The Annals of Statistics
2019, Vol. 47, No. 5, 2472–2503
https://doi.org/10.1214/18-AOS1754

PROPERTY TESTING IN HIGH-DIMENSIONAL ISING MODELS

BY MATEY NEYKOV AND HAN LIU[1]

Carnegie Mellon University and Northwestern University

This paper explores the information-theoretic limitations of graph property testing in zero-field Ising models. Instead of learning the entire graph structure, sometimes testing a basic graph property such as connectivity, cycle presence or maximum clique size is a more relevant and attainable objective. Since property testing is more fundamental than graph recovery, any necessary conditions for property testing imply corresponding conditions for graph recovery, while custom property tests can be statistically and/or computationally more efficient than graph recovery based algorithms. Understanding the statistical complexity of property testing requires the distinction of *ferromagnetic* (i.e., positive interactions only) and general Ising models. Using combinatorial constructs such as graph packing and strong monotonicity, we characterize how target properties affect the corresponding minimax upper and lower bounds within the realm of ferromagnets. On the other hand, by studying the detection of an *antiferromagnetic* (i.e., negative interactions only) Curie–Weiss model buried in Rademacher noise, we show that property testing is strictly more challenging over general Ising models. In terms of methodological development, we propose two types of correlation based tests: computationally efficient screening for ferromagnets, and score type tests for general models, including a fast cycle presence test. Our correlation screening tests match the information-theoretic bounds for property testing in ferromagnets in certain regimes.

1. Introduction. The Ising model is a pairwise binary model introduced by statistical physicists as a model for spin systems with the goal of understanding spontaneous magnetization and phase transitions (Ising (1925)). More recently, the model has found applications in diverse areas such as image analysis (Geman and Geman (1984)), bioinformatics and social networks (Ahmed and Xing (2009)). In statistics, the model is an archetypal example of an undirected graphical model. A central topic of interest in graphical models research is estimating the structure (also known as *structure learning*) of, or inferring questions about, the underlying graph based on a sample of observations. Substantial progress has been made toward understanding structure learning. Popular procedures developed for high-dimensional graph estimation include ℓ_1-regularization methods (Yuan and Lin

Received April 2018.
[1]Supported by NSF BIGDATA 1840866, NSF RI 1408910, NSF CAREER 1841569, NSF TRIPODS 1740735, along with an Alfred P. Sloan Fellowship.
MSC2010 subject classifications. 62F03, 62H15.
Key words and phrases. Ising models, minimax testing, Dobrushin's comparison theorem, antiferromagnetic Curie–Weiss detection, two-point function bounds, correlation tests.

(2007), Rothman et al. (2008), Liu, Lafferty and Wasserman (2009), Ravikumar et al. (2011), Cai, Liu and Luo (2011)), neighborhood selection (Meinshausen and Bühlmann (2006), Bresler, Mossel and Sly (2008)) and thresholding (Montanari and Pereira (2009)). In this paper, instead of focusing on learning the structure of the entire graph, we study the weaker inferential problem of *property testing*, that is, testing whether the graph structure obeys certain properties based on a sample of n observations. Specifically, we study the *zero-field* Ising model.

Formally, a zero-field Ising model is a collection of d binary ± 1 valued random variables $\boldsymbol{X} = (X_1, X_2, \dots X_d)$, hereto referred to as spins, which are distributed according to the law

$$\mathbb{P}_{\theta, G_{\mathbf{w}}}(\boldsymbol{X} = \mathbf{x}) \propto \exp\Bigg(\theta \sum_{(u,v)\in E(G)} w_{uv} x_u x_v\Bigg), \tag{1.1}$$

where $\theta \geq 0$, $G_{\mathbf{w}} = (G, \mathbf{w})$ where $G = ([d], E)$ is a simple graph, and $\mathbf{w} \in \mathbb{R}^{\binom{d}{2}}$ are weights on the graph's edges (i.e., for each edge $(u, v) \in E(G)$, $\mathbf{w}$ specifies the edge weight w_{uv}, and for any $(u, v) \notin E(G)$: $w_{uv} = 0$). Using (1.1), it is easily seen that the vector $\boldsymbol{X}$ is Markov to the graph G, or in other words, any two non-adjacent spins X_u and X_v ($(u, v) \notin E(G)$) are independent given the values of all the remaining spins.

Note that model (1.1) is overparametrized. However, when w_{uv} are viewed as fixed constants, this specification allows one to study the behavior of $\boldsymbol{X}$ for different values of θ. In statistical physics, the parameter $\theta = \frac{1}{T}$ where T stands for temperature, and is often referred to as the *inverse temperature* of the system. The temperature plays an important role in changing the "balance" of the distribution of the spins, and is the main cause for the system to undergo phase transitions. The complicated behavior of the Ising model at different temperatures suggests that the difficulty of property testing is related to θ. The main focus of this paper is uncovering necessary and sufficient conditions on the temperature, sample size, dimensionality and graph properties, allowing one to conduct property tests even when the data is sampled from the most challenging models. Understanding such limitations is practically useful, since necessary conditions can provide a benchmark for algorithm comparisons, while mismatches between sufficient and necessary conditions can prompt to searching for better algorithms.

To elaborate on the type of problems, we study, let $[d] = \{1, \dots, d\}$ be a vertex set of cardinality d and let $\mathcal{G}_d$ be the set of all graphs over the vertex set $[d]$. A binary graph property $\mathcal{P}$ is a map $\mathcal{P} : \mathcal{G}_d \mapsto \{0, 1\}$. Given a sample of n observations from a zero-field Ising model with an underlying simple graph G, the goal of property testing is to test the hypotheses

$$\mathbf{H}_0 : \mathcal{P}(G) = 0 \quad \text{versus} \quad \mathbf{H}_1 : \mathcal{P}(G) = 1. \tag{1.2}$$

Below we give three specific instances of property tests. We furthermore give informal summaries of our findings, which are presented more rigorously in Sections 2 and 3.

Connectivity. A graph is connected if and only if each pair of its vertices is connected via a path. Define $\mathcal{P}$ as $\mathcal{P}(G) = 0$ if G is disconnected and $\mathcal{P}(G) = 1$ otherwise. Testing for connectivity is equivalent to testing whether the variables can be partitioned into two independent sets. It turns out that in *simple ferromagnets* (i.e., models whose spin-spin interactions satisfy $w_{uv} \in \{0, 1\}$ for all (u, v)) connectivity testing is possible iff

$$\sqrt{\frac{\log d}{n}} \lesssim \theta,$$

where $\lesssim$ is inequality up to constants. Note that there is no upper bound on θ and as long as θ is large enough connectivity testing is always possible.

Cycle presence. If a graph is a forest, that is, a graph containing no cycles, its structure can be estimated efficiently using a graph selection procedure based on a maximum spanning tree construction proposed by Chow and Liu (1968). It is therefore sometimes of interest to test whether the underlying graph is a forest. In this example, $\mathcal{P}$ satisfies $\mathcal{P}(G) = 0$ if G is a forest and $\mathcal{P}(G) = 1$ otherwise. We will also refer to forest testing as cycle testing, since it is equivalent to testing whether the graph contains cycles. In simple ferromagnets, cycle testing is possible iff

$$\sqrt{\frac{\log d}{n}} \lesssim \theta \lesssim \log \frac{n}{\log d}.$$

In contrast to connectivity, there appears to be an upper bound on the temperature when one tests for cycle presence.

Clique size. Another relevant question is to test whether the size of a maximum clique (i.e., a maximum complete subgraph) contained in the graph is less than or equal to some integer $m - 1$ versus the alternative that a maximum clique is of size at least m. This is a relevant question since Hammersley–Clifford's theorem (Grimmett (2018)) ensures that the Ising distribution can be factorized over the cliques in the graph, and hence knowing the maximal size of any clique puts a restriction on this factorization. In this example, set $\mathcal{P}(G) = 0$ if G contains no m-clique, and $\mathcal{P}(G) = 1$ otherwise. Let the maximum degree[2] of the graph G be s. It turns out that testing the clique size is impossible in simple ferromagnets unless

$$\sqrt{\frac{\log d}{n}} \lesssim \theta \lesssim \frac{\log \frac{n}{\log d}}{s}.$$

[2]The largest number of neighbors of any vertex of G.

Different from before, the maximum degree appears in the upper bound on θ. We will show that testing the maximum clique size is possible when

$$\sqrt{\frac{\log d}{n}} \lesssim \theta \lesssim \frac{1}{s} \quad \text{and when } m = s+1 \text{ and } \quad \frac{\log s}{s} \ll \theta \lesssim \frac{\log \frac{n}{\log d}}{s},$$

where $\ll$ is used in the sense "much larger than" (for a precise definition see the notation section below). This matches the previous two bounds up to constants.

By definition, property testing is a statistically simpler task compared to learning the entire graph structure, since if a graph estimate is available, property testing can be done via a deterministic procedure (although possibly a computationally challenging one). An important implication of this observation is that any quantification on how hard testing a particular graph property is, immediately implies that estimating the entire graph is at least as hard. Conversely, any algorithm capable of learning the graph structure with high confidence can be applied to test any property while preserving the same confidence. Importantly, however, there could exist tests geared toward particular graph properties which can statistically and/or computationally outperform generic graph learning methods.

Under the assumption that the maximum degree of G is at most s, foundational results on the limitations of structure learning of Ising models were given by Santhanam and Wainwright (2012). In view of the relationship between property tests and structure learning, our work can be seen as a generalization of necessary conditions for structure learning. Our results also help to paint a more complete picture of the statistical complexity of testing in Ising models. Unlike in structure learning, understanding property testing requires the distinction of ferromagnetic and general Ising models, of which the latter exhibit strictly stronger limitations. In terms of methodological development, we formalize correlation based property tests which can be customized to target any graph property. We now outline the three major contributions of this work.

1.1. *Summary of contributions.* Our first contribution is to provide necessary conditions for property testing in ferromagnets. We give a generic lower bound on the inverse temperature (Theorem 2.4), demonstrating that property testing is difficult in high-temperature regimes. A key role in the proof is played by Dobrushin's comparison theorem (Föllmer (1988)), which a is powerful tool for comparing discrepancies between Gibbs measures based on their local specifications. We further formalize the class of *strongly monotone* graph properties, and show that when the temperature drops below a certain property dependent threshold, testing strongly monotone properties becomes challenging (Theorem 2.6). We also provide an analogue of Theorem 2.4 specialized for strongly monotone properties (Proposition 2.7). Our general results are applied to obtain bounds on testing connectivity, cycles and maximum clique size.

Our second contribution is to design several correlation based tests and understand their limitations. First, we formalize and study a generic correlation screening algorithm for ferromagnets. We show that this algorithm works well at high-temperature regimes (Remark 3.4 and Corollary 3.3), and could be successful even beyond this regime for some properties (Section 3.2). To analyze the algorithms at low temperature regimes, we develop a novel "no-edge" correlation bound for graphs of bounded degree (see Proposition 3.3), which may be of independent interest. We apply those algorithms to testing connectivity, cycles and maximum clique size and discover that they match the derived lower bounds in certain regimes. Second, we adapt the correlation decoders of Santhanam and Wainwright (2012) to property testing for general Ising models, and we develop a computationally tractable cycle test (Section 4.3).

Our third contribution is to study necessary conditions for general Ising models, that is, models including both *ferromagnetic and antiferromagnetic*[3] interactions. Specifically, we argue that testing strongly monotone properties over general models requires more stringent conditions than performing the same tests over ferromagnets (Theorem 4.1 and Proposition 4.3). In order to prove this result, we demonstrate that it is very difficult to detect the presence of an antiferromagnetic Curie–Weiss[4] (e.g., see Kochmański, Paszkiewicz and Wolski (2013)) model buried in Rademacher noise, which to the best of our knowledge is the first attempt to analyze this problem. The detection problem we consider is in part inspired by the works Addario-Berry et al. (2010), Arias-Castro, Bubeck and Lugosi (2012, 2015), Arias-Castro et al. (2018).

1.2. *Related work.* Recent works on Ising models related to the Curie–Weiss model include Berthet, Rigollet and Srivastava (2016), Mukherjee, Mukherjee and Yuan (2018). An interesting paper on testing goodness-of-fit in Ising models by Daskalakis, Dikkala and Kamath (2018), uses tests based on minimal pairwise correlations which are similar in spirit to some of the tests we consider. In a related work, Gheissari, Lubetzky and Peres (2017) demonstrated that sums of pairwise correlations concentrate for general Ising models. Pseudo-likelihood parameter estimation and inference of the inverse temperature for Ising models of given structures was studied by Bhattacharya and Mukherjee (2018). Property testing is a fundamentally different problem, and our work is in part inspired by Neykov, Lu and Liu (2016). We show that the graph packing constructions introduced by Neykov, Lu and Liu (2016) for Gaussian models, can also be used to give upper bounds on the temperature for property testing in Ising models (Theorem 2.4). Unlike Neykov, Lu and Liu (2016) however, we do not restrict our study to graphs of

[3]Inspired by statistical physics jargon, throughout the paper we use the terms ferromagnetic and antiferromagnetic to refer to positive and negative interactions, respectively.

[4]That is, an antiferromagnetic model with a complete graph.

bounded degree, and we give a more complete picture of the complicated landscape of property testing in Ising models, by distinguishing ferromagnetic from general models (see Theorems 2.6 and 4.1, and Propositions 2.7 and 4.3).

Structure learning is very relevant to property testing. Restricted to the class of ferromagnetic models, Shanmugam et al. (2014) related structural conditions of the graph with information-theoretic bounds. Santhanam and Wainwright (2012) suggested correlation decoders, which are computationally inefficient but to the best of our knowledge have the smallest sample size requirements for general models. Anandkumar et al. (2012) gave a polynomial time neighborhood selection method for models whose graphs obey special properties. The first polynomial time algorithm which works for general Ising models was given by Bresler (2015) and was motivated by earlier works on structure recovery (Bresler, Mossel and Sly (2008), Bresler, Gamarnik and Shah (2014)). Inspired by the simplicity of the correlation algorithms studied by Montanari and Pereira (2009), Santhanam and Wainwright (2012), we use similar ideas to develop property tests, and demonstrate that for some properties our tests work in vastly different regimes compared to graph recovery.

1.3. *Notation.* For convenience of the reader, we summarize the notation used throughout the paper. For a vector $\mathbf{v} = (v_1, \dots, [d])^T \in \mathbb{R}^d$, let $\|\mathbf{v}\|_q = (\sum_{i=1}^d v_i^q)^{1/q}$, $1 \le q < \infty$ with the usual extension for $q = \infty$: $\|\mathbf{v}\|_\infty = \max_i |v_i|$. Moreover, for a matrix $\mathbf{A} \in \mathbb{R}^{d\times d}$ we denote $\|\mathbf{A}\|_p = \max_{\|\mathbf{v}\|_p=1} \|\mathbf{A}\mathbf{v}\|_p$ for $p \ge 1$. For any $n \in \mathbb{N}$, we use the shorthand $[n] = \{1, \dots, n\}$. We denote $\mathbb{N}_0 = \mathbb{N} \cup \{0\}$. For a set $N \subset \mathbb{N}$, we define $\binom{N}{2} = \{(u, v) \mid u < v, u, v \in N\}$ to be the set of ordered pairs of numbers in N. For a graph $G = (V, E)$, we use $V(G) = V$, $E(G) = E$, $\mathrm{maxdeg}(G)$ to refer to the vertex set, edge set and maximum degree of G, respectively. For two graphs G, G', we use $G' \trianglelefteq G$ if G' is a *spanning subgraph* of G, that is, $V(G') = V(G)$ and $E(G') \subseteq E(G)$; we use $G' \subseteq G$ if G' is a subgraph of G but not necessarily a spanning one, that is, $V(G') \subseteq V(G)$ and $E(G') \subseteq E(G)$. For a graph $G = (V, E)$ and an edge e, we will write $e \in G$, $e \in E$ or $e \in E(G)$ interchangeably whenever this does not cause confusion.

For a probability measure $\mathbb{P}$, the notation $\mathbb{P}^{\otimes n}$ means the product measure of n independent and identically distributed (i.i.d.) samples from $\mathbb{P}$. For two functions $f(x)$ and $g(x)$, we use the notation $f(x) \approx g(x)$ in the sense that $\lim_{x\downarrow 0} \frac{f(x)}{g(x)} = 1$. Given two sequences $\{a_n\}, \{b_n\}$, we write $a_n = O(b_n)$ if for large enough n there exists a constant $C < \infty$ such that $a_n \le C b_n$; $a_n = \Omega(b_n)$ if there exists a positive constant $c > 0$ such that $a_n \ge c b_n$; $a_n = o(b_n)$ if $a_n/b_n \to 0$, and $a_n \asymp b_n$ if there exists positive constants c and C such that $c < a_n/b_n < C$; $a_n \gtrsim b_n$ if there exists an absolute constant $c > 0$ so that $a_n \ge c b_n$. Finally, we use $\wedge$ and $\vee$ for min and max of two numbers, respectively. For positive sequences a_n and b_n, we denote $a_n \gg b_n$ if $b_n/a_n = o(1)$.

1.4. *Organization.* The remainder of the paper is structured as follows. Minimax bounds for ferromagnetic models are given in Section 2. Section 3 is dedicated to correlation screening algorithms for testing in ferromagnets. Section 4 provides minimax bounds for general models and studies correlation based algorithms for general models. The proofs of two results on strongly monotone properties, Theorem 2.6 and Proposition 2.7, are given in Section 5. Discussion is postponed to the final Section 6. Most proofs are relegated to the Appendices A–E in the Supplementary Material (Neykov and Liu (2019)).

2. Bounds for ferromagnets.

This section discusses lower and upper bounds on the temperature for ferromagnetic models. We begin by formally introducing the *simple zero-field ferromagnetic* Ising models. Given a $\theta \geq 0$, the simple zero-field ferromagnetic Ising model with signal θ is given by

$$\mathbb{P}_{\theta,G}(\boldsymbol{X} = \mathbf{x}) = \frac{1}{Z_{\theta,G}} \exp\biggl(\theta \sum_{(u,v)\in E(G)} x_u x_v\biggr), \tag{2.1}$$

where the vector of spins $\mathbf{x} \in \{\pm 1\}^d$ and

$$Z_{\theta,G} = \sum_{\mathbf{x}\in\{\pm 1\}^d} \exp\biggl(\theta \sum_{(u,v)\in E(G)} x_u x_v\biggr),$$

denotes the normalizing constant, also known as *partition function*. Model (2.1) is equivalent to (1.1), where the spin-spin interactions w_{uv} are either equal to 0 or 1; hence the term "simple." The term "zero-field" refers to the fact that all "main-effects" parameters of the spins x_u have been set to zero, and "ferromagnetic" refers to the fact that all spin-spin interactions are non-negative. As discussed in the Introduction, the parameter θ is the inverse temperature but will also be referred to as *signal strength* interchangeably.

2.1. *General results.* A key concept allowing us to quantify the difficulty of testing a graph property $\mathcal{P}$ under the worst possible scenario is the minimax risk. Formally, given data generated from model (2.1) and a property $\mathcal{P}$, testing (1.2) is equivalent to testing $\mathbf{H}_0 : G \in \mathcal{G}_0(\mathcal{P})$ versus $\mathbf{H}_1 : G \in \mathcal{G}_1(\mathcal{P})$ where

$$\mathcal{G}_0(\mathcal{P}) := \bigl\{G \in \mathcal{G}_d \mid \mathcal{P}(G) = 0\bigr\}, \qquad \mathcal{G}_1(\mathcal{P}) := \bigl\{G \in \mathcal{G}_d \mid \mathcal{P}(G) = 1\bigr\}. \tag{2.2}$$

The minimax risk of testing $\mathcal{P}$ is defined as

$$R_n(\mathcal{P}, \theta) := \inf_{\psi}\Bigl[\sup_{G\in\mathcal{G}_0(\mathcal{P})} \mathbb{P}^{\otimes n}_{\theta,G}(\psi = 1) + \sup_{G\in\mathcal{G}_1(\mathcal{P})} \mathbb{P}^{\otimes n}_{\theta,G}(\psi = 0)\Bigr], \tag{2.3}$$

where the infimum is taken over all measurable binary valued test functions ψ, and recall the notation $^{\otimes n}$ for a product measure of n i.i.d. observations. Criteria (2.3) evaluates the sum of the worst possible type I and type II errors under the best possible test function ψ. One can always generate $\psi \sim \mathrm{Ber}(\frac{1}{2})$ independently of the data, which yields a minimax risk equal to 1. In the remainder of this section, we derive upper and lower bounds on the temperature beyond which $R_n(\mathcal{P}, \theta)$ asymptotically equals 1, which implies that asymptotically the best test of $\mathcal{P}$ would be as good as a random guess. Importantly, here and throughout the manuscript, we implicitly assume the high-dimensional regime $d := d(n)$, so that asymptotically $d \to \infty$ as $n \to \infty$.

To formalize our general signal strength bound for combinatorial properties in Ising models, we need several definitions. Similar definitions were previously used by Neykov, Lu and Liu (2016) to understand the limitations of combinatorial inference in Gaussian graphical models. The first definition allows us to measure a graph based pre-distance between edges.

DEFINITION 2.1 (Edge geodesic pre-distance). Let G be a graph and $\{e, e'\}$ be a pair of edges which need not belong to G. The edge geodesic pre-distance is given by

$$d_G(e, e') := \min_{u \in e, v \in e'} d_G(u, v),$$

where $d_G(u, v)$ denotes the geodesic distance[5] between vertices u and v on G. If such a path does not exist, $d_G(e, e') = \infty$.

Here, we use the term *pre-distance* since $d_G(e, e')$ does not obey the triangle inequality. Having defined a pre-distance, we can define edge packing sets and packing numbers.

DEFINITION 2.2 (Packing number). Given a graph $G = (V, E)$ and a collection of edges $\mathcal{C}$ with vertices in V, a r-packing of $\mathcal{C}$ is any subset of edges S, that is, $S \subseteq \mathcal{C}$ such that each pair of edges $e, e' \in S$ satisfy $d_G(e, e') \geq r$. We define the r-packing number:

$$N(\mathcal{C}, d_G, r) = \max\{|S| \mid S \subseteq \mathcal{C}, S \text{ is } r\text{-packing}\},$$

that is, $N(\mathcal{C}, d_G, r)$ is the maximum cardinality of an r-packing set.

A large r-packing number implies that the set $\mathcal{C}$ has a large collection of edges that are far away from each other. Hence the packing number can be understood as a complexity measure of an edge set. The final definition before we state our first result formalizes constructions of graphs belonging to the null and alternative hypothesis and differing in a single edge.

[5]The geodesic distance between u and v is the number of edges on the shortest path connecting u and v.

DEFINITION 2.3 (Null-alternative divider). For a binary graph property $\mathcal{P}$, let $G_0 = ([d], E_0) \in \mathcal{G}_0(\mathcal{P})$. We refer to an edge set

$$\mathcal{C} = \{e_1, \dots, e_m\},$$

as a *null-alternative divider* (or simply *divider* for short) with a null base G_0 if for any $e \in \mathcal{C}$ the graphs $G_e := ([d], E_0 \cup \{e\}) \in \mathcal{G}_1(\mathcal{P})$.

Intuitively, a large divider set $\mathcal{C}$ implies that testing $\mathcal{P}$ is difficult since there exist multiple graphs G_e with which one can confuse the graph G_0. We make this intuition precise in the following.

THEOREM 2.4 (Signal strength general lower bound). *Given a binary graph property $\mathcal{P}$, let $G_0 \in \mathcal{G}_0(\mathcal{P})$, and the set $\mathcal{C}$ be a divider set with a null base G_0. Suppose that $|\mathcal{C}| \to \infty$ asymptotically. If we have*

$$\theta \le \frac{1}{2}\sqrt{\frac{\log N(\mathcal{C}, d_{G_0}, \log\log|\mathcal{C}|)}{n}} \wedge \operatorname{atanh}\biggl(\frac{e^{-2}}{\operatorname{maxdeg}(G_0) + 1}\biggr),$$

then $\liminf_{n\to\infty} R_n(\mathcal{P}, \theta) = 1$.

Theorem 2.4 gives a strategy for obtaining lower bounds on θ using purely combinatorial constructions. Its proof utilizes Dobrushin's comparison theorem (Föllmer (1988)) to bound the χ^2 divergence between Ising measures deferring in a single edge. The second inequality on θ is required to ensure that the system is in a "high-temperature regime" which is where Dobrushin's theorem holds. If one can select a graph G_0 of constant maximum degree, the real obstruction on θ will be given by the entropy term. Theorem 2.4 is reminiscent of Theorem 2.1 of Neykov, Lu and Liu (2016); remarkably, similar constructions can be used to give lower bounds on the signal strength in both the Gaussian and Ising models. Even though the statements of the two results are related, their proofs are vastly different. The proof in the Gaussian case heavily relies on the fact that the partition functions can be evaluated in closed form, which is generally impossible in Ising models. We demonstrate the usefulness of Theorem 2.4 in Section 2.2 where we apply it to a connectivity testing example.

We complement Theorem 2.4 by an upper bound on the inverse temperature θ above which the minimax risk cannot be controlled. The need for such bounds arises due to identifiability issues in Ising models at low temperatures. In such regimes, the model develops long range correlations, that is, even spins which are not neighbors on the graph can become highly correlated. A simple implication of this fact for instance is that it is challenging to tell apart a triangle graph from a vertex with its two disconnected neighbors at low temperatures (see Figure 2). To formalize the statement, we first define a class of graph properties. To this end, recall the distinction between the spanning subgraph and subgraph inclusions $\trianglelefteq$, $\subseteq$ introduced in Section 1.3.

DEFINITION 2.5 (Monotone and strongly monotone properties). A binary graph property $\mathcal{P} : \mathcal{G}_d \mapsto \{0, 1\}$, is called *monotone*, if for any two graphs $G' \trianglelefteq G$ we have $\mathcal{P}(G') \le \mathcal{P}(G)$. A binary property $\mathcal{P}$ is called *strongly monotone* if for any two graphs $G' \subseteq G$ we have $\mathcal{P}(G') \le \mathcal{P}(G)$.

By definition, any strongly monotone property is monotone; however, the converse is not true. An example of a strongly monotone property is the size of the largest clique in a graph. On the other hand, an example of a monotone property which is not strongly monotone is graph connectivity. We now state our result giving an upper bound on θ when testing strongly monotone properties. We have the following.

THEOREM 2.6 (Strongly monotone properties upper bound). *Let $\mathcal{P}$ be a strongly monotone property, and $H_0 \in \mathcal{G}_0(\mathcal{P})$. Assume there exists an $l \times r$ biclique*[6] *B with $r \ge 2$ such that $B \trianglelefteq H_0$. Suppose there are two vertices u, v belonging to the right-hand side of B, so that adding (u, v) to H_0 gives a graph $H_1 \in \mathcal{G}_1(\mathcal{P})$. Let θ satisfy $\theta \ge \frac{2}{l}$ and $\theta \ge \frac{3}{r-2}$ when $r > 2$ or $\theta \ge \log 2$ for $r = 2$. Then if for some $\kappa > 1$, we have*

$$\theta \ge \frac{\log \frac{2\kappa n r}{\log\lfloor d/(l+r)\rfloor}}{l}, \tag{2.4}$$

it holds that $\liminf_{n\to\infty} R_n(\mathcal{P}, \theta) = 1$.

Theorem 2.6 shows how to prove upper bounds on θ using graph constructions. One needs to find a graph H_0 containing a large biclique B, so that adding edges to H_0 transfers it to an alternative graph. The number of "left" vertices l of B appears in (2.4) and, therefore, the larger B is the harder it is to test $\mathcal{P}$ in the worst case. The intuition behind this is as follows. The existence of the biclique B is a measure of how dense H_0 is. The denser H_0 is the harder it is to tell it apart from H_1 when θ is large. On the other hand, the strong monotonicity of $\mathcal{P}$ ensures that if a subgraph H_1 of G satisfies $\mathcal{P}(H_1) = 1$ then $\mathcal{P}(G) = 1$. Therefore, if G contains H_0 as a subgraph it becomes hard to test for $\mathcal{P}$ when the value of θ is large.

We end this section with a result, which shows a simple lower bound on θ for strongly monotone properties. One may use this result in place of Theorem 2.4, when handling strongly monotone properties.

PROPOSITION 2.7 (Strongly monotone properties lower bound). *Let $\mathcal{P}$ be a strongly monotone property, and the graph $H_0 = ([m], E_0) \in \mathcal{G}_0(\mathcal{P})$, be such that*

[6] A complete bipartite graph.

if one adds the edge e to H_0 *the resulting graph* $H_1 = ([m], E_0 \cup \{e\}) \in \mathcal{G}_1(\mathcal{P})$. *Suppose* $\log\lfloor d/m \rfloor \le n$. *Then if*

$$\theta < \operatorname{atanh}\biggl(\sqrt{\frac{\log\lfloor d/m \rfloor}{n}}\biggr), \tag{2.5}$$

we have $\liminf_{n\to\infty} R_n(\mathcal{P}, \theta) = 1$. *Furthermore*, $\liminf_{n\to\infty} R_n(\mathcal{P}, \theta) = 1$ *if* $\log\lfloor d/m \rfloor \gtrsim n$ *for a sufficiently large constant.*

Notice that for positive θ one has $\theta > \tanh(\theta)$ and, therefore, (2.5) implies that $\theta > \sqrt{\frac{\log\lfloor d/m \rfloor}{n}}$ in order for cycle testing to be possible. Examples 2.9 and 2.10 of the following section illustrate how to apply Theorem 2.6 and Proposition 2.7 in practice.

2.2. *Examples*. In this section, we apply Theorems 2.4, 2.6 and Proposition 2.7 to establish necessary conditions on θ for the three examples discussed in the Introduction. In the first example, we derive a lower bound on θ for graph connectivity testing.

EXAMPLE 2.8 (Connectivity). Define "graph connectivity" $\mathcal{P}$ as $\mathcal{P}(G) = 0$ if G is disconnected and $\mathcal{P}(G) = 1$ otherwise. Then if

$$\theta < \kappa\sqrt{\log d/n} \wedge \operatorname{atanh}\bigl(1/\bigl(3e^2\bigr)\bigr), \tag{2.6}$$

we have $\liminf_{n\to\infty} R_n(\mathcal{P}, \theta) = 1$. Furthermore, if $\log d \gtrsim n$ for a sufficiently large absolute constant and if

$$\tanh(\theta) < 1, \tag{2.7}$$

we have $\liminf_{n\to\infty} R_n(\mathcal{P}, \theta) = 1$.

PROOF OF EXAMPLE 2.8. Note that since connectivity is not a strongly monotone property we cannot apply Proposition 2.7, and will use Theorem 2.4 instead. Construct a base graph $G_0 := ([d], E_0)$ where

$$E_0 := \bigl\{(j, j+1)_{j=1}^{\lfloor d/2 \rfloor - 1}, \bigl(\lfloor d/2 \rfloor, 1\bigr), (j, j+1)_{j=\lfloor d/2 \rfloor+1}^{d}, \bigl(\lfloor d/2 \rfloor + 1, d\bigr)\bigr\},$$

and let

$$\mathcal{C} := \bigl\{\bigl(j, \lfloor d/2 \rfloor + j\bigr)_{j=1}^{\lfloor d/2 \rfloor}\bigr\} \qquad \text{(see Figure 1).}$$

Adding any edge from $\mathcal{C}$ to G_0 results in a connected graph, so $\mathcal{C}$ is a divider with a null base G_0. To construct a packing set of $\mathcal{C}$, we collect all edges $(j, \lfloor d/2 \rfloor + j)$ for $j \le \lfloor d/2 \rfloor - \lceil \log\log|\mathcal{C}| \rceil$ satisfying $\lceil \log\log|\mathcal{C}| \rceil$ divides j. This procedure

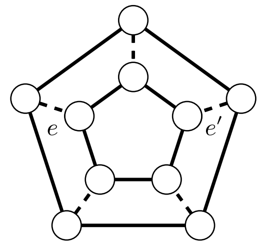

FIG. 1. *The graph G_0 with two edges $e, e' \in \mathcal{C} : d_{G_0}(e, e') = 2$, $d = 10$.*

results in a packing set with radius at least $\lceil \log\log|\mathcal{C}| \rceil$ and cardinality of at least $\lfloor \frac{|\mathcal{C}|}{\lceil \log\log|\mathcal{C}| \rceil} \rfloor - 1$. Therefore,

$$\log N(\mathcal{C}, d_{G_0}, \log\log|\mathcal{C}|) \geq \log\left[\left\lfloor \frac{|\mathcal{C}|}{\lceil \log\log|\mathcal{C}| \rceil} \right\rfloor - 1\right] \asymp \log|\mathcal{C}| \asymp \log d.$$

By Theorem 2.4, we conclude that the asymptotic risk of connectivity testing is 1 if for some absolute constant $\kappa > 0$ we have that (2.6) holds. The second conclusion of this example does not follow directly from our general results. Its proof is deferred to Appendix B. □

EXAMPLE 2.9 (Cycle presence). Consider testing the property $\mathcal{P}$ "cycle presence," that is, $\mathcal{P}(G) = 0$ if G is a forest and $\mathcal{P}(G) = 1$ otherwise. Suppose $\log\lfloor d/3 \rfloor \leq n$. If either

$$\theta < \operatorname{atanh}\left(\sqrt{\log\lfloor d/3 \rfloor / n}\right) \tag{2.8}$$

or

$$\theta \geq 2 \vee \log \frac{4\kappa n}{\log\lfloor d/3 \rfloor}, \tag{2.9}$$

for some absolute constant $\kappa > 0$, we have $\liminf_{n\to\infty} R_n(\mathcal{P}, \theta) = 1$. Furthermore, $\liminf_{n\to\infty} R_n(\mathcal{P}, \theta) = 1$ if for some sufficiently large constant, we have $\log d \gtrsim n$.

PROOF OF EXAMPLE 2.9. By definition, cycle presence is a strongly monotone property. Figure 2 shows an example of a graph H_0 satisfying the conditions of Theorem 2.6 and Proposition 2.7. Concretely, H_0 is a 1×2 biclique which contains no cycle and has the property that adding one edge on its right side gives a graph with a cycle. By Proposition 2.7, we immediately confirm (2.8) and the final conclusion. Furthermore, by a direct application of Theorem 2.6 it follows that if there exists a constant $\kappa > 1$ such that if (2.9) holds the minimax risk of cycle testing is asymptotically 1. □

FIG. 2. *The graph H_0 in the left panel is a triangle with a missing edge. H_0 contains a biclique with no cycles. On the other hand, if we add the dashed edge we obtain a triangle graph H_1, which has a cycle. In terms of the notation of Theorem 2.6, we have $l = 1$ and $r = 2$.*

EXAMPLE 2.10 (Clique size). In our final example, we consider testing the "maximum clique size" property $\mathcal{P}$, where $\mathcal{P}$ is such that $\mathcal{P}(G) = 0$ if G has no m-clique and $\mathcal{P}(G) = 1$ otherwise. Suppose that the maximum degree of G satisfies $\operatorname{maxdeg}(G) \le s$ where s is a known integer such that $m \le s + 1$. Let $\log\lfloor d/m \rfloor \le n$. If either

$$\theta < \operatorname{atanh}\big(\sqrt{\log\lfloor d/m\rfloor / n}\big), \tag{2.10}$$

or

$$\theta \gtrsim \frac{12}{s-9} \vee \frac{\log \frac{\kappa n s}{\log\lfloor 2d/s \rfloor}}{(s-1)/4}, \tag{2.11}$$

for some absolute constant $\kappa > 0$, we have $\liminf_{n\to\infty} R_n(\mathcal{P}, \theta) = 1$. Furthermore, $\liminf_{n\to\infty} R_n(\mathcal{P}, \theta) = 1$ if $\log\lfloor d/m\rfloor \gtrsim n$ for a sufficiently large constant.

PROOF OF EXAMPLE 2.10. Since $\mathcal{P}$ is a strongly monotone property, we can apply Theorem 2.6 and Proposition 2.7 to upper and lower bound θ respectively. We start first with the lower bound. Construct H_0 as an m-clique with a missing edge, as shown in Figure 3. By Proposition 2.7, we immediately deduce (2.10) and the final conclusion of the statement.

The following construction of the graph H_0 from the statement of Theorem 2.6 is inspired by Turán's theorem (e.g., Bollobás (2004)). We build H_0 by taking $\lfloor \frac{s-1}{m-2} \rfloor (m-1) + 1$ vertices, splitting them in $m - 1$ approximately equally sized groups (one group will have 1 more vertex than the others) and connecting any two vertices belonging to different groups; see Figure 4 for a visualization of H_0. It is

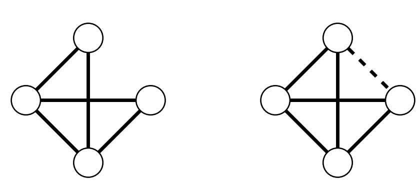

FIG. 3. *For this figure, let $m = 4$. In the left panel, we show an example of a graph H_0, while on the right panel we add one edge to transfer H_0 to H_1 which satisfies the property $\mathcal{P}$.*

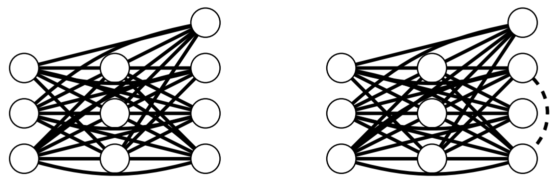

FIG. 4. *In the left panel, we show an example of the graph H_0. The concrete values of s and m used are $s = 7$ and $m = 4$. H_0 contains no 4-clique, and its maximum degree is 7. On the other hand, adding any edge on the rightmost side (such as the dashed edge in the figure on the right) to H_0 results in a graph H_1 which contains a 4-clique, and whose maximum degree remains bounded by 7. The graph H_0 contains a 3×7 biclique, whose left side consists of taking the three leftmost vertices.*

simple to check that H_0 does not contain an m-clique, and adding certain edges to H_0, gives a graph containing an m-clique with maximum degree bounded by s. Furthermore, H_0 contains a $\lfloor \frac{s-1}{m-2} \rfloor \lfloor \frac{m-1}{2} \rfloor \times (\lfloor \frac{s-1}{m-2} \rfloor \lceil \frac{m-1}{2} \rceil + 1)$ biclique; to see this split, the $m-1$ vertex groups into 2 vertex groups: one with all vertices in the first $\lfloor \frac{m-1}{2} \rfloor$ groups, and the other with all remaining vertices.

We are now in a position to apply Theorem 2.6. To render bound (2.4) in a reader friendly form, we use that the terms $\lfloor \frac{s-1}{m-2} \rfloor \lfloor \frac{m-1}{2} \rfloor \geq \frac{s-1}{4}$ and $\lfloor \frac{s-1}{m-2} \rfloor \lceil \frac{m-1}{2} \rceil + 1 \geq \frac{s+3}{4}$. We have that the minimax risk is asymptotically 1 if for any $\kappa > 1/2$ (2.11) holds. □

Notably, the maximum degree s of the graph appears in (2.11) unlike in the previous two examples. The bigger the maximum degree is allowed to be, the smaller the signal θ has to be in order for meaningful clique size tests to exist.

3. Correlation screening for ferromagnets. In this section, we formulate and study the limitations of a greedy correlation screening algorithm on monotone property testing problems. We pay special attention to the examples discussed in Section 2.2. Unlike correlation based decoders, such as the ones studied by Santhanam and Wainwright (2012), this algorithm is designed to directly target the graph property of interest, and also has polynomial runtime for many instances. Moreover, for different properties, the regimes in which the algorithm works differ vastly from graph recovery algorithms. For generality, we expand model class (2.1) to include all zero-field ferromagnetic models such that

$$\mathbb{P}_{\theta, G_{\mathbf{w}}}(\boldsymbol{X} = \mathbf{x}) \propto \exp\biggl(\theta \sum_{(u,v) \in E(G)} w_{uv} x_u x_v\biggr), \tag{3.1}$$

where $\mathbf{x} \in \{\pm 1\}^d$, $\theta \geq 0$, $G_{\mathbf{w}} := (G, \mathbf{w})$ is a weighted graph, and for $(u, v) \in E(G)$: $w_{uv} > 0$. In contrast to (2.1), in (3.1) the weights $\mathbf{w}$ allow for the interactions to have different magnitude.

3.1. *General correlation screening algorithm.* We now define a class of graphs "witnessing" the alternative. For a monotone property $\mathcal{P}$, define the collection of graphs

$$\mathcal{W}(\mathcal{P}) := \{G \in \mathcal{G}_d \mid \mathcal{P}(G) = 1\}.$$[7]

We refer to graphs in $\mathcal{W}(\mathcal{P})$ as *witnesses* of $\mathcal{P}$. It is clear by the monotonicity of $\mathcal{P}$ for two graphs G and G' such that $\mathcal{P}(G) = 0$ and $G' \in \mathcal{W}(\mathcal{P})$ that the set $E(G') \setminus E(G) \neq \varnothing$. Define the sets of weighted graphs

$$\mathcal{G}_0(\mathcal{P}) := \{G_{\mathbf{w}} \mid \mathbf{w} \in \mathbb{R}^{+\binom{d}{2}}, \mathcal{P}(G) = 0\},$$

$$\mathcal{G}_1(\mathcal{P}) := \Big\{G_{\mathbf{w}} \mid \mathbf{w} \in \mathbb{R}^{+\binom{d}{2}}, \mathcal{P}(G) = 1, \max_{G' \in \mathcal{W}(\mathcal{P}), G' \trianglelefteq G} \min_{(u,v) \in E(G')} w_{uv} \geq 1\Big\}.$$

The set $\mathcal{G}_0(\mathcal{P})$ imposes no signal strength restrictions, while $\mathcal{G}_1(\mathcal{P})$ requires the existence of at least one witness of $\mathcal{P}$, each edge of which corresponds to an interaction with magnitude at least θ. For future reference, we omit the dependency on $\mathcal{P}$ if this does not cause confusion. In Section 2.2, we saw that some property tests, such as cycle testing and clique size testing, necessitate further restrictions on their parameters [see (2.9) and (2.11)]. Let $\mathcal{R}$ be an appropriately chosen for the property $\mathcal{P}$ restriction set on the weighted graph pair $G_{\mathbf{w}}$. For instance, an appropriate set $\mathcal{R}$ for cycle testing could be $\mathcal{R} = \{G_{\mathbf{w}} | \|\mathbf{w}\|_\infty \leq \Theta/\theta\}$ for some $\Theta \geq \theta$.

To this end, it is useful to first define the extremal correlation

$$\mathcal{T} := \mathcal{T}(\mathcal{P}, \mathcal{R}, \theta) = \min_{G_{\mathbf{w}} \in \mathcal{G}_1 \cap \mathcal{R}} \max_{G' \in \mathcal{W}} \min_{(u,v) \in E(G')} \mathbb{E}_{\theta, G_{\mathbf{w}}} X_u X_v. \tag{3.2}$$

$\mathcal{T}$ is the maximal smallest possible correlation between neighboring vertices in a witness graph given any model from the alternative. In the following, we give a simple universal lower bound on $\mathcal{T}$.

LEMMA 3.1. *For any monotone property $\mathcal{P}$, we have*

$$\mathcal{T} \geq \tanh(\theta).$$

PROOF OF LEMMA 3.1. Observe that by Griffith's inequality (see Theorem A.2 in Appendix A) deleting any edge can only reduce the correlation between a pair of vertices. Therefore, one can prune the graph G without increasing $\mathcal{T}$, until it becomes a minimal witness W, that is, if we delete any edge from W the resulting graph does not satisfy $\mathcal{P}$. On the graph W, we have $\mathcal{T} \geq \min_{(u,v)\in E(W)} \mathbb{E}_{\theta, W_{\mathbf{w}}} X_u X_v$. Next, one can prune further edges from W until only the minimum edge remains. Since the correlation of a pair of vertices with a graph consisting of the single edge between them is precisely $\tanh(\theta)$ (see Lemma A.7 in Appendix A) the inequality follows. □

[7]This is simply a redefinition of the set $\mathcal{G}_1(\mathcal{P})$ from (2.2).

Since, in practice, $\mathcal{T}$ might be hard to estimate, we assume that we have a lower bound on $\mathcal{T}$: $\underline{\mathcal{T}}$ in closed form (we allow for $\underline{\mathcal{T}} = \mathcal{T}$). Provided that we have sufficiently many samples, and the data is generated under an alternative model, many empirical correlations between neighboring vertices should be approximately at least $\mathcal{T}$ (and hence at least $\underline{\mathcal{T}}$). To formally define the empirical correlations, let $\boldsymbol{X}^{(1)}, \boldsymbol{X}^{(2)}, \dots, \boldsymbol{X}^{(n)} \sim \mathbb{P}_{\theta, G_{\mathbf{w}}}$ be n i.i.d. samples from the ferromagnetic Ising model (3.1). Define the empirical measure $\widehat{\mathbb{P}}$, so that for any Borel set $A \subset \mathbb{R}^d$: $\widehat{\mathbb{P}}(A) = n^{-1} \sum_{i=1}^n \mathbb{1}(\boldsymbol{X}^{(i)} \in A)$. Put $\widehat{\mathbb{E}}$ for the expectation under $\widehat{\mathbb{P}}$. To this end, for a given $\delta > 0$ define the universal threshold

$$\tau := \tau(n, d, \delta) = \sqrt{\frac{4 \log d + \log \delta^{-1}}{n}}, \tag{3.3}$$

and consider the following correlation screening Meta-Algorithm 1 for monotone property testing in ferromagnetic Ising models.

Algorithm 1 Correlation Screening Test

Input: $\{\boldsymbol{X}^{(i)}\}_{i \in [n]}, \theta, \mathcal{R}, \mathcal{P}$
Set $\psi = 0$
Calculate the matrix $\mathbf{M} := \{\widehat{\mathbb{E}} X_u X_v\}_{u,v \in [d]}$
Solve

$$\widehat{G} = \operatorname*{argmax}_{G' \in \mathcal{W}} \min_{(u,v) \in E(G')} M_{uv} \tag{3.4}$$

Set $\psi = 1$ if $\min_{e \in \widehat{G}} M_e > \underline{\mathcal{T}} - \tau$.
return ψ

The only potentially computationally intensive task in Algorithm 1 is optimization (3.4), which aims to find a witness whose smallest empirical correlation is the largest. However, for many properties solving (3.4) can be done in polynomial time via greedy procedures. We remark that step (3.4) treats M_{uv} as a surrogate of θw_{uv}. Instead, one could opt to substitute M_{uv} with an estimate of the parameter θw_{uv}, which can be obtained via a procedure such as ℓ_1-regularized vertex-wise logistic regressions (Ravikumar, Wainwright and Lafferty (2010)), for example. Here, we prefer to focus on correlation screening due to its simplicity, while we recognize that the estimate M_{uv} may not be a good proxy of θw_{uv} in models at low temperature regimes, which are known to develop long range correlations. To this end, define the extremal quantity

$$\mathcal{Q}(\mathcal{P}, \mathcal{R}, \theta) := \max_{G_{\mathbf{w}} \in \mathcal{G}_0 \cap \mathcal{R}} \max_{G' \in \mathcal{W}} \min_{(u,v) \in E(G')} \mathbb{E}_{\theta, G_{\mathbf{w}}} X_u X_v$$

The term $\mathcal{Q}$, selects a weighted graph $G_{\mathbf{w}}$ under the null and a witness G', which yields the largest possible minimal correlation on any of the edges of G'. The following result holds regarding the performance of Algorithm 1.

THEOREM 3.2 (Correlation screening sufficient conditions). *Suppose that* (θ, n, d) *satisfy*

$$\underline{\mathcal{T}} - \mathcal{Q} > 2\tau. \tag{3.5}$$

Then Algorithm 1 satisfies

$$\sup_{G_{\mathbf{w}} \in \mathcal{G}_0 \cap \mathcal{R}} \mathbb{P}_{\theta, G_{\mathbf{w}}}(\psi = 1) \le \delta \quad \textit{and} \quad \sup_{G_{\mathbf{w}} \in \mathcal{G}_1 \cap \mathcal{R}} \mathbb{P}_{\theta, G_{\mathbf{w}}}(\psi = 0) \le \delta. \tag{3.6}$$

Condition (3.5) ensures that the gap between the minimal correlations in models under the null and alternative hypothesis is sufficiently large even in worst case situations. Theorem 3.2 is a straightforward consequence of Hoeffding's inequality, and the real difficulty when applying it is controlling the quantities $\mathcal{T}$ and $\mathcal{Q}$. Recall that Lemma 3.1 showed a simple universal lower bound on $\mathcal{T}$. Below we give two general upper bounds on $\mathcal{Q}$. Given a sparsity level s and a real number Θ, define the ratio

$$R(s, \Theta) := \frac{\cosh(2s\Theta) + 2se^{-2(s-1)\Theta}\cosh(2(s-1)\Theta)}{2se^{-2(s-1)\Theta}\cosh(2\Theta) + 1}.$$

The following holds.

PROPOSITION 3.3 (No edge correlation upper bounds). *Assume that the graph* $G_{\mathbf{w}} \in \mathcal{R}$, *where the restriction set* $\mathcal{R}$ *is* $\mathcal{R} = \{G_{\mathbf{w}} \mid \operatorname{maxdeg}(G) \le s, \|\mathbf{w}\|_\infty \le \Theta/\theta\}$. *Then the following two results hold*:

(i) *Let* $s \ge 3$.[8] *Then*

$$\mathcal{Q} \le \frac{R(s, \Theta) - 1}{R(s, \Theta) + 1}.$$

(ii) *Let* $(s-1)\tanh(\Theta) < 1$. *Then*

$$\mathcal{Q} \le \frac{s\tanh^2(\Theta)}{1 - (s-1)\tanh(\Theta)}.$$

REMARK 3.4. We will now argue that Proposition 3.3(ii) and Lemma 3.1 ensure that Algorithm 1 satisfies (3.6) in the high-temperature regime $s\tanh(\Theta) \lesssim 1$ when the entries of $\mathbf{w}$ are approximately equal. By Lemma 3.1, we have

$$\mathcal{T} \ge \tanh(\theta).$$

Suppose now that $\theta = \Theta$ (equivalently $w_{uv} = 1$ for all nonzero weights). If $s\tanh(\theta) < 1/3$ and $\tanh(\theta) > 4\tau$, by (ii),

$$\mathcal{T} - \mathcal{Q} \ge \tanh(\theta) - \tanh(\theta)/2 > 2\tau.$$

[8] Similar bound on $\mathcal{Q}$ holds for the case $s = 2$. For details, refer to the proof.

Hence when $\tanh(\theta) > 4\tau$, we have that (3.5), and consequently (3.6) hold. More generally, if $\theta \asymp \Theta$, $\tanh(\theta) \geq \Omega(\tau)$ and $s\tanh(\Theta)$ is sufficiently small, implies that Algorithm 1 controls the type I and type II errors. This fact, coupled with the results of Section 2 suggests that correlation screening is optimal up to scalars for many properties in the high-temperature regime (where $s\tanh(\theta)$ is small).

Below we study three specific instances of Algorithm 1 to obtain better understanding of its limitations. Importantly, we observe that the correlation screening test can be constant optimal beyond the high-temperature regime for some properties.

3.2. *Examples*. We now revisit the three examples of Section 2.2.

EXAMPLE 3.5 (Connectivity). Here, we implement the correlation screening algorithm for connectivity testing (see Algorithm 2), and we take the opportunity to contrast property testing to graph recovery. We will argue that correlation screening can test graph connectivity even in graphs of unbounded degree. In contrast, correlation based algorithms fail to learn the structure even in unconnected graphs when the signal strength $\theta \geq \Omega(\frac{1}{s})$, where s denotes the maximum degree of the graph, as argued by Montanari and Pereira (2009).

Algorithm 2 Connectivity Test

Input: $\{X^{(i)}\}_{i\in[n]}$
Set $\psi = 0$
Calculate the matrix $\mathbf{M} := \{\widehat{\mathbb{E}} X_u X_v\}_{u,v\in[d]}$
Estimate $\widehat{T}$ the maximum spanning tree (MST) on $\mathbf{M}$[9]// Equivalent to solving (3.4)
Set $\psi = 1$ if $\min_{e\in\widehat{T}} M_e > \tanh(\theta) - \tau$
return ψ

It is simple to see that $\mathcal{G}_0, \mathcal{G}_1$ reduce to

$$\mathcal{G}_0 := \{G_\mathbf{w} \mid G \text{ is disconnected}\}, \qquad \mathcal{G}_1 := \Big\{G_\mathbf{w} \mid \max_{T \underset{\text{tree}}{\trianglelefteq} G} \min_{(u,v)\in T} w_{uv} \geq 1\Big\},$$

and there are no further parameter restrictions, that is, $\mathcal{R}$ is all weighted graphs. We have the following.

COROLLARY 3.6 (Connectivity). *Assume that* $\tanh(\theta) > 2\tau$. *Then Algorithm 2 satisfies* (3.6).

[9]Finding an MST can be done efficiently.

Corollary 3.6 underscores the difference between property testing and structure learning. Montanari and Pereira (2009) and Santhanam and Wainwright (2012), showed that one cannot recover the graph structure in a ferromagnetic model when the parameter θ exceeds a critical threshold. We also note that the condition $\tanh(\theta) \geq 2\tau$ matches the lower bound prediction (2.6) up to constant terms when τ is sufficiently small.

It is worth mentioning that Algorithm 2 is no longer optimal when $\log d \gtrsim n$ due to lack of concentration. If $\log d \gtrsim n$ for a sufficiently large constant, by (2.7) $\tanh(\theta)$ has to equal 1 asymptotically. It is simple to devise a test that works when $\tanh(\theta) = 1$, namely: reject the null hypothesis if all spins have the same signs through each of the n trials. If the graph is connected this will happen with probability 1; if the graph is disconnected this event happens with probability at most $1/2^n$. Finally, we remark that whether one can devise a finite sample connectivity test when $\log d \gtrsim n$ and $\tanh(\theta) < 1$ remains an open question which merits further investigation.

EXAMPLE 3.7 (Cycle presence). Here, we revisit cycle testing. The sets $\mathcal{G}_0$ and $\mathcal{G}_1$ reduce to

$$\mathcal{G}_0 := \{G_{\mathbf{w}} \mid G \text{ is a forest}\}, \qquad \mathcal{G}_1 := \Big\{G_{\mathbf{w}} \mid \max_{C \underset{\text{cycle}}{\subseteq} G} \min_{(u,v) \in C} \theta_{uv} \geq 1\Big\}.$$

Motivated by (2.9), we take the restriction set as $\mathcal{R} = \{\mathbf{w} | \|\mathbf{w}\|_\infty \leq \Theta/\theta\}$. The correlation screening algorithm for cycle testing is given in Algorithm 3. We have the following corollary of Theorem 3.2.

Algorithm 3 Cycle Test

Input: $\{X^{(i)}\}_{i\in[n]}, \theta$
Set $\psi = 0$
Calculate the matrix $\mathbf{M} := \{\widehat{\mathbb{E}} X_u X_v\}_{u,v\in[d]}$
Add edges with weights from $\mathbf{M}$ from high to low until a cycle $\widehat{C}$ emerges[10]// that is, solve (3.4)
Set $\psi = 1$ if $\min_{e\in\widehat{C}} M_e > \tanh(\theta) - \tau$
return ψ

COROLLARY 3.8 (Cycle presence). *Assume that* $\tanh(\theta) - \tanh^2(\Theta) > 2\tau$. *Then Algorithm* 3 *satisfies* (3.6).

Below we derive a more direct result for the special case when $\theta \equiv \Theta$.

[10]Finding the cycle $\widehat{C}$ takes at most d steps, and can be done efficiently.

COROLLARY 3.9 (Cycle presence $\theta = \Theta$). *Suppose $\theta = \Theta$. When τ is sufficiently small, if*

$$\tau \lesssim \theta \lesssim \log(1/\tau), \tag{3.7}$$

Algorithm 3 satisfies (3.6).

PROOF OF COROLLARY 3.9. In this setting, the condition of Corollary 3.8 reduces to the quadratic inequality $t - t^2 > 2\tau$, where we put $t := \tanh(\theta)$ for brevity. Equivalently, the feasible values of θ satisfy

$$\frac{1 - \sqrt{1 - 8\tau}}{2} \le t \le \frac{1 + \sqrt{1 - 8\tau}}{2}.$$

To make the calculation more accessible, we will now use the notation $f(x) \approx g(x)$ in the sense that $\lim_{x \downarrow 0} \frac{f(x)}{g(x)} = 1$. When τ is sufficiently small, it is simple to check that $\frac{1-\sqrt{1-8\tau}}{2} \approx 2\tau$ and $\frac{1+\sqrt{1-8\tau}}{2} \approx 1 - 2\tau$. It therefore follows that Algorithm 3 is successful when

$$2\tau \approx \operatorname{atanh}(2\tau) \le \theta \le \operatorname{atanh}(1 - 2\tau) \approx \log(1/\tau)/2. \qquad \square$$

Up to scalars, (3.7) agrees with bounds (2.8) and (2.9) given in Section 2.2. An alternative correlation based cycle test which works for general models is given in Section 4.3.

EXAMPLE 3.10 (Clique size). Finally, we revisit clique size testing. The parameter sets reduce to

$$\mathcal{G}_0 := \{G_\mathbf{w} \mid G \text{ has no } m\text{-clique}\}, \qquad \mathcal{G}_1 := \Big\{G_\mathbf{w} \mid \max_{\substack{C \subseteq G \\ m\text{-clique}}} \min_{(u,v) \in C} w_{uv} \ge 1\Big\},$$

and let the restriction set $\mathcal{R} = \{G_\mathbf{w} \mid \|\mathbf{w}\|_\infty \le \Theta/\theta, \operatorname{maxdeg}(G) \le s\}$, where $2 \le m \le s + 1$. We summarize the correlation screening implementation for clique size testing in Algorithm 4. To this end, for a $Z \sim N(0, 1)$ define

$$r(m, \theta) := \frac{e^{2\theta} \mathbb{E} \cosh^{m-2}(\sqrt{\theta} Z + 2\theta)}{\mathbb{E} \cosh^{m-2}(\sqrt{\theta} Z)}.$$

The following holds.

[11] In this footnote, we show an example of an algorithm for checking for m-clique presence. When a new edge (u, v) is added, walk over common neighbors of both u and v, and check for an m-clique. There are at most $\frac{ds}{2}$ steps and at each step we have to check at most $\binom{s-1}{m-2}$ m-cliques, giving a runtime bound of $O(ds(s + m^2\binom{s-1}{m-2}))$.

Algorithm 4 Clique Size Test

Input: $\{X^{(i)}\}_{i\in[n]}, \theta$
Set $\psi = 0$
Calculate the matrix $\mathbf{M} := \{\widehat{\mathbb{E}} X_u X_v\}_{u,v\in[d]}$
Add edges with weights from $\mathbf{M}$ from high to low until an m-clique $\widehat{C}$ emerges[11]// that is, solve (3.4)
Set $\psi = 1$ if $\min_{e\in\widehat{C}} M_e > \frac{r(m,\theta)-1}{r(m,\theta)+1} - \tau$
return ψ

COROLLARY 3.11 (Clique size). *For $G_{\mathbf{w}} \in \mathcal{G}_1 \cap \mathcal{R}$, we have $\mathcal{T} = \frac{r(m,\theta)-1}{r(m,\theta)+1}$. Hence if $(s, m, d, n, \theta, \Theta)$ are such that either*

$$\mathcal{T} - \frac{R(s,\Theta)-1}{R(s,\Theta)+1} \geq 2\tau,$$

or

$$\mathcal{T} - \frac{s\tanh^2(\Theta)}{1-(s-1)\tanh(\Theta)} \geq 2\tau \quad \text{and} \quad (s-1)\tanh(\Theta) < 1,$$

Algorithm 4 satisfies (3.6).

Below we derive a more direct result for the special case when $\theta \equiv \Theta$.

COROLLARY 3.12 (Clique size $\theta = \Theta$). *Suppose $\theta = \Theta$ and that τ and $\frac{1}{s}$ are sufficiently small. Then if*

$$\tau \lesssim \theta \lesssim \frac{1}{s}, \tag{3.8}$$

Algorithm 4 satisfies (3.6). Next, suppose that $m = s+1$. If $e^{s\theta} \gg s$ and

$$\theta \leq \frac{\log(2/\tau)}{4(s-1)}, \tag{3.9}$$

Algorithm 4 satisfies (3.6).

PROOF OF COROLLARY 3.12. In Remark 3.4, we already argued that if $4\tau \leq \tanh(\theta) \leq \frac{1}{3s}$, Algorithm 4 controls the type I and type II errors. Since $\tanh(x) \approx x$ when x is small inequality (3.8) follows. To show the second part, we resort to the first bound of Corollary 3.11. Since the proof is more involved, we show it in Remark C.4 of Appendix C. $\square$

From (3.8), it follows that in the high-temperature regime Algorithm 4 matches the lower bound (2.10). Furthermore, note that (3.9) matches bound (2.11) up to scalars. Hence the special case of clique size testing when $m = s+1$ is yet another example confirming that correlation screening can be useful for property testing even at low temperatures.

4. Results for general models. So far we studied model classes admitting only ferromagnetic interactions. The situation drastically changes if one considers the general class of zero-field Ising models, which includes models with antiferromagnetic, that is, negative interactions.

4.1. *Minimax bounds.* The main result of this section is an impossibility theorem, which shows that testing strongly monotone properties over the general class of models requires boundedness of a certain maximum functional of the property. We also argue more specifically, that unlike in the ferromagnetic case, connectivity testing is not feasible at low temperatures over the general case unless the degree of the graph is bounded. Both of these results sharply contrast what we have seen in the previous sections of the paper.

Concretely, we will work with the simple zero-field models specified by the parameters $\theta > 0$ and $\mathbf{w} \in \{\pm 1\}^{\binom{d}{2}}$ as

$$\mathbb{P}_{\theta, G_{\mathbf{w}}}(\boldsymbol{X} = \mathbf{x}) \propto \exp\biggl(\theta \sum_{(u,v)\in E(G)} w_{uv} x_u x_v\biggr). \tag{4.1}$$

Expression (4.1) has more degrees of freedom compared to (2.1) since the spin-spin interactions in (4.1) are allowed to be negative. Intuitively, interactions corresponding to $w_{uv} = -1$ have a "repelling" effect on the corresponding spins u and v, whereas interactions with $w_{uv} = 1$ have an "attracting" effect.

Given a monotone property $\mathcal{P}$, recall definition (2.2) of the collections of graphs $\mathcal{G}_0(\mathcal{P}), \mathcal{G}_1(\mathcal{P})$ (below we omit the dependence on $\mathcal{P}$). Let $\mathcal{R}$ be a suitable restriction set on the graph G. We redefine the minimax risk to reflect the model class expansion as follows. Let

$$\begin{aligned} &R_n(\mathcal{P}, \mathcal{R}, \theta) \\ &\quad := \inf_{\psi}\Bigl[\sup_{\mathbf{w}} \sup_{G\in\mathcal{G}_0\cap\mathcal{R}} \mathbb{P}^{\otimes n}_{\theta, G_{\mathbf{w}}}(\psi = 1) + \sup_{\mathbf{w}} \sup_{G\in\mathcal{G}_1\cap\mathcal{R}} \mathbb{P}^{\otimes n}_{\theta, G_{\mathbf{w}}}(\psi = 0)\Bigr], \end{aligned} \tag{4.2}$$

where $\mathbb{P}^{\otimes n}_{\theta, G_{\mathbf{w}}}$ denotes the product measure of n i.i.d. observations of (4.1) and the supremum on $\mathbf{w}$ is taken over the set $\{\pm 1\}^{\binom{d}{2}}$. Armed with this new definition, we have the following.

THEOREM 4.1 (Strongly monotone properties general lower bound). *Assume $G_{\mathbf{w}}$ belongs to the restriction set $\mathcal{R}_s = \{G_{\mathbf{w}} \mid \operatorname{maxdeg}(G) \le s\}$, where $s = o(\sqrt{d})$. Suppose that the strongly monotone property $\mathcal{P}$ satisfies $\mathcal{P}(\varnothing) = 0$ and $\mathcal{P}(C_s) = 1$, where C_s denotes an s-clique graph. Then if for some small $\varepsilon > 0$,*

$$\frac{s \log d/s^2}{n} > 2 + \varepsilon \quad \textit{and} \quad \frac{s \log d/(2s)}{n \log \sqrt{2s}} \ge 1 + \varepsilon, \tag{4.3}$$

we have $\liminf_{n\to\infty} \inf_{\theta \ge 0} R_n(\mathcal{P}, \mathcal{R}_s, \theta) = 1$.

REMARK 4.2. We would like to contrast our result with similar known bounds such Theorem 1 of Santhanam and Wainwright (2012) and Theorem 1 of Bresler, Mossel and Sly (2008). The key differences between our result and these known bounds are that, first (4.3) is valid for property testing and even more generally for a certain detection problem (see the proof in Section D of the supplement for more details), while previous results are valid only for structure recovery; and, therefore, second, the worst cases are very different. In fact, both previously known bounds remain valid in the smaller class of ferromagnetic models, while as we saw in Section 3, some strongly monotone property tests such as cycle presence do not exhibit such limitations.

Note that for any nonconstant strongly monotone $\mathcal{P}$ one has $\mathcal{P}(\varnothing) = 0$. Further, the requirement that $\mathcal{P}$ holds true on C_s is mild, since for any non-zero strongly monotone $\mathcal{P}$ one can always find a sufficiently large s for which $\mathcal{P}$ is satisfied. The only true restriction of Theorem 4.1 on $\mathcal{P}$ is thus that one has to be able to find s in the sparse regime $s \ll \sqrt{d}$.

Loosely speaking, Theorem 4.1 shows that when the quantity $\frac{s \log d/s}{n}$ (up to a log factor) is large, strongly monotone property testing over the model class (4.1) is very difficult in the sparse regime when $s \ll \sqrt{d}$. What is more, this statement remains valid regardless of the magnitude of the signal strength parameter $\theta \geq 0$. This contrasts sharply with our results in the ferromagnetic case, where we have already seen an example which did not require such a condition. Take the cycle testing example in Section 3.7. In this example, if s denotes the maximum degree of the graph, we can always take an s-clique C_s (which certainly contains a cycle), and hence s has to satisfy (4.3) in order for tests with reasonable minimax risk (4.2) to exist. In contrast, Corollary 3.8 shows that controlling the minimax risk is possible without requirements on the maximum degree of the graph. Theorem 4.1 shows that this is no longer the case over the broader model class (4.1).

Theorem 4.1 sheds some light on the complexity involved in testing within model class (4.1). However, it also leaves something to be desired, namely it does not quantify the effect θ has, and it does not address specific properties which may potentially exhibit different complexity. Moreover, it only applies to strongly monotone properties and not to all monotone properties, and thus in particular it does not apply to connectivity testing. Below we give an explicit upper bound on the parameter θ for connectivity testing within the model class (4.1). We show a particularly hard case for connectivity testing in Figure 5 and include a brief explanation in its caption.

PROPOSITION 4.3 (Connectivity testing general upper bound). *Let $\mathcal{P}$ be graph connectivity. Assume $G_{\mathbf{w}}$ belongs to the restriction set $\mathcal{R}_s = \{G_{\mathbf{w}} \mid \operatorname{maxdeg}(G) \leq s\}$. Let s, n, d be sufficiently large, and suppose $\theta \geq \frac{3}{2\lfloor s/4 \rfloor - 2}$ and*

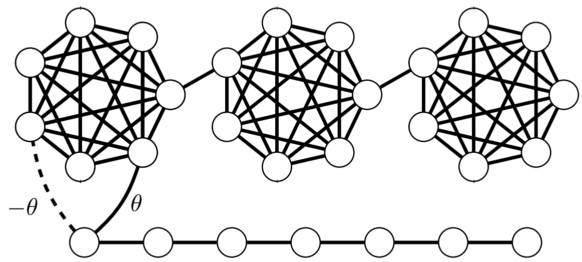

FIG. 5. *Testing connectivity on a model with a connected graph as above is difficult. Solid edges correspond to positive interactions with magnitude θ, while the dashed edge corresponds to a negative interaction with magnitude $-\theta$. The cliques are of size s so that the total degree remains at most s. When the value of θ is large, the majority of the spins in each clique tend to have the same sign. Hence the two interactions of the leftmost clique with the leftmost node in the path graph are "likely" to cancel out, which will make it hard to tell this graph from the disconnected graph consisting of the connected cliques and the path graph. We exploit this construction in the proof of Proposition* 4.3.

there exists a $\kappa > 1$ *so that*

$$\theta > \frac{2\log\frac{\kappa s n}{\log(ds)}}{s-16}. \tag{4.4}$$

Then the minimax risk (4.2) *satisfies* $\liminf_{n\to\infty} R_n(\mathcal{P}, \mathcal{R}_s, \theta) = 1$.

Importantly, condition (4.4) implies that the maximum degree cannot be too large with respect to the other parameters if we hope for a connectivity test with a good control over both type I and type II errors to exist. Recall that no such conditions were needed in Corollary 3.6 when testing connectivity in ferromagnets.

4.2. *Correlation testing for general models.* Section 4.1 made it apparent that property testing is more challenging in the enlarged model space (4.1). In this section, we work with an even larger class of zero-field models compared to (4.1) which are specified as

$$\mathbb{P}_{\theta,G_{\mathbf{w}}}(\boldsymbol{X}=\mathbf{x}) \propto \exp\biggl(\theta \sum_{(u,v)\in E(G)} w_{uv}x_u x_v\biggr), \tag{4.5}$$

where $1 \le |w_{uv}| \le \Theta/\theta$, $(u,v) \in E(G)$ are unknown parameters and $\mathbf{x} \in \{\pm 1\}^d$. For a monotone property $\mathcal{P}$, define the sets of weighted graphs

$$\mathcal{G}_0(\mathcal{P},\theta,\Theta) := \{G_{\mathbf{w}} \mid 1 \le |w_{uv}| \le \Theta/\theta, \mathcal{P}(G) = 0\},$$
$$\mathcal{G}_1(\mathcal{P},\theta,\Theta) := \{G_{\mathbf{w}} \mid 1 \le |w_{uv}| \le \Theta/\theta, \mathcal{P}(G) = 1\}.$$

Different from Section 3, here we impose signal strength restrictions even in the null set $\mathcal{G}_0(\mathcal{P},\Theta)$ and leave the more general setting for future work. We note that one can no longer rely on the screening algorithms of Section 3 to perform a property test for data generated by (4.5); the success of correlation screening hinges on the fact that in ferromagnets deleting edges reduces correlations, which no longer holds in the model class (4.5). One alternative would be to perform exact structure recovery, and check whether the graph property in question holds on the desired graph. Possibilities of exact graph recovery include methods developed in (Bresler, Mossel and Sly (2008), Ravikumar et al. (2011), Santhanam and Wainwright (2012), Anandkumar et al. (2012), Bresler (2015)).

Below we take a different route, and modify the correlation decoders of Santhanam and Wainwright (2012) by specializing them to property testing. Specifically, we consider a score test type of approach, which only involves model fitting assuming the null hypothesis holds. Suppose there exists an algorithm $\mathcal{A}$ mapping the data input as

$$\mathcal{A}(\{\boldsymbol{X}^{(i)}\}_{i\in[n]}, \theta, \Theta, \mathcal{P}) \mapsto \widetilde{G}_{\widetilde{\mathbf{w}}}, \tag{4.6}$$

so that the output $\widetilde{G}_{\widetilde{\mathbf{w}}} \in \mathcal{G}_0(\mathcal{P}, \theta, \Theta)$ and in addition if the true underlying graph G satisfies $\mathcal{P}(G) = 0$ then

$$\max_{u,v\in[d]} |\widehat{\mathbb{E}} X_u X_v - \mathbb{E}_{\theta,\widetilde{G}_{\widetilde{\mathbf{w}}}} X_u X_v| \leq \varepsilon(\delta), \tag{4.7}$$ [12]

holds with probability at least $1 - \delta$. Define the test

$$\psi_\rho(\{\boldsymbol{X}^{(i)}\}_{i\in[n]}, \theta, \widetilde{G}_{\widetilde{\mathbf{w}}}) := \mathbb{1}\Big(\max_{u,v} |\widehat{\mathbb{E}} X_u X_v - \mathbb{E}_{\theta,\widetilde{G}_{\widetilde{\mathbf{w}}}} X_u X_v| \geq \rho\Big). \tag{4.8}$$

Recall the definition of the threshold τ (3.3) and let

$$\mathcal{T} := \mathcal{T}(\theta, \Theta, s) = \frac{\sinh^2(\theta/4)}{2s\Theta(3\exp(2s\Theta) + 1)}. \tag{4.9}$$

The following holds.

PROPOSITION 4.4 (General tests sufficient conditions). *Suppose that $\mathcal{A}$ is an algorithm satisfying* (4.7), *and* $(s, n, d, \theta, \Theta)$ *are such that*

$$\mathcal{T} \geq \tau + \varepsilon(\delta), \tag{4.10}$$

for a small $\delta > 0$. Then the test $\psi_{\varepsilon(\delta)}$ given in (4.8) *satisfies*

$$\sup_{G_{\mathbf{w}}\in\mathcal{G}_0(\mathcal{P},\theta,\Theta)} \mathbb{P}(\psi_{\varepsilon(\delta)} = 1) \leq \delta \quad \textit{and} \quad \sup_{G_{\mathbf{w}}\in\mathcal{G}_1(\mathcal{P},\theta,\Theta)} \mathbb{P}(\psi_{\varepsilon(\delta)} = 0) \leq \delta. \tag{4.11}$$

[12]Here, $\widehat{\mathbb{E}}$ is the empirical expectation defined in Section 3, while $\mathbb{E}_{\theta,\widetilde{G}_{\widetilde{\mathbf{w}}}}$ is the expectation with respect to the measure $\mathbb{P}_{\theta,\widetilde{G}_{\widetilde{\mathbf{w}}}}$ from (4.5).

A key component for the existence of a successful test (4.8) is the algorithm $\mathcal{A}$ satisfying condition (4.7). One example how to construct such an algorithm, is to solve the following optimization problem:

$$\mathcal{A}(\{\boldsymbol{X}^{(i)}\}_{i\in[n]}, \theta, \Theta, \mathcal{P}) = \underset{G_{\mathbf{w}}\in\mathcal{G}_0(\mathcal{P},\theta,\Theta)}{\operatorname{argmin}} \max_{u,v\in[d]} |\widehat{\mathbb{E}}X_u X_v - \mathbb{E}_{\theta,G_{\mathbf{w}}} X_u X_v|. \tag{4.12}$$

In the following lemma, we show that (4.7) indeed holds.

LEMMA 4.5 (Algorithm (4.12) sufficient condition). *If the algorithm $\mathcal{A}$ is defined by* (4.12), *then* (4.7) *holds with $\varepsilon(\delta) = \tau$.*

The proof of Lemma 4.5 is a direct consequence of Hoeffding's inequality. By combining the statements of Proposition 4.4 and Lemma 4.5, we arrive at an abstract generic property test summarized in Algorithm 5.

Algorithm 5 Generic Property Test

Input: $\{\boldsymbol{X}^{(i)}\}_{i\in[n]}, \theta, \Theta, \mathcal{P}$
Calculate the matrix $\{\widehat{\mathbb{E}}X_u X_v\}_{u,v\in[d]}$
Solve $\widetilde{G}_{\widetilde{\mathbf{w}}} = \operatorname{argmin}_{G_{\mathbf{w}}\in\mathcal{G}_0(\mathcal{P},\theta,\Theta)} \max_{u,v\in[d]} |\widehat{\mathbb{E}}X_u X_v - \mathbb{E}_{\theta,G_{\mathbf{w}}} X_u X_v|$
Output $\psi_\tau(\{\boldsymbol{X}^{(i)}\}_{i\in[n]}, \theta, \widetilde{G}_{\widetilde{\mathbf{w}}})$

A sufficient condition for Algorithm 5 to satisfy (4.11) is $\mathcal{T} \geq 2\tau$, where $\mathcal{T}$ is defined in (4.9). One potential problem with Algorithm 5 is that solving (4.12) in general requires combinatorial optimization, which will likely result in nonpolynomial runtime complexity for most properties. However, unlike the structure learning procedure of Santhanam and Wainwright (2012) which also requires combinatorial optimization, Algorithm 5 has the advantage that it does not need to optimize over the entire set of graphs $\mathcal{G}_d$ but only over the smaller set $\{G \in \mathcal{G}_d | \mathcal{P}(G) = 0\}$. We conclude this section by proposing a custom variant of this algorithm specialized to cycle testing which uses a different algorithm $\mathcal{A}$ and can be ran in polynomial time.

4.3. *A computationally efficient cycle test.* In this section, we propose an efficient algorithm $\mathcal{A}$ satisfying (4.7) for cycle testing. Having computationally efficient algorithms for cycle testing is beneficial in practice, since if we have enough evidence that the graph is a forest, we can recover its structure efficiently (Chow and Liu (1968)). We summarize the algorithm $\mathcal{A}$ called "cycle test map" in Algorithm 6.

Algorithm 6 Cycle Test Map

Input: $\{X^{(i)}\}_{i\in[n]}, \theta, \Theta$
Calculate the matrix $\mathbf{M} := \{\widehat{\mathbb{E}} X_u X_v\}_{u,v\in[d]}$
Find a MST $\widetilde{T}$ with edge weights $|M_{uv}|$
for $1 \le u < v \le d$ **do**
  if $(u, v) \notin E(\widetilde{T})$ or $|M_{uv}| < \tanh(\theta) - \tau$ **then**
    $\widetilde{w}_{uv} \leftarrow 0$; $E(\widetilde{T}) \leftarrow E(\widetilde{T}) \setminus \{(u, v)\}$
  else
    $\widetilde{w}_{uv} \leftarrow \operatorname{sign}(M_{uv})((\operatorname{atanh}(|M_{uv}|) \wedge \Theta) \vee \theta)/\theta$
  end if
end for
return $\widetilde{T}_{\widetilde{\mathbf{w}}}$

Given the output $\widetilde{T}_{\widetilde{\mathbf{w}}}$ of Algorithm 6, evaluating the expectations $\mathbb{E}_{\theta,\widetilde{T}_{\widetilde{\mathbf{w}}}} X_u X_v$ needed in (4.8) can be done in polynomial time via the simple formula

$$\mathbb{E}_{\theta,\widetilde{T}_{\widetilde{\mathbf{w}}}} X_u X_v = \prod_{(k,\ell)\in\mathcal{P}^{\widetilde{T}}_{u\to v}} \mathbb{E}_{\theta,\widetilde{T}_{\widetilde{\mathbf{w}}}} X_k X_\ell = \prod_{(k,\ell)\in\mathcal{P}^{\widetilde{T}}_{u\to v}} \tanh(\theta \widetilde{w}_{k\ell}),$$ [13]

where $\mathcal{P}^{\widetilde{T}}_{u\to v}$ denotes the path between vertices u and v in the forest $\widetilde{T}$. Next, we show the validity of the test in (4.8).

PROPOSITION 4.6 (Fast cycle test sufficient conditions). *Suppose that* $\tanh(\theta)(1 - \tanh(\Theta)) > 2\tau$. *Then the output of Algorithm* 6 *satisfies* (4.7) *with*

$$\varepsilon(\delta) = \tau \frac{2 - \tanh(\Theta)}{1 - \tanh(\Theta)}. \tag{4.13}$$

By combining Propositions 4.4 and 4.6, we immediately conclude that if $\mathcal{T} \ge \tau \frac{3-2\tanh(\Theta)}{1-\tanh(\Theta)}$, and the constraints of Proposition 4.6 hold, using the output $\widetilde{T}_{\widetilde{\mathbf{w}}}$ of Algorithm 6 with the test $\psi_{\varepsilon(\delta)}$ of (4.8) with $\varepsilon(\delta)$ as in (4.13), satisfy (4.11).

5. Strongly monotone properties proofs from Section 2. In this section, we give the proofs of the general results on strongly monotone properties: Theorem 2.6 and Proposition 2.7. Other proofs from Section 2 including the proof of Theorem 2.4 can be found in Appendix B. Since the signal strength is uniformly equal to θ in all measures that we consider in this section, we will suppress the dependency on θ whenever that does not cause confusion. For the convenience of the reader, below is a definition of χ^2-divergence which we use in the proofs.

[13]The validity of this formula follows by Proposition A.6 and Lemma A.7 which can be found in the supplement.

DEFINITION 5.1 (χ^2-divergence). For two measures $\mathbb{P}$ and $\mathbb{Q}$ satisfying $\mathbb{P} \ll \mathbb{Q}$, the χ^2-divergence is defined by

$$D_{\chi^2}(\mathbb{P}, \mathbb{Q}) = \mathbb{E}_{\mathbb{Q}}\left(\frac{d\mathbb{P}}{d\mathbb{Q}} - 1\right)^2.$$

Before we prove Theorem 2.6, we state a key lemma whose proof is given in Appendix B.

LEMMA 5.2 (Low temperature bound). *Let $G = (V, E)$ be a graph, and $k, \ell \in V$ be vertices such that $(k, \ell) \notin E$. Let $r \geq 2$ and l be integers such that there exists an $l \times r$ biclique $B \trianglelefteq G$, containing k and ℓ on its right (i.e., r side). Then for values of $\theta \geq \frac{2}{l}$, $\theta \geq \frac{3}{r-2}$ for $r > 2$ and $\theta \geq \log 2$ when $r = 2$ we have*

$$\mathbb{E}_G X_k X_\ell \geq 1 - \frac{2(r-1)}{\exp(\theta l) + r - 1}.$$

PROOF OF THEOREM 2.6. First, we point out a simple implication of (2.4). If $\theta \geq \frac{\log \frac{2\kappa n r}{\log\lfloor d/(l+r)\rfloor}}{l}$, then certainly

$$\theta \geq \frac{\log[\frac{2\kappa n(r-1)}{\log\lfloor d/(l+r)\rfloor} - (r-1)]}{l}. \tag{5.1}$$

We will now argue that even if the above holds the asymptotic minimax risk is still 1. Note that since $B \trianglelefteq H_0$ we have $|V(H_0)| = |V(B)| = l + r$. Consider a graph G_0 based on the union of $m = \lfloor d/(l + r)\rfloor$ disconnected copies of the graph H_0 (see Figure 6). Let $e_k = (u_k, v_k)$ be the local copy of the edge $e = (u, v)$ in the kth copy of H_0. Since the property $\mathcal{P}$ can be represented as a maximum over subgraphs of G_0, by the assumption of the theorem $\mathcal{P}(G_0) = 0$, while adding the edge e_k for any $k \in [m]$ to the kth copy of H_0 transfers G_0 to a graph G_k satisfying $\mathcal{P}(G_k) = 1$. For each graph in $i \in \{0, 1, \dots, m\}$, let $\mathbb{P}_i$[14] be the measure corresponding to Ising models with graphs $\{G_0, G_1, \dots, G_m\}$, and $\mathbb{E}_k$ be the corresponding expectation under $\mathbb{P}_k$. Define the mixture measure $\overline{\mathbb{P}}^{\otimes n} = \frac{1}{m}\sum_{k\in[m]} \mathbb{P}_k^{\otimes n}$. Using Lemma A.4,

$$D_{\chi^2}(\overline{\mathbb{P}}^{\otimes n}, \mathbb{P}_0^{\otimes n}) + 1 \leq \frac{1}{m^2} \sum_{j,k\in[m]} (1 + \tanh(\theta)[\mathbb{E}_j X_{u_k} X_{v_k} - \mathbb{E}_0 X_{u_k} X_{v_k}])^n.$$

By Proposition A.6, since the copies of H_0 are disconnected, for $k \neq j$ we have

$$\mathbb{E}_j X_{u_k} X_{v_k} - \mathbb{E}_0 X_{u_k} X_{v_k} = 0,$$

[14] That is, $\mathbb{P}_i$ is a shorthand for $\mathbb{P}_{\theta, G_i}$ as defined in (2.1).

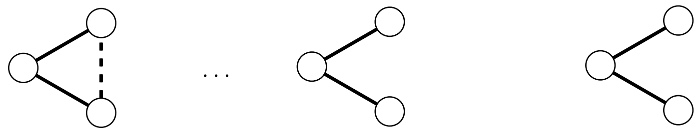

FIG. 6. *For this figure, let $\mathcal{P}$ be cycle testing. The figure shows an example of $d/3$ incomplete triangle graphs (G_0), and takes a mixture of distributions adding one edge to complete the triangles one at a time.*

while by Lemma 5.2 if $k \equiv j$

$$\mathbb{E}_j X_{u_k} X_{v_k} - \mathbb{E}_0 X_{u_k} X_{v_k} \le 1 - \mathbb{E}_0 X_{u_k} X_{v_k} \le \frac{2(r-1)}{\exp(\theta l) + r - 1}.$$

We conclude that

$$\begin{aligned} D_{\chi^2}(\overline{\mathbb{P}}^{\otimes n}, \mathbb{P}_0^{\otimes n}) + 1 &\le \frac{m-1}{m} + \frac{1}{m}\left(1 + \tanh(\theta)\frac{2(r-1)}{\exp(\theta l) + r - 1}\right)^n \\ &\le \frac{m-1}{m} + \frac{1}{m}\left(1 + \frac{2(r-1)}{\exp(\theta l) + r - 1}\right)^n \\ &\le \frac{m-1}{m} + \frac{\exp(\frac{2n(r-1)}{\exp(\theta l)+r-1})}{m}. \end{aligned}$$

Using (5.1), a simple calculation shows that

$$\limsup_{n\to\infty} D_{\chi^2}(\overline{\mathbb{P}}^{\otimes n}, \mathbb{P}_0^{\otimes n}) = 0.$$

Recall that by Le Cam's lemma we have the bound

$$R_n(\mathcal{P}, \theta) \ge \inf_{\psi}[\mathbb{P}_0^{\otimes n}(\psi = 1) + \overline{\mathbb{P}}^{\otimes n}(\psi = 0)] \ge 1 - \frac{1}{2}\sqrt{D_{\chi^2}(\overline{\mathbb{P}}^{\otimes n}, \mathbb{P}_0^{\otimes n})}, \tag{5.2}$$

which completes the proof. □

Below we prove Proposition 2.7. The proof utilizes a similar construction to the one used in the proof of Theorem 2.6.

PROOF OF PROPOSITION 2.7. Construct a null graph G_0 by repeating H_0 $\lfloor d/m \rfloor$ times. Let $\overline{\mathbb{P}}^{\otimes n}$ be the mixture of measures $\mathbb{P}_j^{\otimes n}$, where $\mathbb{P}_j$ is the measure corresponding to adding an edge to one of the "clones" of H_0, thus transferring G_0 to a graph G_j such that $\mathcal{P}(G_j) = 1$, and let $\mathbb{P}_0$ be the Ising measure under the uniform signal model with graph G_0. Let $\mathbb{E}_j$ and $\mathbb{E}_0$ be the expectations under $\mathbb{P}_j$ and $\mathbb{P}_0$, respectively. Let $e_j = (u_j, v_j) = E(G_j) \setminus E(G_0)$ be the edge that distinguishes G_j from G_0. An example of G_0 and one of the alternative graphs G_j is given on Figure 6.

Using Lemma A.4, we can evaluate the divergence D_{χ^2}:

$$D_{\chi^2}(\overline{\mathbb{P}}^{\otimes n}, \mathbb{P}_0^{\otimes n}) = \frac{1}{\lfloor d/m \rfloor^2} \sum_{j,k \in [\lfloor d/m \rfloor]} (1 + \tanh(\theta)[\mathbb{E}_j X_{u_k} X_{v_k} - \mathbb{E}_0 X_{u_k} X_{v_k}])^n - 1$$

$$= \frac{1}{\lfloor d/m \rfloor}(1 + \tanh(\theta)[\mathbb{E}_j X_{u_j} X_{v_j} - \mathbb{E}_0 X_{u_j} X_{v_j}])^n - \frac{1}{\lfloor d/m \rfloor}.$$

By Lemma 4 of Shanmugam et al. (2014), we know that $\mathbb{E}_j X_{u_j} X_{v_j} - \mathbb{E}_0 X_{u_j} X_{v_j} \leq \tanh(\theta)$. Therefore,

$$D_{\chi^2}(\overline{\mathbb{P}}^{\otimes n}, \mathbb{P}_0^{\otimes n}) \leq \frac{1}{\lfloor d/m \rfloor}(1 + \tanh^2(\theta))^n \leq \frac{2^n}{\lfloor d/m \rfloor},$$

whereby by the first inequality if $\tanh(\theta) < \kappa\sqrt{\frac{\log\lfloor d/m \rfloor}{n}}$ for some $\kappa < 1$ the above $\to 0$. The second inequality proves the last implication of the Proposition after an application of Le Cam's argument (5.2). □

6. Discussion. In this paper we formalized necessary and sufficient conditions for property testing in Ising models. Specifically, we showed lower and upper information-theoretic bounds on the temperature for ferromagnetic models. Furthermore, we argued that greedy correlation screening works well at high-temperature regimes, and can also be useful in low temperature regimes for certain properties. We also demonstrated that testing strongly monotone properties over the class of general Ising models is strictly more difficult than testing in ferromagnets. We discussed generic property tests based on correlation decoding, and developed a computationally efficient cycle test.

Important problems that we plan to investigate in future work include—searching for more sophisticated algorithms than correlation screening which will work at low temperature regimes for testing any property in ferromagnets; relating the temperature of the system to the information-theoretic limits of strongly monotone property testing in models with antiferromagnetic interactions; utilizing computationally efficient structure recovery algorithms, such as those of Bresler (2015), to obtain tractable algorithms for property testing in general models. Finally, several further problems merit further work: can one test for connectivity in the regime $\log d \gtrsim n$ if $\tanh(\theta) < 1$ in finite samples in ferromagnets; are there algorithms matching the information-theoretic limitations when testing over general models; is there a more general version of Theorem 2.6 which works for general properties not only strongly monotone properties.

Acknowledgments. The authors would like to express their gratitude to Professors Michael Aizenman, Ramon van Handel and Allan Sly for thought-provoking discussions, their insights and encouragements. Special thanks are due to Ramon who suggested the interpolating measure $\mathbb{P}_{G \oplus t G'}$ toward the proof of

Theorem D.1. Further thanks are due to the Editor, Associate Editor and two referees whose comments and constructive suggestions led to significant improvements of the manuscript.

SUPPLEMENTARY MATERIAL

Supplementary Material to "Property testing in high-dimensional Ising models" (DOI: 10.1214/18-AOS1754SUPP; .pdf). The supplement contains several auxiliary results, minimax risk lower bound proofs for ferromagnets (including that of Theorem 2.4), proofs for the correlation screening algorithm, hardness results for general Ising models and the proofs for the correlation testing algorithms for general models.

REFERENCES

ADDARIO-BERRY, L., BROUTIN, N., DEVROYE, L. and LUGOSI, G. (2010). On combinatorial testing problems. *Ann. Statist.* **38** 3063–3092. MR2722464

AHMED, A. and XING, E. P. (2009). Estimating time-varying networks. In *Proc. Natl. Acad. Sci. USA* **106** 11878–11883.

ANANDKUMAR, A., TAN, V. Y. F., HUANG, F. and WILLSKY, A. S. (2012). High-dimensional structure estimation in Ising models: Local separation criterion. *Ann. Statist.* **40** 1346–1375. MR3015028

ARIAS-CASTRO, E., BUBECK, S. and LUGOSI, G. (2012). Detection of correlations. *Ann. Statist.* **40** 412–435. MR3014312

ARIAS-CASTRO, E., BUBECK, S. and LUGOSI, G. (2015). Detecting positive correlations in a multivariate sample. *Bernoulli* **21** 209–241. MR3322317

ARIAS-CASTRO, E., BUBECK, S., LUGOSI, G. and VERZELEN, N. (2018). Detecting Markov random fields hidden in white noise. *Bernoulli* **24** 3628–3656. MR3788184

BERTHET, Q., RIGOLLET, P. and SRIVASTAVA, P. (2016). Exact recovery in the Ising blockmodel. Preprint. Available at arXiv:1612.03880.

BHATTACHARYA, B. B. and MUKHERJEE, S. (2018). Inference in Ising models. *Bernoulli* **24** 493–525. MR3706767

BOLLOBÁS, B. (2004). *Extremal Graph Theory*. Dover Publications, Inc., Mineola, NY. MR2078877

BRESLER, G. (2015). Efficiently learning Ising models on arbitrary graphs. In *STOC'15—Proceedings of the* 2015 *ACM Symposium on Theory of Computing* 771–782. ACM, New York. MR3388257

BRESLER, G., GAMARNIK, D. and SHAH, D. (2014). Structure learning of antiferromagnetic Ising models. In *Advances in Neural Information Processing Systems*.

BRESLER, G., MOSSEL, E. and SLY, A. (2008). Reconstruction of Markov random fields from samples: Some observations and algorithms. In *Approximation, Randomization and Combinatorial Optimization. Lecture Notes in Computer Science* **5171** 343–356. Springer, Berlin. MR2538799

CAI, T., LIU, W. and LUO, X. (2011). A constrained ℓ_1 minimization approach to sparse precision matrix estimation. *J. Amer. Statist. Assoc.* **106** 594–607. MR2847973

CHOW, C. and LIU, C. (1968). Approximating discrete probability distributions with dependence trees. *IEEE Trans. Inform. Theory* **14** 462–467.

DASKALAKIS, C., DIKKALA, N. and KAMATH, G. (2018). Testing Ising models. In *Proceedings of the Twenty-Ninth Annual ACM-SIAM Symposium on Discrete Algorithms* 1989–2007. SIAM, Philadelphia, PA. MR3775918

FÖLLMER, H. (1988). Random fields and diffusion processes. In *École D'Été de Probabilités de Saint-Flour XV–XVII*, 1985–87. *Lecture Notes in Math*. **1362** 101–203. Springer, Berlin. MR0983373

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

GHEISSARI, R., LUBETZKY, E. and PERES, Y. (2017). Concentration inequalities for polynomials of contracting Ising models. Preprint. Available at arXiv:1706.00121.

GRIMMETT, G. (2018). *Probability on Graphs*: *Random Processes on Graphs and Lattices*. *Institute of Mathematical Statistics Textbooks* **8**. Cambridge Univ. Press, Cambridge. MR3751350

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

KOCHMAŃSKI, M., PASZKIEWICZ, T. and WOLSKI, S. (2013). Curie–Weiss magnet—A simple model of phase transition. *Eur. J. Phys*. **34** 1555.

LIU, H., LAFFERTY, J. and WASSERMAN, L. (2009). The nonparanormal: Semiparametric estimation of high dimensional undirected graphs. *J. Mach*. *Learn*. *Res*. **10** 2295–2328. MR2563983

MEINSHAUSEN, N. and BÜHLMANN, P. (2006). High-dimensional graphs and variable selection with the lasso. *Ann*. *Statist*. **34** 1436–1462. MR2278363

MONTANARI, A. and PEREIRA, J. A. (2009). Which graphical models are difficult to learn? In *Advances in Neural Information Processing Systems*.

MUKHERJEE, R., MUKHERJEE, S. and YUAN, M. (2018). Global testing against sparse alternatives under Ising models. *Ann*. *Statist*. **46** 2062–2093. MR3845011

NEYKOV, M. and LIU, H. (2019). Supplement to "Property testing in high dimensional Ising models." DOI:10.1214/18-AOS1754SUPP.

NEYKOV, M., LU, J. and LIU, H. (2016). Combinatorial inference for graphical models. Preprint. Available at arXiv:1608.03045.

RAVIKUMAR, P., WAINWRIGHT, M. J. and LAFFERTY, J. D. (2010). High-dimensional Ising model selection using ℓ_1-regularized logistic regression. *Ann*. *Statist*. **38** 1287–1319. MR2662343

RAVIKUMAR, P., WAINWRIGHT, M. J., RASKUTTI, G. and YU, B. (2011). High-dimensional covariance estimation by minimizing ℓ_1-penalized log-determinant divergence. *Electron*. *J. Stat*. **5** 935–980. MR2836766

ROTHMAN, A. J., BICKEL, P. J., LEVINA, E. and ZHU, J. (2008). Sparse permutation invariant covariance estimation. *Electron*. *J. Stat*. **2** 494–515. MR2417391

SANTHANAM, N. P. and WAINWRIGHT, M. J. (2012). Information-theoretic limits of selecting binary graphical models in high dimensions. *IEEE Trans*. *Inform*. *Theory* **58** 4117–4134. MR2943079

SHANMUGAM, K., TANDON, R., RAVIKUMAR, P. K. and DIMAKIS, A. G. (2014). On the information theoretic limits of learning Ising models. In *Advances in Neural Information Processing Systems*.

YUAN, M. and LIN, Y. (2007). Model selection and estimation in the Gaussian graphical model. *Biometrika* **94** 19–35. MR2367824

DEPARTMENT OF STATISTICS AND DATA SCIENCE
CARNEGIE MELLON UNIVERSITY
PITTSBURGH, PENNSYLVANIA 15213
USA
E-MAIL: mneykov@stat.cmu.edu

DEPARTMENT OF ELECTRICAL ENGINEERING AND COMPUTER SCIENCE
DEPARTMENT OF STATISTICS
NORTHWESTERN UNIVERSITY
EVANSTON, ILLINOIS 60208
USA
E-MAIL: hanliu@northwestern.edu