Computational-Statistical Tradeoffs
in Inferring Combinatorial Structures of Ising Model

Ying Jin [1]   Zhaoran Wang [2]   Junwei Lu [3]

Abstract

We study the computational and statistical trade-offs in inferring combinatorial structures of high dimensional simple zero-field ferromagnetic Ising model. Under the framework of oracle computational model where an algorithm interacts with an oracle that discourses a randomized version of truth, we characterize the computational lower bounds of learning combinatorial structures in polynomial-time, under which no algorithms within polynomial-time can distinguish between graphs with and without certain structures. This hardness of learning with limited computational budget is shown to be characterized by a novel quantity called vertex overlap ratio. Such quantity is universally valid for many specific graph structures including cliques and nearest neighbors. On the other side, we attain the optimal rates for testing these structures against empty graph by proposing the quadratic testing statistics to match the lower bounds. We also investigate the relationship between computational bounds and information-theoretic bounds for such problems, and found gaps between the two boundaries in inferring some particular structures, especially for those with dense edges.

1. Introduction

In various problems, data are presented and interpreted in the form of a graph $G = (V, E)$, where we observe features X_v for each vertex $v \in V$ representing an individual, and the inter-dependency among them are encoded by the graph edges in E. For example, in bioinformatics, (Fried-

---

[1]Department of Statistics, Stanford University, Stanford, CA, USA [2]Department of Electrical Engineering and Computer Science, Northwestern University, Evanston, IL, USA [3]Department of Biostatistics, Harvard University, Boston, MA, USA. Correspondence to: Junwei Lu <junweilu@hsph.harvard.edu>.

man, 2004) studies relationships among observed cellulars using graphical models, and in social network analysis like (Lazega et al., 1995), (Grabowski & Kosiński, 2006), relationships of people are encoded by a graph. To represent probablity structures of observations, or features of the individuals (vertices) in graphical models, Markov random field (MRF) family is widely used, among which Gaussian graphical model and Ising model are the most popular.

The main focus of this paper is to test whether the underlying graph has a certain structure property. There have been a lot of work on algorithms and information-theoretic limits for such structural inference for graphical model, especially for Gaussian graphical model and Ising model, which are fundamental models in the Markov random field family. See (Louigi et al., 2010), (Ery & Verzelen, 2014), (Bresler et al., 2014), (Neykov et al., 2016), (Neykov & Liu, 2017). However, to the best of our knowledge, it remains unclear whether it is possible and how to achieve the information-theoretic limits by efficient algorithms with certain computational budgets under the Ising model. Due to the discrete nature of the Ising model, the theoretical analysis is more challenging than the Gaussian distribution.

Our paper aims to solve this problem from two major perspectives: (1) the theoretical gap between the computational and statistical rates for recovering various combinatorial structures in Ising model; and (2) polynomial-time algorithms to detect these structures efficiently. When considering the computational budgets, we employ the computational oracle model developed and explored in (Kearns, 1998), (Feldman et al., 2013), (Wang et al., 2015), etc.

Related work: Our work follows the nature of recent work on computational-statistical tradeoffs for various problems like PCA, sparse linear regression, etc., see (Feldman et al., 2013), (Hajek et al., 2014), (Zhang et al., 2014), (Chen & Xu, 2014), (Wang et al., 2016), (Wang et al., 2015), (Cai et al., 2017), (Fan et al., 2018). In comparison, we focus on a family of structural property testing problems in undirected graphical models. (Neykov et al., 2016) provides the information-theoretic limits while (Lu et al., 2018) provides the computation-efficient limits for Gaussian graphical model. But for Ising model, another fundamental model in the family of Markov random field, there are only results re-

lated to information-theoretic bounds for structure property inference, see (Neykov & Liu, 2017) and (Daskalakis et al., 2016).

To characterize the computational complexity, our analysis in this paper is carried out under the framework of oracle computational model, which is proposed by (Kearns, 1998) and generalized by (Feldman et al., 2013), (Wang et al., 2015), (Fan et al., 2018). This general framework captures the computational properties of a wide range of algorithms including moments-based methods, stochastic optimization methods, local search, Markov chain Monte Carlo and most other learning alrogithms. The same oracle computational model is adopted in (Lu et al., 2018), but it focuses on computational efficiency of Gaussian graphical model, which has different probability structure and therefore relies on different approach of analysis. A more recent work (Cao et al., 2018) also studies the efficiency of structure inference in Ising model without focus on comparison with information-theoretic bounds. We also differ in that they investigate additive statistics while we focus on oracle computational models.

Contributions: (1) Firstly, under the oracle computational model, we establish the general computational efficient lower bounds for inferring structural properties in simple zero-field Ising model, under which no polynomial-time queries can distinguish between two hypotheses. In this result, we propose a novel topological quantity on the complexity of the structure of interest: the vertex overlap ratio. We show that it is essential to the trade-off for various graph properties. (2) Secondly, we propose query functions and test functions with polynomial computational budgets which attain the lower bounds for specific property testing problems. The two results together form the computational-efficient boundary for these problems. (3) Thirdly, we discuss the relationship between information-theoretic limits and computational-efficient limits for inferring particular structures. We find that for clique detection and nearest neighbor graph detection problem, there is a gap between computational efficiency and statistical accuracy. However, for relatively sparse structure like perfect matching, we provide information lower bounds that are the same as the computational-efficient limits. This means that for such problem, there is no gap between computational efficiency and statistical accuracy.

Notation: We use the following notations in the paper. For a set D, we use $|D|$ to denote its cardinaliry. For any positive integer n, we use $[n]$ to denote the set $\{1, \ldots, n\}$. For a graph $G = (V, E)$, we use $V = V(G)$ to denote the vertex set and $E = E(G)$ the edge set of G. An element in $E(G)$ is denoted by (u, v) with $u, v \in V(G)$. The maximum degree of G is denoted by $\deg(G) := \max_{u \in V(G)} |\{v \in V(G) : (u, v) \in E(G)\}|$. Similarly, for edge set E in G, we denote

its vertex set $V(E) := \{u \in V \mid \exists v \in V, s.t.(u, v) \in E\} \subset V$ where V is the vertex set of G. For two quantities f and g (usually some order associated with (n, s, d)), we say $f(x) = O(g(x))$ if there exists positive number M such that $|f(x)| \leq M|g(x)|$ for sufficiently large x. We say $f(x) = \Omega(g(x))$ or $f \asymp g$ if $f(x) = O(g(x))$ and $g(x) = O(f(x))$.

2. Background

2.1. Simple zero-field ferromagnetic Ising model

Consider a d-dimensional random vector $\mathbf{X} = (X_1, \ldots, X_d)^\top \in \mathbb{R}^d$ following the ferromagnetic Ising model parameterized by $\boldsymbol{\theta} = (\theta_{uv})_{u,v \in [d]}$, where

$$\mathbb{P}_{\boldsymbol{\theta}}(\mathbf{X} = \mathbf{x}) \propto \exp\big(\sum_{1 \leq u < v \leq d} \theta_{uv} x_u x_v \big),$$

where $\mathbf{X}$ takes value in $\{\pm 1\}^d$ and $\theta_{uv} \geq 0$.

More specifically, we consider the *simple zero-field ferromagnetic* Ising models, where a graph $G = (V, E)$ is encoded by $\boldsymbol{\theta} = (\theta_{uv})_{u,v \in [d]}$ and $\theta_{uv} \in \{0, \theta\}$ for some $\theta \in \mathbb{R}^+$. Thus the graph has edge set $E = \{(u, v) \in V : \theta_{uv} = \theta\}$, and a *simple zero-field ferromagnetic* Ising model is encoded by $\boldsymbol{\theta}$ with distribution

$$\mathbb{P}_{\boldsymbol{\theta}}(\mathbf{X} = \mathbf{x}) = Z(\boldsymbol{\theta})^{-1} \exp(\theta \sum_{(u,v) \in E} x_u x_v) \qquad (1)$$

over all vectors of spins $\mathbf{x} \in \{\pm 1\}^d$, and $Z(\boldsymbol{\theta})$ is the normalizing constant, also known as *partition function*, such that $\mathbb{P}_{\theta}$ is a probability distribution. One usual challenge in examining Ising model is due to the calculation of the partition function, which is often intractable.

Hereafter, we consider graphs containing d vertices, and denote the parameter space $\boldsymbol{\Theta}$ of interest as all encoding vectors $\boldsymbol{\theta} = (\theta_{uv})_{u,v \in [d]}$ whose elements are either 0 or $\theta > 0$. For subset of vertice $V_0 \subset [d]$, denote $\boldsymbol{\theta}_{V_0}$ as the sub-vector containing θ_{uv} for all pairs $u, v \in V_0$.

2.2. Oracle computational model

In this section we introduce an oracle computational model which describes the interactions between algorithms and data and characterizes algorithmic complexity. Let $\mathcal{X}$ be the domain of $\mathbf{X}$, the random vector of interst, and $\mathscr{A}$ be an algorithm.

Definition 2.1 (Oracle Computational Model). *Under the oracle model, $\mathscr{A}$ interacts with an oracle $\mathcal{O}$ for T rounds. In each round, the algorithm $\mathscr{A}$ sends a query function $q : \mathcal{X} \mapsto [-M, M]$ to an oracle $\mathcal{O}$, where $q \in \mathcal{Q}_{\mathscr{A}}$, called the query space of $\mathscr{A}$. The oracle $\mathcal{O}$ responds the algorithm with a realization $z_q \in \mathbb{R}$ of a random variable Z_q which*

satisfies

$$\mathbb{P}\left(\bigcap_{q \in \mathcal{Q}_{\mathscr{A}}} \{|Z_q - \mathbb{E}[q(\mathbf{X})]| \leq \tau_q\} \right) \geq 1 - 2\xi, \quad (2)$$

where

$$\tau_q = \max \left\{ \frac{2M}{3n} \cdot [\eta(\mathcal{Q}_{\mathscr{A}}) + \log(1/\xi)], \right.$$

$$\left. \sqrt{\frac{2}{n} Var[q(\mathbf{X})] \cdot [\eta(\mathcal{Q}_{\mathscr{A}}) + \log(1/\xi)]} \right\}. \quad (3)$$

Here $\xi \in [0, 1/4)$, τ_q is the tolerance parameter and $\eta(\mathcal{Q}_{\mathscr{A}}) \geq 0$ measures the capacity of query space. If $\mathcal{Q}_{\mathscr{A}}$ is finite, then $\eta(\mathcal{Q}_{\mathscr{A}}) = \log |\mathcal{Q}_{\mathscr{A}}|$. T is called the oracle complexity.

Note that in Definition 2.1, n can be any relevent quantity such that Equations (2) and (3) hold, while in many situations n is actually the sample size.

Intuitively, Definition 2.1 describes the concentration property of the query. For example, for i.i.d. samples X_i, $i \in [n]$, for query function q, an oracle may return the sample mean of $q(x)$ as $Z(q) = \frac{1}{n} \sum_{i=1}^{n} q(X_i)$. By Bernstein's inequality combined with the union bounds, the deviation agrees with that in Equation (2). Such a definition is quite general, and more concrete examples can be found in (Wang et al., 2015).

2.3. Combinatorial Structure Inference

In this section, we formally define the test problem of interest in this paper. Denoting the set of all possible graphs over vertex set V as $\mathcal{G}$, a binary graph property is a map $\mathcal{P} : \mathcal{G} \mapsto \{0, 1\}$. Consider two disjoint sets of graphs $\mathcal{G}_0 \cap \mathcal{G}_1 = \varnothing$, $\mathcal{G}_0, \mathcal{G}_1 \subset \mathcal{G}$ with different properties $\mathcal{P}(G_0) = 0, \forall G_0 \in \mathcal{G}_0$ and $\mathcal{P}(G_1) = 1, \forall G_1 \in \mathcal{G}_1$. Given a sample of size n, property testing problems aim to test the hypotheses:

$$\mathbf{H}_0 : G \in \mathcal{G}_0 \text{ versus } \mathbf{H}_1 : G \in \mathcal{G}_1.$$

Or equivalently, based on the correspondance between graph G and its encoding vector $\boldsymbol{\theta}(G)$, the goal is to test the hypotheses:

$$\mathbf{H}_0 : \boldsymbol{\theta} \in \mathcal{C}_0 \text{ versus } \mathbf{H}_1 : \boldsymbol{\theta} \in \mathcal{C}_1,$$

where $\mathcal{C}_0$ and $\mathcal{C}_1$ is the set of encoding vectors for $\mathcal{G}_0$ and $\mathcal{G}_1$. We list a few concrete examples of tests as follows (see Figure 1 for the illustration).

Clique Detection: Let $\mathcal{G}_0 = \{(V, \varnothing)\}$ and $\mathcal{G}_1 : \{G : G \text{ is an } s\text{-clique}\}$. We aim to detect whether G is a clique (fully connected subgraph) with s vertices. See Figure 1(a) for a clique of $s = 5$. Note that the number of vertices d can be much larger than s.

Perfect Matching: Let $\mathcal{G}_0 = \{(V, \varnothing)\}$ and $\mathcal{G}_1 = \{G : |V(G)| = s$ and each vertex is incident to exactly one edge in $E(G)\}$. It means that each vertex has degree 1, and is paired with another unique vertex via one edge. See Figure 1(b) for an illustration.

Nearest neighbor graph detection: Let $\mathcal{G}_0 = \{(V, \varnothing)\}$ and $\mathcal{G}_1 = \{G : G \text{ is an } s/4\text{-nearest neighbor graph}\}$. A $s/4$-nearest neighbor graph is defined by first constructing a cycle with s vertices, and then connect each vertex in this cycle with vertices with distance $\leq s/4$ to it. Without loss of generality we assume $s/4$ is an integer. See Figure 1(c) for an illustration for $s = 8$.

For testing $\mathcal{C}_0$ against $\mathcal{C}_1$, we define the minimax testing risk $\mathcal{R}_n(\mathcal{C}_0, \mathcal{C}_1)$ as

$$\mathcal{R}_n(\mathcal{C}_0, \mathcal{C}_1) = \inf_{\psi} \left[\sup_{\boldsymbol{\theta} \in \mathcal{C}_0} \mathbb{P}_{\boldsymbol{\theta}}(\psi = 1) + \sup_{\boldsymbol{\theta} \in \mathcal{C}_1} \mathbb{P}_{\boldsymbol{\theta}}(\psi = 0) \right], \quad (4)$$

where the infimum is taken over all possible test functions ψ based on observations $\{x_i\}_{i \in [n]}$. Formally, we are interested in the conditions for any test to be asymptoticall powerless, which means $\liminf_{n \to \infty} \mathcal{R}_n(\mathcal{C}_0, \mathcal{C}_1) = 1$. With limited computational budgets, define a minimax risk for testing $\mathcal{C}_0$ against $\mathcal{C}_1$ with oracle $\mathcal{O}$ given n observations under the oracle computational model as

$$\mathcal{R}_n(\mathcal{C}_0, \mathcal{C}_1, \mathscr{A}, \mathcal{O}, T)$$
$$= \inf_{\psi \in \mathcal{H}(\mathscr{A}, \mathcal{O}, T)} \left[\sup_{\boldsymbol{\theta} \in \mathcal{C}_0} \mathbb{P}_{\boldsymbol{\theta}}(\psi = 1) + \sup_{\boldsymbol{\theta} \in \mathcal{C}_1} \mathbb{P}_{\boldsymbol{\theta}}(\psi = 0) \right], \quad (5)$$

where $\mathcal{H}(\mathscr{A}, \mathcal{O}, T)$ is the set of all tests based on algorithm $\mathscr{A}$ and oracle $\mathcal{O}$, where $\mathscr{A}$ interacts with $\mathcal{O}$ for at most T rounds. By this definition, we are interested about under which condition will there be an oracle $\mathcal{O}$ such that $\liminf_{n \to \infty} \mathcal{R}_n(\mathcal{C}_0, \mathcal{C}_1, \mathscr{A}, \mathcal{O}, T) = 1$ which means any hypothesis test based on algorithm $\mathscr{A}$ with at most T rounds of query would be asyptotically powerless. Note that T can grow with parameters n, d, s.

3. Main Results

In this section we present the main results of this paper, the computational and information bounds of detecting combinatorial structures and analysis of the existence of gap between them. In simple zero-field ferromagnetic Ising model, the parameter θ can be viewed as signal strength. Specifically, when θ is large, the realizations with more pairs of nodes taking the same value will be of higher probability, making it easier to identify certain structures.

3.1. Computational Lower Bounds

In this section we present the computational lower bound under the oracle model. We show that under some threshold, no algorithms using polynomial-time queries can distinguish the hypotheses. First, define null-alternative separator as follows.

Definition 3.1 (Null-Alternative Separator). *Let $G_0 = (V, E_0) \in \mathcal{G}_0$ be some graph under null hypothesis. A collection of edge sets $\mathcal{E}$ is called a null-alternative separator with null base G_0 for $\mathcal{G}_1$, if for all edge sets $S \in \mathcal{E}$, we have $S \cap E_0 = \varnothing$ and $(V, E_0 \cup S) \in \mathcal{G}_1$. Hereafter, with some abuse of notation we do not distinguish $\mathcal{G}_1$ and its corresponding set of encoding vectors $\mathcal{C}_1$, and also call $\mathcal{E}$ a null-alternative separator with null base G_0 for $\mathcal{C}_1$.*

In other words, a null-alternative separator is a collection of edge sets that do not overlap with the null base, and would change the graph G_0 to the alternative if added. For example, for the case of clique detection where $G_0 = (V, \varnothing)$, a null-alternative separator would include some edge sets that contain all edges among s nodes.

Given two edge sets S and S', we write $V_{S,S'} = V(S) \cap V(S')$ for simplicity. We define the vertex overlap ratio of a null-alternative separator $\mathcal{E}$ as follows.

Definition 3.2 (Vertex Overlap Ratio). *Given a null-alternative separator $\mathcal{E}$ with null base G_0 for $\mathcal{C}_1$, let $k = \max_{S,S' \in \mathcal{E}} |V_{S,S'}|$, the vertex overlap ratio is defined as*

$$\zeta = \min_{S \in \mathcal{E}} \inf_{0 \le j \le k-1} \frac{|\{S' \in \mathcal{E} : |V_{S,S'}| = j\}|}{|\{S' \in \mathcal{E} : |V_{S,S'}| = j+1\}|}, \quad (6)$$

where the quotient is defined as $+\infty$ when the denominator is zero. Naturally there are some symmetry for the edge sets in null-alternative separator $\mathcal{E}$. For example, they are the same up to permutations of vertices. Intuitively, the quantity $|V_{S,S'}|$ characterizes how correlated two edge sets are: if there are more shared vertices, they may be more related. Therefore the vertex overlap ratio quantifies the growth of correlation between edge sets in a null-alternative separator.

To be specific, we consider the case where the property testing problem focuses on some graphs where there are s vertices among all d vertices involved in a interdenpendent structure. We can say they 'work' in the graph. For example, in a graph with d vertices has nothing but an s-clique, there are s fully connected vertices, and the remaining $d - s$ vertices are of degree 0, thus independent to others.

We have the following theorem on computational lower bounds for Ising model.
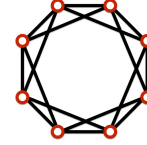
Theorem 3.1. *Suppose we have a null-alternative separator $\mathcal{E}$ with null base G_0 the empty graph for $\mathcal{C}_1$, and let $\mathcal{C}_0$ be the singleton of encoding vector for empty graph. Under the oracle computational model, if the computation budgets are*
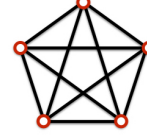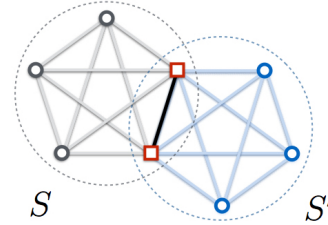

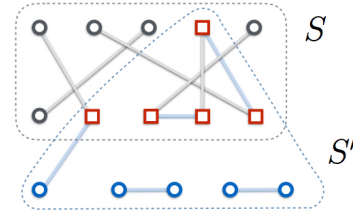(a) Clique, $s = 5$
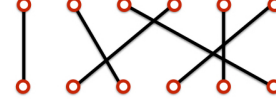
(b) Perfect matching, $s = 12$

(c) Nearest neighbor graph, $s = 8$

Figure 1: Examples of graph properties for the combinatorial inference of Ising model.

(a) Cliques S, S' with $|V_{S,S'}| = 2$

(b) Perfect mathching graphs with $|V_{S,S'}| = 5$

Figure 2: Examples of overlapping $S, S' \in \mathcal{E}$: In (a), two edge sets S, S' (colored grey and blue respectively) in $\mathcal{E}$ are both 5-cliques, and satisfies $|V_{S,S'}| = 2$, where the red vertices are $V_{S,S'}$. In (b), two edge sets $S, S' \in \mathcal{E}$ are colored grey and blue respectively, and are both perfect matching graphs. The red vertices belong to $V_{S,S'}$ with $|V_{S,S'}| = 5$.

polynomial, i.e. $T \leq d^\eta$ for some fixed $\eta > 0$, then if the vertex overlap ratio for $\mathcal{C}_1 \subset \Theta$ satisfies $\liminf_{d \to \infty} \zeta > 1$ and the parameter $\theta > 0$ for simple zero-field ferromagnetic Ising model satisfies

$$\theta \leq \frac{\kappa \log \zeta}{\log d + \log \zeta} \sqrt{\frac{\log(1/\xi)}{n}} \wedge \frac{1}{4s}, \tag{7}$$

where κ is a sufficiently small positive constant that only depends on η, then for any algorithm $\mathscr{A}$ which interacts with the oracle for at most T rounds, there exists an oracle $\mathcal{O}$ such that $\liminf_{n \to \infty} \mathcal{R}_n(\mathcal{C}_0, \mathcal{C}_1, \mathscr{A}, \mathcal{O}, T) = 1$.

Proof of Theorem 3.1. See Appendix A. □

Note that even though this theorem only covers the hardness of distinguishing alternative against an empty graph, it forms the basis for describing the hardness for more general null hypotheses, due to the monotonicity of minimax risk $\mathcal{R}_n(\mathcal{C}_0, \mathcal{C}_1, \mathscr{A}, \mathcal{O}, T)$ in $\mathcal{C}_0, \mathcal{C}_1$.

In the cases we dicussed above, we can take $\zeta \geq d/s^2$. Therefore the bound in Equation (7) can be taken to be

$$\theta \leq \kappa \sqrt{\frac{\log(1/\xi)}{n}} \wedge \frac{1}{4s}, \tag{8}$$

for some sufficiently small κ that only depends on η.

With Theorem 3.1 in hand, we have three corollaries of specific properties.

Corollary 3.1 (Clique Detection). *For testing whether the graph has an s-clique, define*

$$\mathcal{C}_0 = \{\boldsymbol{\theta} \in \Theta : \nexists V_0 \subset [d], |V_0| = s,$$
$$\boldsymbol{\theta}_{V_0} \text{ encodes an s-clique}\}$$

the set of all graphs without an s-clique, and

$$\mathcal{C}_1 = \{\boldsymbol{\theta} \in \Theta : \exists V_0 \subset [d], |V_0| = s,$$
$$\boldsymbol{\theta}_{V_0} \text{ encodes an s-clique}\}$$

the set of all graphs containing an s-cliqeu. Under the oracle computational model where an algorithm $\mathscr{A}$ interacts with the oracle $\mathcal{O}$ for at most $T = d^\eta$ rounds for some $\eta > 0$, then if $s = O(d^\alpha)$ for some $\alpha \in (0, 1/2)$ and $\theta \leq \min\{\kappa/\sqrt{n}, 1/(4s)\}$ where κ is a sufficiently small constant that only depends on η, there exists an oracle $\mathcal{O}$ such that the risk $\liminf_{n \to \infty} \mathcal{R}_n(\mathcal{C}_0, \mathcal{C}_1, \mathscr{A}, \mathcal{O}, T) = 1$.

Proof of Corollary 3.1. Let $\mathcal{C}_0^* = \{\boldsymbol{\theta} : \theta_{uv} = 0, \forall u, v\}$ the empty graph, and $\mathcal{C}_1^* = \{\boldsymbol{\theta} : \exists V_0 \subset [d], |V_0| = s, E(\boldsymbol{\theta}) = \{(u, v) : u, v \in V_0\}\}$ the collection of exact s-cliques. Then since $\mathcal{C}_0^* \subset \mathcal{C}_0$, $\mathcal{C}_1^* \subset \mathcal{C}_1$, by definition in Equation (5) we know $\mathcal{R}_n(\mathcal{C}_0, \mathcal{C}_1, \mathscr{A}, \mathcal{O}, T) \geq \mathcal{R}_n(\mathcal{C}_0^*, \mathcal{C}_1^*, \mathscr{A}, \mathcal{O}, T)$.

So it suffices to show $\liminf_{n \to \infty} \mathcal{R}_n(\mathcal{C}_0^*, \mathcal{C}_1^*, \mathscr{A}, \mathcal{O}, T) = 1$.

Let $G_0 = (V, \varnothing)$ with $|V| = s$, then $\mathcal{E} = \{E(G) : G \in \mathcal{C}_1^*\}$ is a null-alternative separator with null base G_0. We first compute the vertex overlap ratio ζ and then apply Theorem 3.1. Clearly, for any fixed $S \in \mathcal{C}_1$, there are $n_j = \binom{s}{j}\binom{d-s}{s-j}$ elements $S' \in \mathcal{C}_1$ with $|\mathcal{V}_{S,S'}| = j$. Therefore the vertex overlap ratio is

$$\zeta = \inf_{0 \leq j \leq s-1} \frac{(j+1)(d-2s+j+1)}{(s-j)^2} = \Omega\left(\frac{d}{s^2}\right).$$

Therefore if $s = O(d^\alpha)$ for some $\alpha \in (0, \frac{1}{2})$, the ratio ζ satisfies the condition for Theorem 3.1. Plugging in this quantity we get the desired result. □

Corollary 3.2 (Perfect Matching). *For testing whether the graph is perfect matching, define*

$$\mathcal{C}_0 = \{\boldsymbol{\theta} \in \Theta : \nexists V_0 \subset [d], |V_0| = s,$$
$$\boldsymbol{\theta}_{V_0} \text{ encodes a perfect matching}\}$$

the set of all graphs without an s-perfect matching, and

$$\mathcal{C}_1 = \{\boldsymbol{\theta} \in \Theta : \exists V_0 \subset [d], |V_0| = s,$$
$$\boldsymbol{\theta}_{V_0} \text{ encodes a perfect matching}\}$$

the set of all graphs containing an s-perfect matching. Under the oracle computational model where an algorithm $\mathscr{A}$ interacts with the oracle for at most $T = d^\eta$ rounds for some $\eta > 0$, if $s = O(d^\alpha)$ for some $\alpha \in (0, 1/2)$ and $\theta \leq \min\{\kappa/\sqrt{n}, 1/(4s)\}$ where κ is a sufficiently small constant that depends only on η, there exists an oracle $\mathcal{O}$ such that the risk $\liminf_{n \to \infty} \mathcal{R}_n(\mathcal{C}_0, \mathcal{C}_1, \mathscr{A}, \mathcal{O}, T) = 1$.

Proof of Corollary 3.2. Similar to the proof of Corollary 3.1, it suffices to show $\liminf_{n \to \infty} \mathcal{R}_n(\mathcal{C}_0^*, \mathcal{C}_1^*, \mathscr{A}, \mathcal{O}, T) = 1$, where $\mathcal{C}_0^*$ contains the empty graph and $\mathcal{C}_1^*$ is the collection of exact s-perfect matching graphs (i.e., containing a subgraph of s vertices that is a perfect matching, while the other $d - s$ vertices have degree 0).

For any fixed $S \in \mathcal{C}_1^*$, there are

$$n_j = \binom{s}{j}\binom{d-s}{s-j}\binom{s}{s/2} \cdot (s/2)!$$

elements $S' \in \mathcal{C}_1^*$ with $|\mathcal{V}_{S,S'}| = j$. Therefore similar to the proof of 3.1, we have $\zeta = \Omega(\frac{d}{s^2})$. Then if $s = O(d^\alpha)$ for some $\alpha \in (0, 1/2)$, the ratio ξ satisfies the conditions for Theorem 3.1. Plugging in this quantity yields the desired result. □

Corollary 3.3 (Nearest Neighbor Graph Detection). *For testing whether the graph is a nearest neighbor graph, define*

$$\mathcal{C}_0 = \{\boldsymbol{\theta} \in \boldsymbol{\Theta} : \nexists V_0 \subset [d], |V_0| = s,$$
$$\boldsymbol{\theta}_{V_0} \text{ encodes an } s/4\text{-NN graph}\}$$

the set of all graphs containing an s/4-nearest neighbor graph, and

$$\mathcal{C}_1 = \{\boldsymbol{\theta} \in \boldsymbol{\Theta} : \exists V_0 \subset [d], |V_0| = s,$$
$$\boldsymbol{\theta}_{V_0} \text{ encodes an } s/4\text{-NN graph}\}$$

the set of all graphs containing an s/4-nearest neighbor graph. Suppose we test the hypotheses under the oracle computational model where an algorithm $\mathcal{A}$ interacts with the oracle $\mathcal{O}$ for at most $T = d^\eta$ rounds for some $\eta > 0$. Then if $s = O(d^\alpha)$ for some $\alpha \in (0, 1/2)$ and $\theta \leq \min\{\kappa/\sqrt{n}, 1/(4s)\}$ where κ is a sufficiently small constant that depends only on η, there exists an oracle $\mathcal{O}$ such that the risk $\liminf_{n \to \infty} \mathcal{R}_n(\mathcal{C}_0, \mathcal{C}_1, \mathcal{A}, \mathcal{O}, T) = 1$.

Proof of Corollary 3.3. Similar to the proof of Corollary 3.1, it suffices to show $\liminf_{n \to \infty} \mathcal{R}_n(\mathcal{C}_0^*, \mathcal{C}_1^*, \mathcal{A}, \mathcal{O}, T) = 1$, where $\mathcal{C}_0^*$ contains the empty graph and $\mathcal{C}_1^*$ is the collection of exact $s/4$-nearest neighbor graphs (i.e., containing a subgraph of s vertices that is an $s/4$-nearest neighbor graph, while the other $d - s$ vertices have degree 0).

For any fixed $S \in \mathcal{C}_1^*$, there are $n_j = \binom{s}{j}\binom{d-s}{s-j} \cdot (s-1)!$ elements $S' \in \mathcal{C}_1^*$ with $|\mathcal{V}_{S,S'}| = j$. Therefore similar to the proof of 3.1, we have $\zeta = \Omega(\frac{d}{s^2})$. Then if $s = O(d^\alpha)$ for some $\alpha \in (0, 1/2)$, the ratio ξ satisfies the conditions for Theorem 3.1. Plugging in the quantity yields the desired result. $\square$

In the above examples, due to symmetry in the structure, all of the vertex overlap ratios can be set as $\zeta = \Omega(d/s^2)$, yielding similar results for computational lower bounds.

3.2. Computational Upper Bounds

In this section we provide upper bounds for clique detection problem that matches the lower bound in Section 3.1, which is obtained by pair-wise correlation test defined as follows. Here we consider testing empty graph against a graph containing an s-clique.

Given n i.i.d. observations from a simple zero-field ferromagnetic Ising model $G = (V, E)$ with $|V| = d$, for each $j, k \in [d]$ with $j \neq k$ we define query functions

$$q_{jk}(x) = x_j x_k, \tag{9}$$

and test function $\psi(\cdot)$ as

$$\psi(X) = \mathbb{I}\left[\max_{j \neq k} q_{jk}(X) - \min_{j \neq k} q_{jk}(X) \right.$$
$$\left. \geq \sqrt{\frac{8}{n} \log\left(\frac{d(d-1)}{2\xi}\right)} \right]. \tag{10}$$

Then we have the following general computational upper bound attained by the aforementioned query fuctions and test function.

Theorem 3.2. *Suppose we use the aforementioned query functions in Equation (9) and test function in Equation (10) to test empty graph against some particular structure. If $\log d/n = o(1)$ and*

$$\theta \geq 4\sqrt{\frac{2}{n} \log\left(\frac{d(d-1)}{2\xi}\right)} \tag{11}$$

for some sufficiently large constant $c > 0$ that does not depend on ξ, we have

$$\sup_{\boldsymbol{\theta} \in \mathcal{C}_0} \mathbb{P}_{\boldsymbol{\theta}}[\psi = 1] + \sup_{\boldsymbol{\theta} \in \mathcal{C}_1} \mathbb{P}_{\boldsymbol{\theta}}[\psi = 0] \leq 4\xi.$$

Proof of Theorem 3.2. See Appendix B.1. $\square$

Note that this theorem allows all situations where there is at least one edge in the structure that interests us. The idea is that under $\mathbf{H}_0$, all queries has expectation of zero and variance 1, and under $\mathbf{H}_1$, at least one query has expectation $\mathbb{E}_S[X_j X_k] \geq \theta$, and meanwhile all queries are with variance no greater than 1. Therefore the deviation of the realization $q_{jk}(X)$ will be bounded, and once the signal strength θ overwhelms the statistical deviation, the two hypotheses can be distinguished.

Theorem 3.2 can be applied to perfect matching problem, leading to the following corollary.

Corollary 3.4 (Computational Upper Bound for Perfect Matching). *Suppsose we employ the query functions in Equation (9) and test function in Equation (10) to test empty graph against s-perfect matching graphs. If $\log d/n = o(1)$ and*

$$\theta \geq 4\sqrt{\frac{2}{n} \log\left(\frac{d(d-1)}{2\xi}\right)},$$

then we have

$$\sup_{\boldsymbol{\theta} \in \mathcal{C}_0} \mathbb{P}_{\boldsymbol{\theta}}[\psi = 1] + \sup_{\boldsymbol{\theta} \in \mathcal{C}_1} \mathbb{P}_{\boldsymbol{\theta}}[\psi = 0] \leq 4\xi.$$

By Equation (3.4), if we omit the logarithm term and $s^2/n = O(1)$, the computational upper bound matches the computational lower bound provided in Corollary 3.2.

However, when the structure of interest involves sufficiently many edges, the correlation $\mathbb{E}_S[X_j X_k]$ has a better lower bound than merely θ. In some cases like clique detection and $s/4$-nearest neighbor graph detection, the computational upper bound may be sharper than what is provided by Theorem 3.2, as specified by the following two theorems.

Theorem 3.3 (Computational Upper Bound for Clique Detection). *Suppose we employ the query functions in Equation (9) and test function in Equation (10) to test $\mathcal{C}_0$ the empty graph against $\mathcal{C}_1$ all graphs containing an s-clique. If $\log d/n = o(1)$ and*

$$\theta \geq 4\sqrt{\frac{2}{n} \log\Big(\frac{d(d-1)}{2\xi}\Big)} \wedge \Big(\frac{c}{s}\Big) \qquad (12)$$

for some sufficiently large constant $c > 0$ that does not depend on ξ, we have

$$\sup_{\boldsymbol{\theta} \in \mathcal{C}_0} \mathbb{P}_{\boldsymbol{\theta}}[\psi = 1] + \sup_{\boldsymbol{\theta} \in \mathcal{C}_1} \mathbb{P}_{\boldsymbol{\theta}}[\psi = 0] \leq 4\xi.$$

Proof of Theorem 3.3. See Appendix B.2. □

By Equation (12), if we omit the logarithm term, the computational upper bound matches the computational lower bound provided in Corollary 3.1. This bound is the efficient detection boundary for testing nearest neighbor graphs using polynomial-time algorithms.

For detecting nearest neighbor graphs, the following computational lower bound is attained by employing aforementioned query functions and test function.

Theorem 3.4 (Computational Upper Bound for Nearest Neighbor Graph Detection). *Suppsose we employ the aforementioned query functions in Equation (9) and test in Equation (10) to test $\mathcal{C}_0$ the empty graph against $\mathcal{C}_1$ all graphs containing an $s/4$-nearest neighbor graph. If $\log d/n = o(1)$ and*

$$\theta \geq 4\sqrt{\frac{2}{n} \log\Big(\frac{d(d-1)}{2\xi}\Big)} \wedge \Big(\frac{c}{s}\Big) \qquad (13)$$

for some sufficiently large constant c, then we have

$$\sup_{\boldsymbol{\theta} \in \mathcal{C}_0} \mathbb{P}_{\boldsymbol{\theta}}[\psi = 1] + \sup_{\boldsymbol{\theta} \in \mathcal{C}_1} \mathbb{P}_{\boldsymbol{\theta}}[\psi = 0] \leq 4\xi.$$

Proof of Theorem 3.4. See Appendix B.3. □

Note that this computational upper bound matches the computational lower bound provided in Corollary 3.3. It is the efficient detecting boundary for polynomial-time algorithms, and is attained by queries and test functions defined before.

3.3. Information Upper Bounds

In this section, we consider another test function ψ with no limits on computation budgets. With unlimited computation budgets, we can potentially better investigate the structure of the underlying graph. Here we define a new set of query functions and the corresponding test function ψ. The detecting ability of ψ gives an information upper bound for some kind of property testing problems.

Given n i.i.d. observations $\{X_i\}_{i \in [n]}$ from a simple zero-field ferromagnetic Ising model $G = (V, E)$ with $|V| = d$, for each subset $S \subset [d]$, $|S| = s$, we consider the query functions

$$q_S(X) = \Big(\frac{1}{s} \sum_{i \in S} X_i\Big)^2 \qquad (14)$$

and the test function

$$\psi = \mathbb{I}\Big[\max_{S \subset [d], |S| = s} q_S(X) - \frac{1}{s} \geq r(n, d, s, \xi)\Big] \qquad (15)$$

for some $\xi \in (0, 1/2)$, where

$$r(n, d, s, \xi) = c \cdot \max\Big\{\frac{s}{n} \log\Big(\frac{d}{s\xi}\Big), 2\sqrt{\frac{1}{ns} \log\Big(\frac{d}{s\xi}\Big)}\Big\}$$

for some constant c, then we have the following theorem providing general information upper bound for detecting structures with sufficiently many edges.

Theorem 3.5. *Suppose we use the query functions introduced in Equation (14) and test function in Equation (15) to test empty graph in $\mathcal{C}_0$ against graphs in $\mathcal{C}_1$. Suppose $(s/n) \log(d/s) = o(1)$, and for each $G \in \mathcal{C}_1$ we have $|V(G)| = s$ and $|E(G)| \geq \gamma \cdot s^2$ for some constant $\gamma > 0$, then for any $\xi \in (0, 1/4)$, if*

$$\theta \geq \Big(\frac{c \cdot s^2}{n} \log\Big(\frac{d}{s\xi}\Big)\Big) \vee c \cdot \sqrt{\frac{1}{ns} \log\Big(\frac{d}{s\xi}\Big)}$$

for a sufficiently large constant $c > 0$, then

$$\sup_{\boldsymbol{\theta} \in \mathcal{C}_0} \mathbb{P}_{\boldsymbol{\theta}}(\psi = 1) + \sup_{\boldsymbol{\theta} \in \mathcal{C}_1} \mathbb{P}_{\boldsymbol{\theta}}(\psi = 0) \leq 4\xi.$$

Proof of Theorem 3.5. See Appendix C. □

The conditions in Theorem 3.5 apply for structures with dense edges, yielding the following two corollaries.

Corollary 3.5 (Information Upper Bound for Clique Detection). *Suppose we use the aforementioned query functions $q_S(X)$ in Equation (14) and test function $\phi(X)$ in Equation (15) for s-clique detection. Suppose $(s/n)\log(d/s) = o(1)$, then for any $\xi \in (0, 1/4)$, if*

$$\theta \geq \left(\frac{c \cdot s^2}{n} \log \left(\frac{d}{s\xi} \right) \right) \vee c \cdot \sqrt{\frac{1}{ns} \log \left(\frac{d}{s\xi} \right)}$$

for a sufficiently large constant $c > 0$, then

$$\sup_{\boldsymbol{\theta} \in \mathcal{C}_0} \mathbb{P}_{\boldsymbol{\theta}}(\psi = 1) + \sup_{\boldsymbol{\theta} \in \mathcal{C}_1} \mathbb{P}_{\boldsymbol{\theta}}(\psi = 0) \leq 4\xi.$$

Proof of Corollary 3.5. Directly apply Theorem 3.5 with $\gamma = 1/3$, since there are $s(s-1)/2$ edges in an s-clique, then the results hold. □

Note that when n is sufficiently large compared to s, for example $n = O(s^4)$, the detecting boundary for the test in Equation (15) is lower than the detecting boundary for the test in Equation (10) when computational budgets are limited to polynomial. This means that there is a gap between the computational efficient bound and information-theoretic bound, where exponential-time algorithms can be used.

For detecting $s/4$-nearest neighbor graph, we have the following information upper bound, which is also attained by the query functions specified in Equation (14) and test function specified in Equation (15).

Corollary 3.6 (Information Upper Bound for Nearest Neighbor Graph Detection). *Suppose we use the aforementioned query functions and test function $q_S(X)$, ψ for $S \subset [d], |S| = s$ for s/4-nearest neighbor graph detection (assume that $s/4$ is integer). Supose $(s/n)\log(d/s) = o(1)$, then for any $\xi \in (0, 1/4)$, if*

$$\theta \geq \left(\frac{c \cdot s^2}{n} \log \left(\frac{d}{s\xi} \right) \right) \vee c \cdot \sqrt{\frac{1}{ns} \log \left(\frac{d}{s\xi} \right)}$$

for a sufficiently large constant $c > 0$, then

$$\sup_{\boldsymbol{\theta} \in \mathcal{C}_0} \mathbb{P}_{\boldsymbol{\theta}}(\psi = 1) + \sup_{\boldsymbol{\theta} \in \mathcal{C}_1} \mathbb{P}_{\boldsymbol{\theta}}(\psi = 0) \leq 4\xi.$$

Proof of Corollary 3.6. Directly use Theorem 3.5 with $\gamma = 1/4$, since there are $s^2/4$ edges in an $s/4$-nearest neighbor graph. □

Note that by Corollary 3.6, when n is sufficiently large compared to s, the information upper bound is lower than the computational-efficient boundary. This means that there is a gap between computational efficiency and statistical accuracy in the nearest neighbor graph detection problem.

3.4. Information Lower Bound for Perfect Matching

In this section we consider the information lower bound for perfect matching, saying that when the signal strength θ is below some threshold, no algorithm can distinguish empty graph against perfect matching graph of size s with probability larger than some small constant. This completes our discussion on computational-statistical gaps for particular examples. Specifically, we have the following theorem.

Theorem 3.6 (Information Lower Bound for Perfect Matching). *Consider the problem of testing perfect matching graph with $\mathcal{C}_0 = \{(V, \varnothing)\}$ and $\mathcal{C}_1 = \{\boldsymbol{\theta} : \boldsymbol{\theta}$ is exact perfect matching$\}$. Given n i.i.d. samples from $\mathbb{P}_{\boldsymbol{\theta}}$, if the parameter θ for ferromagnetic Ising model satisfies*

$$\theta \leq \frac{c}{\sqrt{n}} \wedge \frac{c}{s} \qquad (16)$$

for some sufficiently small constant $c > 0$, then no algorithm can distinguish between the two hypotheses with probability larger than a small constant $\widetilde{c}$ that depends on c.

Proof of Theorem 3.6. See Appendix D. □

Note that the information lower bound for perfect matching provided in Equation (3.6) is of the same order as the computational efficient boundary, which means that for detecting perfect matching graphs, there is no such gap between computational efficiency and statistical efficiency.

4. Conclusion

We characterize the computational lower bounds for testing structure in simple zero-field ferromagnetic Ising models via a novel quantity called vertex overlap ratio. We show that such quantity could capture the compuatational boundaries of a large family of graph properties, which are actually obtained by efficient algorithms when testing some particular structures against empty graph. We also investigate information-theoretic boundaries in these problems, and find a computational-statistical gap in detecting cliques and nearest neighbor graphs. Meanwhile, we provide information lower bound for testing perfect matching, which shows that there is no such gap for this problem. The difference in the gaps is shown to be related to the density of edges within the structure of interest.

In the future, we aim to generalize our framework to a larger family of discrete graphical models, including the anti-ferromagnetic Ising model, Poisson graphical model and the mixed-value graphical models.

References

Alon, N. and Spencer, J. H. *The probabilistic method*. John Wiley & Sons, 2004.

Bresler, G., Gamarnik, D., and Shah, D. Structure learning of antiferromagnetic ising models. 2014.

Cai, T. T., Liang, T., and Rakhlin, A. Computational and statistical boundaries for submatrix localization in a large noisy matrix. *Ann Statistics*, 45(4):1403–1430, 2017. ISSN 0090-5364. doi: 10.1214/16-AOS1488.

Cao, Y., Neykov, M., and Liu, H. High temperature structure detection in ferromagnets, 2018.

Chen, Y. and Xu, J. Statistical-Computational tradeoffs in planted problems and submatrix localization with a growing number of clusters and submatrices. 2014.

Daskalakis, C., Dikkala, N., and Kamath, G. Testing ising models. 2016.

Ery, A. and Verzelen, N. Community detection in dense random networks. *Ann Statistics*, 42(3):940–969, 2014. ISSN 0090-5364. doi: 10.1214/14-AOS1208.

Fan, J., Liu, H., Wang, Z., and Yang, Z. Curse of heterogeneity: Computational barriers in sparse mixture models and phase retrieval. 2018.

Feldman, V., Grigorescu, E., Reyzin, L., Vempala, S., and Xiao, Y. *Statistical algorithms and a lower bound for detecting planted cliques*. acm, 2013. ISBN 9781450320290. doi: 10.1145/2488608.2488692.

Friedman, N. Inferring cellular networks using probabilistic graphical models. *Science*, 303(5659):799–805, 2004. ISSN 0036-8075. doi: 10.1126/science.1094068.

Grabowski, A. and Kosiński, R. Ising-based model of opinion formation in a complex network of interpersonal interactions. *Phys Statistical Mech Appl*, 361(2):651–664, 2006. ISSN 0378-4371. doi: 10.1016/j.physa.2005.06.102.

Griffiths, R. B. Correlations in ising ferromagnets. i. *Journal of Mathematical Physics*, 8(3):478–483, 1967.

Hajek, B., Wu, Y., and Xu, J. Computational lower bounds for community detection on random graphs. 2014.

Kearns, M. Efficient noise-tolerant learning from statistical queries. *J Acm Jacm*, 45(6):983–1006, 1998. ISSN 0004-5411. doi: 10.1145/293347.293351.

Lazega, E., Wasserman, S., and Faust, K. Social network analysis: Methods and applications. *Revue Française De Sociol*, 36(4):781, 1995. ISSN 0035-2969. doi: 10.2307/3322457.

Louigi, A., Broutin, N., Devroye, L., and Lugosi, G. On combinatorial testing problems. *Ann Statistics*, 38(5):3063–3092, 2010. ISSN 0090-5364. doi: 10.1214/10-AOS817.

Lu, H., Cao, Y., Yang, Z., Lu, J., Liu, H., and Wang, Z. The edge density barrier: Computational-statistical tradeoffs in combinatorial inference. In Dy, J. and Krause, A. (eds.), *Proceedings of the 35th International Conference on Machine Learning*, volume 80 of *Proceedings of Machine Learning Research*, pp. 3247–3256, Stockholmsmässan, Stockholm Sweden, 10–15 Jul 2018. PMLR. URL http://proceedings.mlr.press/v80/lu18a.html.

Neykov, M. and Liu, H. Property testing in high dimensional ising models. 2017.

Neykov, M., Lu, J., and Liu, H. Combinatorial inference for graphical models. 2016.

Wang, T., Berthet, Q., and Samworth, R. J. Statistical and computational trade-offs in estimation of sparse principal components. *Ann Statistics*, 44(5):1896–1930, 2016. ISSN 0090-5364. doi: 10.1214/15-AOS1369.

Wang, Z., Gu, Q., and Liu, H. Sharp Computational-Statistical phase transitions via oracle computational model. 2015.

Zhang, Y., Wainwright, M. J., and Jordan, M. I. Lower bounds on the performance of polynomial-time algorithms for sparse linear regression. 2014.