The Annals of Applied Probability
2024, Vol. 34, No. 1A, 1–45
https://doi.org/10.1214/23-AAP1929

DETECTING STRUCTURED SIGNALS IN ISING MODELS

BY NABARUN DEB[1,a], RAJARSHI MUKHERJEE[2,d], SUMIT MUKHERJEE[1,b] AND MING YUAN[1,c]

[1]*Department of Statistics, Columbia University,* [a]*nd2560@columbia.edu,* [b]*sm3949@columbia.edu,* [c]*ming.yuan@columbia.edu*

[2]*Department of Biostatistics, Harvard University,* [d]*ram521@mail.harvard.edu*

In this paper we study the effect of dependence on detecting a class of signals in Ising models, where the signals are present in a structured way. Examples include Ising models on lattices, and mean-field type Ising models (Erdős–Rényi, Random regular, and dense graphs). Our results rely on correlation decay and mixing type behavior for Ising models, and demonstrate the beneficial behavior of criticality in detection of strictly lower signals. As a by-product of our proof technique, we develop sharp control on mixing and spin-spin correlation for several mean-field type Ising models in all regimes of temperature–which might be of independent interest.

1. Introduction. Understanding probabilistic models describing collections of dependent random variables arises across many scientific disciplines such as spatial statistics, biology, image segmentation, and social science, among others (Ahmed and Xing (2009), Geman and Geman (1984), Järpe (1999), Stauffer (2008)). For binary outcomes, the Ising model Ising (1925) constitutes the simplest yet one of the most fundamental frameworks to explore the effect of dependence on the collective behavior of Bernoulli random variables. In spite of its origins in statistical physics, the Ising model has enjoyed continued enthusiasm across the fields of applied probability, statistics, and computer science–and has provided us a with a rich theory at their intersection. Indeed, as has been demonstrated in recent literature, understanding statistical inference or exploring optimal algorithms for Ising models requires a marriage of ideas from applied probability (e.g., establishing correlation decay and concentration phenomenon) and statistical methodology (e.g., sharp analysis of pseudo-likelihood type methods). In a similar vein, in this paper, we consider a class of high-dimensional hypothesis testing problem involving dependent binary observations following an Ising model and completely characterize how such correlation bounds, which we establish for a large class of mean-field type models, imply different phase transitions of the hypothesis testing problem as a function of dependence in the models considered.

Specifically, we let $\mathbf{X} = (X_1, \dots, X_n)^\top \in \{\pm 1\}^n$ be a random vector with the joint distribution of $\mathbf{X}$ given by an Ising model defined as

$$\mathbb{P}_{\beta,\mathbf{Q},\boldsymbol{\mu}}(\mathbf{X} = \mathbf{x}) := \frac{1}{Z_n(\beta, \mathbf{Q}, \boldsymbol{\mu})} \exp\left(\frac{\beta}{2}\mathbf{x}^\top \mathbf{Q}\mathbf{x} + \boldsymbol{\mu}^\top \mathbf{x}\right), \quad \forall \mathbf{x} \in \{\pm 1\}^n. \tag{1}$$

Here $\mathbf{Q}$ is an $n \times n$ symmetric matrix with 0's on the diagonal, $\boldsymbol{\mu} := (\mu_1, \dots, \mu_n)^\top \in \mathbb{R}^n$ is an unknown parameter vector to be referred to as the external magnetization vector, $\beta \in \mathbb{R}$ is a real number usually referred to as the "inverse temperature," and $Z_n(\beta, \mathbf{Q}, \boldsymbol{\mu})$ is the normalizing constant. The pair $(\beta, \mathbf{Q})$ characterizes the dependence among the coordinates of $\mathbf{X}$, and X_i's are independent if $\beta\mathbf{Q} = \mathbf{0}_{n\times n}$. The most popular class of examples for $\mathbf{Q}$ would be *adjacency matrices* of some underlying graph, with an appropriate scaling. Examples such

Received December 2020; revised July 2022.
MSC2020 subject classifications. 62G10, 62G20, 62C20.
Key words and phrases. Ising model, signal detection, structured sparsity, correlation decay.

as complete graphs, Erdős–Rényi graphs, lattices, etc. have been studied extensively in the context of the Ising model described above (Dembo and Montanari (2010b), Duminil-Copin (2020), Ellis and Newman (1978)). We are interested in understanding the role of dependence $(\beta, \mathbf{Q})$ in testing against a collection of alternatives defined by a class of subsets $\mathcal{C}_n$ of $\{1, 2, \ldots, n\}$ each of which is of size s. More precisely, given any class of subsets $\mathcal{C}_n$ of $\{1, 2, \ldots, n\}$ of size $s \in [n]$, we consider testing the following hypotheses:

$$H_0 : \boldsymbol{\mu} = \mathbf{0} \quad \text{vs} \quad H_1 : \boldsymbol{\mu} \in \Xi(\mathcal{C}_n, s, A), \tag{2}$$

where

$$\Xi(\mathcal{C}_n, s, A) := \Big\{\boldsymbol{\mu} \in \mathbb{R}_+^n : \operatorname{supp}(\boldsymbol{\mu}) \in \mathcal{C}_n, \min_{i \in \operatorname{supp}(\boldsymbol{\mu})} \mu_i \geq A\Big\}, \quad \text{and}$$
$$\operatorname{supp}(\boldsymbol{\mu}) := \big\{i \in \{1, \ldots, n\} : \mu_i \neq 0\big\}.$$

Thus the class of alternatives $\Xi(\mathcal{C}_n, s, A)$ puts nonzero signals on one of the candidate sets in $\mathcal{C}_n$ where each signal set has size s. Throughout we shall assume that there exists a $\upsilon > 0$ such that $s \leq n^{1-\upsilon}$. However, some of our results go through for s as large as $\frac{n}{\log n}$. Finally, although we only consider one-directional signals, our results should go through for any noncritical β (see Section 3) for bi-directional signals as well.

Of primary interest here is to explore the effect of $(\beta, \mathbf{Q})$ on testing (2) for some structured signal classes $\mathcal{C}_n$. Examples of such signals will include geometric structures such as block signals on a lattice or suitable classes of low entropy signals (e.g., class of signals having enough disjoint sets—see Section 3.1 for precise definitions) on graphs with no inherent geometry. In this regard, previously, Arias-Castro, Candès and Durand (2011), Arias-Castro, Donoho and Huo (2005) studied the detection of block-sparse and thick shaped signals on lattices while Addario-Berry et al. (2010) considered general class of signals of combinatorial nature—however both these papers assume independent outcomes which corresponds to $\beta = 0$ in (1). Several other papers have also considered detection of contiguous signals over lattices and networks (see, e.g., Arias-Castro et al. (2018), Butucea and Ingster (2013), Enikeeva, Munk and Werner (2018), Hall and Jin (2008), Hall and Jin (2010), König, Munk and Werner (2020), Sharpnack, Rinaldo and Singh (2016), Walther (2010), Zou, Liang and Poor (2017) and references therein). On the other hand, if $\mathbf{Q}$ is taken to be the adjacency matrix of a line graph, then problem (2) can be viewed as a change point detection problem under dependence, which has also attracted a lot of attention in the literature; see Bhamidi, Jin and Nobel (2018), Cho (2016), Horváth and Hušková (2012), Liu, Gao and Samworth (2021), Ray and Tsay (2002). However, in overwhelming majority of the literature, the networks in question have only been used to describe the nature of signals—such as rectangles or thick clusters in lattices (Arias-Castro, Candès and Durand (2011)). A fundamental question however remains—"how does dependence characterized by a network modulate the behavior of such detection problems?" In this regard, Enikeeva et al. (2020) recently explored the effect of dependence on such detection problems for stationary Gaussian processes—with examples including linear lattices studied through the lens of Gaussian auto-regressive observation schemes. Dependence structures beyond Gaussian random variables are often more challenging to analyze (due to possible lack of closed form expressions of resulting distributions) and allow for interesting and different behavior of such testing problems—see, for example, Mukherjee, Mukherjee and Yuan (2018). One of the motivations of this paper is to fill this gap in the literature and show how dependent binary outcomes can substantially change the results for detecting certain classes of structured signals.

To this end, we adopt a standard asymptotic minimax framework as follows. Let a statistical test for H_0 versus H_1 be a measurable $\{0, 1\}$ valued function of the data $\mathbf{X}$, with

1 denoting rejecting the null hypothesis H_0 and 0 otherwise. The worst case risk of a test $T : \{\pm 1\}^n \to \{0, 1\}$ for testing (2) is defined as

$$\text{(3)} \qquad \operatorname{Risk}\big(T, \Xi(\mathcal{C}_n, s, A), \beta, \mathbf{Q}\big) := \mathbb{P}_{\beta,\mathbf{Q},\mathbf{0}}\big(T(\mathbf{X}) = 1\big) + \sup_{\boldsymbol{\mu} \in \Xi(\mathcal{C}_n, s, A)} \mathbb{P}_{\beta,\mathbf{Q},\boldsymbol{\mu}}\big(T(\mathbf{X}) = 0\big).$$

We say that a sequence of tests T_n corresponding to a sequence of model-problem pair (1) and (3), to be asymptotically powerful, asymptotically not powerful, and asymptotically powerless against $\Xi(\mathcal{C}_n, s, A)$ respectively, if

$$\limsup_{n\to\infty} \operatorname{Risk}\big(T_n, \Xi(\mathcal{C}_n, s, A), \beta, \mathbf{Q}\big) = 0,$$
$$\liminf_{n\to\infty} \operatorname{Risk}\big(T_n, \Xi(\mathcal{C}_n, s, A), \beta, \mathbf{Q}\big) > 0,$$
$$\liminf_{n\to\infty} \operatorname{Risk}\big(T_n, \Xi(\mathcal{C}_n, s, A), \beta, \mathbf{Q}\big) = 1.$$

The goal of the current paper is to characterize how the sparsity s and strength A of the signal jointly determine if there is an asymptotically powerful test, and how the behavior changes with $(\beta, \mathbf{Q})$. In this regard, the main results of this paper are summarized below.

- *General upper bounds*

  (I) For a general class of $(\beta, \mathbf{Q})$, we show that a scan statistic can detect a certain class of sparse signals (2) as soon as $\tanh(A) \gg \sqrt{\log n/s}$; see Theorem 1 (in fact the change happens at a constant level).

  (II) A natural question is what happens if $(\beta, \mathbf{Q})$ are unknown, that is, can we construct tests that are adaptive to the unknown parameters $(\beta, \mathbf{Q})$? In this direction, we get the same detection boundary as above if the unknown $(\beta, \mathbf{Q})$ satisfies some additional assumptions (see Theorem 2).

- *General lower bounds*

  (I) If the signal set has large cardinality ($s \geq C \log n$), we provide a general lower bound by showing that no test is asymptotically powerful, under assumptions on correlations between spins (see Theorem 3) for ferromagnetic Ising models, if the signal A is small.

  (II) The upper bound results suggest that testing is impossible if the signal set has small cardinality ($s \leq c \log n$). We confirm this intuition by showing that no test is asymptotically powerful irrespective of signal strength A for small s, in ferromagnetic Ising models (see Theorem 4).

- *Examples*

  (I) *Mean-field type Ising models:* We apply our general results to several popular examples of mean-field Ising models. These include Ising models on dense regular graphs, random regular graphs with "large" degree, and Erdős–Rényi graphs with "large" edge density. For $\beta \neq 1$ (which is the critical point for these Ising models), detection is impossible for small s for any value of A, and detection is possible for large s with the detection boundary $\tanh(A) \sim \sqrt{\frac{\log n}{s}}$. On the other hand, at criticality the detection boundary has three distinct regimes depending on the length of the signal set s, which we refer to informally as small, medium, and large. For s small, again no testing is possible for any signal strength A. For s medium, detection is possible with detection boundary $\tanh(A) \sim \sqrt{\frac{\log n}{s}}$, which is the same as the case $\beta \neq 1$. Finally if s is large, the detection boundary shifts to $\tanh(A) \sim \frac{n^{1/4}}{s}$ instead, and thus allows for detection of much smaller signals only for

$\beta = 1$. This improved upper bound at criticality for s large does not follow from Theorem 1 (which is based on a scan test), but instead utilizes a test based on sum of spins. The proof of the lower bound requires bounds on correlation between spins at all temperatures (see, e.g., Lemma 10 and other supporting results in Section 6). The proof of the upper bound at criticality follows from a careful analysis the sum of spins (see Lemma 12). To the best of our knowledge, these results are new, and might be of independent interest.

(II) *Ising models on lattices:* We show that for the classical Ising model on any fixed d-dimensional lattice the detection boundary again scales like $\tanh(A) \sim \sqrt{\frac{\log n}{s}}$ throughout the high temperature regime (right up to the critical temperature). The proof uses finite volume correlation decay, and ratio-scale mixing results. We note that similar arguments should apply in the low temperature positive pure-phase regime (plus boundary conditions). The case of free boundary conditions in the low temperature regime remains open.

1.1. *Organization.* The rest of the paper is organized as follows. In Section 2.1 we present some general upper bounds—including both the case of known and unknown dependence parameters $(\beta, \mathbf{Q})$. Section 2.2 contains a general lower bound result under ferromagnetic condition (positive β, $\mathbf{Q}$) and correlation decay type conditions. Subsequently, in Section 3 we apply these general results to demonstrate sharp upper and lower bounds for detecting signals in several commonly studied classes of Ising models. Section 4 contains the proofs of the main results from Sections 2.1 and 2.2. In Section 5, we present the proofs of the theorems stated in Section 3. Section 6 contains proofs of additional technical lemmas which may be of independent interest, pertaining to bounds on mixing, spin-spin correlations, and asymptotic analysis of the sum of spins at critical temperature.

1.2. *Notation.* Throughout, $\mathbb{E}_{\beta,\mathbf{Q},\boldsymbol{\mu}}$, $\mathrm{Var}_{\beta,\mathbf{Q},\boldsymbol{\mu}}$, $\mathrm{Cov}_{\beta,\mathbf{Q},\boldsymbol{\mu}}$ will denote the expectation, variance, and covariance operators corresponding to the measure $\mathbb{P}_{\beta,\mathbf{Q},\boldsymbol{\mu}}$. For a vector $\mathbf{X}$ following the Ising model as in (1), the ith coordinate will be denoted by X_i, and will often be referred to as the spin for vertex i. For a given sequence of symmetric matrices $\mathcal{Q} = \{\mathbf{Q}_{n\times n}\}_{n\geq 2}$ with nonnegative entries, we define the critical temperature as

$$\beta_c(\mathcal{Q}) = \inf\left\{\beta > 0 : \lim_{h\downarrow 0}\lim_{n\to\infty} \mathbb{E}_{\beta,\mathbf{Q},\boldsymbol{\mu}(h)}\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) > 0\right\}, \tag{4}$$

where $\boldsymbol{\mu}(h) = (h, \dots, h)^T \in \mathbb{R}^n$ denotes the vector with all coordinates equal to h. Similarly, for any $S \in \mathcal{C}_n$ and real number η let $\boldsymbol{\mu}_S(\eta)$ denote the vector which has $\mu_i = \eta\mathcal{I}(i \in S)$. In the examples pursued in the rest of the paper the existence of the limit is a part of classical statistical physics literature, and we shall note relevant references whenever talking about critical temperature in our examples. We shall refer to $(0, \beta_c(\mathcal{Q}))$ as high temperature regime, and $\beta \in (\beta_c(\mathcal{Q}), \infty)$ as low temperature regime.

For any $a, b \in \mathbb{N}$, we let $[a : b] = \{a, a+1, \dots, b\}$ and $[a] = \{1, \dots, a\}$. We also denote the m-dimensional 1-vector $(1, 1, \dots, 1) \in \mathbb{R}^m$ by $\mathbf{1}_m$. Also for any finite set S we use $|S|$ to denote the number of elements in S. For any two vectors $\mathbf{v}_1$, $\mathbf{v}_2$ of same dimension and $1 \leq p \leq \infty$ we let $\|\mathbf{v}_1 - \mathbf{v}_2\|_p$ denote the Euclidean ℓ_p norm. For any real matrix $\mathbf{M}$ and $1 \leq p \leq \infty$ we define the p-matrix norm of $\mathbf{M}$ as $\|\mathbf{M}\|_{p\to p} = \sup_{\|\mathbf{v}\|_p=1} \|\mathbf{M}\mathbf{v}\|_p$. For $p = 2$, we drop the the subscript to use $\|\mathbf{M}\|$ as the spectral norm of $\mathbf{M}$. For $p = \infty$ we use the fact that $\|\mathbf{Q}\|_{\infty\to\infty} = \sup_{i\in[n]} \sum_{j\in[n]} |\mathbf{Q}_{ij}|$. Finally for any vector $\mathbf{v} \in \mathbb{R}^m$ and subset $S \subset [m]$ we let $\mathbf{v}_S$ to be the $|S|$-dimensional vector obtained by restricting $\mathbf{v}$ to the coordinates in S. For $\mathbf{X} \sim \mathbb{P}_{\beta,\mathbf{Q},\boldsymbol{\mu}}$ as in (1), let $m_i = m_i(\mathbf{X}) := \sum_{j=1}^{n} \mathbf{Q}_{ij} X_j$.

The results in this paper are mostly asymptotic (in n) in nature and thus requires some standard asymptotic notation. If a_n and b_n are two sequences of real numbers then $a_n \ll$

b_n and $a_n = o(b_n)$ implies that $a_n/b_n \to 0$ as $n \to \infty$. Similarly $a_n \lesssim b_n$ and $a_n = O(b_n)$ implies that $\limsup_{n\to\infty} a_n/b_n < \infty$. We also say $a_n = \Theta(b_n)$ or $a_n \sim b_n$ if both $a_n = O(b_n)$ and $b_n = O(a_n)$. Finally for any integer $m \geq 1$ and real $0 \leq p \leq 1$ we let $\mathrm{Bin}(m, p)$ denote the distribution of a binomial distribution with m trials and success probability p.

2. General results.

We divide our general results into two subsections pertaining to upper and lower bounds.

2.1. *Upper bounds*: *General coupling matrix.*

Let us assume that the class of signals $\mathcal{C}_n$ satisfies

$$\log|\mathcal{C}_n| \leq C_u \log n, \quad |\mathcal{C}_n| \to \infty, \tag{5}$$

for some constant $C_u > 0$. We now begin with a general result which pertains to pinning down a signal strength necessary for detection in a general class of $\mathbf{Q}$. To describe the test, define for any $S \in \mathcal{C}_n$

$$L_S(\boldsymbol{\mu}) := \frac{1}{\sqrt{|S|}} \sum_{i\in S} (X_i - \tanh(\beta m_i + \mu_i)),$$

where $m_i = \sum_{j=1}^n \mathbf{Q}_{ij} X_j$. For a fixed $\delta \in (0, 1)$, consider the test rejects the null hypothesis when

$$L_n := \sup_{S\in\mathcal{C}_n} |L_S(\mathbf{0})| > 2(1 + \beta\|\mathbf{Q}\|_{\infty\to\infty})\sqrt{2(1+\delta)\log|\mathcal{C}_n|}.$$

THEOREM 1. *Suppose $\mathbf{X} \sim \mathbb{P}_{\beta,\mathbf{Q},\boldsymbol{\mu}}$ with any $\beta \in \mathbb{R}$ and $\mathbf{Q}$ such that $\|\mathbf{Q}\|_{\infty\to\infty} \leq C_u'$ for some constant C_u' and (5) holds. Consider testing hypotheses about $\boldsymbol{\mu}$ as described by (2). Then there exists a constant $C' > 0$ such that if $\tanh(A) \geq C'\sqrt{\frac{\log n}{s}}$, then the test based on L_n defined above is asymptotically powerful.*

REMARK 1 (Choosing δ as a function of level). The risk criterion used in this paper is the sum of the type I and the type II error. Our goal is to obtain minimax detection thresholds such that this sum converges to 0, that is, both the type I and the type II error converges to 0 (see equation (3) in the paper). In this framework, we do not control type I error at a prespecified significance level. Therefore, the tests are parametrized by some $\delta > 0$. This framework is quite popular and has been used in Arias-Castro, Candès and Durand (2011), Mukherjee, Mukherjee and Yuan (2018). It is also possible to change the criteria to control the type I error at level α and obtain minimax detection thresholds such that the type II error converges to 0. In that case, we should choose δ depending on α. For instance, in the test discussed in Theorem 1, we should choose

$$\delta \equiv \delta_n := \frac{\log 2 - \log\alpha}{\log|\mathcal{C}_n|}.$$

This will control the type I error at α without changing the minimax detection boundary (where the loss function is now the type II error).

The fact that the test that attains the performance claimed in Theorem 1 is, not surprisingly, a scan type procedure. However, to attain optimal separation rates across all regimes of dependence (i.e., β) we need to conditionally center the scanning elements instead of unconditional centering prevalent for hypothesis testing literature with independent outcomes. This is in fact, based on a maximum pseudo-likelihood score statistic. As an intuition, note that

$$\mathbb{P}_{\beta,\mathbf{Q},\boldsymbol{\mu}}(X_i = 1|X_j, j \neq i) = \frac{\exp(\beta m_i + \mu_i)}{\exp(\beta m_i + \mu_i) + \exp(-\beta m_i - \mu_i)}.$$

Therefore, the maximum (log) pseudo-likelihood function is given by

$$f_N^*(\boldsymbol{\mu}) := -N\log 2 + \sum_{i=1}^{n}\bigl[\bigl(X_i(\beta m_i + \mu_i)\bigr) - \log\cosh(\beta m_i + \mu_i)\bigr],$$

where we treat β, $\mathbf{Q}$ as fixed. Consequently $\frac{\partial f_N^*(\boldsymbol{\mu})}{\partial \mu_i}|_{\boldsymbol{\mu}=\mathbf{0}} = X_i - \tanh(\beta m_i)$. Therefore, the test statistic L_n can be viewed as

$$L_n = \sup_{S\in\mathcal{C}_n} \frac{1}{\sqrt{|S|}} \sum_{i\in S} \frac{\partial f_N^*(\boldsymbol{\mu})}{\partial \mu_i}\bigg|_{\boldsymbol{\mu}=\mathbf{0}}.$$

The above version of the scan test relies explicitly on the full knowledge of the null distribution. Especially this requires that β, $\mathbf{Q}$ are known. As we shall show in Section 3, this test is indeed optimal for any noncritical $\beta > 0$ for a large class of underlying graphs. If however, the values of $(\beta, \mathbf{Q})$ are unknown but "small", there exists a sequence of tests with the same detection thresholds as above, which does not depend on the knowledge of β, $\mathbf{Q}$.

Fixing $\delta, \eta \in (0, 1)$, consider the test which rejects the null hypothesis when

$$\tilde{L}_n := \sup_{S\in\mathcal{C}_n} \frac{1}{\sqrt{|S|}}\left|\sum_{i\in S} X_i\right| > \sqrt{\frac{(1+\delta)\log|\mathcal{C}_n|}{1-\eta}}.$$

THEOREM 2. *Suppose* $\mathbf{X} \sim \mathbb{P}_{\beta,\mathbf{Q},\boldsymbol{\mu}}$ *with* $0 \le \beta\|\mathbf{Q}\|_{\infty\to\infty} \le \eta$ *for some* $0 \le \eta < 1$ *fixed*, $\min_{i,j} \beta\mathbf{Q}_{i,j} \ge 0$ *and* (5) *holds*. *Consider testing hypotheses about* $\boldsymbol{\mu}$ *as described by* (2). *Then for* $C' > 0$ *large enough*, *if* $\tanh(A) \ge C'\sqrt{\frac{\log n}{s}}$, *the test based on* $\tilde{L}_n$ *above is asymptotically powerful*.

REMARK 2. We note that the condition $\eta < 1$ is sharp, in that the detection threshold when $\eta = 1$ can be very different compared to what is stated in Theorem 2, depending on the value of s. A counter example is provided by the Curie–Weiss model, (Ising model on the complete graph) for which $\eta = 1$ allows for the choice $\beta = \beta_c = 1$, which is the critical temperature for this model (see the discussions after Theorem 6 in Section 3 for exact details).

Examples of $\mathbf{Q}$ which satisfy the assumption stated in Theorem 1 and Theorem 2 include some prototypical examples of Ising models studied in literature. We discuss them in detail in Section 3.

2.2. *Lower bounds*: *Ferromagnetic models*. In this section we present results on lower bounds to demonstrate sharpness of Theorem 1 for Ising models having $\min_{i,j} \beta\mathbf{Q}_{ij} \ge 0$—traditionally referred to as ferromagnetic model. In this regard, according to Theorem 1, successful detection is possible by a conditionally centered scan test provided $\tanh(A) \ge C'\sqrt{\frac{\log n}{s}}$ for a constant $C' > 0$ which depends on the class of signals $\mathcal{C}_n$ and $\|\mathbf{Q}\|_{\infty\to\infty}$ through the constants C_u and C'_u. Since $\tanh(A) \in (-1, 1)$, it seems that one might need s to at least be of order $\log n$ for the success of this test. This intuition turns out to be true and there exists a phase transition in the possibility of testing depending on the behavior of s w.r.t $\log n$. In particular, there exists constants $0 < c \le C < \infty$ (depending on the problem sequence $(\beta, \mathbf{Q})$ and class of alternatives $\mathcal{C}_n$) such that the detection problem behaves differently depending on whether $s \le c\log n$ or $s \ge C\log n$. Before formally stating the relevant results, let us assume that there exists a constant $C_l > 0$ and some subcollection $\mathcal{C}'_n \subseteq \mathcal{C}_n$ of disjoint sets such that

$$\log|\mathcal{C}'_n| \ge C_l \log n. \tag{6}$$

Note that (6) immediately implies $\min(|\mathcal{C}_n|, |\mathcal{C}'_n|) \to \infty$, and so it need not be assumed separately. As the proofs of the results in the two regimes (s small/large) involve substantially different ideas, we divide their presentation in separate subsections.

2.2.1. *Large signal size* s. The following theorem will be used to verify sharpness of the upper bound presented in Theorem 1 for s large.

THEOREM 3. *Suppose* $\mathbf{X} \sim \mathbb{P}_{\beta,\mathbf{Q},\boldsymbol{\mu}}$ *such that* $\min_{i,j} \beta \mathbf{Q}_{ij} \geq 0$, $\|\mathbf{Q}\|_{\infty\to\infty} \leq C_u'$ *for some constant* $C_u' > 0$ *and* (6) *holds. Consider testing* (2). *Then there exists fixed constants* $c', C > 0$ *such that the following hold*:

(I) *Suppose there exists sequences* r_n, r_n' *diverging to* $+\infty$ *with* $r_n \geq C \log n$, *and*

(a) $\sup_{S\in\mathcal{C}_n'} \mathrm{Var}_{\beta,\mathbf{Q},\mathbf{0}}(\sum_{i\in S} X_i) \leq r_n$.
(b) $\sup_{S_1\neq S_2\in\mathcal{C}_n'} \mathrm{Cov}_{\beta,\mathbf{Q},\mathbf{0}}(\sum_{i\in S_1} X_i, \sum_{j\in S_2} X_j) = o(r_n')$.

Then all tests are asymptotically powerless when $\tanh(A) \leq c' \min\{\sqrt{\frac{\log n}{r_n}}, \sqrt{\frac{1}{r_n'}}\}$.

(II) *Suppose there exists an increasing set* $\Omega_n \subseteq \{-1,+1\}^n$ *and sequences* r_n, r_n' *diverging to* $+\infty$ *with* $r_n \geq C\log n$, *such that*:

(a) $\sup_{\eta\in[0,A]} \sup_{S\in\mathcal{C}_n'} \mathrm{Var}_{\beta,\mathbf{Q},\boldsymbol{\mu}_S(2\eta)}(\sum_{i\in S} X_i|\Omega_n) \leq r_n$.
(b) $\sup_{\eta\in[0,A]} \sup_{S_1\neq S_2\in\mathcal{C}_n'} |\mathrm{Var}_{\beta,\mathbf{Q},\boldsymbol{\mu}_{S_1\cup S_2}(\eta)}(\sum_{i\in S_1\cup S_2} X_i|\Omega_n) - \mathrm{Var}_{\beta,\mathbf{Q},\boldsymbol{\mu}_{S_1}(\eta)}(\sum_{i\in S_1} X_i|\Omega_n) - \mathrm{Var}_{\beta,\mathbf{Q},\boldsymbol{\mu}_{S_2}(\eta)}(\sum_{i\in S_2} X_i|\Omega_n)| = o(r_n')$,
(c) $\liminf_{n\to\infty} \mathbb{P}_{\beta,\mathbf{Q},\mathbf{0}}(\Omega_n) > 0$.

Then no test is asymptotically powerful if $\tanh(A) \leq c' \min\{\sqrt{\frac{\log n}{r_n}}, \sqrt{\frac{1}{r_n'}}\}$.

A few remarks are in order about the assumptions and implications of Theorem 3. The results are presented in two parts since they will eventually be applied (in Section 3) to prove sharp lower bounds for high and low temperature regimes separately. To fix ideas, let us focus on the case of the Curie–Weiss model, which corresponds to (1) with $\mathbf{Q} = \frac{1}{n}\mathbf{1}\mathbf{1}^\top$, that is,

$$\mathbb{P}_{\beta,\boldsymbol{\mu}}(\mathbf{X}=\mathbf{x}) = \frac{1}{Z_n(\beta,\boldsymbol{\mu})} \exp\left(\frac{\beta}{2n}\left(\sum_{i=1}^n x_i\right)^2 + \sum_{i=1}^n \mu_i x_i\right). \tag{7}$$

Under (7), when $\mu_i = 0$ for all i, the coordinates of $\mathbf{X}$ are exchangeable. Further, X_i's can be viewed as *independently distributed* conditioned on a appropriate latent variable (see Mukherjee, Mukherjee and Yuan ((2018), Lemma 3)). Consequently, the model (7) has a lot of additional structure. Therefore, let us try to illustrate Theorem 3 under (7).

Indeed, for $\beta < \beta_c = 1$ (β_c, the critical temperature, can be different in different examples) we will appeal to part I of the theorem while part II of the theorem will be used for the low temperature regime $\beta > \beta_c = 1$ with the increasing event Ω_n being $\bar{\mathbf{X}} \geq 0$. In fact, under (7), $\beta < 1$, it is easy to check that

$$\sup_{S\in\mathcal{C}_n'} \mathrm{Var}_{\beta,\mathbf{Q},\mathbf{0}}\left(\sum_{i\in S} X_i\right) \lesssim s,$$

$$\sup_{S_1\neq S_2\in\mathcal{C}_n'} \mathrm{Cov}_{\beta,\mathbf{Q},\mathbf{0}}\left(\sum_{i\in S_1} X_i, \sum_{j\in S_2} X_j\right) = o\left(\frac{s}{\log n}\right).$$

Consequently, we invoke Theorem 3, part (I), with $r_n = O(s)$ and $r_n' = O(s/\log n)$, to get that the detection boundary is $\tanh(A) \sim \sqrt{\frac{\log n}{s}}$. The same choices work if $\beta > 1$ and invoking Theorem 3, part (II), gives the same detection boundary. When $\beta = 1$, the correlation bounds are quite different, that is,

$$\sup_{S\in\mathcal{C}_n'} \mathrm{Var}_{\beta,\mathbf{Q},\mathbf{0}}\left(\sum_{i\in S} X_i\right) \lesssim s + \frac{s^2}{\sqrt{n}},$$

$$\sup_{S_1 \neq S_2 \in \mathcal{C}_n'} \mathrm{Cov}_{\beta,\mathbf{Q},\mathbf{0}}\Big(\sum_{i\in S_1} X_i, \sum_{j\in S_2} X_j\Big) \lesssim \frac{s^2}{\sqrt{n}}.$$

Both these bounds are potentially larger than when $\beta < 1$, as long as $s \gg \sqrt{n}/\log n$. This leads to a substantially different detection boundary at the critical point when s is "large."

In general, the main quantity that decides the validity of the lower bound presented above happens to be the correlation between spins X_i and X_j for suitable pairs i, j. Indeed, such correlation control is an area of active research and eventual verification of these conditions requires establishing correlation bounds on the graphs in our examples. We derive several such bounds in Section 6.

2.2.2. *Small signal size s.* The following theorem will be used to verify sharpness of the upper bound presented in Theorem 1 for s small.

THEOREM 4. *Suppose $\mathbf{X} \sim \mathbb{P}_{\beta,\mathbf{Q},\boldsymbol{\mu}}$ such that $\min_{i,j} \beta\mathbf{Q}_{ij} \geq 0$ and (6) holds. Consider testing (2). Then there exists $c > 0$ such that whenever $s \leq c\log n$, the following holds*:

(I) *If*

$$\lim_{n\to\infty} \sup_{S_1\neq S_2\in\mathcal{C}_n'} \left| \frac{\mathbb{P}_{\beta,\mathbf{Q},\mathbf{0}}(\mathbf{X}_{S_1} = 1, \mathbf{X}_{S_2} = 1)}{\mathbb{P}_{\beta,\mathbf{Q},\mathbf{0}}(\mathbf{X}_{S_1} = 1)\mathbb{P}_{\beta,\mathbf{Q},\mathbf{0}}(\mathbf{X}_{S_2} = 1)} - 1\right| = 0,$$

then all tests are asymptotically powerless irrespective of A.

(II) *If there exists an increasing set Ω_n such that $\liminf_{n\to\infty} \mathbb{P}(\Omega_n) > 0$ and the following holds*:

$$\lim_{n\to\infty} \sup_{S_1\neq S_2\in\mathcal{C}_n'} \left| \frac{\mathbb{P}_{\beta,\mathbf{Q},\mathbf{0}}(\mathbf{X}_{S_1} = 1, \mathbf{X}_{S_2} = 1|\Omega_n)}{\mathbb{P}_{\beta,\mathbf{Q},\mathbf{0}}(\mathbf{X}_{S_1} = 1|\Omega_n)\mathbb{P}_{\beta,\mathbf{Q},\mathbf{0}}(\mathbf{X}_{S_2} = 1|\Omega_n)} - 1\right| = 0,$$

then no test is asymptotically powerful irrespective of A.

A few remarks are in order regarding the conditions and results in Theorem 4. Once again we have presented the results in two parts since they will be applicable to two different regimes of β. Also, we have once more kept the conditions of the theorem somewhat general so that we can verify them across different classes of Ising models on different graph families in Section 3. As for the intuition behind the conditions, the main idea behind the expressions appearing in the statements of Theorem 4 is to capture a notion of dependence between disjoint groups of coordinates of the Ising model and hence conveys a similar essence to the correlation decay conditions presented in Theorem 3. Instead of correlation decay, here we need to verify precise deviations as specified by these mixing type conditions. Indeed, under these conditions, Theorem 4 verifies our intuition from Theorem 1 that no signal can be detected for $s \leq c\log n$ for suitably small $c > 0$. In Section 3, we verify the strong mixing type conditions in Theorem 4 for different classes of Ising models.

3. Specific examples. In this section we provide examples of coupling strengths $(\beta, \mathbf{Q})$ and class of signals $\mathcal{C}_n$ which when looked through the lens of Theorems 1, 3, and 4 yield matching (in terms of rate) upper and lower bounds.

3.1. *Mean-field interactions.* In this section we verify the validity and optimality of our upper and lower bounds for some examples of mean-field type Ising models. Mean-field Ising models can be generally characterized by positing conditions on $\mathbf{Q}$ under which mean-field approximation holds (see, e.g., Basak and Mukherjee (2017), Jain, Koehler and Mossel

(2018) for exact definitions and details). Our results on rate optimal detection boundaries will be verified for some important subclasses of such mean-field Ising models. Before stating these results, let us present the rates for the $\beta = 0$, that is, the independent case. This will serve as a useful baseline for comparison later.

PROPOSITION 5. *Suppose* $\beta = 0$ *in* (1). *Then there exists constants* $c, C, c', C' > 0$ *such that if*:

(a) $s \le c \log n$, *all tests are asymptotically powerless for any* $A > 0$.

(b) $s \ge C \log n$, *then all tests are asymptotically powerless if* $\tanh(A) \le c' \sqrt{\frac{\log n}{s}}$. *On the other hand, if* $\tanh(A) \ge C' \sqrt{\frac{\log n}{s}}$, *then the sequence of tests based on* L_n, *as described in Theorem* 1, *is asymptotically powerful*.

The proof of the above Proposition follows from Arias-Castro, Candès and Durand (2011)). It is also a direct consequence of Theorem 6, and hence we omit the details. With the $\beta = 0$ case in mind, let us proceed with the more nontrivial examples. In the rest of the section, throughout, the matrix $\mathbf{Q}$ will usually be associated with a certain sequence of simple labeled graphs $\mathbb{G}_n = (\mathcal{V}_n, \mathcal{E}_n)$ with vertex set $\mathcal{V}_n = \{1, \dots, n\}$ and edge set $\mathcal{E}_n \subseteq \mathcal{V}_n \times \mathcal{V}_n$ and corresponding $\mathbf{Q} = \mathbf{G}_n / \overline{d}$ where we define the average degree of the graph $\mathbb{G}_n$ to be $\overline{d} = |\mathcal{E}_n| / |\mathcal{V}_n|$. Here $\mathbf{G}_n$ is the adjacency matrix of $\mathbb{G}_n$. Also given any square matrix $\mathbf{M}$, we use $\lambda_i(\mathbf{M})$ to denote the ith largest eigenvalue of $\mathbf{M}$.

Since such models have no apparent geometry, there is less restriction on the choice of signal classes $\mathcal{C}_n$. Consequently, in this section we discuss testing against sparse alternatives of size s defined by $\Xi(\mathcal{C}_n, s, A)$, where $\mathcal{C}_n$ is any collection of subsets of $\mathcal{V}_n$ of size s such that (5) and (6) hold.

In all the examples to be considered in this subsection, the critical temperature corresponds to $\beta_c = 1$ (see, e.g., Basak and Mukherjee (2017)), and we demonstrate a double phase transition on the detection boundary in terms of the signal size s at the critical temperature (compared to only one phase transition at $s \sim \log n$ for noncritical temperatures), at $s \sim \log n$ and $s \sim \sqrt{n}/\log n$. In particular, for $s \gtrsim \sqrt{n}/\log n$ the behavior of the testing problem changes at the critical temperature, and one is able to detect lower signals using a simple test based on total number of spins.

Our first example in this regard is for dense regular graphs.

THEOREM 6. *Suppose* $\mathbb{G}_n$ *corresponds to a* d_n*-regular graph, which is dense, that is,* $d_n = \Theta(n)$ *and* (5) *and* (6) *hold. Then there exists constants* $c, C > 0$ *such that if*,

(a) $s \le c \log n$, *then the following conclusions hold*:
   (I) *If* $\beta \le 1$, *all tests are asymptotically powerless for any* $A > 0$.
   (II) *If* $\beta > 1$, *no test is asymptotically powerful for any* $A > 0$, *provided* $\limsup_{n \to \infty} \frac{\lambda_2(\mathbf{G}_n)}{d_n} < 1$.

(b) $s \ge C \log n$, *then there exists constants* $c', C' > 0$ *such that the following conclusions hold*:
   (I) *If* $\beta < 1$, *all tests are asymptotically powerless if* $\tanh(A) \le c' \sqrt{\frac{\log n}{s}}$. *On the other hand, if* $\tanh(A) \ge C' \sqrt{\frac{\log n}{s}}$, *there the sequence of tests based on* L_n *as described in Theorem* 1, *is asymptotically powerful*.
   (II) *Suppose* $\beta = 1$ *and* $\limsup_{n \to \infty} \lambda_2(\mathbf{G}_n)/d_n < 1$.
      i. *If* $s \ll \sqrt{n}/\log n$, *all tests are asymptotically powerless if* $\tanh(A) \le c' \sqrt{\frac{\log n}{s}}$, *and the sequence of tests based on* L_n *is asymptotically powerful if* $\tanh(A) \ge C' \sqrt{\frac{\log n}{s}}$.

ii. *If* $s \gtrsim \sqrt{n}/\log n$, *all tests are asymptotically powerless if* $s\tanh(A) \ll n^{1/4}$ *and there exists a* $\delta > 0$ *such that the test which rejects when* $n^{1/4}\bar{\mathbf{X}} \geq \delta(s\tanh(A)/n^{1/4})^{1/3}$, *is asymptotically powerful if* $s\tanh(A) \gg n^{1/4}$.

(III) *If* $\beta > 1$ *and* $\limsup_{n\to\infty} \lambda_2(\mathbf{G}_n)/d_n < 1$, *there are no asymptotically powerful tests if* $\tanh(A) \leq c'\sqrt{\frac{\log n}{s}}$. *On the other hand, if* $\tanh(A) \geq C'\sqrt{\frac{\log n}{s}}$, *the sequence of tests based on* L_n *is asymptotically powerful.*

Theorem 6, which includes the classical Curie–Weiss model as a special case (see (7)), demonstrates the benefit of critical temperature in detecting lower signals for $s \gtrsim \sqrt{n}/\log n$. It should be mentioned that a lot of existing literature on testing in the context of Ising models at critical temperature with mean-field interactions, focuses on the Curie–Weiss case. This case has a lot of additional structure (see Mukherjee, Mukherjee and Yuan ((2018), Lemma 3)), which is not present in general dense graphs, that are covered in Theorem 6 above. Therefore, the proof of Theorem 6 requires considerably different techniques.

Note that, for any $\beta \neq \beta_c$, the detection thresholds resemble that of $\beta = 0$ (i.e., independent observations), see Proposition 5; and only for $\beta = \beta_c$ one can detect lower signals—and that too only when the number of signals is large enough. Proving that the detection thresholds are the same as the independent case, whenever $\beta \neq \beta_c$, is itself a very challenging exercise. This is particularly evident if $\beta > \beta_c$ when the correlations between spins, that is, $\mathrm{Corr}(X_i, X_j)$ is asymptotically $O(1)$ as opposed to the independent case where this is 0. Overcoming this challenge requires new correlation decay arguments which may be of independent interest; see Lemma 13.

REMARK 3 (Computational complexity). The computational complexity of our test depends on the size of $|\mathcal{C}_n|$, and by our assumption $|\mathcal{C}_n| \leq n^C$ for some $C > 0$ (see (5)). Thus our computational complexity is at most polynomial in n. Note that even in the independent case ($\beta = 0$), the same scan test (with the same computational complexity) is known to be optimal (Arias-Castro, Candès and Durand (2011)), and so one cannot hope for a test with much better computational complexity in all regimes. Note, however, that when $\beta = 1$ and $s \gtrsim \sqrt{n}/\log n$, we show that the optimal test is based on the sum $\sum_{i=1}^n X_i$, which is computationally much more feasible. Such a phenomenon is not present in the independent case, and can be seen as a computational gain because of dependence.

The assumption of denseness of the regular graph can be removed under randomness. In particular, a similar result holds for sparser but random regular graphs. We state this in our next result.

THEOREM 7. *Suppose* $\mathbb{G}_n$ *is the adjacency matrix of a* d_n *random regular graph, with*

$$\theta := \liminf_{n\to\infty} \frac{\log d_n}{\log n} \in [0, 1]$$

and (5), (6) *hold. Then there exists fixed constants* $c, C > 0$ *such that if,*

(a) $s \leq c\log n$, *then the following conclusions hold*:
  (I) *If* $\beta \leq 1$, *all tests are asymptotically powerless for any* A, *provided* $\theta > 0$.
  (II) *If* $\beta > 1$, *no test is asymptotically powerful for any* A, *provided* $\theta > 0$.

(b) $s \geq C\log n$, *then there exists constants* $c', C' > 0$ *such that the following conclusions hold*:
  (I) *If* $\beta < 1$, *all tests are asymptotically powerless if* $\tanh(A) \leq c'\sqrt{\frac{\log n}{s}}$ *and* $\theta > 1/2$. *On the other hand, if* $\tanh(A) \geq C'\sqrt{\frac{\log n}{s}}$, *there is a sequence of asymptotically powerful tests for any* $\theta \geq 0$.

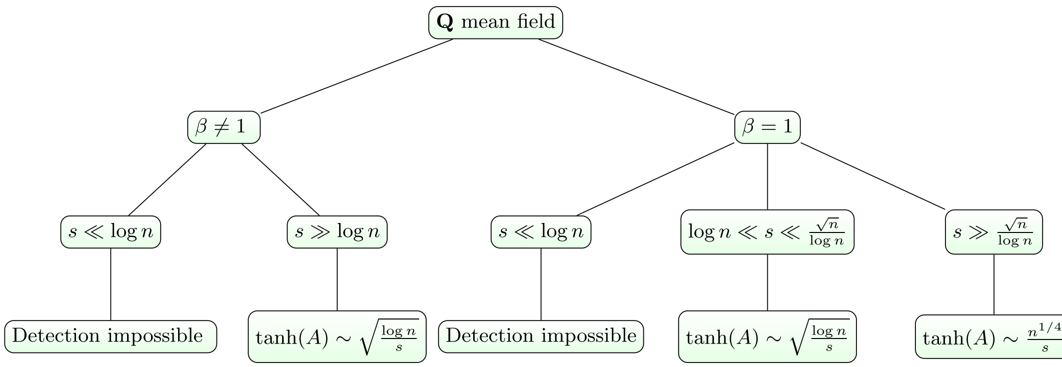

FIG. 1. *Summary of detection boundary in mean field models.*

(II) *Suppose* $\beta = 1$.

i. *If* $s \ll \sqrt{n}/\log n$, *all tests are asymptotically powerless if* $\tanh(A) \le c'\sqrt{\frac{\log n}{s}}$, $\theta > 1/2$ *and there is a sequence of asymptotically powerful tests if* $\tanh(A) \ge C'\sqrt{\frac{\log n}{s}}$ *and* $\theta \ge 0$.

ii. *If* $s \gtrsim \sqrt{n}/\log n$ *and* $\theta > 1/2$, *then all tests are asymptotically powerless if* $s\tanh(A) \ll n^{1/4}$, *and there exists a sequence of asymptotically powerful tests if* $s\tanh(A) \gg n^{1/4}$.

(III) *If* $\beta > 1$, *there are no asymptotically powerful tests if* $\tanh(A) \le c'\sqrt{\frac{\log n}{s}}$ *and* $\theta > 2/3$. *On the other hand, if* $\tanh(A) \ge C'\sqrt{\frac{\log n}{s}}$, *there is a sequence of asymptotically powerful tests for any* $\theta \ge 0$.

The minimax optimal upper bounds are attained by the same tests as in Theorem 6.

It is intuitive that the results for random regular graphs should naturally extend to suitable Erdős–Rényi graphs as well. This intuition is indeed correct—as verified by our next result.

THEOREM 8. *Suppose* $\mathbb{G}_n$ *is the adjacency matrix of an Erdős–Rényi random graph with parameter* p_n, *such that*

$$\theta := \liminf_{n\to\infty} \frac{\log(np_n)}{\log n} \in [0, 1]$$

as before and (5), (6) *hold. Then the same conclusions hold as in Theorem* 7 *except that every occurrence of the condition* $\theta \ge 0$ *in part* (b) *of Theorem* 7 *is replaced with* $\theta > 0$.

A summary of the detection boundary for mean field Ising models is given in the tree in Figure 1. Even though all the transitions happen at a constant level, we remove all constants to make the results more transparent.

The proofs of the theorems above, mostly rely on verifying the conditions of Theorems 1, 3, and 4 for the respective graphs. Only for $\beta = \beta_c = 1$ and $s \gtrsim \sqrt{n}/\log n$, the optimal upper bound does not follow from Theorem 1. In this case the optimal test is not based on a scan test but rather simply on the total magnetization $\sum_{i=1}^n X_i$. The sharp analysis of the test based $\sum_{i=1}^n X_i$ requires several additional technical details. We develop the necessary ingredients in Section 6.

3.2. *Short range interactions.* Indeed, the most classical example of an Ising model corresponds to nearest neighbor interactions on a lattice in dimension d (Ising (1925), Onsager

(1944)). To introduce this model it is convenient to rewrite the vertices of the graph as the vertices of a lattice as follows. Given positive integer d, consider a growing sequence of integer lattice hypercubes of dimension d defined by $\Lambda_n(d) := [-n^{1/d}, n^{1/d}]^d \cap \mathbb{Z}^d$, $n \geq 1$, where $\mathbb{Z}^d$ denotes the d-dimensional integer lattice. For any two distinct elements $i, j \in \Lambda_n(d)$ we put a weight $\mathbf{Q}_{ij}$, with the restriction that $\mathbf{Q}_{ij} = \mathbf{Q}_{ji}$. Thus $\mathbf{Q}$ is a symmetric array with zeros on the diagonal. With this notation, we say $\mathbf{Q}$ is short range if there exists $L \geq 1$ such that $\mathbf{Q}_{ij} = \mathbf{Q}_{ij}(\Lambda_n(d), L) = \mathcal{I}(0 < \|i - j\|_1 \leq L)$. Since such a model has an inherent geometry given by the lattic structure in d-dimensions, it is natural to consider signals which can be described by such geometry. Similar to one of the emblematic cases considered in Arias-Castro, Candès and Durand (2011), Arias-Castro, Donoho and Huo (2005), Butucea and Ingster (2013), König, Munk and Werner (2020), Walther (2010), here we discuss testing against block sparse alternatives of size s define by $\Xi(\mathcal{C}_n, s, A)$ with

$$\mathcal{C}_n = \left\{ \prod_{j=1}^{d} [a_j : b_j] \cap \Lambda_n(d) : \ b_j - a_j = \lceil s^{1/d} \rceil \right\}. \tag{8}$$

Although we only present the results for subcube detection in this paper, one can easily extend the results to detection of thick clusters (see Arias-Castro, Candès and Durand (2011) for details) with minor modifications of the arguments presented here.

We now argue that the conditions of Theorems 3 and 4 hold right up to the critical temperature in such model and class of signals problem pair and therefore we have sharp matching lower bounds corresponding to the upper bounds presented after Theorem 1. In the following analyses, we let $\beta_c(d, L)$ denote the critical temperature of an Ising model with $\mathcal{Q} = \mathbf{Q}(\Lambda_n(d), L)$ in (4). Although analytic forms of $\beta_c(d, L)$ are intractable for $d \geq 3$, the existence of such critical temperatures has been classically established—see for example, Duminil-Copin (2020), Ellis and Newman (1978), Friedli and Velenik (2018) for more details. Note that $\beta_c(d, L)$ can potentially be infinite.

THEOREM 9. *Let* $\mathbf{X} \sim \mathbb{P}_{\beta,\mathbf{Q},\boldsymbol{\mu}}$ *with* $\mathbf{Q}_{ij} = \mathcal{I}(0 < \|i - j\|_1 \leq L)$ *for* $i, j \in \Lambda_n(d)$ *and* $0 \leq \beta < \beta_c(d, L) \leq \infty$, *and consider testing* (2) *with* $\mathcal{C}_n$ *as in* (8). *Then there exists positive constants* $c, C > 0$ *depending on* β, L, d *such that the following hold.*

(a) *Suppose* $s \leq c \log n$. *Then all tests are asymptotically powerless irrespective of* A.
(b) *Suppose* $s \geq C \log n$. *Then there exists constants* $c', C' > 0$ *such that if* $\tanh(A) \geq C'\sqrt{\frac{\log n}{s}}$, *then the sequence of tests based on* L_n *in Theorem* 1 *is asymptotically powerful. On the other hand, if* $\tanh(A) \leq c'\sqrt{\frac{\log n}{s}}$, *then all tests are asymptotically powerless.*

We note that at the critical point $\beta = \beta_c(d, L)$ we do not expect Theorem 9 to hold, and the detection boundary to be lower (see the discussion in Mukherjee and Ray (2022) for heuristics in this regard).

We conclude this example by considering the case of one-dimensional Ising model that is, $d = 1$. This is the earliest studied Ising model and has $\mathbf{Q}$ correspond to the adjacency matrix of the line graph on n vertices (Ising (1925)). It is well known that the Ising model on the line graph does not exhibit a thermodynamic phase transition that is, $\beta_c(1, 1) = +\infty$ (see, e.g., Ising (1925) and Friedli and Velenik ((2018), Section 3.3)). As an immediate corollary to Theorem 3 and Theorem 4, we get that the detection boundary remains the same for any $\beta \geq 0$, and is the same as the independent case that is, $\beta = 0$.

COROLLARY 1. *Let* $\mathbf{X} \sim \mathbb{P}_{\beta,\mathbf{Q},\boldsymbol{\mu}}$ *with* $d = 1$, $\mathbf{Q}_{ij} = \mathcal{I}(|i - j| = 1)$ *for* $i, j \in \Lambda_n(d)$ *and* $\beta \in \mathbb{R}^+$, *and consider testing* (2) *with* $\mathcal{C}$ *as in* (8). *Then there exists positive constants* $c, C > 0$ *depending on* β *such that the following hold.*

(a) *Suppose* $s \leq c \log n$. *Then all tests are asymptotically powerless irrespective of* A.

(b) *Suppose* $s \geq C \log n$. *Then there exists constants* $c', C' > 0$ *such that if* $\tanh(A) \geq C'\sqrt{\frac{\log n}{s}}$, *then the sequence of tests described in Theorem* 1 *is asymptotically powerful. On the other hand, if* $\tanh(A) \leq c'\sqrt{\frac{\log n}{s}}$, *then all tests are asymptotically powerless.*

We note that for this particular case, a different proof using exact expressions for the log partition function can be used to show the validity of Corollary 1 for any $\beta \in \mathbb{R}$.

3.3. *Future scope.* In this paper we have explored how the level of dependence in Ising models can modulate the behavior of detection problems for testing certain structured anomalies. This minimax hypothesis testing problem provides a natural next step in a rich area of research of detecting contiguous signals of geometric nature—yet mostly under independence of the outcomes. Although we pinpoint the rates of minimax separation in this paper, we believe that there is a sharp constant threshold at which the transition happens from test-ability to non-test-ability (see, e.g., Arias-Castro, Candès and Durand (2011) for independent outcomes). In a different direction, one can study whether it is possible to attain optimal rates of detection for all β, if β is unknown (but $\mathcal{Q}$ is known, say).

4. Proofs of main results.

4.1. *Some supporting lemmas.* In this section we collect the lemmas (whose proofs we defer to Section 4.6) which will be used in the proofs of Theorems 1–4.

LEMMA 1 (GHS inequality (Lebowitz (1974))). *Suppose* $X \sim \mathbb{P}_{\beta,\mathbf{Q},\boldsymbol{\mu}}$ *with* $\beta > 0$, $\mathbf{Q}_{ij} \geq 0$ *for all* $i, j \in [n]$ *and* $\boldsymbol{\mu} \in (\mathbb{R}^+)^n$. *Then for any* $(i_1, i_2, i_3) \in [n]^{\otimes 3}$ *one has*

$$\frac{\partial^3 \log Z_n(\beta, \mathbf{Q}, \boldsymbol{\mu})}{\partial \mu_{i_1} \partial \mu_{i_2} \partial \mu_{i_3}} \leq 0.$$

Consequently, for any $\boldsymbol{\mu}_1 \succcurlyeq \boldsymbol{\mu}_2 \succcurlyeq \mathbf{0}$ (*i.e., coordinatewise inequality*) *one has*

$$\mathrm{Cov}_{\beta,\mathbf{Q},\boldsymbol{\mu}_1}(X_i, X_j) \leq \mathrm{Cov}_{\beta,\mathbf{Q},\boldsymbol{\mu}_2}(X_i, X_j), \tag{9}$$

whenever $\beta\mathbf{Q}_{ij} \geq 0$ *for all* $i, j \in [n]$.

LEMMA 2 (GKS inequality (Friedli and Velenik (2018))). *Suppose* $X \sim \mathbb{P}_{\beta,\mathbf{Q},\boldsymbol{\mu}}$ *with* $\beta > 0$, $\mathbf{Q}_{ij} \geq 0$ *for all* $i, j \in [n]$ *and* $\boldsymbol{\mu} \in (\mathbb{R}^+)^n$. *Then the following hold for any* $i, j \in [n]$:

$$\mathrm{Cov}_{\beta,\mathbf{Q},\boldsymbol{\mu}}(X_i, X_j) \geq 0; \qquad \mathbb{E}_{\beta,\mathbf{Q},\boldsymbol{\mu}}(X_i) \geq 0.$$

LEMMA 3 (Lemma 8 of Daskalakis, Dikkala and Kamath (2019)). *Suppose* $X^{(k)} \sim \mathbb{P}_{\beta^{(k)},\mathbf{Q}^{(k)},\mathbf{0}}$ *for* $k = 1, 2$ *with* $\beta^{(1)}\mathbf{Q}^{(1)}_{ij} \geq \beta^{(2)}\mathbf{Q}^{(2)}_{ij} \geq 0$ *for all* i, j. *Then*

$$\mathrm{Cov}_{\beta^{(1)},\mathbf{Q}^{(1)},\mathbf{0}}(X_i, X_j) \geq \mathrm{Cov}_{\beta^{(2)},\mathbf{Q}^{(2)},\mathbf{0}}(X_i, X_j), \quad \forall i, j.$$

LEMMA 4 (See Theorem 1.5 of Chatterjee (2007) and Lemma 1 of Mukherjee, Mukherjee and Yuan (2018)). *Let* $\mathbf{X} \sim \mathbb{P}_{\beta,\mathbf{Q},\boldsymbol{\mu}}$. *Then for any* $t > 0$ *and* $S \subset [n]$ *we have*

$$\mathbb{P}_{\beta,\mathbf{Q},\boldsymbol{\mu}}\big(|L_S(\boldsymbol{\mu})| > 2(1 + \beta\|\mathbf{Q}\|_{\infty\to\infty})t\big) \leq 2e^{-t^2/2},$$

where $L_S(\boldsymbol{\mu}) := \frac{1}{\sqrt{|S|}}\sum_{i\in S}(X_i - \tanh(\beta m_i + \mu_i))$ *with* $m_i = \sum_{j=1}^n \mathbf{Q}_{ij} X_j$.

LEMMA 5. *Let* $\mathbf{X} \sim \mathbb{P}_{\beta,\mathbf{Q},\boldsymbol{\mu}}$ *such that* $\beta\|\mathbf{Q}\|_{\infty\to\infty} < 1$ *and* $\beta \min_{i,j} \mathbf{Q}_{i,j} \geq 0$. *Then for any* $S \subset [n]$ *and* $t > 0$ *we have*

$$\mathbb{P}_{\beta,\mathbf{Q},\boldsymbol{\mu}}\left(|\tilde{L}_S(\boldsymbol{\mu})| > \frac{t}{\sqrt{1-\beta\|\mathbf{Q}\|_{\infty\to\infty}}}\right) \leq 2e^{-t^2},$$

where $\tilde{L}_S(\boldsymbol{\mu}) := \frac{1}{\sqrt{|S|}} \sum_{i\in S}(X_i - \mathbb{E}_{\beta,\mathbf{Q},\boldsymbol{\mu}} X_i)$.

LEMMA 6. *Consider a finite set* $\boldsymbol{\mu}_1, \ldots, \boldsymbol{\mu}_k$ *of* $(\mathbb{R}^+)^n$ *and let* π *be the uniform prior on them. If* $\beta\mathbf{Q}_{ij} \geq 0$ *for all* $i, j \in [n]$ *and* L_π *denotes the likelihood ratio of* (2) *w.r.t* π, *then* $\mathbb{E}_{\boldsymbol{\beta},\mathbf{Q},\mathbf{0}}(L_\pi^2)$, *viewed as a function of* $k \times n$ *coordinates of* $\boldsymbol{\mu}_1, \ldots, \boldsymbol{\mu}_k$ *is coordinatewise increasing.*

LEMMA 7. *Suppose* $X \sim \mathbb{P}_{\beta,\mathbf{Q},\boldsymbol{\mu}}$ *with* $\boldsymbol{\mu} \in (\mathbb{R}^+)^n$, *and* $\beta\mathbf{Q}_{i,j} \geq 0$ *for all* $i \neq j$. *Setting* $\rho := (1 - \tanh(\beta\|\mathbf{Q}\|_{\infty\to\infty})$ *we have* $\mathbb{E}_{\beta,\mathbf{Q},\boldsymbol{\mu}}(X_i) \geq \rho \tanh(\mu_i)$.

LEMMA 8. *For every* $n \geq 1$, *let* $(\mathcal{X}_n, \mathcal{F}_n)$ *be a measure space and* $\mathbb{P}_n$, $\mathbb{Q}_n$ *be two probability distributions on this measure space. Also assume that* $\mathbb{P}_n$ *and* $\mathbb{Q}_n$ *have densities* $\mathfrak{p}_n$ *and* $\mathfrak{q}_n$ *with respect to some common dominating measure. Define* $L_n := \mathfrak{q}_n/\mathfrak{p}_n$ *whenever the denominator is nonzero. Then the following conclusions hold*:

1. *If* $\mathbb{E}_{\mathbb{P}_n} L_n^2 = 1 + o(1)$, *then all tests are asymptotically powerless for testing* $\mathbb{P}_n$ *versus* $\mathbb{Q}_n$.
2. *If* $\mathbb{E}_{\mathbb{P}_n}[L_n^2|\Omega_n] = O(1)$ *for some event* Ω_n *for which*

$$\min\left\{\liminf_{n\to\infty} \mathbb{P}_n(\Omega_n), \liminf_{n\to\infty} \mathbb{Q}_n(\Omega_n)\right\} > 0,$$

then no test is asymptotically powerful for testing $\mathbb{P}_n$ *versus* $\mathbb{Q}_n$.

4.2. *Proof of Theorem* 1. By Lemma 4 we get

$$\mathbb{P}_{\beta,\mathbf{Q},\mathbf{0}}(|L_S(\mathbf{0})| > 2(1+\beta\|\mathbf{Q}\|_{\infty\to\infty})\sqrt{2(1+\delta)\log|\mathcal{C}_n|}) \leq 2\exp(-(1+\delta)\log|\mathcal{C}_n|).$$

A union bound then gives

$$\mathbb{P}_{\beta,\mathbf{Q},\mathbf{0}}(L_n > 2(1+\beta\|\mathbf{Q}\|_{\infty\to\infty})\sqrt{2(1+\delta)\log|\mathcal{C}_n|}) \leq \frac{2}{|\mathcal{C}_n|^\delta} \overset{n\to\infty}{\to} 0,$$

yielding a control over the Type I error of the test. In order to control the Type II error of the test suppose $\text{supp}(\boldsymbol{\mu}) = S$ for some $S \in \mathcal{C}_n$. Then we show that $|L_S(\mathbf{0})|$ beats the the null cut-off. To this end note that

$$|L_S(\mathbf{0})| \geq \left|\frac{1}{\sqrt{|S|}} \sum_{i\in S}(\tanh(\beta m_i + \mu_i) - \tanh(\beta m_i))\right| - |L_S(\boldsymbol{\mu})|.$$

Again by Lemma 4 we have

$$\mathbb{P}_{\beta,\mathbf{Q},\boldsymbol{\mu}}(|L_S(\boldsymbol{\mu})| \leq 2(1+\beta\|\mathbf{Q}\|_{\infty\to\infty})\sqrt{2(1+\delta)\log|\mathcal{C}_n|}) \geq 1 - 2\exp(-(1+\delta)\log|\mathcal{C}_n|).$$

On the other hand, observe that from elementary calculus,

$$\sup_{x\in[0,K], y\geq 0} \frac{\tanh(x+y) - \tanh(x)}{\tanh(y)} \gtrsim 1,$$

where $K > 0$ and the (hidden) constant on the right is strictly positive and depends on K. Therefore there exists a constant $M > 0$ (depending on β, $\|\mathbf{Q}\|_{\infty\to\infty}$) such that

$$\sum_{i\in S}(\tanh(\beta m_i + \mu_i) - \tanh(\beta m_i)) \geq M \tanh(A)|S|.$$

The desired control on the Type II error therefore follows on noting the following string of inequalities:

$$M\tanh(A)\sqrt{|S|}\geq C'\cdot M\sqrt{\log n}\geq 4\big(1+\beta\|\mathbf{Q}\|_{\infty\to\infty}\big)\sqrt{2(1+\delta)\log|\mathcal{C}_n|},$$

where the last inequality holds for all n large (using the fact that $\log|\mathcal{C}_n|\leq C_u\log n$) for $C':=\frac{8(1+\beta C'_u)\sqrt{C_u}}{M}$.

4.3. *Proof of Theorem* 2. Define for any $S\in\mathcal{C}_n$ set

$$\tilde{L}_S(\boldsymbol{\mu}):=\frac{1}{\sqrt{|S|}}\sum_{i\in S}(X_i-\mathbb{E}_{\beta,\mathbf{Q},\boldsymbol{\mu}}X_i),$$

and note that $\tilde{L}_n=\sup_{S\in\mathcal{C}_n}\tilde{L}_S(\mathbf{0})$.

Then, a union bound along with Lemma 5 gives

$$\mathbb{P}_{\beta,\mathbf{Q},\mathbf{0}}\bigg(\tilde{L}_n>\sqrt{\frac{(1+\delta)\log|\mathcal{C}_n|}{1-\eta}}\bigg)\leq\frac{2}{|\mathcal{C}_n|^{\delta}}\overset{n\to\infty}{\to}0,$$

yielding a control over the Type I error of the test.

Proceeding to control Type II error, with $\rho:=1-\tanh(\eta)$ we have for the signal set $S\in\mathcal{C}_n$

$$\big|\tilde{L}_S(\mathbf{0})\big|\geq\frac{1}{\sqrt{|S|}}\sum_{i\in S}\mathbb{E}_{\beta,\mathbf{Q},\boldsymbol{\mu}}(X_i)-\big|\tilde{L}_S(\boldsymbol{\mu})\big|\geq\rho\tanh(A)\sqrt{|S|}-\big|\tilde{L}_S(\boldsymbol{\mu})\big|,$$

where the last inequality uses Lemma 7. Also, again invoking Lemma 5 we have

$$\mathbb{P}_{\beta,\mathbf{Q},\boldsymbol{\mu}}\bigg(\big|\tilde{L}_S(\boldsymbol{\mu})\big|\leq\sqrt{\frac{(1+\delta)\log|\mathcal{C}_n|}{1-\eta}}\bigg)\geq 1-2\exp\big(-(1+\delta)\log|\mathcal{C}_n|\big).$$

This gives that

$$\tilde{L}_n\geq\tilde{L}_S(\mathbf{0})\geq\rho\tanh(A)\sqrt{|S|}-\sqrt{\frac{(1+\delta)\log|\mathcal{C}_n|}{1-\eta}}$$

with probability tending to 1 under $\mathbb{P}_{\beta,\mathbf{Q},\boldsymbol{\mu}}$, and so Type II error converges to 0 as soon as we have

$$\rho\tanh(A)\sqrt{s}\geq 2\sqrt{\frac{(1+\delta)\log|\mathcal{C}_n|}{1-\eta}},$$

which can be achieved by choosing $C'=\frac{4\sqrt{C_u}}{\rho\sqrt{1-\eta}}$ (since $\log|\mathcal{C}_n|\leq C_u\log n$), which depends only on η, C_u.

4.4. *Proof of Theorem* 3.

(I) Let $\mathcal{C}'_n$ be the subclass of $\mathcal{C}_n$ described in the statement of Theorem 3. Recall that for any $S\in\mathcal{C}_n$ and real number η, $\boldsymbol{\mu}_S(\eta)$ denotes the vector which has $\mu_i=\eta\mathcal{I}(i\in S)$. Let π denote the uniform prior on $\{\mathbb{P}_{\beta,\mathbf{Q},\boldsymbol{\mu}_S(A)},S\in\mathcal{C}'_n\}$ and let L_π be the likelihood ratio of π against $\mathbb{P}_{\beta,\mathbf{Q},\mathbf{0}}$. By Lemma 8, it suffices to show $\mathbb{E}_{\beta,\mathbf{Q},\mathbf{0}}(L_\pi^2)=1+o(1)$. it is easy to see that the corresponding second moment of the likelihood ratio is given by (owing to the disjointedness of the sets in $\mathcal{C}'_n$),

$$\begin{aligned}\mathbb{E}_{\beta,\mathbf{Q},\mathbf{0}}(L_\pi^2)&=\frac{1}{|\mathcal{C}'_n|^2}\sum_{S\in\mathcal{C}'_n}\frac{Z_n^2(\beta,\mathbf{Q},\mathbf{0})Z_n(\beta,\mathbf{Q},\boldsymbol{\mu}_S(2A))}{Z_n^2(\beta,\mathbf{Q},\boldsymbol{\mu}_S(A))Z_n(\beta,\mathbf{Q},\mathbf{0})}\\&\quad+\frac{1}{|\mathcal{C}'_n|^2}\sum_{S_1\neq S_2\in\mathcal{C}'_n}\frac{Z_n^2(\beta,\mathbf{Q},\mathbf{0})Z_n(\beta,\mathbf{Q},\boldsymbol{\mu}_{S_1\cup S_2}(A))}{Z_n(\beta,\mathbf{Q},\boldsymbol{\mu}_{S_1}(A))Z_n(\beta,\mathbf{Q},\boldsymbol{\mu}_{S_2}(A))Z_n(\beta,\mathbf{Q},\mathbf{0})}.\end{aligned}\tag{10}$$

Now for any $S \in \mathcal{C}_n'$, a two term Taylor expansion in A around 0 gives the existence of $\eta \in [0, A]$ (depending on S, A, β) such that

$$\begin{aligned}
&\frac{Z_n^2(\beta, \mathbf{Q}, \mathbf{0}) Z_n(\beta, \mathbf{Q}, \boldsymbol{\mu}_S(2A))}{Z_n^2(\beta, \mathbf{Q}, \boldsymbol{\mu}_S(A)) Z_n(\beta, \mathbf{Q}, \mathbf{0})} \\
&\quad = \exp\big(\log Z_n(\beta, \mathbf{Q}, \mathbf{0}) + \log Z_n\big(\beta, \mathbf{Q}, \boldsymbol{\mu}_S(2A)\big) - 2\log Z_n\big(\beta, \mathbf{Q}, \boldsymbol{\mu}_S(A)\big)\big) \\
&\quad = \exp\bigg(\frac{A^2}{2}\bigg[4\frac{\partial^2 \log Z_n(\beta, \mathbf{Q}, \boldsymbol{\mu}_S(2h))}{\partial h^2}|_{h=\eta} - 2\frac{\partial^2 \log Z_n(\beta, \mathbf{Q}, \boldsymbol{\mu}_S(h))}{\partial h^2}|_{h=\eta}\bigg]\bigg) \\
&\quad = \exp\bigg(\frac{A^2}{2}\bigg[4\mathrm{Var}_{\beta,\mathbf{Q},\boldsymbol{\mu}_S(2\eta)}\bigg(\sum_{i\in S} X_i\bigg) - 2\mathrm{Var}_{\beta,\mathbf{Q},\boldsymbol{\mu}_S(\eta)}\bigg(\sum_{i\in S} X_i\bigg)\bigg]\bigg) \\
&\quad \le \exp\bigg(2A^2\mathrm{Var}_{\beta,\mathbf{Q},\boldsymbol{\mu}_S(2\eta)}\bigg(\sum_{i\in S} X_i\bigg)\bigg).
\end{aligned}$$

Now the main challenge is to understand these spin-spin covariances at arbitrary magnetization η at locations S. To deal with this we employ GHS inequality (Lemma 1) to get that for any $\eta \ge 0$, any $S \in \mathcal{C}_n'$, and any i, j we have

$$\mathrm{Cov}_{\beta,\mathbf{Q},\boldsymbol{\mu}_S(\eta)}(X_i, X_j) \le \mathrm{Cov}_{\beta,\mathbf{Q},\mathbf{0}}(X_i, X_j).$$

Therefore, using the condition of the theorem we have that

$$\frac{Z_n^2(\beta, \mathbf{Q}, \mathbf{0}) Z_n(\beta, \mathbf{Q}, \boldsymbol{\mu}_S(2A))}{Z_n^2(\beta, \mathbf{Q}, \boldsymbol{\mu}_S(A)) Z_n(\beta, \mathbf{Q}, \mathbf{0})} \le \exp\bigg(2A^2\mathrm{Var}_{\beta,\mathbf{Q},\mathbf{0}}\sum_{i\in S} X_i\bigg) \le \exp(2A^2 r_n). \tag{11}$$

Next note that, once again for any $S_1, S_2 \in \mathcal{C}_n'$ we have for some $\eta \in [0, A]$ (possibly different) such that the following hold by a two term Taylor expansion in A around 0:

$$\begin{aligned}
&\frac{Z_n^2(\beta, \mathbf{Q}, \mathbf{0}) Z_n(\beta, \mathbf{Q}, \boldsymbol{\mu}_{S_1\cup S_2}(\eta))}{Z_n(\beta, \mathbf{Q}, \boldsymbol{\mu}_{S_1}(A)) Z_n(\beta, \mathbf{Q}, \boldsymbol{\mu}_{S_2}(A)) Z_n(\beta, \mathbf{Q}, \mathbf{0})} \\
&\quad = \exp\left(\frac{A^2}{2}\left[\begin{array}{c}\mathrm{Var}_{\beta,\mathbf{Q},\boldsymbol{\mu}_{S_1\cup S_2}(\eta)}\Big(\sum_{i\in S_1\cup S_2} X_i\Big) \\ -\mathrm{Var}_{\beta,\mathbf{Q},\boldsymbol{\mu}_{S_1}(\eta)}\Big(\sum_{i\in S_1} X_i\Big) - \mathrm{Var}_{\beta,\mathbf{Q},\boldsymbol{\mu}_{S_2}(\eta)}\Big(\sum_{i\in S_2} X_i\Big)\end{array}\right]\right).
\end{aligned}$$

Again by GHS inequality (Lemma 1) one has that for disjoint S_1, S_2 and $\eta \ge 0$

$$\mathrm{Var}_{\beta,\mathbf{Q},\boldsymbol{\mu}_{S_1\cup S_2}(\eta)}\bigg(\sum_{i\in S_1} X_i\bigg) \le \mathrm{Var}_{\beta,\mathbf{Q},\boldsymbol{\mu}_{S_1}(\eta)}\bigg(\sum_{i\in S_1} X_i\bigg),$$

$$\mathrm{Var}_{\beta,\mathbf{Q},\boldsymbol{\mu}_{S_1\cup S_2}(\eta)}\bigg(\sum_{i\in S_2} X_i\bigg) \le \mathrm{Var}_{\beta,\mathbf{Q},\boldsymbol{\mu}_{S_2}(\eta)}\bigg(\sum_{i\in S_2} X_i\bigg).$$

Consequently,

$$\begin{aligned}
&\frac{Z_n^2(\beta, \mathbf{Q}, \mathbf{0}) Z_n(\beta, \mathbf{Q}, \boldsymbol{\mu}_{S_1\cup S_2}(A))}{Z_n(\beta, \mathbf{Q}, \boldsymbol{\mu}_{S_1}(A)) Z_n(\beta, \mathbf{Q}, \boldsymbol{\mu}_{S_2}(A)) Z_n(\beta, \mathbf{Q}, \mathbf{0})} \\
&\quad \le \exp\bigg(\frac{A^2}{2}\sum_{i\in S_1, j\in S_2}\mathrm{Cov}_{\beta,\mathbf{Q},\boldsymbol{\mu}_{S_1\cup S_2}(\eta)}(X_i, X_j)\bigg) \qquad (12) \\
&\quad \le \exp\bigg(\frac{A^2}{2}\sum_{i\in S_1, j\in S_2}\mathrm{Cov}_{\beta,\mathbf{Q},\mathbf{0}}(X_i, X_j)\bigg),
\end{aligned}$$

where the second to last line follows, as before, by GHS inequality. Therefore, combining (10) and (11), we have

$$\mathbb{E}_{\beta,\mathbf{Q},\mathbf{0}}(L_\pi^2) \le \frac{1}{|\mathcal{C}_n'|^2}\sum_{S\in\mathcal{C}_n'}\exp(2A^2 r_n) + \frac{1}{|\mathcal{C}_n'|^2}\sum_{S_1\neq S_2\in\mathcal{C}_n'}\exp\big(o(A^2 r_n')\big). \tag{13}$$

As $\log|\mathcal{C}_n'| \ge C_l \log n$, there exists constant $c' > 0$ such that if $A \le c' \min\{\sqrt{\frac{\log n}{r_n}}, \frac{1}{\sqrt{r_n'}}\}$ then one has $\mathbb{E}_{\beta,\mathbf{Q},\mathbf{0}}(L_\pi^2) = 1 + o(1)$ by (13). Since $r_n \ge C\log n$, the same conclusion holds if $\tanh(A) \le c'\min\{\sqrt{\frac{\log n}{r_n}}, \frac{1}{\sqrt{r_n'}}\}$ for a different constant $c' > 0$.

(II) To prove this part of the theorem, we consider the same prior π as in part (I) and denote $P_{\beta,\mathbf{Q},\pi}$ to be the corresponding mixture of probability measures. Since Ω_n is an increasing event for every n, we also have by monotonicity of measures w.r.t. $\boldsymbol{\mu}$ that $\liminf_{n\to\infty}\mathbb{P}_{\beta,\mathbf{Q},\pi}(\Omega_n) > 0$.

By Lemma 8, part 2, it suffices to show that $\mathbb{E}_{\beta,\mathbf{Q},\mathbf{0}}(L_{\pi,\Omega_n}^2|\Omega_n)$ stays bounded, where L_{π,Ω_n} is the likelihood ratio of the probability measure $\mathbb{P}_{\beta,\mathbf{Q},\pi}(\cdot|\Omega_n)$ is contiguous w.r.t. $\mathbb{P}_{\beta,\mathbf{Q},\mathbf{0}}(\cdot|\Omega_n)$. To this end, one has that the conditional probability measure $\mathbb{P}_{\beta,\mathbf{Q},\boldsymbol{\mu}}(\cdot|\Omega_n)$ corresponding to any $\mathbb{P}_{\beta,\mathbf{Q},\boldsymbol{\mu}}$ in (1) is given by

$$\mathbb{P}_{\beta,\mathbf{Q},\boldsymbol{\mu}}(\mathbf{X}=\mathbf{x}|\Omega) = \frac{\exp(\frac{\beta}{2}\mathbf{x}^T\mathbf{Q}\mathbf{x} + \boldsymbol{\mu}^T\mathbf{X})\mathbf{1}(x\in\Omega_n)}{Z_n(\beta,\mathbf{Q},\boldsymbol{\mu}|\Omega_n)},$$

where $Z_n(\beta,\mathbf{Q},\boldsymbol{\mu}|\Omega_n) = \sum_{\mathbf{x}\in\Omega_n}\exp(\frac{\beta}{2}\mathbf{x}^T\mathbf{Q}\mathbf{x} + \boldsymbol{\mu}^T\mathbf{X})$. Therefore, by direct calculations similar to part (I) one has

$$\begin{aligned}
&\mathbb{E}_{\beta,\mathbf{Q},\mathbf{0}}\big(L_{\pi,\Omega_n}^2|\Omega_n\big)\\
&\quad= \frac{1}{|\mathcal{C}_n'|^2}\sum_{S\in\mathcal{C}_n'}\frac{Z_n^2(\beta,\mathbf{Q},\mathbf{0}|\Omega_n)Z_n(\beta,\mathbf{Q},\boldsymbol{\mu}_S(2A)|\Omega_n)}{Z_n^2(\beta,\mathbf{Q},\boldsymbol{\mu}_S(A)|\Omega_n)Z_n(\beta,\mathbf{Q},\mathbf{0}|\Omega_n)}\\
&\qquad+ \frac{1}{|\mathcal{C}_n'|^2}\sum_{S_1\neq S_2\in\mathcal{C}_n'}\frac{Z_n^2(\beta,\mathbf{Q},\mathbf{0}|\Omega_n)Z_n(\beta,\mathbf{Q},\boldsymbol{\mu}_{S_1\cup S_2}(\eta)|\Omega_n)}{Z_n(\beta,\mathbf{Q},\boldsymbol{\mu}_{S_1}(A)|\Omega_n)Z_n(\beta,\mathbf{Q},\boldsymbol{\mu}_{S_2}(A))Z_n(\beta,\mathbf{Q},\mathbf{0}|\Omega_n)} \qquad (14)\\
&\quad= \mathrm{I} + \mathrm{II}. \qquad (15)
\end{aligned}$$

Now, it is easy to check by direct calculations that for any $S\in\mathcal{C}_n'$,

$$\frac{\partial\log Z_n(\beta,\mathbf{Q},\boldsymbol{\mu}_S(h)|\Omega_n)}{\partial h}\bigg|_{h=\eta} = \mathbb{E}_{\beta,\mathbf{Q},\boldsymbol{\mu}_S(\eta)}\bigg(\sum_{i\in S}X_i|\Omega_n\bigg),$$

$$\frac{\partial^2\log Z_n(\beta,\mathbf{Q},\boldsymbol{\mu}_S(h)|\Omega_n)}{\partial h^2}\bigg|_{h=\eta} = \mathrm{Var}_{\beta,\mathbf{Q},\boldsymbol{\mu}_S(\eta)}\bigg(\sum_{i\in S}X_i|\Omega_n\bigg).$$

This implies that, the first term of (15) can be bounded similar to part (I) as

$$\mathrm{I} \le \frac{1}{|\mathcal{C}_n'|^2}\sum_{S\in\mathcal{C}'}\exp\bigg(4\frac{A^2}{2}\mathrm{Var}_{\beta,\mathbf{Q},\mu_S(2\eta)}\bigg(\sum_{i\in S}X_i|\Omega_n\bigg)\bigg) \tag{16}$$

for some $0\le\eta\le A$. Similarly, the second term of (15) can be written as

$$\mathrm{II} = \exp\left(\frac{A^2}{2}\left[\begin{array}{c}\mathrm{Var}_{\beta,\mathbf{Q},\boldsymbol{\mu}_{S_1\cup S_2}(\eta)}\bigg(\sum_{i\in S_1\cup S_2}X_i|\Omega_n\bigg)\\ -\mathrm{Var}_{\beta,\mathbf{Q},\boldsymbol{\mu}_{S_1}(\eta)}\bigg(\sum_{i\in S_1}X_i|\Omega_n\bigg) - \mathrm{Var}_{\beta,\mathbf{Q},\boldsymbol{\mu}_{S_2}(\eta)}\bigg(\sum_{i\in S_2}X_i|\Omega_n\bigg)\end{array}\right]\right). \tag{17}$$

The conclusion then follows using the given assumptions, in a similar manner as in part (I).

4.5. *Proof of Theorem* 4. The proof proceeds by controlling appropriate likelihood ratios in the same way as in the proof of Theorem 3.

(a) With $\pi(A)$ denoting the same prior as in the proof of Theorem 3, the likelihood ratio $L_{\pi(A)}$ is given by

$$L_{\pi(A)}(\mathbf{x}) = \frac{1}{|\mathcal{C}_n'|}\sum_{S\in\mathcal{C}_n'}\frac{\mathbb{P}_{\boldsymbol{\beta},\mathbf{Q},\boldsymbol{\mu}_S(A)}(\mathbf{X}=\mathbf{x})}{\mathbb{P}_{\boldsymbol{\beta},\mathbf{Q},\mathbf{0}}(\mathbf{X}=\mathbf{x})} = \frac{1}{|\mathcal{C}_n'|}\sum_{S\in\mathcal{C}_n'}\frac{\mathbb{P}_{\boldsymbol{\beta},\mathbf{Q},\boldsymbol{\mu}_S(A)}(\mathbf{X}_S=\mathbf{x}_S)}{\mathbb{P}_{\boldsymbol{\beta},\mathbf{Q},\mathbf{0}}(\mathbf{X}_S=\mathbf{x}_S)},$$

where the second inequality follows on noting that the conditional distribution of $\mathbf{X}_{S^c}$ given $\mathbf{X}_S$ is the same under both the measures $\mathbb{P}_{\beta,\mathbf{Q},\boldsymbol{\mu}_S(A)}$ and $\mathbb{P}_{\beta,\mathbf{Q},\mathbf{0}}$. On letting $A\to\infty$ gives

$$\lim_{A\to\infty} L_\pi(\mathbf{x}) = \frac{1}{|\mathcal{C}_n'|}\sum_{S\in\mathcal{C}_n'}\frac{\mathbf{1}(\mathbf{x}_s=\mathbf{1})}{\mathbb{P}_{\boldsymbol{\beta},\mathbf{Q},\mathbf{0}}(\mathbf{X}_S=\mathbf{x}_S)} =: L_\infty, \quad \text{say.}$$

Since, by Lemma 6, $\mathbb{E}_{\beta,\mathbf{Q},\mathbf{0}}L^2_{\pi(A)}$ is a nondecreasing function of A, and $L_{\pi(A)} \le \frac{1}{\mathbb{P}_{\beta,Q,0}(\mathbf{X}_S=\mathbf{x}_S)}$ which is bounded in A, using dominated convergence we have

$$\begin{aligned}
\mathbb{E}_{\beta,\mathbf{Q},\mathbf{0}}L^2_{\pi(A)} &\le \lim_{A\to\infty}\mathbb{E}_{\beta,\mathbf{Q},\mathbf{0}}L^2_{\pi(A)} \\
&= \mathbb{E}_{\beta,\mathbf{Q},\mathbf{0}}L^2_\infty \\
&= \frac{1}{|\mathcal{C}_n'|^2}\sum_{S\in\mathcal{C}_n'}\frac{1}{\mathbb{P}_{\boldsymbol{\beta},\mathbf{Q},\mathbf{0}}(\mathbf{X}_S=\mathbf{1})} \\
&\quad + \frac{1}{|\mathcal{C}_n'|^2}\sum_{S_1,S_2\in\mathcal{C}_n'}\frac{\mathbb{P}_{\boldsymbol{\beta},\mathbf{Q},\mathbf{0}}(\mathbf{X}_{S_1\cup S_2}=\mathbf{1})}{\mathbb{P}_{\boldsymbol{\beta},\mathbf{Q},\mathbf{0}}(\mathbf{X}_{S_1}=\mathbf{1})\mathbb{P}_{\boldsymbol{\beta},\mathbf{Q},\mathbf{0}}(\mathbf{X}_{S_2}=\mathbf{1})} \qquad (18)\\
&\le \frac{2^s}{|\mathcal{C}_n'|} + \sup_{S_1\ne S_2}\frac{\mathbb{P}_{\boldsymbol{\beta},\mathbf{Q},\mathbf{0}}(\mathbf{X}_{S_1\cup S_2}=\mathbf{1})}{\mathbb{P}_{\boldsymbol{\beta},\mathbf{Q},\mathbf{0}}(\mathbf{X}_{S_1}=\mathbf{1})\mathbb{P}_{\boldsymbol{\beta},\mathbf{Q},\mathbf{0}}(\mathbf{X}_{S_2}=\mathbf{1})}. \qquad (19)
\end{aligned}$$

The first term in the RHS of (18) is small since $\log|\mathcal{C}_n'| \ge C_l \log n$ and $s \le c\log n$ for a small enough $c>0$. The second term converges to 1 using the given hypothesis. Thus we have $\mathbb{E}L^2_\infty = 1+o(1)$, and so the proof is complete.

(b) It suffices to show that $\mathbb{E}_{\beta,\mathbf{Q},\mathbf{0}}L^2_{\pi(A)} = O(1)$. We can assume that $\mathbb{P}_{\beta,\mathbf{Q},\mathbf{0}}(\Omega_n) \ge \kappa/2$ for all large n and some $\kappa>0$. Using (18) it suffices to show that

$$\sup_{S_1\ne S_2\in\mathcal{C}_n'}\frac{\mathbb{P}_{\beta,\mathbf{Q},\mathbf{0}}(\mathbf{X}_{S_1}=\mathbf{1},\mathbf{X}_{S_2}=\mathbf{1})}{\mathbb{P}_{\beta,\mathbf{Q},\mathbf{0}}(\mathbf{X}_{S_1}=\mathbf{1})\mathbb{P}_{\beta,\mathbf{Q},\mathbf{0}}(\mathbf{X}_{S_2}=\mathbf{1})} \le \frac{4}{\kappa^2}+o(1). \qquad (20)$$

To this effect, for $i=1,2$ a simple inclusion gives

$$\mathbb{P}_{\beta,\mathbf{Q},\mathbf{0}}(\mathbf{X}_{S_i}=\mathbf{1}) \ge \mathbb{P}_{\beta,\mathbf{Q},\mathbf{0}}(\mathbf{X}_{S_i}=\mathbf{1},\Omega_n) \ge \frac{\kappa}{2}\mathbb{P}_{\beta,\mathbf{Q},\mathbf{0}}(\mathbf{X}_{S_i}=\mathbf{1}|\Omega_n),$$

and the FKG inequality subsequently gives

$$\mathbb{P}_{\beta,\mathbf{Q},\mathbf{0}}(\mathbf{X}_{S_1}=\mathbf{1},\mathbf{X}_{S_2}=\mathbf{1}) \le \mathbb{P}_{\beta,\mathbf{Q},\mathbf{0}}(\mathbf{X}_{S_1}=\mathbf{1},\mathbf{X}_{S_2}=\mathbf{1}|\Omega_n).$$

Combining these two observations along with the given hypothesis, (20) follows.

4.6. *Proofs of lemmas from Section* 4.1. The proofs of Lemmas 1–4 follow from the references cited in the statements themselves.

4.6.1. *Proof of Lemma* 5. A direct computation gives

$$d_{\mathrm{TV}}\big(\mathbb{P}^{(i)}(\cdot|\mathbf{x}_{-i}), \mathbb{P}^{(i)}(\cdot|\mathbf{y}_{-i})\big) = \frac{1}{2}\left|\tanh\Big(\beta\sum_j \mathbf{Q}_{ij}x_j + \mu_i\Big) - \tanh\Big(\beta\sum_j \mathbf{Q}_{ij}y_j + \mu_i\Big)\right|$$
$$\leq \beta\sum_j \mathbf{Q}_{ij}1\{x_j \neq y_j\}.$$

This, along with assumption $0 \leq \beta\|\mathbf{Q}\|_2 < 1$, on invoking Chatterjee ((2005), Theorem 4.3) gives the desired conclusion.

4.6.2. *Proof of Lemma* 6. Note that

$$\begin{aligned}
k^2\mathbb{E}_{\boldsymbol{\beta},\mathbf{Q},\mathbf{0}}(L_\pi^2) &= \sum_l \frac{Z_n(\beta,\mathbf{Q},\mathbf{0})Z_n(\beta,\mathbf{Q},2\boldsymbol{\mu}_l)}{Z_n^2(\beta,\mathbf{Q},\boldsymbol{\mu}_l)} + \sum_{l_1\neq l_2}\frac{Z_n(\beta,\mathbf{Q},\mathbf{0})Z_n(\beta,\mathbf{Q},\boldsymbol{\mu}_{l_1}+\boldsymbol{\mu}_{l_2})}{Z_n(\beta,\mathbf{Q},\boldsymbol{\mu}_{l_1})Z_n(\beta,\mathbf{Q},\boldsymbol{\mu}_{l_2})}\\
&= \sum_l \exp\big(\log Z_n(\beta,\mathbf{Q},\mathbf{0})\big) + \log Z_n(\beta,\mathbf{Q},2\boldsymbol{\mu}_l) - 2\log Z_n(\beta,\mathbf{Q},\boldsymbol{\mu}_l))\\
&\quad + \sum_{l_1\neq l_2}\exp\big(\log Z_n(\beta,\mathbf{Q},\mathbf{0})\big) + \log Z_n(\beta,\mathbf{Q},\boldsymbol{\mu}_{l_1}+\boldsymbol{\mu}_{l_2})\\
&\quad - \log Z_n(\beta,\mathbf{Q},\boldsymbol{\mu}_{l_1}) - \log Z_n(\beta,\mathbf{Q},\boldsymbol{\mu}_{l_2})),
\end{aligned}$$

fix any coordinate $l = 1,\dots,k$ and consider $k^2\mathbb{E}_{\boldsymbol{\beta},\mathbf{Q},\mathbf{0}}(L_\pi^2)$ as a function of the n coordinates of $\boldsymbol{\mu}_l$ fixing the rest of the coordinates. We note that it is enough to show that each coordinate of this gradient is nonnegative in the direction of any vector in $(\mathbb{R}^+)^n$. Being a sum of exponentials, it is sufficient to individually consider each exponent and show the same conclusion desired above. A typical such term is one of two types:

$$2\frac{\partial\log Z_n(\beta,\mathbf{Q},\boldsymbol{\mu})}{\partial\boldsymbol{\mu}}\bigg|_{\boldsymbol{\mu}=2\boldsymbol{\mu}_l} - 2\frac{\partial\log Z_n(\beta,\mathbf{Q},\boldsymbol{\mu})}{\partial\boldsymbol{\mu}}\bigg|_{\boldsymbol{\mu}=\boldsymbol{\mu}_l} \quad \text{or} \tag{21}$$

$$\frac{\partial\log Z_n(\beta,\mathbf{Q},\boldsymbol{\mu}+\boldsymbol{\mu}_{l'})}{\partial\boldsymbol{\mu}}\bigg|_{\boldsymbol{\mu}=\boldsymbol{\mu}_l} - \frac{\partial\log Z_n(\beta,\mathbf{Q},\boldsymbol{\mu})}{\partial\boldsymbol{\mu}}\bigg|_{\boldsymbol{\mu}=\boldsymbol{\mu}_l}, \quad l\neq l'. \tag{22}$$

However, by mean value theorem, the ith coordinate of $2\frac{\partial\log Z_n(\beta,\mathbf{Q},\boldsymbol{\mu})}{\partial\boldsymbol{\mu}}\big|_{\boldsymbol{\mu}=2\boldsymbol{\mu}_l} - 2\frac{\partial\log Z_n(\beta,\mathbf{Q},\boldsymbol{\mu})}{\partial\boldsymbol{\mu}}\big|_{\boldsymbol{\mu}=\boldsymbol{\mu}_l}$ equals $\boldsymbol{\mu}_l^T\mathrm{Var}_{\beta,\mathbf{Q},\boldsymbol{\eta}}(\mathbf{X})\mathbf{e}_i$ for some $\boldsymbol{\eta}$ lying on the line joining $\boldsymbol{\mu}_l$ and $2\boldsymbol{\mu}_l$ and $\mathbf{e}_i$ denoting the ith unit vector in $\mathbb{R}^n$. This implies that all coordinates of $\boldsymbol{\eta}$ are positive. Consequently, each coordinate of $\mathrm{Var}_{\beta,\mathbf{Q},\boldsymbol{\eta}}(\mathbf{X})$ is positive, since $\beta\mathbf{Q}_{j_1j_2} \geq 0$ for all $j_1, j_2 \in [n]$ which gives $\mathrm{Cov}_{\beta,\mathbf{Q},\boldsymbol{\eta}}(X_{j_1}X_{j_2}) \geq 0$ by GKS inequality (Lemma 2). This proves that the first term (21) has positive coordinates. A similar proof works for the second term (22).

4.6.3. *Proof of Lemma* 7. Note that

$$\begin{aligned}
\mathbb{E}_{\beta,\mathbf{Q},\boldsymbol{\mu}}(X_i) &= \mathbb{E}_{\beta,\mathbf{Q},\boldsymbol{\mu}}\tanh\Big(\beta\sum_{j\in[n]}\mathbf{Q}_{ij}X_j + \mu_i\Big)\\
&\geq \mathbb{E}_{\beta,\mathbf{Q},\boldsymbol{\mu}}\tanh\Big(\beta\sum_{j\in[n]}\mathbf{Q}_{ij}X_j\Big) + \big(1-\tanh\big(\beta\|\mathbf{Q}\|_{\infty\to\infty}\big)\tanh(\mu_i)\big),
\end{aligned}$$

from which the result follows on noting that

$$\mathbb{E}_{\beta,\mathbf{Q},\boldsymbol{\mu}}\tanh\Big(\beta\sum_{j\in[n]}\mathbf{Q}_{ij}X_j\Big) \geq \mathbb{E}_{\beta,\mathbf{Q},\mathbf{0}}\tanh\Big(\beta\sum_{j\in[n]}\mathbf{Q}_{ij}X_j\Big) = 0.$$

In the above display, the first inequality follows the fact that the Ising model is stochastically nondecreasing in $\boldsymbol{\mu}$, along with the observation that the function $(x_j, j \in [n]) \mapsto \tanh(\beta \sum_{j\in[n]} \mathbf{Q}_{ij} x_j)$ is nondecreasing, and the second equality follows by symmetry of the Ising model when $\boldsymbol{\mu} = \mathbf{0}$.

4.6.4. *Proof of Lemma 8*.

1. Let $\{\phi_n(\cdot)\}_{n\geq 1}$ be a sequence of $\{0, 1\}$-valued test functions and let $\mathcal{R}_n$ denote the corresponding rejection regions, that is, $\mathcal{R}_n := \{x \in \mathcal{X}_n : \phi_n(x) = 1\}$. Observe that the following equality holds:

$$\mathbb{P}_n(\mathcal{R}_n) + \mathbb{Q}_n(\mathcal{R}_n^c) = 1 - \int_{\mathcal{R}_n} (L_n - 1)\, d\mathbb{P}_n.$$

As $\mathbb{E}_{\mathbb{P}_n} L_n = 1$ and $\mathbb{E}_{\mathbb{P}_n} L_n^2 = 1 + o(1)$, we have $L_n \xrightarrow{\mathbb{P}_n} 1$. By the dominated convergence theorem,

$$\mathbb{E}_{\mathbb{P}_n} |L_n - 1| \mathbf{1}_{\mathcal{R}_n} \to 0.$$

Combining the two displays above, completes the proof.

2. We first claim that it is enough to prove that $\mathbb{Q}_n(\cdot|\Omega_n)$ is contiguous w.r.t. $\mathbb{P}_n(\cdot|\Omega_n)$ (where for any distribution $\mathbb{P}$, $\mathbb{P}(\cdot|\Omega)$ is used denote conditional distribution given $\mathbf{X} \in \Omega$). To verify this by contradiction, suppose we have a sequence of rejection regions $\mathcal{R}_n$ such that

$$\mathbb{P}_n(\mathcal{R}_n) \to 0, \qquad \mathbb{Q}_n(\mathcal{R}_n) \to 1.$$

We will show that if $\mathbb{Q}_n(\cdot|\Omega_n)$ is contiguous w.r.t. $\mathbb{P}_n(\cdot|\Omega_n)$ then $\limsup_{n\to\infty} \mathbb{Q}_n(\mathcal{R}_n) < 1$, which will give a contradiction. To see this, first note that since $\mathbb{P}_n(\mathcal{R}_n) \to 0$ and $\liminf_{n\to\infty} \mathbb{P}_n(\Omega_n) > 0$, we have $\mathbb{P}_n(\mathcal{R}_n|\Omega_n) \to 0$. Consequently, by contiguity one must have $\mathbb{Q}_n(\mathcal{R}_n|\Omega_n) \to 0$. Thus, writing

$$\mathbb{Q}_n(\mathcal{R}_n) = \mathbb{Q}_n(\mathcal{R}_n|\Omega_n)\mathbb{Q}_n(\Omega_n) + \mathbb{Q}_n(\mathcal{R}_n \cap \Omega_n^c),$$

the first term of the right hand side of the display above goes to 0 by contiguity and the second term satisfies

$$\limsup_{n\to\infty} \mathbb{Q}_n(\mathcal{R}_n \cap \Omega_n^c) \leq \limsup_{n\to\infty} \mathbb{Q}_n(\Omega_n^c) < 1.$$

It follows that $\limsup_{n\to\infty} \mathbb{Q}_n(\mathcal{R}_n) < 1$, as desired.

The final step is to show that $\mathbb{E}_{\mathbb{P}_n}[L_n^2|\Omega_n] = O(1)$ implies that $\mathbb{Q}_n(\cdot|\Omega_n)$ is contiguous w.r.t. $\mathbb{P}_n(\cdot|\Omega_n)$. So, we will show that if $\mathcal{A}_n$ is a sequence of events such that $\mathbb{P}_n(\mathcal{A}_n|\Omega_n) \to 0$, then $\mathbb{Q}_n(\mathcal{A}_n|\Omega_n) \to 0$. Towards this direction, observe that for any $K > 0$, the following holds:

$$\begin{aligned}\mathbb{Q}_n(\mathcal{A}_n \cap \Omega_n) &= \mathbb{E}_{\mathbb{P}_n}[L_n \mathbf{1}_{\mathcal{A}_n\cap\Omega_n} \mathbf{1}_{L_n\geq K}] + \mathbb{E}_{\mathbb{P}_n}[L_n \mathbf{1}_{\mathcal{A}_n\cap\Omega_n} \mathbf{1}_{L_n\leq K}] \\ &\leq K^{-1}\mathbb{E}_{\mathbb{P}_n}[L_n^2 \mathbf{1}_{\Omega_n}] + K\mathbb{P}_n(\mathcal{A}_n \cap \Omega_n).\end{aligned}$$

As $\mathbb{P}_n(\mathcal{A}_n|\Omega_n) \to 0$, the second term in the above display converges to 0 as $n \to \infty$ for every fixed K. As $\mathbb{E}_{\mathbb{P}_n}[L_n^2|\Omega_n] = O(1)$, the first term converges to 0 by taking $n \to \infty$, followed by $K \to \infty$. Therefore,

$$\limsup_{n\to\infty} \mathbb{Q}_n(\mathcal{A}_n|\Omega_n) \leq \frac{\limsup_{n\to\infty} \mathbb{Q}_n(\mathcal{A}_n \cap \Omega_n)}{\liminf_{n\to\infty} \mathbb{Q}_n(\Omega_n)} = 0.$$

This completes the proof.

5. Proofs of Theorems 6–9. This section will be devoted to proving Theorems 6–9. Towards that direction, we first mention a collection of lemmas, whose proofs we defer.

5.1. *Some auxiliary lemmas.* The first lemma describes some relevant properties of a fixed-point equation which arises naturally in mean-field Ising models (see Basak and Mukherjee (2017) for details) and will be useful for the subsequent discussion.

LEMMA 9 (See Page 10 in Dembo and Montanari (2010a)). *Consider the fixed point equation*

$$\phi(x) = 0, \quad \text{where } \phi(x) := x - \tanh(\beta x + B). \tag{23}$$

(a) (*High temperature*) *If* $\beta < 1$, *then* (23) *has a unique solution at* $t = 0$, *and* $\phi'(0) > 0$.

(b) (*Low temperature*) *If* $\beta > 1$, *then* (23) *has two nonzero roots* $\pm t$ *of this equation, where* $t > 0$, *and* $\phi'(\pm t) > 0$.

(c) (*Critical temperature*) *If* $\beta = 1$, *then* (23) *has a unique solution at* $t = 0$, *and* $\phi'(0) = 0$.

In the rest of the paper, t will always denote the nonnegative root of $\phi(\cdot)$ as defined in 9.

For the remaining results we need a few notation. For a graph $\mathbb{G}_n = (\mathcal{V}_n, \mathcal{E}_n)$ with vertex set $\mathcal{V}_n$ and edge-set $\mathcal{E}_n$, let $\mathbf{G}_n$ denote the adjacency matrix with its (i, j)th element denoted by $\mathbf{G}_n(i, j)$. Let d_i denote the degree of vertex $i \in \mathcal{V}_n$, $\overline{d}$ the average degree, and $d_{\max}$ the maximum degree. For an Ising model $\mathbb{P}_{\beta,\mathbf{Q},\boldsymbol{\mu}}$ defined on $\mathbb{G}_n$ we will use the convention that $\mathbf{Q} = \mathbf{G}/\overline{d}$. Finally, we denote the ith largest eigenvalue of a square matrix $\mathbf{M}$ by $\lambda_i(\mathbf{M})$. With these notation, the following lemma establishes sharp bounds on the spin-spin correlations for some mean-field type Ising models. These will serve as quintessential ingredients for verifying the conditions of Theorem 3 for the examples in Section 3.

LEMMA 10. *Let* $\alpha_n := \sqrt{\frac{\log n}{\overline{d}}}$ *and assume that* $\max_{i\in[n]} |\frac{d_i}{\overline{d}} - 1| \to 0$.

(a) *If* $0 \le \beta < 1$ *and* $\overline{d} \gtrsim (\log n)^2$ *then we have*

$$\big|\mathbb{E}_{\beta,\mathbf{Q},\mathbf{0}}(X_i X_j)\big| \lesssim \begin{cases} \dfrac{1}{\overline{d}} & \text{if } (i,j) \in \mathcal{E}_n, \\ \Big(\max_{(i,j)} (\mathbf{Q}^3)_{ij}\Big) + \alpha_n^4 & \text{if } (i,j) \notin \mathcal{E}_n, \end{cases}$$

where $\mathcal{E}_n$ *denotes the set of edges in* $\mathbb{G}_n$.

(b) *If* $\beta > 1$ *and* $\limsup_{n\to\infty} \max_{i\in[n]} \beta \operatorname{sech}^2(\beta t + \mu_i) < 1$ *and the assumptions in Lemma* 14 [*part* (*b*)(*i*) *or* (*b*)(*ii*)] *hold, then we have*:

(*i*)

$$|\mathrm{Cov}_{\beta,\mathbf{Q},\boldsymbol{\mu}}(X_i, X_j|\bar{\mathbf{X}} \ge 0) \lesssim \begin{cases} \dfrac{1}{\overline{d}} & \text{if } (i,j) \in \mathcal{E}_n, \\ \Big(\max_{(i,j)} (\mathbf{Q}^3)_{ij}\Big) + \alpha_n^3 & \text{if } (i,j) \notin \mathcal{E}_n. \end{cases}$$

(*ii*)

$$\max_{i\in[n]} \big|\mathrm{Var}_{\beta,\mathbf{Q},\boldsymbol{\mu}}(X_i|\bar{\mathbf{X}} \ge 0) - \operatorname{sech}^2(\beta t + \mu_i)\big| \lesssim \alpha_n.$$

(c) *If* $\beta = 1$, (39) *holds, and the graph* $\mathbb{G}_n$ *satisfies*

$$\overline{d} \gtrsim \sqrt{n}(\log n)^5, \qquad \max_{i\in[n]}\left|\frac{d_i}{\overline{d}} - 1\right| \lesssim \alpha_n,$$

then we have $|\mathbb{E}_{\beta,\mathbf{Q},\mathbf{0}}[X_i X_j]| \lesssim n^{-1/2}$.

Our next result establishes some crucial probability estimates for mean-field type Ising models. These will serve as quintessential ingredients for verifying the conditions of Theorem 4 for the examples in Section 3.

LEMMA 11. *Assume that* $\max_{i\in[n]} |d_i/\overline{d} - 1| \to 0$.

(a) *If* $0 \le \beta < 1$, *then for any fixed* $c > 0$ *and any* s *satisfying* $s \le c\log n$, *we have*

$$\lim_{n\to\infty} \sup_{S:|S|=s,\ \mathbf{a}\in\{-1,1\}^s} \left| \frac{\mathbb{P}_{\beta,\mathbf{Q},\mathbf{0}}(\mathbf{X}_S = \mathbf{a})}{2^{-s}} - 1 \right| = 0, \tag{24}$$

provided $\overline{d} \gg (\log n)^4$.

(b) *If* $\beta = 1$, *the same conclusion as part* (a) *holds provided* $\overline{d} \gg (\log n)^{10}$.

(c) *If* $\beta > 1$ *and the assumptions in either part* (b)(i) *or part* (b)(ii) *of Lemma* 14 *hold, then we get*

$$\begin{aligned} &\lim_{n\to\infty} \sup_{S:|S|=s,\mathbf{a}\in\{-1,1\}^s} \left| \frac{\mathbb{P}_{\beta,\mathbf{Q},\mathbf{0}}(\mathbf{X}_S = \mathbf{a}|\bar{\mathbf{X}} \ge 0)}{g(\mathbf{a}, s)} - 1 \right| = 0, \\ &g(\mathbf{a}, s) := \frac{\exp(\beta t \sum_{i=1}^s a_i)}{\sum_{\mathbf{b}\in\{-1,1\}^s} \exp(\beta t \sum_{i=1}^s b_i)}, \end{aligned} \tag{25}$$

provided $\overline{d} \gg (\log n)^4$.

Our final result in this section establishes precise behavior of average magnetization $\bar{X}$ at critical temperature for some mean-field type Ising models—under the presence of asymptotically vanishing, yet detectable, external magnetization $\boldsymbol{\mu}$. The application of this result for $s \gtrsim \sqrt{n}/\log n$ in these models yields matching sharp upper bounds to the lower bounds developed in Theorem 3.

LEMMA 12. *Suppose* $\boldsymbol{\mu} \in \Xi(\mathcal{C}_n, s, A)$, $\beta = 1$, $\overline{d} \gg \sqrt{n}(\log n)^5$, *and* $sA \gg n^{1/4}$. *Further assume that* $\max_{i\in[n]} |d_i/\overline{d} - 1| \lesssim \sqrt{\frac{\log n}{\overline{d}}}$, *and* $\limsup_{n\to\infty} \lambda_2(\mathbf{Q}) < 1$. *Then there exists a constant* $\delta > 0$ *such that*

$$\lim_{n\to\infty} \mathbb{P}_{\beta,\mathbf{Q},\boldsymbol{\mu}}\big(n^{1/4}\overline{\mathbf{X}} \ge \delta k_n\big) \to 1,$$

where $k_n := (n^{-1/4} s A)^{1/3}$.

Our final lemma concerns the correlation decay property in the context of Ising models on lattices which serves as the main tool in the proof of Theorem 9. We refer to Aizenman, Barsky and Fernández (1987), Duminil-Copin (2020), Duminil-Copin, Raoufi and Tassion (2019), Duminil-Copin and Tassion (2016), Mukherjee and Ray (2022) for more details. We use the notation used in Section 3.2 for denoting the vertices of the d-dimensional lattice.

LEMMA 13 (Correlation Decay). *Suppose* $X \sim \mathbb{P}_{\beta,\mathbf{Q},\mathbf{0}}$ *with* $\beta > 0$ *and* $\mathbf{Q}_{ij} = \mathcal{I}(0 < \|i - j\|_1 \le L)$ *for some* $L \ge 1$ *and* $i, j \in \Lambda_n(d)$ (*see Section* 3.2 *for precise definitions*). *Then there exists a* $\beta_c(d, L) > 0$ *such that for all* $0 \le \beta < \beta_c(d, L)$ *one has*

$$\mathrm{Cov}_{\beta,\mathbf{Q},\mathbf{0}}(X_i, X_j) \le \exp\big(-c(\beta, d, L)\|i - j\|_1\big), \tag{26}$$

for some $c(\beta, d, L) > 0$ *depending on* β, d, L.

5.2. *Proof of Theorem 6.*

(a) • *High temperature and critical point* ($0 \le \beta \le 1$)

In this case, note that the average degree $\bar{d} = d_n \gg (\log n)^\gamma$ for any $\gamma > 0$ and so we have $\mathbb{P}_{\beta,\mathbf{Q},\mathbf{0}}(\mathbf{X}_S = \mathbf{1}) = 2^{-|S|}(1 + o(1))$ for any S with $|S| \le 2s$, by Lemma 11 ((a) and (b)), where the $o(1)$ term depends on S only through its cardinality. Consequently we have

$$\sup_{S_1 \cap S_2 = \phi, |S_1| = |S_2| = s} \left| \frac{\mathbb{P}_{\beta,\mathbf{Q},\mathbf{0}}(\mathbf{X}_{S_1} = 1, \mathbf{X}_{S_2} = 1)}{\mathbb{P}_{\beta,\mathbf{Q},\mathbf{0}}(\mathbf{X}_{S_1} = 1)\mathbb{P}_{\beta,\mathbf{Q},\mathbf{0}}(\mathbf{X}_{S_2} = 1)} - 1 \right| = o(1).$$

The desired conclusion then follows by using Theorem 4, part (I).

• *Low temperature* ($\beta > 1$)

In this case we have

$$\mathbb{P}_{\beta,\mathbf{Q},\mathbf{0}}(\mathbf{X}_S = \mathbf{1} | \bar{\mathbf{X}} \ge 0) = \lambda_{|S|}^{-1} e^{\beta t |S|}\big(1 + o(1)\big),$$

where $\lambda_{|S|} := \sum_{\mathbf{b} \in \{-1,1\}^s} \exp(\beta t \sum_{i=1}^s b_i)$ for any set S with $|S| \le 2s$, by Lemma 11 (c). Consequently we have

$$\sup_{S_1 \cap S_2 = \phi, |S_1| = |S_2| = s} \left| \frac{\mathbb{P}_{\beta,\mathbf{Q},\mathbf{0}}(\mathbf{X}_{S_1} = 1, \mathbf{X}_{S_2} = 1 | \bar{\mathbf{X}} \ge 0)}{\mathbb{P}_{\beta,\mathbf{Q},\mathbf{0}}(\mathbf{X}_{S_1} = 1 | \bar{\mathbf{X}} \ge 0)\mathbb{P}_{\beta,\mathbf{Q},\mathbf{0}}(\mathbf{X}_{S_2} = 1 | \bar{\mathbf{X}} \ge 0)} - 1 \right| = o(1).$$

Next, set $\Omega_n := \{\bar{\mathbf{X}} \ge 0\}$ and note that $\liminf_{n\to\infty} \mathbb{P}_{\beta,\mathbf{Q},\mathbf{0}}(\Omega_n) \ge 1/2$ by symmetry. The desired conclusion then follows by invoking Theorem 4, part (II).

(b) • *High temperature* ($0 \le \beta < 1$) Since $\|\mathbf{Q}\|_{\infty\to\infty} = 1$ for regular graphs, the upper bounds follows from Theorem 1. For the lower bound, note that

$$\max_{i,j} (\mathbf{Q}^3)_{ij} = \max_{i,j} d_n^{-3} \sum_{k,l} \mathbf{Q}_{ik}\mathbf{Q}_{kl}\mathbf{Q}_{lj} \lesssim \frac{d_n^2}{d_n^3} \lesssim n^{-1}. \tag{27}$$

Combining the above observation with Lemma 10 (a), we have

$$\mathrm{Cov}_{\beta,\mathbf{Q},\mathbf{0}}(X_i, X_j) = \mathbb{E}_{\beta,\mathbf{Q},\mathbf{0}}(X_i X_j) \lesssim \frac{\log n}{n},$$

which immediately gives

$$\sup_{S \in \mathcal{C}_n} \mathrm{Var}_{\beta,\mathbf{Q},\mathbf{0}}\Big(\sum_{i \in S} X_i\Big) \le s + \sum_{i \ne j} \mathrm{Cov}_{\beta,\mathbf{Q},\mathbf{0}}(X_i, X_j) \le s + s^2 \frac{\log n}{n} \lesssim s, \tag{28}$$

where the last line follows from the fact that $s \le n^{1-\upsilon}$ for some $\upsilon > 0$ which is a standing assumption throughout the paper. For the same reason, we also have

$$\frac{\log n}{s} \sup_{S_1 \ne S_2} \sum_{i \in S_1, j \in S_2} \mathrm{Cov}_{\beta,\mathbf{Q},\mathbf{0}}(X_i, X_j) \lesssim s^2 \frac{\log n}{s} \frac{\log n}{n} = \frac{s(\log n)^2}{n} = o(1).$$

Thus, invoking Theorem 3, part (I) with $r_n := s$, $r_n' := \frac{s}{\log n}$ then shows that testing is impossible if $\tanh(A) \le c' \min\{\sqrt{\frac{\log n}{s}}, \sqrt{\frac{\log n}{s}}\} = c'\sqrt{\frac{\log n}{s}}$, for some small constant $c' > 0$, as desired.

• *Critical point* ($\beta = 1$) The upper bound for $s(\log n)/\sqrt{n} \ll 1$ follows from Theorem 1 as before. For $s \gtrsim \sqrt{n}/\log n$, it suffices to show that there is a sequence of asymptotically powerful tests for $sA \gg n^{1/4}$. Note that by Lemma 12, there exists a sequence $k_n := \delta(n^{-1/4}sA)^{1/3} \to \infty$ for some $\delta > 0$, such that $\mathbb{P}_{\beta,\mathbf{Q},\boldsymbol{\mu}}(n^{1/4}\bar{\mathbf{X}} \ge k_n) \to 1$ as $n \to \infty$. Further by Deb and Mukherjee ((2023), Theorem 1.3), $\mathbb{P}_{\beta,\mathbf{Q},\mathbf{0}}(n^{1/4}\bar{\mathbf{X}} \ge k_n) \to 0$ as $n \to \infty$. Therefore the test which rejects H_0 if $n^{1/4}\bar{\mathbf{X}} \ge k_n$ is asymptotically powerful.

For the lower bound, as before, using Lemma 10 (a) we have

$$\mathrm{Cov}_{\beta,\mathbf{Q},\mathbf{0}}(X_i, X_j) = \mathbb{E}_{\beta,\mathbf{Q},\mathbf{0}}(X_i X_j) \lesssim \frac{1}{\sqrt{n}} \quad \Rightarrow \quad \sup_{S \in \mathcal{C}_n} \sum_{i \in S} \mathrm{Var}_{\beta,\mathbf{Q},0}(X_i) \leq s + \frac{s^2}{\sqrt{n}}.$$

The same bound also gives

$$\sup_{S_1 \neq S_2} \sum_{i \in S_1, j \in S_2} \mathrm{Cov}_{\beta,\mathbf{Q},0}(X_i, X_j) \leq \frac{s^2}{\sqrt{n}}.$$

We will now consider two cases depending on the value of s.

– $s \ll \frac{\sqrt{n}}{\log n}$. In this case we can invoke Theorem 3 with $r_n := s$ and $r'_n := \frac{s}{\log n}$ as before to get the detection boundary $\tanh(A) \lesssim \sqrt{\frac{\log n}{s}}$.

– $s \gtrsim \frac{\sqrt{n}}{\log n}$. In this case setting $\epsilon_n := \frac{sA}{n^{1/4}}$, $r_n := s + \frac{s^2}{\sqrt{n}}$ and $r'_n := \frac{s^2}{\epsilon_n \sqrt{n}}$, we note that

$$\min\left(\sqrt{\frac{\log n}{r_n}}, \sqrt{\frac{1}{r'_n}}\right) = \frac{\sqrt{\epsilon_n} n^{1/4}}{s}, \tag{29}$$

where the last equality uses the fact that $s \gtrsim \frac{\sqrt{n}}{\log n}$ which in turn implies

$$\sqrt{\frac{\log n}{s}} \geq \frac{\sqrt{\epsilon_n} n^{1/4}}{2s}, \qquad \sqrt{\frac{\sqrt{n}\log n}{s^2}} \geq \frac{\sqrt{\epsilon_n} n^{1/4}}{2s},$$

for all large enough n. Combining (29) with Theorem 3, part (I) shows that testing is impossible if $sA = o(n^{1/4})$.

- *Low temperature ($\beta > 1$)*

As in the low temperature regime for part (a), set $\Omega_n := \{\bar{\mathbf{X}} \geq 0\}$. It is an increasing set of probability at least $1/2$ under $\mathbb{P}_{\beta,\mathbf{Q},\mathbf{0}}$ (by symmetry). Fix $S \subseteq [n]$ and set $\boldsymbol{\mu} \equiv \boldsymbol{\mu}_S(\eta)$ for some $\eta \in [0, 2A]$. For any $A \leq c'\sqrt{\log n/s}$ for some $c' > 0$, note that

$$\sum_{i=1}^n \mu_i \leq 2sA \leq 2c'\sqrt{s \log n} \leq 2c'\sqrt{n \log n},$$

$$\|\mathbf{Q}\boldsymbol{\mu}\|_\infty \leq \max_{i \in [n]} \sqrt{\sum_{j=1}^n \mathbf{Q}_{ij}^2} \sqrt{\sum_{j=1}^n \mu_j^2} \leq 2A\sqrt{\frac{s}{d_n}} \leq 2c'\sqrt{\frac{\log n}{d_n}},$$

which implies that the conditions for Lemma 14, part (b)(i) hold. Also by choosing $c' > 0$ small enough, using $s \geq C \log n$ and Lemma 9, we have, without loss of generality, $\limsup_{n\to\infty} \max_{i \in [n]} \beta \operatorname{sech}^2(\beta t + \mu_i) < 1$. Therefore, by Lemma 10, part (b)(i) and (27), we have $\mathrm{Cov}_{\beta,\mathbf{Q},\boldsymbol{\mu}}(X_i, X_j | \Omega_n) \lesssim \frac{\log n}{n}$ uniformly in i, j. Therefore we can choose $r_n = s$ by the same argument as in (28). With $\tilde{S} := S_1 \cup S_2$, observe that

$$\begin{aligned}\mathrm{Var}_{\beta,\mathbf{Q},\boldsymbol{\mu}_{\tilde{S}}(\eta)}\Big(\sum_{i \in S_1 \cup S_2} X_i | \Omega_n\Big) &= \mathrm{Var}_{\beta,\mathbf{Q},\boldsymbol{\mu}_{\tilde{S}}(\eta)}\Big(\sum_{i \in S_1} X_i | \Omega_n\Big) + \mathrm{Var}_{\beta,\mathbf{Q},\boldsymbol{\mu}_{\tilde{S}}(\eta)}\Big(\sum_{i \in S_2} X_i | \Omega_n\Big) \\ &\quad + \mathrm{Cov}_{\beta,\mathbf{Q},\boldsymbol{\mu}_{\tilde{S}}(\eta)}\Big(\sum_{i \in S_1} X_i, \sum_{i \in S_2} X_i | \Omega_n\Big) \\ &= 2s \operatorname{sech}^2(\beta t + \eta) + O\big(s\sqrt{\log n/d_n} + s^2 \log n/n\big),\end{aligned}$$

where the last line follows from Lemma 10, part (b), (i), and (ii), and the error term is uniform over S_1, S_2 and $\eta \in [0, A]$. Similarly,

$$\mathrm{Var}_{\beta,\mathbf{Q},\boldsymbol{\mu}_{S_1}(\eta)}\Big(\sum_{i\in S_1} X_i|\Omega_n\Big) = s\,\mathrm{sech}^2(\beta t+\eta) + O(s\sqrt{\log n/d_n})$$

and the same conclusion holds with S_1 replaced by S_2 above, with all error terms being uniform in $S_1, S_2, \eta \in [0, A]$. Consequently,

$$\begin{aligned}\sup_{\eta\in[0,A]}\sup_{S_1\neq S_2\in\mathcal{C}'_n}\Big|&\mathrm{Var}_{\beta,\mathbf{Q},\boldsymbol{\mu}_{S_1\cup S_2}(\eta)}\Big(\sum_{i\in S_1\cup S_2} X_i|\Omega_n\Big) - \mathrm{Var}_{\beta,\mathbf{Q},\boldsymbol{\mu}_{S_1}(\eta)}\Big(\sum_{i\in S_1} X_i|\Omega_n\Big)\\ &- \mathrm{Var}_{\beta,\mathbf{Q},\boldsymbol{\mu}_{S_2}(\eta)}\Big(\sum_{i\in S_2} X_i|\Omega_n\Big)\Big| \lesssim O(s\sqrt{\log n/d_n} + s^2\log n/n) = o(s/\log n).\end{aligned} \tag{30}$$

Then, by setting $r'_n = \frac{s}{\log n}$ and applying Theorem 3, part (II) completes the proof.

5.3. *Proof of Theorem 7*. Observe that $\|\mathbf{Q}\|_{\infty\to\infty} = 1$ and $\max\{|\lambda_2(\mathbf{Q})|, |\lambda_n(\mathbf{Q})|\} = O_p(d_n^{-1/2})$ (see Friedman, Kahn and Szemerédi (1989), Theorem A). Further if $d_n = \Theta(n)$, the result follows from Theorem 6. Therefore we will assume $d_n = o(n)$ in the rest of the proof. The proof of the whole of part (a), and the critical case $\beta = 1$ of part (b), then follows similar to the proof of Theorem 6. In fact, the proofs of the upper bounds for the high and low temperature regimes are also the same as in Theorem 6. We prove the lower bounds for the high and low temperature regimes below.

- *High temperature* ($0 \le \beta < 1$)

  To begin note that the conditions for Lemma 10, part (a) hold if $\theta > 1/2$. Using (27), we therefore have the bound

$$\sup_{i,j\in S}\mathrm{Cov}_{\beta,\mathbf{Q},\mathbf{0}}(X_i, X_j) \lesssim \frac{1}{d_n},$$

and so, if $s \ll \frac{d_n}{\log n}$,

$$\begin{aligned}&\sup_{S\in\mathcal{C}_n}\mathrm{Var}_{\beta,\mathbf{Q},\mathbf{0}}\Big(\sum_{i\in S} X_i\Big) \lesssim s + s^2\frac{\log n}{d_n} \lesssim s,\\ &\frac{\log n}{s}\sup_{S_1\neq S_2}\sum_{i\in S_1, j\in S_2}\mathrm{Cov}_{\beta,\mathbf{Q},\mathbf{0}}(X_i, X_j) \lesssim \frac{s\log n}{d_n} = o(1).\end{aligned} \tag{31}$$

Thus invoking Theorem 3 with $r_n = s$, $r'_n = \frac{s}{\log n}$ gives the desired conclusion. Thus without loss of generality we will assume $s \gtrsim \frac{d_n}{\log n}$ throughout the remainder of the proof.

In the remaining part of the proof we will denote the adjacency matrix any graph $\mathbb{G}_n = (\mathcal{V}_n, \mathcal{E}_n)$ on n-vertices as $\mathbf{G}_n$ and its i, jth element as $\mathbf{G}_n(i, j)$. Also conditioning on a random graph $\mathbb{G}_n$ will imply conditioning w.r.t to the sigma field generated by the random variables involved in $\mathbb{G}_n$ (i.e., the random edges in case of a simple random graph with fixed vertex set).

Now, if $\mathbb{G}_n$ is a random d_n-regular graph on vertices $\{1, \dots, n\}$, then using Gao, Isaev and McKay ((2020), Theorem 1.5 (b)), it follows that $\mathbb{G}_n$ is stochastically dominated by an Erdős–Renyi graph $\widetilde{\mathbb{G}}_n$ (whose corresponding adjacency matrix will be denoted by $\widetilde{\mathbf{G}}_n(i, j)$ for its i, jth element) with parameter $p_n := \kappa\frac{d_n\log(n/d_n)}{n}$ for some fixed $\kappa > 0$, and so we have

$$\mathbb{P}\left((\mathbf{Q}^3)_{ij} \geq 8\kappa^3 \frac{(\log(n/d_n))^3}{n}\right) = \mathbb{P}\left(\mathbf{G}_n(i,j)^3 \geq 8\kappa^3 \frac{d_n^3(\log(n/d_n))^3}{n}\right)$$
$$\leq \mathbb{P}(\widetilde{\mathbf{G}}_n^3(i,j) \geq 8n^2 p_n^3).$$

Let F_i denote the neighbors of i and note that $\widetilde{\mathbf{G}}_n^3(i,j) = \sum_{k\in F_i,\ell\in F_j} \widetilde{\mathbf{G}}_n(k,\ell)$. Also we have $|F_i| \sim \text{Bin}(n-1, p_n)$, $|F_j| \sim \text{Bin}(n-1,p_n)$, and given the sets F_i, F_j we further have $\sum_{k\in F_i,\ell\in F_j} \widetilde{\mathbf{G}}_n(k,\ell)$ is stochastically dominated by the $\text{Bin}(|F_i||F_j|, p_n)$ distribution. Using this we have

$$\begin{aligned}
&\mathbb{P}(\widetilde{\mathbf{G}}_n^3(i,j) \geq 8n^2 p_n^3)\\
&\quad\leq 2\mathbb{P}(|F_i| > 2np_n)\\
&\qquad + \mathbb{E}\left[\mathbb{P}\left(\sum_{k\in F_i,\ell\in F_j} \widetilde{\mathbf{G}}_n(k,\ell) > 8n^2p_n^3 | F_i, F_j\right) 1\{\max\{|F_i|, |F_j| \leq 2np_n\}\right]\\
&\quad\leq 2\mathbb{P}(\text{Bin}(n,p_n) > 2np_n) + \mathbb{P}(\text{Bin}(4n^2p_n^2, p_n) > 8n^2p_n^3) \leq e^{-\delta np_n} + e^{-\delta n^2 p_n^3}
\end{aligned}$$

for some $\delta > 0$, where the last step uses standard Chernoff bounds for a binomial distribution. A union bound along with the assumption $\theta > 1/2$, which translates to $p_n \geq n^{-\alpha}$ for some $\alpha < 1/2$ shows that, on setting $D_n := \{\max_{i,j\in[n]} (\mathbf{Q}^3)_{ij} \geq 8\kappa^3 \frac{(\log(n/d_n))^3}{n}\}$ we have

$$\mathbb{P}(D_n) \leq n^2(e^{-\delta np_n} + e^{-\delta n^2 p_n^3}) = o(1).$$

Also, if $\mathbb{G}_n \in D_n^c$, then for any $i,j\in[n]$ we have the following bound from Lemma 10, part (a):

$$\sup_{(i,j)\in\mathcal{E}_n} \text{Cov}_{\beta,\mathbf{Q},\mathbf{0}}(X_i,X_j|\mathbb{G}_n) \lesssim \frac{1}{d_n}, \qquad \sup_{(i,j)\notin\mathcal{E}_n} \text{Cov}_{\beta,\mathbf{Q},\mathbf{0}}(X_i,X_j|\mathbb{G}_n) \lesssim \frac{(\log n)^3}{n}. \tag{32}$$

Let $\mathcal{L}_{n,2s}$ denote the collection of all subsets of $[n]$ of size $2s$. Then for any sets $S \in \mathcal{L}_{n,2s}$ let $E(S,\mathbb{G}_n)$ denote the number of edges in $\mathbb{G}_n$ within the vertices in S. Then we have

$$\mathbb{P}(E(S,\mathbb{G}_n) > 4s^2p_n) \leq \mathbb{P}(E(S,\widetilde{\mathbb{G}}_n) > 4s^2p_n) = \mathbb{P}\left(\text{Bin}\left(\binom{2s}{2}, p_n\right) > 4s^2 p_n\right) \leq e^{-\delta s^2 p_n}.$$

Setting $E_n := \{\sup_{S\in\mathcal{L}_{n,2s}} E(S,\mathbb{G}_n) > 4s^2 p_n\}$, a union bound then gives

$$\mathbb{P}(E_n) \leq \binom{n}{2s} e^{-\delta s^2 p_n} \leq n^{2s} e^{-\delta s^2 p_n} = o(1),$$

where we use the bound

$$s^2 p_n \geq s \cdot \frac{d_n}{\log n} \cdot \frac{\kappa d_n \log(n/d_n)}{n} = s \cdot \frac{d_n^2}{n\log n} \gg 2s\log n.$$

Combining we have $\mathbb{P}(D_n \cup E_n) = o(1)$. For $\mathbb{G}_n \in D_n^c \cap E_n^c$, using the bound (32) gives

$$\sup_{S\in\mathcal{C}_n} \text{Var}_{\beta,\mathbf{Q},\mathbf{0}}\left(\sum_{i\in S} X_i | \mathbb{G}_n\right) \lesssim s + \frac{s^2 p_n}{d_n} + s^2 \frac{(\log n)^3}{n} \lesssim s,$$

$$\frac{\log n}{s} \sup_{S_1\neq S_2} \sum_{i\in S_1, j\in S_2} \text{Cov}_{\beta,\mathbf{Q},\mathbf{0}}(X_i,X_j|\mathbb{G}_n) \lesssim \frac{\log n}{s}\left(\frac{s^2p_n}{d_n} + s^2\frac{(\log n)^3}{n}\right) = o(1).$$

Thus again we have verified (31), and so invoking Theorem 3 with $r_n = s$, $r_n' = \frac{s}{\log n}$ gives the desired conclusion as before using Theorem 3, part (I).

- *Low temperature*

Set $\Omega_n := \{\bar{\mathbf{X}} \geq 0\}$. The proof is similar to Theorem 6, part (b) for the low temperature regime. Without loss of generality, we assume $d_n = o(n)$ as before. Note that, on the set D_n^c defined above we have by the same calculation as in the high temperature regime:

$$\sup_{(i,j)\in\mathcal{E}_n} \mathrm{Cov}_{\beta,\mathbf{Q},\boldsymbol{\mu}}(X_i, X_j|\bar{\mathbf{X}} \geq 0, \mathbb{G}_n) \lesssim \frac{1}{d_n},$$
$$\sup_{(i,j)\notin\mathcal{E}_n} \mathrm{Cov}_{\beta,\mathbf{Q},\boldsymbol{\mu}}(X_i, X_j|\bar{\mathbf{X}} \geq 0, \mathbb{G}_n) \lesssim \frac{(\log n)^3}{n}, \tag{33}$$

by Lemma 10, part (b) where we have used the fact that $\theta > 2/3$. Using (33) and $s \ll \frac{d_n}{\log n}$, we get

$$\sup_{\eta\in[0,2A]} \sup_{S\in\mathcal{C}_n} \mathrm{Var}_{\beta,\mathbf{Q},\boldsymbol{\mu}_S(\eta)}\Big(\sum_{i\in S} X_i|\bar{\mathbf{X}} \geq 0, \mathbb{G}_n\Big) \lesssim s + \frac{s^2 \log n}{d_n} \lesssim s. \tag{34}$$

Also by the same calculation as in (30), we see that

$$\begin{aligned}
&\sup_{\eta\in[0,A]} \sup_{S_1\neq S_2\in\mathcal{C}_n'} \Big|\mathrm{Var}_{\beta,\mathbf{Q},\boldsymbol{\mu}_{S_1\cup S_2}(\eta)}\Big(\sum_{i\in S_1\cup S_2} X_i|\Omega_n, \mathbb{G}_n\Big) \\
&\quad - \mathrm{Var}_{\beta,\mathbf{Q},\boldsymbol{\mu}_{S_1}(\eta)}\Big(\sum_{i\in S_1} X_i|\Omega_n, \mathbb{G}_n\Big) - \mathrm{Var}_{\beta,\mathbf{Q},\boldsymbol{\mu}_{S_2}(\eta)}\Big(\sum_{i\in S_2} X_i|\Omega_n, \mathbb{G}_n\Big)\Big| \\
&\lesssim s\sqrt{\frac{\log n}{d_n}} + \frac{s^2}{d_n} = o_p\Big(\frac{s}{\log n}\Big).
\end{aligned} \tag{35}$$

Therefore by choosing $r_n = s$, $r_n' = s/\log n$ and using Theorem 3, part (II) then completes the proof. We can therefore assume $s \gtrsim \frac{\log n}{d_n}$. In this case, the same conclusions as in (34) and (35) follow as in the proof of the high temperature regime if we further restrict to $G_n \in D_n^c \cap E_n^c$. We omit the details for brevity.

5.4. *Proof of Theorem* 8. The proof goes through exactly as the proof of Theorem 7, with the only change being that now $\mathbb{G}_n$ is not regular but "approximately regular." In order to carry out the same proof, we need to show that $\mathbb{G}_n$ (having adjacency matrix $\mathbf{G}_n$ and $\mathbf{Q} = \mathbf{G}_n/\overline{d}$) with and $\theta > 0$, satisfies:

(1) $\max_{i\in[n]} |\frac{d_i}{\overline{d}} - 1| = O_p(\sqrt{\frac{\log n}{\overline{d}}})$.
(2) $\|\mathbf{Q}\|_{\infty\to\infty} = O_p(1)$.
(3) $\max\{|\lambda_2(\mathbf{Q})|, |\lambda_n(\mathbf{Q})|\} = o_p(1)$.

Here we as usual have defined $d_i = \sum_{j=1}^n \mathbf{G}_n(i,j)$ to be the degree of the ith vertex and $\overline{d}$ the average degree of the graph. As $d_1 \sim \mathrm{Bin}(n-1, p_n)$, a standard Chernoff's inequality yields that $\sqrt{np_n}|\frac{d_1}{np_n} - 1| = O_p(1)$ and a union bound then yields $\sqrt{\frac{np_n}{\log n}} \max_{i\in[n]} |\frac{d_i}{np_n} - 1| = O_p(1)$. A similar argument also shows that $\sqrt{np_n}|\frac{\overline{d}}{np_n} - 1| = O_p(1)$. Combining these observations yields

$$\sqrt{\frac{\overline{d}}{\log n}} \max_{i\in[n]}\Big|\frac{d_i}{\overline{d}} - 1\Big| \lesssim \frac{np_n}{\sqrt{\overline{d}\log n}}\Big(\max_{i\in[n]}\Big|\frac{d_i}{np_n} - 1\Big| + \Big|\frac{\overline{d}}{np_n} - 1\Big|\Big) = O_p(1),$$

which establishes (1). Note that (2) follows from (1), and (3) follows from Feige and Ofek ((2005), Theorem 1.1).

5.5. *Proof of Theorem* 9. In this proof we follow the notation introduced in Section 3.2.

(a) $s \geq C \log n$

For this part, it suffices to verify the conditions of Theorem 3 for some large constant $C > 0$. To this end, we first define a subcollection $\mathcal{C}_n'$ of $\mathcal{C}_n$ as follows. Throughout we assume that $s^{1/d}$ and $n^{1/d}$ are integers for the sake of notational convenience. The analyses works verbatim otherwise by working with the corresponding ceiling functions. Also assume without loss of generality that $3s^{1/d}$ divides $n^{1/d}$. First, let $\mathcal{C}_n''$ be the class of disjoint subcubes of $\Lambda_n(d)$ obtained by translating along each axis (by $3s^{1/d}$ in each direction each time) the cube of side lengths $3s^{1/d}$ from the bottom left corner of $\Lambda_n(d) = [-n^{1/d}, n^{1/d}]^d \cap \mathbb{Z}^d$. Consequently, subdivide each cube in $\mathcal{C}_n'$ into 3^d cubes of side length $s^{1/d}$ each and take the center subcube of each cube in $\mathcal{C}_n''$ to be elements of our class $\mathcal{C}_n'$. It is easy to see that $|\mathcal{C}_n'| = |\mathcal{C}_n''| = (2/3)^d \frac{n}{s}$ and also that $\min_{\substack{i \in S_1, j \in S_2: \\ S_1 \neq S_2 \in \mathcal{C}_n}} \|i - j\|_1 \geq 4s^{1/d}$.

First note that $\|\mathbf{Q}\|_{\infty \to \infty} = 2dL$. Also note that by Lemma 13, for all $0 \leq \beta < \beta_c(d, L)$ one has

$$\operatorname{Cov}_{\beta,\mathbf{Q},\mathbf{0}}(X_i, X_j) \leq \exp\bigl(-c(\beta, d, L)\|i - j\|_1\bigr), \tag{36}$$

for some $c(\beta, d, L) > 0$ depending on β, d, L. Therefore for any $S \in \mathcal{C}_n'$ we have

$$\begin{aligned} \operatorname{Var}_{\beta,\mathbf{Q},\mathbf{0}}\biggl(\sum_{i \in S} X_i\biggr) &\leq \sum_{i,j \in S} \exp\bigl(-c(\beta, d, L)\|i - j\|_1\bigr) \\ &\leq \sum_{l=0}^{ds^{1/d}} \sum_{i,j \in S : \|i-j\|_1 = l} \exp\bigl(-c(\beta, d, L)\|i - j\|_1\bigr) \leq C's, \end{aligned}$$

for some constant $C' > 0$. Also for any $S_1 \neq S_2 \in \mathcal{C}'$

$$\begin{aligned} \sum_{i \in S_1, j \in S_2} \operatorname{Cov}_{\beta,\mathbf{Q},\mathbf{0}}(X_i, X_j) &\leq \sum_{i \in S_1, j \in S_2} \exp\bigl(-c(\beta, d, L)\|i - j\|_1\bigr) \\ &\leq s^2 \exp\bigl(-4c(\beta, d, L)s^{1/d}\bigr). \end{aligned}$$

This completes the verification of the conditions of Theorem 3, thus verifying part (a).

(b) $s \leq c \log n$

It suffices to verify the conditions of Theorem 4 for some small constant $c > 0$. To this end, we again define a subcollection of $\mathcal{C}_n$ as follows. First, let $\mathcal{C}_n'$ be the class of disjoint subcubes of $\Lambda_n(d)$ obtained by translating along each axis (by $\log n$ in each direction each time) the cube of side lengths $s^{1/d}$ from the bottom left corner of $\Lambda_n(d) = [-n^{1/d}, n^{1/d}]^d \cap \mathbb{Z}^d$. It is easy to see that $|\mathcal{C}_n'| \gtrsim (\frac{n}{s + \log^d n}) \gtrsim \frac{n}{\log^d n}$, and also $\min_{\substack{i \in S_1, j \in S_2: \\ S_1 \neq S_2 \in \mathcal{C}_n'}} \|i - j\|_1 \geq \log n$. Consider now any two $S_1 \neq S_2 \in \mathcal{C}_n'$ and consider the ratio

$$\frac{\mathbb{P}_{\beta,\mathbf{Q},\mathbf{0}}(\mathbf{X}_{S_1} = 1, \mathbf{X}_{S_2} = 1)}{\mathbb{P}_{\beta,\mathbf{Q},\mathbf{0}}(\mathbf{X}_{S_1} = 1)\mathbb{P}_{\beta,\mathbf{Q},\mathbf{0}}(\mathbf{X}_{S_2} = 1)}.$$

To analyze this ratio, let $\tilde{\mathbb{P}}$ denote the Edward–Sokal coupling measure between the Ising model $\mathbb{P}_{\beta,\mathbf{Q},\mathbf{0}}$ and the corresponding random cluster model (see, e.g., Duminil-Copin (2020), Grimmett (2006)). In the following argument, for any two sets $A, B \in \Lambda_n(d)$ we denote $A \leftrightarrow B$ (respectively $A \not\leftrightarrow B$) to denote the event that there is an open path between the sets A and B (respectively there is not open path between A and B). Then

$$\begin{aligned} &\mathbb{P}_{\beta,\mathbf{Q},\mathbf{0}}(\mathbf{X}_{S_1} = 1, \mathbf{X}_{S_2} = 1) \\ &\quad = \tilde{\mathbb{P}}(\mathbf{X}_{S_1} = 1, \mathbf{X}_{S_2} = 1) \\ &\quad = \tilde{\mathbb{P}}(\mathbf{X}_{S_1} = 1, \mathbf{X}_{S_2} = 1, S_1 \leftrightarrow S_2) + \tilde{\mathbb{P}}(\mathbf{X}_{S_1} = 1, \mathbf{X}_{S_2} = 1, S_1 \not\leftrightarrow S_2). \end{aligned}$$

Now, since $0 \leq \beta < \beta_c(d)$ and $\min_{\substack{i \in S_1, j \in S_2: \\ S_1 \neq S_2 \in \mathcal{C}'_n}} \|i - j\|_1 \geq \log n$, there exists a constant $\rho > 0$ such that

$$\tilde{\mathbb{P}}(\mathbf{X}_{S_1} = 1, \mathbf{X}_{S_2} = 1, S_1 \leftrightarrow S_2) \leq \tilde{\mathbb{P}}(S_1 \leftrightarrow S_2) \leq \exp(-\rho \log n). \tag{37}$$

Similarly, since under the Edward–Sokal coupling disjoint clusters are assigned spins independent of one another, we have

$$\begin{aligned}\tilde{\mathbb{P}}(\mathbf{X}_{S_1} = 1, \mathbf{X}_{S_2} = 1, S_1 \not\leftrightarrow S_2) &= \tilde{\mathbb{P}}(\mathbf{X}_{S_1} = 1, \mathbf{X}_{S_2} = 1 | S_1 \not\leftrightarrow S_2) \tilde{P}(S_1 \not\leftrightarrow S_2) \\ &= \tilde{\mathbb{P}}(\mathbf{X}_{S_1} = 1 | S_1 \not\leftrightarrow S_2) \tilde{\mathbb{P}}(\mathbf{X}_{S_2} = 1 | S_1 \not\leftrightarrow S_2) \tilde{\mathbb{P}}(S_1 \not\leftrightarrow S_2).\end{aligned}$$

Moreover, by FKG inequality we have

$$\mathbb{P}_{\beta,\mathbf{Q},\mathbf{0}}(\mathbf{X}_{S_1} = 1) \mathbb{P}_{\beta,\mathbf{Q},\mathbf{0}}(\mathbf{X}_{S_2} = 1) \geq 2^{-2s}. \tag{38}$$

Therefore,

$$\begin{aligned}\frac{\mathbb{P}_{\beta,\mathbf{Q},\mathbf{0}}(\mathbf{X}_{S_1} = 1, \mathbf{X}_{S_2} = 1)}{\mathbb{P}_{\beta,\mathbf{Q},\mathbf{0}}(\mathbf{X}_{S_1} = 1) \mathbb{P}_{\beta,\mathbf{Q},\mathbf{0}}(\mathbf{X}_{S_2} = 1)} &= \frac{\tilde{\mathbb{P}}(\mathbf{X}_{S_1} = 1 | S_1 \not\leftrightarrow S_2) \tilde{\mathbb{P}}(\mathbf{X}_{S_2} = 1 | S_1 \not\leftrightarrow S_2) \tilde{\mathbb{P}}(S_1 \not\leftrightarrow S_2)}{\mathbb{P}_{\beta,\mathbf{Q},\mathbf{0}}(\mathbf{X}_{S_1} = 1) \mathbb{P}_{\beta,\mathbf{Q},\mathbf{0}}(\mathbf{X}_{S_2} = 1)} \\ &\quad + \frac{\tilde{\mathbb{P}}(\mathbf{X}_{S_1} = 1, \mathbf{X}_{S_2} = 1, S_1 \leftrightarrow S_2)}{\mathbb{P}_{\beta,\mathbf{Q},\mathbf{0}}(\mathbf{X}_{S_1} = 1) \mathbb{P}_{\beta,\mathbf{Q},\mathbf{0}}(\mathbf{X}_{S_2} = 1)}.\end{aligned}$$

The second term in the display above is smaller than $\exp(-\rho \log n + s \log 4)$ by (37) and (38). Therefore, for $s \leq c \log n$ with c small enough, this converges to 0, and we only need to show that the first term above is $1 + o(1)$ uniformly in $S_1 \neq S_2$. To this end, note that

$$\frac{\tilde{\mathbb{P}}(\mathbf{X}_{S_1} = 1 | S_1 \not\leftrightarrow S_2) \tilde{\mathbb{P}}(\mathbf{X}_{S_2} = 1 | S_1 \not\leftrightarrow S_2) \tilde{\mathbb{P}}(S_1 \not\leftrightarrow S_2)}{\mathbb{P}_{\beta,\mathbf{Q},\mathbf{0}}(\mathbf{X}_{S_1} = 1) \mathbb{P}_{\beta,\mathbf{Q},\mathbf{0}}(\mathbf{X}_{S_2} = 1)} = T_1 + T_2 + T_3,$$

where

$$\begin{aligned}T_1 &= \frac{\tilde{\mathbb{P}}(\mathbf{X}_{S_1} = 1) \tilde{\mathbb{P}}(\mathbf{X}_{S_2} = 1)}{\tilde{\mathbb{P}}(S_1 \not\leftrightarrow S_2) \mathbb{P}_{\beta,\mathbf{Q},\mathbf{0}}(\mathbf{X}_{S_1} = 1) \mathbb{P}_{\beta,\mathbf{Q},\mathbf{0}}(\mathbf{X}_{S_2} = 1)}, \\ T_2 &= -\frac{\tilde{\mathbb{P}}(\mathbf{X}_{S_1} = 1, S_1 \leftrightarrow S_2) \tilde{\mathbb{P}}(X_{S_2} = 1) + \tilde{\mathbb{P}}(\mathbf{X}_{S_2} = 1, S_1 \leftrightarrow S_2) \tilde{\mathbb{P}}(X_{S_1} = 1)}{\tilde{\mathbb{P}}(S_1 \not\leftrightarrow S_2) \mathbb{P}_{\beta,\mathbf{Q},\mathbf{0}}(\mathbf{X}_{S_1} = 1) \mathbb{P}_{\beta,\mathbf{Q},\mathbf{0}}(\mathbf{X}_{S_2} = 1)}, \\ T_3 &= \frac{\tilde{\mathbb{P}}(\mathbf{X}_{S_1} = 1, S_1 \leftrightarrow S_2) \tilde{\mathbb{P}}(\mathbf{X}_{S_2} = 1, S_1 \leftrightarrow S_2)}{\tilde{\mathbb{P}}(S_1 \not\leftrightarrow S_2) \mathbb{P}_{\beta,\mathbf{Q},\mathbf{0}}(\mathbf{X}_{S_1} = 1) \mathbb{P}_{\beta,\mathbf{Q},\mathbf{0}}(\mathbf{X}_{S_2} = 1)}.\end{aligned}$$

Once again, it is easy to see from (37) and (38) that there exists a constant $\tilde{\rho}$ such that for large enough n one has $|T_2 + T_3| \lesssim C_1 \exp(-\tilde{\rho} \log n + s \log 4)$. For T_1, note that by definition of the coupling we have

$$\begin{aligned}T_1 &= \frac{\mathbb{P}_{\beta,\mathbf{Q},\mathbf{0}}(\mathbf{X}_{S_1} = 1) \mathbb{P}_{\beta,\mathbf{Q},\mathbf{0}}(\mathbf{X}_{S_2} = 1)}{\tilde{\mathbb{P}}(S_1 \not\leftrightarrow S_2) \mathbb{P}_{\beta,\mathbf{Q},\mathbf{0}}(\mathbf{X}_{S_1} = 1) \mathbb{P}_{\beta,\mathbf{Q},\mathbf{0}}(\mathbf{X}_{S_2} = 1)} \\ &= \frac{1}{\tilde{\mathbb{P}}(S_1 \not\leftrightarrow S_2)}.\end{aligned}$$

From (37) we immediately have $T_1 = 1 + o(1)$ uniformly in $S_1 \neq S_2 \in \mathcal{C}'_n$. This completes the verification of the condition of Theorem 4 for $0 \leq \beta < \beta_c(d)$.

6. Proofs of auxiliary lemmas. This section is devoted to proving Lemmas 10, 11, and 12.

In the sequel, we will use $d_{\max}$ to represent the maximum degree of the graph $\mathbb{G}_n$ and $\mathbf{I}_m$ for the $m \times m$ identity matrix. Finally, throughout we let $\alpha_n = \sqrt{\frac{\log n}{\overline{d}}}$, for $\mathbf{X} \sim \mathbb{P}_{\beta,\mathbf{Q},\boldsymbol{\mu}}$ let $m_i = m_i(\mathbf{X}) := \sum_{j=1}^n \mathbf{Q}_{ij} X_j$, $\bar{\mathbf{m}} = \sum_{i=1}^n m_i/n$, and use the letter t to denote the nonnegative root of $x = \tanh(\beta x)$ for $\beta > 1$.

Our first result yields a sharp control on the tail behavior of $m_i, i \geq 1$—which serves as a crucial building block for proving Lemmas 10, 11, and 12.

LEMMA 14. *Suppose that* $\mathbf{Q}$ *is the scaled adjacency matrix of a graph, such that* $\max_{1\leq i\leq n} |\frac{d_i}{\overline{d}} - 1| \to 0$, *and let* $\lambda \geq 1$. *Then we have the following conclusions*:

(a) *For* $0 \leq \beta < 1$ *we have*

$$\log \mathbb{P}_{\beta,\mathbf{Q},\mathbf{0}}\Big(\max_{i\in[n]} |m_i| > \lambda\alpha_n\Big) \lesssim -\lambda^2,$$

for all large enough n.

(b) *Let* $\beta > 1$, *and* $\max_{i\in[n]} |\frac{d_i}{\overline{d}} - 1| \lesssim \alpha_n$.

(i) *If* $\mathbb{G}_n$ *satisfies* $\overline{d} \gg \sqrt{n \log n}$ *and*

$$\limsup_{n\to\infty} \lambda_2(\mathbf{Q}) < 1, \tag{39}$$

and $\boldsymbol{\mu}$ *satisfies*

$$\sum_{i=1}^n \mu_i \lesssim \sqrt{n \log n}, \quad \|\mathbf{Q}\boldsymbol{\mu}\|_\infty \lesssim \alpha_n, \tag{40}$$

then we have

$$\log \mathbb{P}_{\beta,\mathbf{Q},\boldsymbol{\mu}}\Big(\max_{i\in[n]} |m_i - t| \geq \lambda\alpha_n, \bar{\mathbf{X}} \geq 0\Big) \lesssim -\lambda^2,$$

for all large enough n.

(ii) *If* $\mathbb{G}_n$ *satisfies* $\overline{d} \geq n^\gamma$ *for some* $\gamma > 0$, *and* $\max_{2\leq i\leq n} |\lambda_i(\mathbf{Q})| \to 0$, *and* $\boldsymbol{\mu}$ *satisfies* (40), *then we have*

$$\log \mathbb{P}_{\beta,\mathbf{Q},\mathbf{0}}\Big(\max_{i\in[n]} |m_i - t| \geq \lambda\alpha_n, \bar{\mathbf{X}} \geq 0\Big) \lesssim -\min\big(\lambda^2, \overline{d}^{1-\gamma}\big),$$

for all large enough n.

(c) *For* $\beta = 1$, *assume that* $\max_{i\in[n]} |\frac{d_i}{\overline{d}} - 1| \lesssim \alpha_n$.

(i) *Then we have*

$$\log \mathbb{P}_{\beta,\mathbf{Q},\mathbf{0}}\Big(\max_{i\in[n]} |m_i| > \lambda\alpha_n^{1/3}\Big) \lesssim -\lambda^2$$

for all large enough n.

(ii) *If* $\mathbb{G}_n$ *satisfies* (39), *and* $\boldsymbol{\mu} \in (\mathbb{R}^+)$ *satisfies* $\|\mathbf{Q}\boldsymbol{\mu}\|_\infty \lesssim \alpha_n$, *then we have*

$$\log \mathbb{P}_{\beta,\mathbf{Q},\boldsymbol{\mu}}\left(\max_{i\in[n]} |m_i - \bar{\mathbf{m}}| > \lambda\frac{(\log n)^{3/2}}{\overline{d}} + \lambda\sqrt{\bar{\boldsymbol{\mu}}}\right) \lesssim -\lambda^2,$$

for all large enough n.

PROOF. Part (a) follows directly from part (a) of Deb and Mukherjee ((2023), Lemma 2.3). Part (c)(i) follows by combining Deb and Mukherjee ((2023), equations (4.9), (4.10)). Here we prove the remaining parts (b)(i) (ii), and (c)(ii).

Part (b)(i). Without loss of generality we can assume $\lambda\alpha_n \le 1$, as otherwise the bound is trivial on noting that $\lambda\alpha_n > 1 \gtrsim \max_{i\in[n]} |m_i|$. We now claim that

$$\log \mathbb{P}_{\beta,\mathbf{Q},\boldsymbol{\mu}}(A_{n,\lambda}) \lesssim -\lambda^2, \quad A_{n,\lambda} := \left\{\max_{i\in[n]}\left|m_i - \sum_{j=1}^n \mathbf{Q}_{ij}\tanh(\beta m_j)\right| > \lambda\alpha_n\right\}. \tag{41}$$

Indeed, this follows on using Lemma 4 along with the bound $\|\mathbf{Q}\boldsymbol{\mu}\|_\infty \lesssim \alpha_n$.

A two term Taylor expansion of $\tanh(\beta m_j)$ gives

$$\left|\sum_{j=1}^n \mathbf{Q}_{ij}\tanh(\beta m_j) - \frac{d_i}{\overline{d}}t - \beta(1-t^2)\sum_{j=1}^n \mathbf{Q}_{ij}(m_j - t)\right| \lesssim \sum_{j=1}^n \mathbf{Q}_{ij}(m_j-t)^2,$$

and so on the set $A^c_{n,\lambda}$ we have

$$\begin{aligned} &\max_{i\in[n]}|m_i - t|\Big[1 - \beta(1-t^2)\max_{i\in[n]} d_i/\overline{d}\Big] \lesssim \lambda\alpha_n + \max_{i\in[n]}\left[\sum_{j=1}^n \mathbf{Q}_{ij}(m_j-t)^2 + \left|\frac{d_i}{\overline{d}} - 1\right|\right] \\ &\Rightarrow \quad \max_{i\in[n]}|m_i - t| \lesssim \max_{i\in[n]}\sum_{j=1}^n Q_{ij}(m_j - t)^2 + \lambda\alpha_n, \end{aligned} \tag{42}$$

where the last line uses the fact that $\max_{i\in[n]}\frac{d_i}{\overline{d}} \to 1$, and $\beta(1-t^2) < 1$. We now claim that for every $\varepsilon > 0$ we have

$$\log \mathbb{P}_{\beta,\mathbf{Q},\boldsymbol{\mu}}(B_{n,\varepsilon}) \lesssim -\frac{1}{\alpha_n^2}, \quad B_{n,\varepsilon} := \Big\{\max_{i\in[n]}|m_i - t| > \varepsilon\Big\}. \tag{43}$$

Given (43), choosing $\varepsilon > 0$ small enough, on the set $A^c_{n,\lambda} \cap B^c_{n,\varepsilon}$, using (42) we have

$$\max_{i\in[n]}|m_i - t| \lesssim \lambda\alpha_n + \max_{i\in[n]}|m_i - t|^2 \quad \Rightarrow \quad \max_{i\in[n]}|m_i - t| \lesssim \lambda\alpha_n.$$

Combining the above observations, we get

$$\begin{aligned} \log \mathbb{P}_{\beta,\mathbf{Q},\boldsymbol{\mu}}\Big(\max_{i\in[n]}|m_i - t| \gtrsim \lambda\alpha_n\Big) &\le \log\big[\mathbb{P}_{\beta,\mathbf{Q},\boldsymbol{\mu}}(A_{n,\lambda}) + \mathbb{P}_{\beta,\mathbf{Q},\boldsymbol{\mu}}(B_{n,\varepsilon})\big] \\ &\lesssim -\min\left[\lambda^2, \frac{1}{\alpha_n^2}\right] \end{aligned} \tag{44}$$

using (41) and (43). The desired conclusion follows from this on recalling that $\lambda\alpha_n \le 1$.

It thus suffices to verify (43). To this effect, use (41) with $\lambda\alpha_n = \varepsilon$ to note that

$$\log \mathbb{P}_{\beta,\mathbf{Q},\boldsymbol{\mu}}(A_{n,\varepsilon/\alpha_n}) = \log \mathbb{P}_{\beta,\mathbf{Q},\boldsymbol{\mu}}\left(\max_{i\in[n]}\left|m_i - \sum_{j=1}^n \mathbf{Q}_{ij}\tanh(\beta m_j)\right| > \varepsilon\right) \lesssim -\frac{1}{\alpha_n^2}. \tag{45}$$

We now claim that for any $0 < M < \infty$ and for all $L \ge M\alpha_n^{-2}\log n$ (i.e., $L \ge M\overline{d}$), we have

$$\log \mathbb{P}_{\beta,\mathbf{Q},\boldsymbol{\mu}}(C_{n,L}) \lesssim -L, \quad \text{where } C_{n,L} := \left\{\sum_{i=1}^n (m_i - t)^2 > L, \bar{\mathbf{X}} \ge 0\right\}. \tag{46}$$

On the set $C^c_{n,\varepsilon\overline{d}}$ we have

$$\sum_{j=1}^n \mathbf{Q}_{ij}(m_j - t)^2 \le \frac{1}{\overline{d}}\sum_{j=1}^n (m_j - t)^2 \le \varepsilon,$$

which along with (42) implies that there exists constants $M_1 > 0$ such that

$$\log \mathbb{P}_{\beta,\mathbf{Q},\boldsymbol{\mu}}\Big(\max_{i\in[n]} |m_i - t| > \varepsilon, \bar{\mathbf{X}} \geq 0\Big)$$
$$\leq \log \mathbb{P}_{\beta,\mathbf{Q},\boldsymbol{\mu}}\left(\max_{i\in[n]} \sum_{j=1}^n \mathbf{Q}_{ij}(m_j - t)^2 \geq M_1\varepsilon, \bar{\mathbf{X}} \geq 0\right)$$
$$\leq \log \max\big[\mathbb{P}_{\beta,\mathbf{Q},\boldsymbol{\mu}}(A_{n,M_1\varepsilon/\alpha_n}), \mathbb{P}_{\beta,\mathbf{Q},\boldsymbol{\mu}}(C_{n,\varepsilon\overline{d}})\big] \lesssim -\min\left[\frac{1}{\alpha_n^2}, \overline{d}\right] = -\alpha_n^{-2},$$

where the last inequality uses (45) and (46). This verifies (43). Finally, (46) follows using (40) to note that $\sup_{\mathbf{x}\in[-1,1]^n} |\sum_{i=1}^n \mu_i x_i| \leq \sum_{i=1}^n \mu_i \lesssim \sqrt{n \log n}$, and consequently for δ small enough we have

$$\log \mathbb{E}_{\beta,\mathbf{Q},\boldsymbol{\mu}}\left[\exp\left(\delta \sum_{i=1}^n (m_i - t)^2\right) \middle| \bar{\mathbf{X}} \geq 0\right]$$
$$\leq \sqrt{n \log n} + \log \mathbb{E}_{\beta,\mathbf{Q},\mathbf{0}}\left[\exp\left(\delta \sum_{i=1}^n (m_i - t)^2\right) \middle| \bar{\mathbf{X}} \geq 0\right] \tag{47}$$
$$\lesssim \sqrt{n \log n} + \sum_{i=1}^n \left(\frac{d_i}{\overline{d}} - 1\right)^2 + \frac{n}{\overline{d}} = o(\overline{d}),$$

where the last line uses Deb and Mukherjee ((2023), Lemma 2.2, part (b)).

Part (b)(ii). We begin by claiming that for every $\varepsilon > 0$ we have

$$\log \mathbb{P}_{\beta,\mathbf{Q},\boldsymbol{\mu}}\Big(\max_{i\in[n]} |m_i - t| > \varepsilon\Big) \lesssim -\overline{d}^{1-\gamma}. \tag{48}$$

Given this claim, note that (48) is the analogue to (43) above, which is the only place where we use the fact that $\overline{d} \gg \sqrt{n \log n}$ in the proof of part (c)(i). Thus, following the above proof for the derivation of (44) gives

$$\log \mathbb{P}_{\beta,\mathbf{Q},\boldsymbol{\mu}}\Big(\max_{i\in[n]} |m_i - t| > \lambda\alpha_n\Big) \lesssim \max\big[\log \mathbb{P}_{\beta,\mathbf{Q},\boldsymbol{\mu}}(A_{n,\lambda}), \log \mathbb{P}_{\beta,\mathbf{Q},\boldsymbol{\mu}}(B_{n,\varepsilon})\big]$$
$$= -\min\big[\lambda^2, \overline{d}^{1-\gamma}\big],$$

as desired. It thus remains to verify (48).

To this effect, a one term Taylor's series expansion of $\tanh(\beta m_j)$ gives

$$\sum_{j=1}^n \mathbf{Q}_{ij} \tanh(\beta m_j) = t\frac{d_i}{\overline{d}} + \beta \sum_{j=1}^n \mathbf{Q}_{ij}(m_j - t)\operatorname{sech}^2(\beta\xi_j)$$
$$= t\frac{d_i}{\overline{d}} + \beta(1 - t^2)\frac{1}{n}\sum_{j=1}^n (m_j - t) + \sum_{j=1}^n \mathbf{D}_{ij}^{(n)}(m_j - t),$$

where $\mathbf{D}_{ij}^{(n)} := \beta[\mathbf{Q}_{ij}\operatorname{sech}^2(\beta\xi_j) - \frac{1-t^2}{n}]$, for some ξ_j lying between m_j and t. On the set $A^c_{n,\sqrt{\overline{d}^{1-\gamma}}} \cap C^c_{n,n\log n/\overline{d}^\gamma}$ this gives

$$\left|(m_i - t) - \sum_{j=1}^n \mathbf{D}_{ij}^{(n)}(m_j - t)\right| \lesssim \left|\frac{d_i}{\overline{d}} - 1\right| + \left|\frac{1}{n}\sum_{i=1}^n (m_i - t)\right| + \sqrt{\frac{\log n}{\overline{d}^\gamma}}$$
$$\leq \left|\frac{d_i}{\overline{d}} - 1\right| + \sqrt{\frac{1}{n}\sum_{i=1}^n (m_i - t)^2} + \sqrt{\frac{\log n}{\overline{d}^\gamma}} \leq 3\sqrt{\frac{\log n}{\overline{d}^\gamma}}$$

for all n large enough. With K denoting the implied constant in the display above we have

$$\max_{i\in[n]}\left|(m_i - t) - \sum_{j=1}^{n} \mathbf{D}_{ij}^{(n)}(m_j - t)\right| \le K\sqrt{\frac{\log n}{\overline{d}^{\gamma}}}. \tag{49}$$

Consequently, for every $\ell \ge 1$ we have

$$\begin{aligned}&\left|\sum_{j=1}^{n}(\mathbf{D}_{ij}^{(n)})^{\ell}(m_j - t) - \sum_{j=1}^{n}\mathbf{D}_{ij}^{(n)}\sum_{k=1}^{n}(\mathbf{D}_{jk}^{(n)})^{\ell}(m_k - t)\right| \\ &\quad\le \max_{j\in[n]}\left|(m_j - t) - \sum_{k=1}^{n}\mathbf{D}_{jk}^{(n)}(m_j - t)\right| \max_{i\in[n]}\left|\sum_{j=1}^{n}(\mathbf{D}_{ij}^{(n)})^{\ell}\right| \le (2\beta)^{\ell} K\sqrt{\frac{\log n}{\overline{d}^{\gamma}}},\end{aligned}$$

where the last inequality uses the bound $\max_{i\in[n]}\sum_{j=1}^{n}|\mathbf{D}_{ij}^{(n)}| \le 2\beta$ for all n large enough. Combining the last two displays show that for any $\ell \ge 1$ we have

$$\max_{i\in[n]}\left|(m_i - t) - \sum_{j=1}^{n}(\mathbf{D}_{ij}^{(n)})^{\ell}(m_j - t)\right| \le K\overline{d}^{-\gamma}\sum_{r=0}^{\ell-1}(2\beta)^{r} \le (2\beta)^{\ell}K\sqrt{\frac{\log n}{\overline{d}^{\gamma}}},$$

and so

$$\max_{i\in[n]}|m_i - t| \le \|(\mathbf{D}^{(n)})^{\ell}\|_{\infty}\max_{i\in[n]}|m_i - t| + (2\beta)^{\ell}K\overline{d}^{-\gamma} \lesssim \|(\mathbf{D}^{(n)})^{\ell}\|_{\infty} + (2\beta)^{\ell}\sqrt{\frac{\log n}{\overline{d}^{\gamma}}}. \tag{50}$$

We now claim that on the set $A^{c}_{n,\sqrt{\overline{d}^{1-\gamma}}} \cap C^{c}_{n,n\log n/\overline{d}^{\gamma}}$ we have

$$\lim_{n\to\infty}\|\mathbf{D}^{(n)}\|_2 = 0. \tag{51}$$

Given (51), using spectral theorem write $(\mathbf{D}^{(n)})^{\top}\mathbf{D}^{(n)} = \sum_{i=1}^{n}\lambda_i \mathbf{p}_i\mathbf{p}_i$, where $\max_{i\in[n]}|\mu_i| = o(1)$, and so

$$|\{((\mathbf{D}^{(n)})^{\top}\mathbf{D}^{(n)})^{\ell}\}_{ij}| = \left|\left(\sum_{k=1}^{n}\lambda_i^{\ell}\mathbf{p}_k\mathbf{p}_k'\right)_{ij}\right| \le \max_{k\in[n]}|\lambda_k|^{\ell}.$$

This immediately shows that setting $\ell = \delta\log n$ with $\delta = \frac{\gamma}{2\log(2\beta)}$, using (50) we have

$$\max_{i\in[n]}|m_i - t| \lesssim n\max_{i\in[n]}|\lambda_i|^{\ell} + (2\beta)^{\delta\log n}\sqrt{\frac{\log n}{\overline{d}^{\gamma}}} \le \sqrt{\frac{\log n}{\overline{d}^{\gamma/2}}} \to 0$$

for all n large enough. Thus for any $\varepsilon > 0$, for all n large we have

$$\mathbb{P}_{\beta,\mathbf{Q},\boldsymbol{\mu}}\Big(\max_{i\in[n]}|m_i - t| > \varepsilon\Big) \le \mathbb{P}_{\beta,\mathbf{Q},\boldsymbol{\mu}}(A_{n,\sqrt{\overline{d}^{1-\gamma}}}) + \mathbb{P}_{\beta,\mathbf{Q},\boldsymbol{\mu}}(C_{n,n\log n/\overline{d}^{\gamma}}),$$

which along with (41) and (46) gives

$$\log\mathbb{P}_{\beta,\mathbf{Q},\boldsymbol{\mu}}\Big(\max_{i\in[n]}|m_i - t| \ge \varepsilon\Big) \lesssim -\min\left[\overline{d}^{1-\gamma}, \frac{n\log n}{\overline{d}^{\gamma}}\right] = -\overline{d}^{1-\gamma},$$

which verifies (48), and hence completes the proof of part (b)(ii).

It thus suffices to verify (51). To this effect, setting $\mathbf{J} := \frac{1}{n}\mathbf{1}\mathbf{1}'$ and $\boldsymbol{\Delta}^{(n)}$ denote a diagonal matrix with entries $\boldsymbol{\Delta}_{ii}^{(n)} := \operatorname{sech}^2(\beta\xi_i)$ we have

$$\|\mathbf{D}^{(n)}\|_2 = \|\beta\mathbf{Q}\boldsymbol{\Delta}^{(n)} - \beta(1-t^2)\mathbf{J}\|_2 \le \beta\|\mathbf{Q}\boldsymbol{\Delta}^{(n)} - \mathbf{J}\boldsymbol{\Delta}^{(n)}\|_2 + \beta\|\mathbf{J}[\boldsymbol{\Delta}^{(n)} - (1-t^2)\mathbf{I}_n]\|_2,$$

from which (51) follows on noting that $\|\mathbf{Q}-\mathbf{J}\|_2 = o(1)$ by assumption, and

$$\|\mathbf{J}[\boldsymbol{\Delta}^{(n)}-(1-t^2)I]\|_2^2 = \|\mathbf{J}[\boldsymbol{\Delta}^{(n)}-(1-t^2)I]^2\mathbf{J}\|_2 = \frac{1}{n}\sum_{i=1}^n [\boldsymbol{\Delta}_{ii}^{(n)}-(1-t^2)]^2 \to 0,$$

where the last limit uses the fact that we are working in the set $C_{n,n\log n/\overline{d}^\gamma}$. This verifies (51), and hence completes the proof of part (c).

Part (c)(ii). By Deb and Mukherjee ((2023), equation (4.12)), on the set $A_{n,\lambda}$ for any $\ell \ge 1$ we have

$$\max_{i\in[n]}\left|m_i - \bar{\mathbf{m}} - \sum_{j=1}^n \tilde{\mathbf{Q}}_{ij}^\ell(m_j-\bar{\mathbf{m}})\right| \lesssim \lambda\ell\sqrt{\log n}/\sqrt{\overline{d}},$$

where $\tilde{\mathbf{Q}}_{ij} := \mathbf{Q}_{ij}$ for $i \neq j$, and $\tilde{\mathbf{Q}}_{ii} := \frac{d_{\max}}{\overline{d}} - 1$ satisfies $\tilde{\mathbf{Q}}\mathbf{1} = \mathbf{1}$. Set $\ell = D\log n$ for D fixed but large enough so that $\max_{i\in[n]} \tilde{\mathbf{Q}}_{ii}^\ell \le \frac{3}{n}$ (such a D exists by Deb and Mukherjee ((2023), Lemma 5.2(a))). Then we have

$$\left|\max_{i\in[n]}\sum_{j=1}^n \tilde{\mathbf{Q}}_{ij}^\ell(m_j-\bar{\mathbf{m}})\right| \le \sqrt{\max_{i\in[n]}\tilde{\mathbf{Q}}_{ii}}\sqrt{\sum_{j=1}^n(m_j-\bar{\mathbf{m}})^2} \le \sqrt{3}\sqrt{\frac{1}{n}\sum_{i=1}^n(m_i-\bar{\mathbf{m}})^2},$$

and so

$$\begin{aligned}&\mathbb{P}_{\beta,\mathbf{Q},\boldsymbol{\mu}}\left(\max_{i\in[n]}|m_i-\bar{\mathbf{m}}| \ge \lambda(\log n)^{3/2}/\sqrt{\overline{d}} + \lambda\sqrt{n^{-1}\sum_i \mu_i}\right)\\ &\quad\le \mathbb{P}_{\beta,\mathbf{Q},\boldsymbol{\mu}}(A_{n,\lambda}^c) + \mathbb{P}_{\beta,\mathbf{Q},\boldsymbol{\mu}}\left(\sum_{i=1}^n(m_i-\bar{\mathbf{m}})^2 \gtrsim \lambda^2 n(\log n)^3/\overline{d} + \lambda^2\sum_{i=1}^n\mu_i\right).\end{aligned} \tag{52}$$

The desired conclusion follows from this on noting that for δ small enough, using a similar argument as (47), we have

$$\log\mathbb{E}_{\beta,\mathbf{Q},\boldsymbol{\mu}}\left[\exp\left(\delta\sum_{i=1}^n(m_i-\bar{\mathbf{m}})^2\right)\right] \lesssim \frac{n}{\overline{d}}\log n + \sum_{i=1}^n\mu_i,$$

where we have used Deb and Mukherjee ((2023), Lemma 2.2). □

6.1. *Proof of Lemma 10*. *Part (a)*. For $i \neq j$ setting $m_i^{(j)} := \sum_{k\neq j}\mathbf{Q}_{ik}X_k$ we have $\mathbb{E}_{\beta,\mathbf{Q},\mathbf{0}}[(X_j - \tanh(\beta m_j))\tanh(\beta m_i^{(j)})] = 0$, and so

$$\begin{aligned}&\left|\mathbb{E}_{\beta,\mathbf{Q},\mathbf{0}}[X_iX_j] - \mathbb{E}_{\beta,\mathbf{Q},\mathbf{0}}[\tanh(\beta m_i)\tanh(\beta m_j)]\right|\\ &\quad= \left|\mathbb{E}_{\beta,\mathbf{Q},\mathbf{0}}[(X_j-\tanh(\beta m_j))\tanh(\beta m_i)]\right| \le 2\beta\mathbf{Q}_{ij}.\end{aligned} \tag{53}$$

Now, using part (a) of Lemma 14, for any positive integer k we have

$$\mathbb{E}_{\beta,\mathbf{Q},\mathbf{0}}\left[\max_{i\in[n]}|m_i|^k\right] \lesssim \alpha_n^k. \tag{54}$$

Using this, a Taylor's series expansion gives $|\tanh(\beta m_i) - \beta m_i| \lesssim |m_i|^3$, which gives

$$\left|\mathbb{E}_{\beta,\mathbf{Q},\mathbf{0}}[\tanh(\beta m_i)\tanh(\beta m_j)] - \beta^2\mathbb{E}_{\beta,\mathbf{Q},\mathbf{0}}[m_im_j]\right| \lesssim \alpha_n^4. \tag{55}$$

Equipped with (53), (54) and (55), we now complete the proof of part (a). To verify the first estimate, setting $\rho_n^{(1)} := \max_{k\neq\ell}\mathbb{E}_{\beta,\mathbf{Q},\mathbf{0}}[X_kX_\ell]$ we have

$$\mathbb{E}_{\beta,\mathbf{Q},\mathbf{0}}[m_im_j] = \sum_{k,\ell=1}^n \mathbf{Q}_{ik}\mathbf{Q}_{j\ell}\mathbb{E}_{\beta,\mathbf{Q},\mathbf{0}}[X_kX_\ell] \le \rho_n^{(1)}\max_{i\in[n]}\left(\frac{d_i}{\overline{d}}\right)^2$$

which along with (53), (54) and (55) gives the existence of a constant M such that

$$\rho_n^{(1)} \le \beta^2 \rho_n^{(1)} \max_{i\in[n]} \left(\frac{d_i}{\overline{d}}\right)^2 + M\Big[\max_{i\neq j} \mathbf{Q}_{ij} + \alpha_n^4\Big] \quad \Rightarrow \quad \rho_n^{(1)} \lesssim \frac{1}{\overline{d}} + \alpha_n^4 \lesssim \frac{1}{\overline{d}}, \tag{56}$$

where the last bound uses $\overline{d} \gtrsim (\log n)^2$. Proceeding to show the second bound, setting $\rho_n^{(2)} := \max_{(k,\ell)\notin\mathcal{E}_n} \mathbb{E}_{\beta,\mathbf{Q},\mathbf{0}}[X_k X_\ell]$ and using (56) gives the existence of a finite constant $\tilde{M}$ free of n such that for all $(i,j) \notin \mathcal{E}_n^c$ we have

$$\begin{aligned}
\mathbb{E}_{\beta,\mathbf{Q},\mathbf{0}}[m_i m_j] &= \sum_{k,\ell=1}^n \mathbf{Q}_{ik}\mathbf{Q}_{j\ell}\mathbb{E}_{\beta,\mathbf{Q},\mathbf{0}}[X_k X_\ell] \\
&\le \frac{\tilde{M}}{\overline{d}} \sum_{(k,\ell)\in\mathcal{E}_n} \mathbf{Q}_{ik}\mathbf{Q}_{j\ell}\mathbf{G}_n(k,l) + \beta^2 \max_{(k,\ell)\notin\mathcal{E}_n} \mathbb{E}X_k X_\ell \sum_{(k,\ell)\notin\mathcal{E}_n} \mathbf{Q}_{ik}\mathbf{Q}_{j\ell} \\
&\le \tilde{M}(\mathbf{Q}^3)_{ij} + \beta^2 \rho_n^{(2)} \frac{d_i d_j}{\overline{d}^2} = \tilde{M}(\mathbf{Q}^3)_{ij} + \beta^2 \rho_n^{(3)} \frac{d_i d_j}{\overline{d}^2}.
\end{aligned}$$

Since $\max_{i\in[n]} \frac{d_i}{\overline{d}} \to 1$ and $\beta < 1$, using (53) and (55) along with the above display gives $\rho_n^{(2)} \lesssim \max_{i,j} \mathbf{Q}_{ij}^3 + \alpha_n^4$, which is the second conclusion of part (a).

Part (b)(i). A Taylor's series expansion of $g_i(x) := \tanh(\beta x + \mu_i)$ gives

$$g_i(m_i) = g_i(t) + (m_i - t)g_i'(t) + \frac{g_i''(t)}{2}(m_i - t)^2 + \frac{g_i'''(\xi)}{3!}(m_i - t)^3, \tag{57}$$

where ξ_i lies between m_i and t. Also, using part (b) of Lemma 14, for any positive integer k we have

$$\mathbb{E}_{\beta,\mathbf{Q},\boldsymbol{\mu}} \max_{i\in[n]} |m_i - t|^k \lesssim \alpha_n^k, \tag{58}$$

and so

$$\begin{aligned}
\mathbb{E}_{\beta,\mathbf{Q},\boldsymbol{\mu}}\big[g_i(m_i)g_j(m_j)|\bar{\mathbf{X}} \ge 0\big] &= g_i(t)g_j(t) + g_i'(t)g_j(t)\mathbb{E}_{\beta,\mathbf{Q},\boldsymbol{\mu}}[m_i - t|\bar{\mathbf{X}} \ge 0] \\
&\quad + g_i(t)g_j'(t)\mathbb{E}_{\beta,\mathbf{Q},\boldsymbol{\mu}}[m_j - t|\bar{\mathbf{X}} \ge 0] \\
&\quad + \frac{g_i''(t)g_j(t)}{2}\mathbb{E}_{\beta,\mathbf{Q},\boldsymbol{\mu}}\big[(m_i - t)^2|\bar{\mathbf{X}} \ge 0\big] \\
&\quad + \frac{g_i(t)g_j''(t)}{2}\mathbb{E}_{\beta,\mathbf{Q},\boldsymbol{\mu}}\big[(m_j - t)^2|\bar{\mathbf{X}} \ge 0\big] \\
&\quad + g_i(t)g_j(t)\mathbb{E}_{\beta,\mathbf{Q},\boldsymbol{\mu}}\big[(m_i - t)(m_j - t)|\bar{\mathbf{X}} \ge 0\big] + O(\alpha_n^3), \\
\mathbb{E}_{\beta,\mathbf{Q},\boldsymbol{\mu}}\big[g_i(m_i)|\bar{\mathbf{X}} \ge 0\big] &= g_i(t) + g_i'(t)\mathbb{E}_{\beta,\mathbf{Q},\boldsymbol{\mu}}[m_i - t|\bar{\mathbf{X}} \ge 0] \\
&\quad + \frac{g_i''(t)}{2}\mathbb{E}_{\beta,\mathbf{Q},\boldsymbol{\mu}}\big[(m_i - t)^2|\bar{\mathbf{X}} \ge 0\big] + O(\alpha_n^3).
\end{aligned}$$

A direct multiplication using the last display gives

$$\big|\mathrm{Cov}_{\beta,\mathbf{Q},\boldsymbol{\mu}}\big(g_i(m_i), g_j(m_j)|\bar{\mathbf{X}} \ge 0\big) - g_i'(t)g_j'(t)\mathrm{Cov}_{\beta,\mathbf{Q},\boldsymbol{\mu}}(m_i - t, m_j - t|\bar{\mathbf{X}} \ge 0)\big| \lesssim \alpha_n^3. \tag{59}$$

We now claim that there exists $\rho > 0$ such that for all n large enough we have

$$\begin{aligned}
&\max_{i\neq j}\big|\mathrm{Cov}_{\beta,\mathbf{Q},\boldsymbol{\mu}}(X_i, X_j|\bar{\mathbf{X}} \ge 0) - \mathrm{Cov}_{\beta,\mathbf{Q},\boldsymbol{\mu}}\big(\tanh(\beta m_i + \mu_i), \tanh(\beta m_j + \mu_j)|\bar{\mathbf{X}} \ge 0\big)\big| \\
&\quad \lesssim \mathbf{Q}_{ij} + e^{-\rho n}.
\end{aligned} \tag{60}$$

Note that (60), (58) and (59) are the analogues of (53), (54) and (55). Given these estimates, the rest of the proof follows along similar lines as in part (a), on noting that

$$\max_{1\le i,j\le n} g_i'(t)g_j'(t) = \beta^2 \max_{1\le i,j\le n} \operatorname{sech}^2(\beta t+\mu_i)\operatorname{sech}^2(\beta t+\mu_j)$$
$$= \beta^2\operatorname{sech}^4\Big(\beta t+\max_{i\in[n]}\mu_i\Big) < 1-\epsilon$$

for some fixed $\epsilon > 0$ and all large enough n by assumption. It only remains to verify (60). To this effect, we first claim that there exists a (different) constant $\rho > 0$ free of n, such that for any function $f : \{-1,1\}^n \mapsto [-1,1]$ we have

$$\max_{i\in[n]}\big|\mathbb{E}_{\beta,\mathbf{Q},\boldsymbol{\mu}}[f(\mathbf{X})|\bar{\mathbf{X}}\ge 0] - \mathbb{E}_{\beta,\mathbf{Q},\boldsymbol{\mu}}[f(\mathbf{X})|\bar{\mathbf{X}}_i > 0]\big| \le 5e^{-\rho n}, \tag{61}$$

where $\bar{\mathbf{X}}_i := \frac{1}{n}\sum_{j\ne i} X_j$. Given (61), noting that $\mathbb{P}_{\beta,\mathbf{Q},\boldsymbol{\mu}}(\bar{\mathbf{X}}\ge 0)\ge 1/3$ for all large enough n, we have

$$\begin{aligned}
&\mathbb{E}_{\beta,\mathbf{Q},\boldsymbol{\mu}}(X_iX_j|\bar{\mathbf{X}}\ge 0)\\
&\quad= \frac{\mathbb{E}_{\beta,\mathbf{Q},\boldsymbol{\mu}}[X_iX_j\mathbf{1}\{\bar{\mathbf{X}}\ge 0\}]}{\mathbb{P}_{\beta,\mathbf{Q},\boldsymbol{\mu}}(\bar{\mathbf{X}}\ge 0)}\\
&\quad= \frac{\mathbb{E}_{\beta,\mathbf{Q},\boldsymbol{\mu}}[X_iX_j\mathbf{1}\{\bar{\mathbf{X}}_j > 0\}]}{\mathbb{P}_{\beta,\mathbf{Q},\boldsymbol{\mu}}(\bar{\mathbf{X}}\ge 0)} + O(e^{-\rho n}) \text{ [By (61)]}\\
&\quad= \frac{\mathbb{E}_{\beta,\mathbf{Q},\boldsymbol{\mu}}[\tanh(\beta m_i+\mu_i)X_j\mathbf{1}\{\bar{\mathbf{X}}_j > 0\}]}{\mathbb{P}_{\beta,\mathbf{Q},\boldsymbol{\mu}}(\bar{\mathbf{X}}\ge 0)} + O(e^{-\rho n})\\
&\quad= \frac{\mathbb{E}_{\beta,\mathbf{Q},\boldsymbol{\mu}}[\tanh(\beta m_i^{(j)}+\mu_i)X_j\mathbf{1}\{\bar{\mathbf{X}}_i > 0\}]}{\mathbb{P}_{\beta,\mathbf{Q},\boldsymbol{\mu}}(\bar{\mathbf{X}}\ge 0)} + O(e^{-\rho n}) + O(\mathbf{Q}_{ij}) \text{ [By (61)]}\\
&\quad= \frac{\mathbb{E}_{\beta,\mathbf{Q},\boldsymbol{\mu}}[\tanh(\beta m_i^{(j)}+\mu_i)\tanh(\beta m_j+\mu_j)\mathbf{1}\{\bar{\mathbf{X}}_i > 0]}{\mathbb{P}_{\beta,\mathbf{Q},\boldsymbol{\mu}}(\bar{\mathbf{X}}\ge 0)} + O(e^{-\rho n}) + O(\mathbf{Q}_{ij})\\
&\quad= \frac{\mathbb{E}_{\beta,\mathbf{Q},\boldsymbol{\mu}}[\tanh(\beta m_i+\mu_i)\tanh(\beta m_j+\mu_j)\mathbf{1}\{\bar{\mathbf{X}}\ge 0]}{\mathbb{P}_{\beta,\mathbf{Q},\boldsymbol{\mu}}(\bar{\mathbf{X}}\ge 0)} + O(e^{\rho n}) + O(\mathbf{Q}_{ij}) \quad \text{[By (61)]}\\
&\quad= \mathbb{E}_{\beta,\mathbf{Q},\boldsymbol{\mu}}[\tanh(\beta m_i+\mu_i)\tanh(\beta m_j+\mu_j)|\bar{\mathbf{X}}\ge 0] + O(e^{-\rho n}) + O(\mathbf{Q}_{ij}).
\end{aligned}$$

A similar calculation gives

$$\mathbb{E}_{\beta,\mathbf{Q},\boldsymbol{\mu}}[X_i|\bar{\mathbf{X}}\ge 0] = \mathbb{E}_{\beta,\mathbf{Q},\boldsymbol{\mu}}[\tanh(\beta m_i+\mu_i)|\bar{\mathbf{X}}\ge 0] + O(e^{-\rho n}) + O(\mathbf{Q}_{ij}), \tag{62}$$

which along with the above display gives (60), as desired. To complete the proof, we need to verify (61). To this effect, use Deb and Mukherjee ((2023), equation (2.8)), to note that

$$\mathbb{P}_{\beta,\mathbf{Q},\boldsymbol{\mu}}(\bar{\mathbf{X}}\ge 0, \bar{\mathbf{X}}_i\le 0) + \mathbb{P}_{\beta,\mathbf{Q},\boldsymbol{\mu}}(\bar{\mathbf{X}}\le 0, \bar{\mathbf{X}}_i > 0) \le e^{-\rho n}$$
$$\Rightarrow \quad \big|\mathbb{P}_{\beta,\mathbf{Q},\boldsymbol{\mu}}(\bar{\mathbf{X}}\ge 0) - \mathbb{P}_{\beta,\mathbf{Q},\boldsymbol{\mu}}(\bar{\mathbf{X}}_i > 0)\big| \le e^{-\rho n},$$

and so

$$\begin{aligned}
&\big|\mathbb{E}_{\beta,\mathbf{Q},\boldsymbol{\mu}}[f(\mathbf{X})|\bar{\mathbf{X}}\ge 0] - \mathbb{E}_{\beta,\mathbf{Q},\boldsymbol{\mu}}[f(\mathbf{X})|\bar{\mathbf{X}}_i > 0]\big|\\
&\quad\le \left|\frac{1}{\mathbb{P}_{\beta,\mathbf{Q},\boldsymbol{\mu}}(\bar{\mathbf{X}}\ge 0)} - \frac{1}{\mathbb{P}_{\beta,\mathbf{Q},\boldsymbol{\mu}}(\bar{\mathbf{X}}_i > 0)}\right| + \mathbb{P}_{\beta,\mathbf{Q},\boldsymbol{\mu}}(\bar{\mathbf{X}}\ge 0, \bar{\mathbf{X}}_i < 0)\\
&\qquad+ \mathbb{P}_{\beta,\mathbf{Q},\boldsymbol{\mu}}(\bar{\mathbf{X}} < 0, \bar{\mathbf{X}}_i > 0) \le 5e^{-\rho n},
\end{aligned}$$

which verifies (61).

Part (b)(ii). It suffices to show that $\max_{i\in[n]} |\mathbb{E}_{\beta,\mathbf{Q},\boldsymbol{\mu}}(X_i|\bar{\mathbf{X}} \geq 0) - \tanh(\beta t + \mu_i)| \lesssim \alpha_n$. But this follows on using (57) and (58) to note that

$$\big|\mathbb{E}_{\beta,\mathbf{Q},\boldsymbol{\mu}}\big(\tanh(\beta m_i + \mu_i)|\bar{\mathbf{X}} \geq 0\big) - \tanh(\beta t + \mu_i)\big| \lesssim \mathbb{E}_{\beta,\mathbf{Q},\boldsymbol{\mu}}|m_i - t| \lesssim \alpha_n.$$

Part (c). To begin, use Deb and Mukherjee ((2023), equation (4.25)) and Deb and Mukherjee ((2023), Lemma 2.4(c)), coupled with the assumption $\overline{d} \gg \sqrt{n}(\log n)^5$ to note that

$$\mathbb{E}_{\beta,\mathbf{Q},\mathbf{0}}\bar{\mathbf{m}}^2 \lesssim \frac{1}{n} + \mathbb{E}\bar{\mathbf{X}}^2 \lesssim \frac{1}{n} + \frac{1}{\sqrt{n}} \lesssim \frac{1}{\sqrt{n}}.$$

Also, using part (c)(ii) of Lemma 14 with $\boldsymbol{\mu} = \mathbf{0}$ gives

$$\begin{aligned}
&\big|\mathbb{E}_{\beta,\mathbf{Q},\mathbf{0}} \tanh(m_i)\tanh(m_j)\big| \\
&\quad = \big|\mathbb{E}_{\beta,\mathbf{Q},\mathbf{0}}\big[\big(\tanh(m_i) - \tanh(\bar{\mathbf{m}}) + \tanh(\beta\bar{\mathbf{m}})\big)\big(\tanh(m_j) - \tanh(\bar{\mathbf{m}}) + \tanh(\beta\bar{\mathbf{m}})\big)\big]\big| \\
&\quad \leq \mathbb{E}_{\beta,\mathbf{Q},\mathbf{0}}\big|(m_i - \bar{\mathbf{m}})(m_j - \bar{\mathbf{m}})\big| + \mathbb{E}_{\beta,\mathbf{Q},\mathbf{0}}\big|(m_i - \bar{\mathbf{m}})\bar{\mathbf{m}}\big| \\
&\qquad + \mathbb{E}_{\beta,\mathbf{Q},\mathbf{0}}\big|(m_j - \bar{\mathbf{m}})\bar{\mathbf{m}}\big| + \mathbb{E}_{\beta,\mathbf{Q},\mathbf{0}}\bar{\mathbf{m}}^2 \\
&\quad \leq \mathbb{E}_{\beta,\mathbf{Q},\mathbf{0}} \max_{i\in[n]} |m_i - \bar{\mathbf{m}}|^2 + 2\sqrt{\mathbb{E}_{\beta,\mathbf{Q},\mathbf{0}} \max_{i\in[n]} |m_i - \bar{\mathbf{m}}|^2}\sqrt{\mathbb{E}_{\beta,\mathbf{Q},\mathbf{0}}\bar{\mathbf{m}}^2} + \mathbb{E}_{\beta,\mathbf{Q},\mathbf{0}}\bar{\mathbf{m}}^2 \\
&\quad \lesssim \alpha_n^2 + 2\frac{\alpha_n}{n^{1/4}} + \frac{1}{\sqrt{n}} \lesssim \frac{1}{\sqrt{n}}.
\end{aligned}$$

6.2. *Proof of Lemma* 11. *Part (a).* To begin, note that for any $i \in [n]$ we have

$$\mathbb{P}_{\beta,\mathbf{Q},\mathbf{0}}(X_i = x_i | X_j = x_j, j \neq i) = \frac{e^{\beta m_i}}{e^{\beta m_i} + e^{-\beta m_i}} \geq \frac{1}{1 + e^{-2C_u'\beta}} =: p,$$

where $\max_{i\in[n]} |m_i(\mathbf{X})| \leq \max_{1\leq i\leq n} \sum_{j=1}^n \mathbf{Q}_{ij} \leq C_u'$. The above display on taking expectation gives $\mathbb{P}_{\beta,\mathbf{Q},\mathbf{0}}(X_i = x_i | X_j = x_j, j \in A) \geq p$ for any $A \subset [n]$, and so $\mathbb{P}_{\beta,\mathbf{Q},\mathbf{0}}(\mathbf{X}_S = \mathbf{a}) \geq p^s$. On the other hand, setting

$$m_i(S) := \sum_{j\in S^c} Q_{ij}x_j, \qquad \Omega(S) := \left\{\mathbf{x} \in \{-1, 1\}^{n-s} : \max_{i\in[n]}\big|m_i(S)\big| \leq \lambda\frac{\log n}{\sqrt{\overline{d}}}\right\},$$

we have

$$\big|m_i - m_i(S)\big| \leq \frac{s}{\overline{d}} \lesssim \frac{\log n}{\overline{d}} \quad \Rightarrow \quad \mathbb{P}_{\beta,\mathbf{Q},\mathbf{0}}\big(\mathbf{X}_{S^c} \notin \Omega(S)\big) \leq n^{-\rho\lambda}$$

for some $\rho > 0$, where we use part (a) of Lemma 14. Consequently, for λ sufficiently large, for any $\mathbf{a} \in \{-1, 1\}^s$ we have

$$\mathbb{P}_{\beta,\mathbf{Q},\mathbf{0}}(\mathbf{X}_S = \mathbf{a}) \geq p^s \geq p^{c\log n} \gg n^{-\rho\lambda}, \quad \text{and so}$$
$$\lim_{n\to\infty} \sup_{S:|S|=s,\ \mathbf{a}\in\{-1,1\}^s} \left|\frac{\mathbb{P}_{\beta,\mathbf{Q},\mathbf{0}}(\mathbf{X}_S = \mathbf{a})}{\mathbb{P}_{\beta,\mathbf{Q},\mathbf{0}}(\mathbf{X}_S = \mathbf{a}, \mathbf{X}_{S^c} \in \Omega(S))} - 1\right| = 0.$$

It thus suffices to estimate $\mathbb{P}_{\beta,\mathbf{Q},\mathbf{0}}(\mathbf{X}_S=\mathbf{a},\mathbf{X}_{S^c}\in\Omega(S))$. To this effect, we have

$$\begin{aligned}
&\mathbb{P}_{\beta,\mathbf{Q},\mathbf{0}}\big(\mathbf{X}_S=\mathbf{a},\mathbf{X}_{S^c}\in\Omega(S)\big)\frac{1}{Z_n(\beta,\mathbf{Q},\mathbf{0})}\\
&\quad\times\sum_{\mathbf{x}\in\Omega(S)}\exp\bigg(\frac{\beta}{2}\sum_{i,j\in S^c}\mathbf{Q}_{ij}x_ix_j+\beta\sum_{i\in S,j\in S^c}\mathbf{Q}_{ij}a_ix_j+\frac{\beta}{2}\sum_{i,j\in S}\mathbf{Q}_{ij}a_ia_j\bigg)\\
&=\frac{1}{Z_n(\beta,\mathbf{Q},\mathbf{0})}\sum_{\mathbf{x}\in\Omega(S)}\exp\bigg(\frac{\beta}{2}\sum_{i,j\in S^c}\mathbf{Q}_{ij}x_ix_j+\beta\sum_{i\in S}a_im_i(S)+\frac{\beta}{2}\sum_{i,j\in S}\mathbf{Q}_{ij}a_ia_j\bigg)\\
&\le\frac{\exp(\lambda\beta s\frac{\log n}{\sqrt{\overline{d}}}+\frac{\beta s^2}{2\overline{d}})}{Z_n(\beta,\mathbf{Q},\mathbf{0})}\sum_{\mathbf{x}\in\Omega(S)}\exp\bigg(\frac{\beta}{2}\sum_{i,j\in S^c}\mathbf{Q}_{ij}x_ix_j\bigg).
\end{aligned}\tag{63}$$

A similar calculation gives

$$\mathbb{P}_{\beta,\mathbf{Q},\mathbf{0}}\big(\mathbf{X}_S=\mathbf{a},\mathbf{X}_{S^c}\in\Omega(S)\big)\ge\frac{\exp(-\lambda\beta s\frac{\log n}{\sqrt{\overline{d}}}-\frac{\beta s^2}{2\overline{d}})}{Z_n(\beta,\mathbf{Q},\mathbf{0})}\sum_{\mathbf{x}\in\Omega(S)}\exp\bigg(\frac{\beta}{2}\sum_{i,j\in S^c}\mathbf{Q}_{ij}x_ix_j\bigg).\tag{64}$$

Note that both the bounds in (63) and (64) are free of $\mathbf{a}$ and depend on S only through its cardinality, that is, s. The ratio of these two bounds converge to 1 using the fact that $\overline{d}\gg(\log n)^4$. The conclusion in (24) then follows.

Part (b). The proof of part (a) goes through verbatim after replacing the term $\frac{\log n}{\sqrt{\overline{d}}}$ in the definition of $\Omega(S)$ by $\frac{(\log n)^{2/3}}{\overline{d}^{1/6}}$.

Part (c). Again the proof is similar, except that we now use the bound $\sum_{i\in S}|a_im_i(S)-a_it|\le\lambda s\frac{\log n}{\sqrt{\overline{d}}}$ for $\mathbf{x}\in\Omega(S)$, where the revised $\Omega(S)$ is defined as

$$\Omega(S):=\bigg\{\mathbf{x}\in\{-1,1\}^{n-s}:\max_{i\in[n]}\big|m_i(S)-t\big|\le\lambda\frac{\log n}{\sqrt{\overline{d}}},\ \frac{1}{n}\sum_{i\in S^c}X_i>0\bigg\}.$$

6.3. *Proof of Lemma* 12. Since the probability distribution $\mathbb{P}_{\beta,\mathbf{Q},\boldsymbol{\mu}}$ is monotonic in $\boldsymbol{\mu}$ (coordinatewise), without loss of generality we can replace μ by $\tilde{\mu}$, where $\tilde{\mu}_i=\min(A,\frac{\sqrt{\overline{d}}}{s\sqrt{\log n}})$ for $i\in S$ and $\tilde{\mu}_i=0$ otherwise. Then we have

$$\sum_{i=1}^n\tilde{\mu}_i=\min\bigg(sA,\sqrt{\frac{\overline{d}}{\log n}}\bigg)\quad\Rightarrow\quad n^{1/4}\ll\sum_{i=1}^n\tilde{\mu}_i\le\sqrt{\frac{\overline{d}}{\log n}}.$$

Therefore, more generally, we will show the existence of $\eta>0$ such that

$$\mathbb{P}_{\beta,\mathbf{Q},\boldsymbol{\mu}}\bigg(n\overline{\mathbf{X}}^3>\eta\sum_{i=1}^n\mu_i\bigg)\to1\quad\text{whenever } n^{1/4}\ll\sum_{i=1}^n\mu_i\lesssim\sqrt{\frac{\overline{d}}{\log n}},\quad\max_{i\in[n]}\mu_i\to0.\tag{65}$$

This choice gives

$$\sum_{j=1}^nQ_{ij}\mu_j\le\frac{1}{\overline{d}}\sum_{j=1}^n\mu_j\le\frac{\sqrt{\overline{d}}}{\overline{d}\sqrt{\log n}}=\frac{1}{\sqrt{\overline{d}\log n}}\ll\sqrt{\frac{\log n}{\overline{d}}},$$

and so Lemma 14 part (c)(ii) applies.

We begin by claiming the following, whose proofs we defer:

$$\mathbb{E}_{\beta,\mathbf{Q},\boldsymbol{\mu}}(n^{1/4}\overline{\mathbf{X}})^6 \lesssim n^{-1/2}\left(\sum_{i=1}^n \mu_i\right)^2, \tag{66}$$

$$\mathbb{E}_{\beta,\mathbf{Q},\boldsymbol{\mu}}\left[\sum_{i=1}^n (d_i/\overline{d}-1)X_i\right]^2 \lesssim n^{1/3}\left(\sum_{i=1}^n \mu_i\right)^{2/3}, \tag{67}$$

$$\mathbb{E}_{\beta,\mathbf{Q},\boldsymbol{\mu}}\left[\sum_{i=1}^n (X_i-\tanh(m_i+\mu_i))\right]^2 \lesssim n^{1/3}\left(\sum_{i=1}^n \mu_i\right)^{2/3}. \tag{68}$$

An application of the triangle inequality gives:

$$\begin{aligned}\sum_{i=1}^n (X_i-\tanh(m_i)) &\geq \sum_{i=1}^n \big(\tanh(m_i+\mu_i)-\tanh(m_i)\big) - \left|\sum_{i=1}^n (X_i-\tanh(m_i+\mu_i))\right| \\ &\geq \rho\sum_{i=1}^n \mu_i - \left|\sum_{i=1}^n (X_i-\tanh(m_i+\mu_i))\right|\end{aligned} \tag{69}$$

for some positive constant ρ free of n. Also, since $\sum_{i=1}^n \mu_i \gg n^{1/4}$, using (68) gives

$$\left|\sum_{i=1}^n (X_i-\tanh(m_i+\mu_i))\right| = O_p\left(n^{1/3}\left(\sum_{i=1}^n \mu_i\right)^{2/3}\right) = o_p\left(\sum_{i=1}^n \mu_i\right). \tag{70}$$

Finally, a Taylor's series expansion of $\tanh(m_i)$ at $\bar{\mathbf{m}}$ gives

$$\left|\sum_{i=1}^n \tanh(m_i) - n\tanh(\bar{\mathbf{m}})\right| \lesssim |\bar{\mathbf{m}}|\sum_{i=1}^n (m_i-\bar{\mathbf{m}})^2 + \sum_{i=1}^n |m_i-\bar{\mathbf{m}}|^3, \tag{71}$$

and so

$$\begin{aligned}&\left|\sum_{i=1}^n (X_i-\tanh(m_i))\right| \\ &\lesssim n|\bar{\mathbf{X}}|^3 + n|\bar{\mathbf{m}}-\bar{\mathbf{X}}| + |\bar{\mathbf{X}}|\sum_{i=1}^n (m_i-\bar{\mathbf{m}})^2 + \sum_{i=1}^n |m_i-\bar{\mathbf{m}}|^3 \\ &= n\overline{\mathbf{X}}^3 + O_p\left(n^{1/6}\left(\sum_{i=1}^n \mu_i\right)^{1/3}\right) + O_p\left(n^{-1/3}\left(\sum_{i=1}^n \mu_i\right)^{1/3}\left(\frac{n(\log n)^3}{\overline{d}^2} + n\overline{\boldsymbol{\mu}}\right)\right) \\ &\quad + O_p\left(\frac{n(\log n)^{9/2}}{\overline{d}^{3/2}} + n(\overline{\boldsymbol{\mu}})^{3/2}\right) \\ &= n\overline{\mathbf{X}}^3 + o_p\left(\sum_{i=1}^n \mu_i\right),\end{aligned} \tag{72}$$

where the last line uses (66), (67) and Lemma 14 part (c)(ii). Combining (72) along with (69) and (70) completes the proof of (65).

We now verify the three claims (66), (67), (68). To this effect, note that

$$\mathbb{E}_{\beta,\mathbf{Q},\boldsymbol{\mu}}(n^{1/4}\bar{\mathbf{X}})^6 \lesssim \frac{1}{\sqrt{n}}\left\{\left(\sum_{i=1}^n \mu_i\right)^2 + n^{-1/2}\mathbb{E}_{\beta,\mathbf{Q},\boldsymbol{\mu}}\left[\sum_{i=1}^n (d_i/\overline{d}-1)X_i\right]^2\right\}, \tag{73}$$

$$\mathbb{E}_{\beta,\mathbf{Q},\boldsymbol{\mu}}\left[\sum_{i=1}^n (d_i/\overline{d}-1)X_i\right]^2 \lesssim \sqrt{n}\left[1+\mathbb{E}_{\beta,\mathbf{Q},\boldsymbol{\mu}}(n^{1/4}\bar{\mathbf{X}})^2\right] \tag{74}$$

together imply (66) and (67). It thus suffices to verify (73), (74), and (68).

- Proof of (73)

The proof follows closely the proof of Deb and Mukherjee ((2023), equation 4.13).

Set $T_n = n^{-3/4}\sum_{i=1}^n X_i$, and form an exchangeable pair $(\mathbf{X}, \mathbf{X}')$ as follows: Let I denote a randomly sampled index from $\{1, 2, \dots, n\}$. Given $I = i$, replace X_i with an independent ± 1 valued random variable X_i' with mean $\tanh(\beta m_i + \mu_i) = \mathbb{E}_{\beta,\mathbf{Q},\boldsymbol{\mu}}[X_i|(X_j, j \neq i)]$, and let $\mathbf{X}' := (X_1, \dots, X_{i-1}, X_i', X_{i+1}, \dots, X_n)$. Then we have

$$\begin{aligned}\mathbb{E}_{\beta,\mathbf{Q},\boldsymbol{\mu}}[T_n - T_n'|\mathbf{X}] &= \frac{1}{n^{7/4}}\sum_{i=1}^n \big(X_i - \tanh(m_i + \mu_i)\big) \\ &= \frac{1}{n^{7/4}}\sum_{i=1}^n \big(X_i - \tanh(m_i)\big) + \frac{1}{n^{7/4}}\sum_{i=1}^n \xi_i \mu_i,\end{aligned}$$

where $\{\xi_i\}_{1\le i\le n}$ are bounded random variables. This, along with the last display gives

$$\begin{aligned}&\big|\mathbb{E}_{\beta,\mathbf{Q},\boldsymbol{\mu}}[T_n - T_n'|\mathbf{X}] - n^{-3/2}T_n^3/3\big| \\ &\quad\le \frac{2}{15}n^{-2}|T_n|^5 + M\left\{n^{-3/4}|\bar{\mathbf{X}} - \bar{\mathbf{m}}| + n^{-2}|T_n|\sum_{i=1}^n (m_i - \bar{\mathbf{m}})^2 + n^{-7/4}\left|\sum_{i=1}^n (m_i - \bar{\mathbf{m}})^3\right|\right\}\end{aligned}$$

for some fixed constant $M > 0$. On multiplying both sides of the above inequality by $|T_n|^3$ and taking expectation gives

$$\begin{aligned}&\mathbb{E}_{\beta,\mathbf{Q},\boldsymbol{\mu}}[T_n^6] \\ &\quad\le (2/5)n^{-1/2}\mathbb{E}_{\beta,\mathbf{Q},\boldsymbol{\mu}}|T_n|^8 \\ &\qquad + 3M\left\{\begin{array}{l} n^{3/4}\mathbb{E}_{\beta,\mathbf{Q},\boldsymbol{\mu}}[|T_n|^3|\overline{\mathbf{X}} - \overline{\mathbf{m}}|] + n^{-1/2}\mathbb{E}_{\beta,\mathbf{Q},\boldsymbol{\mu}}\left[|T_n|^4\sum_{i=1}^n (m_i - \overline{\mathbf{m}})^2\right] \\ + n^{-1/4}\mathbb{E}_{\beta,\mathbf{Q},\boldsymbol{\mu}}\left[|T_n|^3\left|\sum_{i=1}^n (m_i - \overline{\mathbf{m}})^3\right|\right] + n^{-1/4}\mathbb{E}_{\beta,\mathbf{Q},\boldsymbol{\mu}}|T_n|^3\sum_{i=1}^n \mu_i \end{array}\right\} \\ &\qquad + 3n^{3/2}|\mathbb{E}_{\beta,\mathbf{Q},\boldsymbol{\mu}}(T_n - T_n')T_n^3|.\end{aligned}$$

This is the analogue of Deb and Mukherjee ((2023), equation 4.15) in the case when $\boldsymbol{\mu}$ is not necessarily $\mathbf{0}$. Also, using part (c)(ii) of Lemma 14 we have

$$\begin{aligned}&\mathbb{E}_{\beta,\mathbf{Q},\boldsymbol{\mu}}\left[\sum_{i=1}^n (m_i - \bar{\mathbf{m}})^2\right]^p \lesssim (n\alpha_n^2 + n\overline{\boldsymbol{\mu}})^p, \\ &\mathbb{E}_{\beta,\mathbf{Q},\boldsymbol{\mu}} \max_{1\le i\le N} |m_i - \bar{\mathbf{m}}|^p \lesssim (\alpha_n + \sqrt{\overline{\boldsymbol{\mu}}})^p,\end{aligned} \tag{75}$$

where we use the fact that $\bar{\boldsymbol{\mu}} \lesssim \frac{\overline{d}}{n} \le \alpha_n$, as $\overline{d} \gg \sqrt{n}$. This is the analogue of Deb and Mukherjee ((2023), equation 4.17). Hereon, proceeding similarly as in the derivation of Deb and Mukherjee ((2023), equation 4.13) gives (73).

- Proof of (74)

This proof is similar to the derivation of Deb and Mukherjee ((2023), equation 4.14).

With $\tilde{\mathbf{Q}}$ as defined in the proof of Lemma 14 part (c)(ii), set $\mathbf{c}^\top := (d_1/\overline{d} - 1, \dots, d_n/\overline{d} - 1)$, $(\mathbf{c}^{(\ell)})^\top := \mathbf{c}^\top(\tilde{Q})^\ell$ and $x_\ell := \mathbb{E}_{\beta,\mathbf{Q},\boldsymbol{\mu}}[\sum_{i=1}^n c_i^{(\ell)} X_i]^2$. Note that, we can write $x_\ell = T_{1\ell} + T_{2\ell} + T_{3\ell}$ where

$$T_{1\ell} := \mathbb{E}_{\beta,\mathbf{Q},\boldsymbol{\mu}}\left[\sum_{i=1}^n c_i^{(l)}\big(X_i - \tanh(m_i + \mu_i)\big)\right]^2,$$

$$T_{2\ell} := \mathbb{E}_{\beta,\mathbf{Q},\boldsymbol{\mu}}\left[\sum_{i=1}^n c_i^{(l)} \tanh(m_i + \mu_i)\right]^2,$$

$$T_{3\ell} := 2\mathbb{E}_{\beta,\mathbf{Q},\boldsymbol{\mu}}\left[\sum_{i \neq j} c_i^{(\ell)} c_j^{(\ell)} \big(X_i - \tanh(m_i + \mu_i)\big) \tanh(m_i + \mu_i)\right].$$

Using Lemma 4 it follows that

$$T_{1\ell} \lesssim \|\mathbf{c}^{(\ell)}\|_2^2 \le \|\mathbf{c}\|_2^2. \tag{76}$$

For controlling $T_{3\ell}$ setting $m_i^{(j)} := \sum_{k \neq j} \mathbf{Q}_{ik} X_k$ as before we have

$$\begin{aligned} |T_{3\ell}| &= 2\left|\sum_{i \neq j} c_i^{(\ell)} c_j^{(\ell)} \mathbb{E}_{\beta,\mathbf{Q},\boldsymbol{\mu}} \big(X_i - \tanh(m_i + \mu_i)\big)\big(\tanh(m_i + \mu_i) - \tanh\big(m_i^{(j)} + \mu_i\big)\big)\right| \\ &\lesssim \sum_{i \neq j} |c_i^{(\ell)}||c_j^{(\ell)}| \mathbf{Q}_{ij} \lesssim \|\mathbf{c}^{(\ell)}\|_2^2 \le \|\mathbf{c}\|_2^2. \end{aligned} \tag{77}$$

For bounding $T_{2\ell}$, note that

$$\begin{aligned} & n|\bar{\mathbf{X}} - \bar{\mathbf{m}}| \\ &\quad = \left|\sum_{i=1}^n c_i X_i\right| \\ &\quad \le \left|\sum_{i=1}^n c_i \big(X_i - \tanh(m_i + \mu_i)\big)\right| + \left|\sum_{i=1}^n c_i \big(\tanh(m_i + \mu_i) - \tanh(\bar{\mathbf{m}} + \mu_i)\big)\right| \\ &\qquad + \left|\sum_{i=1}^n c_i \tanh(\mu_i)\right| \\ &\quad \lesssim \left|\sum_{i=1}^n c_i \big(X_i - \tanh(m_i + \mu_i)\big)\right| + \|\mathbf{c}\|_2 \sqrt{\sum_{i=1}^n (m_i - \bar{\mathbf{m}})^2} + n\alpha_n \bar{\boldsymbol{\mu}}, \end{aligned}$$

where the last step uses the bound $\max_{i \in [n]} |c_i| = \max_{i \in [n]} |\frac{d_i}{\bar{d}} - 1| \lesssim \alpha_n$. Consequently, for any positive integer p using (75) we have

$$\begin{aligned} \mathbb{E}_{\beta,\mathbf{Q},\boldsymbol{\mu}} (\overline{\mathbf{X}} - \overline{\mathbf{m}})^{2p} &\le n^{-2p} \big(\|\mathbf{c}\|_2^{2p} \big(1 + n^p \big(\alpha_n^2 + \overline{\boldsymbol{\mu}}\big)^p\big) + (n\alpha_n \bar{\boldsymbol{\mu}})^{2p}\big) \\ &\le n^{-2p} \big[n^p \alpha_n^{2p} \big(1 + n^p \alpha_n^{2p} + n^p \overline{\boldsymbol{\mu}}^p\big) + \alpha_n^{2p} \overline{d}^p\big] \\ &\lesssim \frac{(\log n)^{2p}}{\overline{d}^{2p}} \lesssim \frac{1}{n^p}, \end{aligned} \tag{78}$$

where the last line uses the bound

$$\|\mathbf{c}\|_2^2 = \sum_{i=1}^n \left(\frac{d_i}{\overline{d}} - 1\right)^2 \lesssim n\alpha_n^2, \qquad \sum_{i=1}^n \mu_i \le \sqrt{\overline{d}}, \quad \overline{d} \ge \sqrt{n} \log n.$$

In the subsequent proof, unless otherwise stated, (78) will always be invoked with $p = 1$. Combining (78) and (66), we get

$$\nu_n \lesssim 1 + n^{-1/2} \left(\sum_{i=1}^n \mu_i\right)^2 + \frac{n^{3/2} (\log n)^2}{\overline{d}^2} \lesssim \sqrt{n},$$

which again on invoking (78) (with $p=3$) gives

$$\mathbb{E}_{\beta,\mathbf{Q},\boldsymbol{\mu}}\overline{\mathbf{X}}^6 = \frac{\nu_n}{n^{3/2}} \lesssim n^{-1}, \qquad \mathbb{E}_{\boldsymbol{\mu}}\overline{\boldsymbol{\mu}}^6 \lesssim n^{-1}, \qquad \mathbb{E}_{\beta,\mathbf{Q},\boldsymbol{\mu}}\left[\sum_{i=1}^n m_i^6\right] \lesssim 1. \tag{79}$$

Now, a Taylor's series expansion gives $\tanh(m_i+\mu_i) = \tanh(m_i) + \mu_i\xi_i$ for bounded random variables ξ_i, and so

$$\begin{aligned} T_{2\ell} = {} & \mathbb{E}_{\beta,\mathbf{Q},\boldsymbol{\mu}}\left[\sum_{i=1}^n c_i^{(\ell)}\tanh(m_i)\right]^2 + 2\left\{\mathbb{E}_{\beta,\mathbf{Q},\boldsymbol{\mu}}\left[\sum_{i=1}^n c_i^{(\ell)}\tanh(m_i)\right]\right\}\left(\sum_{i=1}^n c_i^{(\ell)}\xi_i\mu_i\right) \\ & + \left(\sum_{i=1}^n c_i^{(\ell)}\xi_i\mu_i\right)^2. \end{aligned} \tag{80}$$

Setting $\theta_n := 1 + \mathbb{E}_{\beta,\mathbf{Q},\boldsymbol{\mu}}(n^{1/4}\overline{\mathbf{X}})^2$ and invoking (78) and (75) the terms in the RHS of (80) can be estimated as

$$\mathbb{E}_{\beta,\mathbf{Q},\boldsymbol{\mu}}\left[\sum_{i=1}^n c_i^{(\ell)}\tanh(m_i)\right]^2 \le \|\mathbf{c}^{(\ell)}\|_2^2\cdot\left[\mathbb{E}_{\mu}\sum_{i=1}^n(m_i-\bar{\mathbf{m}})^2 + n\mathbb{E}_{\beta,\mathbf{Q},\boldsymbol{\mu}}\bar{\mathbf{m}}^2\right] \lesssim \|\mathbf{c}\|_2^2\sqrt{n}\theta_n,$$

$$\left|\sum_{i=1}^n c_i^{(\ell)}\xi_i\mu_i\right| \le \max_{i\in[n]}\sum_{i=1}^n |c_i^{(\ell)}|\cdot\sum_{i=1}^n\mu_i \le \max_{i\in[n]}|c_i|\sum_{i=1}^n\mu_i \lesssim 1.$$

Combining the above estimate with (78) and (79) and repeating the derivation of Deb and Mukherjee ((2023), (4.32)), we get the existence of $M<\infty$ such that for all $\ell\ge 1$ we have

$$x_\ell \le x_{\ell+1} + 2M\sqrt{x_{\ell+1}}\beta_n + M^2\beta_n^2, \quad \beta_n := 1 + \|\mathbf{c}\|_2\sqrt{\theta_n}.$$

The above relation is similar to Deb and Mukherjee ((2023), (4.32)). Proceeding in a similar manner, setting $L = D(\log n)^2$ with D large enough, an inductive argument gives $x_\ell \le (L-\ell+1)^2M^2\beta_n^2$, giving

$$x_0 = \mathbb{E}_{\beta,\mathbf{Q},\boldsymbol{\mu}}\left[\sum_{i=1}^n\left(\frac{d_i}{\overline{d}}-1\right)X_i\right]^2 \le (L+1)^2M^2\beta_n^2 \lesssim (\log n)^4(1+\|\mathbf{c}\|_2^2\theta_n),$$

which verifies (74).

- Proof of (68) A direct expansion gives

$$\begin{aligned} & \mathbb{E}_{\beta,\mathbf{Q},\boldsymbol{\mu}}\left[\sum_{i=1}^n(X_i - \tanh(m_i+\mu_i))\right]^2 \\ & = \mathbb{E}_{\beta,\mathbf{Q},\boldsymbol{\mu}}\left[\sum_{i=1}^n \operatorname{sech}^2(m_i+\mu_i)\right] \\ & \quad + \mathbb{E}_{\beta,\mathbf{Q},\boldsymbol{\mu}}\left[\sum_{i\ne j}(X_i-\tanh(m_i+\mu_i))\left(\tanh(m_j^i+\mu_j)-\tanh(m_j+\mu_j)\right)\right] \\ & = \mathbb{E}_{\beta,\mathbf{Q},\boldsymbol{\mu}}\left[\sum_{i=1}^n \operatorname{sech}^2(m_i+\mu_i)\right] \\ & \quad + \mathbb{E}_{\beta,\mathbf{Q},\boldsymbol{\mu}}\left[\sum_{i\ne j}(1-X_i\tanh(m_i+\mu_i))(-\mathbf{Q}_{ji}\operatorname{sech}^2(m_j^i+\mu_j))\right] \end{aligned}$$

$$+ O\left(\sum_{i,j=1}^{n} \mathbf{Q}_{ij}^2\right)$$

$$= \mathbb{E}_{\beta,\mathbf{Q},\boldsymbol{\mu}}\left[\sum_{i=1}^{n} \operatorname{sech}^2(m_i+\mu_i)\left(1-\sum_{j=1}^{n}\mathbf{Q}_{ji}\operatorname{sech}^2(m_j+\mu_j)\right)\right] + O\left(\sum_{i,j=1}^{n}\mathbf{Q}_{ij}^2\right).$$

The first term of the above display, splits into two terms as follows:

$$\mathbb{E}_{\beta,\mathbf{Q},\boldsymbol{\mu}}\left[\sum_{i=1}^{n}\operatorname{sech}^2(m_i+\mu_i)\left(1-\frac{d_i}{\overline{d}}\right)\right] + \mathbb{E}_{\beta,\mathbf{Q},\boldsymbol{\mu}}\left[\sum_{i,j}\mathbf{Q}_{ij}\operatorname{sech}^2(m_i+\mu_i)\tanh^2(m_j+\mu_j)\right]$$

$$\overset{(a)}{\lesssim} \mathbb{E}_{\beta,\mathbf{Q},\boldsymbol{\mu}}\left[\left|\sum_{i=1}^{n}\left(\operatorname{sech}^2(\overline{\mathbf{m}})+\xi_{i1}\mu_i+\xi_{i2}(m_i-\overline{\mathbf{m}})\right)\left(1-\frac{d_i}{\overline{d}}\right)\right|\right]$$

$$+ \mathbb{E}_{\beta,\mathbf{Q},\boldsymbol{\mu}}\left[\sum_{i,j}\mathbf{Q}_{ij}\tanh^2(m_j+\mu_j)\right]$$

$$\lesssim \max_i\left|\frac{d_i}{\overline{d}}-1\right|\sum_{i=1}^{n}\mu_i + \sum_{i=1}^{n}\mathbb{E}_{\beta,\mathbf{Q},\boldsymbol{\mu}}\left|\left(\frac{d_i}{\overline{d}}-1\right)(m_i-\overline{\mathbf{m}})\right| + \mathbb{E}_{\beta,\mathbf{Q},\boldsymbol{\mu}}\left[\sum_{i,j}\mathbf{Q}_{ij}(m_j^2+\mu_j^2)\right]$$

$$\lesssim 1 + \sqrt{\sum_{i=1}^{n}\left(\frac{d_i}{\overline{d}}-1\right)^2}\sqrt{\mathbb{E}_{\beta,\mathbf{Q},\boldsymbol{\mu}}\sum_{i=1}^{n}(m_i-\bar{\mathbf{m}})^2}$$

$$+ \mathbb{E}_{\beta,\mathbf{Q},\boldsymbol{\mu}}\left[\sum_{i=1}^{n}(m_i-\overline{\mathbf{m}})^2\right] + n\mathbb{E}_{\beta,\mathbf{Q},\boldsymbol{\mu}}\overline{\mathbf{m}}^2 + \sum_{j=1}^{n}\mu_j^2$$

$$\overset{(b)}{\lesssim} 1 + \sqrt{\frac{n\log n}{\overline{d}}}\cdot\sqrt{\frac{n(\log n)^3}{\overline{d}} + \sum_{i=1}^{n}\mu_i} + \frac{n(\log n)^3}{\overline{d}}$$

$$+ \sum_{i=1}^{n}\mu_i + n\mathbb{E}_{\beta,\mathbf{Q},\boldsymbol{\mu}}(\overline{\mathbf{X}}-\overline{\mathbf{m}})^2 + n\mathbb{E}_{\beta,\mathbf{Q},\boldsymbol{\mu}}[\overline{\mathbf{X}}]^2$$

$$+ \left(\max_j \mu_j\right)\sum_{j=1}^{n}\mu_j \lesssim \sqrt{n} + \mathbb{E}_{\beta,\mathbf{Q},\boldsymbol{\mu}}[\overline{\mathbf{X}}]^2 \lesssim n^{1/3}\left(\sum_{i=1}^{n}\mu_i\right)^{2/3}.$$

Here (a) follows from standard Taylor expansions. Note that ξ_{i1} and ξ_{i2} are uniformly bounded random variables. The bounds in (b), (c) and (d) are consequences of Lemma 14 part (c)(ii), (78) and (66) respectively.

Funding. The research of Sumit Mukherjee was supported in part by NSF Grant DMS-1712037.

The research of Ming Yuan was supported in part by NSF Grant DMS-2015285.

REFERENCES

ADDARIO-BERRY, L., BROUTIN, N., DEVROYE, L. and LUGOSI, G. (2010). On combinatorial testing problems. *Ann. Statist.* **38** 3063–3092. MR2722464 https://doi.org/10.1214/10-AOS817

AHMED, A. and XING, E. P. (2009). Recovering time-varying networks of dependencies in social and biological studies. *Proc. Natl. Acad. Sci. USA* **106** 11878–11883.

AIZENMAN, M., BARSKY, D. J. and FERNÁNDEZ, R. (1987). The phase transition in a general class of Ising-type models is sharp. *J. Stat. Phys.* **47** 343–374. MR0894398 https://doi.org/10.1007/BF01007515

ARIAS-CASTRO, E., CANDÈS, E. J. and DURAND, A. (2011). Detection of an anomalous cluster in a network. *Ann. Statist.* **39** 278–304. MR2797847 https://doi.org/10.1214/10-AOS839

ARIAS-CASTRO, E., CASTRO, R. M., TÁNCZOS, E. and WANG, M. (2018). Distribution-free detection of structured anomalies: Permutation and rank-based scans. *J. Amer. Statist. Assoc.* **113** 789–801. MR3832227 https://doi.org/10.1080/01621459.2017.1286240

ARIAS-CASTRO, E., DONOHO, D. L. and HUO, X. (2005). Near-optimal detection of geometric objects by fast multiscale methods. *IEEE Trans. Inf. Theory* **51** 2402–2425. MR2246369 https://doi.org/10.1109/TIT.2005.850056

BASAK, A. and MUKHERJEE, S. (2017). Universality of the mean-field for the Potts model. *Probab. Theory Related Fields* **168** 557–600. MR3663625 https://doi.org/10.1007/s00440-016-0718-0

BHAMIDI, S., JIN, J. and NOBEL, A. (2018). Change point detection in network models: Preferential attachment and long range dependence. *Ann. Appl. Probab.* **28** 35–78. MR3770872 https://doi.org/10.1214/17-AAP1297

BUTUCEA, C. and INGSTER, Y. I. (2013). Detection of a sparse submatrix of a high-dimensional noisy matrix. *Bernoulli* **19** 2652–2688. MR3160567 https://doi.org/10.3150/12-BEJ470

CHATTERJEE, S. (2005). *Concentration Inequalities with Exchangeable Pairs*. ProQuest LLC, Ann Arbor, MI. Thesis (Ph.D.)—Stanford University. MR2707160

CHATTERJEE, S. (2007). Stein's method for concentration inequalities. *Probab. Theory Related Fields* **138** 305–321. MR2288072 https://doi.org/10.1007/s00440-006-0029-y

CHO, H. (2016). Change-point detection in panel data via double CUSUM statistic. *Electron. J. Stat.* **10** 2000–2038. MR3522667 https://doi.org/10.1214/16-EJS1155

DASKALAKIS, C., DIKKALA, N. and KAMATH, G. (2019). Testing Ising models. *IEEE Trans. Inf. Theory* **65** 6829–6852. MR4030862 https://doi.org/10.1109/TIT.2019.2932255

DEB, N. and MUKHERJEE, S. (2023). Fluctuations in mean-field Ising models. *Ann. Appl. Probab.* **33** 1961–2003. MR4583662 https://doi.org/10.1214/22-aap1857

DEMBO, A. and MONTANARI, A. (2010a). Gibbs measures and phase transitions on sparse random graphs. *Braz. J. Probab. Stat.* **24** 137–211. MR2643563 https://doi.org/10.1214/09-BJPS027

DEMBO, A. and MONTANARI, A. (2010b). Ising models on locally tree-like graphs. *Ann. Appl. Probab.* **20** 565–592. MR2650042 https://doi.org/10.1214/09-AAP627

DUMINIL-COPIN, H. (2020). Lectures on the Ising and Potts models on the hypercubic lattice. In *Random Graphs, Phase Transitions, and the Gaussian Free Field. Springer Proc. Math. Stat.* **304** 35–161. Springer, Cham. MR4043224 https://doi.org/10.1007/978-3-030-32011-9_2

DUMINIL-COPIN, H., RAOUFI, A. and TASSION, V. (2019). Sharp phase transition for the random-cluster and Potts models via decision trees. *Ann. of Math.* (2) **189** 75–99. MR3898174 https://doi.org/10.4007/annals.2019.189.1.2

DUMINIL-COPIN, H. and TASSION, V. (2016). A new proof of the sharpness of the phase transition for Bernoulli percolation and the Ising model. *Comm. Math. Phys.* **343** 725–745. MR3477351 https://doi.org/10.1007/s00220-015-2480-z

ELLIS, R. S. and NEWMAN, C. M. (1978). The statistics of Curie–Weiss models. *J. Stat. Phys.* **19** 149–161. MR0503332 https://doi.org/10.1007/BF01012508

ENIKEEVA, F., MUNK, A., POHLMANN, M. and WERNER, F. (2020). Bump detection in the presence of dependency: Does it ease or does it load? *Bernoulli* **26** 3280–3310. MR4140545 https://doi.org/10.3150/20-BEJ1226

ENIKEEVA, F., MUNK, A. and WERNER, F. (2018). Bump detection in heterogeneous Gaussian regression. *Bernoulli* **24** 1266–1306. MR3706794 https://doi.org/10.3150/16-BEJ899

FEIGE, U. and OFEK, E. (2005). Spectral techniques applied to sparse random graphs. *Random Structures Algorithms* **27** 251–275. MR2155709 https://doi.org/10.1002/rsa.20089

FRIEDLI, S. and VELENIK, Y. (2018). *Statistical Mechanics of Lattice Systems: A Concrete Mathematical Introduction*. Cambridge Univ. Press, Cambridge. MR3752129

FRIEDMAN, J., KAHN, J. and SZEMERÉDI, E. (1989). On the second eigenvalue of random regular graphs. In *Proceedings of the Twenty-First Annual ACM Symposium on Theory of Computing*, *STOC'*89 587–598, Association for Computing Machinery, New York, NY, USA. https://doi.org/10.1145/73007.73063

GAO, P., ISAEV, M. and MCKAY, B. D. (2020). Sandwiching random regular graphs between binomial random graphs. In *Proceedings of the* 2020 *ACM-SIAM Symposium on Discrete Algorithms* 690–701. SIAM, Philadelphia, PA. MR4141224

GEMAN, S. and GEMAN, D. (1984). Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. *IEEE Trans. Pattern Anal. Mach. Intell.* **6** 721–741.

GRIMMETT, G. (2006). *The Random-Cluster Model. Grundlehren der Mathematischen Wissenschaften* [*Fundamental Principles of Mathematical Sciences*] **333**. Springer, Berlin. MR2243761 https://doi.org/10.1007/978-3-540-32891-9

HALL, P. and JIN, J. (2008). Properties of higher criticism under strong dependence. *Ann. Statist.* **36** 381–402. MR2387976 https://doi.org/10.1214/009053607000000767

HALL, P. and JIN, J. (2010). Innovated higher criticism for detecting sparse signals in correlated noise. *Ann. Statist.* **38** 1686–1732. MR2662357 https://doi.org/10.1214/09-AOS764

HORVÁTH, L. and HUŠKOVÁ, M. (2012). Change-point detection in panel data. *J. Time Series Anal.* **33** 631–648. MR2944843 https://doi.org/10.1111/j.1467-9892.2012.00796.x

ISING, E. (1925). Beitrag Zur Theorie des Ferromagnetismus. *Z. Phys. Hadrons Nucl.* **31** 253–258.

JAIN, V., KOEHLER, F. and MOSSEL, E. (2018). The mean-field approximation: Information inequalities, algorithms, and complexity. In *Proceedings of the* 31*st Conference on Learning Theory*. *Proceedings of Machine Learning Research* **75** 1326–1347. PMLR. https://proceedings.mlr.press/v75/jain18b.html.

JÄRPE, E. (1999). Surveillance of the interaction parameter of the Ising model. *Comm. Statist. Theory Methods* **28** 3009–3027. MR1729012 https://doi.org/10.1080/03610929908832462

KÖNIG, C., MUNK, A. and WERNER, F. (2020). Multidimensional multiscale scanning in exponential families: Limit theory and statistical consequences. *Ann. Statist.* **48** 655–678. MR4102671 https://doi.org/10.1214/18-AOS1806

LEBOWITZ, J. L. (1974). GHS and other inequalities. *Comm. Math. Phys.* **35** 87–92. MR0339738

LIU, H., GAO, C. and SAMWORTH, R. J. (2021). Minimax rates in sparse, high-dimensional change point detection. *Ann. Statist.* **49** 1081–1112. MR4255120 https://doi.org/10.1214/20-aos1994

MUKHERJEE, R., MUKHERJEE, S. and YUAN, M. (2018). Global testing against sparse alternatives under Ising models. *Ann. Statist.* **46** 2062–2093. MR3845011 https://doi.org/10.1214/17-AOS1612

MUKHERJEE, R. and RAY, G. (2022). On testing for parameters in Ising models. *Ann. Inst. Henri Poincaré Probab. Stat.* **58** 164–187. MR4375621 https://doi.org/10.1214/21-aihp1157

ONSAGER, L. (1944). Crystal statistics. I. A two-dimensional model with an order-disorder transition. *Phys. Rev.* (2) **65** 117–149. MR0010315

RAY, B. K. and TSAY, R. S. (2002). Bayesian methods for change-point detection in long-range dependent processes. *J. Time Series Anal.* **23** 687–705. MR1944606 https://doi.org/10.1111/1467-9892.00286

SHARPNACK, J., RINALDO, A. and SINGH, A. (2016). Detecting anomalous activity on networks with the graph Fourier scan statistic. *IEEE Trans. Signal Process.* **64** 364–379. MR3446216 https://doi.org/10.1109/TSP.2015.2481866

STAUFFER, D. (2008). Social applications of two-dimensional Ising models. *Am. J. Phys.* **76** 470–473.

WALTHER, G. (2010). Optimal and fast detection of spatial clusters with scan statistics. *Ann. Statist.* **38** 1010–1033. MR2604703 https://doi.org/10.1214/09-AOS732

ZOU, S., LIANG, Y. and POOR, H. V. (2017). Nonparametric detection of geometric structures over networks. *IEEE Trans. Signal Process.* **65** 5034–5046. MR3694274 https://doi.org/10.1109/TSP.2017.2718977