERJEE, S. (2023). Fluctuations in mean-field Ising models. *The Annals of Applied Probability* **33** 1961–2003.
- DEB, N., MUKHERJEE, R., MUKHERJEE, S. and YUAN, M. (2024). Detecting Structured Signals in Ising Models. *The Annals of Applied Probability* **34** 1–45.
- DEMUSE, R., EASLICK, T. and YIN, M. (2019). Mixing time of vertex-weighted exponential random graphs. *J. Comput. Appl. Math.* **362** 443–459.
- DESCOMBES, X., MORRIS, R. D., ZERUBIA, J. and BERTHOD, M. (1999). Estimation of Markov random field prior parameters using Markov chain Monte Carlo maximum likelihood. *IEEE Transactions on Image Processing* **8** 954–963.
- DING, J., LUBETZKY, E. and PERES, Y. (2009). The Mixing Time Evolution of Glauber Dynamics for the Mean-Field Ising Model. *Communications in Mathematical Physics* **289** 725–764.
- EICHELSBACHER, P. and MARTSCHINK, B. (2015). On Rates of Convergence in the Curie–Weiss–Potts Model with an External Field. *Annales de l’Institut Henri Poincaré, Probabilités et Statistiques* **51** 252–282.
- ELLIS, R. S. (1985). *Entropy, Large Deviations, and Statistical Mechanics*. Springer, New York.
- ELLIS, R. S., NEWMAN, C. M. and ROSEN, J. S. (1980). Limit theorems for sums of dependent random variables occurring in statistical mechanics. *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete* **51** 153–169.
- ELLIS, R. S. and WANG, K. (1990). Limit theorems for the empirical vector of

- the Curie-Weiss-Potts model. *Stochastic Processes and their Applications* **35** 59–79.
- ELLIS, R. S. and WANG, K. (1992). Limit theorems for maximum likelihood estimators in the Curie-Weiss-Potts model. *Stochastic Processes and their Applications* **40** 251–288.
- GANDOLFO, D., RUIZ, J. and MARC, W. (2010). Limit theorems and coexistence probabilities for the Curie-Weiss Potts model with an external field. *Stochastic Processes and their Applications* **120** 84–104.
- GHOSAL, P., MUKHERJEE, S. et al. (2020). Joint estimation of parameters in Ising model. *The Annals of Statistics* **48** 785–810.
- GIMENEZ, J., FRERY, A. C. and FLESIA, A. G. (2013). Inference strategies for the smoothness parameter in the Potts model. In *2013 IEEE International Geoscience and Remote Sensing Symposium-IGARSS* 2539–2542. IEEE.
- GRANER, F. and GLAZIER, J. A. (1992). Simulation of biological cell sorting using a two-dimensional extended Potts model. *Physical Review Letters* **69** 2013–2016.
- HANDCOCK, M. S., ROBINS, G., SNIJDERS, T., MOODY, J. and BESAG, J. (2003). Assessing degeneracy in statistical models of social networks Technical Report, Working paper.
- HE, R. and LOK, J. (2025). On Approximating the Potts Model with Contracting Glauber Dynamics. *Probability in the Engineering and Informational Sciences*. Published online November 7, 2025.
- ISING, E. (1925). Beitrag zur Theorie des Ferromagnetismus. *Zeitschrift für Physik* **31** 253–258.
- KENNA, R. (2005). The XY Model and the Berezinskii-Kosterlitz-Thouless Phase Transition. *arXiv preprint arXiv:cond-mat/0512356*.
- KIRKPATRICK, K. and NAWAZ, T. (2016). Asymptotics of Mean-Field O(N) Models. *Journal of Statistical Physics* **165** 1114–1140.
- LEVADA, A. L., MASCARENHAS, N. D. and TANNÚS, A. (2008). Pseudolikelihood equations for potts mrf model parameter estimation on higher order neighborhood systems. *IEEE Geoscience and Remote Sensing Letters* **5** 522–526.
- LEVADA, A. L. M., MASCARENHAS, N. D. A. and TANNÚS, A. (2009). Pseudolikelihood equations for Potts model on higher-order neighborhood systems: A quantitative approach for parameter estimation in image analysis. *Brazilian Journal of Probability and Statistics* **23** 120–140.
- LEVADA, A. L., MASCARENHAS, N. D., TANNÚS, A. and SALVADEO, D. H. (2008). Spatially non-homogeneous potts model parameter estimation on higher-order neighborhood systems by maximum pseudo-likelihood. In *Proceedings of the 2008 ACM symposium on Applied computing* 1733–1737.
- LEVIN, D. A., LUCZAK, M. J. and PERES, Y. (2010). Glauber Dynamics for the Mean-Field Ising Model: Cutoff, Critical Power Law, and Metastability. *Probability Theory and Related Fields* **146** 223–265.
- LOKHOV, A. Y., VUFFRAY, M., MISRA, S. and CHERTKOV, M. (2018). Optimal structure and parameter learning of Ising models. *Science Advances* **4** e1700791.

- LOVÁSZ, L. (2012). *Large Networks and Graph Limits. American Mathematical Society Colloquium Publications* **60**. American Mathematical Society, Providence, RI. [MR3012035](#)
- MCGRORY, C. A., TITTERINGTON, D. M., REEVES, R. and PETTITT, A. N. (2009). Variational Bayes for estimating the parameters of a hidden Potts model. *Statistics and Computing* **19** 329–340.
- MOLTCHANOVA, E. V., PITKÄNIEMI, J. and HAAPALA, L. (2005). Potts model for haplotype associations. *BMC Genetics* **6** S64.
- MUKHERJEE, S. (2020). Degeneracy in sparse ERGMs with functions of degrees as sufficient statistics. *Bernoulli* **26** 1016–1043.
- MUKHERJEE, R., MUKHERJEE, S. and YUAN, M. (2018). Global testing against sparse alternatives under Ising models. *The Annals of Statistics* **46** 2062–2093.
- MUKHERJEE, S., SON, J. and BHATTACHARYA, B. B. (2021). Fluctuations of the Magnetization in the p-Spin Curie–Weiss Model. *Communications in Mathematical Physics* **387** 681–728.
- MUKHERJEE, S., SON, J. and BHATTACHARYA, B. B. (2022). Estimation in Tensor Ising Models. *Information and Inference: A Journal of the IMA* **11** 1457–1500.
- MUKHERJEE, S., SON, J. and BHATTACHARYA, B. B. (2025). Phase Transitions of the Maximum Likelihood Estimators in the p-Spin Curie–Weiss Model. *Bernoulli* **31** 1502–1526.
- MUKHERJEE, S. and XU, Y. (2023). Statistics of the Two-Star ERGM. *Bernoulli* **29** 24–51.
- MUKHERJEE, S., SON, J., GHOSH, S. and MUKHERJEE, S. (2024). Efficient Estimation in Tensor Curie–Weiss and Erdős–Rényi Ising Models. *Electronic Journal of Statistics* **18** 2405–2449.
- NEYKOV, M. and LIU, H. (2019). Property Testing in High-Dimensional Ising Models. *The Annals of Statistics* **47** 2472–2503.
- OKABAYASHI, S., JOHNSON, L. and GEYER, C. J. (2011). Extending pseudo-likelihood for Potts models. *Statistica Sinica* 331–347.
- PEREYRA, M., DOBIGEON, N., BATATIA, H. and TOURNERET, J.-Y. (2013). Estimating the granularity coefficient of a Potts-Markov random field within a Markov chain Monte Carlo algorithm. *IEEE Transactions on Image Processing* **22** 2385–2397.
- PEREYRA, M., WHITELEY, N., ANDRIEU, C. and TOURNERET, J.-Y. (2014). Maximum marginal likelihood estimation of the granularity coefficient of a Potts-Markov random field within an MCMC algorithm. In *2014 IEEE Workshop on Statistical Signal Processing (SSP)* 121–124. IEEE.
- POTTS, R. B. (1952). Some generalized order-disorder transformations. In *Mathematical proceedings of the cambridge philosophical society* **48** 106–109. Cambridge University Press.
- RAVIKUMAR, P., WAINWRIGHT, M. J. and LAFFERTY, J. D. (2010). High-dimensional Ising model selection using ℓ_1 -regularized logistic regression. *The Annals of Statistics* **38** 1287–1319.
- ROSU, R.-G., GIOVANNELLI, J.-F., GIREMUS, A. and VACAR, C. (2015). Potts

- model parameter estimation in Bayesian segmentation of piecewise constant images. In *2015 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)* 4080–4084. IEEE.
- SAMANTA, R. J., MUKHERJEE, S. and ZHANG, J. (2024). Mixing Phases and Metastability for the Glauber Dynamics on the p -Spin Curie–Weiss Model. *arXiv preprint arXiv:2412.16952*. Version 2, revised February 20, 2025.
- SNIJDERS, T. A. B., PATTISON, P. E., ROBINS, G. L. and HANDCOCK, M. S. (2006). New Specifications for Exponential Random Graph Models. *Sociological Methodology* **36** 99–153.
- SONG, S., SI, B., HERRMANN, J. M. and FENG, X. (2016). Local autoencoding for parameter estimation in a hidden Potts-Markov random field. *IEEE Transactions on Image Processing* **25** 2324–2336.
- STIVALA, A., ROBINS, G. and LOMI, A. (2020). Exponential random graph model parameter estimation for very large directed networks. *PloS one* **15** e0227804.
- TAKAISHI, T. (2005). Simulations of Financial Markets in a Potts-like Model. *International Journal of Modern Physics C* **16**.
- VUFFRAY, M., MISRA, S., LOKHOV, A. Y. and CHERTKOV, M. (2016). Interaction Screening: Efficient and Sample-Optimal Learning of Ising Models. In *Advances in Neural Information Processing Systems 29 (NeurIPS 2016)*.
- WU, F. Y. (1982). The Potts Model. *Reviews of Modern Physics* **54** 235–268.
- ZUKOVIC, M. (2008). Simulations of Environmental Spatial Data Using Ising and Potts Models. In *SigmaPhi Conference*.

Appendix A: Proof of Theorem 1.3

In view of Theorem 1.1 (c), it suffices to show that under condition (13), we have $T_N(\mathbf{X})^{-1} = O_{\mathbb{P}}(1)$. Now, for every $\delta > 0$, define the set:

$$E_N(\delta) := \{\mathbf{x} \in [q]^N : T_N(\mathbf{x}) \leq \delta\}. \quad (29)$$

In order to prove that $T_N(\mathbf{X})^{-1} = O_{\mathbb{P}}(1)$, it suffices to show the stronger conclusion that there exists $\delta \in (0, 1)$ such that $\mathbb{P}_{\beta, \mathbf{B}}(\mathbf{X} \in E_N(\delta)) = o(1)$. Towards this, using the definition of T_N from (10), we have:

$$\begin{aligned} T_N(\mathbf{x}) &= \sum_{1 \leq r < s \leq q} \left(\frac{1}{N} \sum_{i=1}^N (m_{i,r}(\mathbf{x}) - m_{i,s}(\mathbf{x}))^2 - (\bar{m}_r(\mathbf{x}) - \bar{m}_s(\mathbf{x}))^2 \right) \\ &= \frac{1}{N} \sum_{i=1}^N \sum_{1 \leq r < s \leq q} (m_{i,r}(\mathbf{x}) - m_{i,s}(\mathbf{x}))^2 - \sum_{1 \leq r < s \leq q} (\bar{m}_r(\mathbf{x}) - \bar{m}_s(\mathbf{x}))^2 \quad (30) \\ &= \frac{1}{N} \sum_{i=1}^N \left(q \sum_{r=1}^q m_{i,r}(\mathbf{x})^2 - \left(\sum_{r=1}^q m_{i,r}(\mathbf{x}) \right)^2 \right) - \left(q \sum_{r=1}^q \bar{m}_r(\mathbf{x})^2 - \bar{R}^2 \right) \\ &= \frac{q}{N} \sum_{i=1}^N \sum_{r=1}^q \left[\left(m_{i,r}(\mathbf{x}) - \frac{R_i}{q} \right)^2 - \left(\bar{m}_r(\mathbf{x}) - \frac{\bar{R}}{q} \right)^2 \right] \\ &= \frac{q}{N} \sum_{i=1}^N \sum_{r=1}^q \left(m_{i,r}(\mathbf{x}) - \bar{m}_r(\mathbf{x}) - \frac{1}{q}(R_i - \bar{R}) \right)^2. \quad (31) \end{aligned}$$

where $R_i := \sum_{j=1}^N a_{ij}$ and $\bar{R} := \frac{1}{N} \sum_{i=1}^N R_i$. (as in (14)). Hence, for $\mathbf{x} \in E_N(\delta)$ (as defined in (29)), we have by the Cauchy-Schwarz inequality:

$$\begin{aligned} &\left| \sum_{i=1}^N \sum_{r=1}^q m_{i,r}(\mathbf{x}) x_{i,r} - \sum_{i=1}^N \sum_{r=1}^q \left(\bar{m}_r(\mathbf{x}) + \frac{1}{q}(R_i - \bar{R}) \right) x_{i,r} \right| \\ &\leq \sqrt{qN} \sqrt{\sum_{i=1}^N \sum_{r=1}^q \left(m_{i,r}(\mathbf{x}) - \bar{m}_r(\mathbf{x}) - \frac{1}{q}(R_i - \bar{R}) \right)^2} \\ &\leq N\sqrt{\delta}. \end{aligned}$$

Since $\sum_{i=1}^N \sum_{r=1}^q (R_i - \bar{R}) x_{i,r} = \sum_{i=1}^N (R_i - \bar{R}) = 0$, the above display gives

$$\left| \sum_{i=1}^N \sum_{r=1}^q m_{i,r}(\mathbf{x}) x_{i,r} - \sum_{r=1}^q \bar{m}_r(\mathbf{x}) \sum_{i=1}^N x_{i,r} \right| \leq N\sqrt{\delta}. \quad (32)$$

Next, define the following function:

$$f_r(t_1, \dots, t_q) := \frac{\exp\{\beta t_r + B_r\}}{\sum_{s=1}^q \exp\{\beta t_s + B_s\}}. \quad (33)$$

Recalling that:

$$\theta_{i,r}(\mathbf{x}) = \frac{\exp\{\beta m_{i,r}(\mathbf{x}) + B_r\}}{\sum_{s=1}^q \exp\{\beta m_{i,s}(\mathbf{x}) + B_s\}} = f_r(m_{i,1}(\mathbf{x}), \dots, m_{i,q}(\mathbf{x})) \quad (\text{see (4)})$$

and noting that the first order partial derivatives of f_r are bounded by β , we have the following from the mean-value theorem:

$$\left| f_r(m_{i,1}(\mathbf{x}), \dots, m_{i,q}(\mathbf{x})) - f_r(n_{i,1}(\mathbf{x}), \dots, n_{i,q}(\mathbf{x})) \right| \lesssim \sum_{u=1}^q |m_{i,u}(\mathbf{x}) - n_{i,u}(\mathbf{x})|$$

where

$$n_{i,r}(\mathbf{x}) := \bar{m}_r(\mathbf{x}) + \frac{1}{q}(R_i - \bar{R}).$$

This implies that for $\mathbf{x} \in E_N(\delta)$,

$$\begin{aligned} \sum_{i=1}^N \left| \theta_{i,r}(\mathbf{x}) - f_r(n_{i,1}(\mathbf{x}), \dots, n_{i,q}(\mathbf{x})) \right| &\lesssim \sum_{i=1}^N \sum_{u=1}^q |m_{i,u}(\mathbf{x}) - n_{i,u}(\mathbf{x})| \\ &\leq \sqrt{qN} \sqrt{\sum_{i=1}^N \sum_{u=1}^q (m_{i,u}(\mathbf{x}) - n_{i,u}(\mathbf{x}))^2} \\ &\leq N\sqrt{\delta}. \end{aligned} \quad (34)$$

Invoking (34) together with the observation that

$$f_r(n_{i,1}(\mathbf{x}), \dots, n_{i,q}(\mathbf{x})) = \frac{\exp\{\beta \bar{m}_r(\mathbf{x}) + B_r\}}{\sum_{s=1}^q \exp\{\beta \bar{m}_s(\mathbf{x}) + B_s\}}$$

implies that for $\mathbf{x} \in E_N(\delta)$ we have:

$$\left| \sum_{i=1}^N \theta_{i,r}(\mathbf{x}) - \frac{N \exp\{\beta \bar{m}_r(\mathbf{x}) + B_r\}}{\sum_{s=1}^q \exp\{\beta \bar{m}_s(\mathbf{x}) + B_s\}} \right| \lesssim N\sqrt{\delta} \quad (35)$$

Again using (34) and the fact that $R_i \leq \gamma$ (see (8)), for $\mathbf{x} \in E_N(\delta)$ we have:

$$\begin{aligned} \left| \sum_{i=1}^N R_i \left[\theta_{i,r}(\mathbf{x}) - \frac{\exp\{\beta \bar{m}_r(\mathbf{x}) + B_r\}}{\sum_{s=1}^q \exp\{\beta \bar{m}_s(\mathbf{x}) + B_s\}} \right] \right| &\lesssim N\sqrt{\delta}, \quad \text{i.e.} \\ \left| \sum_{i=1}^N \sum_{j=1}^N a_{ij} \theta_{i,r}(\mathbf{x}) - N\bar{R} \frac{\exp\{\beta \bar{m}_r(\mathbf{x}) + B_r\}}{\sum_{s=1}^q \exp\{\beta \bar{m}_s(\mathbf{x}) + B_s\}} \right| &\lesssim N\sqrt{\delta}. \end{aligned} \quad (36)$$

Next, define the sets

$$C_N(\delta) := \left\{ \mathbf{x} \in [q]^N : \max_{r \in [q]} \left| \sum_{i=1}^N (x_{i,r} - \theta_{i,r}(\mathbf{x})) \right| \leq N\delta \right\},$$

$$D_N(\delta) := \left\{ \mathbf{x} \in [q]^N : \max_{r \in [q]} \left| \sum_{i=1}^N \left(m_{i,r}(\mathbf{x}) - \sum_{j=1}^N a_{ij} \theta_{i,r}(\mathbf{x}) \right) \right| \leq N\delta \right\}.$$

By the triangle inequality, we have:

$$\begin{aligned} & \left| \bar{R} \sum_{i=1}^N x_{i,r} - \sum_{i=1}^N m_{i,r}(\mathbf{x}) \right| \\ & \leq \left| \bar{R} \sum_{i=1}^N x_{i,r} - \bar{R} \sum_{i=1}^N \theta_{i,r}(\mathbf{x}) \right| + \left| \bar{R} \sum_{i=1}^N \theta_{i,r}(\mathbf{x}) - \frac{N\bar{R} \exp\{\beta\bar{m}_r(\mathbf{x}) + B_r\}}{\sum_{s=1}^q \exp\{\beta\bar{m}_s(\mathbf{x}) + B_s\}} \right| \\ & + \left| \frac{N\bar{R} \exp\{\beta\bar{m}_r(\mathbf{x}) + B_r\}}{\sum_{s=1}^q \exp\{\beta\bar{m}_s(\mathbf{x}) + B_s\}} - \sum_{i=1}^N \sum_{j=1}^N a_{ij} \theta_{i,r}(\mathbf{x}) \right| + \left| \sum_{i=1}^N \sum_{j=1}^N a_{ij} \theta_{i,r}(\mathbf{x}) - \sum_{i=1}^N m_{i,r}(\mathbf{x}) \right|. \end{aligned} \quad (37)$$

Suppose now, that $\mathbf{x} \in F_N(\delta) := C_N(\delta) \cap D_N(\delta) \cap E_N(\delta)$ for some $\delta \in (0, 1)$. Since $\mathbf{x} \in C_N(\delta)$, we have:

$$\left| \bar{R} \sum_{i=1}^N x_{i,r} - \bar{R} \sum_{i=1}^N \theta_{i,r}(\mathbf{x}) \right| \leq N\delta\gamma \lesssim N\sqrt{\delta}.$$

Next, in view of (35) (since $\mathbf{x} \in E_N(\delta)$), we have:

$$\left| \bar{R} \sum_{i=1}^N \theta_{i,r}(\mathbf{x}) - \frac{N\bar{R} \exp\{\beta\bar{m}_r(\mathbf{x}) + B_r\}}{\sum_{s=1}^q \exp\{\beta\bar{m}_s(\mathbf{x}) + B_s\}} \right| \lesssim N\sqrt{\delta}\gamma \lesssim N\sqrt{\delta}.$$

By (36) (since $\mathbf{x} \in E_N(\delta)$), we have:

$$\left| \frac{N\bar{R} \exp\{\beta\bar{m}_r(\mathbf{x}) + B_r\}}{\sum_{s=1}^q \exp\{\beta\bar{m}_s(\mathbf{x}) + B_s\}} - \sum_{i=1}^N \sum_{j=1}^N a_{ij} \theta_{i,r}(\mathbf{x}) \right| \lesssim N\sqrt{\delta}.$$

Finally, since $\mathbf{x} \in D_N(\delta)$, we have:

$$\left| \sum_{i=1}^N \sum_{j=1}^N a_{ij} \theta_{i,r}(\mathbf{x}) - \sum_{i=1}^N m_{i,r}(\mathbf{x}) \right| \leq N\delta \leq N\sqrt{\delta}.$$

Plugging these inequalities in (37), we have:

$$\left| \bar{R} \sum_{i=1}^N x_{i,r} - \sum_{i=1}^N m_{i,r}(\mathbf{x}) \right| \lesssim 4N\sqrt{\delta}. \quad (38)$$

Denoting $\bar{x}_r := N^{-1} \sum_{i=1}^N x_{i,r}$, for all $\mathbf{x} \in [q]^N$ we have:

$$\left| \sum_{i=1}^N \sum_{r=1}^q x_{i,r} m_{i,r}(\mathbf{x}) - N\bar{R} \sum_{r=1}^q \bar{x}_r^2 \right|$$

$$\begin{aligned}
&\leq \left| \sum_{i=1}^N \sum_{r=1}^q x_{i,r} m_{i,r}(\mathbf{x}) - N \sum_{r=1}^q \bar{m}_r(\mathbf{x}) \bar{x}_r \right| + \left| N \sum_{r=1}^q \bar{x}_r (\bar{m}_r(\mathbf{x}) - \bar{R} \bar{x}_r) \right| \\
&\lesssim N\sqrt{\delta},
\end{aligned}$$

where we use (32) and (38) respectively, along with the trivial bound $\bar{x}_r \leq 1$. With K denoting the implied constant in the above display, we get

$$\begin{aligned}
\mathbb{P}_{\beta, \mathbf{B}}(\mathbf{X} \in F_N(\delta)) &= \frac{1}{Z_N(\beta, \mathbf{B})} \sum_{\mathbf{x} \in F_N(\delta)} \exp\left(\frac{\beta}{2} \sum_{i=1}^N \sum_{r=1}^q m_{i,r}(\mathbf{x}) x_{i,r} + \sum_{i=1}^N \sum_{r=1}^q B_r x_{i,r}\right) \\
&\leq \frac{e^{KN\sqrt{\delta}}}{Z_N(\beta, \mathbf{B})} \sum_{\mathbf{x} \in F_N(\delta)} \exp\left(\frac{N\beta\bar{R}}{2} \sum_{r=1}^q \bar{x}_r^2 + N \sum_{r=1}^q B_r \bar{x}_r\right) \quad (39) \\
&= e^{KN\sqrt{\delta}} \frac{Z_N^{\text{CW}}(\beta\bar{R}, \mathbf{B})}{Z_N(\beta, \mathbf{B})} \mathbb{P}_{\beta\bar{R}, \mathbf{B}}^{\text{CW}}(\mathbf{X} \in F_N(\delta)) \\
&\leq e^{KN\sqrt{\delta}} \frac{Z_N^{\text{CW}}(\beta\bar{R}, \mathbf{B})}{Z_N(\beta, \mathbf{B})} \mathbb{P}_{\beta\bar{R}, \mathbf{B}}^{\text{CW}}(\mathbf{X} \in E_N(\delta))
\end{aligned}$$

where for $(\beta, \mathbf{B}) \in (0, \infty) \times \mathbb{R}^{q-1}$, we define the Curie-Weiss Potts model with parameter (β, \mathbf{B}) as:

$$\mathbb{P}_{\beta, \mathbf{B}}^{\text{CW}}(\mathbf{x}) = \frac{1}{Z_N^{\text{CW}}(\beta, \mathbf{B})} \exp\left(\frac{N\beta}{2} \sum_{r=1}^q \bar{x}_r^2 + N \sum_{r=1}^q B_r \bar{x}_r\right). \quad (40)$$

Now, it follows from the Gibbs variational principle/mean field lower bound (see Equation (1.8) in Basak and Mukherjee (2017)) that:

$$\begin{aligned}
\log Z_N(\beta, \mathbf{B}) &\geq \sup_{\mathbf{t} \in \mathcal{P}([q])} \frac{\beta}{2} \sum_{i,j} a_{ij} \sum_{r=1}^q t_r^2 - N \sum_{r=1}^q t_r \log t_r + N \sum_{r=1}^q B_r t_r \\
&= \sup_{\mathbf{t} \in \mathcal{P}([q])} \frac{N\beta}{2} \bar{R} \sum_{r=1}^q t_r^2 - N \sum_{r=1}^q t_r \log t_r + N \sum_{r=1}^q B_r t_r.
\end{aligned}$$

Also, it follows from equation (2.3) in Basak and Mukherjee (2017) that:

$$\log Z_N^{\text{CW}}(\beta\bar{R}, \mathbf{B}) = \sup_{\mathbf{t} \in \mathcal{P}([q])} \left[\frac{N\beta}{2} \bar{R} \sum_{r=1}^q t_r^2 - N \sum_{r=1}^q t_r \log t_r + N \sum_{r=1}^q B_r t_r \right] + o(N).$$

Combining the above two displays gives

$$\frac{Z_N^{\text{CW}}(\beta\bar{R}, \mathbf{B})}{Z_N(\beta, \mathbf{B})} \leq e^{o(N)},$$

which along with (39) gives

$$\mathbb{P}_{\beta, \mathbf{B}}(\mathbf{X} \in F_N(\delta)) \leq e^{KN(\sqrt{\delta} + o(1))} \mathbb{P}_{\beta\bar{R}, \mathbf{B}}^{\text{CW}}(\mathbf{X} \in E_N(\delta)). \quad (41)$$

The following lemma bounds $\mathbb{P}_{\beta\bar{R}, \mathbf{B}}^{\text{CW}}(\mathbf{X} \in E_N(\delta))$.

Lemma A.1. *Suppose the interaction matrix \mathbf{A}_N satisfies the conditions (8), (9) and (13). There exist constants $A, B > 0$ depending only on β, B, γ, q , such that:*

$$\mathbb{P}_{\beta, \mathbf{B}}^{\text{CW}}(\mathbf{X} \in E_N(A)) \leq e^{-BN}.$$

Lemma A.1 is proved in Appendix G. Using Lemma A.1 together with (41), we know that for every $\delta \in (0, A)$,

$$\mathbb{P}_{\beta, \mathbf{B}}(\mathbf{X} \in F_N(\delta)) \leq \exp\{N(K\sqrt{\delta} - B + o(1))\}.$$

We can now choose $\delta = \min\left\{\frac{A}{2}, \frac{B^2}{4K^2}\right\}$ small enough, such that

$$\mathbb{P}_{\beta, \mathbf{B}}(\mathbf{X} \in F_N(\delta)) \leq e^{-\frac{NB}{2} + o(1)} = o(1). \quad (42)$$

Recalling that $F_N(\delta) = C_N(\delta) \cap D_N(\delta) \cap E_N(\delta)$, we get

$$\mathbb{P}_{\beta, \mathbf{B}}(\mathbf{X} \in E_N(\delta)) \leq \mathbb{P}_{\beta, \mathbf{B}}(\mathbf{X} \in F_N(\delta)) + \mathbb{P}\left(\mathbf{X} \notin C_N(\delta) \cap D_N(\delta)\right).$$

By (42), the first term in the RHS of the above inequality is $o(1)$. It also follows directly from Lemma F.1 (on taking $b_{itrs} := \mathbb{1}_{t=r}$ and $g \equiv 1$), that $\mathbb{P}_{\beta, \mathbf{B}}(\mathbf{X} \notin C_N(\delta)) = o(1)$. Finally, note that:

$$\sum_{i=1}^N \left(m_{i,r}(\mathbf{x}) - \sum_{j=1}^N a_{ij} \theta_{i,r}(\mathbf{x}) \right) = \sum_{j=1}^N R_j x_{j,r} - \sum_{i=1}^N R_i \theta_{i,r}(\mathbf{x}) = \sum_{i=1}^N R_i (x_{i,r} - \theta_{i,r}(\mathbf{x})).$$

It now follows from Lemma F.1 (on taking $b_{itrs} := R_i \mathbb{1}_{t=r}$ and $g \equiv 1$), that $\mathbb{P}_{\beta, \mathbf{B}}(\mathbf{X} \notin D_N(\delta)) = o(1)$. Hence, $\mathbb{P}_{\beta, \mathbf{B}}(\mathbf{X} \notin C_N(\delta) \cap D_N(\delta)) = o(1)$. We thus conclude that $\mathbb{P}_{\beta, \mathbf{B}}(\mathbf{X} \in E_N(\delta)) = o(1)$, thereby completing the proof of Theorem 1.3. \square

Appendix B: Proof of Theorem 1.4

Since we have already established \sqrt{N} -consistency of the MPL estimator $(\hat{\beta}_N, \hat{\mathbf{B}}_N)$ in the non mean-field setup of Subsection 1.3.1 (Theorem 1.3), we will assume that condition (13) does not hold, i.e. we will assume the following *mean-field* condition:

$$\lim_{N \rightarrow \infty} \frac{1}{N} \sum_{1 \leq i, j \leq N} a_{ij}^2 = 0. \quad (43)$$

Note that if the coupling matrix is the scaled adjacency of a graph (2), then the mean field condition says that the average degree of the graph goes to ∞ . Once again, in view of Theorem 1.1 (b), we just need to show that $T_N(\mathbf{X})^{-1} = O_{\mathbb{P}}(1)$. The sequence of measures $\mu_N := N^{-1} \sum_{i=1}^N \delta_{R_i}$ are supported on the compact set $[0, \gamma]$, and so by Prokhorov's theorem and possibly passing to subsequences,

we can assume that $\mu_N \xrightarrow{w} \mu$ for some probability measure μ on $[0, \gamma]$. The dominated convergence theorem now implies that:

$$\lim_{N \rightarrow \infty} \frac{1}{N} \sum_{i=1}^N (R_i - \bar{R})^2 = \int_{[0, \gamma]} (\theta - \mathbb{E}_\mu(\theta))^2 d\mu(\theta). \quad (44)$$

Using (31) we have

$$\begin{aligned} T_N(\mathbf{X}) &= \frac{q}{N} \sum_{i=1}^N \sum_{r=1}^q \left(m_{i,r}(\mathbf{X}) - \bar{m}_r(\mathbf{X}) - \frac{1}{q}(R_i - \bar{R}) \right)^2 \\ &= \frac{q}{2N^2} \sum_{1 \leq i, j \leq N} \sum_{r=1}^q \left(m_{i,r}(\mathbf{X}) - m_{j,r}(\mathbf{X}) - \frac{1}{q}(R_i - R_j) \right)^2. \end{aligned} \quad (45)$$

Setting $u_{i,r}(\mathbf{X}) := m_{i,r}(\mathbf{X}) - R_i/q$ and recalling the definition of f_r and $\theta_{i,r}(\mathbf{X})$ (see (33) and (4)),

$$f_r(t_1, \dots, t_q) = \frac{\exp\{\beta t_r + B_r\}}{\sum_{s=1}^q \exp\{\beta t_s + B_s\}},$$

we have:

$$\begin{aligned} &|\theta_{i,r}(\mathbf{X}) - \theta_{j,r}(\mathbf{X})| \\ &= |f_r(m_{i,1}(\mathbf{X}), \dots, m_{i,q}(\mathbf{X})) - f_r(m_{j,1}(\mathbf{X}), \dots, m_{j,q}(\mathbf{X}))| \\ &= |f_r(u_{i,1}(\mathbf{X}), \dots, u_{i,q}(\mathbf{X})) - f_r(u_{j,1}(\mathbf{X}), \dots, u_{j,q}(\mathbf{X}))| \\ &\leq \beta \sum_{s=1}^q |u_{i,s}(\mathbf{X}) - u_{j,s}(\mathbf{X})| \end{aligned}$$

where the last step follows from mean-value theorem, and the fact that all the partial derivatives of f_r are bounded by β . Hence, we have:

$$\begin{aligned} \frac{1}{N} \sum_{i=1}^N \sum_{r=1}^q (\theta_{i,r}(\mathbf{X}) - \bar{\theta}_r(\mathbf{X}))^2 &= \frac{1}{2N^2} \sum_{1 \leq i, j \leq N} \sum_{r=1}^q (\theta_{i,r}(\mathbf{X}) - \theta_{j,r}(\mathbf{X}))^2 \\ &\leq \frac{q^2 \beta^2}{2N^2} \sum_{1 \leq i, j \leq N} \sum_{s=1}^q (u_{i,s}(\mathbf{X}) - u_{j,s}(\mathbf{X}))^2 \\ &= q\beta^2 T_N(\mathbf{X}), \end{aligned}$$

where the last equality uses (45). Hence, for showing that $T_N(\mathbf{X})^{-1} = O_{\mathbb{P}}(1)$, it is enough to show that:

$$\left[\sum_{i=1}^N \sum_{r=1}^q (\theta_{i,r}(\mathbf{X}) - \bar{\theta}_r(\mathbf{X}))^2 \right]^{-1} = O_{\mathbb{P}}(N^{-1}). \quad (46)$$

Towards this, let $\mathcal{S}_{N,q}$ denote the set of all $\mathbf{y} := ((y_{i,r}))_{i \in [N], r \in [q]} \in [0, 1]^{Nq}$, such that $\sum_{r=1}^q y_{i,r} = 1$ for all $i \in [N]$. Define functions $h_N : \mathcal{S}_{N,q} \rightarrow \mathbb{R}$ and $I_N : \mathcal{S}_{N,q} \rightarrow \mathbb{R}$ as:

$$h_N(\mathbf{y}) := \frac{\beta}{2} \sum_{1 \leq i, j \leq N} \sum_{r=1}^q a_{ij} y_{i,r} y_{j,r} + \sum_{i=1}^N \sum_{r=1}^q B_r y_{i,r} \quad \text{and} \quad I_N(\mathbf{y}) := \sum_{i=1}^N \sum_{r=1}^q y_{i,r} \log y_{i,r}. \quad (47)$$

Also, define

$$\psi_N(\mathbf{y}) := h_N(\mathbf{y}) - I_N(\mathbf{y}). \quad (48)$$

Then, we have:

$$\frac{\partial \psi_N(\mathbf{y})}{\partial y_{i,r}} = \beta \sum_{j=1}^N a_{ij} y_{j,r} + B_r - 1 - \log y_{i,r}. \quad (49)$$

Denoting $\nabla_{\cdot,r}(\mathbf{y}) := ((\partial \psi_N / \partial y_{i,r}))_{i \in [N]}$, we have:

$$\nabla_{\cdot,r}(\mathbf{y}) = \beta \mathbf{A}_N \mathbf{y}_{\cdot,r} + (B_r - 1) \mathbf{1}_N - \log \mathbf{y}_{\cdot,r} \quad (50)$$

where $\mathbf{y}_{\cdot,r} := ((y_{i,r}))_{i \in [N]}$. Also, define $\bar{\nabla}(\mathbf{y}) \in \mathbb{R}^N$ and $\tilde{\nabla}(\mathbf{y}) \in \mathbb{R}^{Nq}$ as:

$$\bar{\nabla}(\mathbf{y}) := \frac{1}{q} \sum_{r=1}^q \nabla_{\cdot,r}(\mathbf{y}) \quad \text{and} \quad \tilde{\nabla}(\mathbf{y}) := (\bar{\nabla}(\mathbf{y}), \dots, \bar{\nabla}(\mathbf{y})). \quad (51)$$

Note that $\tilde{\nabla}(\mathbf{y})$ is obtained from $\nabla \psi_N(\mathbf{y})$ by replacing each row $(\partial \psi_N(\mathbf{y}) / \partial y_{i,r})_{r \in [q]}$ by the constant vector of its average. Therefore, whenever some vector $\mathbf{y} \in \mathcal{S}_{N,q}$ satisfies the conditions:

$$\sum_{i=1}^N \sum_{r=1}^q (y_{i,r} - \bar{y}_r)^2 \leq N \delta_N \quad \text{and} \quad \min_{i \in [N], r \in [q]} y_{i,r} \geq \alpha \quad (52)$$

for some non-negative sequence $\delta_N \rightarrow 0$ and $\alpha > 0$, then with $\tilde{\mathbf{y}} \in \mathcal{S}_{N,q}$ defined as $\tilde{y}_{i,r} := \bar{y}_r$ for all i, r , we have:

$$\begin{aligned} & \|\nabla_{\cdot,r}(\mathbf{y}) - \bar{\nabla}(\mathbf{y})\| \\ \geq & \|\nabla_{\cdot,r}(\tilde{\mathbf{y}}) - \bar{\nabla}(\tilde{\mathbf{y}})\| - \|\nabla_{\cdot,r}(\mathbf{y}) - \nabla_{\cdot,r}(\tilde{\mathbf{y}})\| - \|\bar{\nabla}(\mathbf{y}) - \bar{\nabla}(\tilde{\mathbf{y}})\| \\ \geq & \|\nabla_{\cdot,r}(\tilde{\mathbf{y}}) - \bar{\nabla}(\tilde{\mathbf{y}})\| - \|\mathbf{y}_{\cdot,r} - \tilde{\mathbf{y}}_{\cdot,r}\| \left(\beta \|\mathbf{A}_N\|_2 + \frac{1}{\alpha} \right) - \frac{1}{q} \sum_{s=1}^q \|\mathbf{y}_{\cdot,s} - \tilde{\mathbf{y}}_{\cdot,s}\| \left(\beta \|\mathbf{A}_N\|_2 + \frac{1}{\alpha} \right) \\ \geq & \|\nabla_{\cdot,r}(\tilde{\mathbf{y}}) - \bar{\nabla}(\tilde{\mathbf{y}})\| - 2\sqrt{N\delta_N}(\beta\gamma + \alpha^{-1}) \\ = & \|\nabla_{\cdot,r}(\tilde{\mathbf{y}}) - \bar{\nabla}(\tilde{\mathbf{y}})\| - o(\sqrt{N}). \end{aligned}$$

In the above display, the second inequality uses the expression of $\psi_N(\mathbf{y})$ in (50), and the third inequality uses (52). To bound the RHS above, again use (50) to note that

$$\|\nabla_{\cdot,r}(\tilde{\mathbf{y}}) - \bar{\nabla}(\tilde{\mathbf{y}})\|^2$$

$$\begin{aligned}
&= \sum_{i=1}^N \left[\beta \bar{y}_r R_i + B_r - 1 - \log \bar{y}_r - \frac{1}{q} \sum_{s=1}^q (\beta \bar{y}_s R_i + B_s - 1 - \log \bar{y}_s) \right]^2 \\
&\geq \inf_{\mathbf{t} \in [\alpha, 1]^q} \sum_{i=1}^N \left(\beta(t_r - \bar{t}) R_i + B_r - \frac{1}{q} \sum_{s=1}^q B_s - (\log t_r - q^{-1} \sum_{s=1}^q \log t_s) \right)^2 \\
&= N \inf_{\mathbf{t} \in [\alpha, 1]^q} \int_0^\gamma \left(\beta(t_r - \bar{t}) \theta + B_r - \frac{1}{q} \sum_{s=1}^q B_s - (\log t_r - q^{-1} \sum_{s=1}^q \log t_s) \right)^2 d\mu(\theta) \\
&+ o(N),
\end{aligned}$$

where the last inequality uses the weak convergence of μ_N to μ . Using the above two displays, and noting that $\theta_{i,r}(\mathbf{X}) \geq \alpha := q^{-1} \exp\{-\beta\gamma - 2\|\mathbf{B}\|_\infty\}$, on the event $\sum_{i=1}^N \sum_{r=1}^q (\theta_{i,r}(\mathbf{X}) - \bar{\theta}_r(\mathbf{X}))^2 \leq N\delta_N$ we have:

$$\begin{aligned}
&N^{-1/2} \|\nabla_{\cdot r}(\boldsymbol{\theta}(\mathbf{X})) - \bar{\nabla}(\boldsymbol{\theta}(\mathbf{X}))\| + o(1) \\
&\geq \sqrt{\inf_{\mathbf{t} \in [\alpha, 1]^q} \int_0^\gamma \left(\beta(t_r - \bar{t}) \theta + B_r - \frac{1}{q} \sum_{s=1}^q B_s - (\log t_r - q^{-1} \sum_{s=1}^q \log t_s) \right)^2 d\mu(\theta)},
\end{aligned} \tag{53}$$

where $\boldsymbol{\theta}(\mathbf{X}) := ((\theta_{i,r}(\mathbf{X})))_{i \in [N], r \in [q]}$. We now claim that the RHS of (53) is strictly positive for at least one $r \in [q]$. Given the claim, on the event $\sum_{i=1}^N \sum_{r=1}^q (\theta_{i,r}(\mathbf{X}) - \bar{\theta}_r(\mathbf{X}))^2 \leq N\delta_N$ we have the existence of a constant $c > 0$, free of N , such that

$$\max_{r \in [q]} \|\nabla_{\cdot r}(\boldsymbol{\theta}(\mathbf{X})) - \bar{\nabla}(\boldsymbol{\theta}(\mathbf{X}))\| \geq (c - o(1))\sqrt{N}.$$

Consequently, for any non-negative sequence $\delta_N \rightarrow 0$ we have

$$\begin{aligned}
&\mathbb{P} \left(\sum_{i=1}^N \sum_{r=1}^q (\theta_{i,r}(\mathbf{X}) - \bar{\theta}_r(\mathbf{X}))^2 \leq N\delta_N \right) \\
&\leq \mathbb{P} \left(\max_{r \in [q]} \|\nabla_{\cdot r}(\boldsymbol{\theta}(\mathbf{X})) - \bar{\nabla}(\boldsymbol{\theta}(\mathbf{X}))\| \geq (c - o(1))\sqrt{N} \right) \rightarrow 0,
\end{aligned}$$

where the last limit uses Lemma I.6 (b). This establishes (46) and completes the proof of Theorem 1.4.

It thus remains to prove the claim that the RHS of (53) is strictly positive. To this effect, assume by contradiction that the RHS is 0 for all $r \in [q]$. By the dominated convergence theorem, the map

$$\mathbf{t} \mapsto \int_0^\gamma \left(\beta(t_r - \bar{t}) \theta + B_r - \frac{1}{q} \sum_{s=1}^q B_s - (\log t_r - q^{-1} \sum_{s=1}^q \log t_s) \right)^2 d\mu(\theta) \tag{54}$$

is continuous on the compact set $[p, 1]^q$, and hence, attains its infimum over $[p, 1]^q$ at some point $\mathbf{t}^{(r)} \in [p, 1]^q$. First suppose that $t_r^{(r)} \neq \bar{t}^{(r)}$ for some $r \in [q]$, whence we must have the following under μ :

$$\theta \stackrel{\text{a.s.}}{=} \beta^{-1} (t_r^{(r)} - \bar{t}^{(r)})^{-1} \left[B_r - \frac{1}{q} \sum_{r=1}^q B_r - (\log t_r^{(r)} - q^{-1} \sum_{s=1}^q \log t_s^{(r)}) \right]$$

a contradiction, since μ is not a degenerate measure in view of (44) and condition (14).

This forces $t_r^{(r)} = \bar{t}^{(r)}$ for all $r \in [q]$. In this case, Jensen's inequality gives

$$\log t_r^{(r)} - q^{-1} \sum_{s=1}^q \log t_s^{(r)} \geq 0$$

for all $r \in [q]$. Suppose further, that there exists $r \in [q]$ such that

$$B_r - q^{-1} \sum_{s=1}^q B_s < 0.$$

In this case, $B_r - \frac{1}{q} \sum_{s=1}^q B_s - (\log t_r^{(r)} - q^{-1} \sum_{s=1}^q \log t_s^{(r)}) < 0$ and hence, the integral (54) is strictly positive for that r , a contradiction.

The only case left for consideration is when $B_r - \frac{1}{q} \sum_{s=1}^q B_s \geq 0$ for all $r \in [q]$. This however forces $B_1 = \dots = B_q = 0$, which was ruled out in the hypothesis of Theorem 1.4 (in fact, in this case, the minimum value is exactly 0, attained at any constant vector $\mathbf{t} \in [\alpha, 1]^q$). This completes the proof of our claim, and subsequently, the theorem. \square

Appendix C: Proof of Theorem 1.5

In this section, we prove Theorem 1.5. To begin with, note that by Lemma H.5, the product measure $\mathbf{m}^N := \otimes_{i=1}^N \mathbf{m}$ is contiguous to the Curie-Weiss Potts measure $\mathbb{P}_{\beta, \mathbf{B}}^{\text{CW}}$ in (40). We next claim that for any $(\beta, \mathbf{B}) \in (0, \infty) \times \mathbb{R}^{q-1}$, the product measure $\mathbb{P}_{\beta, \mathbf{B}}^{\text{CW}} \times \mathcal{G}(N, p)$ is contiguous to the measure $\mathbb{P}_{\beta, \mathbf{B}}^{\text{ER}}$. Towards proving this, note that in view of Proposition 6.1 of [Bhattacharya and Mukherjee \(2018\)](#) it suffices to show that:

$$D(\mathbb{P}_{\beta, \mathbf{B}}^{\text{CW}} \times \mathcal{G}(N, p) \parallel \mathbb{P}_{\beta, \mathbf{B}}^{\text{ER}}) = O(1) \quad (55)$$

where D denotes the Kullback-Leibler divergence. Now, we have:

$$\begin{aligned} & D(\mathbb{P}_{\beta, \mathbf{B}}^{\text{CW}} \times \mathcal{G}(N, p) \parallel \mathbb{P}_{\beta, \mathbf{B}}^{\text{ER}}) \\ &= \sum_{(\mathbf{x}, G) \in [q]^N \times \{0, 1\}^{\binom{N}{2}}} (\mathbb{P}_{\beta, \mathbf{B}}^{\text{CW}} \times \mathcal{G}(N, p))(\mathbf{x}, G) \log \frac{(\mathbb{P}_{\beta, \mathbf{B}}^{\text{CW}} \times \mathcal{G}(N, p))(\mathbf{x}, G)}{\mathbb{P}_{\beta, \mathbf{B}}^{\text{ER}}(\mathbf{x}, G)} \end{aligned}$$

$$\begin{aligned}
&= \mathbb{E}_{\mathcal{G}(N,p)} \left[\sum_{\mathbf{x} \in [q]^N} \mathbb{P}_{\beta, \mathbf{B}}^{\text{CW}}(\mathbf{x}) \log \frac{\mathbb{P}_{\beta, \mathbf{B}}^{\text{CW}}(\mathbf{x})}{\mathbb{P}_{\beta, \mathbf{B}}^{\text{ER}}(\mathbf{x}|G)} \right] \\
&= \mathbb{E}_{\mathcal{G}(N,p)} \log \left(\frac{Z_{\beta, \mathbf{B}}^{\text{ER}}}{Z_{\beta, \mathbf{B}}^{\text{CW}}} \right) - \sum_{\mathbf{x} \in [q]^N} \mathbb{P}_{\beta, \mathbf{B}}^{\text{CW}}(\mathbf{x}) \left[\mathbb{E}_{\mathcal{G}(N,p)} \left(\tilde{H}_{\beta, \mathbf{B}}^{\text{ER}}(\mathbf{x}) \right) - \tilde{H}_{\beta, \mathbf{B}}^{\text{CW}}(\mathbf{x}) \right]
\end{aligned}$$

where $\tilde{H}_{\beta, \mathbf{B}}^{\text{ER}}(\mathbf{x})$ and $\tilde{H}_{\beta, \mathbf{B}}^{\text{CW}}(\mathbf{x})$ follow from the expression

$$\frac{\beta}{2} \sum_{1 \leq i, j \leq k} a_{ij} \mathbb{1}_{x_i = x_j} + \sum_{i=1}^n \sum_{r=1}^q B_r \mathbb{1}_{x_i = r}$$

with a_{ij} replaced by (16) and $1/N$, respectively. Note that $\mathbb{E}_{\mathcal{G}(N,p)} \left(\tilde{H}_{\beta, \mathbf{B}}^{\text{ER}}(\mathbf{x}) \right) = \tilde{H}_{\beta, \mathbf{B}}^{\text{CW}}(\mathbf{x})$ and hence, we have by Jensen's inequality:

$$D(\mathbb{P}_{\beta, \mathbf{B}}^{\text{CW}} \times \mathcal{G}(N,p) \parallel \mathbb{P}_{\beta, \mathbf{B}}^{\text{ER}}) = \mathbb{E}_{\mathcal{G}(N,p)} \log \left(\frac{Z_{\beta, \mathbf{B}}^{\text{ER}}}{Z_{\beta, \mathbf{B}}^{\text{CW}}} \right) \leq \log \left(\frac{\mathbb{E}_{\mathcal{G}(N,p)}(Z_{\beta, \mathbf{B}}^{\text{ER}})}{Z_{\beta, \mathbf{B}}^{\text{CW}}} \right). \quad (56)$$

Finally, note that:

$$\begin{aligned}
&\mathbb{E}_{\mathcal{G}(N,p)}(Z_{\beta, \mathbf{B}}^{\text{ER}}) \\
&= \sum_{\mathbf{x} \in [q]^N} \mathbb{E}_{\mathcal{G}(N,p)} \left[\exp \left(\frac{\beta}{2Np} \sum_{1 \leq i, j \leq N} g_{ij} \mathbb{1}_{x_i = x_j} + \sum_{i=1}^N \sum_{r=1}^q B_r \mathbb{1}_{x_i = r} \right) \right] \\
&= \sum_{\mathbf{x} \in [q]^N} e^{\sum_{i=1}^N \sum_{r=1}^q B_r \mathbb{1}_{x_i = r}} \prod_{1 \leq i < j \leq N} \mathbb{E}_{\mathcal{G}(N,p)} \left[\exp \left(\frac{\beta}{Np} g_{ij} \mathbb{1}_{x_i = x_j} \right) \right] \\
&\leq \sum_{\mathbf{x} \in [q]^N} e^{\sum_{i=1}^N \sum_{r=1}^q B_r \mathbb{1}_{x_i = r}} \prod_{1 \leq i < j \leq N} \exp \left(\frac{\beta}{N} \mathbb{1}_{x_i = x_j} + \frac{\beta^2}{8N^2 p^2} \right) \\
&\leq \sum_{\mathbf{x} \in [q]^N} e^{\beta^2/(16p^2)} \exp \left(\frac{\beta}{N} \sum_{1 \leq i < j \leq N} \mathbb{1}_{x_i = x_j} + \sum_{i=1}^N \sum_{r=1}^q B_r \mathbb{1}_{x_i = r} \right) \\
&\leq e^{\beta^2/(16p^2)} Z_{\beta, \mathbf{B}}^{\text{CW}}
\end{aligned}$$

where in going from the second to the third line, we used the fact that for a Bernoulli random variable Y with $\mathbb{E}Y = p$,

$$\mathbb{E}e^{t(Y-p)} \leq \exp \left(\frac{t^2}{8} \right),$$

which is a direct consequence of Hoeffding's lemma. It now follows from (56) that:

$$D(\mathbb{P}_{\beta, \mathbf{B}}^{\text{CW}} \times \mathcal{G}(N,p) \parallel \mathbb{P}_{\beta, \mathbf{B}}^{\text{ER}}) \leq \frac{\beta^2}{16p^2}$$

thereby establishing (55) and completing the proof of our claim.

Lemma H.5 coupled with our claim, establishes that the product measure $\nu := \mathbf{m}^N \times \mathcal{G}(N, p)$ is contiguous to the measure $\mathbb{P}_{\beta, \mathbf{B}}^{\text{ER}}$ for every $(\beta, \mathbf{B}) \in \Theta_{\mathbf{m}}$, thereby completing the proof of Theorem 1.5. \square

Appendix D: Proof of Theorem 1.7

In this section, we prove Theorem 1.7.

(a) Fixing $\beta > 0$, it follows from Lemma F.2 part (b) that $\nabla_{\mathbf{B}}^2 \ell_N(\beta, \mathbf{B})$ is negative definite, hence $\mathbf{B} \mapsto \ell_N(\beta, \mathbf{B})$ is strictly concave. Define:

$$A_{4,N} := \{\mathbf{x} \in [q]^N : \text{there exists } r \in [q] \text{ such that for all } i \in [N], x_i \neq r\}. \quad (57)$$

Lemma I.1 part (b) gives that if $\mathbf{X} \in A_{4,N}^c$, then $\ell_N(\beta, \mathbf{B}) \rightarrow -\infty$ as $\|\mathbf{B}\| \rightarrow \infty$. Consequently, we have the existence of a unique global maximizer at some $\mathbf{B} \in \mathbb{R}^{q-1}$, i.e. $\hat{\mathbf{B}}_N$ exists on the event $\{\mathbf{X} \in A_{4,N}^c\}$.

Similarly, fixing $\mathbf{B} \in \mathbb{R}^{q-1}$, on the event $\{U_N(\mathbf{X}) \neq 0\}$, the function $\beta \mapsto \ell_N(\beta, \mathbf{B})$ is strictly concave (see Lemma F.2 part (c)). Define:

$$A_{2,N} := \{\mathbf{x} \in [q]^N : m_{i,x_i}(\mathbf{x}) = \min_{r \in [q]} m_{i,r}(\mathbf{x}) \text{ for all } i \in [N]\} \quad (58)$$

$$A_{3,N} := \{\mathbf{x} \in [q]^N : m_{i,x_i}(\mathbf{x}) = \max_{r \in [q]} m_{i,r}(\mathbf{x}) \text{ for all } i \in [N]\}. \quad (59)$$

Once again, it follows from Lemma I.1 part (a) that if $\mathbf{X} \in A_{2,N}^c \cap A_{3,N}^c$ and $U_N(\mathbf{X}) \neq 0$, then the function $\beta \mapsto \ell_N(\beta, \mathbf{B})$ attains a unique global maximum at some $\beta \in \mathbb{R}$, i.e. $\hat{\beta}_N$ exists on the event $\{\mathbf{X} \in A_{2,N}^c \cap A_{3,N}^c\} \cap \{U_N(\mathbf{X}) \neq 0\}$.

To complete the proof of part (a), noting that (21) implies $\mathbb{P}(U_n(\mathbf{X}) = 0) \rightarrow 0$, it suffices to show that

$$\mathbb{P}(\mathbf{X} \in A_{4,N}) = o(1), \quad \mathbb{P}(\mathbf{X} \in A_{2,N} \cup A_{3,N}) = o(1).$$

Towards showing that $\mathbb{P}(\mathbf{X} \in A_{4,N}) = o(1)$, note that by Lemma F.1 with $b_{itrs} := \mathbb{1}_{t \neq r_0}$ and $g \equiv 1$, we have

$$\sum_{i=1}^N \sum_{t \neq r_0} (\mathbb{1}_{X_i=t} - \theta_{i,t}(\mathbf{X})) = O_{\mathbb{P}}(\sqrt{N}), \quad (60)$$

where $\theta_{i,t}(bm\mathbf{x}) = \mathbb{P}(X_i = t | X_j = x_j, j \neq i)$ as in (4). Assume $\mathbf{X} \in A_{4,N}$. Let $r_0 \in [q]$ be such that $X_i \neq r_0$ for all i . This implies that for all i , $\sum_{s \neq r_0} \mathbb{1}_{X_i=s} = 1$. Hence, we have:

$$\sum_{s \neq r_0} (\mathbb{1}_{X_i=s} - \theta_{i,s}(\mathbf{X})) = 1 - \sum_{s \neq r_0} \theta_{i,s}(\mathbf{X}) \geq q^{-1} \exp(-\beta\gamma - 2\|\mathbf{B}\|_{\infty}).$$

This says that the left side of (60) is $\Omega(N)$ whenever $\mathbf{X} \in A_{4,N}$, thereby showing that $\mathbb{P}(\mathbf{X} \in A_{4,N}) = o(1)$.

Next, towards showing that $\mathbb{P}(\mathbf{X} \in A_{3,N}) = o(1)$, invoking Lemma F.1 applied with $b_{itrs} := \mathbb{1}_{t=r}$, $\lambda := 0$ and $g(x) = x$, a union bound gives:

$$\sum_{i=1}^N \sum_{r=1}^q (\mathbb{1}_{X_i=r} - \theta_{i,r}(\mathbf{X})) m_{i,r}(\mathbf{X}) = O_{\mathbb{P}}(\sqrt{N}). \quad (61)$$

Now, suppose that $\mathbf{X} \in A_{3,N}$. Then, $\sum_{r=1}^q m_{i,r}(\mathbf{X}) \mathbb{1}_{X_i=r} = m_{i,X_i}(\mathbf{X}) = \max_r m_{i,r}(\mathbf{X})$, and hence

$$\begin{aligned} & \sum_{i=1}^N \sum_{r=1}^q (\mathbb{1}_{X_i=r} - \theta_{i,r}(\mathbf{X})) m_{i,r}(\mathbf{X}) \\ &= \sum_{i=1}^N \left(\max_r m_{i,r}(\mathbf{X}) - \sum_r m_{i,r}(\mathbf{X}) \theta_{i,r} \right) \\ &\geq \frac{1}{q^2(q-1)\gamma} \exp(-\beta\gamma - 2\|\mathbf{B}\|_{\infty}) \sum_{i=1}^N \sum_{r<s} (m_{i,r}(\mathbf{X}) - m_{i,s}(\mathbf{X}))^2 \\ &= \frac{1}{q^2(q-1)\gamma} \exp(-\beta\gamma - 2\|\mathbf{B}\|_{\infty}) NU_N(\mathbf{X}) = a_N \Omega_{\mathbb{P}}(\sqrt{N}). \end{aligned}$$

In the above display, the inequality on the third line invokes Lemma I.9 with the choices

$$w_r = \theta_{ir}(\mathbf{X}) \geq q^{-1} \exp(-\beta\gamma - 2\|\mathbf{B}\|_{\infty}) = \alpha, \quad t_r = m_{i,r}(\mathbf{X}) \in [0, \gamma],$$

and the last equality holds for some sequence $a_N \rightarrow \infty$, by (21). The above, combined with (61) now gives $\mathbb{P}(\mathbf{X} \in A_{3,N}) = o(1)$. The proof of the fact $\mathbb{P}(\mathbf{X} \in A_{2,N}) = o(1)$ follows similarly, and we skip the details.

(b) The proof will be carried out by an application of Proposition E.1 with $w_N(\mathbf{B}) = \nabla_{\mathbf{B}} \ell_N(\beta, \mathbf{B})$. To begin with, noting the expression of $\nabla_{\mathbf{B}} \ell_N(\beta, \mathbf{B})$ (see (7)), it follows from Lemma F.1 that $\|\nabla_{\mathbf{B}} \ell_N(\beta, \mathbf{B})\|_2 = O_{\mathbb{P}}(\sqrt{N})$. Consequently, Assumption (i) of Proposition E.1 follows with $a_N = \sqrt{N}$. Assumption (ii) of Proposition E.1 follows from Lemma F.2 (b) with $h_N(\mathbf{X}) = N$. This completes the proof of part (b).

(c) The proof of this part is similar to the proof of part (b), so we skip it.

(d) Define the set

$$G_N(\delta) := \{\mathbf{x} \in [q]^N : U_N(\mathbf{x}) \leq \delta\}.$$

Note that for all $\mathbf{x} \in G_N(\delta)$, we have by the Cauchy-Schwarz inequality,

$$\left| \frac{\beta}{2} \sum_{i=1}^N \sum_{r=1}^q x_{i,r} (m_{i,r}(\mathbf{x}) - \bar{m}_i(\mathbf{x})) \right| \leq \frac{\beta}{2} \sqrt{qN} \sqrt{\frac{NU_N(\mathbf{X})}{q}} \leq \frac{\beta N}{2} \sqrt{\delta},$$

i.e.

$$\left| \frac{\beta}{2} \sum_{i=1}^N \sum_{r=1}^q x_{i,r} m_{i,r}(\mathbf{x}) - \frac{\beta N}{2q} \bar{R} \right| \leq \frac{\beta N}{2} \sqrt{\delta}$$

where $\bar{m}_i(\mathbf{x}) := q^{-1} \sum_{r=1}^q m_{i,r}(\mathbf{x})$. Therefore, we have:

$$\begin{aligned} & \mathbb{P}_{\beta, \mathbf{B}}(\mathbf{X} \in G_N(\delta)) \\ &= \sum_{\mathbf{x} \in G_N(\delta)} \frac{\exp \left\{ \frac{\beta}{2} \sum_{i=1}^N \sum_{r=1}^q x_{i,r} m_{i,r}(\mathbf{x}) + \sum_{i=1}^N \sum_{r=1}^q B_r x_{i,r} \right\}}{Z_N(\beta, \mathbf{B})} \\ &\leq \frac{\exp \left\{ \frac{\beta N}{2} \left(\frac{\bar{R}}{q} + \sqrt{\delta} \right) \right\}}{Z_N(\beta, \mathbf{B})} \sum_{\mathbf{x} \in G_N(\delta)} \exp \left\{ \sum_{i=1}^N \sum_{r=1}^q B_r x_{i,r} \right\} \end{aligned}$$

where $R_i = \sum_{j=1}^N a_{ij}$ and $\bar{R} = \frac{1}{N} \sum_{i=1}^N R_i$. Now, note that:

$$\sum_{\mathbf{x} \in [q]^n} \exp \left\{ \sum_{r=1}^q \sum_{i=1}^N B_r x_{i,r} \right\} = \left(\sum_{x \in [q]} \exp \left\{ \sum_{r=1}^q B_r \mathbb{1}_{x=r} \right\} \right)^N = \left(\sum_{r=1}^q e^{B_r} \right)^N.$$

Hence, we have:

$$\mathbb{P}_{\beta, \mathbf{B}}(\mathbf{X} \in G_N(\delta)) \leq \frac{1}{Z_N(\beta, \mathbf{B})} \exp \left\{ \frac{\beta N}{2} \left(\frac{\bar{R}}{q} + \sqrt{\delta} \right) + N \log \left(\sum_{r=1}^q e^{B_r} \right) \right\}, \quad \text{i.e.}$$

$$\frac{1}{N} \log \mathbb{P}_{\beta, \mathbf{B}}(\mathbf{X} \in G_N(\delta)) \leq \frac{\beta}{2} \left(\frac{\bar{R}}{q} + \sqrt{\delta} \right) + \log \left(1 + \sum_{r=1}^{q-1} e^{B_r} \right) - \frac{\log Z_N(\beta, \mathbf{B})}{N} \quad (62)$$

Now, it follows from (1.8) and (1.9) in [Basak and Mukherjee \(2017\)](#) (by choosing $q_i(r) = t_r$) that:

$$\frac{\log Z_N(\beta, \mathbf{B})}{N} \geq \sup_{\mathbf{t} \in \mathcal{P}([q])} \left\{ \frac{\beta}{2} \bar{R} \sum_{r=1}^q t_r^2 + \sum_{r=1}^q B_r t_r - \sum_{r=1}^q t_r \log t_r \right\} \quad (63)$$

We now divide the proof into two cases.

Case-1: $\mathbf{B} \neq \mathbf{0}$. In this case, choosing

$$t_s = \frac{e^{B_s}}{\sum_{r=1}^q e^{B_r}} \quad (\text{for } s \in [q]),$$

Then, $\mathbf{t} := (t_1, t_2, \dots, t_q) \in \mathcal{P}([q])$, we have:

$$\begin{aligned} & \sum_{r=1}^q B_r t_r - \sum_{r=1}^q t_r \log t_r \\ &= \frac{\sum_{r=1}^q B_r e^{B_r}}{\sum_{s=1}^q e^{B_s}} - \sum_{r=1}^q \frac{e^{B_r}}{\sum_{s=1}^q e^{B_s}} \log \left(\frac{e^{B_r}}{\sum_{s=1}^q e^{B_s}} \right) \end{aligned}$$

$$\begin{aligned}
&= \sum_{r=1}^q \frac{e^{B_r}}{\sum_{s=1}^q e^{B_s}} \log \left(\sum_{s=1}^q e^{B_s} \right) \\
&= \log \left(\sum_{s=1}^q e^{B_s} \right).
\end{aligned}$$

Using (63) for this choice of \mathbf{t} gives

$$\frac{\log Z_N(\beta, \mathbf{B})}{N} \geq \frac{\beta}{2} \bar{R} \sum_{r=1}^q t_r^2 + \log \left(\sum_{r=1}^q e^{B_r} \right) > \frac{\beta}{2q} \bar{R} + \log \left(\sum_{r=1}^q e^{B_r} \right),$$

where the last inequality uses the fact that $\mathbf{t} \neq (q^{-1}, \dots, q^{-1})$ for $\mathbf{B} \neq \mathbf{0}$. Along with (62), this gives the existence of $\delta > 0$ such that

$$\limsup_{N \rightarrow \infty} \frac{1}{N} \log \mathbb{P}_{\beta, \mathbf{B}}(\mathbf{X} \in G_N(\delta)) < 0$$

thereby establishing part (d) for the case $\mathbf{B} \neq \mathbf{0}$.

Case-2: $\mathbf{B} = \mathbf{0}, \beta > \beta_c(q)$. In this case, it follows from Theorem 2.3 (iv) in [Gandolfo, Ruiz and Marc \(2010\)](#) that the constant vector $\frac{1}{q} \mathbf{1}_q$ is not a maximizer of the function in the RHS of (63)). With \mathbf{t} denoting a maximizer, using (63) we again have

$$\frac{\log Z_N(\beta, \mathbf{B})}{N} \geq \frac{\beta}{2} \left(\sum_{r=1}^q t_r^2 \right) \bar{R} + \log \left(\sum_{r=1}^q e^{B_r} \right) > \frac{\beta}{2q} \bar{R} + \log \left(\sum_{r=1}^q e^{B_r} \right).$$

As before, this along with (63)) gives the existence of $\delta > 0$ such that

$$\limsup_{N \rightarrow \infty} \frac{1}{N} \log \mathbb{P}_{\beta, \mathbf{B}}(\mathbf{X} \in G_N(\delta)) < 0,$$

thereby completing the proof of part (d).

(e) Finally, the non-existence of consistent estimators for the model $\mathbb{P}_{\beta, \mathbf{0}}^{\text{ER}}$ for $\beta < \beta_c$ follows from the contiguity of $\frac{1}{q} \mathbf{1}_q \times \mathcal{G}(N, p)$ to the measure $\mathbb{P}_{\beta, \mathbf{0}}^{\text{ER}}$ from Theorem 1.5, on noting that $(0, \beta_c) \times \{\mathbf{0}_{q-1}\} \subseteq \Theta_{(\frac{1}{q}, \dots, \frac{1}{q})}$. Uniqueness of the global maximizer follows from ([Gandolfo, Ruiz and Marc, 2010](#), Thm 2.3 (iii)), and positive definiteness follows from ([Gandolfo, Ruiz and Marc, 2010](#), Lemma 4.7 (i)).

Appendix E: Convergence of Z-estimators

In this section, we prove a general result about convergence of Z-estimators of the true parameter $\boldsymbol{\tau}_0$ in a parametric family $(\mathbb{P}_{\boldsymbol{\tau}})_{\boldsymbol{\tau} \in \Theta \subseteq \mathbb{R}^d}$ for some $d \geq 1$. To begin with, suppose that $\hat{\boldsymbol{\tau}}_N := \hat{\boldsymbol{\tau}}_N(\mathbf{X}) \in \mathbb{R}^d$ is an approximate root of a function $w_N : \mathbb{R}^d \times [q]^N \rightarrow \mathbb{R}^d$, i.e. the following is satisfied:

$$w_N(\hat{\boldsymbol{\tau}}_N, \mathbf{X}) = 0.$$

Proposition E.1. *Suppose that the objective function w satisfies the following two conditions:*

- (i) $\|w_N(\boldsymbol{\tau}_0, \mathbf{X})\|_2 = O_P(a_N)$ for some non-negative sequence a_N , where $\boldsymbol{\tau}_0 \in \Theta$ is the true parameter,
- (ii) There exists a function $c : \Theta \rightarrow \mathbb{R}$ which is continuous and strictly positive at $\boldsymbol{\tau}_0$, and a non-negative function $h_N : [q]^N \mapsto [0, \infty)$ on the sample space, satisfying:

$$\lambda_{\min}(-\nabla w_N(\boldsymbol{\tau})) \geq c(\boldsymbol{\tau})h_N(\mathbf{X})$$

for all $\boldsymbol{\tau} \in \Theta$.

Then, as long as $a_N = o_P(h_N(\mathbf{X}))$, we have:

$$\|\hat{\boldsymbol{\tau}}_N - \boldsymbol{\tau}_0\|_2 = O_{\mathbb{P}_{\boldsymbol{\tau}_0}}\left(\frac{a_N}{h_N(\mathbf{X})}\right).$$

Proof. For each $t \in [0, 1]$, define

$$\boldsymbol{\tau}_t := t\hat{\boldsymbol{\tau}}_N + (1-t)\boldsymbol{\tau}_0$$

and introduce the following function $g_N : [0, 1] \rightarrow \mathbb{R}$:

$$g_N(t) := (\hat{\boldsymbol{\tau}}_N - \boldsymbol{\tau}_0)^\top w_N(\boldsymbol{\tau}_t, \mathbf{X}).$$

Then we have

$$g'_N(t) = (\hat{\boldsymbol{\tau}}_N - \boldsymbol{\tau}_0)^\top \nabla w_N(\boldsymbol{\tau}_t, \mathbf{X})(\hat{\boldsymbol{\tau}}_N - \boldsymbol{\tau}_0).$$

Setting $Y_N := \|\hat{\boldsymbol{\tau}}_N - \boldsymbol{\tau}_0\|_2$, using assumption (ii) along with the above display gives

$$g'_N(t) \leq -c(\boldsymbol{\tau}_t)h_N(\mathbf{X})\|\hat{\boldsymbol{\tau}}_N - \boldsymbol{\tau}_0\|_2^2 = -c(\boldsymbol{\tau}_t)h_N(\mathbf{X})Y_N^2.$$

Since c is continuous at $\boldsymbol{\tau}_0$ and $c(\boldsymbol{\tau}_0) > 0$, there exists $r > 0$ such that

$$\inf_{\|\boldsymbol{\tau} - \boldsymbol{\tau}_0\| \leq r} c(\boldsymbol{\tau}) \geq \frac{c(\boldsymbol{\tau}_0)}{2} > 0.$$

For $t \leq \frac{r}{Y_N}$, we have:

$$\|\boldsymbol{\tau}_t - \boldsymbol{\tau}_0\|_2 = tY_N \leq r,$$

and consequently

$$g'_N(t) \leq -\frac{c(\boldsymbol{\tau}_0)}{2}h_N(\mathbf{X})Y_N^2.$$

Using the above bound gives

$$\begin{aligned} |g_N(1) - g_N(0)| &= -\int_0^1 g'_N(t) dt \\ &\geq -\int_0^{\min\{1, r/Y_N\}} g'_N(t) dt \end{aligned}$$

$$\begin{aligned}
&\geq \int_0^{\min\{1, r/Y_N\}} \frac{c(\boldsymbol{\tau}_0)}{2} h_N(\mathbf{X}) Y_N^2 dt \\
&= \frac{c(\boldsymbol{\tau}_0)}{2} h_N(\mathbf{X}) Y_N^2 \min\left\{1, \frac{r}{Y_N}\right\} \\
&= \frac{c(\boldsymbol{\tau}_0)}{2} h_N(\mathbf{X}) Y_N \min\{Y_N, r\}.
\end{aligned}$$

On the other hand, using the definition of $g_N(\cdot)$ gives

$$|g_N(1) - g_N(0)| = |(\hat{\boldsymbol{\tau}}_N - \boldsymbol{\tau})^\top w_N(\boldsymbol{\tau}_0, \mathbf{X})| \leq Y_N \|w_N(\boldsymbol{\tau}_0, \mathbf{X})\|_2 = O_P(a_N Y_N).$$

where the last equality uses assumption (i). Combining the above two displays gives

$$\frac{c(\boldsymbol{\tau}_0)}{2} h_N(\mathbf{X}) Y_N \min\{Y_N, r\} = O_P(a_N Y_N),$$

which implies that

$$\min\{Y_N, r\} = O_{\mathbb{P}}\left(\frac{a_N}{h_N(\mathbf{X})}\right).$$

Proposition E.1 now follows from the fact that $a_N/h_N(\mathbf{X}) \xrightarrow{P} 0$ and $r > 0$ is fixed. \square

Appendix F: Necessary tools for analyzing the derivatives of the log pseudo-likelihood

The goal of this section is to develop necessary tools for bounding the first and second derivatives of the log pseudolikelihood, which (in view of Proposition E.1) is crucial in establishing consistency and rates of convergence of the MPL estimators. The first step towards doing this, is to bound the L^2 -norm of the gradient of the log pseudo-likelihood with high probability. This can be achieved via the following general concentration inequality for sums of the random variables $\mathbb{1}_{X_i=t}$ centered by their conditional means given $(X_j)_{j \neq i}$, which actually holds at all temperatures.

Lemma F.1. *For every constant $M > 0$ (not depending on N) and every differentiable function $g : [-M, M] \rightarrow \mathbb{R}$ such that g' is bounded, there exists a constant $C > 0$ (depending only on $g, \beta, \mathbf{B}, q, \gamma$) such that for every $t > 0$, $\lambda \in [0, 1]$, every array $(b_{itrs})_{i \in [N], t, r, s \in [q]} \in \mathbb{R}^{Nq}$ and every $r, s \in [q]$, we have:*

$$\mathbb{P}\left(\left|\sum_{i=1}^N \sum_{t=1}^q b_{itrs} (\mathbb{1}_{X_i=t} - \theta_{i,t}(\mathbf{X})) g(\bar{m}_i^{r,s}(\mathbf{X}))\right| \geq \sqrt{Nt}\right) \leq 2 \exp\left(-\frac{CtN}{\|\mathbf{L}\|_2^2}\right) \quad (64)$$

where $\bar{m}_i^{r,s}(\mathbf{X}) := m_{i,r}(\mathbf{X}) - \lambda m_{i,s}(\mathbf{X})$, $L_i = L_{i,r,s} := \sum_{t=1}^q |b_{itrs}|$, $\mathbf{L} := (L_1, \dots, L_N)^\top$, and $\theta_{i,t}(\mathbf{x}) = \mathbb{P}(X_i = t | X_j = x_j, j \neq i)$ as in (4).

Proof. Fix $r, s \in [q]$, $c \in F$ and define

$$G(\mathbf{X}) := \sum_{i=1}^N \sum_{t=1}^q b_{itrs} (\mathbb{1}_{X_i=t} - \theta_{i,t}(\mathbf{X})) g(\bar{m}_i^{r,s}(\mathbf{X})).$$

We now construct an exchangeable pair $(\mathbf{X}, \mathbf{X}')$ in the following way:

Let U be a discrete uniform random variable on $[N]$ independent of \mathbf{X} , and conditioned on $U = i$, we simulate X'_i from the conditional distribution of X_i given $(X_j)_{j \neq i}$. Then, we replace the i^{th} entry of \mathbf{X} by X'_i and call the resulting vector \mathbf{X}' . Then it is easy to check that $(\mathbf{X}, \mathbf{X}')$ is an exchangeable pair, i.e. $(\mathbf{X}, \mathbf{X}') \stackrel{D}{=} (\mathbf{X}', \mathbf{X})$. Define:

$$F(\mathbf{x}, \mathbf{x}') := \frac{1}{2} \sum_{i=1}^N \sum_{t=1}^q b_{itrs} (g(\bar{m}_i^{r,s}(\mathbf{x})) + g(\bar{m}_i^{r,s}(\mathbf{x}')))(\mathbb{1}_{x_i=t} - \mathbb{1}_{x'_i=t}).$$

Plugging in $\mathbf{x} = \mathbf{X}$, $\mathbf{x}' = \mathbf{X}'$ we get

$$F(\mathbf{X}, \mathbf{X}') = g(\bar{m}_U^{r,s}(\mathbf{X})) \sum_{t=1}^q b_{Utrs} (\mathbb{1}_{X_U=t} - \mathbb{1}_{X'_U=t}), \quad (65)$$

which in turn gives

$$\mathbb{E}[F(\mathbf{X}, \mathbf{X}') | \mathbf{X}] = \frac{1}{N} \sum_{i=1}^N \sum_{t=1}^q b_{itrs} g(\bar{m}_i^{r,s}(\mathbf{X})) (\mathbb{1}_{X_i=t} - \theta_{i,t}(\mathbf{X})) = \frac{G(\mathbf{X})}{N}.$$

Since $(\mathbf{X}, \mathbf{X}')$ is an exchangeable pair, we have $\mathbb{E}G(\mathbf{X}) = \mathbb{E}F(\mathbf{X}, \mathbf{X}') = 0$. Setting

$$\Delta(\mathbf{X}) := \frac{1}{2N} \mathbb{E} \left(|(G(\mathbf{X}) - G(\mathbf{X}'))F(\mathbf{X}, \mathbf{X}')| \middle| \mathbf{X} \right),$$

if we can show the existence of $K > 0$ such that

$$\Delta(\mathbf{X}) \leq \frac{K}{N^2} \|\mathbf{L}\|_2^2, \quad (66)$$

then (Chatterjee, 2007b, Thm 1.5) gives, for $t \geq 0$,

$$\mathbb{P} \left(\frac{|G(\mathbf{X})|}{N} \geq t \right) \leq 2 \exp \left\{ -\frac{N^2 t^2}{2K \|\mathbf{L}\|_2^2} \right\}$$

which further implies that

$$\mathbb{P}(|G(\mathbf{X})| \geq \sqrt{Nt}) \leq 2 \exp \left\{ -\frac{Nt}{2K \|\mathbf{L}\|_2^2} \right\}$$

thereby establishing (64). Therefore, all that remains to complete the proof of (64), is to show (66). Towards this, for every $1 \leq i \leq N$, $u \in [q]$ and $\mathbf{x} \in [q]^N$, define:

$$\mathbf{x}_u^{(i)} := (x_1, \dots, x_{i-1}, u, x_{i+1}, \dots, x_N)^\top.$$

Then using the definition of $\Delta(\mathbf{X})$ we have:

$$\begin{aligned} 2N\Delta(\mathbf{X}) &= \frac{1}{N} \sum_{i=1}^N \sum_{u=1}^q |(G(\mathbf{X}) - G(\mathbf{X}_u^{(i)}))F(\mathbf{X}, \mathbf{X}_u^{(i)})|\theta_{i,t}(\mathbf{x}) \\ &\lesssim \frac{1}{N} \sum_{i=1}^N L_i \sum_{u=1}^q |(G(\mathbf{X}) - G(\mathbf{X}_u^{(i)})), \end{aligned}$$

where in the last inequality we use (65) to get

$$|F(\mathbf{X}, \mathbf{X}_u^{(i)})| \leq \|g\|_\infty \sum_{t=1}^q |b_{itrs}| = \|g\|_\infty L_i.$$

Now, for any $\mathbf{X}, \mathbf{Y} \in [q]^n$, using the definition of $G(\cdot)$, we have:

$$\begin{aligned} &|G(\mathbf{X}) - G(\mathbf{Y})| \\ &\leq \sum_{j=1}^N \sum_{t=1}^q |b_{jtrs}| |g(\bar{m}_j^{r,s}(\mathbf{X}))X_{j,t} - g(\bar{m}_j^{r,s}(\mathbf{Y}))Y_{j,t}| \\ &+ \sum_{j=1}^N \sum_{t=1}^q |b_{jtrs}| |\theta_{j,t}(\mathbf{X})g(\bar{m}_j^{r,s}(\mathbf{X})) - \theta_{j,t}(\mathbf{Y})g(\bar{m}_j^{r,s}(\mathbf{Y}))| \\ &\leq \sum_{j=1}^N \sum_{t=1}^q g(\bar{m}_j^{r,s}(\mathbf{X}))|b_{jtrs}| |X_{j,t} - Y_{j,t}| + \sum_{j=1}^N \sum_{t=1}^q Y_{j,t}|b_{jtrs}| |g(\bar{m}_j^{r,s}(\mathbf{X})) - g(\bar{m}_j^{r,s}(\mathbf{Y}))| \\ &+ \sum_{j=1}^N \sum_{t=1}^q |b_{jtrs}| |\theta_{j,t}(\mathbf{X})g(\bar{m}_j^{r,s}(\mathbf{X})) - \theta_{j,t}(\mathbf{Y})g(\bar{m}_j^{r,s}(\mathbf{Y}))| \\ &\leq \|g\|_\infty \sum_{j=1}^N \sum_{t=1}^q |b_{jtrs}| |X_{j,t} - Y_{j,t}| + \|g'\|_\infty \sum_{j=1}^N \sum_{t=1}^q Y_{j,t}|b_{jtrs}| |\bar{m}_j^{r,s}(\mathbf{X}) - \bar{m}_j^{r,s}(\mathbf{Y})| \\ &+ \sum_{j=1}^N \sum_{t=1}^q |b_{jtrs}| |\theta_{j,t}(\mathbf{X})g(\bar{m}_j^{r,s}(\mathbf{X})) - \theta_{j,t}(\mathbf{Y})g(\bar{m}_j^{r,s}(\mathbf{Y}))| \end{aligned} \quad (67)$$

We now bound each of the terms in the RHS of (67), for the special choice $\mathbf{Y} = \mathbf{X}_u^{(i)}$. In this case, noting that $X_{j,t} = Y_{j,t}$ for all $j \neq i$, the first term in the RHS of (67) can be bounded as follows:

$$\sum_{j=1}^N \sum_{t=1}^q |b_{jtrs}| |X_{j,t} - Y_{j,t}| = \sum_{t=1}^q |b_{itrs}| |X_{i,t} - Y_{i,t}| \leq \sum_{t=1}^q |b_{itrs}| = L_i.$$

For bounding the second term in the RHS of (67), recalling that $\bar{m}_i^{r,s}(\mathbf{X}) = m_i^r(\mathbf{X}) - \lambda m_i^s(\mathbf{X})$ we get

$$\sum_{j=1}^N \sum_{t=1}^q Y_{j,t}|b_{jtrs}| |\bar{m}_j^{r,s}(\mathbf{X}) - \bar{m}_j^{r,s}(\mathbf{Y})| = \sum_{j=1}^N \sum_{k=1}^N \sum_{t=1}^q Y_{j,t} a_{jk} |b_{jtrs}| |X_{kr} - Y_{kr} - \lambda X_{ks} + \lambda Y_{ks}|$$

$$\begin{aligned}
&= \sum_{j=1}^N \sum_{t=1}^q a_{ji} |b_{jtrs}| |X_{ir} - Y_{ir} - \lambda X_{is} + \lambda Y_{is}| \\
&\leq (1 + \lambda) \sum_{j=1}^N a_{ji} L_j = (1 + \lambda) A_{i*} \mathbf{L},
\end{aligned}$$

where A_{i*} denotes the i^{th} row of \mathbf{A} . Proceeding to bound the third term in the RHS of (67), define the function

$$\psi_t(\alpha_1, \dots, \alpha_q) := g(\alpha_r - \lambda \alpha_s) \frac{\exp\{\beta \alpha_t + B_t\}}{\sum_{s=1}^q \exp\{\beta \alpha_s + B_s\}},$$

and note that $\|\nabla \psi_t\|_\infty$ is bounded, since g has a bounded derivative. Using this definition we can write $\theta_{j,t}(\mathbf{X})g(\bar{m}_j^{r,s}(\mathbf{X})) = \psi_t(m_{j,1}(\mathbf{X}), \dots, m_{j,q}(\mathbf{X}))$, and hence mean-value theorem gives

$$|\theta_{j,t}(\mathbf{X})g(\bar{m}_j^{r,s}(\mathbf{X})) - \theta_{j,t}(\mathbf{Y})g(\bar{m}_j^{r,s}(\mathbf{Y}))| \lesssim \sum_{w=1}^q |m_{j,w}(\mathbf{X}) - m_{j,w}(\mathbf{Y})|.$$

Using this, the the third term in the RHS of (67) can be bounded as follows:

$$\begin{aligned}
&\sum_{j=1}^N \sum_{t=1}^q |b_{jtrs}| |\theta_{j,t}(\mathbf{X})g(\bar{m}_j^{r,s}(\mathbf{X})) - \theta_{j,t}(\mathbf{Y})g(\bar{m}_j^{r,s}(\mathbf{Y}))| \\
&\lesssim \sum_{w=1}^q \sum_{j=1}^N \sum_{t=1}^q |b_{jtrs}| |m_{j,w}(\mathbf{X}) - m_{j,w}(\mathbf{Y})| \\
&= \sum_{w=1}^q \sum_{j=1}^N \sum_{t=1}^q a_{ji} |b_{jtrs}| |X_{i,w} - Y_{i,w}| \\
&\leq \sum_{w=1}^q \sum_{j=1}^N \sum_{t=1}^q a_{ji} |b_{jtrs}| = q \sum_{j=1}^N a_{ji} L_j = q A_{i*} \mathbf{L}.
\end{aligned}$$

Combining the last three bounds to the RHS of (67) we get

$$|G(\mathbf{X}) - G(\mathbf{X}_u^{(i)})| \lesssim L_i + A_{i*} \mathbf{L},$$

and consequently,

$$\begin{aligned}
2N\Delta(\mathbf{X}) &\lesssim \frac{1}{N} \sum_{i=1}^N L_i (L_i + A_{i*} \mathbf{L}) = \frac{1}{N} (\|\mathbf{L}\|_2^2 + \mathbf{L}^\top \mathbf{A} \mathbf{L}) \\
&\leq \frac{1}{N} (1 + \|\mathbf{A}\|_2) \|\mathbf{L}\|_2^2 \leq \frac{1}{N} (1 + \gamma) \|\mathbf{L}\|_2^2.
\end{aligned}$$

This establishes (66), and consequently, (64). \square

The second crucial step towards establishing consistency and rates of convergence of the MPL estimators is to provide a lower bound on the minimum eigenvalue of the negative Hessian of the log pseudo-likelihood function. The following lemma achieves this.

Lemma F.2. *There exists a continuous strictly positive function $C_{q,\gamma}$ on $[0, \infty) \times \mathbb{R}^{q-1}$, such that:*

- (a) $\lambda_{\min}(-\nabla^2 \ell_N(\beta, \mathbf{B})) \geq C_{q,\gamma}(\beta, \mathbf{B})NT_N(\mathbf{X})$.
- (b) $\lambda_{\min}(-\nabla_{\mathbf{B}}^2 \ell_N(\beta, \mathbf{B})) \geq C_{q,\gamma}(\beta, \mathbf{B})N$ and
- (c) $-\frac{\partial^2 \ell_N(\beta, \mathbf{B})}{\partial \beta^2} \geq C_{q,\gamma}(\beta, \mathbf{B})NU_N(\mathbf{X})$

where T_N and U_N are as defined in (10) and (20), respectively.

Proof. It follows from (6), (7) and a simple calculation, that the second-order partial derivatives of ℓ_N are given by:

$$\begin{aligned} \frac{\partial^2 \ell_N(\beta, \mathbf{B})}{\partial \beta^2} &= -\sum_{i=1}^N \sum_{1 \leq a < b \leq q} (m_{i,a}(\mathbf{X}) - m_{i,b}(\mathbf{X}))^2 \theta_{i,a}(\mathbf{X}) \theta_{i,b}(\mathbf{X}), \\ \frac{\partial^2 \ell_N(\beta, \mathbf{B})}{\partial B_s \partial \beta} &= -\sum_{i=1}^N \sum_{a=1}^q (m_{i,s}(\mathbf{X}) - m_{i,a}(\mathbf{X})) \theta_{i,a}(\mathbf{X}) \theta_{i,s}(\mathbf{X}) \quad (1 \leq s \leq q-1), \\ \frac{\partial^2 \ell_N(\beta, \mathbf{B})}{\partial B_s^2} &= -\sum_{i=1}^N \sum_{a \neq s} \theta_{i,s}(\mathbf{X}) \theta_{i,a}(\mathbf{X}) \quad (1 \leq s \leq q-1), \\ \frac{\partial^2 \ell_N(\beta, \mathbf{B})}{\partial B_r \partial B_s} &= \sum_{i=1}^N \theta_{i,r}(\mathbf{X}) \theta_{i,s}(\mathbf{X}) \quad (1 \leq r \neq s \leq q-1) \end{aligned}$$

Here $\theta_{i,t}(\mathbf{x}) = \mathbb{P}(X_i = t | X_j = x_j, j \neq i)$ is as in (4). Using the above expressions, for any $\mathbf{y} := (y_0, y_1, \dots, y_{q-1}) \in \mathbb{R}^q$, we have:

$$\begin{aligned} & -\mathbf{y}^\top \nabla^2 \ell_N(\beta, \mathbf{B}) \mathbf{y} \\ &= \sum_{i=1}^N \left[y_0^2 \sum_{1 \leq a < b \leq q} (m_{i,a}(\mathbf{X}) - m_{i,b}(\mathbf{X}))^2 \theta_{i,a}(\mathbf{X}) \theta_{i,b}(\mathbf{X}) \right. \\ & \quad + 2y_0 \sum_{s=1}^{q-1} y_s \sum_{a=1}^q (m_{i,s}(\mathbf{X}) - m_{i,a}(\mathbf{X})) \theta_{i,a}(\mathbf{X}) \theta_{i,s}(\mathbf{X}) \\ & \quad \left. + \sum_{s=1}^{q-1} y_s^2 \theta_{i,s}(\mathbf{X}) \sum_{a \neq s} \theta_{i,a}(\mathbf{X}) - 2 \sum_{1 \leq r < s \leq q-1} y_r y_s \theta_{i,r}(\mathbf{X}) \theta_{i,s}(\mathbf{X}) \right] \quad (68) \end{aligned}$$

The first term in the RHS of (68) (without y_0^2) can be simplified as follows:

$$\sum_{1 \leq a < b \leq q} (m_{i,a}(\mathbf{X}) - m_{i,b}(\mathbf{X}))^2 \theta_{i,a}(\mathbf{X}) \theta_{i,b}(\mathbf{X})$$

$$\begin{aligned}
&= \sum_{1 \leq r < s \leq q-1} (m_{i,r}(\mathbf{X}) - m_{i,s}(\mathbf{X}))^2 \theta_{i,r}(\mathbf{X}) \theta_{i,s}(\mathbf{X}) \\
&+ \sum_{r=1}^{q-1} (m_{i,r}(\mathbf{X}) - m_{i,q}(\mathbf{X}))^2 \theta_{i,r}(\mathbf{X}) \theta_{i,q}(\mathbf{X})
\end{aligned}$$

The second term in the RHS of (68) (without $2y_0$) can be simplified as follows:

$$\begin{aligned}
&\sum_{s=1}^{q-1} y_s \sum_{a=1}^q (m_{i,s}(\mathbf{X}) - m_{i,a}(\mathbf{X})) \theta_{i,a}(\mathbf{X}) \theta_{i,s}(\mathbf{X}) \\
&= \sum_{s,r=1}^{q-1} (m_{i,s}(\mathbf{X}) - m_{i,r}(\mathbf{X})) y_s \theta_{i,s}(\mathbf{X}) \theta_{i,r}(\mathbf{X}) \\
&+ \sum_{s=1}^{q-1} (m_{i,s}(\mathbf{X}) - m_{i,q}(\mathbf{X})) y_s \theta_{i,s}(\mathbf{X}) \theta_{i,q}(\mathbf{X}) \\
&= \sum_{1 \leq r < s \leq q-1} (m_{i,s}(\mathbf{X}) - m_{i,r}(\mathbf{X})) (y_s - y_r) \theta_{i,s}(\mathbf{X}) \theta_{i,r}(\mathbf{X}) \\
&+ \sum_{s=1}^{q-1} (m_{i,s}(\mathbf{X}) - m_{i,q}(\mathbf{X})) y_s \theta_{i,s}(\mathbf{X}) \theta_{i,q}(\mathbf{X})
\end{aligned}$$

The third term in the RHS of (68) can be simplified as follows:

$$\begin{aligned}
&\sum_{s=1}^{q-1} y_s^2 \theta_{i,s}(\mathbf{X}) \sum_{a \neq s} \theta_{i,a}(\mathbf{X}) \\
&= \sum_{s=1}^{q-1} y_s^2 \theta_{i,s}(\mathbf{X}) \sum_{r=1}^{q-1} \mathbb{1}_{r \neq s}(\mathbf{X}) \theta_{i,r}(\mathbf{X}) + \sum_{s=1}^{q-1} y_s^2 \theta_{i,s}(\mathbf{X}) \theta_{i,q}(\mathbf{X}) \\
&= \sum_{1 \leq r < s \leq q-1} (y_r^2 + y_s^2) \theta_{i,r}(\mathbf{X}) \theta_{i,s}(\mathbf{X}) + \sum_{s=1}^{q-1} y_s^2 \theta_{i,s}(\mathbf{X}) \theta_{i,q}(\mathbf{X}) \\
&= \sum_{1 \leq r < s \leq q-1} (y_r - y_s)^2 \theta_{i,r}(\mathbf{X}) \theta_{i,s}(\mathbf{X}) + 2 \sum_{1 \leq r < s \leq q-1} y_r y_s \theta_{i,r}(\mathbf{X}) \theta_{i,s}(\mathbf{X}) \\
&+ \sum_{s=1}^{q-1} y_s^2 \theta_{i,s}(\mathbf{X}) \theta_{i,q}(\mathbf{X}).
\end{aligned}$$

Using the three displays above the RHS of (68) simplifies to

$$\sum_{i=1}^N \left[y_0^2 \sum_{1 \leq r < s \leq q-1} (m_{i,r}(\mathbf{X}) - m_{i,s}(\mathbf{X}))^2 \theta_{i,r}(\mathbf{X}) \theta_{i,s}(\mathbf{X}) \right]$$

$$\begin{aligned}
& + y_0^2 \sum_{r=1}^{q-1} (m_{i,r}(\mathbf{X}) - m_{i,q}(\mathbf{X}))^2 \theta_{i,r}(\mathbf{X}) \theta_{i,q}(\mathbf{X}) \\
& + 2y_0 \sum_{1 \leq r < s \leq q-1} (m_{i,r}(\mathbf{X}) - m_{i,s}(\mathbf{X})) (y_r - y_s) \theta_{i,r}(\mathbf{X}) \theta_{i,s}(\mathbf{X}) \\
& + 2y_0 \sum_{r=1}^{q-1} (m_{i,r}(\mathbf{X}) - m_{i,q}(\mathbf{X})) y_r \theta_{i,r}(\mathbf{X}) \theta_{i,q}(\mathbf{X}) \\
& + \left[\sum_{1 \leq r < s \leq q-1} (y_r - y_s)^2 \theta_{i,r}(\mathbf{X}) \theta_{i,s}(\mathbf{X}) + \sum_{r=1}^{q-1} y_r^2 \theta_{i,r}(\mathbf{X}) \theta_{i,q}(\mathbf{X}) \right] \\
= & \sum_{i=1}^N \sum_{1 \leq r < s \leq q-1} \{(m_{i,r}(\mathbf{X}) - m_{i,s}(\mathbf{X})) y_0 + y_r - y_s\}^2 \theta_{i,r}(\mathbf{X}) \theta_{i,s}(\mathbf{X}) \\
& + \sum_{i=1}^N \sum_{r=1}^{q-1} \{(m_{i,r}(\mathbf{X}) - m_{i,q}(\mathbf{X})) y_0 + y_r\}^2 \theta_{i,r}(\mathbf{X}) \theta_{i,q}(\mathbf{X}). \quad (69)
\end{aligned}$$

Now, it follows from Condition (8) that $m_{i,r}(\mathbf{X}) \leq \gamma$, and so

$$\theta_{i,r}(\mathbf{X}) \geq \alpha := q^{-1} \exp\{-\beta\gamma - 2\|\mathbf{B}\|_\infty\} > 0.$$

Hence,

$$\begin{aligned}
-\mathbf{y}^\top \nabla^2 \ell_N(\beta, \mathbf{B}) \mathbf{y} & \geq \alpha^2 \sum_{i=1}^N \sum_{1 \leq r < s \leq q-1} \{(m_{i,r}(\mathbf{X}) - m_{i,s}(\mathbf{X})) y_0 + y_r - y_s\}^2 \\
& + \alpha^2 \sum_{i=1}^N \sum_{r=1}^{q-1} \{(m_{i,r}(\mathbf{X}) - m_{i,q}(\mathbf{X})) y_0 + y_r\}^2 \\
& = \alpha^2 \mathbf{y}^\top \mathbf{H}_N \mathbf{y} \quad (70)
\end{aligned}$$

where

$$\mathbf{H}_N := \begin{pmatrix} \sum_i \sum_{1 \leq r < s \leq q} (m_{i,r}(\mathbf{X}) - m_{i,s}(\mathbf{X}))^2 & \sum_i \mathbf{v}_i^\top \\ \sum_i \mathbf{v}_i & N(q\mathbf{I} - \mathbf{J}) \end{pmatrix}$$

with \mathbf{I} and \mathbf{J} being the $(q-1) \times (q-1)$ identity matrix and matrix of all ones, respectively, and

$$\mathbf{v}_i = (qm_{i,1}(\mathbf{X}) - \sum_r m_{i,r}(\mathbf{X}), \dots, qm_{i,q-1}(\mathbf{X}) - \sum_r m_{i,r}(\mathbf{X}))^\top \in \mathbb{R}^{q-1}.$$

The above equality (70) follows from repeating the above calculations, and (69) replacing $\theta_{i,r}(\mathbf{X})$ by 1 throughout. Thus \mathbf{H}_N is non-negative definite, and denoting $\lambda_1 \leq \dots \leq \lambda_q$ to be eigenvalues of \mathbf{H}_N we have:

$$\lambda_1 \operatorname{tr}(\mathbf{H}_N)^{q-1} = \lambda_1 \sum_{1 \leq i_1, \dots, i_{q-1} \leq q} \lambda_{i_1} \dots \lambda_{i_{q-1}} \geq \prod_{i=1}^q \lambda_i = \det(\mathbf{H}_N). \quad (71)$$

Next, using the fact that $(q\mathbf{I} - \mathbf{J})^{-1} = \frac{1}{q}(\mathbf{I} + \mathbf{J})$, along with the expression for the determinant of a block partitioned matrix, we get:

$$\begin{aligned} & \det(\mathbf{H}_N) \\ = & \det(N(q\mathbf{I} - \mathbf{J})) \left[\sum_i \sum_{r < s} (m_{i,r}(\mathbf{X}) - m_{i,s}(\mathbf{X}))^2 - \frac{1}{Nq} \left(\sum_i \mathbf{v}_i \right)^\top (\mathbf{I} + \mathbf{J}) \left(\sum_i \mathbf{v}_i \right) \right]. \end{aligned}$$

Next, on noting that the r^{th} entry of $\sum_i \mathbf{v}_i$ is given by $Nq(\bar{m}_r(\mathbf{X}) - \frac{1}{q} \sum_{s=1}^q \bar{m}_s(\mathbf{X}))$ for $1 \leq r \leq q-1$, we have:

$$\begin{aligned} & \frac{1}{Nq} \left(\sum_{i=1}^N \mathbf{v}_i \right)^\top (\mathbf{I} + \mathbf{J}) \left(\sum_{i=1}^N \mathbf{v}_i \right) \\ = & Nq \left[\sum_{r=1}^{q-1} \left(\bar{m}_r(\mathbf{X}) - \frac{1}{q} \sum_{s=1}^q \bar{m}_s(\mathbf{X}) \right)^2 + \left(\sum_{r=1}^{q-1} \left(\bar{m}_r(\mathbf{X}) - \frac{1}{q} \sum_{s=1}^q \bar{m}_s(\mathbf{X}) \right) \right)^2 \right] \\ = & Nq \left[\sum_{r=1}^{q-1} \left(\bar{m}_r(\mathbf{X}) - \frac{1}{q} \sum_{s=1}^q \bar{m}_s(\mathbf{X}) \right)^2 + \left(-\bar{m}_q(\mathbf{X}) + \frac{1}{q} \sum_{s=1}^q \bar{m}_s(\mathbf{X}) \right)^2 \right] \\ = & Nq \sum_{r=1}^q \left(\bar{m}_r(\mathbf{X}) - \frac{1}{q} \sum_{s=1}^q \bar{m}_s(\mathbf{X}) \right)^2. \end{aligned}$$

Hence, we have:

$$\begin{aligned} & \det(\mathbf{H}_N) \\ = & N^{q-1} q^{q-2} \left[\sum_i \sum_{r < s} (m_{i,r}(\mathbf{X}) - m_{i,s}(\mathbf{X}))^2 - Nq \sum_{r=1}^q \left(\bar{m}_r(\mathbf{X}) - \frac{1}{q} \sum_{s=1}^q \bar{m}_s(\mathbf{X}) \right)^2 \right] \\ = & N^{q-1} q^{q-2} \left[\sum_i \sum_{r < s} (m_{i,r}(\mathbf{X}) - m_{i,s}(\mathbf{X}))^2 - N \sum_{r < s} (\bar{m}_r(\mathbf{X}) - \bar{m}_s(\mathbf{X}))^2 \right] \\ = & N^q q^{q-2} T_N(\mathbf{X}), \end{aligned}$$

where the last equality uses (30). Also, note that:

$$\text{tr}(\mathbf{H}_N) = \sum_i \sum_{r < s} (m_{i,r}(\mathbf{X}) - m_{i,s}(\mathbf{X}))^2 + N(q-1)^2 \leq N \left[\binom{q}{2} \gamma^2 + (q-1)^2 \right].$$

Hence, (71) implies that

$$\lambda_1 \geq C_{q,\gamma} N T_N(\mathbf{X}),$$

where

$$C_{q,\gamma} := \frac{q^{q-2}}{\left(\binom{q}{2} \gamma^2 + (q-1)^2 \right)^{q-1}}.$$

It now follows from (70) that

$$-\mathbf{y}^\top \nabla^2 \ell_N(\beta, \mathbf{B}) \mathbf{y} \geq C_{q,\gamma} \delta^2(\beta, \mathbf{B}, \gamma) \|\mathbf{y}\|_2^2 NT_N(\mathbf{X})$$

for all $\mathbf{y} \in \mathbb{R}^q$, from which part (a) follows on taking \mathbf{y} as an eigenvector of $-\nabla^2 \ell_N(\beta, \mathbf{B})$ corresponding to its minimum eigenvalue. Part (b) follows from (70) on taking $\mathbf{y} = (0, \tilde{\mathbf{y}})$, and noting that the smallest eigenvalue of the $(q-1) \times (q-1)$ matrix $q\mathbf{I} - \mathbf{J}$ is 1. Part (c) follows on taking y in (70) to be the vector $(1, 0, \dots, 0)$ of length q . \square

Appendix G: Proof of Lemma A.1

The proof of Lemma A.1 is based on reducing the Curie-Weiss Potts model to a product measure, by conditioning on a suitable random variable. Towards this, conditional on $\mathbf{X} \sim \mathbb{P}_{\beta, \mathbf{B}}^{\text{CW}}$, let Z_1, \dots, Z_q be independent random variables with $Z_r \sim N(\bar{X}_r, (\beta N)^{-1})$. Define $\mathbf{Z} := (Z_1, \dots, Z_q)$.

Lemma G.1. *If \mathbf{X} follows the Curie-Weiss Potts model $\mathbb{P}_{\beta, \mathbf{B}}^{\text{CW}}$, then the entries of $\mathbf{X}|\mathbf{Z}$ are independent and identically distributed, with*

$$\mathbb{P}_{\beta, \mathbf{B}}^{\text{CW}}(\mathbf{X}_i = r | \mathbf{Z}) = \frac{e^{\beta Z_r + B_r}}{\sum_{s=1}^q e^{\beta Z_s + B_s}}.$$

Proof. Recall from (40) that:

$$\mathbb{P}_{\beta, \mathbf{B}}^{\text{CW}}(\mathbf{X}) \propto \exp \left(\frac{N\beta}{2} \sum_{r=1}^q \bar{X}_r^2 + N \sum_{r=1}^q B_r \bar{X}_r \right).$$

Hence, we have:

$$\begin{aligned} \mathbb{P}_{\beta, \mathbf{B}}^{\text{CW}}(\mathbf{X}, \mathbf{Z}) &\propto \mathbb{P}_{\beta, \mathbf{B}}^{\text{CW}}(\mathbf{X}) \prod_r \mathbb{P}(Z_r | \mathbf{X}) \\ &\propto \exp \left\{ \frac{N\beta}{2} \sum_r (\bar{X}_r^2 - (Z_r - \bar{X}_r)^2) + N \sum_{r=1}^q B_r \bar{X}_r \right\} \\ &= \exp \left\{ -\frac{N\beta}{2} \sum_r Z_r^2 + N\beta \sum_r \bar{X}_r Z_r + N \sum_{r=1}^q B_r \bar{X}_r \right\}. \end{aligned}$$

Therefore,

$$\mathbb{P}_{\beta, \mathbf{B}}^{\text{CW}}(\mathbf{X} | \mathbf{Z}) \propto \exp \left\{ \sum_{i=1}^N \sum_{r=1}^q X_{i,r} (\beta Z_r + B_r) \right\},$$

and so given \mathbf{Z} the random variables (X_1, \dots, X_N) are iid, with

$$\mathbb{P}_{\beta, \mathbf{B}}^{\text{CW}}(\mathbf{X}_i = r | \mathbf{Z}) = \frac{e^{\beta Z_r + B_r}}{\sum_{s=1}^q e^{\beta Z_s + B_s}}.$$

This completes the proof of Lemma G.1. \square

Returning to the proof of Lemma A.1, let us define for every $r \in [q]$ and $\mathbf{x} \in [q]^N$:

$$S_{N,r}(\mathbf{x}) := \sum_{i=1}^N \left(m_{i,r}(\mathbf{x}) - \bar{m}_r(\mathbf{x}) - \frac{1}{q}(R_i - \bar{R}) \right)^2 .$$

For ease of notation, we will abbreviate $S_{N,r}(\mathbf{X})$ by $S_{N,r}$. Then, we have:

$$\begin{aligned} \mathbb{E}(S_{N,r}|\mathbf{Z}) &= \sum_{i=1}^N \mathbb{E} \left(\left(m_{i,r}(\mathbf{X}) - \bar{m}_r(\mathbf{X}) - \frac{1}{q}(R_i - \bar{R}) \right)^2 \middle| \mathbf{Z} \right) \\ &\geq \sum_{i=1}^N \text{Var} \left[m_{i,r}(\mathbf{X}) - \bar{m}_r(\mathbf{X}) - \frac{1}{q}(R_i - \bar{R}) \middle| \mathbf{Z} \right] \\ &= \sum_{i=1}^N \text{Var} \left[\sum_{j=1}^N a_{ij} X_{j,r} - \frac{1}{N} \sum_{j=1}^N R_j X_{j,r} - \frac{1}{q} \sum_{j=1}^N \left(a_{ij} - \frac{R_j}{N} \right) \middle| \mathbf{Z} \right] \\ &= \sum_{i=1}^N \text{Var} \left[\sum_{j=1}^N \left(a_{ij} - \frac{R_j}{N} \right) \left(X_{j,r} - \frac{1}{q} \right) \middle| \mathbf{Z} \right] \\ &= \text{Var}(X_{1,r}|\mathbf{Z}) \sum_{i=1}^N \sum_{j=1}^N \left(a_{ij} - \frac{R_j}{N} \right)^2 \end{aligned}$$

where in going from the second to the third line in the above display, we have used the following identity:

$$\bar{m}_r(\mathbf{X}) := \frac{1}{N} \sum_{i=1}^N \sum_{j=1}^N a_{ij} X_{j,r} = \frac{1}{N} \sum_{j=1}^N X_{j,r} \sum_{i=1}^N a_{ij} = \frac{1}{N} \sum_{j=1}^N R_j X_{j,r} .$$

The RHS above can be bounded below, as follows:

$$\sum_{i=1}^N \sum_{j=1}^N \left(a_{ij} - \frac{R_j}{N} \right)^2 = \sum_{i=1}^N \sum_{j=1}^N a_{ij}^2 - \frac{1}{N} \sum_{i=1}^N R_i^2 \geq \sum_{i=1}^N \sum_{j=1}^N a_{ij}^2 - \gamma^2 = \Omega(N),$$

where the last equality uses the non mean-field condition (13). Hence, we have shown that

$$\mathbb{E}(S_{N,r}|\mathbf{Z}) \geq \Omega(N) \text{Var}(X_{1,r}|\mathbf{Z}) . \quad (72)$$

Next, we will show that for any sequence β_N which is bounded away from 0 and ∞ , the quantity $\text{Var}_{\beta_N, \mathbf{B}}^{\text{CW}}(X_{1,r}|\mathbf{Z})$ is bounded away from 0. For this, it only suffices to show that $\mathbb{P}_{\beta_N, \mathbf{B}}^{\text{CW}}(X_1 = r|\mathbf{Z})$ is bounded away from both 0 and 1. To this effect, Lemma G.1 gives:

$$\mathbb{P}_{\beta_N, \mathbf{B}}^{\text{CW}}(X_1 = r|\mathbf{Z}) = \frac{\exp\{\beta_N Z_r + B_r\}}{\sum_{s=1}^q \exp\{\beta_N Z_s + B_s\}} = \frac{1}{1 + \sum_{s \neq r} \exp\{\beta_N (Z_s - Z_r) + B_s - B_r\}} .$$

Define the event $\mathcal{E} := \{\mathbf{Z} \in [-2, 2]^q\}$, and note that on \mathcal{E} we have:

$$\frac{1}{1 + (q-1) \exp\{4\bar{\beta} + 2\|\mathbf{B}\|_\infty\}} \leq \mathbb{P}_{\beta_N, \mathbf{B}}^{\text{CW}}(X_1 = r | \mathbf{Z}) \leq \frac{1}{1 + (q-1) \exp\{-4\bar{\beta} - 2\|\mathbf{B}\|_\infty\}}.$$

where $\bar{\beta} < \infty$ is an upper bound of β_N . The above bound along with (72) gives:

$$\mathbb{E}_{\beta_N, \mathbf{B}}^{\text{CW}}(S_{N,r} | \mathbf{Z}) \geq CN$$

on the event \mathcal{E} , for some constant $C > 0$ not depending on N .

We will now show, using McDiarmid's inequality, that $S_{N,r}$ concentrates around $\mathbb{E}_{\beta_N, \mathbf{B}}^{\text{CW}}(S_{N,r} | \mathbf{Z})$. For this, fix two vectors \mathbf{x} and \mathbf{x}' in $[q]^N$ that differ exactly in the k^{th} coordinate. Then, we have:

$$\begin{aligned} & |S_{N,r}(\mathbf{x}) - S_{N,r}(\mathbf{x}')| \\ &= \left| \sum_{i=1}^N \left[\left(m_{i,r}(\mathbf{x}) - \bar{m}_r(\mathbf{x}) - \frac{1}{q}(R_i - \bar{R}) \right)^2 - \left(m_{i,r}(\mathbf{x}') - \bar{m}_r(\mathbf{x}') - \frac{1}{q}(R_i - \bar{R}) \right)^2 \right] \right| \\ &\leq 4\gamma \sum_{i=1}^N |m_{i,r}(\mathbf{x}) - \bar{m}_r(\mathbf{x}) - m_{i,r}(\mathbf{x}') + \bar{m}_r(\mathbf{x}')|. \end{aligned}$$

Also we have:

$$\begin{aligned} & \sum_{i=1}^N |m_{i,r}(\mathbf{x}) - \bar{m}_r(\mathbf{x}) - m_{i,r}(\mathbf{x}') + \bar{m}_r(\mathbf{x}')| \\ &= \sum_{i=1}^N \left| a_{ik}(x_{k,r} - x'_{k,r}) - \frac{1}{N} \sum_{\ell=1}^N a_{\ell k}(x_{k,r} - x'_{k,r}) \right| \\ &\leq \sum_{i=1}^N a_{ik} + \sum_{\ell=1}^N a_{\ell k} \leq 2\gamma. \end{aligned}$$

Combining the above two, we thus have:

$$|S_{N,r}(\mathbf{x}) - S_{N,r}(\mathbf{x}')| \leq 8\gamma^2.$$

Hence, by McDiarmid's inequality, we have the following on the event \mathcal{E} :

$$\begin{aligned} \mathbb{P}_{\beta_N, \mathbf{B}}^{\text{CW}} \left(S_{N,r} < \frac{CN}{2} \middle| \mathbf{Z} \right) &\leq \mathbb{P}_{\beta_N, \mathbf{B}}^{\text{CW}} \left(S_{N,r} < \mathbb{E}_{\beta_N, \mathbf{B}}^{\text{CW}}(S_{N,r} | \mathbf{Z}) - \frac{CN}{2} \middle| \mathbf{Z} \right) \\ &\leq \exp \left(-\frac{C^2 N^2}{128 N \gamma^4} \right) = \exp \left(-\frac{NC^2}{128 \gamma^4} \right) \end{aligned}$$

Hence, we have:

$$\mathbb{P}_{\beta_N, \mathbf{B}}^{\text{CW}} \left(S_{N,r} \leq \frac{CN}{2} \right) = \mathbb{E}_{\beta_N, \mathbf{B}}^{\text{CW}} \mathbb{P}_{\beta_N, \mathbf{B}}^{\text{CW}} \left(S_{N,r} \leq \frac{CN}{2} \middle| \mathbf{Z} \right)$$

$$\begin{aligned}
&\leq \mathbb{P}_{\beta_N, \mathbf{B}}(\mathcal{E}^c) + \mathbb{E}_{\beta_N, \mathbf{B}}^{\text{CW}} \mathbb{P}_{\beta_N, \mathbf{B}}^{\text{CW}} \left(S_{N,r} \leq \frac{CN}{2} \mid \mathbf{Z} \right) \mathbb{1}_{\mathcal{E}} \\
&\leq \mathbb{P}_{\beta_N, \mathbf{B}}(\mathcal{E}^c) + \exp\left(-\frac{NC^2}{128\gamma^4}\right).
\end{aligned}$$

Also, we have

$$\begin{aligned}
\mathbb{P}_{\beta_N, \mathbf{B}}^{\text{CW}}(\mathcal{E}^c) &\leq q \max_{1 \leq r \leq q} \mathbb{P}_{\beta_N, \mathbf{B}}(|Z_r| > 2) \\
&= q \max_{1 \leq r \leq q} \mathbb{E}_{\beta_N, \mathbf{B}} \mathbb{P}_{\beta_N, \mathbf{B}}(|Z_r| > 2 \mid \mathbf{X}) \\
&\leq q \max_{1 \leq r \leq q} \mathbb{E}_{\beta_N, \mathbf{B}} \mathbb{P}_{\beta_N, \mathbf{B}}(|Z_r - \bar{X}_r| > 1 \mid \mathbf{X}) \quad (\text{since } |\bar{X}_r| \leq 1) \\
&= q \max_{1 \leq r \leq q} \mathbb{E}_{\beta_N, \mathbf{B}} \mathbb{P}_{\beta_N, \mathbf{B}}\left(\sqrt{\beta_N N} |Z_r - \bar{X}_r| > \sqrt{\beta_N N} \mid \mathbf{X}\right) \\
&= q \mathbb{P}(|Z| > \sqrt{\beta_N N}) \quad (\text{where } Z \sim N(0, 1)) \\
&\leq 2qe^{-\beta N/2},
\end{aligned}$$

where $\underline{\beta} > 0$ is a lower bound of β_N . Combining all these, we conclude:

$$\mathbb{P}_{\beta_N, \mathbf{B}}^{\text{CW}} \left(S_{N,r} \leq \frac{CN}{2} \right) \leq e^{-KN}$$

for some constant $K > 0$ (not depending on N), which invoking (31) gives:

$$\mathbb{P}_{\beta_N, \mathbf{B}}^{\text{CW}} \left(T_N(\mathbf{X}) < \frac{Cq}{2} \right) \leq qe^{-KN} \implies \mathbb{P}_{\beta_N, \mathbf{B}}^{\text{CW}} \left(\mathbf{X} \in E_N \left(\frac{Cq}{2} \right) \right) \leq qe^{-KN},$$

where the set $E_N(\cdot)$ is as in (29).

To complete the proof, it suffices to verify that the sequence $\beta_N = \beta \bar{R}$ is bounded above by $\bar{\beta} < \infty$, and bounded below by $\underline{\beta} > 0$. But this follows on using (8) and (9), and recalling that $\beta > 0$.

Appendix H: Necessary results for proving Theorem 1.5

In this section, we state and prove some lemmas which will be used to verify Theorem 1.5.

Lemma H.1. *Let $\mathbf{X} = (X_1, \dots, X_N) \in [q]^N$ be distributed according to the Curie–Weiss Potts measure $\mathbb{P}_{\beta, \mathbf{B}}^{\text{CW}}$ (40). Let*

$$f_r(\mathbf{t}) = \frac{\exp\{\beta t_r + B_r\}}{\sum_{s=1}^q \exp\{\beta t_s + B_s\}}, \quad \mathbf{t} \in \mathbb{R}^q, \quad r \in [q]$$

be as in (33), and $f : \mathbb{R}^q \rightarrow \mathcal{P}([q])$ (see (17)) as $f(\mathbf{t}) := (f_1(\mathbf{t}), \dots, f_q(\mathbf{t}))$. Then for every $r \in [q]$ and every $t \geq 0$,

$$\mathbb{P}_{\beta, \mathbf{B}}^{\text{CW}} \left(\|\bar{\mathbf{X}} - f(\bar{\mathbf{X}})\|_\infty > \frac{\beta}{2N} + t \right) \leq 2q \exp\left(-\frac{2Nt^2}{2 + \beta}\right).$$

Proof. To begin with, use (4) to note that for all $i \in [N], r \in [q]$ we have:

$$\mathbb{P}(X_i = r \mid \mathbf{X}_{-i}) = f_r(\bar{\mathbf{X}}^{(i)}), \text{ where } \bar{X}_r^{(i)} := \frac{1}{N} \sum_{j \neq i} X_{j,r} \text{ and } \bar{\mathbf{X}}^{(i)} := (\bar{X}_1^{(i)}, \dots, \bar{X}_q^{(i)}). \quad (73)$$

Define \mathbf{X}' from \mathbf{X} by choosing an index I uniformly at random from $[N]$, and updating the I^{th} entry of \mathbf{X} by a sample from the conditional distribution $\mathbb{P}_{\beta, \mathbf{B}}^{\text{CW}}(X_I = \cdot \mid \mathbf{X}_{-I}) = f_r(\bar{\mathbf{X}}^{(I)})$, keeping the other entries unchanged. It is straightforward to verify that $(\mathbf{X}, \mathbf{X}')$ is an exchangeable pair. Fix $r \in [q]$ and define the antisymmetric function

$$F_r(\mathbf{X}, \mathbf{Y}) := N(\bar{X}_r - \bar{Y}_r), \text{ which gives } F_r(\mathbf{X}, \mathbf{X}') = X_{I,r} - X'_{I,r} \in [-1, 1].$$

Set $u_r(\mathbf{X}) := \mathbb{E}[F_r(\mathbf{X}, \mathbf{X}') \mid \mathbf{X}]$, and use the tower property to get

$$u_r(\mathbf{X}) = \frac{1}{N} \sum_{i=1}^N \left[X_{i,r} - \mathbb{P}(X_i = r \mid \mathbf{X}_{-i}) \right] = \bar{X}_r - \frac{1}{N} \sum_{i=1}^N f_r(\bar{\mathbf{X}}^{(i)}), \quad (74)$$

where we used (73) in the last equality. Next, observe that for $\mathbf{t} \in \mathbb{R}^q$ and $r, s \in [q]$ we have

$$\frac{\partial f_r}{\partial t_s}(\mathbf{t}) = \beta f_r(\mathbf{t})(\mathbb{1}\{r = s\} - f_s(\mathbf{t})). \quad (75)$$

Hence

$$\|\nabla f_r(\mathbf{t})\|_1 = \sum_{s=1}^q \left| \frac{\partial f_r}{\partial t_s}(\mathbf{t}) \right| = \beta f_r(\mathbf{t}) \left(1 - f_r(\mathbf{t}) + \sum_{s \neq r} f_s(\mathbf{t}) \right) = 2\beta f_r(\mathbf{t})(1 - f_r(\mathbf{t})) \leq \frac{\beta}{2}.$$

Consequently, we have:

$$f_r(\bar{\mathbf{X}}^{(i)}) - f_r(\bar{\mathbf{X}}'^{(i)}) \leq \frac{\beta}{2} \max_{s \in [q]} |\bar{X}_s^{(i)} - \bar{X}'_s{}^{(i)}| \leq \frac{\beta}{2} \cdot \frac{1}{N} = \frac{\beta}{2N}.$$

Hence, it follows from (74) that:

$$|u_r(\mathbf{X}) - u_r(\mathbf{X}')| \leq \frac{1}{N} + \frac{\beta}{2N}.$$

Consequently, setting

$$v_r(\mathbf{X}) := \frac{1}{2} \mathbb{E} \left(|u_r(\mathbf{X}) - u_r(\mathbf{X}')| \mid F_r(\mathbf{X}, \mathbf{X}') \mid \mathbf{X} \right),$$

the above display gives $v_r(\mathbf{X}) \leq \frac{2+\beta}{4N}$. Hence, by Theorem 1.5 in Chatterjee (2007b) we have:

$$\mathbb{P}_{\beta, \mathbf{B}}^{\text{CW}}(|u_r(\mathbf{X})| > t) \leq 2 \exp\left(-\frac{2Nt^2}{2+\beta}\right)$$

Next, define $g_r(\mathbf{X}) := \bar{X}_r - f_r(\bar{\mathbf{X}})$, and use (74) to get:

$$|g_r(\mathbf{X}) - u_r(\mathbf{X})| \leq \frac{1}{N} \sum_{i=1}^N \left| f_r(\bar{\mathbf{X}}^{(i)}) - f_r(\bar{\mathbf{X}}) \right| \leq \frac{1}{N} \sum_{i=1}^N \frac{\beta}{2} \max_{s \in [q]} |\bar{X}_s^{(i)} - \bar{X}_s| \leq \frac{\beta}{2N}.$$

Therefore, we have:

$$\mathbb{P}_{\beta, \mathbf{B}}^{\text{CW}} \left(|g_r(\mathbf{X})| > \frac{\beta}{2N} + t \right) \leq \mathbb{P}_{\beta, \mathbf{B}}^{\text{CW}} (|u_r(\mathbf{X})| > t) \leq 2 \exp \left(-\frac{2Nt^2}{2 + \beta} \right).$$

The proof of Lemma H.1 is now complete by a further union bound. \square

Lemma H.2. *Suppose the following assumptions hold:*

- (i) Let $\mathbf{m} \in P([q])$ (see (17)) be a solution to the equation $\xi(\mathbf{t}) = 0$, where $\xi(\mathbf{t}) := \mathbf{t} - f(\mathbf{t})$, and $f: \mathbb{R}^q \mapsto P([q])$ is as in (4).
- (ii) Setting $H = H_{\beta, \mathbf{B}}: P([q]) \rightarrow \mathbb{R}$ by

$$H(\mathbf{t}) := \frac{\beta}{2} \sum_{r=1}^q t_r^2 + \sum_{r=1}^q B_r t_r - \sum_{r=1}^q t_r \log t_r,$$

as in (18), we have $\mathbf{u}^\top \nabla^2 H(\mathbf{m}) \mathbf{u} < 0$ for all $\mathbf{u} \in T^* \setminus \{\mathbf{0}\}$, where

$$T^* := \left\{ \mathbf{u} \in \mathbb{R}^q : \sum_{r=1}^q u_r = 0 \right\}. \quad (76)$$

Then, the Jacobian operator $D\xi(\mathbf{m})$ viewed as a linear map on the domain T^* , is injective.

Proof. Since $\mathbf{m} = f(\mathbf{m})$ (assumption (i)), using (75) the Jacobian of f at \mathbf{m} is given by

$$\frac{\partial f_r}{\partial t_s}(\mathbf{m}) = \beta m_r (\mathbf{1}\{r = s\} - m_s),$$

which can be written in matrix form as

$$Df(\mathbf{m}) = \beta(\text{diag}(\mathbf{m}) - \mathbf{m}\mathbf{m}^\top) =: \beta\Sigma(\mathbf{m}) \Rightarrow D\xi(\mathbf{m}) = I - Df(\mathbf{m}) = I - \beta\Sigma(\mathbf{m}).$$

On the other hand, for $\mathbf{t} \in P([q])$ with strictly positive coordinates, the Hessian of H at \mathbf{m} is given by

$$\nabla^2 H(\mathbf{m}) = \beta I - \text{diag}(1/m_1, \dots, 1/m_q).$$

Now, for any $\mathbf{u} \in \mathbb{R}^q$ we have:

$$\mathbf{u}^\top \Sigma(\mathbf{m}) \mathbf{u} = \sum_{r=1}^q m_r u_r^2 - \left(\sum_{r=1}^q m_r u_r \right)^2 = \frac{1}{2} \sum_{r,s=1}^q m_r m_s (u_r - u_s)^2 \geq 0.$$

Thus $\Sigma(\mathbf{m})$ is positive semidefinite. Moreover, $\mathbf{u}^\top \Sigma(\mathbf{m})\mathbf{u} = 0$ if and only if $u_r = u_s$ for all r, s , which implies that the null space of $\Sigma(\mathbf{m})$ is $\text{span}\{\mathbf{1}\}$. Consequently, the map $\mathbf{u} \mapsto \Sigma(\mathbf{m})\mathbf{u}$ restricted to T^* is invertible.

Now, suppose that $\mathbf{u} \in T^*$. Then

$$\Sigma(\mathbf{m}) \text{diag}(1/\mathbf{m})\mathbf{u} = (\text{diag}(\mathbf{m}) - \mathbf{m}\mathbf{m}^\top) \text{diag}(1/\mathbf{m})\mathbf{u} = \mathbf{u} - \mathbf{m} \sum_{r=1}^q u_r = \mathbf{u},$$

since $\sum_r u_r = 0$. Using the explicit form of $\nabla^2 H(\mathbf{m})$, this yields

$$\begin{aligned} \Sigma(\mathbf{m})\nabla^2 H(\mathbf{m})\mathbf{u} &= \Sigma(\mathbf{m})(\beta I - \text{diag}(1/\mathbf{m}))\mathbf{u} \\ &= \beta \Sigma(\mathbf{m})\mathbf{u} - \mathbf{u} \\ &= -(I - \beta \Sigma(\mathbf{m}))\mathbf{u} = -D\xi(\mathbf{m})\mathbf{u}. \end{aligned}$$

Therefore, for all $\mathbf{u} \in T^*$,

$$D\xi(\mathbf{m})\mathbf{u} = -\Sigma(\mathbf{m})\nabla^2 H(\mathbf{m})\mathbf{u}.$$

Now, suppose that $D\xi(\mathbf{m})\mathbf{u} = \mathbf{0}$ for some $\mathbf{u} \in T^*$. Then, the above display gives $\nabla^2 H(\mathbf{m})\mathbf{u} \in \text{Null}(\Sigma(\mathbf{m}))$, which implies that $\nabla^2 H(\mathbf{m})\mathbf{u} = c\mathbf{1}$ for some constant c . Hence, $\mathbf{u}^\top \nabla^2 H(\mathbf{m})\mathbf{u} = 0$, which by assumption (ii) gives $\mathbf{u} = \mathbf{0}$. Hence, $D\xi(\mathbf{m})$ is injective on the domain T^* . This completes the proof of Lemma H.2. \square

Lemma H.3. *Assume the setting of the Curie–Weiss Potts model defined in (40). Suppose that $H_{\beta, \mathbf{B}}$ (as in (18)) admits a unique global maximizer $\mathbf{m} \in \mathcal{P}([q])$ (cf. Theorem 1.5). Then for every $\delta > 0$,*

$$\limsup_{N \rightarrow \infty} \frac{1}{N} \log \mathbb{P}_{\beta, \mathbf{B}}^{\text{CW}}(\|\bar{\mathbf{X}} - \mathbf{m}\|_\infty \geq \delta) < 0.$$

In particular,

$$\bar{\mathbf{X}} \xrightarrow{P} \mathbf{m}.$$

Proof. To begin with, define:

$$\mathcal{P}_{q, N} := \{\mathbf{v} \in \mathcal{P}([q]) : Nv_r \in \mathbb{Z} \text{ for all } r\},$$

where \mathbb{Z} denotes the set of all integers. For $\mathbf{v} \in \mathcal{P}_{q, N}$ define

$$A_N(\mathbf{v}) = \{\mathbf{x} \in [q]^N : \bar{\mathbf{x}} = \mathbf{v}\}.$$

Then for any Borel set $G \subset \mathbb{R}^q$,

$$\mathbb{P}_{\beta, \mathbf{B}}^{\text{CW}}(\bar{\mathbf{X}} \in G) = \frac{\sum_{\mathbf{v} \in \mathcal{P}_{q, N} \cap G} |A_N(\mathbf{v})| \exp\left\{N \left(\frac{\beta}{2} \sum_{r=1}^q v_r^2 + \sum_{r=1}^q B_r v_r\right)\right\}}{\sum_{\mathbf{v} \in \mathcal{P}_{q, N}} |A_N(\mathbf{v})| \exp\left\{N \left(\frac{\beta}{2} \sum_{r=1}^q v_r^2 + \sum_{r=1}^q B_r v_r\right)\right\}}. \quad (\text{A.1})$$

By Lemma S.4.1 in [Bhowal and Mukherjee \(2023\)](#),

$$|A_N(\mathbf{v})| = \exp \left\{ -N \sum_{r=1}^q v_r \log v_r \right\} \cdot e^{o(N)}$$

uniformly over $\mathbf{v} \in \mathcal{P}_{q,N}$. Substituting this estimate for $|A_N(\mathbf{v})|$ into (A.1) and using the fact that $|\mathcal{P}_{q,N}| \leq (N+1)^q = e^{o(N)}$, we have:

$$\mathbb{P}_{\beta, \mathbf{B}}^{\text{CW}}(\bar{\mathbf{X}} \in G) \leq e^{o(N)} \frac{\sup_{\mathbf{v} \in \mathcal{P}_{q,N} \cap G} e^{NH_{\beta, \mathbf{B}}(\mathbf{v})}}{\sup_{\mathbf{v} \in \mathcal{P}_{q,N}} e^{NH_{\beta, \mathbf{B}}(\mathbf{v})}} \quad (77)$$

Now, note that

$$\sup_{\mathbf{v} \in \mathcal{P}_{q,N}} H_{\beta, \mathbf{B}}(\mathbf{v}) \leq \sup_{\mathbf{v} \in \mathcal{P}([q])} H_{\beta, \mathbf{B}}(\mathbf{v}). \quad (78)$$

On the other hand, if \mathbf{m} is the unique maximizer of $H_{\beta, \mathbf{B}}$ on $\mathcal{P}([q])$, then by Lemma S.4.3 in [Bhowal and Mukherjee \(2023\)](#), there exists a sequence $\mathbf{v}_N \in \mathcal{P}_{q,N}$ such that $\mathbf{v}_N \rightarrow \mathbf{m}$. By continuity of $H_{\beta, \mathbf{B}}$, we have:

$$\sup_{\mathbf{v} \in \mathcal{P}_{q,N}} H_{\beta, \mathbf{B}}(\mathbf{v}) \geq H_{\beta, \mathbf{B}}(\mathbf{v}_N) = H_{\beta, \mathbf{B}}(\mathbf{m}) = \sup_{\mathbf{v} \in \mathcal{P}([q])} H_{\beta, \mathbf{B}}(\mathbf{v}). \quad (79)$$

Combining (78) and (79), we have:

$$\sup_{\mathbf{v} \in \mathcal{P}_{q,N}} H_{\beta, \mathbf{B}}(\mathbf{v}) \longrightarrow \sup_{\mathbf{v} \in \mathcal{P}([q])} H_{\beta, \mathbf{B}}(\mathbf{v}) \quad \text{as } N \rightarrow \infty.$$

Hence, we have the following from (77):

$$\limsup_{N \rightarrow \infty} \frac{1}{N} \log \mathbb{P}_{\beta, \mathbf{B}}^{\text{CW}}(\bar{\mathbf{X}} \in G) \leq \sup_{\mathbf{v} \in \mathcal{P}([q]) \cap G} H_{\beta, \mathbf{B}}(\mathbf{v}) - \sup_{\mathbf{v} \in \mathcal{P}([q])} H_{\beta, \mathbf{B}}(\mathbf{v}). \quad (80)$$

Choosing $G := \{\mathbf{v} \in \mathcal{P}([q]) : \|\mathbf{v} - \mathbf{m}\|_\infty \geq \delta\}$, the RHS of (80) equals a negative constant C_ε (since \mathbf{m} is the unique maximizer of $H_{\beta, \mathbf{B}}$ and $\mathbf{m} \notin G$). This completes the proof of Lemma H.3. \square

Lemma H.4. *Suppose that \mathbf{X} is sampled from the Curie-Weiss Potts model (40) and conditional on \mathbf{X} , let Z_1, \dots, Z_q be independent random variables with $Z_r \sim N(\bar{X}_r, (\beta N)^{-1})$. Let $\mathbf{Z} := (Z_1, \dots, Z_q)$ and $\bar{\mathbf{X}} := (\bar{X}_1, \dots, \bar{X}_q)$.*

(i) *Suppose that $(\beta, \mathbf{B}) \in (0, \infty) \times \mathbb{R}^{q-1}$ is such that the function*

$$H_{\beta, \mathbf{B}}(\mathbf{t}) := \frac{\beta}{2} \sum_{r=1}^q t_r^2 + \sum_{r=1}^q B_r t_r - \sum_{r=1}^q t_r \log t_r$$

(as defined in (18)) has a unique global maximizer \mathbf{m} on the set $\mathcal{P}([q])$.

(ii) *Suppose further that the quadratic form $\mathbf{u}^\top \nabla^2 H_{\beta, \mathbf{B}}(\mathbf{m}) \mathbf{u}$ is strictly negative for all $\mathbf{u} \in T := \{\mathbf{u} \in \mathbb{R}^q \setminus \{\mathbf{0}\} : \sum_{r=1}^q u_r = 0\}$.*

Then both the random variables $\sqrt{N}(\bar{\mathbf{X}} - \mathbf{m})$ and $\sqrt{N}(\mathbf{Z} - \mathbf{m})$ are tight.

Proof. Set $f_r : P([q]) \mapsto \mathbb{R}$ as in (33), setting $f = (f_1, \dots, f_q)$, and $\xi(\mathbf{t}) = \mathbf{t} - f(\mathbf{t})$ as in the statement of Lemma H.2. We now claim that there exists open sets $U \subseteq \mathbb{R}^q$ containing \mathbf{m} , and $V \subseteq \mathbb{R}^q$ containing $\xi(\mathbf{m}) = \mathbf{0}$, such that the function $\xi : U \cap P([q]) \mapsto V \cap T^*$ is invertible, with

$$\varepsilon := \inf_{\mathbf{u} \in U \cap P([q])} \inf_{\mathbf{v} \in T^* : \|\mathbf{v}\|_2=1} \|D\xi(\mathbf{u})\mathbf{v}\|_2 > 0,$$

where $D\xi(\mathbf{u})$ is the Jacobian of the function $\xi : \mathbb{R}^q \mapsto \mathbb{R}^q$ at $\mathbf{u} \in \mathbb{R}^q$.

We now complete the proof of the lemma, deferring the proof of the claim. By Lemma H.3 we have $\mathbb{P}(\bar{\mathbf{X}} \in U) \rightarrow 1$, as $\mathbf{m} \in U$, and on the set $\bar{\mathbf{X}} \in U$ we have

$$\begin{aligned} \|\sqrt{N}(\bar{\mathbf{X}} - \mathbf{m})\|_2 &= \left\| \sqrt{N}(\xi^{-1}(\xi(\bar{\mathbf{X}})) - \xi^{-1}(\mathbf{0})) \right\|_2 \\ &\leq \sqrt{N} \|\xi(\bar{\mathbf{X}})\|_2 \sup_{\mathbf{v} \in V} \|D\xi^{-1}(\mathbf{v})\|_2 \\ &= \sqrt{N} \|\xi(\bar{\mathbf{X}})\|_2 \sup_{\mathbf{u} \in U} \|(D\xi(\mathbf{u})|_{T^*})^{-1}\|_2 \leq \varepsilon^{-1} \sqrt{N} \|\xi(\bar{\mathbf{X}})\|_2. \end{aligned}$$

Since $\sqrt{N} \|\xi(\bar{\mathbf{X}})\|_2 = O_p(1)$ by Lemma H.1, tightness of $\sqrt{N}(\bar{\mathbf{X}} - \bar{m})$ follows. Tightness of $\sqrt{N}(\mathbf{Z} - \mathbf{m})$ follows on noting that

$$\left(\sqrt{N}(\mathbf{Z} - \mathbf{m}) | \mathbf{X} \right) \sim N\left(\sqrt{N}(\bar{\mathbf{X}} - \mathbf{m}), \beta^{-1} \right).$$

It thus remains to verify the claim, for which we will invoke the Inverse function theorem. We break the proof into the following steps:

- \mathbf{m} satisfies $\xi(\mathbf{m}) = \mathbf{0}$.

A Lagrangian argument gives that the global maximizer \mathbf{m} of the function $H_{\beta, \mathbf{B}}(\cdot)$ satisfies the fixed point equation

$$m_r = \frac{e^{\beta m_r + B_r}}{\sum_{s=1}^q e^{\beta m_s + B_s}} = f_r(\mathbf{m}), \quad (81)$$

and so we have $\mathbf{m} = f(\mathbf{m})$, i.e. $\xi(\mathbf{m}) = \mathbf{0}$.

- ξ maps $P([q])$ into T^* , where T^* is as in (76).

Since $f(\mathbf{t}) \in P([q])$ for all $\mathbf{t} \in P([q])$, we have $\sum_{r=1}^q \xi_r(\mathbf{t}) = 0$, and hence

$$\xi(\mathbf{t}) \in T^* = \left\{ \mathbf{u} \in \mathbb{R}^q : \sum_{r=1}^q u_r = 0 \right\}.$$

- For $\mathbf{u} \in P([q])$, the Jacobian operator $D\xi(\mathbf{u})$ induces a linear map

$$D\xi(\mathbf{u})|_{T^*} : T^* \rightarrow T^*.$$

To see this, use (75) to note that for any $\mathbf{u} \in \mathbb{R}^q$ we have

$$D\xi(\mathbf{u}) = \mathbf{I} - \beta(\text{diag}(\mathbf{f}) - \mathbf{f}\mathbf{f}^\top)$$

where $\mathbf{f} := f(\mathbf{u}) \in \mathcal{P}([q])$, and hence,

$$\mathbf{1}^\top D\xi(\mathbf{u}) = \mathbf{1}^\top - \beta(\mathbf{f}^\top - \mathbf{1}^\top \mathbf{f} \mathbf{f}^\top) = \mathbf{1}^\top,$$

which implies that for $\mathbf{v} \in T^*$, one has $\mathbf{1}^\top D\xi(\mathbf{u})\mathbf{v} = \mathbf{1}^\top \mathbf{v} = 0$. So, $D\xi(\mathbf{u})|_{T^*}$ indeed maps T^* to T^* .

- *Completing the proof*

By Lemma H.2, the linear map $D\xi(\mathbf{m})|_{T^*} : T^* \rightarrow T^*$ is injective, which along with the rank-nullity theorem and the previous step shows $D\xi(\mathbf{m})|_{T^*} : T^* \rightarrow T^*$ is invertible. By continuity of $D\xi(\mathbf{u})|_{T^*}$ in \mathbf{u} , there exists a non-empty neighborhood $U' \subseteq \mathbb{R}^q$ of \mathbf{m} such that

$$\varepsilon' := \inf_{\mathbf{u} \in U' \cap \mathcal{P}([q])} \inf_{\mathbf{v} \in T^* : \|\mathbf{v}\|_2=1} \|D\xi(\mathbf{u})\mathbf{v}\| > 0,$$

By the inverse function theorem applied to the map $\xi : \mathcal{P}([q]) \rightarrow T^*$, there exist non-empty neighborhoods $U \subseteq U'$ of \mathbf{m} and $V \subseteq T^*$ of $\xi(\mathbf{m}) = \mathbf{0}$ such that the restriction $\xi : U \cap \mathcal{P}([q]) \rightarrow V \cap T^*$ is a bijection (hence, invertible). This verifies the claim, and hence completes the proof of the Lemma. □

Lemma H.5. *Let $(\beta, \mathbf{B}) \in (0, \infty) \times \mathbb{R}^{q-1}$ be such that:*

- (i) *The function $H_{\beta, \mathbf{B}}(\mathbf{t}) = \frac{\beta}{2} \sum_{r=1}^q t_r^2 + \sum_{r=1}^q B_r t_r - \sum_{r=1}^q t_r \log t_r$ (as defined in (18)) has the unique global maximizer \mathbf{m} on the set $\mathcal{P}([q])$,*
- (ii) *$\mathbf{u}^\top \nabla^2 H(\mathbf{m})\mathbf{u} < 0$ for all $\mathbf{u} \in T = \{\mathbf{u} \in \mathbb{R}^q \setminus \{\mathbf{0}\} : \sum_{r=1}^q u_r = 0\}$.*

Then, the product measure $\mathbf{m}^N := \otimes_{i=1}^N \mathbf{m}$ is contiguous to the Curie-Weiss Potts measure $\mathbb{P}_{\beta, \mathbf{B}}^{\text{CW}}$.

Proof. Suppose that \mathbf{X} is sampled from the Curie-Weiss Potts model (40) and conditional on \mathbf{X} , let Z_1, \dots, Z_q be independent random variables with $Z_r \sim N(\bar{X}_r, (\beta N)^{-1})$. Let $\mathbf{Z} := (Z_1, \dots, Z_q)$ and $\bar{\mathbf{X}} := (\bar{X}_1, \dots, \bar{X}_q)$. Also, let \mathbb{P} denote the joint distribution of \mathbf{X} and \mathbf{Z} . By Lemma G.1 and (33), we have $\mathbb{P}(X_i = r | \mathbf{Z}) = f_r(\mathbf{Z})$, where

$$f_r(\mathbf{t}) := \frac{e^{\beta t_r + B_r}}{\sum_{s=1}^q e^{\beta t_s + B_s}}.$$

Next, define

$$W_r := \sqrt{N}(f_r(\mathbf{Z}) - f_r(\mathbf{m})). \quad (82)$$

Note that $\|\nabla f_r(\cdot)\|_\infty < \infty$ and $\sqrt{N}(\mathbf{Z} - \mathbf{m})$ is tight (by Lemma H.4), and hence $W_r = O_{\mathbb{P}}(1)$. Since $m_r = f_r(\mathbf{m})$ (see (81)), we get the following from (82):

$$\mathbb{P}(X_i = r | \mathbf{Z}) = m_r + \frac{W_r}{\sqrt{N}}.$$

Next, define a product measure \mathbb{Q} on $[q]^N \times \mathbb{R}^q$ as $\mathbb{Q} := \mathbf{m}^N \otimes \mu$, where μ denotes the marginal distribution of \mathbf{Z} . Setting $T_{N,r} := |\{i \in [N] : X_i = r\}|$, the standard central limit theorem gives the following under \mathbb{Q} :

$$\frac{T_{N,r} - Nm_r}{\sqrt{N}} = O_{\mathbb{Q}}(1) .$$

Now, for every $K > 0$, on the intersection of the events $|W_r| \leq K$ and $|T_{N,r} - Nm_r| \leq K\sqrt{N}$ for all $r \in [q]$, we have:

$$\begin{aligned} \log \frac{\mathbb{Q}(\mathbf{X}|\mathbf{Z})}{\mathbb{P}(\mathbf{X}|\mathbf{Z})} &= - \sum_{r=1}^q T_{N,r} \log \left(1 + \frac{W_r}{m_r \sqrt{N}} \right) \\ &= - \sum_{r=1}^q T_{N,r} \left(\frac{W_r}{m_r \sqrt{N}} + O\left(\frac{K^2}{N}\right) \right) \\ &= - \sum_{r=1}^q \frac{T_{N,r} W_r}{m_r \sqrt{N}} + O(K^2) \\ &= \sum_{r=1}^{q-1} W_r \left(\frac{T_{N,q}}{m_q \sqrt{N}} - \frac{T_{N,r}}{m_r \sqrt{N}} \right) + O(K^2) \quad (\text{since } \sum_{r=1}^q W_r = 0) \\ &\leq \sum_{r=1}^{q-1} |W_r| \left(\frac{K}{m_r} + \frac{K}{m_q} \right) + O(K^2) \\ &\leq K^2 \sum_{r=1}^q \left(\frac{1}{m_r} + \frac{1}{m_q} \right) + O(K^2) =: \phi_K . \end{aligned}$$

Thus, if $A_N \subseteq [q]^N$ is a sequence of sets such that $\mathbb{P}(\mathbf{X} \in A_N) \rightarrow 0$ as $N \rightarrow \infty$, then:

$$\begin{aligned} &\mathbb{Q} \left(\mathbf{X} \in A_N, |W_r| \leq K \forall r \in [q], |T_{N,r} - Nm_r| \leq K\sqrt{N} \forall r \in [q] \right) \\ &= \mathbb{E} \mathbb{Q} \left(\mathbf{X} \in A_N, |W_r| \leq K \forall r \in [q], |T_{N,r} - Nm_r| \leq K\sqrt{N} \forall r \in [q] \mid \mathbf{Z} \right) \\ &\leq e^{\phi_K} \mathbb{E} \mathbb{P} \left(\mathbf{X} \in A_N, |W_r| \leq K \forall r \in [q], |T_{N,r} - Nm_r| \leq K\sqrt{N} \forall r \in [q] \mid \mathbf{Z} \right) \\ &= e^{\phi_K} \mathbb{P} \left(\mathbf{X} \in A_N, |W_r| \leq K \forall r \in [q], |T_{N,r} - Nm_r| \leq K\sqrt{N} \forall r \in [q] \right) \\ &\leq e^{\phi_K} \mathbb{P}(\mathbf{X} \in A_N) . \end{aligned}$$

Hence, we have:

$$\mathbb{Q}(\mathbf{X} \in A_N) \leq e^{\phi_K} \mathbb{P}(\mathbf{X} \in A_N) + \sum_{r=1}^q \mathbb{Q}(|W_r| > K) + \sum_{r=1}^q \mathbb{Q}(|T_{N,r} - Nm_r| > K\sqrt{N})$$

which on letting $N \rightarrow \infty$ followed by $K \rightarrow \infty$ gives $\mathbb{Q}(\mathbf{X} \in A_N) \rightarrow 0$. This completes the proof of Lemma H.5. \square

Appendix I: Other Technical Lemmas

In this section, we state additional technical lemmas necessary for proving some of the main results of the paper. We start with the following lemma, which is crucial in establishing existence of the partial MPL estimators $\hat{\beta}_N$ and $\hat{\mathbf{B}}_N$.

Lemma I.1. *Define the sets $A_{2,N}, A_{3,N}, A_{4,N}$*

$$A_{2,N} = \{\mathbf{x} \in [q]^N : m_{i,x_i}(\mathbf{x}) = \min_{r \in [q]} m_{i,r}(\mathbf{x}) \text{ for all } i \in [N]\},$$

$$A_{3,N} = \{\mathbf{x} \in [q]^N : m_{i,x_i}(\mathbf{x}) = \max_{r \in [q]} m_{i,r}(\mathbf{x}) \text{ for all } i \in [N]\},$$

$$A_{4,N} = \{\mathbf{x} \in [q]^N : \text{There exists } r \in [q] \text{ such that for all } i \in [N], x_i \neq r\},$$

as in (59), (58) and (57) respectively. Then, the following conclusions hold:

- (a) If $\mathbf{X} \in A_{2,N}^c \cap A_{3,N}^c$, then for every $\mathbf{B} \in \mathbb{R}^{q-1}$, $\ell_N(\beta, \mathbf{B}) \rightarrow -\infty$ as $|\beta| \rightarrow \infty$.
- (b) If $\mathbf{X} \in A_{4,N}^c$, then for every $\beta \in \mathbb{R}$, $\ell_N(\beta, \mathbf{B}) \rightarrow -\infty$ as $\|\mathbf{B}\|_\infty \rightarrow \infty$.

Proof. To begin with, use (5) to note that:

$$\ell_N(\beta, \mathbf{B}) = \sum_{i=1}^N \log(\theta_{i,X_i}),$$

where $\theta_{i,r}$ is as in (4), and satisfies the inequality

$$\theta_{i,X_i} = \frac{\exp\{\beta m_{i,X_i}(\mathbf{X}) + B_{X_i}\}}{\sum_{t=1}^q \exp\{\beta m_{i,t}(\mathbf{X}) + B_t\}} \leq \frac{1}{1 + \exp\{\beta(m_{i,t}(\mathbf{X}) - m_{i,X_i}(\mathbf{X})) + B_t - B_{X_i}\}} \quad (83)$$

for all $t \neq X_i$. Since $\theta_{i,r} \leq 1$ for all i, r , showing $\ell_N(\beta, \mathbf{B}) \rightarrow -\infty$ is equivalent to proving that there exists $i \in [N]$ such that $\theta_{i,X_i} \rightarrow 0$.

- (a) $|\beta| \rightarrow \infty$.

Case I.2 ($\beta \rightarrow -\infty$). If $\mathbf{X} \in A_{2,N}^c$, there exists $i \in [N]$ and $r \in [q]$ such that $m_{i,r}(\mathbf{X}) < m_{i,X_i}(\mathbf{X})$ (and so $X_i \neq r$). Then, (83) with $t = r$ gives:

$$\theta_{i,X_i} \leq \frac{1}{1 + \exp\{\beta(m_{i,r}(\mathbf{X}) - m_{i,X_i}(\mathbf{X})) + B_r - B_{X_i}\}}$$

which implies that $\theta_{i,X_i} \rightarrow 0$ as $\beta \rightarrow -\infty$.

Case I.3 ($\beta \rightarrow \infty$). If $\mathbf{X} \in A_{3,N}^c$, there exists $i \in [N]$ and $r \in [q]$ such that $m_{i,r}(\mathbf{X}) > m_{i,X_i}(\mathbf{X})$ (and so $X_i \neq r$). Then, (83) with $t = r$ gives:

$$\theta_{i,X_i} \leq \frac{1}{1 + \exp\{\beta(m_{i,r}(\mathbf{X}) - m_{i,X_i}(\mathbf{X})) + B_r - B_{X_i}\}}$$

which implies that $\theta_{i,X_i} \rightarrow 0$ as $\beta \rightarrow \infty$.

(b) $\|\mathbf{B}\|_\infty \rightarrow \infty$

This ensures the existence of an $r \in [q]$ such that $|B_r| \rightarrow \infty$. Also $r \neq q$, as $B_q = 0$ by convention.

Case I.4 ($B_r \rightarrow -\infty$). If $\mathbf{X} \in A_{4,N}^c$, there exists $i \in [N]$ such that $X_i = r$, in which case (83) with $t = q$ gives:

$$\theta_{i,X_i} \leq \frac{1}{1 + \exp\{\beta(m_{i,q}(\mathbf{X}) - m_{i,r}(\mathbf{X})) - B_r\}}$$

which implies that $\theta_{i,X_i} \rightarrow 0$ as $B_r \rightarrow -\infty$.

Case I.5 ($B_r \rightarrow \infty$). If $\mathbf{X} \in A_{4,N}^c$, there exists $i \in [N]$ such that $X_i \neq r$ (otherwise the whole vector $\{X_i\}_{1 \leq i \leq N}$ have the same color, which contradicts $A_{4,N}$). Using (83) with $t = r$ we have:

$$\theta_{i,X_i} \leq \frac{1}{1 + \exp\{\beta(m_{i,r}(\mathbf{X}) - m_{i,X_i}(\mathbf{X})) + B_r - B_{X_i}\}}$$

which implies that $\theta_{i,X_i} \rightarrow 0$ as $B_r \rightarrow \infty$.

This completes the proof of Lemma I.1. □

The next result gives a concentration for the vector of conditional probabilities, and will be used to prove Theorem 1.4.

Lemma I.6. Suppose \mathbf{X} is an observation from the Potts model (1), where the coupling matrix \mathbf{A}_N satisfies the assumptions (8), (9) and (43). Let $\mathcal{S}_{N,q}$ denote the set of all $\mathbf{y} := ((y_{i,r}))_{i \in [N], r \in [q]} \in [0, 1]^{Nq}$, such that $\sum_{r=1}^q y_{i,r} = 1$ for all $i \in [N]$. Set the functions $h_N : \mathcal{S}_{N,q} \rightarrow \mathbb{R}$ and $I_N : \mathcal{S}_{N,q} \rightarrow \mathbb{R}$ as:

$$h_N(\mathbf{y}) = \frac{\beta}{2} \sum_{1 \leq i, j \leq N} \sum_{r=1}^q a_{ij} y_{i,r} y_{j,r} + \sum_{i=1}^N \sum_{r=1}^q B_r y_{i,r} \quad \text{and} \quad I_N(\mathbf{y}) = \sum_{i=1}^N \sum_{r=1}^q y_{i,r} \log y_{i,r}$$

(as in (47)), and let $\psi_N(\mathbf{y}) = h_N(\mathbf{y}) - I_N(\mathbf{y})$ (as in (48)), $\boldsymbol{\theta}(\mathbf{X}) = ((\theta_{i,r}(\mathbf{X})))_{i \in [N], r \in [q]}$ and

$$\bar{\nabla}(\mathbf{y}) = \frac{1}{q} \sum_{r=1}^q \nabla_{\cdot r}(\mathbf{y})$$

(as in (51)) where $\nabla_{\cdot r}(\mathbf{y}) := ((\partial \psi_N / \partial y_{i,r}))_{i \in [N]}$ (as in (50)). Then, we have the following.

(a) As $N \rightarrow \infty$,

$$\psi_N(\boldsymbol{\theta}(\mathbf{X})) = \sup_{\mathbf{y} \in \mathcal{S}_{N,q}} \psi_N(\mathbf{y}) + o_{\mathbb{P}}(N).¹$$

¹By a slight abuse of notation, for a function $\ell : [0, 1]^{Nq} \rightarrow \mathbb{R}$ and a vector $\mathbf{x} \in [q]^N$, we will often refer to $\ell((x_{i,r})_{i \in [N], r \in [q]})$ as $\ell(\mathbf{x})$, where $x_{i,r} := \mathbb{1}_{x_i=r}$

(b) For all $r \in [q]$, as $N \rightarrow \infty$,

$$\|\nabla_{\cdot,r}(\boldsymbol{\theta}(\mathbf{X})) - \bar{\nabla}(\boldsymbol{\theta}(\mathbf{X}))\| = o_{\mathbb{P}}(\sqrt{N}).$$

Proof. (a) Choosing $b_{itrs} = \mathbb{1}_{t=r}$ and $g \equiv 1$ in Lemma F.1 gives $L_i = 1$, and so

$$\mathbb{P}\left(\left|\sum_{i=1}^N (X_{ir} - \theta_{i,r}(\mathbf{X}))\right| \geq \sqrt{Nt}\right) \leq 2 \exp(-Ct).$$

Similarly, choosing $b_{itrs} = \mathbb{1}_{t=r}$, $g(x) = x$ and $\lambda = 0$ in Lemma F.1 gives $L_i = 1$, and so

$$\mathbb{P}\left(\left|\sum_{i=1}^N (X_{ir} - \theta_{i,r}(\mathbf{X}))m_{i,r}(\mathbf{X})\right| \geq \sqrt{Nt}\right) \leq 2 \exp(-Ct).$$

Fixing $\varepsilon > 0$, a union bound over all the colors $r \in [q]$ gives $\mathbb{P}(\mathbf{X} \in A_{N,\varepsilon} \cap B_{N,\varepsilon}) \rightarrow 1$, where:

$$A_{N,\varepsilon} := \left\{ \mathbf{x} \in [q]^N : \left| \sum_{i=1}^N (x_{i,r} - \theta_{i,r}(\mathbf{x})) \right| \leq N\varepsilon \text{ for all } r \in [q] \right\},$$

$$B_{N,\varepsilon} := \left\{ \mathbf{x} \in [q]^N : \left| \sum_{i=1}^N (x_{i,r} - \theta_{i,r}(\mathbf{x}))m_{i,r}(\mathbf{x}) \right| \leq N\varepsilon \text{ for all } r \in [q] \right\}.$$

Also, it follows from the proof of Theorem 1.1 in Basak and Mukherjee (2017) (see (Basak and Mukherjee, 2017, Lem 3.2)), that under Conditions (8) and (43) we have:

$$\mathbb{E} \left[(h_N(\mathbf{X}) - h_N(\boldsymbol{\theta}(\mathbf{X})))^2 \right] = o(N^2)$$

and hence $\mathbb{P}(\mathbf{X} \in C_{N,\varepsilon}) \rightarrow 1$, where

$$C_{N,\varepsilon} := \{ \mathbf{x} \in [q]^N : |h_N(\mathbf{x}) - h_N(\boldsymbol{\theta}(\mathbf{x}))| \leq N\varepsilon \}.$$

This shows that $\mathbb{P}(\mathbf{X} \in D_{N,\varepsilon}) \rightarrow 1$, where $D_{N,\varepsilon} = A_{N,\varepsilon} \cap B_{N,\varepsilon} \cap C_{N,\varepsilon}$. Consequently, setting

$$W_{N,\delta} := \{ \mathbf{x} \in [q]^N : \psi_N(\boldsymbol{\theta}(\mathbf{x})) \leq r_N - N\delta \}$$

where $r_N := \sup_{\mathbf{y} \in \mathcal{S}_{N,q}} \psi_N(\mathbf{y})$, we have:

$$\mathbb{P}(\mathbf{X} \in W_{N,\delta}) \leq \mathbb{P}(\mathbf{X} \in W_{N,\delta} \cap D_{N,\varepsilon}) + \mathbb{P}(D_{N,\varepsilon}^c) = \frac{\sum_{\mathbf{x} \in W_{N,\delta} \cap D_{N,\varepsilon}} e^{h_N(\mathbf{x})}}{\sum_{\mathbf{x} \in [q]^N} e^{h_N(\mathbf{x})}} + o(1). \quad (84)$$

Also, the Gibbs variational principle gives a variational mean field lower bound to the denominator of (84) as follows (see, for example, (Basak and Mukherjee, 2017, Eqn 1.8)):

$$\sum_{\mathbf{x} \in [q]^N} e^{h_N(\mathbf{x})} \geq \sup_{\mathbf{x} \in \mathcal{P}([q]^N)} e^{\psi_N(\mathbf{x})} = e^{r_N}. \quad (85)$$

The task now is to bound the numerator of the ratio in the right-hand side of (84). Towards this, define $g_N : [0, 1]^{2Nq} \rightarrow \mathbb{R}$ as

$$g_N(\mathbf{z}, \mathbf{w}) := \sum_{i=1}^N \sum_{r=1}^q z_{i,r} \log w_{i,r}.$$

Note that $I_N(\mathbf{y}) = g_N(\mathbf{y}, \mathbf{y})$. It follows from the proof of (Basak and Mukherjee, 2017, Thm 1.1) (see (Basak and Mukherjee, 2017, Page 575 last display)) that:

$$\begin{aligned} \left| g_N(\tilde{\mathbf{X}}, \boldsymbol{\theta}(\mathbf{X})) - I_N(\boldsymbol{\theta}(\mathbf{X})) \right| &= \left| \sum_{i \in [N], r \in [q]} (X_{i,r} - \theta_{i,r}(\mathbf{X})) (\beta m_{i,r}(\mathbf{X}) + B_r) \right| \\ &\stackrel{D_{N,\varepsilon}}{\leq} (\beta + B) q N \varepsilon \end{aligned} \quad (86)$$

where $B := \|\mathbf{B}\|_\infty$, and for any $\mathbf{x} \in [q]^N$, we denote $\tilde{\mathbf{x}} := (x_{i,r})_{i \in [N], r \in [q]} \in \mathcal{S}_{N,q}$. At this point, we need the following definition:

Definition I.7. For $S \subseteq \mathbb{R}^N$ and $\varepsilon > 0$, a set D is called an ε -net of S , if given any $s \in S$ there exists $d \in D$ such that $\|s - d\|_2 \leq \varepsilon$.

The following result ((Basak and Mukherjee, 2017, Lem 3.4)) guarantees that under Condition (43), the set $\{\mathbf{A}_N \mathbf{v}, \mathbf{v} \in [0, 1]^N\}$ has an $\varepsilon\sqrt{N}$ -net of size $e^{o(N)}$.

Proposition I.8. *If \mathbf{A}_N satisfies Condition (43), then for every $\varepsilon > 0$, the set*

$$\{\mathbf{A}_N \mathbf{v} : \mathbf{v} \in [0, 1]^N\}$$

has an $\varepsilon\sqrt{N}$ -net $J_{N,\varepsilon}$ of cardinality $e^{o(N)}$.

For every $\mathbf{p} \in J_{N,\varepsilon}$, define:

$$L_r(\mathbf{p}) := \left\{ \mathbf{x} \in [q]^N : \|\mathbf{p} - m_{\cdot r}(\mathbf{x})\| \leq \varepsilon\sqrt{N} \right\}$$

where $m_{\cdot r}(\mathbf{x}) := (m_{i,r}(\mathbf{x}))_{i \in [N]}$, and $m_{i,r}(\mathbf{x}) = \sum_{j=1}^N a_{ij} \mathbb{1}_{x_j=r}$. Then, we have:

$$\begin{aligned} &\sum_{\mathbf{x} \in W_{N,\delta} \cap D_{N,\varepsilon}} e^{h_N(\mathbf{x})} \\ &\leq_{C_{N,\varepsilon}} e^{N\varepsilon} \sum_{\mathbf{x} \in W_{N,\delta} \cap D_{N,\varepsilon}} e^{h_N(\boldsymbol{\theta}(\mathbf{x}))} \\ &\leq e^{N\varepsilon(1+q(\beta+B))} \sum_{\mathbf{x} \in W_{N,\delta} \cap D_{N,\varepsilon}} e^{h_N(\boldsymbol{\theta}(\mathbf{x})) + g_N(\tilde{\mathbf{x}}, \boldsymbol{\theta}(\mathbf{x})) - I_N(\boldsymbol{\theta}(\mathbf{x}))} \quad (\text{by (86)}) \\ &\leq e^{N\varepsilon(1+q(\beta+B))} \sup_{\mathbf{x} \in W_{N,\delta}} e^{\psi_N(\boldsymbol{\theta}(\mathbf{x}))} \sum_{\mathbf{x} \in [q]^N} e^{g_N(\tilde{\mathbf{x}}, \boldsymbol{\theta}(\mathbf{x}))} \\ &\leq_{W_{N,\delta}} e^{N\varepsilon(1+q(\beta+B)) + r_N - N\delta} \sum_{\mathbf{x} \in [q]^N} e^{g_N(\tilde{\mathbf{x}}, \boldsymbol{\theta}(\mathbf{x}))} \end{aligned}$$

Since $J_{N,\varepsilon}$ is a $\varepsilon\sqrt{N}$ net of the set $\{\mathbf{A}_N \mathbf{v}, \mathbf{v} \in [0, 1]^N\}$, and $\|\mathbf{A}_N \mathbf{v}\|_\infty \leq \gamma$ for all $\mathbf{v} \in [0, 1]^N$ (see (8)), by replacing ε by a factor of 2 if necessary, without loss of generality we can assume $\|\mathbf{p}\|_\infty \leq \gamma$ for all $\mathbf{p} \in J_{N,\varepsilon}$. Thus, noting that $m_{\cdot r}(\mathbf{x}) = \mathbf{A}_N \mathbf{x}_{\cdot r}$ where $x_{ir} = \mathbb{1}_{x_i=r}$, for every $\mathbf{x} \in [q]^N$ and every $r \in [q]$, there exists $\mathbf{p} \in J_{N,\varepsilon}$ such that $\|\mathbf{p} - m_{\cdot r}(\mathbf{x})\| \leq \varepsilon\sqrt{N}$, i.e. $\mathbf{x} \in L_r(\mathbf{p})$. For $\mathbf{P} := (\mathbf{p}_1, \dots, \mathbf{p}_q) \in J_{N,\varepsilon}^q$ we can write:

$$\sum_{\mathbf{x} \in W_{N,\delta} \cap D_{N,\varepsilon}} e^{h_N(\mathbf{x})} \leq e^{N\varepsilon(1+q(\beta+B))+r_N-N\delta} \sum_{\mathbf{P} \in J_{N,\varepsilon}^q} \sum_{\mathbf{x} \in L(\mathbf{P})} e^{g_N(\tilde{\mathbf{x}}, \theta(\mathbf{x}))} \quad (87)$$

where $L(\mathbf{P}) := \cap_{r \in [q]} L_r(\mathbf{p}_r)$. Next, define the matrix $\mathbf{u}(\mathbf{P}) := (u_{i,r}(\mathbf{P}))_{i \in [n], r \in [q]}$, where:

$$u_{i,r}(\mathbf{P}) := \frac{\exp\{\beta p_{i,r} + B_r\}}{\sum_{t=1}^q \exp\{\beta p_{i,t} + B_t\}} \geq q^{-1} \exp(-\beta\gamma - 2B) = \alpha,$$

where α is as in (27). Since $\theta(\mathbf{x}) \geq \alpha$ as well, mean-value theorem gives

$$|g_N(\tilde{\mathbf{x}}, \theta(\mathbf{x})) - g_N(\tilde{\mathbf{x}}, \mathbf{u}(\mathbf{P}))| \leq C_1 \sum_{i=1}^N \sum_{r=1}^q |\theta_{i,r}(\mathbf{x}) - u_{i,r}(\mathbf{P})|.$$

Also observe that

$$\theta_{i,r}(\mathbf{x}) = f_r(m_{i,1}(\mathbf{x}), \dots, m_{i,q}(\mathbf{x})), \quad u_{i,r}(\mathbf{P}) = f_r(p_{i,1}, \dots, p_{i,q}),$$

where

$$f_r(t_1, \dots, t_q) = \frac{\exp(\beta t_r + B_r)}{\sum_{s=1}^q \exp(\beta t_s + B_s)}$$

as in (33). Since $\|\nabla f_r\|_\infty < \infty$, another mean value theorem gives

$$|\theta_{i,r}(\mathbf{x}) - u_{i,r}(\mathbf{P})| \leq C_2 \sum_{s=1}^q |m_{i,s}(\mathbf{x}) - p_{i,s}|$$

Combining the last two bounds we get

$$\begin{aligned} |g_N(\tilde{\mathbf{x}}, \theta(\mathbf{x})) - g_N(\tilde{\mathbf{x}}, \mathbf{u}(\mathbf{P}))| &\leq C_1 C_2 q \sum_{i=1}^N \sum_{s=1}^q |m_{i,s}(\mathbf{x}) - p_{i,s}| \\ &\leq C_1 C_2 q \sqrt{N} \sum_{s=1}^q \|m_{\cdot s}(\mathbf{x}) - \mathbf{p}_s\| \leq C q \varepsilon N, \end{aligned}$$

where the last equality uses the fact that $\mathbf{x} \in L(\mathbf{P})$, and $C = C_1 C_2 q$.

Hence, from (87),

$$\sum_{\mathbf{x} \in W_{N,\delta} \cap D_{N,\varepsilon}} e^{h_N(\mathbf{x})}$$

$$\begin{aligned}
&\leq e^{N\varepsilon(1+q(\beta+B+C))+r_N-N\delta} \sum_{\mathbf{P} \in J_{N,\varepsilon}^q} \sum_{\mathbf{x} \in L(\mathbf{P})} e^{g_N(\tilde{\mathbf{x}}, \mathbf{u}(\mathbf{P}))} \\
&\leq e^{N\varepsilon(1+q(\beta+B+C))+r_N-N\delta} \sum_{\mathbf{P} \in J_{N,\varepsilon}^q} \sum_{\mathbf{x} \in [q]^N} e^{g_N(\tilde{\mathbf{x}}, \mathbf{u}(\mathbf{P}))}
\end{aligned}$$

Finally, since $u_{i,r}(\mathbf{P}) = v_{i,r}$ satisfies $\sum_{r=1}^q v_{i,r} = 1$ for any $\mathbf{P} \in J_{N,\varepsilon}^q$, we have

$$\begin{aligned}
\sum_{\mathbf{x} \in [q]^N} e^{g_N(\tilde{\mathbf{x}}, \mathbf{u}(\mathbf{P}))} &= \sum_{\mathbf{x} \in [q]^N} \exp \left\{ \sum_{i=1}^N \sum_{r=1}^q x_{i,r} \log v_{i,r} \right\} \\
&= \sum_{\mathbf{x} \in [q]^N} \prod_{i=1}^N \prod_{r=1}^q v_{i,r}^{x_{i,r}} \\
&= \prod_{i=1}^N \left(\sum_{\mathbf{x} \in [q]} \prod_{r=1}^q v_{i,r}^{x_r} \right) \\
&= \prod_{i=1}^N \sum_{r=1}^q v_{i,r} = 1.
\end{aligned}$$

Hence, using the fact that $|J_{N,\varepsilon}| = e^{o(N)}$ we have:

$$\begin{aligned}
\sum_{\mathbf{x} \in W_{N,\delta} \cap D_{N,\varepsilon}} e^{h_N(\mathbf{x})} &\leq e^{N\varepsilon(1+q(\beta+B+C))+r_N-N\delta} |J_{N,\varepsilon}|^q \\
&= e^{N\varepsilon(1+q(\beta+B+C))+r_N-N\delta+qo(N)} \tag{88}
\end{aligned}$$

Combining (85) and (88), the above display gives

$$\mathbb{P}(\mathbf{X} \in W_{N,\delta}) \leq e^{N\varepsilon(1+q(\beta+B+C))-N\delta+qo(N)} + o(1)$$

Since $\varepsilon > 0$ is arbitrary, we conclude that $\mathbb{P}(\mathbf{X} \in W_{N,\delta}) = o(1)$, i.e.

$$\mathbb{P} \left(\sup_{\mathbf{y} \in \mathcal{S}_{N,q}} \psi_N(\mathbf{y}) - \psi_N(\theta(\mathbf{X})) \geq N\delta \right) = o(1)$$

for all $\delta > 0$, which implies that

$$\sup_{\mathbf{y} \in \mathcal{S}_{N,q}} \psi_N(\mathbf{y}) - \psi_N(\theta(\mathbf{X})) = o_{\mathbb{P}}(N)$$

and completes the proof of part (a) of Lemma I.6.

(b) Fixing $\delta > 0$, note that it suffices to show the following for all $r \in [q]$:

$$\mathbb{P}(\|\nabla_{\cdot,r}(\theta(\mathbf{X})) - \bar{\nabla}(\theta(\mathbf{X}))\|^2 > N\delta) = o(1).$$

To begin with, use the expression of $\theta_{i,r}(\mathbf{X})$ in (4) to note that for all $i \in [N]$ and $r \in [q]$,

$$0 < p_1 := \frac{1}{1 + (q-1)e^{\beta\gamma+2B}} \leq \theta_{i,r}(\mathbf{X}) \leq \frac{1}{1 + (q-1)e^{-\beta\gamma-2B}} =: p_2 < 1,$$

where $B := \|\mathbf{B}\|_\infty$. Now, suppose that $\mathbf{y} \in [p_1, p_2]^{Nq} \cap \mathcal{S}_{N,q}$ is such that $\|\nabla_{\cdot r}(\mathbf{y}) - \bar{\nabla}(\mathbf{y})\|^2 > N\delta$ for some $r \in [q]$, where $\nabla_{\cdot r}$ and $\bar{\nabla}$ are as in the statement of the lemma. Also set $\tilde{\nabla}(\mathbf{y}) := (\bar{\nabla}(\mathbf{y}), \dots, \bar{\nabla}(\mathbf{y}))$ as in (51). Then, setting $\mathbf{y}^{(t)} := \mathbf{y} + t(\nabla\psi_N(\mathbf{y}) - \tilde{\nabla}(\mathbf{y}))$, we claim that $\mathbf{y}^{(t)} \in [0, 1]^{Nq}$ for $t \in [0, \varepsilon]$ for some fixed $\varepsilon > 0$ not depending on N . Towards showing this, note from (49) that:

$$\|\nabla\psi_N(\mathbf{y})\|_\infty \leq \beta\gamma + B + 1 - \log p_1$$

and hence, we can take $\varepsilon := \frac{1}{2}(\beta\gamma + B + 1 - \log p_1)^{-1} \min\{p_1, 1 - p_2\}$. Since $\mathbf{y} \in \mathcal{S}_{N,q}$, and

$$\sum_{r=1}^q \left(\nabla\psi_N(\mathbf{y}) - \tilde{\nabla}(\mathbf{y}) \right)_{i,r} = 0 \text{ for all } i \in [N] \text{ using (51),}$$

we also have $\mathbf{y}^{(t)} \in \mathcal{S}_{N,q}$. Now, a two-term Taylor expansion of the function $t \mapsto \psi_N(\mathbf{y}^{(t)})$ for $t \in [0, \varepsilon]$ gives some $\xi \in (0, t)$ satisfying:

$$\begin{aligned} & \psi_N(\mathbf{y}^{(t)}) - \psi_N(\mathbf{y}) \\ &= t\|\nabla\psi_N(\mathbf{y}) - \tilde{\nabla}(\mathbf{y})\|^2 + \frac{t^2}{2} \left(\nabla\psi_N(\mathbf{y}) - \tilde{\nabla}(\mathbf{y}) \right)^\top \nabla^2\psi_N(\mathbf{y}^{(\xi)}) \left(\nabla\psi_N(\mathbf{y}) - \tilde{\nabla}(\mathbf{y}) \right) \\ &\geq t\|\nabla\psi_N(\mathbf{y}) - \tilde{\nabla}(\mathbf{y})\|^2 - \frac{t^2}{2} \|\nabla\psi_N(\mathbf{y}) - \tilde{\nabla}(\mathbf{y})\|^2 \lambda_{\max}(-\nabla^2\psi_N(\mathbf{y}^{(\xi)})) \\ &\geq t\|\nabla\psi_N(\mathbf{y}) - \tilde{\nabla}(\mathbf{y})\|^2 - \frac{t^2}{2} \|\nabla\psi_N(\mathbf{y}) - \tilde{\nabla}(\mathbf{y})\|^2 \|\nabla^2\psi_N(\mathbf{y}^{(\xi)})\|_1 \\ &\geq t\|\nabla\psi_N(\mathbf{y}) - \tilde{\nabla}(\mathbf{y})\|^2 - \frac{t^2}{2} \|\nabla\psi_N(\mathbf{y}) - \tilde{\nabla}(\mathbf{y})\|^2 \left(\beta\gamma + \frac{1}{p_1} \right) \\ &= t\|\nabla\psi_N(\mathbf{y}) - \tilde{\nabla}(\mathbf{y})\|^2 \left[1 - \frac{t}{2} \left(\beta\gamma + \frac{1}{p_1} \right) \right]. \end{aligned}$$

In the above display, the last inequality uses the fact

$$\begin{aligned} \nabla^2\psi_N(\mathbf{z})_{ir,js} &= -\frac{1}{z_{ir}} & \text{if } i = j \text{ and } r = s, \\ &= \beta a_{ij} & \text{if } i \neq j \text{ and } r = s, \\ &= 0 & \text{if } r \neq s \end{aligned}$$

to get the bound $\|\nabla^2\psi_N(\mathbf{z})\|_1 \leq \beta\gamma + \frac{1}{p_1}$ uniformly in $\mathbf{z} \in \mathcal{S}_{N,q}$. Choosing $t \in [0, \varepsilon]$ sufficiently small such that $1 - \frac{t}{2}(\beta\gamma + \frac{1}{p_1}) \geq \frac{1}{2}$, we can thus conclude that:

$$\psi_N(\mathbf{y}^{(t)}) - \psi_N(\mathbf{y}) \geq \frac{Nt\delta}{2}, \quad \text{i.e.} \quad \sup_{\mathbf{y} \in [p_1, p_2]^{Nq}: \|\nabla_{\cdot r}(\mathbf{y}) - \bar{\nabla}(\mathbf{y})\|^2 > N\delta} \psi_N(\mathbf{y}) \leq \sup_{\mathbf{y} \in \mathcal{S}_{N,q}} \psi_N(\mathbf{y}) - \frac{Nt\delta}{2}.$$

Therefore, on the event $F_r := \{\|\nabla_{\cdot r}(\boldsymbol{\theta}(\mathbf{X})) - \bar{\nabla}(\boldsymbol{\theta}(\mathbf{X}))\|^2 > N\delta\}$, we have:

$$\psi_N(\boldsymbol{\theta}(\mathbf{X})) \leq \sup_{\mathbf{y} \in \mathcal{S}_{N,q}} \psi_N(\mathbf{y}) - \frac{Nt\delta}{2}.$$

Therefore, we have by part (a),

$$\mathbb{P}(F_r) \leq \mathbb{P}\left(\sup_{\mathbf{y} \in \mathcal{S}_{N,q}} \psi_N(\boldsymbol{\theta}(\mathbf{X})) - \psi_N(\boldsymbol{\theta}(\mathbf{X})) \geq \frac{Nt\delta}{2}\right) = o(1)$$

which completes the proof of part (b). \square

The following is a technical lemma needed in the proof of Theorem 1.7.

Lemma I.9. *Suppose that $w_1, \dots, w_q \in (\alpha, 1]$ with $\alpha > 0$, and $\sum_{r=1}^q w_r = 1$. Then, for positive real numbers t_1, \dots, t_q bounded above by γ , we have:*

$$\max_r t_r - \sum_r w_r t_r \geq \frac{\alpha}{q(q-1)\gamma} \sum_{r < s} (t_r - t_s)^2.$$

Proof. Without loss of generality, suppose that $t_1 = \max_r t_r$ and $t_2 = \min_r t_r$. Then,

$$\begin{aligned} \max_r t_r - \sum_r w_r t_r &= t_1(1 - w_1) - t_2 w_2 - \sum_{r \geq 3} t_r w_r \\ &\geq t_1 \left(1 - w_1 - \sum_{r \geq 3} w_r\right) - t_2 w_2 \\ &= t_1 w_2 - t_2 w_2 \\ &\geq \alpha |t_r - t_s| \end{aligned}$$

for every r, s . Hence, using the fact $\max_{r < s} |t_r - t_s| \leq 2\gamma$, we have

$$\begin{aligned} \max_r t_r - \sum_r w_r t_r &\geq \frac{2\alpha}{q(q-1)} \sum_{r < s} |t_r - t_s| \\ &\geq \frac{\alpha}{q(q-1)\gamma} \sum_{r < s} (t_r - t_s)^2 \end{aligned}$$

This completes the proof of Lemma I.9. \square