Information and Inference: A Journal of the IMA (2022) **11**, 55–102
https://doi.org/10.1093/imaiai/iaaa032
Advance Access publication on 26 December 2020

High-temperature structure detection in ferromagnets

YUAN CAO
Department of Computer Science, University of California, Los Angeles, CA 90095, USA

MATEY NEYKOV†
Department of Statistics & Data Science, Carnegie Mellon University, Pittsburgh, PA 15213, USA
†Corresponding author: Email: mneykov@stat.cmu.edu

AND

HAN LIU
Department of Electrical Engineering and Computer Science and Department of Statistics, Northwestern University, Evanston, IL 60208, USA

[Received on 8 January 2020; revised on 26 September 2020; accepted on 5 October 2020]

This paper studies structure detection problems in high-temperature ferromagnetic (positive interaction only) Ising models. The goal is to distinguish whether the underlying graph is empty, i.e., the model consists of independent Rademacher variables, vs. the alternative that the underlying graph contains a subgraph of a certain structure. We give matching upper and lower minimax bounds under which testing this problem is possible/impossible, respectively. Our results reveal that a key quantity called graph arboricity drives the testability of the problem. On the computational front, under a conjecture of the computational hardness of sparse principal component analysis, we prove that, unless the signal is strong enough, there are no polynomial time tests which are capable of testing this problem. In order to prove this result, we exhibit a way to give sharp inequalities for the even moments of sums of i.i.d. Rademacher random variables which may be of independent interest.

Keywords: graph structure detection; minimax testing; ferromagnetic Ising model; total variation distance.

1. Introduction

Graphical models are a powerful tool in high-dimensional statistical inference. The graph structure of a graphical model gives a simple way to visualize the dependency among the variables in multivariate random vectors. The analysis of graph structures plays a fundamental role in a wide variety of applications, including information retrieval, bioinformatics, image processing and social networks [**10**,**18**,**30**,**59**]. Motivated by these applications, theoretical results on graph estimation [**6**,**15**,**40**,**45**,**52**], single edge inference [**32**,**36**,**50**,**53**] and combinatorial inference [**48**,**49**] have been studied in the literature.

In this paper, we are concerned with the distinct problem of *structure detection*. In structure detection problems, one is interested in testing whether the underlying graph is empty (i.e., the random variables are independent) vs. the alternative that the graph contains a subgraph of a certain structure. A variety of detection problems have been previously considered in the literature [see for example**1**,**3**–**5**]. These works mainly focus on covariance or precision matrix detection problems and establish minimax lower and upper bounds.

While covariance and precision matrix detection problems are inherently related to the Gaussian graphical model, in this paper, we focus on detection problems under the *zero-field ferromagnetic Ising model*. The Ising model is a probability model for binary data originally developed in statistical mechanics [35] and has wide range of modern applications including image processing [28], social networks and bioinformatics [2]. Below, we formally introduce the model and problems of interest.

Zero-field ferromagnetic Ising model. Under a zero-field Ising model, the binary vector $\boldsymbol{X} \in \{\pm 1\}^d$ follows a distribution with probability mass function given by

$$\mathbb{P}_{\Theta}(\boldsymbol{X}) = \frac{1}{Z_{\Theta}} \exp\left(\sum_{i,j=1}^{d} \theta_{ij} X_i X_j\right),$$

where $\Theta = (\theta_{ij})_{d\times d}$ is a symmetric interaction matrix with zero diagonal entries and Z_{Θ} is the partition function defined as

$$Z_{\Theta} = \sum_{\boldsymbol{X}\in\{\pm 1\}^d} \exp\left(\sum_{i,j=1}^{d} \theta_{ij} X_i X_j\right).$$

The non-zero elements of the symmetric matrix Θ specify a graph $G(\Theta) = G = (\overline{V}, E)$ with the vertex set $\overline{V} = \{1, \ldots, d\}$ and the edge set $E = \{(i,j) : \theta_{ij} \neq 0\}$. We will refer to the graph $G(\Theta)$ as G whenever it is clear what the underlying matrix Θ is. It is not hard to check that by the definition of G, the vector $\boldsymbol{X}$ is Markov with respect to G, that is, each two elements X_i and X_j are independent given the remaining values of $\boldsymbol{X}_{-(i,j)}$ if and only if $(i,j) \notin E$.

Here, the term *zero-field* specifies that there is no external magnetic field affecting the system, meaning that the energy function $\sum_{i,j=1}^{d} \theta_{ij} X_i X_j$ consists purely the terms of degree 2 (i.e., there are no main effects). In this paper, we further focus on zero-field *ferromagnetic* models, where we also assume that $\theta_{ij} \geqslant 0$, $i,j \in \{1, \ldots, d\}$. In addition, our analysis is under the high-temperature setting, where the magnitudes of θ_{ij}s are under a certain level. More specifically, throughout this paper, we assume that $\|\Theta\|_F \leqslant \frac{1}{2}$, where $\|\Theta\|_F = \left[\sum_{i,j=1}^{d} \theta_{ij}^2\right]^{1/2}$ is the Frobenius norm of Θ.

Structure detection problems. As described in the previous paragraph, a zero-field ferromagnetic Ising model specifies a graph $G = (\overline{V}, E)$. In a structure detection problem, we are interested in testing whether the underlying graph G is an empty graph vs. the alternative that G belongs to a set of graphs with a certain structure. Specifically, let $G_{\varnothing} = (\overline{V}, \emptyset)$ be the empty graph, and let $\mathscr{G}_1$ be a class of graphs not containing $G_{\varnothing}$. The following hypothesis testing problem is an example of a detection problem. Given a sample of n independent observations $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n \in \mathbb{R}^d$ from a zero-field ferromagnetic Ising model, we aim to test

$$\mathrm{H}_0 : G = G_{\varnothing} \ \text{ vs. } \ \mathrm{H}_1 : G \in \mathscr{G}_1. \tag{1.1}$$

The term detection here is used in the sense that if one rejects the null hypothesis, the presence of a non-null graph has been detected. In (1.1), the graph class $\mathscr{G}_1$ can be arbitrary, which makes the hypothesis testing problem (1.1) a very general problem. We now give a specific instance of this problem

which is of particular importance. Let G_* be a fixed graph with $s = o(\sqrt{d})$[1] non-isolated vertices which represents some specific graph structure. The structure detection problem that considers all possible 'positions' of G_* is of the following form:

$$\mathrm{H}_0 : G = G_\varnothing \text{vs.} \mathrm{H}_1 : G \in \mathscr{G}_1(G_*), \tag{1.2}$$

where $\mathscr{G}_1(G_*)$ is the class of all graphs that contain a subgraph isomorphic to G_*.

While problems (1.1) and (1.2) give a good intuition what a detection problem is, in order to facilitate testing, we need to impose certain assumptions on the matrix Θ, as otherwise even with graphs vastly different from the empty graph there might not be enough 'separation' between the null and the alternative hypothesis. Since the underlying graph G is specified by the matrix Θ, we can reformulate problems (1.1) and (1.2) into testing problems on Θ. Given a class of graphs $\mathscr{G}_1$, we define the corresponding parameter space with minimum signal strength $\theta > 0$ as

$$\mathscr{S}(\mathscr{G}_1, \theta) = \left\{ \Theta = (\theta_{ij})_{d\times d} : \Theta = \Theta^T, G(\Theta) \in \mathscr{G}_1, \|\Theta\|_F \leqslant 1/2, \min_{(i,j)\in E[G(\Theta)]} \theta_{ij} \geqslant \theta \right\}. \tag{1.3}$$

We now reformulate the hypothesis testing problems (1.1) and (1.2) as follows:

$$\mathrm{H}_0 : \Theta = \mathbf{0} \text{ vs. } \mathrm{H}_1 : \Theta \in \mathscr{S}(\mathscr{G}_1, \theta), \tag{1.4}$$

$$\mathrm{H}_0 : \Theta = \mathbf{0} \text{ vs. } \mathrm{H}_1 : \Theta \in \mathscr{S}[\mathscr{G}_1(G_*), \theta]. \tag{1.5}$$

The results of our paper cover the following examples.

Empty graph vs. non-empty graph. We consider testing whether the underlying graph of the Ising model is empty or not. Clearly, since our null hypothesis is that the graph is empty, this is a detection problem. We have $\mathscr{G}_1 = \{G : E(G) \neq \emptyset\}$.

Clique detection. A clique is a set of vertices such that every two distinct vertices are adjacent. We consider detecting graphs that contain a clique of size s. We have $\mathscr{G}_1 = \{G = (\overline{V}, E) : \exists V \subseteq \overline{V}$ such that $|V| = s$ and $(i,j) \in E$ for all $i,j \in V\}$. This is a more general version of the previous example, since one can think of a non-empty graph as a graph containing a clique of size $s = 2$.

Star detection. A star is a tree in which all leaves are connected to the same node. We consider detecting graphs that contain an $s-1$ star. In this example, we have $\mathscr{G}_1 = \{G = (\overline{V}, E)$: there exist distinct $i_0, i_1, \ldots, i_{s-1} \in \overline{V}$ such that $(i_0, i_1), (i_0, i_2), \ldots, (i_0, i_{s-1}) \in E\}$.

Community structure detection. In this example, we consider a class of graphs with more complex structure. Let k and l be positive integers. A community $\mathscr{C}$ is represented by a k-clique, which means that every two members in the same community are connected. For a community $\mathscr{C}$, we select one fixed representative vertex and denote it as $v(\mathscr{C})$. We consider the class of graphs $\mathscr{G}_1$ that contains graphs with

[1] For two positive sequences a_n and b_n, we write $a_n = o(b_n)$ if $\lim_{n\to\infty} a_n/b_n = 0$.

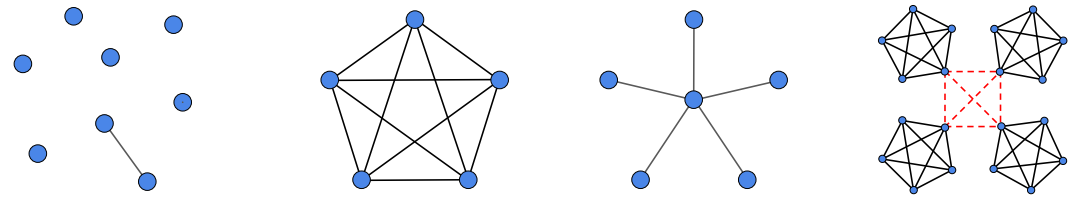

FIG. 1. Illustration of the examples considered in this paper. (a) shows a single-edge graph, (b) is a 5-clique, (c) is a 5-star, (d) is an example of a graph that has community structure with $k = 5$ and $l = 4$. We can write the detection problems as (1.5) by defining the corresponding shown graphs as G_*.

at least l disjoint communities, such that for every two different communities $\mathscr{C}$ and $\mathscr{C}'$, there exists an edge connecting $v(\mathscr{C})$ and $v(\mathscr{C}')$. In this example, we set $s = kl$.

All of the above examples are of the type (1.5). We show examples of these detection problems in Fig. 1. In the following section, we outline the main contributions of our work.

1.1 *Main contributions*

There are three major contributions of this paper.

First, we develop a novel technique to derive minimax lower bounds of structure detection problems in Ising models. Our proof technique relates the Ising model probability mass function and the χ^2-divergence between two distributions to the number of certain Eulerian subgraphs of the underlying graph. With this technique, we are able to obtain a general information-theoretic lower bound for arbitrary alternative hypothesis, which can be immediately applied to examples including any of the four examples described in the previous section.

Second, we propose a linear scan test on the sample covariance matrix that matches our minimax lower bound for arbitrary structure detection problems in certain regimes. Along with our general minimax lower bound result, this procedure reveals the fact that a quantity called *arboricity* (i.e., a certain maximum edge to vertex ratio of graphs in the alternative hypothesis) essentially determines the information-theoretic limit of the testing problem. This matches the intuition that in order to distinguish a graph with small signal strength from the empty graph, one need to examine the densest part of the graph. Furthermore, the denser the graph is, the easier it is to detect it, where the precise measurement of graph density turns out to be graph arboricity.

In addition, we also study the computational lower bound of the structure detection problems. Based on a conjecture on the computational hardness of sparse principal component analysis (PCA), which has been studied by recent works [**7**,**8**,**12**,**13**,**27**], we prove that no polynomial time test can detect structures successfully unless there is a sufficiently large signal strength. In order to prove this result, we exhibit a way to give sharp inequalities for the even moments of sums of i.i.d. Rademacher random variables which may be of independent interest. Furthermore, in addition to this result, we also derive another computational lower-bound result under the oracle computational model studied by [**23**,**24**,**58**].

1.2 *Related work*

Plenty of work has been done on graph estimation (also known as graph selection) in Ising models. [**54**] gave the first information-theoretic lower bounds of graph selection problems for bounded edge cardinality and bounded vertex degree models. Later, [**55**] proposed a general framework for obtaining

information-theoretic lower bounds for graph selection in ferromagnetic Ising models and showed that the lower bound is specified by certain structural conditions. On the other hand, [51] proposed an algorithm for structure learning based on l_1-regularized logistic regression that works in the high-temperature regime [6]. [14] gave a polynomial time algorithm that works for both low- and high-temperature regimes. In addition, see [41,57] for the best polynomial time structure learning algorithms. Compared with graph estimation, structure detection is a statistically easier problem. As a consequence, the limitations on signal strength that we exhibit in this paper are weaker than the corresponding requirements used in the graph estimation literature.

Structure detection problems have been studied in [1,3–5]. However, all these works focus on Gaussian random vectors. Specifically, [1] study testing the existence of specific subsets of components in a Gaussian vector whose means are non-zero based on a single observation. [3] consider the correlation graph of a Gaussian random vector and establish upper and lower bounds for detecting certain classes of fully connected cliques based on one sample. In a follow-up work, [4] generalize the result to multiple i.i.d. samples. [5] give another related result on detecting a region of a Gaussian Markov random field against a background of white noise. The major difference between these existing works and our work is that we focus on detection in the Ising model, and our results work not only for cliques but also for general graph structures. Recently, [43,48,49] proposed a novel problem where one considers testing whether the underlying graph obeys certain combinatorial properties. We stress that while related to structure detection, these problems are fundamentally different as structure detection is a statistically simpler task. It is not surprising, therefore, that the algorithms we develop are very different from those in the aforementioned works and the proofs of our lower bounds use different techniques.

Our result on computational lower bound follows the recent line of work on computational barriers for statistical models [7,8,12,13,27,44] based on the planted clique conjecture. [8] focus on the testing method based on minimum dual perturbation and semidefinite programming and prove that such polynomial time testing methods cannot attain the minimax optimal rate for sparse PCA. [7] prove the computational lower bound on a generalized sparse PCA problem which includes all multivariate distributions with certain tail probability assumptions on the quadratic form. [44] consider the Gaussian submatrix detection problem and propose a framework to analyze computational limits of continuous random variables via constructing a sequence of asymptotically equivalent discretized models. Inspired by the results in [44,27] consider the computational lower bound for Gaussian sparse canonical correlation analysis as well as sparse PCA problems. Our computational lower bound result is based on the previous studies on the sparse PCA problem. We summarize these results and directly base our result for Ising models on a sparse PCA conjecture.

Other related works on Ising models include the following. [9] study the Ising block model by providing efficient methods for block structure recovery as well as information-theoretic lower bounds. [46] study the upper and lower bounds for the detection of a sparse external magnetic field in Ising models. [16] consider goodness-of-fit and independence testing in Ising models using pairwise correlations. [29] establish concentration inequalities for polynomials of a random vector in contracting Ising models.

1.3 *Notation*

We use the following notations in our paper. For a vector $\mathrm{v} = (\mathrm{v}_1, \ldots, \mathrm{v}_d)^T \in \mathbb{R}^d$ and a number $1 \leqslant p < \infty$, let $\|\mathrm{v}\|_p = \left(\sum_{i=1}^d |\mathrm{v}_i|^p\right)^{1/p}$. We also define $\|\mathrm{v}\|_\infty = \max_i |\mathrm{v}_i|$. For a matrix A, we denote $\|\mathrm{A}\|_{\max} = \max_{j,k} |\mathrm{A}_{jk}|$, $\|\mathrm{A}\|_F = \left(\sum_{i,j=1}^d \mathrm{A}_{ij}^2\right)^{1/2}$ and $\|\mathrm{A}\|_p = \max_{\|\mathrm{v}\|_p=1} \|\mathrm{Av}\|_p$ for $p \geqslant 1$.

We also use the standard asymptotic notations $O(\cdot)$ and $o(\cdot)$. Let a_n and b_n be two sequences and assume that b_n is non-zero for large enough n. We write $a_n = O(b_n)$ if $\limsup_{n\to\infty} |a_n/b_n| < \infty$ and $a_n = o(b_n)$ if $\lim_{n\to\infty} a_n/b_n = 0$.

Let $\overline{V} = \{1, \ldots, d\}$ be the complete vertex set. In this paper, we consider graphs with d vertices over the vertex set $\overline{V}$. For a graph G, let $E(G) = \{(i,j) : G \text{ has an edge connecting vertex } i \text{ and } j\}$, where $(i,j) = (j,i)$ are undirected pairs. Moreover, we denote by $V(G) = \{i \in \overline{V} : G \text{ has an edge connecting vertex} i\}$ the set of non-isolated vertices of G.

1.4 *Organization of the paper*

Our paper is organized as follows. In Section 2, we present our main information-theoretic lower bound result as well as its applications to various detection problems. In Section 3, we develop a general procedure to construct optimal linear scan tests on the sample covariance matrix. In Section 4, we examine the computational limit of the polynomial time tests by comparing the Ising and sparse PCA models. Sections 5 and 6 contain the proofs of the main results of Sections 2 and 3, respectively. The remaining detailed proofs are all placed in Section A. In Section B, we provide an additional proof of a computational lower bound under the oracle computational model.

2. Lower bounds

The minimax risk of detection problem (1.4) is defined as

$$\gamma[\mathscr{S}(\mathscr{G}_1, \theta)] := \inf_{\psi} \left[\mathbb{P}_{0,n}(\psi = 1) + \max_{\Theta \in \mathscr{S}(\mathscr{G}_1, \theta)} \mathbb{P}_{\Theta,n}(\psi = 0) \right], \tag{2.1}$$

where $\mathbb{P}_{0,n}$ and $\mathbb{P}_{\Theta,n}$ are the joint probability measures of n i.i.d. samples under null and alternative hypotheses, respectively. The infimum in (2.1) is taken over all measurable test functions $\psi : \{\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n\} \mapsto \{0, 1\}$. If $\liminf_{n\to\infty} \gamma[\mathscr{S}(\mathscr{G}_1, \theta)] = 1$, we say that any test is asymptotically powerless.

In this section, we derive necessary conditions on the signal strength θ required for detection problems to admit tests which are not asymptotically powerless. Our results will show that the difficulty of testing an empty graph against $\mathscr{G}_1$ is determined by a quantity called arboricity, which was originally introduced in graph theory by [**47**] to quantify the minimum number of disjoint forests into which the edges of a given graph can be partitioned.

For a graph $G \in \mathscr{G}_1$ and a vertex set $V \subseteq \overline{V}$, let G_V be the graph obtained by restricting G on the vertices in V (i.e., removing all edges which are connected to vertices $\overline{V} \setminus V$). The arboricity of G is defined as follows:

$$\mathscr{R}(G) := \left\lceil \max_{V \subseteq \overline{V}} \frac{|E(G_V)|}{|V| - 1} \right\rceil, \tag{2.2}$$

where $\lceil \cdot \rceil$ is the ceiling function and $0/0$ is understood as 0. The arboricity of a graph measures how dense the graph is. For an illustration of arboricity, see Fig. 2. Let $G_\emptyset = (\overline{V}, \emptyset)$ denote the empty graph. By definition, $\mathscr{R}(G_\emptyset) = 0$.

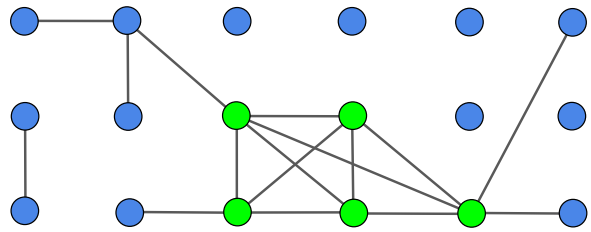

FIG. 2. Illustration of arboricity. Here, the nodes and black lines represent the vertices and edges of graph G, respectively. The vertex set V that maximizes $|E(G_V)|/(|V|-1)$ is denoted by green nodes, which also gives the densest subgraph of G. We have $\mathscr{R}(G)=3$.

For a given graph G the larger $\mathscr{R}(G)$ is, the more different $G_\emptyset$ and G are. We further define

$$\mathscr{R} := \min_{G\in\mathscr{G}_1} \mathscr{R}(G)$$

to measure the difference in graph density between $G_\emptyset$ and $\mathscr{G}_1$ in a worst-case sense. Let $\mathscr{G}^*$ be a non-empty subset of $\mathscr{G}_1$ such that all graphs in $\mathscr{G}^*$ have arboricity $\mathscr{R}$. By the definition of $\mathscr{R}$, such non-empty $\mathscr{G}^*$ exists and may not be unique. Our analysis works for arbitrary choices of $\mathscr{G}^*$ which satisfy the incoherence condition [49] defined as follows.

DEFINITION 2.1 (Negative association and incoherence condition) For $k \geqslant 0$, we say the random variables $Y_1,\dots,Y_k$ are negatively associated if for any $k_1,k_2 \geqslant 0$ with $k_1+k_2 \leqslant k$, any distinct indices $i_1,i_2,\dots,i_{k_1},j_1,\dots,j_{k_2}$ and any coordinate-wise non-decreasing functions f and g, we have

$$\mathrm{Cov}[f(Y_{i_1},\dots,Y_{i_{k_1}}),g(Y_{j_1},\dots,Y_{j_{k_2}})] \leqslant 0.$$

We say that the graph set $\mathscr{G}^*$ is incoherent if for any fixed graph G, the binary random variables

$$\{\mathbb{1}[i\in V(G')]\}_{i\in V(G)}$$

are negatively associated with respect to uniformly sampling $G'\in\mathscr{G}^*$.

For a graph G, we denote by A_G the adjacency matrix of G. Then, given $\mathscr{G}^*$, we define the corresponding parameter set with minimal signal strength θ as

$$\mathscr{S}^* = \{\Theta = \theta \mathrm{A}_G : G\in\mathscr{G}^*\}.$$

Let $V_{\max} = \max_{G\in\mathscr{G}^*}|V(G)|$, $\Lambda = \max_{G\in\mathscr{G}^*}\|\mathrm{A}_G\|_F$, $\Gamma = \max_{G\in\mathscr{G}^*}\|\mathrm{A}_G\|_1$ and $\mathscr{B} = 512\{\Lambda^4 \wedge [V_{\max}(\Gamma\vee\Lambda)^2]\}$. Then, for $\theta \leqslant (2\Lambda)^{-1}$, by definition (recall (1.3)), we have $\mathscr{S}^* \subseteq \mathscr{S}(\mathscr{G}_1,\theta)$, and therefore,

$$\gamma[\mathscr{S}(\mathscr{G}_1,\theta)] \geqslant \gamma(\mathscr{S}^*) := \inf_{\psi}\left[\mathbb{P}_{0,n}(\psi=1) + \max_{\Theta\in\mathscr{S}^*}\mathbb{P}_{\Theta,n}(\psi=0)\right]. \tag{2.3}$$

By (2.3), it follows that to give a lower bound on $\gamma[\mathscr{S}(\mathscr{G}_1, \theta)]$, it suffices to lower bound $\gamma(\mathscr{S}^*)$. We are ready to introduce our main theorem.

THEOREM 2.2 Let $\mathscr{G}^*$ be a non-empty subset of $\mathscr{G}_1$ such that all graphs in $\mathscr{G}^*$ have arboricity $\mathscr{R}$. Define $N(\mathscr{G}^*) := \max_{G \in \mathscr{G}^*} \mathbb{E}_{G' \sim U(\mathscr{G}^*)} |V(G) \cap V(G')|$, where $U(\mathscr{G}^*)$ is the uniform distribution over $\mathscr{G}^*$. If $\mathscr{G}^*$ is incoherent, $N(\mathscr{G}^*) = o(1)$ and

$$\theta \leqslant \sqrt{\frac{\log[N^{-1}(\mathscr{G}^*)]}{6n\mathscr{R}}} \wedge \sqrt{\frac{\mathscr{R}}{\mathscr{B}}} \wedge \frac{1}{8(\Lambda \vee \Gamma)}, \tag{2.4}$$

then we have

$$\liminf_{n \to \infty} \gamma(\mathscr{S}^*) = 1.$$

The following is a proof sketch of Theorem 2.2. The full proof is given in Section 5.

Proof Sketch. The proof of Theorem 2.2 follows the following three steps.

Step 1. Lower bounding the minimax risk using Le Cam's method. We put uniform prior on the graphs in the alternative hypothesis and use Le Cam's method of fuzzy hypotheses to derive a lower bound of $\gamma(\mathscr{S}^*)$. The resulting lower bound of the minimax risk is inversely related to the χ^2-divergence between the null distribution and the uniform mixture of the distributions characterized by $\mathscr{S}^*$. This converts the minimax lower bound problem into the problem of upper bounding the χ^2-divergence.

Step 2. Connecting the minimax risk lower bound to a Eulerian subgraphs counting problem. We translate the χ^2-divergence obtained in Step 1 to a combinatorial quantity to ease the calculation. Specifically, based on the *high-temperature expansion* [**26**] of Ising model densities, we show that the χ^2-divergence can be bounded by a polynomial of $t = \tanh(\theta)$, whose coefficients are related to the number of certain Eulerian graphs defined by $\mathscr{G}^*$.

Step 3. Finalizing the proof using the incoherence condition of $\mathscr{G}^*$. In this step, we further bound the polynomial derived in Step 2 by neglecting the higher degree terms and increasing the coefficients in front of the lower degree terms (namely only the terms of degree 2, 3 and 4 survive). In addition, using that $\theta \leqslant \sqrt{\mathscr{R}/\mathscr{B}} \wedge [8(\Lambda \vee \Gamma)]^{-1}$ we can further fold the third and fourth degree terms into the second degree term. It turns out that the coefficient in front of the second degree term is bounded by $\mathscr{R}|V(G) \cap V(G')|$, where the quantity $|V(G) \cap V(G')|$ is the number of overlapping vertices between graphs G, G' in $\mathscr{G}^*$. We then invoke the incoherence property of $\mathscr{G}^*$ and show that under the conditions of Theorem 2.2, $\liminf_{n \to \infty} \gamma(\mathscr{S}^*) = 1$. □

REMARK 2.1 The goal of this remark is to explain the intuition behind the first term on the right-hand side of (2.4). It is intuitive that some sort of quantity measuring the density of the graph appears in our bound, since the denser a graph is, the easier it is to detect it. Moreover, it is clear that a single test examining the overlapping part of different graphs in $\mathscr{G}^*$ can simultaneously distinguish these overlapped graphs from the null. As an extreme example, consider the case where all graphs in $\mathscr{G}^*$ are perfectly overlapped, then $\mathscr{G}^*$ is essentially a singleton, which is easy to detect. Therefore, the degree of overlaps among graphs in $\mathscr{G}^*$ also affects the difficulty of the test. In our bound in Theorem 2.2, the density of the graph is measured by the arboricity $\mathscr{R}$ and the degree of overlaps is measured by the quantity $N(\mathscr{G}^*)$.

REMARK 2.2 Inequality (2.4) shows that the necessary signal strength of detection problems is determined by the minimum of three terms. While the first term $\sqrt{\frac{\log[N^{-1}(\mathscr{G}^*)]}{6n\mathscr{R}}}$ is related to both the structural properties of graphs in $\mathscr{G}_1$ and the sample size n, the second term $\sqrt{\frac{\mathscr{R}}{\mathscr{B}}}$ and the third term $\frac{1}{8(\Lambda\vee\Gamma)}$ are independent of n. Therefore, when the sample size is large enough, $\sqrt{\frac{\log[N^{-1}(\mathscr{G}^*)]}{6n\mathscr{R}}}$ is the leading term determining the necessary signal strength and the other two terms mainly serve as scaling conditions of θ.

REMARK 2.3 The condition (2.4) given by Theorem 2.2 is comparable to the 'multi-edge' results given in [**49**], where the authors give minimax lower bounds of combinatorial inference problems in Gaussian graphical models. Unlike our results in Theorem 2.2, the necessary signal strength for Gaussian graphical models given by [**49**] does not explicitly involve graph arboricity. It is also not very clear under what condition the lower bound given by [**49**] is sharp. In comparison, in this paper, we show that graph arboricity is an appropriate quantity that gives sharp lower bounds for *any* structure detection problems under the incoherence condition and the sparsity assumption $s = O(d^{1/2-c})$ for some $c > 0$. It is also worth comparing Theorem 2.2 with the results of [**48**]. The lower bounds on the signal θ of [**48**], typically involve the quantity $\sqrt{\frac{\log d}{n}}$ which is generally much larger than the right-hand side of (2.4) when $\mathscr{R}$ is large enough. This is intuitively clear since detection problems are statistically easier than graph property testing. Our proof strategy is also completely different than the one used by [**48**] and relies on high-temperature expansions rather than Dobrushin's comparison theorem.

In Theorem 2.2, the incoherence condition of $\mathscr{G}^*$ is not always easy to check. However, it is known that this condition is satisfied by a various discrete distributions including the multinomial and hypergeometric distributions [**17**,**37**]. In particular, [**37**, Theorem 2.11] states that negative association holds for all permutation distributions. Therefore, for detection problems of the form (1.5), incoherence condition is always satisfied by picking $\mathscr{G}^*$ to be the set of all graphs isomorphic to G_*. This leads to the following corollary (recall that we are assuming $s = o(\sqrt{d})$).

COROLLARY 2.1 Let G_* be a graph with s vertices and $\mathscr{G}_1(G_*)$ be the class of all graphs that contain a size-s subgraph isomorphic to G_*. Let $\mathscr{B}(G_*) = 512\{\|\mathrm{A}_{G_*}\|_F^4 \wedge [(\|\mathrm{A}_{G_*}\|_1 \vee \|\mathrm{A}_{G_*}\|_F)^2 s]\}$. If

$$\theta \leqslant \sqrt{\frac{\log(d/s^2)}{6n\mathscr{R}(G_*)}} \wedge \sqrt{\frac{\mathscr{R}(G_*)}{\mathscr{B}(G_*)}} \wedge \frac{1}{8(\|\mathrm{A}_{G_*}\|_F \vee \|\mathrm{A}_{G_*}\|_1)},$$

then we have

$$\liminf_{n\to\infty} \gamma\{\mathscr{S}[\mathscr{G}_1(G_*), \theta]\} = 1.$$

2.1 *Examples*

In this section, we apply Corollary 2.1 to specific detection problems.

EXAMPLE 2.3 (Empty graph vs. non-empty graph). Consider testing empty graph vs. non-empty graph defined in Section 1. If

$$\theta \leqslant \sqrt{\frac{\log(d/4)}{6n}} \wedge \frac{1}{32\sqrt{2}}, \tag{2.5}$$

we have $\liminf_{n\to\infty} \gamma[\mathscr{S}(\mathscr{G}_1, \theta)] = 1$.

Proof. In this example, $s = 2, \mathscr{G}_1 = \mathscr{G}_1(G_*)$, where G_* is a single-edge graph, and we have $\mathscr{R}(G_*) = 1$. By a direct calculation, we have $\|\mathrm{A}_{G_*}\|_F = \sqrt{2}$, $\|\mathrm{A}_{G_*}\|_1 = 1$ and $\mathscr{B}(G_*) = 2048$. By Corollary 2.1, if (2.5) holds we have $\liminf_{n\to\infty} \gamma(\mathscr{S}^*) = 1$. □

EXAMPLE 2.4 (Clique detection). For the clique detection problem defined in Section 1, if

$$\theta \leqslant \sqrt{\frac{\log(d/s^2)}{6ns}} \wedge \frac{1}{32s}, \tag{2.6}$$

we have $\liminf_{n\to\infty} \gamma[\mathscr{S}(\mathscr{G}_1, \theta)] = 1$.

Proof. In this example, $\mathscr{G}_1 = \mathscr{G}_1(G_*)$ with G_* being an s-clique graph. We have $\mathscr{R}(G_*) = \lceil s/2 \rceil$ and therefore $s/2 \leqslant \mathscr{R}(G_*) \leqslant s$. By direct calculation, we have $\|\mathrm{A}_{G_*}\|_F = \sqrt{s(s-1)} \leq s$, $\|\mathrm{A}_{G_*}\|_1 = s - 1 \leqslant s$, and therefore, $\mathscr{B}(G_*) \leqslant 512s^3$. By Corollary 2.1, if (2.6) holds, we have $\liminf_{n\to\infty} \gamma(\mathscr{S}^*) = 1$. □

EXAMPLE 2.5 (Star detection). For the star detection problem defined in Section 1, if $s \geqslant 4$ and

$$\theta \leqslant \sqrt{\frac{\log(d/s^2)}{6n}} \wedge \frac{1}{32\sqrt{2}s}, \tag{2.7}$$

we have $\liminf_{n\to\infty} \gamma[\mathscr{S}(\mathscr{G}_1, \theta)] = 1$.

Proof. In this example, G_* is a star graph and we have $\mathscr{R}(G_*) = 1$. By direct calculation, we have $\|\mathrm{A}_{G_*}\|_F = \sqrt{2(s-1)} \leq \sqrt{2s}$, $\|\mathrm{A}_{G_*}\|_1 = s - 1 \leqslant s$. If $s \geqslant 4$, we have $\mathscr{B}(G_*) = 2048s^2$. By Corollary 2.1, if (2.7) holds, we have $\liminf_{n\to\infty} \gamma(\mathscr{S}^*) = 1$. □

EXAMPLE 2.6 (Community structure detection). For the community structure detection problem defined in Section 1, if $k \geqslant 4$, $l \geqslant 2$ and

$$\theta \leqslant \sqrt{\frac{\log(d/s^2)}{6n(l \vee k)}} \wedge \frac{1}{32\sqrt{2}s}, \tag{2.8}$$

we have $\liminf_{n\to\infty} \gamma[\mathscr{S}(\mathscr{G}_1, \theta)] = 1$.

Proof. We have $\mathscr{G}_1 = \mathscr{G}_1(G_*)$, where G_* is a community structure graph defined in Section 1. To calculate $\mathscr{R}(G_*)$, we utilize the fact that arboricity equals the minimum number of forests into which the edges of a given graph can be partitioned [**47**]. Let $\mathscr{C}_1, \ldots, \mathscr{C}_l$ be the communities. For $i = 1, \ldots, l$, we know that $\mathscr{C}_i$ is a k-clique and the arboricity is $\lceil k/2 \rceil$. Therefore, inside $\mathscr{C}_i$, we can partition the graph into $\lceil k/2 \rceil$ forests. There is also an l-clique in G_* consisting of the cross-community edges. This clique can be partitioned into $\lceil l/2 \rceil$ forests. Note that this l-clique shares only one vertex $v(\mathscr{C}_i)$ with the community $\mathscr{C}_i$. Therefore, for any forest in the partition of this l-clique and any forest in the partition of $\mathscr{C}_i$, we can merge them into a single forest because the resulting graph is still acyclic. We can keep merging forests from other communities. Eventually, we can merge l forests from distinct communities to a forest in the l-clique, without introducing any cycles. If $\lceil l/2 \rceil \geqslant \lceil k/2 \rceil$, we will obtain $\lceil l/2 \rceil$ forests that form a partition of G_*; if $\lceil l/2 \rceil < \lceil k/2 \rceil$, then the partition will contain $\lceil k/2 \rceil$ forests. Therefore, by

the equivalent definition of arboricity given in [47], we have $\mathscr{R}(G_*) \leqslant \lceil (l \vee k)/2 \rceil$. On the other hand, since G_* contains an l clique, obviously $\mathscr{R}(G_*) \geqslant \lceil l/2 \rceil$. Similarly, $\mathscr{R}(G_*) \geqslant \lceil k/2 \rceil$ and hence we have $\mathscr{R}(G_*) = \lceil (l \vee k)/2 \rceil$. Therefore, $(l \vee k)/2 \leqslant \mathscr{R}(G_*) \leq l \vee k$.

By direct calculation, we have $\|\mathrm{A}_{G_*}\|_F = \sqrt{lk(k-1) + l(l-1)} \leq \sqrt{lk^2 + l^2}$, $\|\mathrm{A}_{G_*}\|_1 = k - 1 + l - 1 \leqslant k + l$. We now compare the upper bounds of $\|\mathrm{A}_{G_*}\|_F$ and $\|\mathrm{A}_{G_*}\|_1$. If $k \geqslant 4$ and $l \geqslant 2$, we have $l \geqslant 1 + l/2$ and

$$lk^2 + l^2 \geqslant (1 + l/2)k^2 + l^2 = k^2 + lk^2/2 + l^2 \geqslant k^2 + 2kl + l^2 = (k + l)^2.$$

Therefore,

$$\mathscr{B}(G_*) \leqslant 512[(lk^2 + l^2)^2 \wedge (lk^2 + l^2)lk] = 512(lk^2 + l^2)lk = 512(s^2k + sl^2)$$

and

$$\sqrt{\frac{\mathscr{R}(G_*)}{\mathscr{B}(G_*)}} \geqslant \sqrt{\frac{l \vee k}{1024(s^2k + sl^2)}} \geqslant \sqrt{\frac{l \vee k}{1024(s^2k + s^2l)}} \geqslant \sqrt{\frac{l \vee k}{2048s^2(k \vee l)}} = \frac{1}{32\sqrt{2}s}.$$

Moreover,

$$\frac{1}{8(\|\mathrm{A}_{G_*}\|_F \vee \|\mathrm{A}_{G_*}\|_1)} \geqslant \frac{1}{8\sqrt{lk^2 + l^2}} \geqslant \frac{1}{8\sqrt{2}s}.$$

Therefore, by Corollary 2.1, if (2.8) holds, we have $\liminf_{n\to\infty} \gamma(\mathscr{S}^*) = 1$. □

3. Upper bounds

In this section, we construct upper bounds for the hypothesis testing problem (1.1). We propose a general framework for testing an empty graph $G_\varnothing$ against an arbitrary graph set $\mathscr{G}_1$. We remind the reader that the arboricity of a graph G is defined in (2.2) as

$$\mathscr{R}(G) := \left\lceil \max_{V \subseteq \overline{V}} \frac{|E(G_V)|}{|V| - 1} \right\rceil,$$

where G_V is the graph obtained by restricting G on the vertex set V. The arboricity $\mathscr{R}$ of $\mathscr{G}_1$ is then defined as

$$\mathscr{R} := \min_{G \in \mathscr{G}_1} \mathscr{R}(G).$$

We now introduce the concept of witnessing subgraph and witnessing set. Before that, we remind the reader that in this paper all graphs have d vertices (i.e., all graphs are over the vertex set $\overline{V}$), unless otherwise specified. Therefore, a subgraph G' of a graph $G = (\overline{V}, E)$ is a graph with d vertices whose edge set is a subset of the edge set of the larger graph, i.e., $G' = (\overline{V}, E')$ where $E' \subseteq E$. Importantly, the notation $V(G)$ and $V(G')$ refer to the non-isolated vertices of G and G' which may be strict subsets of $\overline{V}$.

DEFINITION 3.1 (Witnessing subgraph). For a graph $G \in \mathscr{G}_1$, we call the graph H a witnessing subgraph of G with respect to $\mathscr{G}_1$, if H is a subgraph of G and $\lceil |E(H)|/[|V(H)| - 1] \rceil \geqslant \mathscr{R}$.

Here, we remark that for H to be a witnessing subgraph of G, it is unnecessary to have $\lceil |E(H)|/[|V(H)| - 1] \rceil = \mathscr{R}(G)$. Instead, we only require that $\lceil |E(H)|/[|V(H)| - 1] \rceil \geqslant \mathscr{R}$, which is a weaker requirement since by definition, we have $\mathscr{R} \leqslant \mathscr{R}(G)$ for any $G \in \mathscr{G}_1$. This implies that every graph $G \in \mathscr{G}_1$ has at least one witnessing graph, which may be obtained from the densest subgraph of G (with potential edge pruning).

DEFINITION 3.2 (Witnessing set). We call the collection of graphs $\mathscr{H}$ a witnessing set of $\mathscr{G}_1$, if for every $G \in \mathscr{G}_1$, there exists $H \in \mathscr{H}$ such that H is a witnessing subgraph of G.

By the definition of $\mathscr{R}$, and as we previously argued, every graph $G \in \mathscr{G}_1$ must have at least one witnessing subgraph. Therefore, at least one witnessing set $\mathscr{H}$ of $\mathscr{G}_1$ exists. We define the set of witnessing graphs in order to facilitate the development of *scan tests*. Below, we will formalize a test statistic which scans over all graphs in $\mathscr{H}$. Importantly, in order to match the lower bound result given by Theorem 2.2, it is not sufficient to scan directly over the graphs from the set $\mathscr{G}_1$. This is because the graphs in $\mathscr{G}_1$ may contain non-essential edges which may introduce noise during the testing. In contrast, the graphs from $\mathscr{H}$ trim down those non-essential edges and focus only on the essential parts of the graphs in $\mathscr{G}_1$.

We now introduce our general testing procedure. Our test is based on a witnessing set $\mathscr{H}$. For $H \in \mathscr{H}$, we define

$$\hat{W}_H := \frac{1}{n} \cdot \sum_{l=1}^{n} \left(\frac{1}{|E(H)|} \sum_{(i,j) \in E(H)} X_{l,i} X_{l,j} \right), \tag{3.1}$$

where $\boldsymbol{X}_l$ is the lth sample and $X_{l,i}$, $X_{l,j}$ are the ith and jth components of $\boldsymbol{X}_l$, respectively. Our test then scans over all possible $H \in \mathscr{H}$ and calculates the corresponding $\hat{W}_H$. We define

$$\psi := \mathbb{1}\left[\max_{H \in \mathscr{H}} \hat{W}_H > \frac{\kappa}{4} \sqrt{\frac{M(\mathscr{H})}{\mathscr{R} n}} \right], \tag{3.2}$$

where κ is a large enough absolute constant, and

$$M(\mathscr{H}) := \frac{\log(|\mathscr{H}|)}{m(\mathscr{H})}, \text{with } m(\mathscr{H}) := \min_{H \in \mathscr{H}} |V(H)|.$$

The following theorem justifies the usage of the test defined in (3.2).

THEOREM 3.3 Given any fixed $\alpha \in (0, 1)$, suppose that $\log(|\mathscr{H}|)/n = o(1)$ and $|\mathscr{H}| \geqslant 2/\alpha$. If

$$\theta > \kappa \sqrt{\frac{M(\mathscr{H})}{\mathscr{R} n}}$$

for a large enough absolute constant κ, when n is large enough, we have that the test ψ of (3.2) satisfies

$$\mathbb{P}_{0,n}(\psi = 1) + \max_{\Theta \in \mathscr{S}(\mathscr{G}_1,\theta)} \mathbb{P}_{\Theta,n}(\psi = 0) \leqslant \alpha.$$

We give a proof sketch of Theorem 3.3 as follows. The detailed proof is given in Section 6.

Proof Sketch. The proof of Theorem 3.3 follows the following three steps.

Step 1. Bounding the ψ_1-norm of the test variable. The key observation in the proof is to establish that each of the i.i.d. summands of $\hat{W}_H$ (3.1) is a sub-exponential random variable. To prove this, we use a result of [**11**] and our assumption that $\|\Theta\|_F \leqslant 1/2$.

Step 2. Upper/lower bounding $\hat{W}_H$ under the null/alternative. Using the result of Step 1 and the second Griffith's inequality, we establish an upper bound on $\hat{W}_H$ under the null and a lower bound on $\hat{W}_H$ under any graph from the alternative hypothesis.

Step 3. Proof completion. We show that under the assumptions of Theorem 3.3, there is a sufficient gap between the upper and lower bounds established in Step 2, which renders the final claim of the theorem. □

REMARK 3.1 We can compare our upper bound result with Corollary 2.1. For the testing problems of form (1.5), we can always choose a subgraph H_* of G_* as a witnessing subgraph (if there are multiple such subgraphs pick any of them) and construct $\mathscr{H}$ to be the set consisting of all graphs isomorphic to H_*. For this $\mathscr{H}$, we have $|\mathscr{H}| \leqslant \frac{d!}{(d-|V(H_*)|)!}$. Therefore,

$$M(\mathscr{H}) \leqslant |V(H_*)|^{-1} \log[d!/(d - |V(H_*)|)!] \leqslant \log(d).$$

If $s = O(d^{1/2-c})$ for some $c > 0$, $\log(d/s^2)$ is also of order $\log(d)$. Therefore, the rate given by Theorem 3.3 matches Corollary 2.1. When applying Theorem 3.3 to specific detection problems, potentially there might exist different choices of $\mathscr{H}$, which may result in lower values of $M(\mathscr{H})$.

3.1 *Examples*

In this section, we apply Theorem 3.3 to the examples we discussed in Sections 1 and 2.1.

EXAMPLE 3.4 (Empty graph vs. non-empty graph). Consider testing empty graph vs. non-empty graph defined in Section 1. If $\log(d)/n = o(1)$, $4/[d(d-1)] \leqslant \alpha$ and

$$\theta > \kappa \sqrt{\frac{\log d}{n}} \tag{3.3}$$

for a large enough constant κ, then when n is large enough, we have

$$\mathbb{P}_{0,n}(\psi = 1) + \max_{\Theta \in \mathscr{S}(\mathscr{G}_1,\theta)} \mathbb{P}_{\Theta,n}(\psi = 0) \leqslant \alpha. \tag{3.4}$$

Proof. In this example, $\mathscr{R} = 1$, and hence $\mathscr{H} = \{\text{single-edge graphs}\}$ is a witnessing set of $\mathscr{G}_1$. We have $|\mathscr{H}| = d(d-1)/2$, $m(\mathscr{H}) = 2$ and $M(\mathscr{H}) = \log(|\mathscr{H}|)/m(\mathscr{H}) \leq \log d$. Therefore, by Theorem 3.3, if (3.3) holds for a large enough constant κ, then when n is large enough, we have that (3.4) holds. □

EXAMPLE 3.5 (Clique detection). For the clique detection problem defined in Section 1, if $s\log(ed/s)/n = o(1)$, $(d/s)^s \geqslant 2/\alpha$ and

$$\theta > \kappa\sqrt{\frac{\log(ed/s)}{sn}} \tag{3.5}$$

for a large enough constant κ, then when n is large enough, we have

$$\mathbb{P}_{0,n}(\psi = 1) + \max_{\Theta \in \mathscr{S}(\mathscr{G}_1,\theta)} \mathbb{P}_{\Theta,n}(\psi = 0) \leqslant \alpha. \tag{3.6}$$

Proof. In this example, we have $\mathscr{R} = \lceil s/2 \rceil$ and $\mathscr{H} = \{s\text{-cliques}\}$ is a witnessing set of $\mathscr{G}_1$. We have $\binom{|\mathscr{H}|=}{ds}$, and therefore, $(d/s)^s \leqslant |\mathscr{H}| \leqslant (ed/s)^s$. We have $m(\mathscr{H}) = s$ and $M(\mathscr{H}) = \log(|\mathscr{H}|)/m(\mathscr{H}) \leq \log(ed/s)$. Therefore, by Theorem 3.3, if (3.5) holds for a large enough constant κ, then when n is large enough, we have that (3.6) holds. □

EXAMPLE 3.6 (Star detection). For the star detection problem defined in Section 1, if $\log(d)/n = o(1)$, $4/[d(d-1)] \leq \alpha$ and

$$\theta > \kappa\sqrt{\frac{\log(ed/s)}{n}} \tag{3.7}$$

for a large enough constant κ, then when n is large enough, we have

$$\mathbb{P}_{0,n}(\psi = 1) + \max_{\Theta \in \mathscr{S}(\mathscr{G}_1,\theta)} \mathbb{P}_{\Theta,n}(\psi = 0) \leqslant \alpha. \tag{3.8}$$

Proof. In this example, we have $\mathscr{R} = 1$ and $\mathscr{H} = \{(s-1)\text{-stars}\}$ is a witnessing set of $\mathscr{G}_1$. We have $\binom{|\mathscr{H}|=s}{ds}$, and therefore, $s(d/s)^s \leqslant |\mathscr{H}| \leqslant s(ed/s)^s$. We have $m(\mathscr{H}) = s$. When $s = o(\sqrt{d})$, we have $s \leqslant (ed/s)^s$ and $M(\mathscr{H}) = \log(|\mathscr{H}|)/m(\mathscr{H}) \leq 2\log(ed/s)$. Therefore, by Theorem 3.3, if (3.7) holds for a large enough constant κ, then when n is large enough, we have that (3.8) holds. □

EXAMPLE 3.7 (Community structure detection). Consider the community structure detection problem defined in Section 1. If $(l \vee k)\log[ed/(l \vee k)]/n = o(1)$, $[d/(l \vee k)]^{(l \vee k)} \geqslant 2/\alpha$ and

$$\theta > \kappa\sqrt{\frac{\log[ed/(l \vee k)]}{(l \vee k)n}}$$

for a large enough constant κ, then when n is large enough, we have

$$\mathbb{P}_{0,n}(\psi = 1) + \max_{\Theta \in \mathscr{S}(\mathscr{G}_1,\theta)} \mathbb{P}_{\Theta,n}(\psi = 0) \leqslant \alpha.$$

Proof. If $l \geqslant k$, we have $\mathscr{R} = \lceil l/2 \rceil$, and we can choose $\mathscr{H} = \{l\text{-cliques}\}$ as a witnessing set of $\mathscr{G}_1$; if $l < k$, we have $\mathscr{R} = \lceil k/2 \rceil$ and $\mathscr{H} = \{k\text{-cliques}\}$ is a witnessing set of $\mathscr{G}_1$. The rest of the proof is identical to the clique detection problem, and we omit the details. □

4. Computational lower bound

In this section, we give our main result on the computational lower bound of the structure testing problems in Ising models. Our result is based on a sparse PCA conjecture. Denote by $1_{i_1,\dots,i_s} = e_{i_1} + \cdots + e_{i_s} \in \mathbb{R}^d$ the vector whose $i_1,\dots,i_s$th entries are 1 and other entries are 0. Let

$$\mathscr{S}_\sigma = \left\{ \mathrm{I} + \sigma 1_{i_1,\dots,i_s} 1_{i_1,\dots,i_s}^T \in \mathbb{R}^{d\times d} : i_1,\dots,i_s \in \{1,\dots,d\} \text{ are distinct} \right\}$$

be the set of covariance matrices from the Gaussian spiked model. In sparse PCA, we consider the hypothesis testing problem for n i.i.d samples $\boldsymbol{Z}_1,\dots,\boldsymbol{Z}_n \in \mathbb{R}^d$:

$$\mathrm{H}_0^{\mathrm{PCA}} : \boldsymbol{Z}_1,\dots,\boldsymbol{Z}_n \sim N(0,\mathrm{I}) \text{vs.} \mathrm{H}_1^{\mathrm{PCA}} : \boldsymbol{Z}_1,\dots,\boldsymbol{Z}_n \sim N(0,\Sigma), \Sigma \in \mathscr{S}_\sigma. \tag{4.1}$$

We denote by $\mathbb{P}_{\mathrm{I},n}$ and $\mathbb{P}_{\Sigma,n}$ the probability measure under $\mathrm{H}_0^{\mathrm{PCA}}$ and $\mathrm{H}_1^{\mathrm{PCA}}$, respectively.

Conjecture 4.1 (Computational hardness of sparse PCA). Let $\delta > 0$ be any fixed small constant. If $\sigma \leqslant \eta[n^{-(1/2+\delta)} \wedge s^{-(1+\delta)}]$ for some small enough constant η, then for any polynomial time test ψ, we have

$$\liminf_{n\to\infty} \left[\mathbb{P}_{\mathrm{I},n}(\psi = 1) + \max_{\Sigma\in\mathscr{S}_\sigma} \mathbb{P}_{\Sigma,n}(\psi = 0) \right] \geqslant \frac{1}{4}.$$

Conjecture 4.1 is derived by [27] under the widely believed planted clique conjecture and additional assumptions which essentially require that $2n \leqslant d \leqslant n^a$ for some constant $a > 1$ and $n[\log(n)]^5 \leqslant Cs^4$ for some small enough constant $C > 0$. It is also studied in [7], [13] and [12]. In particular, the latter two papers prove Conjecture 4.1 for the two regimes $s \gg \sqrt{d}$ [13] and $s < \sqrt{d}$ [12], respectively, based on the planted clique conjecture.

In the following, based on the Gaussian random vectors from the sparse PCA problem, we propose a polynomial time reduction algorithm that constructs n d-dimensional random vectors which cannot be distinguished from n samples from the d-dimensional Ising model with a parameter matrix Θ. Importantly, this reduction only needs to be done for clique graphs because detecting the s-clique containing G_* is always easier than directly detecting G_* (recall that we are testing whether the underlying graph is empty or contains a graph isomorphic to G_*). Furthermore, in the sparse PCA problem each $\Sigma \in \mathscr{S}_\sigma$ corresponds to an s-clique. More specifically, for any index set $\mathscr{I} \subseteq \{1,\dots,d\}$ of size s representing the position of a clique, we consider the probability measure $\mathbb{P}_{\Theta,n}$ of the Ising model with parameter matrix $\Theta = \theta \cdot [1(i,j \in \mathscr{I}, i \neq j)]_{d\times d}$ (when $\theta = 0$ we denote $\mathbb{P}_{0,n}$ the Ising clique model under the null hypothesis) and $\mathbb{P}_{\Sigma,n}$ for the multivariate Gaussian distribution with mean 0 and covariance matrix $\Sigma = \mathrm{I} + \sigma 1_{\mathscr{I}} 1_{\mathscr{I}}^T$.

We are now ready to state the main result of this section. We have the following theorem.

Theorem 4.2 (Computational hardness of Ising clique detection). Suppose that Conjecture 4.1 holds. If $\theta \leqslant \eta[n^{-(1/2+\delta)} \wedge s^{-(1+\delta)}]$ for some small enough constant η, then for any polynomial time test ψ and an s-clique graph G_*, we have

$$\liminf_{n\to\infty} \left[\mathbb{P}_{0,n}(\psi = 1) + \max_{\Theta\in\mathscr{S}(\mathscr{G}_1(G_*),\theta)} \mathbb{P}_{\Theta,n}(\psi = 0) \right] \geqslant \frac{1}{4}.$$

In order to prove the computational hardness result above, we need to formalize a polynomial time transformation which (approximately in total variation) maps the null hypothesis of a sparse PCA model to the null hypothesis of the Ising clique model and simultaneously maps the corresponding alternative hypothesis of sparse PCA to the alternative hypothesis of the Ising clique model up to a small total variation.

Our construction is surprisingly simple: taking the signs of Gaussians and showing that they are close enough to an Ising model in total variation. Clearly, for this type of reduction, the null hypothesis of a sparse PCA model is mapped to exactly the null hypothesis of Ising clique model. Similarly, under the alternative hypothesis of the sparse PCA model, the vertices out of the clique are still mapped to Rademacher random variables, which also correspond to the vertices out of the Ising clique. Therefore, we only need to focus on the vertices in the Gaussian and Ising cliques under the alternative hypotheses. We introduce the following notation for the random Gaussian/Ising vectors corresponding to the vertices in the clique. Let $\boldsymbol{W}_1, \dots \boldsymbol{W}_n$ and $\boldsymbol{V}_1, \dots, \boldsymbol{V}_n$ be s-dimensional i.i.d. Gaussian and Ising random vectors with parameter matrices $\mathrm{I}_{s\times s} + \sigma 1_{[s]} 1_{[s]}^T$ and $\theta(1_{[s]} 1_{[s]}^T - \mathrm{I}_{s\times s})$, respectively. Denote $\boldsymbol{U}_i = \operatorname{sign}(\boldsymbol{W}_i)$, $i \in [n]$.

Before we introduce the proof details, it is necessary to determine the parameter σ for any fixed θ. Part of our choice of σ is based on matching the first terms in the Taylor expansions of the functions $x \mapsto \Phi(x)$ (here, $\Phi(x)$ is the standard normal cumulative density function) and $x \mapsto \exp(cx)/[\exp(cx) + \exp(-cx)]$ for some constant c (turns out that the 'best' $c = \sqrt{2/\pi}$). Supposing that $s\theta < \frac{1}{2}$, we set $\sigma = \pi\theta/(1 - 2s\theta)$.

Recall now that if one has a model $\boldsymbol{W} \sim N(0, \mathrm{I}_{s\times s} + \sigma 1_{[s]} 1_{[s]}^T)$, this is equivalent to having generated i.i.d. $Y \sim N(0,1)$ and $Y_i \sim N(0,1)$ for $i \in [s]$, and set $W_i = Y_i + \sqrt{\sigma} Y$. This means that in terms of signs of $\boldsymbol{W}$, i.e., $\operatorname{sign}(W_i)$, the generation is as follows: first, generate $Y \sim N(0,1)$ and then generate 1 or -1 conditionally i.i.d. on Y with probability $\Phi(\sqrt{\sigma} Y)$ or $1 - \Phi(\sqrt{\sigma} Y)$, respectively. Our next lemma, which is the key to the proof of bounding the total variation argues that similar conditional i.i.d. decomposition holds for the Curie–Weiss model. Recall that the Curie–Weiss model is given by

$$\mathbb{P}_\theta(\boldsymbol{V} = \boldsymbol{v}) \propto \exp\left(\theta \sum_{i \neq j} v_i v_j\right) \propto \exp\left(\theta \left(\sum_i v_i\right)^2\right),$$

for $\boldsymbol{v} \in \{\pm 1\}^s$. Clearly the Curie–Weiss model corresponds to the 'clique' part of the Ising clique model with parameter matrix Θ.

LEMMA 4.1 (Generating Curie–Weiss as conditional i.i.d.). The Curie–Weiss model with s vertices and parameter $\theta \geqslant 0$ can be generated as (1) generate Y' coming from a distribution with density proportional to $p_{Y'}(y) \propto \cosh(\sqrt{2\theta} y)^s \exp(-y^2/2)$ and (2) generate $V_i | Y'$ i.i.d. with probabilities $\mathbb{P}(V_i = 1 | Y') = \frac{\exp(\sqrt{2\theta} Y')}{\exp(\sqrt{2\theta} Y') + \exp(-\sqrt{2\theta} Y')}$. In addition, the normalizing constant $Z_{Y'}$ of $p_{Y'}(y)$ satisfies that

$$\frac{Z_{Y'}}{\sqrt{2\pi}} = \sum_{k=0}^{s} \frac{\binom{s}{k} \exp(\theta(2k - s)^2)}{2^s}. \tag{4.2}$$

The proof of Lemma 4.1 is deferred to the supplementary material. Call the distribution with density $p_{Y'}(y) \propto \cosh(\sqrt{2\theta} y)^s \exp(-y^2/2)$ Curie–Weiss normal (CWN) with parameters s and θ. Next, we

will argue that n samples from the CWN distribution are close in total variation to n samples from the Gaussian distribution with variance $(1-2s\theta)^{-1}$ provided that θ is small.

THEOREM 4.3 (CWN is close to Gaussian). Let Y' be a CWN random variable with parameters s and θ, and let $Y \sim N(0,(1-2s\theta)^{-1})$ be a Gaussian random variable. Suppose we have n i.i.d. copies of each of the two random variables. We have that

$$\mathrm{TV}(\mathscr{L}(\{Y_i\}_{i\in[n]}),\mathscr{L}(\{Y_i'\}_{i\in[n]})) \leqslant C\sqrt{n\theta^2\cdot\sum_{i=1}^{5}\frac{(s\theta)^i}{(1-2s\theta)^{i-1/2}}},$$

where C is an absolute constant, provided that $s\theta < c < \frac{1}{2}$ for some sufficiently small constant c.

The proof of Theorem 4.3 is given in Section 7. In order the prove this result, we develop novel bounds on the even moments of sums of i.i.d. Rademacher random variables. We are now ready to give the intuition of the proof of the main theorem of this section. Since both the normal clique and the Curie–Weiss model can be generated as conditional i.i.d., where the corresponding conditional variables have close (in total variation) distributions, and the functions $\Phi(x)$ and $\frac{\exp(\sqrt{2/\pi}x)}{\exp(\sqrt{2/\pi}x)+\exp(-\sqrt{2/\pi}x)}$ are close to each other it ought to follow that the total variation between the signs of the normal clique and the Curie–Weiss models are close. The proof of Theorem 4.2, which is presented in Section 7, makes this intuition precise by calculating a bound on the total variation.

5. Proof of Theorem 2.2

In this section, we give the proof of Theorem 2.2. Note that by the definition of $\mathscr{S}^*$, we only need to consider the simple zero-field ferromagnetic Ising model where all non-zero entries in Θ are the same. Let $G=(\overline{V},E)$ be the underlying graph and $\theta=\theta_{ij}$, $(i,j)\in E$ be the parameter. Let $t=\tanh(\theta)$ and $\mathbb{E}_0$ denote the expectation under the probability measure that $X_1,\dots,X_d$ are i.i.d. Rademacher variables. The following lemma gives an equivalent form of the probability mass function in simple zero-field ferromagnetic Ising models.

LEMMA 5.1 For a simple zero-field Ising model with underlying graph $G=(\overline{V},E)$ and parameter θ, we have

$$\mathbb{P}_{\Theta}(\boldsymbol{X}) = \frac{\prod_{(i,j)\in E}(1+tX_iX_j)}{2^d\mathbb{E}_0\left[\prod_{(i,j)\in E}(1+tX_iX_j)\right]}, \tag{5.1}$$

where $t=\tanh(\theta)$.

We now apply Le Cam's method. Let $\mathbb{P}_{\Theta,n}$ be the joint probability mass function of n i.i.d. samples of Ising model with parameter Θ, $\mathbb{E}_{\Theta,n}$ denote the expectation under $\mathbb{P}_{\Theta,n}$ and $\overline{\mathbb{P}}=\frac{1}{|\mathscr{S}^*|}\sum_{\Theta\in\mathscr{S}^*}\mathbb{P}_{\Theta,n}$ be the averaged probability measure among $\Theta\in\mathscr{S}^*$ under the alternative. Then, the result of Le Cam's method is given in the following lemma.

LEMMA 5.2 For the risk $\gamma(\mathscr{S}^*)$ defined in (2.3), we have $\gamma(\mathscr{S}^*) \geqslant 1 - \frac{1}{2}\sqrt{D_{\chi^2}(\overline{\mathbb{P}}, \mathbb{P}_{0,n})}$, where $D_{\chi^2}(\overline{\mathbb{P}}, \mathbb{P}_{0,n})$ is the χ^2-divergence between $\overline{\mathbb{P}}$ and $\mathbb{P}_{0,n}$ defined as

$$D_{\chi^2}(\overline{\mathbb{P}}, \mathbb{P}_{0,n}) := \frac{1}{|\mathscr{S}^*|^2} \sum_{\Theta, \Theta' \in \mathscr{S}^*} \mathbb{E}_{0,n}\left[\frac{\mathbb{P}_{\Theta,n}}{\mathbb{P}_{0,n}} \frac{\mathbb{P}_{\Theta',n}}{\mathbb{P}_{0,n}}\right] - 1. \tag{5.2}$$

Lemma 5.2 is a direct result of the Le Cam's method. By Lemma 5.2, $\liminf_{n\to\infty} \gamma(\mathscr{S}^*) = 1$ is implied by $\limsup_{n\to\infty} D_{\chi^2}(\overline{\mathbb{P}}, \mathbb{P}_{0,n}) = 0$. To prove this, we use a method similar to the high-temperature expansion of Ising model [**26**,**33**]. By Lemma 5.1 and the fact that the n samples are independent, for $\Theta, \Theta' \in \mathscr{S}^*$ with corresponding graphs G, G', we can rewrite the term $\mathbb{E}_{0,n}\left[\mathbb{P}_{\Theta,n}\mathbb{P}_{\Theta',n}/\mathbb{P}_{0,n}^2\right]$ as follows:

$$\mathbb{E}_{0,n}\left[\frac{\mathbb{P}_{\Theta,n}}{\mathbb{P}_{0,n}} \frac{\mathbb{P}_{\Theta',n}}{\mathbb{P}_{0,n}}\right] = \frac{\mathbb{E}_0^n\left[\prod_{(i,j)\in E(G)}(1 + tX_iX_j) \cdot \prod_{(i,j)\in E(G')}(1 + tX_iX_j)\right]}{\mathbb{E}_0^n\left[\prod_{(i,j)\in E(G)}(1 + tX_iX_j)\right] \cdot \mathbb{E}_0^n\left[\prod_{(i,j)\in E(G')}(1 + tX_iX_j)\right]}, \tag{5.3}$$

where $t = \tanh(\theta)$. Each expectation on the right-hand side above is a polynomial of t. For any $G, G' \in \mathscr{G}^*$, we define

$$f_G(t) = \mathbb{E}_0\left[\prod_{(i,j)\in E(G)}(1 + tX_iX_j)\right],$$

$$f_{G,G'}(t) = \mathbb{E}_0\left[\prod_{(i,j)\in E(G)}(1 + tX_iX_j) \prod_{(i,j)\in E(G')}(1 + tX_iX_j)\right].$$

Plugging the definitions above into (5.3), we obtain

$$\mathbb{E}_{0,n}\left[\frac{\mathbb{P}_{\Theta,n}}{\mathbb{P}_{0,n}} \frac{\mathbb{P}_{\Theta',n}}{\mathbb{P}_{0,n}}\right] = \left[1 + \frac{f_{G,G'}(t) - f_G(t)f_{G'}(t)}{f_G(t)f_{G'}(t)}\right]^n. \tag{5.4}$$

We now analyze the coefficients of each polynomial in (5.4). Let

$$f_G(t) = \sum_{k=0}^{\infty} a_k t^k,\ f_{G'}(t) = \sum_{k=0}^{\infty} b_k t^k,\ f_{G,G'}(t) = \sum_{k=0}^{\infty} c_k t^k.$$

We also define

$$f_{G,G'}(t) - f_G(t)f_{G'}(t) = \sum_{k=0}^{\infty}\left(c_k - \sum_{k_1+k_2=k} a_{k_1}b_{k_2}\right)t^k = \sum_{k=0}^{\infty} u_k t^k.$$

For $f_G(t)$, note that after expanding $\prod_{(i,j)\in E(G)}(1+tX_iX_j)$, the terms with non-zero expectations must have the form $t^k X_{i_1}^2 \cdots X_{i_k}^2$, where $i_1, \ldots, i_k \in \overline{V}$. Therefore, by Lemma 5.3, the coefficient of t^k is equal to the number of k-edge subgraphs of G where every vertex has an even degree. Similar arguments also applies to $f_{G'}(t)$ and $f_{G,G'}(t)$. This observation motivates us to introduce the definitions of multigraphs and Eulerian graphs.

DEFINITION 5.1 (Multigraph). A multigraph is a graph which is permitted to have multiple edges connecting two vertices. We denote $G = (V, E)$, where V is the vertex set and E is the edge multiset.

For a multigraph G with d vertices, we define its adjacency matrix to be $\mathrm{A} = (\mathrm{A}_{ij})_{d\times d}$, where $\mathrm{A}_{ij} = \mathrm{A}_{ji} =$'the number of edges connecting vertices i and j'. A symmetric matrix $\mathrm{A} \in \mathbb{R}^{d\times d}$ with non-negative integer off-diagonal entries and zero diagonal entries naturally represents a multigraph with vertex set $\overline{V}$. Given two multigraphs G and G' with adjacency matrices A and A′, we define $G \oplus G'$ to be the multigraph defined by $\mathrm{A} + \mathrm{A}'$.

DEFINITION 5.2 (Eulerian graph). A Eulerian circuit on a multigraph is a closed walk that uses each edge exactly once. We say that a multigraph is Eulerian if every connected component has a Eulerian circuit.

Note that in graph theory, the term Eulerian graph has different meanings. Sometimes, Eulerian graph is referred to as a graph that has a Eulerian circuit. This is different from our definition because in this paper, we do not require a Eulerian graph to be connected. The following famous lemma on the Eulerian graph is first given by [19] and then completely proved by [34].

LEMMA 5.3 A graph is Eulerian if and only if all vertices in the graph have an even degree.

Based on our previous discussion, Lemma 5.3 relates a_k, b_k and c_k to the number of k-edge Eulerian graphs. Define

$$\mathscr{E}(k, G) := \left\{\tilde{G} = (\overline{V}, \tilde{E}) : \tilde{E} \subseteq E,\ |\tilde{E}| = k,\ \tilde{G} \text{ is a Eulerian graph}\right\}.$$

In words $\mathscr{E}(k, G)$ is the set of k-edge Eulerian subgraphs of G. By Lemma 5.3 and our previous discussion, we have $a_k = |\mathscr{E}(k, G)|$, $b_k = |\mathscr{E}(k, G')|$ and $c_k = |\mathscr{E}(k, G \oplus G')|$, and therefore,

$$f_G(t) = \sum_{k\geqslant 0} |\mathscr{E}(k, G)| t^k,\ f_{G'}(t) = \sum_{k\geqslant 0} |\mathscr{E}(k, G')| t^k, \tag{5.5}$$
$$\text{and } f_{G,G'}(t) = \sum_{k\geqslant 0} |\mathscr{E}(k, G \oplus G')| t^k.$$

Figure 3 gives an example of how to calculate $|\mathscr{E}(k, G)|$ for a given multigraph G.

We now proceed to analyze u_k. Apparently, $u_0 = u_1 = 0$. For $k \geqslant 2$, by the definition of u_k, we can see that, if a k-edge Eulerian subgraph of $G \oplus G'$ can be split into two graphs G_1 and G_2 such that G_1 and G_2 are Eulerian subgraphs of G and G', respectively, then it is also counted in the sum $\sum_{k_1+k_2=k} a_{k_1} b_{k_2}$ and therefore is not counted in u_k. Figure 4 gives examples of Eulerian subgraphs that are counted and not counted in u_k.

Using this type of argument, the following two lemmas together calculate and bound u_k for $k \geqslant 2$.

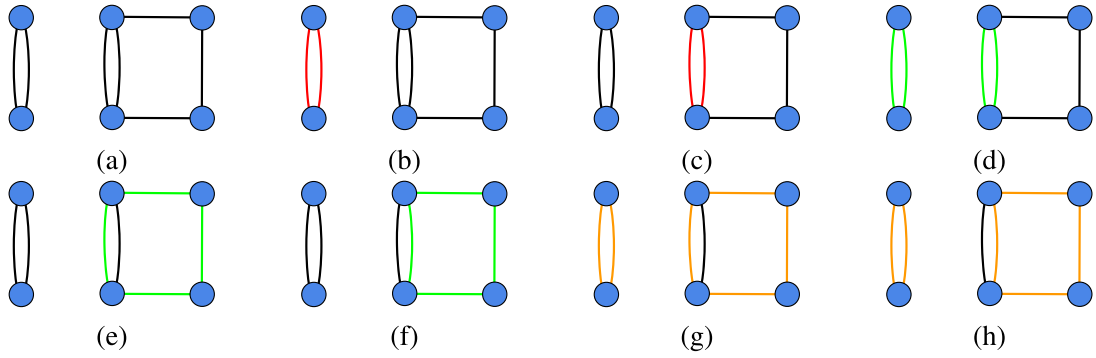

FIG. 3. An example of the calculation of $\{|\mathscr{E}(k,G)|\}_{k\geqslant 1}$ for a multigraph G is given in (a). We use red, green and orange edges to highlight 2-edge, 4-edge and 6-edge Eulerian subgraphs of G, respectively. (b), (c) give the 2-edge Eulerian subgraphs; (d), (e), (f) give the 4-edge Eulerian subgraphs; and (g), (h) give the 6-edge Eulerian subgraphs. We have $|\mathscr{E}(2,G)| = 2$, $|\mathscr{E}(4,G)| = 3$, $|\mathscr{E}(6,G)| = 2$ and $|\mathscr{E}(k,G)| = 0$ for $k \neq 2,4,6$.

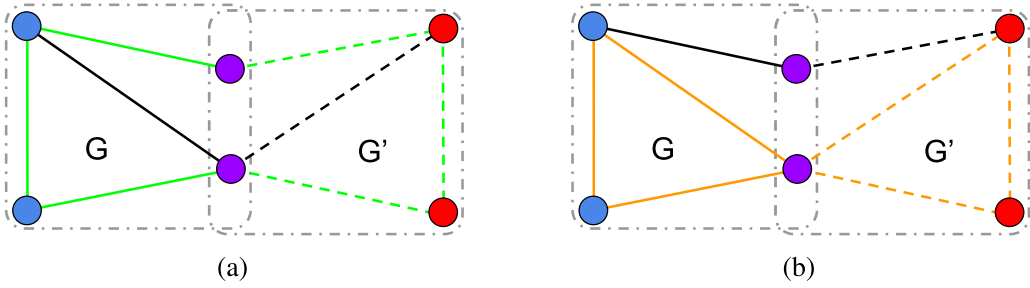

FIG. 4. Illustration of graphs counted and not counted in u_k. The gray dot-dashed squares highlight the non-isolated vertices of G and G'. The solid and dashed lines are edges in G and G', respectively. We use purple vertices to represent the common non-isolated vertices of G and G'. The blue vertices are non-isolated in G but isolated in G', and the red vertices are non-isolated in G' but isolated in G. The green edges in (a) give an example of a 6-edge Eulerian subgraph of $G \oplus G'$ counted in u_6, while the orange edges in (b) form a 6-edge Eulerian subgraph of $G \oplus G'$ that is not counted in u_6.

LEMMA 5.4 We have

$$u_2 = |E(G) \cap E(G')|,\ u_3 = \Delta_{G,G'},\ \text{and}$$
$$u_k \leqslant q_k[G \oplus G', V(G) \cap V(G')] \text{ for } k \geqslant 4,$$

where $\Delta_{G,G'}$ denotes the number of triangles that contains at least one edge from G and one edge from G', and the function $q_k(\cdot,\cdot)$ is defined as follows:

$$q_k(G,V) := \left|\left\{\tilde{G} \in \mathscr{E}(k,G) : \exists i,j \in V, i,j \text{ are in one connected component of } \tilde{G}\right\}\right|.$$

LEMMA 5.5 We have

$$|\mathscr{E}(k,G)| \leqslant 2^k \|\mathrm{A}_G\|_F^k,\ \Delta_{G,G'} \leqslant 2|V(G) \cap V(G')| \cdot \mathscr{R} \cdot \Gamma.$$

Moreover, for any multigraph G and vertex set $V \subseteq \overline{V}$, we have

$$q_k(G, V) \leqslant \left(2^k \cdot |V| \cdot \|\mathrm{A}_G\|_F^k\right) \wedge \left[k \cdot 2^{k-2} \cdot |V|^2 \cdot (\|\mathrm{A}_G\|_1 \vee \|\mathrm{A}_G\|_F)^{k-2}\right].$$

The upper bound of $|\mathscr{E}(k, G)|$ in Lemma 5.5 and the assumption that $\theta \leqslant [8(\Lambda \vee \Gamma)]^{-1}$ together show that $f_G(t), f_{G'}(t)$ and $f_{G,G'}(t)$, as power series, all converge. Moreover, by the definition of $\mathscr{B}$, the upper bound for $q_k(\cdot, \cdot)$ in Lemma 5.5 and the assumption that $\theta \leqslant [8(\Lambda \vee \Gamma)]^{-1}$, we have

$$\sum_{k \geqslant 4} q_k[G \oplus G', V(G) \cap V(G')]\theta^k \leqslant |V(G) \cap V(G')| \cdot \mathscr{B}\theta^4.$$

By Lemmas 5.4 and 5.5 and the fact that $t = \tanh(\theta) \leqslant \theta$, we have

$$\begin{aligned} f_{G,G'}(t) - f_G(t) f_{G'}(t) &\leqslant |E(G) \cap E(G)|\theta^2 + \Delta_{G,G'}\theta^3 + |V(G) \cap V(G')|\mathscr{B}\theta^4 \\ &\leqslant |V(G) \cap V(G')| \cdot (\mathscr{R} + 2\mathscr{R} \cdot \Gamma\theta + \mathscr{B}\theta^2) \cdot \theta^2. \end{aligned} \tag{5.6}$$

Note that $f_G(t) f_{G'}(t) \geqslant 1$ for $t \geqslant 0$ since all coefficients a_k and b_k are non-negative. By (5.4), (5.6) and the assumption that $\theta \leqslant [8(\Lambda \vee \Gamma)]^{-1} \leqslant (2\Gamma)^{-1}$ and $\theta \leqslant \sqrt{\mathscr{R}/\mathscr{B}}$, we have

$$\mathbb{E}_0\left[\frac{\mathbb{P}_\Theta}{\mathbb{P}_0}\frac{\mathbb{P}_{\Theta'}}{\mathbb{P}_0}\right] \leqslant 1 + 3|V(G) \cap V(G')|\mathscr{R}\theta^2.$$

Plugging the inequality above into the definition of χ^2-divergence (5.2) gives

$$D_{\chi^2}(\overline{\mathbb{P}}, \mathbb{P}_{0,n}) \leqslant \frac{1}{|\mathscr{S}^*|^2} \sum_{\Theta, \Theta' \in \mathscr{S}^*} \left[1 + 3|V(G) \cap V(G')| \cdot \mathscr{R}\theta^2\right]^n - 1. \tag{5.7}$$

To complete the proof, we invoke the incoherence condition of $\mathscr{G}^*$. We summarize the result as the following lemma.

LEMMA 5.6 If $\mathscr{G}^*$ is incoherent, then the following inequality holds:

$$\frac{1}{|\mathscr{S}^*|^2} \sum_{\Theta, \Theta' \in \mathscr{S}^*} \exp[3n\mathscr{R}|V(G) \cap V(G')|\theta^2] \leqslant \exp[N(\mathscr{G}^*) \cdot \exp(3n\mathscr{R}\theta^2)].$$

Now, by (5.7), Lemmas 5.2 and 5.6, if $\theta \leqslant \sqrt{\frac{\log[N^{-1}(\mathscr{G}^*)]^{-1}\}}{6nR}}$ and $N(\mathscr{G}^*) = o(1)$, we have

$$\liminf_{n \to \infty} \gamma(\mathscr{S}^*) = 1.$$

6. Proof of Theorem 3.3

In this section, we give the proof of Theorem 3.3. The key part of our proof is to derive concentration inequalities for W_H. Following the definition in [**56**], we define the ψ_1-norm of the random variable Z

as follows.

$$\|Z\|_{\psi_1} := \sup_{p \geqslant 1} p^{-1} (\mathbb{E}|Z|^p)^{1/p}.$$

If a random variable Z has finite ψ_1-norm, we say Z is a sub-exponential random variable. The following lemma gives bounds for the ψ_1-norm of W_H.

LEMMA 6.1 Let $\boldsymbol{X} \in \{\pm 1\}^d$ be a random vector generated from the high-temperature ferromagnetic Ising model with parameter matrix Θ. For any graph H, define

$$W_H := \frac{1}{|E(H)|} \sum_{(i,j) \in E(H)} X_i X_j.$$

If $\|\Theta\|_F \leqslant 1/2$, then we have $\|W_H\|_{\psi_1} \leqslant C|E(H)|^{-1/2}$, where $C > 0$ is an absolute constant.

We first prove that $\mathbb{P}_{0,n}(\psi = 1) < \alpha/2$. Under the null, $X_1, \ldots, X_d$ are independent Rademacher random variables. Therefore, for every $H \in \mathscr{H}$ we have $\mathbb{E}_{0,n} \hat{W}_H = 0$. By Lemma 6.1, with $\Theta = (0)_{d \times d}$, we have $\|W_H\|_{\psi_1} \leqslant C_1 |E(H)|^{-1/2}$ for an absolute constant $C_1 > 0$. By [**56**, Proposition 5.16], for $\varepsilon \leqslant |E(H)|^{-1/2}$, we have

$$\mathbb{P}_{0,n}\left(\left|\hat{W}_H - \mathbb{E}_{0,n}\hat{W}_H\right| > \varepsilon\right) \leqslant 2\exp(-C_2 \cdot |E(H)| \cdot n\varepsilon^2),$$

where C_2 is an absolute constant. Setting the right-hand side above to be $\alpha/(2|\mathscr{H}|)$ and solving for ε shows that under the null hypothesis, with probability at least $1 - \alpha/(2|\mathscr{H}|)$, we have

$$\hat{W}_H \leqslant C_3 \sqrt{\frac{\log|\mathscr{H}| + \log(2/\alpha)}{|E(H)|n}} \leqslant C_3 \sqrt{\frac{2\log|\mathscr{H}|}{|E(H)|n}},$$

for absolute constant C_3. Note that the condition $\varepsilon \leqslant |E(H)|^{-1/2}$ is satisfied since we assume that $\log(|\mathscr{H}|)/n = o(1)$. By definition, we have $m(\mathscr{H}) \leqslant |V(H)|$. Moreover, we have

$$\mathscr{R} \leqslant \frac{|E(H)|}{|V(H)| - 1} + 1 \leqslant \frac{2|E(H)|}{|V(H)|} + 1 \leqslant \frac{4|E(H)|}{|V(H)|}, \tag{6.1}$$

where the last inequality follows by $2|E(H)| \geqslant |V(H)|$. Therefore, with probability at least $1 - \alpha/(2|\mathscr{H}|)$,

$$\hat{W}_H \leqslant C_4 \sqrt{\frac{\log|\mathscr{H}|}{m(\mathscr{H})} \cdot \frac{|V(H)|}{|E(H)|} \cdot \frac{1}{n}} \leq C_5 \sqrt{\frac{M(\mathscr{H})}{\mathscr{R}n}},$$

where C_4, C_5 are absolute constants. Therefore, by union bound, when κ is chosen to be a large enough constant, we have $\mathbb{P}_{0,n}(\psi = 1) \leqslant \alpha/2$.

For any $\Theta \in \mathscr{S}(\mathscr{G}_1, \theta)$ with the corresponding graph G, we now prove that $\mathbb{P}_{\Theta,n}(\psi = 0) < \alpha/2$. By the definition of witnessing set and (6.1), there exists $H \in \mathscr{H}$ which is a subgraph of G and we have $|E(H)|/|V(H)| \geqslant \mathscr{R}/4$. It now suffices to prove that

$$\mathbb{P}_{\Theta,n}\left(\hat{W}_H \leqslant \frac{\kappa}{4}\sqrt{\frac{M(\mathscr{H})}{\mathscr{R}n}}\right) \leqslant \frac{\alpha}{2}.$$

Since H is a subgraph of G, each edge in $E(H)$ is also an edge in $E(G)$. By the second Griffith's inequality [see**31**,**39**], for $\theta \leqslant 1$, we have

$$\mathbb{E}_{\Theta,n}\hat{W}_H \geqslant \tanh(\theta) \geqslant \theta/2.$$

Applying Lemma 6.1 gives $\|W_H\|_{\psi_1} \leqslant C_6|E(H)|^{-1/2}$ for an absolute constant C_6. By [**56**, Proposition 5.16], for $\eta \leqslant |E(H)|^{-1/2}$, we have

$$\mathbb{P}_{\Theta,n}\left(\left|\hat{W}_H - \mathbb{E}_{\Theta,n}\hat{W}_H\right| > \eta\right) \leqslant 2\exp(-C_7 \cdot |E(H)| \cdot n\eta^2),$$

for an absolute constant C_7. Therefore, with probability at least $1 - \alpha/2$, we have

$$\hat{W}_H \geqslant \mathbb{E}_{\Theta}\hat{W}_H - C_8\sqrt{\frac{\log(2/\alpha)}{|E(H)|n}} \geqslant \frac{\theta}{2} - C_9\sqrt{\frac{|V(H)|}{m(\mathscr{H})} \cdot \frac{\log(|\mathscr{H}|)}{|E(H)|n}} > \left(\frac{\kappa}{2} - C_{10}\right)\sqrt{\frac{M(\mathscr{H})}{\mathscr{R}n}},$$

where C_8, C_9 and C_{10} are absolute constants. Therefore, when κ is chosen as a large enough constant, we have $\mathbb{P}_{\Theta,n}(\psi = 0) \leqslant \alpha/2$ and

$$\mathbb{P}_{0,n}(\psi = 1) + \sup_{\Theta \in \mathscr{S}(\mathscr{G}_1,\theta)} \mathbb{P}_{\Theta,n}(\psi = 0) \leqslant \alpha.$$

This completes the proof.

7. Proof of the theorems in Section 4

Proof. of Theorem 4.3 We consider the normal distribution $N(0, (1 - 2s\theta)^{-1})$ which has density

$$\left(\frac{1 - 2s\theta}{2\pi}\right)^{1/2} \exp(s\theta y^2)\exp(-y^2/2).$$

Denote with

$$\int_{-\infty}^{\infty} \cosh(\sqrt{2\theta}y)^s \frac{\exp(-y^2/2)}{\sqrt{2\pi}}\,\mathrm{d}y = \frac{Z_{Y'}}{\sqrt{2\pi}} =: C(\theta, s),$$

where $Z_{Y'}$ is the notation used in Lemma 4.1.

The proof begins by using Pinsker's inequality to bound the total variation with the square root of the KL divergence. We need to control

$$n\int_{-\infty}^{\infty}\log\frac{(\frac{1-2s\theta}{2\pi})^{1/2}\exp(s\theta y^2)\exp(-y^2/2)}{\cosh(\sqrt{2\theta}y)^s\frac{\exp(-y^2/2)}{C(\theta,s)\sqrt{2\pi}}}\left(\frac{1-2s\theta}{2\pi}\right)^{1/2}\exp(s\theta y^2)\exp(-y^2/2)\,\mathrm{d}y$$

$$=n\left(\log(C(\theta,s)\sqrt{1-2s\theta})+\int_{-\infty}^{\infty}(s\theta y^2-s\log\cosh(\sqrt{2\theta}y))\left(\frac{1-2s\theta}{2\pi}\right)^{1/2}\exp(s\theta y^2)\exp(-y^2/2)\,\mathrm{d}y\right).$$

One can check that $x^2/2-\log(\cosh(x))-x^4/12\leqslant 0$ (we give a proof in the supplement); therefore, the above is bounded as

$$n\left(\log(C(\theta,s)\sqrt{1-2s\theta})+\int_{-\infty}^{\infty}s\frac{4\theta^2y^4}{12}\left(\frac{1-2s\theta}{2\pi}\right)^{1/2}\exp(s\theta y^2)\exp(-y^2/2)\,\mathrm{d}y\right)$$

$$=n(\log(C(\theta,s)\sqrt{1-2s\theta})+s\theta^2(1-2s\theta)^{-2}).$$

We will now bound $C(\theta,s)$ from above. By Lemma 4.1, it follows that

$$C(\theta,s)=\frac{Z_{Y'}}{\sqrt{2\pi}}=\sum_{k=0}^{s}\frac{\binom{s}{k}\exp(\theta(2k-s)^2)}{2^s}=\sum_{m=0}^{\infty}\frac{\theta^m\mathbb{E}(2k-s)^{2m}}{m!},$$

where $k\sim\mathrm{Bin}(s,.5)$.

We will now show that the central moments of the balanced binomial distribution, $\mathrm{Bin}(s,.5)$, satisfy certain sharp inequalities in terms of polynomials in s. Note that the moments $\mathbb{E}(2k-s)^{2m}$ can be thought of as $\mathbb{E}(X_1+\ldots+X_s)^{2m}$ where X_i are i.i.d. Rademacher random variables. Denote with $P_{2m}(s)=\mathbb{E}(X_1+\ldots+X_s)^{2m}$. We will now show the following recursive formula for $P_{2m}(s)$.

LEMMA 7.1 (Recursion for $P_{2m}(s)$).

$$P_{2m}(s)=\sum_{k=0}^{m-1}(-1)^k\binom{2m-1}{2k+1}E_{2k+1}sP_{2m-2k-2}(s),$$

where E_{2k+1} are the tangent (aka zag) numbers, the first few of which are given by $E_1=1$, $E_3=2$, $E_5=16$, $E_7=272$ and so on. In addition, it holds that (for $s\in\mathbb{N}$),

$$\sum_{k=2l+1}^{m-1}(-1)^k\binom{2m-1}{2k+1}E_{2k+1}sP_{2m-2k-2}(s)\leqslant 0, \tag{7.1}$$

$$\sum_{k=2l}^{m-1}(-1)^k\binom{2m-1}{2k+1}E_{2k+1}sP_{2m-2k-2}(s)\geqslant 0. \tag{7.2}$$

The proof of Lemma 7.1 (and all subsequent lemmas) is deferred to the supplement. Here, we give a little comment on the tangent numbers: first, they appear in the Taylor expansion of $\tan(x)$ around 0. Second, they are the number of alternating (aka zigzag) permutations of $2k+1$ numbers, i.e., permutations of the type $a_1 < a_2 > a_3 < a_4 > a_5 < \ldots > a_{2k+1}$. Using Lemma 7.1 and the fact that $\mathbb{E}(2k-s)^0 = 1$, it is simple to verify that $\mathbb{E}(2k-s)^{2m}$ are polynomials of s with integer coefficients of degree m. Suppose that $P_{2m}(s) = \sum_{k=0}^{m} a_k^{2m} s^{m-k}$. Next, we need the following.

LEMMA 7.2 (Sharp bounds for $P_{2m}(s)$). We have that for any $s \in \mathbb{N}$ and l, the following is true (provided that the index in the summation makes sense):

$$\sum_{k=0}^{2l+1} a_k^{2m} s^{m-k} \leqslant P_{2m}(s) \leqslant \sum_{k=0}^{2l} a_k^{2m} s^{m-k}.$$

We are now ready to begin our calculations. We will bound the moments of $\mathbb{E}(2k-s)^{2m}$ with the first three coefficients of the polynomial $P_{2m}(s)$. In order to do so, we need to derive the values of these coefficients. We do so in the following lemma.

LEMMA 7.3 (Coefficients of $P_{2m}(s)$). We have

$$a_0^{2m} = (2m)!!, a_1^{2m} = -m(m-1)(2m)!!/3$$
$$a_2^{2m} = (2m-1)a_2^{2m-2} + 2\binom{2m-1}{3}(m-2)(m-3)\frac{(2m-4)!!}{3} + 16\binom{2m-1}{5}(2m-6)!!,$$

where a_0^{2m} is defined as 1 for $m = 0$, a_2^{2m} is defined for $m \geqslant 3$ and is 0 otherwise.

Next, we will control the coefficient a_2^{2m} from above. We have the following lemma.

LEMMA 7.4 We have that for a sufficiently large absolute constant $C > 0$,

$$a_2^{2m} \leqslant C(m-2)(m-1)m(m+1)(2m)!!.$$

Using Lemma 7.4, we conclude that

$$\mathbb{E}(2k-s)^{2m} \leqslant (2m)!!\, s^m - m(m-1)(2m)!!/3 s^{m-1} + C(m-2)(m-1)m(m+1)(2m)!!\, s^{m-2}.$$

Hence,

$$C(\theta, s) \leqslant \sum_{m=0}^{\infty} \frac{(2m)!!\,(\theta s)^m}{m!} - \theta \sum_{m=2}^{\infty} \frac{m(m-1)(2m)!!\,(\theta s)^{m-1}}{3m!} + \theta^2 \sum_{m=3}^{\infty} \frac{C(m-2)(m-1)m(m+1)(2m)!!\,(\theta s)^{m-2}}{m!}.$$

Simple algebra verifies that $\frac{(2m)!!}{m!} = \binom{-1/2}{m}(-2)^m$ and $\frac{m(m-1)(2m)!!}{3m!} = \binom{-5/2}{m-2}(-2)^{m-2}$ and therefore (using $(2m)!\,!/m! \leqslant 2^m$),

$$C(\theta, s) \leqslant (1-2s\theta)^{-1/2} - s\theta^2(1-2s\theta)^{-5/2} + \theta^2(\theta s)\sum_{m=3}^{\infty} C(m-2)(m-1)m(m+1)(2\theta s)^{m-3}.$$

Now, we use a simple trick that

$$\sum_{m=3}^{\infty}(m-2)(m-1)m(m+1)x^{m-3} = \frac{d^4}{dx^4}\Big(\sum_{m=3}^{\infty} x^{m+1}\Big) = \frac{d^4}{dx^4}\frac{x^4}{1-x} = \sum_{i=1}^{5} C_i \frac{x^{i-1}}{(1-x)^i},$$

where C_i are some absolute constants. Finally, we obtain

$$C(\theta, s) \leqslant (1-2s\theta)^{-1/2} - s\theta^2(1-2s\theta)^{-5/2} + \theta^2 \sum_{i=1}^{5} C_i' \frac{(s\theta)^i}{(1-2s\theta)^i},$$

for some absolute constants. The above can clearly be made very small as $s\theta$ is made small.

Finally, using $\log(1-x) \leqslant -x$ for $x \geqslant 0$ small enough, we have

$$n\log(C(\theta, s)\sqrt{1-2s\theta}) + ns\theta^2(1-2s\theta)^{-2} \asymp n\theta^2 \sum_{i=1}^{5} C_i' \frac{(s\theta)^i}{(1-2s\theta)^{i-1/2}},$$

which is going to be very small given that $\theta \ll \frac{1}{\sqrt{n}} \wedge \frac{1}{s}$. □

Proof of Theorem 4.2. Recall that $\boldsymbol{U}_i$ and $\boldsymbol{V}_i$ for $i \in [n]$, denote the n i.i.d. copies of the models $\mathrm{sign}(N(0, I + \sigma 1_{[s]}1_{[s]}^T))$ and n i.i.d. copies of the Curie–Weiss model with parameters s and θ. Let $\mathbf{U} = \{\boldsymbol{U}_i\}_{i\in[n]}$ and $\mathbf{V} = \{\boldsymbol{V}_i\}_{i\in[n]}$ be the collections of the variables $\boldsymbol{U}_i$ and $\boldsymbol{V}_i$ for $i \in [n]$, and recall that $\boldsymbol{Y}$ and $\boldsymbol{Y}'$ are the collections of the variables $\{Y_i\}_{i\in[n]}$ and $\{Y_i'\}_{i\in[n]}$, i.e., $\boldsymbol{Y}$ is a vector consisting of n i.i.d copies from a Gaussian distribution with variance $(1-2s\theta)^{-1}$ and $\boldsymbol{Y}'$ is a vector consisting of n i.i.d. copies from a CWN distribution with parameters s, θ. As we argued earlier, it suffices to prove that the total variation $\mathrm{TV}(\mathscr{L}(\mathbf{U}), \mathscr{L}(\mathbf{V}))$ is small.

To do so we will use [**12**, Fact 3.1. Part 5 in the arXiv version], which states that

$$\mathrm{TV}(\mathscr{L}(\mathbf{U}), \mathscr{L}(\mathbf{V})) \leqslant \mathrm{TV}(\mathscr{L}(\boldsymbol{Y}), \mathscr{L}(\boldsymbol{Y}')) + \mathbb{E}_{\boldsymbol{y}\sim\boldsymbol{Y}}\,\mathrm{TV}(\mathscr{L}(\mathbf{U}|\boldsymbol{Y}=\boldsymbol{y}), \mathscr{L}(\mathbf{V}|\boldsymbol{Y}'=\boldsymbol{y})).$$

By Theorem 4.3, we know that $\mathrm{TV}(\mathscr{L}(\boldsymbol{Y}), \mathscr{L}(\boldsymbol{Y}'))$ is small under the condition $\theta \ll \frac{1}{\sqrt{n}} \wedge \frac{1}{s}$. We will now argue that the second term $\mathrm{TV}(\mathscr{L}(\mathbf{U}|\boldsymbol{Y}=\boldsymbol{y}), \mathscr{L}(\mathbf{V}|\boldsymbol{Y}'=\boldsymbol{y}))$ is also small under the same condition which will complete the proof.

Recall that $\sigma = \frac{\pi\theta}{1-2s\theta}$, and let $\kappa = \pi\theta$. Let

$$g(\sqrt{\kappa 2/\pi}\,y) := \frac{\exp(\sqrt{\kappa 2/\pi}\,y)}{\exp(\sqrt{\kappa 2/\pi}\,y) + \exp(-\sqrt{\kappa 2/\pi}\,y)}.$$

We will now control

$$\begin{aligned}
&\mathrm{TV}(\mathscr{L}(\mathbf{U}|\boldsymbol{Y}=\boldsymbol{y}),\mathscr{L}(\mathbf{V}|\boldsymbol{Y}'=\boldsymbol{y}))^2\\
&\quad=\Bigg(\sum_{\{k_i\}_{i=1}^n}\Bigg|\prod_{i\in[n]}\binom{s}{k_i}(\Phi(\sqrt{\kappa}y_i))^{k_i}(1-\Phi(\sqrt{\kappa}y_i))^{s-k_i}\\
&\quad-\prod_{i\in[n]}\binom{s}{k_i}g(\sqrt{\kappa 2/\pi}y_i)^{k_i}(1-g(\sqrt{\kappa 2/\pi}y_i)^{s-k_i}\Bigg|\Bigg)^2\\
&\quad\leqslant\frac{n}{2}\sum_{k_i=0}^{s}\Bigg[k_i(\log\Phi(\sqrt{\kappa}y_i)-\log g(\sqrt{\kappa 2/\pi}y_i))\\
&\quad+(s-k_i)(\log(1-\Phi(\sqrt{\kappa}y_i))-\log(1-g(\sqrt{\kappa 2/\pi}y_i))\Bigg]\times\\
&\quad\times\binom{s}{k_i}\Phi(\sqrt{\kappa}y_i)^{k_i}(1-\Phi(\sqrt{\kappa}y_i))^{s-k_i},
\end{aligned}\tag{7.3}$$

where the last inequality is Pinsker's inequality. We now write the following inequalities which can be verified with a direct calculation from the Taylor expansions of the functions (we provide short proofs in the supplement)

$$\begin{aligned}
\log\Phi(x)&\leqslant-\log(2)+\sqrt{\frac{2}{\pi}}x-\frac{x^2}{\pi}-\frac{\pi-4}{3\sqrt{2}\pi^{3/2}}x^3+\frac{\pi-3}{3\pi^2}x^4+\frac{96-40\pi+3\pi^2}{60\sqrt{2}\pi^{5/2}}x^5+Cx^6,\\
\log(1-\Phi(x))&\leqslant-\log(2)-\sqrt{\frac{2}{\pi}}x-\frac{x^2}{\pi}+\frac{\pi-4}{3\sqrt{2}\pi^{3/2}}x^3+\frac{\pi-3}{3\pi^2}x^4-\frac{96-40\pi+3\pi^2}{60\sqrt{2}\pi^{5/2}}x^5+Cx^6,\\
\log g(\sqrt{2/\pi}x)&\geqslant-\log(2)+\sqrt{\frac{2}{\pi}}x-\frac{x^2}{\pi}+\frac{x^4}{3\pi^2}-\frac{8x^6}{45\pi^3},\\
\log(1-g(\sqrt{2/\pi}x))&\geqslant-\log(2)-\sqrt{\frac{2}{\pi}}x-\frac{x^2}{\pi}+\frac{x^4}{3\pi^2}-\frac{8x^6}{45\pi^3},
\end{aligned}$$

for a sufficiently large absolute constant $C>0$. Hence, continuing the bound from (7.3), we obtain the following:

$$\begin{aligned}
&\sum_{k_i=0}^{s}\Bigg[(s-2k_i)\frac{\pi-4}{3\sqrt{2}\pi^{3/2}}(\sqrt{\kappa}y_i)^3+s\frac{\pi-4}{3\pi^2}(\sqrt{\kappa}y_i)^4+(2k_i-s)\frac{96-40\pi+3\pi^2}{60\sqrt{2}\pi^{5/2}}(\sqrt{\kappa}y_i)^5\\
&\quad+(\frac{8}{45\pi^3}+C)s(\sqrt{\kappa}y_i)^6\Bigg]\binom{s}{k_i}\Phi(\sqrt{\kappa}y_i)^{k_i}(1-\Phi(\sqrt{\kappa}y_i))^{s-k_i}\\
&\quad=\frac{\pi-4}{3\sqrt{2}\pi^{3/2}}s(1-2\Phi(\sqrt{\kappa}y_i))(\sqrt{\kappa}y_i)^3+s\frac{\pi-4}{3\pi^2}(\sqrt{\kappa}y_i)^4+\\
&\quad+s(2\Phi(\sqrt{\kappa}y_i)-1)\frac{96-40\pi+3\pi^2}{60\sqrt{2}\pi^{5/2}}(\sqrt{\kappa}y_i)^5+s(\sqrt{\kappa}y_i)^6(\frac{8}{45\pi^3}+C).
\end{aligned}\tag{7.4}$$

Next, we note that (and give proofs in the supplement)

$$(1-2\Phi(x))x^3 \leqslant -\sqrt{\frac{2}{\pi}}x^4 + \frac{x^6}{3\sqrt{2\pi}},$$

$$(2\Phi(x)-1)x^5 \leqslant \sqrt{\frac{2}{\pi}}x^6.$$

Putting everything together, we conclude that there exists a universal constant C such that (7.4) is bounded as $Cs\kappa^3 y_i^6$. Then, using the fact that square root is a concave function, by Jensen's inequality we can bring the expectations over y_i inside the square root to obtain

$$\mathbb{E}_{y\sim \boldsymbol{Y}} \mathrm{TV}(\mathscr{L}(\mathbf{U}|\boldsymbol{Y}=\boldsymbol{y}), \mathscr{L}(\mathbf{V}|\boldsymbol{Y}'=\boldsymbol{y})) \leqslant \sqrt{ns\kappa^3 C\mathbb{E}y_i^6} \asymp \sqrt{ns\kappa^3/(1-2s\theta)^3}.$$

which completes the proof. □

8. Discussion

In this paper, we studied structure detection problems in zero-field ferromagnetic Ising models. Our upper and lower bounds demonstrated that graph arboricity is a key concept which drives the testability of structure detection. We furthermore argued that under a sparse PCA conjecture no polynomial time tests can test the problem unless the signal strength is of the order of $\frac{1}{\sqrt{n}}$, which is statistically sub-optimal for graphs with high arboricity.

There are several important questions which we leave for future work. First, our upper bound results are derived under the assumption that $\|\Theta\|_F \leqslant \frac{1}{2}$. This assumption is needed to ensure that the terms (3.1) concentrate around their mean value. This may not be a necessary condition, and we anticipate that the tests we develop might work beyond this regime.

Second, an interesting question that is left open is whether one can develop upper and lower bounds for problems of the type (1.5) in the dense regime when $s \gg \sqrt{d}$. We believe that this regime may require completely different tests than the ones we developed in this paper.

Acknowledgements

We would like to thank the anonymous reviewer, associate editor and editor for their helpful comments.

Funding

National Science Foundation (BIGDATA 1840866, RI 1408910, CAREER 1841569, TRIPODS 1740735 to H.L.); Alfred P Sloan Fellowship to H.L.

Data Availability Statement

No new data were generated or analyzed in support of this research.

References

1. Addario-Berry, L., Broutin, N., Devroye, L. & Lugosi, G. (2010) On combinatorial testing problems. *Ann. Statist.*, **38**, 3063–3092.

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

3. Arias-Castro, E., Bubeck, S. & Lugosi, G. (2012) Detection of correlations. *Ann. Statist.*, **40**, 412–435.

4. Arias-Castro, E., Bubeck, S., Lugosi, G., (2015) Detecting positive correlations in a multivariate sample. *Bernoulli*, **21**, 209–241.

5. Arias-Castro, E., Bubeck, S., Lugosi, G. & Verzelen, N. (2018) Detecting Markov random fields hidden in white noise. Bernouli 24(4B) 3628–3656.

6. Bento, J. & Montanari, A. (2009) Which graphical models are difficult to learn? *Advances in Neural Information Processing Systems*, pp. 1303–1311.

7. Berthet, Q. & Rigollet, P. (2013a) Complexity theoretic lower bounds for sparse principal component detection. *Conference on Learning Theory*, pp. 1046–1066.

8. Berthet, Q. & Rigollet, P. (2013b) Optimal detection of sparse principal components in high dimension. *Ann. Statist.*, **41**, 1780–1815.

9. Berthet, Q., Rigollet, P. & Srivastava, P. (2019) Exact recovery in the Ising blockmodel. Ann. Statist.. **47**(4), 1805–1834.

10. Besag, J. (1993) Statistical analysis of dirty pictures. *J. Appl. Statist.*, **20**, 63–87.

11. Bhattacharya, B. B. & Mukherjee, S. (2018) Inference in Ising models. Bernouli. **24**(1), 493–525.

12. Brennan, M. & Bresler, G. (2019) Optimal average-case reductions to sparse PCA: from weak assumptions to strong hardness. In Conference on Learning Theory, pp. 469–470.

13. Brennan, M., Bresler, G. & Huleihel, W. (2018) Reducibility and computational lower bounds for problems with planted sparse structure. In Conference On Learning Theory, pp. 48–166.

14. Bresler, G. (2015) Efficiently learning Ising models on arbitrary graphs. *Proceedings of the Forty-Seventh Annual ACM on Symposium on Theory of Computing*. ACM, pp. 771–782.

15. Cai, T. T., Liu, W. & Luo, X. (2011) A constrained ℓ_1 minimization approach to sparse precision matrix estimation. *J. Amer. Statist. Assoc.*, **106**, 594–607.

16. Daskalakis, C., Dikkala, N. & Kamath, G. (2019) Testing Ising models. *IEEE Transactions on Information Theory* **65**(11), 6829–6852.

17. Dubhashi, D. & Ranjan, D. (1998) Balls and bins: a study in negative dependence. *Random Structures Algorithms*, **13**, 99–124.

18. Durbin, R., Eddy, S. R., Krogh, A. & Mitchison, G. (1998) *Biological Sequence Analysis: Probabilistic Models of Proteins and Nucleic Acids*. Cambridge University Press.

19. Euler, L. (1741) Solutio problematis ad geometriam situs pertinentis. *Comment. Acad. Sci. Petropol.*, **8**, 128–140.

20. Fan, J., Liu, H., Wang, Z. & Yang, Z. (2018) Curse of heterogeneity: computational barriers in sparse mixture models and phase retrieval. arXiv preprint arXiv:1808.06996.

21. Feldman, V., Grigorescu, E., Reyzin, L., Vempala, S. & Xiao, Y. (2017) Statistical algorithms and a lower bound for detecting planted cliques. Journal of the ACM (JACM),. **64**(2), 1–37.

22. Feldman, V., Grigorescu, E., Reyzin, L., Vempala, S. S. & Xiao, Y. (2017) Statistical algorithms and a lower bound for detecting planted cliques. *J. ACM*, **64**, 1–37.

23. Feldman, V., Guzman, C. & Vempala, S. (2017) Statistical query algorithms for mean estimation and stochastic convex optimization. In SIAM Symposium on Discrete Algorithms..

24. Feldman, V., Perkins, W. & Vempala, S. (2018) On the complexity of random satisfiability problems with planted solutions. SIAM Journal on Computing **47**(4), 1294–1338.

25. Feldman, V., Perkins, W. & Vempala, S. (2018) On the complexity of random satisfiability problems with planted solutions. *SIAM J. Comput.*, **47**, 1294–1338.

26. Fisher, M. E. (1967) Critical temperatures of anisotropic Ising lattices. II. General upper bounds. *Phys. Rev.*, **162**, 480.

27. Gao, C., Ma, Z. & Zhou, H. H. (2017) Sparse CCA: adaptive estimation and computational barriers. Ann. Statist. **45**(5), 2074–2101.

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

29. Gheissari, R., Lubetzky, E. & Peres, Y. (2018) Concentration inequalities for polynomials of contracting Ising models Electronic Communications in Probability, **23**.

30. Grabowski, A. & Kosiński, R. (2006) Ising-based model of opinion formation in a complex network of interpersonal interactions. *Phys. A*, **361**, 651–664.

31. Griffiths, R. B. (1967) Correlations in Ising ferromagnets. I. *J. Math. Phys.*, **8**, 478–483.

32. Gu, Q., Cao, Y., Ning, Y. & Liu, H. (2015) Local and global inference for high dimensional Gaussian copula graphical models. arXiv preprint arXiv:1502.02347.

33. Guttman, A. (1989) Asymptotic analysis of power-series expansions, in *Phase Transitions and Critical Phenomena edited by C Domb and J Lebowitz, Academic Press, vol. ch. 1, pp*. 1–234.

34. Hierholzer, C. & Wiener, C. (1873) Über die möglichkeit, einen linienzug ohne wiederholung und ohne unterbrechung zu umfahren. *Math. Ann.*, **6**, 30–32.

35. Ising, E. (1925) Beitrag zur theorie des ferromagnetismus. *Z. Phys. A Hadrons Nuclei*, **31**, 253–258.

36. Jankova, J., Van De Geer, S. (2015) Confidence intervals for high-dimensional inverse covariance estimation. *Electron. J. Stat.*, **9**, 1205–1229.

37. Joag-Dev, K. & Proschan, F. (1983) Negative association of random variables with applications. *Ann. Statist.*, **11**, 286–295.

38. Kearns, M. (1998) Efficient noise-tolerant learning from statistical queries. *J. ACM*, **45**, 983–1006.

39. Kelly, D. G. & Sherman, S. (1968) General Griffiths' inequalities on correlations in Ising ferromagnets. *J. Math. Phys.*, **9**, 466–484.

40. Liu, H., Lafferty, J. & Wasserman, L. (2009) The nonparanormal: semiparametric estimation of high dimensional undirected graphs. *J. Mach. Learn. Res.*, **10**, 2295–2328.

41. Lokhov, A. Y., Vuffray, M., Misra, S. & Chertkov, M. (2018) Optimal structure and parameter learning of Ising models. *Sci. Adv.*, **4**, e1700791.

42. Lu, H., Cao, Y., Lu, J., Liu, H. & Wang, Z. (2018) The edge density barrier: computational-statistical tradeoffs in combinatorial inference. *International Conference on Machine Learning*, pp. 3253–3262.

43. Lu, J., Neykov, M. & Liu, H. (2017) Adaptive inferential method for monotone graph invariants. arXiv preprint arXiv:1707.09114.

44. Ma, Z., Wu, Y. (2015) Computational barriers in minimax submatrix detection. *Ann. Statist.*, **43**, 1089–1116.

45. Meinshausen, N. & Bühlmann, P. (2006) High dimensional graphs and variable selection with the lasso. *Ann. Statist.*, **34**, 1436–1462.

46. Mukherjee, R., Mukherjee, S., Yuan, M. (2018) Global testing against sparse alternatives under Ising models. *Ann. Statist.*, **46**, 2062–2093.

47. Nash-Williams, C. (1961) Edge-disjoint spanning trees of finite graphs. *J. Lond. Math. Soc. (2)*, **1**, 445–450.

48. Neykov, M. & Liu, H. (2019) Property testing in high dimensional Ising models. Ann. Statist. **47**(5), 2472–2503.

49. Neykov, M., Lu, J. & Liu, H. (2019) Combinatorial inference for graphical models. Ann. Statist. **47**(2), 795–827.

50. Neykov, M., Ning, Y., Liu, J. S. & Liu, H. (2018) A unified theory of confidence regions and testing for high dimensional estimating equations. Statistical Science. **33**(3), 427–443.

51. Ravikumar, P., Wainwright, M. J., Lafferty, J. D. (2010) High-dimensional Ising model selection using ℓ_1-regularized logistic regression. *Ann. Statist.*, **38**, 1287–1319.

52. Ravikumar, P., Wainwright, M. J., Raskutti, G. & Yu, B. (2011) High-dimensional covariance estimation by minimizing ℓ_1-penalized log-determinant divergence. *Electron. J. Stat.*, **5**, 935–980.

53. Ren, Z., Sun, T., Zhang, C.-H., Zhou, H. H. (2015) Asymptotic normality and optimalities in estimation of large Gaussian graphical models. *Ann. Statist.*, **43**, 991–1026.

54. Santhanam, N. P. & Wainwright, M. J. (2012) Information-theoretic limits of selecting binary graphical models in high dimensions. *IEEE Trans. Inform. Theory*, **58**, 4117–4134.

55. Tandon, R., Shanmugam, K., Ravikumar, P. K. & Dimakis, A. G. (2014) On the information theoretic limits of learning Ising models. *Advances in Neural Information Processing Systems*, pp. 2303–2311.
56. Vershynin, R. (2010) Introduction to the non-asymptotic analysis of random matrices. arXiv preprint arXiv:1011.3027.
57. Vuffray, M., Misra, S., Lokhov, A. & Chertkov, M. (2016) Interaction screening: efficient and sample-optimal learning of Ising models. *Advances in Neural Information Processing Systems*, pp. 2595–2603.
58. Wang, Z., Gu, Q. & Liu, H. (2015) Sharp computational-statistical phase transitions via oracle computational model. arXiv preprint arXiv:1512.08861.
59. Wasserman, S. & Faust, K. (1994) *Social Network Analysis: Methods and Applications*, vol. 8. Cambridge University Press.
60. Yi, X., Wang, Z., Yang, Z., Caramanis, C. & Liu, H. (2016) More supervision, less computation: statistical-computational tradeoffs in weakly supervised learning. *Advances in Neural Information Processing Systems*, pp. 4482–4490.

Appendix

A. Proofs

A.1 *Lower bound proofs*

We first introduce two important lemmas.

Lemma A.1 For a multigraph $G = (\overline{V}, E)$, define the following two classes of Eulerian spanning subgraphs and connected Eulerian subgraphs of G with k edges.

$$\mathscr{E}_c(k, G) := \big\{\tilde{G} = (\tilde{V}, \tilde{E}) : \tilde{V} \subseteq \overline{V},\ \tilde{E} \subseteq E,\ |\tilde{E}| = k,\ \tilde{G} \text{ is a connected Eulerian graph}\big\},$$

$$\mathscr{E}(k, G) := \left\{\tilde{G} = (\overline{V}, \tilde{E}) : \tilde{E} \subseteq E,\ |\tilde{E}| = k,\ \tilde{G} \text{ is a Eulerian graph}\right\}.$$

Let A be the adjacency matrix of G. Then, for $k \geqslant 2$, we have

$$|\mathscr{E}_c(k, G)| \leqslant \|\mathrm{A}\|_F^k, \text{ and } |\mathscr{E}(k, G)| \leqslant 2^k \|\mathrm{A}\|_F^k.$$

Proof. For the first inequality, note that we have

$$\begin{aligned}(\mathrm{A}^k)_{(i,i)} &= \sum_{r_1,\ldots,r_{k-1} \in \overline{V}} \mathrm{A}_{ir_1}\mathrm{A}_{r_1r_2}\cdots\mathrm{A}_{r_{k-2}r_{k-1}}\mathrm{A}_{r_{k-1}i} \\ &= \text{the number of length-}k \text{ closed walks starting at vertex } i. \end{aligned} \tag{A.1}$$

Summing up all possible starting vertices, we get

$$|\mathscr{E}_c(k, G)| \leqslant |\{\text{legnth-}k \text{ closed walks in} G\}| \leqslant \mathrm{Tr}\left(\mathrm{A}^k\right) \leqslant \|\mathrm{A}\|_F^k.$$

This proves the first inequality. For the second inequality, we use induction. First, for $|\mathscr{E}(2, G)|$, we have

$$|\mathscr{E}(2, G)| = |\mathscr{E}_c(2, G)| \leqslant \|\mathrm{A}\|_F^2 \leqslant 2^2 \|\mathrm{A}\|_F^2.$$

Suppose that for $l \leqslant k$, we have $|\mathscr{E}(l,G)| \leqslant 2^l \|\mathrm{A}\|_F^l$. Then, for $|\mathscr{E}(k+1,G)|$, by the fact that $\mathscr{E}(1,G) = \mathscr{E}_c(1,G) = \emptyset$, we have

$$|\mathscr{E}(k+1,G)| \leqslant \sum_{l=2}^{k-1} |\mathscr{E}_c(l,G)| \cdot |\mathscr{E}(k+1-l,G)| + |\mathscr{E}_c(k+1,G)|.$$

Plugging in the inequalities for $|\mathscr{E}(l,G)|$, we get

$$\begin{aligned} |\mathscr{E}(k+1,G)| &\leqslant \sum_{l=2}^{k-1} \|\mathrm{A}\|_F^l \cdot 2^{k+1-l} \|\mathrm{A}\|_F^{k+1-l} + \|\mathrm{A}\|_F^{k+1} \\ &\leqslant \|\mathrm{A}\|_F^{k+1} \cdot \left(\sum_{l=2}^{k-1} 2^{k+1-l} + 1 \right) \\ &\leqslant 2^{k+1} \|\mathrm{A}\|_F^{k+1}. \end{aligned}$$

Therefore, by induction, we get the second inequality. □

LEMMA A.2 Let G be a multigraph with vertex set $\overline{V} = \{1, \ldots, d\}$ and adjacency matrix A. Let $V \subseteq \overline{V}$ be a vertex set. For $k \geqslant 2$, we define

$$\begin{aligned} p_k(G,V) &= \left| \left\{ \tilde{G} \in \mathscr{E}_c(k,G) : \tilde{G} \text{ contains at least two distinct vertices in} V \right\} \right|, \\ q_k(G,V) &= \left| \left\{ \tilde{G} \in \mathscr{E}(k,G) : \exists i,j \in V, i,j \text{ are contained in one connected} \right. \right. \\ &\qquad \left. \left. \text{component of} \tilde{G} \right\} \right|. \end{aligned}$$

Then, we have

$$p_k(G,V) \leqslant (k-1) \cdot |V|^2 \cdot \|\mathrm{A}\|_1^{k-2}, \tag{A.2}$$

$$q_k(G,V) \leqslant \left(2^k \cdot |V| \cdot \|\mathrm{A}\|_F^k \right) \wedge \left[k \cdot 2^{k-2} \cdot |V|^2 \cdot (\|\mathrm{A}\|_1 \vee \|\mathrm{A}\|_F)^{k-2} \right]. \tag{A.3}$$

Proof. We first prove (A.2). By definition, we have

$$\begin{aligned} p_k(G,V) &\leqslant |V| \cdot (|V|-1) \cdot \max_{i,j \in V} \left| \left\{ \tilde{G} \in \mathscr{E}_c(k,G) : \tilde{G} \text{ contains vertices } i \text{ and } j \right\} \right| \\ &\leqslant |V|^2 \cdot \max_{i,j \in V} |\{\text{length-}k \text{ closed walks in } G \text{ starting at } i \text{ and traversing } j\}|. \end{aligned}$$

Note that each vertex can have at most $\|\mathrm{A}\|_1$ neighbors. Therefore, we can bound the number of length-k Eulerian circuits starting at vertex i and containing vertex j by counting the possible vertices on the walk:

- the number of possible positions of vertex j in V is $k-1$;
- the number of choices of the rest $k-2$ vertices is at most $\|\mathrm{A}\|_1^{k-2}$.

This completes the proof of (A.2).

Now, we prove (A.3). Suppose that $\tilde{G}$ is a subgraph of G with k edges such that one of its connected components contains at least two distinct vertices in V. Let l be the number of edges of this connected

component. Then, by definition, clearly the rest connected components form a graph in $\mathscr{E}(k-l,G)$. Therefore, we have

$$q_k(G,V) \leqslant \sum_{l=2}^{k-2} p_l(G,V)\cdot|\mathscr{E}(k-l,G)| + p_k(G,V).$$

By (A.2) and Lemma A.1, we have

$$\begin{aligned} q_k(G,V) &\leqslant \sum_{l=2}^{k-2}(l-1)\cdot|V|^2\|\mathrm{A}\|_1^{l-2}\cdot 2^{k-l}\|\mathrm{A}\|_F^{k-l} + k\cdot|V|^2\|\mathrm{A}\|_1^{k-2} \\ &\leqslant |V|^2\cdot(\|\mathrm{A}\|_1\vee\|\mathrm{A}\|_F)^{k-2}\cdot\left(\sum_{l=2}^{k-2}(l-1)\cdot 2^{k-l}+k\right) \\ &\leqslant k\cdot 2^{k-2}\cdot|V|^2\cdot(\|\mathrm{A}\|_1\vee\|\mathrm{A}\|_F)^{k-2}, \end{aligned}$$

where the last inequality holds because for $l\geqslant 2$, we have $l-1\leqslant 2^{l-2}$. Moreover, for $V\neq\emptyset$, by Lemma A.1, clearly we have $q_k(G,V)\leq|\mathscr{E}(k,G)|\leqslant 2^k\|\mathrm{A}\|_F^k\leqslant 2^k\cdot|V|\cdot\|\mathrm{A}\|_F^k$. When $V=\emptyset$, by definition we have $q_k(G,V)=2^k\cdot|V|\cdot\|\mathrm{A}\|_F^k=0$. This completes the proof. □

Proof of Lemma 5.1. For any $i,j\in\overline{V}$, we have

$$\exp(\theta X_iX_j)=\cosh(\theta X_iX_j)+\sinh(\theta X_iX_j)=\cosh(\theta X_iX_j)[1+\tanh(\theta X_iX_j)].$$

Note that $\cosh(x)$ is an even function and X_iX_j is binary. Therefore, we have $\cosh(\theta X_iX_j)\equiv\cosh(\theta)$. Similarly, $\tanh(x)$ is an odd function, by checking the function values at $X_iX_j=1$ and $X_iX_j=-1$, we obtain $\tanh(\theta X_iX_j)=\tanh(\theta)X_iX_j$. Therefore, we have

$$\exp(\theta X_iX_j)=c(1+tX_iX_j), \tag{A.4}$$

where $c=\cosh(\theta)$ and $t=\tanh(\theta)$. Plugging (A.4) into the definition of $\mathbb{P}_\Theta(\boldsymbol{X})$ proves (5.1). □

Proof of Lemma 5.2. Define

$$\overline{\mathbb{P}}=\frac{1}{|\mathscr{S}^*|}\sum_{\Theta\in\mathscr{S}^*}\mathbb{P}_{\Theta,n},$$

then by Neyman–Pearson's lemma, we have

$$\gamma(\mathscr{S}^*)\geqslant\inf_{\psi}\left[\mathbb{P}_0(\psi=1)+\overline{\mathbb{P}}(\psi=0)\right]=1-\mathrm{TV}(\overline{\mathbb{P}},\mathbb{P}_{0,n}),$$

where $\mathrm{TV}(\overline{\mathbb{P}},\mathbb{P}_{0,n}):=\max_{A\subseteq\{\pm1\}^{n\times d}}|\overline{\mathbb{P}}(A)-\mathbb{P}_{0,n}(A)|$ is the total variation distance between $\overline{\mathbb{P}}$ and $\mathbb{P}_{0,n}$. Note that for total variation distance, we have

$$\begin{aligned} \mathrm{TV}(\overline{\mathbb{P}},\mathbb{P}_{0,n}) &= \frac{1}{2}\sum_{X\in\{\pm1\}^{n\times d}}|\overline{\mathbb{P}}(X)-\mathbb{P}_{0,n}(X)| \\ &= \frac{1}{2}\sum_{X\in\{\pm1\}^{n\times d}}\left|\frac{\overline{\mathbb{P}}(X)}{\mathbb{P}_{0,n}(X)}-1\right|\cdot\mathbb{P}_{0,n}(X). \end{aligned}$$

Applying Cauchy–Schwartz inequality to the right-hand side above gives

$$\mathrm{TV}(\overline{\mathbb{P}}, \mathbb{P}_{0,n}) \leqslant \frac{1}{2}\sqrt{\mathbb{E}_{0,n}\left\{\left[\frac{\overline{\mathbb{P}}(X)}{\mathbb{P}_{0,n}(X)} - 1\right]^2\right\}} = \frac{1}{2}\sqrt{\mathbb{E}_{0,n}\left[\frac{\overline{\mathbb{P}}^2(X)}{\mathbb{P}^2_{0,n}(X)}\right] - 1}.$$

It then suffices to show that

$$\mathbb{E}_{0,n}\left[\frac{\overline{\mathbb{P}}^2(X)}{\mathbb{P}^2_{0,n}(X)}\right] = \frac{1}{|\mathscr{S}^*|^2}\sum_{\Theta,\Theta'\in\mathscr{S}^*}\mathbb{E}_{0,n}\left[\frac{\mathbb{P}_{\Theta,n}}{\mathbb{P}_{0,n}}\frac{\mathbb{P}_{\Theta',n}}{\mathbb{P}_{0,n}}\right],$$

which follows by direct calculation. □

Proof of Lemma 5.4. Since there cannot be multiple edges in G connecting the same two vertices, the coefficient of t^2 in $f_G(t)$ is 0. For the same reason the coefficient of t^2 in $f_{G'}(t)$ is also 0. In $f_{G,G'}(t)$, the only possible way to form a two-edge Eulerian circuit is to pick one edge from $E(G)$ and to pick another edge from $E(G')$ connecting to the same two vertices. Therefore, $u_2 = |E(G) \cap E(G')|$.

For u_3, note that 3-edge Eulerian subgraphs must be triangles. If a triangle only uses edges in $E(G)$, then it is counted in the coefficient of t^3 in $f_G(t)$. Similarly, if a triangle only uses edges in G', it is also counted in the coefficient of t^3 in $f_{G'}(t)$. Therefore, u_3 is the number of triangles that use at least one edge in $E(G)$ and another edge in $E(G')$, which is defined as $\Delta_{G.G'}$.

We denote by $\mathscr{E}(G)$ and $\mathscr{E}(G')$ the sets of Eulerian subgraphs of G and G', respectively. For $k \geqslant 4$, by (5.5), the coefficient of t^k in $f_G(t)f_{G'}(t)$ is equal to

$$|\{\tilde{G} \in \mathscr{E}(k, G \oplus G') : \exists G_1 \in \mathscr{E}(G), G_2 \in \mathscr{E}(G') \; s.t. \; \tilde{G} = G_1 \oplus G_2\}|.$$

We now prove that, for $\tilde{G} \in \mathscr{E}(k, G \oplus G')$, if each connected component contains at most one vertex in $V(G) \cap V(G')$, then there exist $G_1 \in \mathscr{E}(G)$ and $G_2 \in \mathscr{E}(G')$ such that $\tilde{G} = G_1 \oplus G_2$. To prove this statement, take a fixed connected component of $\tilde{G}$. Suppose first that the connected component does not contain any vertices in $V(G) \cap V(G')$. Then, it follows that all of its edges must be contained either in $E(G)$ or $E(G')$. Next, consider the case when the connected component contains only one vertex $v \in V(G) \cap V(G')$. Since this connected component must be a connected Eulerian graph, we can consider the Eulerian circuit starting and ending at v. If we start walking along the circuit on an edge in $E(G)$, then since v is the only vertex contained in the intersection $V(G) \cap V(G')$, we cannot reach vertices in $E(G')$ until we return to v. Upon returning to v, we have completed a closed walk using purely edges in G. We can continue this process to obtain closed walks on G and G' starting and ending at v. Concatenating all the closed walks on G gives G_1. Similarly, concatenating all the closed walks on G' gives G_2. We have proved that

$$\mathscr{E}(k, G \oplus G')\backslash\{\tilde{G} \in \mathscr{E}(k, G \oplus G') : \exists G_1 \in \mathscr{E}(G), G_2 \in \mathscr{E}(G') \; s.t. \; \tilde{G} = G_1 \oplus G_2\} \subseteq$$
$$\left\{\tilde{G} \in \mathscr{E}(k, G) : \exists i, j \in V(G) \cap V(G'), i, j \text{ are in one connected component of } \tilde{G}\right\}.$$

Therefore, by the definition of $q_k(\cdot,\cdot)$, we have $u_k \leqslant q_k[G \oplus G', V(G) \cap V(G')]$. □

Proof of Lemma 5.5. The bounds for $\mathscr{E}(k, G)$ and $q_k(G, V)$ are included in Lemmas A.1 and A.2. We now prove the bound for $\Delta_{G,G'}$. We remind the reader that for a graph G and a vertex set V, G_V denotes the graph obtained by restricting G on the vertex set V. Note that if a triangle has one edge in $E(G)$ and two edges in $E(G')$, then the two vertices of the edge in $E(G)$ must be in $V(G) \cap V(G')$. Therefore, an

upper bound of the number of triangles that have one edge in $E(G)$ and two edges in $E(G')$ is given by the following procedure:

- pick an edge e from $E[G_{V(G)\cap V(G')}]$;
- pick a common neighbor of the two vertices of edge e.

Since all graphs in $\mathscr{G}^*$ have arboricity $\mathscr{R}$, by the definition of arboricity, we have

$$\begin{aligned}\Delta_{G,G'} &\leqslant \left|E[G_{V(G)\cap V(G')}]\right| \cdot \|\mathrm{A}_{G'}\|_1 + \left|E[G'_{V(G)\cap V(G')}]\right| \cdot \|\mathrm{A}_G\|_1 \\ &\leqslant 2|V(G)\cap V(G')| \cdot \mathscr{R} \cdot \Gamma.\end{aligned}$$

This completes the proof. □

Proof of Lemma 5.6. Let

$$A(\mathscr{G}^*) = \frac{1}{|\mathscr{S}^*|^2} \sum_{\Theta,\Theta'\in\mathscr{S}^*} \exp[3n\mathscr{R}|V(G)\cap V(G')|\theta^2].$$

Then, we have

$$A(\mathscr{G}^*) \leqslant \max_{\Theta\in\mathscr{S}^*} \frac{1}{|\mathscr{S}^*|} \sum_{\Theta'\in\mathscr{S}^*} \exp\left\{3n\mathscr{R}\theta^2 \cdot \sum_{v\in V(G)} \mathbb{1}\left[v \in V(G')\right]\right\}.$$

Consider drawing Θ' uniformly from $\mathscr{S}^*$, and let $\mathbb{P}_{\Theta'\sim U(\mathscr{S}^*)}$ be the probability measure. By assumption, the random variables $\{\mathbb{1}[v \in V(G')] \mid v \in V(G)\}$ are negatively associated. Therefore,

$$\begin{aligned}A(\mathscr{G}^*) &\leqslant \max_{\Theta\in\mathscr{S}^*} \mathbb{E}_{\Theta'\sim U(\mathscr{S}^*)} \prod_{v\in V(G)} \exp\left\{3n\mathscr{R}\theta^2 \cdot \mathbb{1}\left[v \in V(G')\right]\right\} \\ &\leqslant \max_{\Theta\in\mathscr{S}^*} \prod_{v\in V(G)} \mathbb{E}_{\Theta'\sim U(\mathscr{S}^*)} \exp\left\{3n\mathscr{R}\theta^2 \cdot \mathbb{1}\left[v \in V(G')\right]\right\}.\end{aligned}$$

Expanding the expectation and applying the inequality $1 + x \leqslant \exp(x)$ give

$$\begin{aligned}A(\mathscr{G}^*) &\leqslant \max_{\Theta\in\mathscr{S}^*} \prod_{v\in V(G)} \left\{\exp\left(3n\mathscr{R}\theta^2\right)\mathbb{P}_{\Theta'\sim U(\mathscr{S}^*)}[v \in V(G')] + 1 \right. \\ &\qquad\qquad \left. - \mathbb{P}_{\Theta'\sim U(\mathscr{S}^*)}[v \in V(G')]\right\} \\ &\leqslant \max_{\Theta\in\mathscr{S}^*} \prod_{v\in V(G)} \exp\left\{\left[\exp\left(3n\mathscr{R}\theta^2\right) - 1\right]\mathbb{P}_{\Theta'\sim U(\mathscr{S}^*)}[v \in V(G')]\right\}.\end{aligned}$$

Rearranging terms, we get

$$\begin{aligned} A(\mathscr{G}^*) &\leqslant \max_{\Theta\in\mathscr{S}^*} \exp\left\{\left[\exp\left(3n\mathscr{R}\theta^2\right)-1\right]\cdot \sum_{v\in V(G)} \mathbb{P}_{\Theta'\sim U(\mathscr{S}^*)}[v\in V(G')]\right\} \\ &\leqslant \exp\left\{\exp\left(3n\mathscr{R}\theta^2\right)\cdot \max_{\Theta\in\mathscr{S}^*} \mathbb{E}_{\Theta'\sim U(\mathscr{S}^*)}|V(G)\cap V(G')|\right\} \\ &= \exp[N(\mathscr{G}^*)\cdot\exp(3n\mathscr{R}\theta^2)]. \end{aligned}$$

This completes the proof. □

Proof of Corollary 2.1. Let $\mathscr{G}^*$ be the set of graphs isomorphic to G_*. Then, clearly, if G' is uniformly sampling from $\mathscr{G}^*$, then $\{\mathbb{1}[i\in V(G')]\}_{i=1}^d$ is just a permutation of s 1s and $d-s$ 0s. Therefore, by [37, Theorem 2.11], the incoherence condition is satisfied. For any $G\in\mathscr{G}^*$ and $v\in V(G)$, we have

$$\mathbb{E}_{G'\sim U(\mathscr{G}^*)}|V(G)\cap V(G')| = \sum_{i\in V(G)} \mathbb{E}_{G'\sim U(\mathscr{G}^*)}\mathbb{1}[i\in V(G')] = s\cdot s/d = s^2/d.$$

And therefore, $N(\mathscr{G}^*)=s^2/d$. Moreover, by definition, we have

$$\mathscr{R}=\mathscr{R}(G_*),\ V_{\max}=s,\ \Lambda=\|\mathrm{A}_{G_*}\|_F,\ \Gamma=\|\mathrm{A}_{G_*}\|_1 \text{and} \mathscr{B}=\mathscr{B}(G_*).$$

Therefore, by Theorem 2.2, if

$$\theta\leqslant\sqrt{\frac{\log(d/s^2)}{6n\mathscr{R}(G_*)}}\wedge\sqrt{\frac{\mathscr{R}(G_*)}{\mathscr{B}(G_*)}}\wedge\frac{1}{8(\|\mathrm{A}_{G_*}\|_F\vee\|\mathrm{A}_{G_*}\|_1)},$$

then we have

$$\liminf_{n\to\infty}\gamma(\mathscr{S}^*)=1.$$

□

A.2 *Upper bound proofs*

The following lemma given by [11] is helpful for bounding the ψ_1-norm of W_H.

LEMMA A.3 Let J be a $d\times d$ symmetric matrix with non-negative off-diagonal entries and zeros on the diagonal. If $\|J\|_2\leqslant 1$, then we have

$$\sum_{1\leqslant i,j\leqslant d}\log\cosh(J_{ij})\leqslant\log\mathbb{E}_0\exp\left(\frac{1}{2}X^TJX\right)\leqslant-\frac{1}{2}\sum_{i=1}^n\log[1-\lambda_i(J)],$$

where $\lambda_1(J),\ldots,\lambda_d(J)$ are the eigenvalues of J.

Proof of Lemma 6.1. By [56, (5.16)] as an equivalent definition of ψ_1-norm, it suffices to prove

$$\mathbb{E}_\Theta\exp\left(\frac{\sqrt{2}|E(H)|^{1/2}}{8}\cdot W_H\right)\leqslant e. \tag{A.5}$$

To prove (A.5), first note that we have $\|\mathrm{A}_H\|_F^2 = 2|E(H)|$. By definition of the Ising model, we have

$$\mathbb{E}_\Theta \exp\left(\frac{\sqrt{2}|E(H)|^{1/2}}{8} \cdot W_H\right) = \mathbb{E}_\Theta \exp\left(\frac{\|\mathrm{A}_H\|_F}{8} \cdot W_H\right) = \frac{\mathbb{E}_0 \exp(X^T JX/2)}{\mathbb{E}_0 \exp(X^T \Theta X/2)},$$

where $J := \Theta + \mathrm{A}_H/(4\|\mathrm{A}_H\|_F)$. Therefore,

$$\log \mathbb{E}_\Theta \exp\left(\frac{\|\mathrm{A}_H\|_F}{8} \cdot W_H\right) = \log \mathbb{E}_0 \exp\left(\frac{1}{2}X^T JX\right) - \log \mathbb{E}_0 \exp\left(\frac{1}{2}X^T \Theta X\right). \tag{A.6}$$

By Lemma A.3, we have

$$\log \mathbb{E}_0 \exp\left(\frac{1}{2}X^T JX\right) \leqslant -\frac{1}{2}\sum_{i=1}^n \log[1-\lambda_i(J)] \leqslant \frac{1}{2}\sum_{i=1}^n \left[\lambda_i(J) + 2\lambda_i^2(J)\right],$$

where the second inequality holds because for $|x| \leqslant 3/4$ we have $-\log(1-x) = \sum_{k\geqslant 1} x^k/k \leqslant x + 2x^2$ and by assumption, we have $\|J\|_2 \leqslant \|J\|_F \leqslant 3/4$. Since $\mathrm{Tr}(J) = \mathrm{Tr}(\mathrm{A}_H)/(2\|\mathrm{A}_H\|_F) = 0$, we have

$$\log \mathbb{E}_0 \exp\left(\frac{1}{2}X^T JX\right) \leqslant \|J\|_F^2 \leqslant \frac{9}{16}. \tag{A.7}$$

Moreover, since $\theta_{ij} \geqslant 0$ for all $i,j = 1,\ldots,d$, by Lemma A.3, clearly, we have

$$\log \mathbb{E}_0 \exp(X^T \Theta X/2) \geqslant 0. \tag{A.8}$$

Plugging (A.7) and (A.8) into (A.6), we obtain

$$\log \mathbb{E}_\Theta \exp\left(\frac{\|\mathrm{A}_H\|_F}{8} \cdot W_H\right) \leqslant \frac{9}{16}.$$

Therefore, by [**56**, (5.16)] as an equivalent definition of ψ_1-norm, we have $\|W_H\|_{\psi_1} \leqslant C|E_H|^{-1/2}$ for an absolute constant C. □

A.3 *Computational lower bound*

Proof of Lemma 4.1. Suppose that $\boldsymbol{V}$ is generated according to the procedure described in Lemma 4.1. Then, we can calculate

$$\begin{aligned}
&\mathbb{P}(V_1 = 1,\ldots,V_k = 1, V_{k+1} = -1,\ldots,V_s = -1)\\
&\quad = \int_{-\infty}^{\infty} \frac{\exp(\sqrt{2\theta}y(2k-s))}{2^s \cosh(\sqrt{2\theta}y)^s} \frac{\cosh(\sqrt{2\theta}y)^s \exp(-y^2/2)}{Z_{Y'}}\,\mathrm{d}y\\
&\quad = \frac{\sqrt{2\pi}}{2^s Z_{Y'}} \exp(\theta(2k-s)^2),
\end{aligned}$$

where $Z_{Y'}$ is the normalizing constant of $p_{Y'}(y)$. This is proportional precisely to $\exp(\theta(2k-s)^2)$ which is what the Curie–Weiss model is also proportional to. Moreover, from the above reasoning it follows that (4.2) is satisfied. With this the proof is complete. □

Proof of Lemma 7.1 The proof will begin from the last number and work its way through the chain of additions and subtractions. By induction, we will prove that $\sum_{k=l}^{m-1}(-1)^k\binom{2m-1}{2k+1}E_{2k+1}sP_{2m-2k-2}(s)$ can

be associated with $(-1)^l$ times the number of all sequences of the type $a_1 < a_2 > a_3 < \ldots < a_{2l} > a_{2l+1} > a_{2l+2} > a_{2l+3} > \ldots > a_{2l+2r+1}$ for $r = 0, \ldots, m-(l+1)$. The number s multiplying the summation indicates which number X_t we have selected from the first of the $2m$ brackets. The sequence $a_1 < a_2 > a_3 < \ldots < a_{2l} > a_{2l+1} > a_{2l+2} > a_{2l+3} > \ldots > a_{2l+2r+1}$ indicates the positions of the brackets that have been selected to choose X_t from. The remaining number $2m-2l-2r-2$ of the bracket positions have selected all possible X_{-t} elements so that they appear in an even power (without over-counting).

When $l = m-1$, we have $(-1)^{m-1}E_{2m-1}s$ can be simply equated to all sequences that form an alternating permutation on $2m-1$ numbers, so the induction step holds. Suppose now this is true for $l+1$. We will show it for l.

Consider the element $(-1)^l\binom{2m-1}{2l+1}E_{2l+1}sP_{2m-2l-2}(s)$. It consists of all sequences selecting brackets for X_t such that $a_1 < a_2 > a_3 < \ldots < a_{2l} > a_{2l+1}$ and $a_{2l+2} > a_{2l+3} > \ldots > a_{2l+2r+1}$ for $r = 0, \ldots, m-(l+1)$ (which come from $P_{2m-2l-2}(s)$; they are ordered in a decreasing manner since $P_{2m-2l-2}(s)$ counts bracket permutations for choosing X_t uniquely). By the induction step the previous summation consists of all sequences such that $a_1 < a_2 > a_3 < \ldots < a_{2l} > a_{2l+1} < a_{2l+2} > a_{2l+3} \ldots > a_{2l+2r+1}$ for $r = 0, \ldots, m-(l+1)$. Hence, after the subtraction all that remains are sequences for selecting brackets such that $a_1 < a_2 > a_3 < \ldots < a_{2l} > a_{2l+1} > a_{2l+2} > a_{2l+3} \ldots > a_{2l+2r+1}$ which is what we wanted to show.

It follows immediately from this discussion that (7.1) and (7.2) hold (since the number of sequences is multiplied by $(-1)^l$, and hence it will be negative with odd l and positive with even l). In addition, the last line (i.e., $l = 0$) selects brackets for X_t such that $a_1 > a_2 > \ldots > a_{2r+1}$ for $r = 0, \ldots, m-1$, and hence it does not double count summations. With this the proof is complete. □

Proof of Lemma 7.2. We will only show the right-hand side inequality with the other one being similar (however, in the induction step, we will assume it holds in both directions). We will use induction in m. For $m = 1$, we have $P_2(s) = s+0$ and hence the inequality is true by default in both directions. Suppose now it is true for $m \leqslant r$, we will show it for $m = r+1$. WLOG assume that $2l < m = r+1$ (if $2l = m = r+1$ was we trivially have equality). Using Lemma 7.1 and subtracting off $\sum_{k=0}^{2l} a_k^{2(r+1)} s^{r+1-k}$, we obtain

$$P_{2(r+1)}(s) - \sum_{k=0}^{2l} a_k^{2(r+1)} s^{r+1-k} = \sum_{k=0}^{r} (-1)^k \binom{2r+1}{2k+1} E_{2k+1} s P_{2r-2k}(s) - \sum_{k=0}^{2l} a_k^{2(r+1)} s^{r+1-k}.$$

Next, observe that this subtraction will erase the first $2l+1$ coefficients in the term multiplied by $sP_{2r}(s)$ and the first $2l$ coefficients in the term multiplied by $sP_{2r-2}(s)$ and so on. Now, by the induction hypothesis we know that $(2r+1)E_1 sP_{2r}(s)$ minus its $2l+1$ coefficients is non-positive, while $\binom{2r+1}{3}E_3 sP_{2r-2}(s)$ minus its first $2l$ coefficients is non-negative (hence $-s\binom{2r+1}{3}E_3 P_{2r-2}(s)$ is non-positive) and so on. It remains to notice that the remaining part of the summation which is not affected by the polynomial subtraction is non-positive by (7.1) of Lemma 7.1. This completes the proof. □

Proof of Lemma 7.3. This lemma simply compares the coefficients of the polynomials given in Lemma 7.1 and can be verified by a direct calculation. We omit the details. □

Proof of Lemma 7.4. The proof of this fact is by induction. For $m = 3$, it suffices to use $C \geqslant 16/360$ so that $a_2^6 = 16 \leqslant C(3-2)(3-1)3(3+1)(6)!! = C360$. Suppose that $a_2^{2m-2} \leqslant C(m-3)(m-2)(m-1)m(2m-2)!!$. We have that $P_3(m) := (m-2)(m-1)m(m+1) - (m-3)(m-2)(m-1)m =$

$4(m-2)(m-1)m$ is a polynomial of degree 3. We have

$$\begin{aligned}(2m-1)a_2^{2m-2} &+ 2\binom{2m-1}{3}(m-2)(m-3)\frac{(2m-4)!!}{3} + 16\binom{2m-1}{5}(2m-6)!!\\ &\leqslant C(m-2)(m-1)m(m+1)(2m)!! - CP_3(m)(2m)!!\\ &+ 2\binom{2m-1}{3}(m-2)(m-3)\frac{(2m-4)!!}{3} + 16\binom{2m-1}{5}(2m-6)!!\end{aligned}$$

Note that when C is sufficiently large, $CP_3(m)(2m-1)(2m-3)$ will dominate the remaining polynomial in m which is of degree 5 for all $m \geqslant 3$ which completes the proof. □

Below, we prove several of the inequalities given in the paper. We start with

$$f(x) = x^2/2 - \log(\cosh(x)) - x^4/12 \leqslant 0.$$

These are the first two terms in the Taylor expansion of $\log(\cosh(x))$. We will show that the first derivative of f is decreasing, with a unique 0 at 0. This will imply that 0 is a point of maximum of f and since $f(0) = 0$ this will complete the proof. Note that since the terms match the Taylor expansion, we have that $f'(0) = 0$.

To see that $d/dxf(x)$ is decreasing, we will look at the second derivative of $f(x)$. Direct calculation verifies that

$$\frac{d^2}{dx^2}f(x) = -x^2 - 1/\cosh^2(x) + 1.$$

Using $\cosh(x) \leqslant \exp(x^2/2)$ (which can be verified by a Taylor expansion), we conclude that

$$\frac{d^2}{dx^2}f(x) \leqslant -x^2 - \exp(-x^2) + 1.$$

Now, we consider the function $g(y) = -y - \exp(-y) + 1$ for $y \geqslant 0$. Taking the derivative, we conclude $g'(y) = -1 + \exp(-y) \leqslant 0$ for $y \geqslant 0$ with equality when $y = 0$. Hence, when $y \geqslant 0$ it follows that $-y - \exp(-y) + 1 \leqslant 0$. Hence,

$$\frac{d^2}{dx^2}f(x) \leqslant -x^2 - \exp(-x^2) + 1 \leqslant 0,$$

which completes the proof.

Next, we will show that

$$f(x) = \log(\cosh(x)) - x^2/2 + x^4/12 - x^6/45 \leqslant 0,$$

which will show the second two inequalities that we used. The strategy is similar to the one above.

The second derivative of $f(x)$ is

$$d^2/dx^2 f(x) = -\frac{2}{3}x^4 + x^2 + \frac{1}{\cosh^2(x)} - 1.$$

Now, we use that

$$\begin{aligned}-\frac{2}{3}x^4+x^2+\frac{1}{\cosh^2(x)}-1 &= -\frac{2}{3}x^4+4x^2-3x^2+\frac{1}{\cosh^2(x)}-1\\ &= 12\frac{2}{3}(-\frac{x^4}{12}+\frac{x^2}{2})-3x^2+\frac{1}{\cosh^2(x)}-1\\ &\leqslant 8\log(\cosh(x))+\frac{1}{\cosh^2(x)}-1-3x^2,\end{aligned}$$

where we used the previous part that $\log(\cosh(x)) \geqslant x^2/2-x^4/12$. Now, evaluating the second derivative of $g(x) := 8\log(\cosh(x))+\frac{1}{\cosh^2(x)}-1-3x^2$, we obtain $d^2/dx^2 g(x) = -6\tanh^4(x) \leqslant 0$, so the function is concave, and in addition, its first derivative (which equals to $-6x-2\tanh(x)(\operatorname{sech}^2(x)-4)$) has a 0 at 0. Hence, we conclude that

$$d^2/dx^2 f(x) \leqslant 0,$$

with a 0 at 0. So that $f(x)$ is concave and since its first derivative $(-2/15x^5+x^3/3-x+\tanh(x))$ has a 0 at 0 we conclude that $f(x) \leqslant 0$.

Next, we will show the inequality for $\log \Phi(x)$. The first five terms are from the Taylor expansion of $\log \Phi(x)$. Clearly, then $\log \Phi(0)-\sum_{i=0}^{5} C_i 0^i - C0^6 = 0$ (where C_i are the constants from the main text). Furthermore, by increasing the constant C it is clear that the outside of some small interval $[-c_0, c_0]$ (c_0 depends on C) the function $\sum_{i=1}^{5} C_i x^i + Cx^6$ will dominate $\log \Phi(x)$. Next, by the choice of the constants C_i we have that $\frac{d}{dx}\log \Phi(x) - \sum_{i=1}^{5} iC_i x^{i-1} = C'x^5 + o(x^5)$ locally around 0. Hence, for a small enough c_0 we will have that $\frac{d}{dx}\log \Phi(x)-\sum_{i=1}^{5} iC_i x^{i-1}-6Cx^5 = (C'-6C)x^5+o(x^5)$, and hence the sign of the derivative will equal to $\operatorname{sign}(C'-6C)\operatorname{sign}(x^5) = -\operatorname{sign}(x)$ (the last identity holding for C large enough). We will therefore have that the function is increasing for x in $[-c_0, 0]$ and decreasing on $[0, c_0]$ which implies that 0 is a local maximum on $[-c_0, c_0]$ which shows that the inequality holds.

Next, we will show that $(2\Phi(x)-1)x^5 \leqslant \sqrt{\frac{2}{\pi}}x^6$. For $x \geqslant 0$, this is equivalent to $2\Phi(x)-1-\sqrt{\frac{2}{\pi}}x \leqslant 0$. The derivative of the function is $\sqrt{\frac{2}{\pi}}\exp(-x^2/2)-\sqrt{\frac{2}{\pi}} \leqslant 0$; hence, 0 is the maximum on $x \geqslant 0$ which completes the proof for the $x \geqslant 0$ case. Next, for $x < 0$, the above reasoning says that $2\Phi(x)-1-\sqrt{\frac{2}{\pi}}x \geqslant 2\Phi(0)-1-\sqrt{\frac{2}{\pi}}0 = 0$ which completes the proof.

Next, we will argue that $(1-2\Phi(x))x^3 \leqslant -\sqrt{\frac{2}{\pi}}x^4+\frac{x^6}{3\sqrt{2\pi}}$. For $x \geqslant 0$, this is equivalent to $(1-2\Phi(x)) \leqslant -\sqrt{\frac{2}{\pi}}x+\frac{x^3}{3\sqrt{2\pi}}$. Taking the derivative of $2\Phi(x)-1-\sqrt{\frac{2}{\pi}}x+\frac{x^3}{3\sqrt{2\pi}}$ results in $\sqrt{\frac{2}{\pi}}\exp(-x^2/2)-\sqrt{\frac{2}{\pi}}+\frac{x^2}{\sqrt{2\pi}} \geqslant 0$ since $\exp(-y) \geqslant 1-y$ for all $y \geqslant 0$. Hence, the minimum of $2\Phi(x)-1-\sqrt{\frac{2}{\pi}}x+\frac{x^3}{3\sqrt{2\pi}}$ for $x \geqslant 0$ is reached at 0 and therefore $2\Phi(x)-1-\sqrt{\frac{2}{\pi}}x+\frac{x^3}{3\sqrt{2\pi}} \geqslant 0$ for $x \geqslant 0$. For $x < 0$, we need to show $1-2\Phi(x) \geqslant -\sqrt{\frac{2}{\pi}}x+\frac{x^3}{3\sqrt{2\pi}}$. Taking the derivative of $2\Phi(x)-1-\sqrt{\frac{2}{\pi}}x+\frac{x^3}{3\sqrt{2\pi}}$ equals to $\sqrt{\frac{2}{\pi}}\exp(-x^2/2)-\sqrt{\frac{2}{\pi}}+\frac{x^2}{\sqrt{2\pi}} \geqslant 0$ as before hence for $x \leqslant 0$ we have $2\Phi(x)-1-\sqrt{\frac{2}{\pi}}x+\frac{x^3}{3\sqrt{2\pi}} \leqslant 2\Phi(0)-1-\sqrt{\frac{2}{\pi}}0+\frac{0^3}{3\sqrt{2\pi}} = 0$ which implies that $1-2\Phi(x) \geqslant -\sqrt{\frac{2}{\pi}}x+\frac{x^3}{3\sqrt{2\pi}}$ which is what we wanted to show.

B. Computational lower bound under oracle computational model

In this section, we consider an oracle computational model [21–23,25,38,58,60], based on which we derive another computational lower bound result for detection problems in Ising model. The main idea of oracle computational model is to use the number of rounds of interactions between data and a certain algorithm to represent the algorithmic complexity of this algorithm. In specific, let $\boldsymbol{X}$ be the random vector of interest and $\mathscr{X}$ be the domain of $\boldsymbol{X}$. We define

$$\mathscr{Q}^* = \{q : q(\boldsymbol{X}) \text{ is a sub-exponential variable}\}. \tag{B.1}$$

We call every subset $\mathscr{Q} \subseteq \mathscr{Q}^*$ a query space. Next, we define the statistical query oracle.

Definition B.1 (Statistical query oracle). Let n be the sample size of a testing problem. A statistical query oracle r_n on a query space $\mathscr{Q} \subseteq \mathscr{Q}^*$ is a random mapping from $\mathscr{Q}$ to $\mathbb{R}$. Given a query $q \in \mathscr{Q}$, the oracle r_n returns an output $Z_q \in \mathbb{R}$, such that for any tail probability $\xi \in [0, 1)$,

$$\mathbb{P}\left(\bigcap_{q\in\mathscr{Q}} \left\{|Z_q - \mathbb{E}[q(\boldsymbol{X})]| \leqslant \|q(\boldsymbol{X})\|_{\psi_1} \cdot \tau\right\}\right) \geqslant 1 - 2\xi, \text{where}$$

$$\tau = \max\left\{\frac{\eta(\mathscr{Q}) + \log(1/\xi)}{n}, \sqrt{\frac{2[\eta(\mathscr{Q}) + \log(1/\xi)]}{n}}\right\}. \tag{B.2}$$

Here, we call $\eta(\mathscr{Q}) > 0$ the capacity measure of $\mathscr{Q}$. When $\mathscr{Q}$ is finite, we define $\eta(\mathscr{Q}) = \log(|\mathscr{Q}|)$.

Given a query space $\mathscr{Q} \subseteq \mathscr{Q}^*$, we define $R_n(\mathscr{Q})$ to be the set of all statical query oracles on $\mathscr{Q}$ with sample size n. We now give the definition of oracle computational model.

Definition B.2 (Oracle computational model). An oracle computational model Ψ is defined as a tuple $\Psi = \Psi(\mathscr{Q}_\Psi, T_\Psi, q_{\text{init}}, \{\delta_t\}_{t=1}^{T_\Psi}, \psi)$, where

- $\mathscr{Q}_\Psi$ is a subset of $\mathscr{Q}^*$ that contains all queries the test will potentially use;
- T_Ψ is the maximum number of rounds the model queries an oracle;
- $q_{\text{init}} \in \mathscr{Q}_\Psi$ is the initial query;
- $\delta_t : (\mathscr{Q}_\Psi \times \mathbb{R})^{t-1} \to \mathscr{Q}_\Psi \cup \{\text{HALT}\}$ is the transition function at the tth round. If δ_t returns HALT, then the model stops querying the oracle;
- $\psi : (\mathscr{Q}_\Psi \times \mathbb{R})^{T_\Psi} \to \{0, 1\}$ is the test function that takes the results of at most T_Ψ queries as input and returns the test result as binary output.

Each instance of $\Psi(\mathscr{Q}_\Psi, T_\Psi, q_{\text{init}}, \{\delta_t\}_{t=1}^{T_\Psi}, \psi)$ refers to a test algorithm. The parameter T_Ψ is the query complexity of algorithm Ψ. We define $\mathscr{A}(T) = \{\Psi : T_\Psi \leqslant T\}$ to be the set of all algorithms with query complexity at most T. Under oracle computational model, the risk of detection problem (1.5) with maximum query complexity T is defined as

$$\gamma_{\text{oracle}}\{\mathscr{S}[\mathscr{G}_1(G_*), \theta]\} = \inf_{\Psi\in\mathscr{A}(T)} \sup_{r_n\in R_n(\mathscr{Q}_\Psi)} \left\{\mathbb{P}_0(\psi = 1) + \max_{\Theta\in\mathscr{S}[\mathscr{G}_1(G_*),\theta]} \mathbb{P}_\Theta(\psi = 0)\right\}. \tag{B.3}$$

Note that in (B.3), the supreme over $r \in R_n(\mathscr{Q}_\Psi)$ implies that we consider the worst oracle. If $\liminf_{n\to\infty} \gamma_{\text{oracle}}\{\mathscr{S}[\mathscr{G}_1(G_*), \theta]\} = 1$, then when n is large enough, for any algorithm that queries at most T rounds, there exists an oracle r_n such that the algorithm cannot distinguish the null and alternative hypotheses. We now give our main result.

THEOREM B.3 Let G_* be a graph with s vertices. Under the statistical query model, if $T \leqslant d^p$ for some constant $p > 0$, $s \leqslant d^{(1-\eta)/2}$ for some constant $\eta > 0$ and

$$\theta \leqslant \kappa \sqrt{\frac{1}{n}} \wedge \frac{1}{16s}, \tag{B.4}$$

where κ is some sufficiently small positive constant, then

$$\liminf_{n\to\infty} \gamma_{\text{oracle}}\{\mathscr{S}[\mathscr{G}_1(G_*), \theta]\} = 1.$$

Proof of Theorem B.3. We denote by $G_\emptyset$ the empty graph, Similar to the computational lower bound analysis in Section 4, we only need to consider the case where G_* is an s-clique. Therefore, we set $\mathscr{G}^*$ to be the set of graphs isomorphic to G_*, and let $S^* = \{\theta \mathrm{A}_G : G \in \mathscr{G}^*\}$. Each parameter matrix $\Theta \in S^*$ can be represented by a graph $G \in \mathscr{G}^*$. In the following, we always denote by Θ the parameter matrix with underlying graph G, and by Θ', the parameter matrix with underlying graph G'. For a graph G, in order to successfully detect it with the worst-case oracle, a test has to utilize at least one query q that can distinguish G from $G_\emptyset$. We define

$$\mathscr{G}(q) = \{G \in \mathscr{G}^* : |\mathbb{E}_\Theta q(\boldsymbol{X}) - \mathbb{E}_0 q(\boldsymbol{X})| \geqslant \|q(\boldsymbol{X})\|_{\psi_1,0} \cdot \tau\},$$

where $\|q(\boldsymbol{X})\|_{\psi_1,0}$ is the ψ_1-norm of $q(\boldsymbol{X})$ when $\boldsymbol{X}$ follows the distribution $\mathbb{P}_0$ and τ is defined in Definition B.1. By the definition of $\mathscr{G}(q)$, if $T \cdot \sup_{q\in\mathscr{Q}_\Psi} |\mathscr{G}(q)| < |\mathscr{G}^*|$, then there must be some $G' \in \mathscr{G}^*$ such that none of the T queries used by the test can distinguish G from $G_\emptyset$. Therefore, the worst-case oracle that returns $\mathbb{E}_{\Theta'} q(\boldsymbol{X})$ when $\boldsymbol{X} \sim \mathbb{P}_0$ can still satisfy Definition B.1 but will make all the tests powerless. This gives the following lemma.

LEMMA B.1 For any algorithm Ψ that queries the oracle at most T rounds, if $T \cdot \sup_{q\in\mathscr{Q}_\Psi} |\mathscr{G}(q)| < |\mathscr{G}^*|$, then there exists an oracle $r_n \in R_n(\mathscr{Q}_\Psi)$ defined in Definition B.1 such that $\liminf_{n\to\infty} \gamma_{\text{oracle}}(S^*) \geqslant 1$.

Proof. See Section B.1 for a detailed proof. □

By Lemma B.1, to prove $\liminf_{n\to\infty} \gamma_{\text{oracle}}\{\mathscr{S}[\mathscr{G}_1(G_*), \theta]\} = 1$, it suffices to show that $T \cdot \sup_{q\in\mathscr{Q}_\Psi} |\mathscr{G}(q)|/|\mathscr{G}^*|$ is asymptotically smaller than one. In the rest of the proof, for any $q \in \mathscr{Q}_\Psi$, we derive an upper bound on $|\mathscr{G}(q)|$. To do so, we first split $\mathscr{G}(q)$ into two subsets $\mathscr{G}^+(q)$ and $\mathscr{G}^-(q)$, which are given by

$$\mathscr{G}^+(q) = \{G \in \mathscr{G}^* : \mathbb{E}_\Theta[q(\boldsymbol{X})] - \mathbb{E}_0[q(\boldsymbol{X})] > \|q(\boldsymbol{X})\|_{\psi_1,0} \cdot \tau\}, \tag{B.5}$$

$$\mathscr{G}^-(q) = \{G \in \mathscr{G}^* : \mathbb{E}_0[q(\boldsymbol{X})] - \mathbb{E}_\Theta[q(\boldsymbol{X})] > \|q(\boldsymbol{X})\|_{\psi_1,0} \cdot \tau\}. \tag{B.6}$$

We now bound $|\mathscr{G}^+(q)|$. $|\mathscr{G}^-(q)|$ can be bounded in exactly the same way. The following lemma summarizes an inequality derived from the definition (B.5).

LEMMA B.2 For any query function q,

$$\frac{1}{|\mathscr{G}^+(q)|^2}\sum_{G,G'\in\mathscr{G}^+(q)}\mathbb{E}_0\left[\frac{\mathrm{d}\mathbb{P}_\Theta}{\mathrm{d}\mathbb{P}_0}\frac{\mathrm{d}\mathbb{P}_{\Theta'}}{\mathrm{d}\mathbb{P}_0}\right] > 1+\frac{1}{n}. \tag{B.7}$$

Proof. See Section B.1 for a detailed proof. □

It remains to calculate the left-hand side of (B.7). By Lemma 5.4, we have

$$\begin{aligned}\mathbb{E}_0\left[\frac{\mathbb{P}_\Theta}{\mathbb{P}_0}\frac{\mathbb{P}_{\Theta'}}{\mathbb{P}_0}\right] &\leqslant 1+|E(G)\cap E(G')|\theta^2+\Delta_{G,G'}\theta^3\\ &\quad+\sum_{k\geqslant 4}q_k[G\oplus G', V(G)\cap V(G')]\theta^k.\end{aligned}$$

For $|E(G)\cap E(G')|$, we use the trivial bound that $|E(G)\cap E(G')|\leqslant |V(G)\cap V(G')|^2/2$. For $q_k[G\oplus G', V(G)\cap V(G')]$, $k\geqslant 4$, we apply the bound given by Lemma 5.5 and obtain

$$\begin{aligned}q_k[G\oplus G', V(G)\cap V(G')] &\leqslant k\cdot 2^{k-2}\cdot|V(G)\cap V(G')|^2\cdot(\|\mathrm{A}_{G\oplus G'}\|_1\vee\|\mathrm{A}_{G\oplus G'}\|_F)^{k-2}\\ &\leqslant k\cdot 2^{k-2}\cdot|V(G)\cap V(G')|^2\cdot(2s)^{k-2}\\ &\leqslant 2^{k-2}\cdot 2^{k-2}\cdot|V(G)\cap V(G')|^2\cdot(2s)^{k-2}\\ &= 8^{k-2}\cdot s^{k-2}\cdot|V(G)\cap V(G')|^2.\end{aligned}$$

Therefore, by the assumption that $\theta\leqslant(16s)^{-1}$, we have

$$\begin{aligned}\sum_{k\geqslant 4}q_k[G\oplus G', V(G)\cap V(G')]\theta^k &\leqslant 64|V(G)\cap V(G')|^2s^2\theta^4\\ &\leqslant |V(G)\cap V(G')|^2\theta^2/4.\end{aligned}$$

For $\Delta_{G,G'}$, we use a bound similar to Lemma 5.5 but more specific for cliques. If a triangle has one edge in $E(G)$ and two edges in $E(G')$, then the two vertices of the edge in $E(G)$ must be in $V(G)\cap V(G')$. Therefore, an upper bound of the number of triangles that have one edge in $E(G)$ and two edges in $E(G')$ is given by the following procedure:

- pick an edge e from $E[G_{V(G)\cap V(G')}]$;
- pick a common neighbor of the two vertices of edge e.

Therefore, by the trivial bound $\left|E[G_{V(G)\cap V(G')}]\right|, \left|E[G'_{V(G)\cap V(G')}]\right|\leqslant|V(G)\cap V(G')|^2/2$, we have

$$\begin{aligned}\Delta_{G,G'} &\leqslant \left|E[G_{V(G)\cap V(G')}]\right|\cdot\|\mathrm{A}_{G'}\|_1+\left|E[G'_{V(G)\cap V(G')}]\right|\cdot\|\mathrm{A}_G\|_1\\ &\leqslant |V(G)\cap V(G')|^2\cdot s.\end{aligned}$$

Therefore, we have

$$\begin{aligned}\mathbb{E}_0\left[\frac{\mathbb{P}_{\Theta}}{\mathbb{P}_0}\frac{\mathbb{P}_{\Theta'}}{\mathbb{P}_0}\right] &\leqslant 1+|V(G)\cap V(G')|^2\theta^2/2+|V(G)\cap V(G')|^2\cdot s\cdot\theta^3\\ &\quad+|V(G)\cap V(G')|^2\theta^2/4\\ &\leqslant 1+|V(G)\cap V(G')|^2\theta^2.\end{aligned}$$

Denote by $U[\mathscr{G}^+(q)]$ uniformly choosing a graph in $\mathscr{G}^+(q)$. Then, by Lemma B.2, we get

$$\begin{aligned}\frac{1}{n} &< \frac{1}{|\mathscr{G}^+(q)|^2}\sum_{G,G'\in\mathscr{G}^+(q)}|V(G)\cap V(G')|^2\theta^2\\ &\leqslant \theta^2\cdot\sup_{G\in\mathscr{G}^*}\mathbb{E}_{G'\sim U[\mathscr{G}^+(q)]}|V(G)\cap V(G')|^2.\end{aligned}\tag{B.8}$$

Equation (B.8) gives an lower bound of the expectation defined on the right-hand side. In the following, we utilize this lower bound to derive an upper bound of $|\mathscr{G}^+(q)|$. Inspired my similar results given in [20,42], we give the following lemma.

LEMMA B.3 For $j=0,\ldots,s$, define $m_j=\max_{G\in\mathscr{G}^*}|\{G'\in\mathscr{G}^*:|V(G)\cap V(G')|=s-j\}|$. For $k\leqslant|\mathscr{G}^*|$, let $\mathscr{G}(k)=\{\mathscr{G}\subseteq\mathscr{G}^*:|\mathscr{G}|=k\}$ and $l(k)=\max\{r\leqslant s:\sum_{j=0}^{r}m_j\leqslant k\}$. Then, we have

$$\sup_{G\in\mathscr{G}^*}\sup_{\mathscr{G}\in\mathscr{G}(k)}\mathbb{E}_{G'\sim U(\mathscr{G})}|V(G)\cap V(G')|^2\leqslant\frac{\sum_{j=0}^{l(k)}(s-j)^2m_j}{\sum_{j=0}^{l(k)}m_j}.$$

The intuition of Lemma B.3 is that, among all sets of graphs $\mathscr{G}$ with cardinality k (i.e., sets of graphs $\mathscr{G}\in\mathscr{G}(k)$), the ones that maximize the expectation $\mathbb{E}_{G'\sim U(\mathscr{G})}|V(G)\cap V(G')|^2$ consist of graphs that make $|V(G)\cap V(G')|^2$ as large as possible. Let $\zeta=\inf_{0\leqslant j\leqslant s-1}m_{j+1}/m_j$. Then, for clique detection problem we have

$$\zeta=\inf\frac{m_{j+1}}{m_j}=\inf\left[\binom{s}{s-j-1}\binom{d-s}{j+1}\right]\Big/\left[\binom{s}{s-j}\binom{d-s}{j}\right]\geqslant\frac{d}{s^2}\geqslant d^{\eta}.$$

Clearly, for large enough d we have $\zeta>2$. Let $h(j)=(s-j)^2$. Then, by assumption, for $i<j$, we have $m_i\zeta^j-m_j\zeta^i<0$, $h(i)-h(j)>0$ and therefore $(m_i\zeta^j-m_j\zeta^i)[h(i)-h(j)]\leqslant 0$. Similarly, for $i\geqslant j$ the same inequality $(m_i\zeta^j-m_j\zeta^i)[h(i)-h(j)]\leqslant 0$ still holds. Therefore, we have $\sum_{0\leqslant i,j\leqslant l(k)}(m_i\zeta^j-m_j\zeta^i)[h(i)-h(j)]\leqslant 0$. Rearranging terms gives

$$\frac{\sum_{j=0}^{l(k)}h(j)m_j}{\sum_{j=0}^{l(k)}m_j}\leqslant\frac{\sum_{j=0}^{l(k)}h(j)\zeta^j}{\sum_{j=0}^{l(k)}\zeta^j}=\frac{\sum_{j=0}^{l(k)}h(j)\zeta^{-(s-j)}}{\sum_{j=0}^{l(k)}\zeta^{-(s-j)}}.\tag{B.9}$$

We now bound the right-hand side of (B.9). Note that $\zeta^{-1}\leqslant 1/8$ for large enough d. For the numerator, we have

$$\sum_{j=0}^{l(k)}h(j)\zeta^{-(s-j)}=\sum_{i=s-l(k)}^{s}i^2\zeta^{-i}\leqslant[s-l(k)]^2\zeta^{-[s-l(k)]}+\sum_{i=s-l(k)+1}^{s}i^2\zeta^{-i}.$$

Since $s - l(k) + 1 \geqslant 1$, for $i \geqslant s - l(k) + 1$, we have $i^2 \leqslant [s - l(k) + 1]^2 4^{i-s+l(k)-1}$. Therefore,

$$\begin{aligned}\sum_{j=0}^{l(k)} h(j)\zeta^{-(s-j)} &\leqslant [s - l(k) + 1]^2 \zeta^{-[s-l(k)+1]} \cdot \sum_{i=s-l(k)+1}^{s} (4\zeta^{-1})^{i-s+l(k)-1} \\ &\quad + [s - l(k)]^2 \zeta^{-[s-l(k)]} \\ &\leqslant 2[s - l(k) + 1]^2 \zeta^{-[s-l(k)+1]} + [s - l(k)]^2 \zeta^{-[s-l(k)]} \\ &\leqslant 2[s - l(k) + 1]^2 \zeta^{-[s-l(k)]}.\end{aligned}$$

For the denominator of the right-hand side of (B.9), we have $\sum_{j=0}^{l(k)} \zeta^{-(s-j)} \geqslant \zeta^{-[s-l(k)]}$. Therefore, we have

$$\frac{\sum_{j=0}^{l(k)} h(j)\zeta^{-(s-j)}}{\sum_{j=0}^{l(k)} \zeta^{-(s-j)}} \leqslant 2[s - l(k) + 1]^2. \tag{B.10}$$

By (B.10), (B.8) and Lemma B.3, for $k = |\mathscr{G}^+(q)|$, we have

$$2[s - l(k) + 1]^2 \geqslant \frac{1}{n}.$$

Therefore, for large enough d, we have

$$s - l(k) \geqslant \sqrt{\frac{1}{2\theta^2 n}} - 1. \tag{B.11}$$

On the other hand, by the definition of $l(k)$, we have

$$|\mathscr{G}^+(q)| \leqslant \sum_{j=0}^{l(k)+1} m_j \leqslant m_s \cdot \sum_{j=0}^{l(k)+1} \zeta^{j-s} \leqslant \frac{\zeta^{-[s-l(k)-1]}|\mathscr{G}^*|}{1 - \zeta^{-1}} \leqslant 2\zeta^{-[s-l(k)-1]}|\mathscr{G}^*|, \tag{B.12}$$

where the last inequality follows from the fact that $\zeta^{-1} \leqslant 1/2$ for large enough d. Plugging (B.11) into (B.12) gives

$$|\mathscr{G}^+(q)| \leqslant 2|\mathscr{G}^*| \exp\left[-\log(\zeta) \cdot \left(\sqrt{\frac{1}{2\theta^2 n}} - 2\right)\right].$$

Applying the same analysis to $|\mathscr{G}^-(q)|$, we obtain

$$|\mathscr{G}^-(q)| \leqslant 2|\mathscr{G}^*| \exp\left[-\log(\zeta) \cdot \left(\sqrt{\frac{1}{2\theta^2 n}} - 2\right)\right].$$

Therefore, we have

$$|\mathscr{G}(q)| \leqslant 4|\mathscr{G}^*| \exp\left[-\log(\zeta) \cdot \left(\sqrt{\frac{1}{2\theta^2 n}} - 2\right)\right].$$

Since the inequality above holds for all $q \in \mathscr{Q}_\Psi$, we have

$$T \cdot \frac{\sup_{q\in\mathscr{Q}_\Psi} |\mathscr{G}(q)|}{|\mathscr{G}^*|} \leqslant 4 \exp\left[\log(T) - \left(\sqrt{\frac{1}{2\theta^2 n}} - 2\right) \cdot \log \zeta\right].$$

If $T \leqslant d^p$, then

$$T \cdot \frac{\sup_{q\in\mathscr{Q}_\Psi} |\mathscr{G}(q)|}{|\mathscr{G}^*|} \leqslant \exp\left[\log(4) + p\log(d) - \left(\sqrt{\frac{1}{2\theta^2 n}} - 2\right) \cdot \log \zeta\right].$$

Let $\kappa < [\sqrt{2}(2 + p/\eta)]^{-1}$. Then, if $\theta \leqslant \kappa\sqrt{\frac{1}{n}}$, for large enough d, we have

$$\begin{aligned}
&\log(4) + p\log(d) - \left(\sqrt{\frac{1}{2\theta^2 n}} - 2\right) \cdot \log \zeta \\
&\quad \leqslant \log(4) + p\log(d) - \eta\left(\sqrt{\frac{1}{2\theta^2 n}} - 2\right) \log d \leqslant -1,
\end{aligned}$$

and therefore, $T \cdot \sup_{q\in\mathscr{Q}_\Psi} |\mathscr{G}(q)|/|\mathscr{G}^*| < 1$. By Lemma B.1, there exists an oracle r such that $\liminf_{n\to\infty} \gamma_{\text{oracle}}(S^*) \geqslant 1$. This completes the proof. □

B.1 *Proofs of auxiliary lemmas*

Proof of Lemma B.1. We consider an algorithm Ψ with query space $\mathscr{Q}_\Psi$ and $T_\Psi = T$. If $T \cdot \sup_{q\in\mathscr{Q}_\Psi} |\mathscr{G}(q)| < |\mathscr{G}^*|$, then for any T queries $q_1, \ldots, q_T \in \mathscr{Q}_\Psi$, there exists $G_0 \in \mathscr{G} \setminus \bigcup_{t=1}^T \mathscr{G}(q_t)$. Let $\Theta_0 = \theta \mathrm{A}_{G_0}$ be the parameter matrix with underlying graph G_0. Then, by definition, for $t = 1, \ldots, T$, we have

$$|\mathbb{E}_{\Theta_0} q_t(\boldsymbol{X}) - \mathbb{E}_0 q_t(\boldsymbol{X})| \leqslant \|q_t(\boldsymbol{X})\|_{\psi_1,0} \cdot \tau.$$

We set r to be the oracle that returns Z_{q_t} such that

$$\begin{aligned}
&\mathbb{P}_0(Z_{q_t} = \mathbb{E}_{\Theta_0}[q_t(\boldsymbol{X})]) = 1, \\
&\mathbb{P}_\Theta(Z_{q_t} = \mathbb{E}_\Theta[q_t(\boldsymbol{X})]) = 1,\ G \in \mathscr{G}_1.
\end{aligned}$$

Then, clearly,

$$\mathbb{P}_0(|Z_{q_t} - \mathbb{E}_{\Theta_0}[q_t(\boldsymbol{X})]| \leqslant \|q_t(\boldsymbol{X})\|_{\psi_1,0} \cdot \tau_{q_t}) = 1,$$

and hence r satisfies Definition B.2. However, for $t = 1, \ldots, T$, the oracle always returns the same Z_{q_t} under $\mathbb{P}_0$ and $\mathbb{P}_{\Theta_0}$. Therefore, we have

$$\mathbb{P}_0(\psi = 1) + \mathbb{P}_{\Theta_0}(\psi = 0) = 1.$$

This completes the proof. □

Proof of Lemma B.2. By (B.5), we have

$$\|q(\boldsymbol{X})\|_{\psi_1,0}\cdot\tau < \frac{1}{|\mathscr{G}^+(q)|}\sum_{G\in\mathscr{G}^+(q)}\left\{\mathbb{E}_{\Theta}[q(\boldsymbol{X})]-\mathbb{E}_0[q(\boldsymbol{X})]\right\}$$
$$=\mathbb{E}_0\left\{q(\boldsymbol{X})\cdot\frac{1}{|\mathscr{G}^+(q)|}\sum_{G\in\mathscr{G}^+(q)}\left[\frac{\mathrm{d}\mathbb{P}_{\Theta}}{\mathrm{d}\mathbb{P}_0}(\boldsymbol{X})-1\right]\right\}.$$

Applying Cauch–Schwartz inequality on the right-hand side above gives

$$\|q(\boldsymbol{X})\|_{\psi_1,0}\cdot\tau < \underbrace{\{\mathbb{E}_0[q^2(\boldsymbol{X})]\}^{1/2}}_{\text{(i)}}\cdot\underbrace{\left(\mathbb{E}_0\left\{\frac{1}{|\mathscr{G}^+(q)|}\sum_{G\in\mathscr{G}^+(q)}\left[\frac{\mathrm{d}\mathbb{P}_{\Theta}}{\mathrm{d}\mathbb{P}_0}(\boldsymbol{X})-1\right]\right\}^2\right)^{1/2}}_{\text{(ii)}}. \tag{B.13}$$

For term (i), by the definition of ψ_1-norm, we have

$$(\mathbb{E}_0\{[q(\boldsymbol{X})]^2\})^{1/2}\leqslant 2\|q(\boldsymbol{X})\|_{\psi_1,0}. \tag{B.14}$$

For term (ii), we have

$$\left[\mathbb{E}_0\left(\left\{\frac{1}{|\mathscr{G}^+(q)|}\sum_{G\in\mathscr{G}^+(q)}\left[\frac{\mathrm{d}\mathbb{P}_{\Theta}}{\mathrm{d}\mathbb{P}_0}(\boldsymbol{X})-1\right]\right\}^2\right)\right]^{1/2}$$
$$=\left(\frac{1}{|\mathscr{G}^+(q)|^2}\sum_{G,G'\in\mathscr{G}^+(q)}\mathbb{E}_0\left\{\left[\frac{\mathrm{d}\mathbb{P}_{\Theta}}{\mathrm{d}\mathbb{P}_0}(\boldsymbol{X})-1\right]\cdot\left[\frac{\mathrm{d}\mathbb{P}_{\Theta'}}{\mathrm{d}\mathbb{P}_0}(\boldsymbol{X})-1\right]\right\}\right)^{1/2}$$
$$=\left\{\frac{1}{|\mathscr{G}^+(q)|^2}\sum_{G,G'\in\mathscr{G}^+(q)}\mathbb{E}_0\left[\frac{\mathrm{d}\mathbb{P}_{\Theta}}{\mathrm{d}\mathbb{P}_0}\frac{\mathrm{d}\mathbb{P}_{\Theta'}}{\mathrm{d}\mathbb{P}_0}(\boldsymbol{X})\right]-1\right\}^{1/2}. \tag{B.15}$$

Plugging (B.14) and (B.15) into (B.13) and using the bound $\tau\geqslant\sqrt{\frac{1}{n}}$, we obtain

$$\frac{1}{|\mathscr{G}^+(q)|^2}\sum_{G,G'\in\mathscr{G}^+(q)}\mathbb{E}_0\left[\frac{\mathrm{d}\mathbb{P}_{\Theta}}{\mathrm{d}\mathbb{P}_0}\frac{\mathrm{d}\mathbb{P}_{\Theta'}}{\mathrm{d}\mathbb{P}_0}\right]>1+\frac{1}{n}.$$

Therefore, we conclude the proof. □

Proof of Lemma B.3. For any $G\in\mathscr{G}^*$, we have

$$\mathbb{E}_{G'\sim U(\mathscr{G})}|V(G)\cap V(G')|^2=\sum_{j=0}^{s}(s-j)^2|\{G'\in\mathscr{G}:|V(G)\cap V(G')|=s-j\}|.$$

We define

$$\overline{m} = k - \sum_{j=0}^{l(k)} m_j.$$

Note that $h(j) := (s-j)^2$ is a decreasing function of j and $\sum_{G' \in \mathscr{G}} |V(G) \cap V(G')|^2$ is a sum of $m_1 + \ldots + m_{l(k)} + \overline{m} = k$ terms, with at most m_j terms being $h(j)$. Therefore, by (B.8), we have

$$\begin{aligned}\sup_{\mathscr{G} \in \mathscr{G}(k)} \mathbb{E}_{G' \sim U(\mathscr{G})} |V(G) \cap V(G')|^2 &\leqslant \frac{\sum_{j=0}^{l(k)} h(j) \cdot m_j + h[l(k)+1] \cdot \overline{m}}{\sum_{j=0}^{l(k)} m_j + \overline{m}} \\ &\leqslant \frac{\sum_{j=0}^{l(k)} h(j) \cdot m_j}{\sum_{j=0}^{l(k)} m_j}.\end{aligned}$$

This finishes the proof. □