Lower bounds for testing graphical models: colorings and antiferromagnetic Ising models

Ivona Bezáková IB@CS.RIT.EDU
Rochester Institute of Technology

Antonio Blanca ABLANCA@CSE.PSU.EDU
Pennsylvania State University

Zongchen Chen CHENZONGCHEN@GATECH.EDU
Georgia Institute of Technology

Daniel Štefankovič STEFANKO@CS.ROCHESTER.EDU
University of Rochester

Eric Vigoda VIGODA@GATECH.EDU
Georgia Institute of Technology

Editors: Alina Beygelzimer and Daniel Hsu

Abstract

We study the identity testing problem in the context of spin systems or undirected graphical models, where it takes the following form: given the parameter specification of the model M and a sampling oracle for the distribution μ_{M^*} of an unknown model M^*, can we efficiently determine if the two models M and M^* are the same? We consider identity testing for both soft-constraint and hard-constraint systems. In particular, we prove hardness results in two prototypical cases, the *Ising model* and *proper colorings*, and explore whether identity testing is easier than structure learning.

For the ferromagnetic (attractive) Ising model, Daskalakis et al. (2018) presented a polynomial time algorithm for identity testing. We prove hardness results in the antiferromagnetic (repulsive) setting in the same regime of parameters where structure learning is known to require a super-polynomial number of samples. Specifically, for n-vertex graphs of maximum degree d, we prove that if $|\beta|d = \omega(\log n)$ (where β is the inverse temperature parameter), then there is no identity testing algorithm for the antiferromagnetic Ising model that runs in polynomial time unless $RP = NP$. We also establish computational lower bounds for a broader set of parameters under the (randomized) exponential time hypothesis. In our proofs, we use random graphs as gadgets; this is inspired by similar constructions in seminal works on the hardness of approximate counting.

In the hard-constraint setting, we present hardness results for identity testing for proper colorings. Our results are based on the presumed hardness of #BIS, the problem of (approximately) counting independent sets in bipartite graphs. In particular, we prove that identity testing for colorings is hard in the same range of parameters where structure learning is known to be hard, which in turn matches the parameter regime for NP-hardness of the corresponding decision problem.

Keywords: distribution testing, structure learning, graphical models, Ising model, colorings.

[1] Extended abstract. Full version appears as [4].

1. Introduction

We study the *identity testing* problem in the context of *spin systems*. Spin systems, also known as Markov random fields or undirected graphical models, are a general framework in statistical physics, theoretical computer science and machine learning for modeling interacting systems of simple elements. In this type of model, the identity testing problem, sometimes also called *goodness-of-fit testing*, takes the following form: given the parameter specification of the model M and a sampling oracle for the distribution μ_{M^*} of an unknown model M^*, can we efficiently determine if the two models M and M^* are the same?

A spin system consists of a finite graph $G = (V, E)$ and a set S of *spins*; a *configuration* $\sigma \in S^V$ assigns a spin value to each vertex $v \in V$. The probability of finding the system in a given configuration σ is given by the *Gibbs* (or *Boltzmann*) distribution

$$\mu_{G,\mathcal{H}}(\sigma) = \frac{\mathrm{e}^{-\mathcal{H}(\sigma)}}{Z},$$

where Z is the normalizing factor known as the partition function and the Hamiltonian $\mathcal{H} : S^V \to \mathbb{R}$ contains terms that depend on the spin values at each vertex (a "vertex potential") and at each pair of adjacent vertices (an "edge potential").

When $\mu_{G,\mathcal{H}}(\sigma) > 0$ for every configuration $\sigma \in S^V$ (i.e., the Gibbs distribution has full support), the spin system is known as a *soft-constraint model*; otherwise, it is called a *hard-constraint model*. This is a fundamental distinction among spin systems, as it determines their application domains and the computational complexity of several inherent problems. We provide here hardness results for identity testing for both soft-constraint and hard-constraint models by considering two prototypical systems: the *Ising model* and *proper colorings*.

A naive approach to the identity testing problem is to learn first the unknown model $(G^*, \mathcal{H}^*)$ and then check whether $(G, \mathcal{H}) = (G^*, \mathcal{H}^*)$. The problem of learning G^* from samples is known as *structure learning* and has received tremendous attention; see, e.g., [16, 18, 43, 2, 46, 7, 9, 6, 55, 38, 42]. Once the graph G^* is known, it is often a simpler task to estimate $\mathcal{H}^*$ [6]; this is known as the *parameter estimation* problem. Hence, one may be inclined to conjecture that identity testing is in fact easier than structure learning, and we investigate whether or not this is the case. The main takeaway from our results is evidence that identity testing is as hard as structure learning for antiferromagnetic (repulsive) systems, as we show that the cases (i.e., parameter regimes) where these two problems are hard, in both the Ising model and proper colorings, coincide.

Lower bounds for the Ising model. The Ising model is the quintessential example of a soft-constraint system and is studied in a variety of fields, including phylogeny [27, 19], computer vision [31, 47], statistical mechanics [32, 28] and deep learning, where it appears under the guise of Boltzmann machines [1, 50, 49]. The Ising model on a graph $G = (V, E)$ is parameterized by the inverse temperature β which controls the strength of the nearest-neighbor interactions. Configurations of the model are the assignments of spins $S = \{+, -\}$ to the vertices of G. The probability of a configuration $\sigma \in S^V$ is given by the Gibbs distribution:

$$\mu_{G,\beta}(\sigma) = \frac{\mathrm{e}^{\beta \cdot A(\sigma)}}{Z_{G,\beta}}, \tag{1}$$

where $A(\sigma)$ is the number of edges of G connecting vertices with the same spin and $Z_{G,\beta} = \sum_{\sigma \in S^V} \exp(\beta \cdot A(\sigma))$ is the partition function; the associated Hamiltonian is $\mathcal{H}(\sigma) = -\beta \cdot A(\sigma)$.

In the *ferromagnetic* case ($\beta > 0$) neighboring vertices prefer to align to the same spin, whereas the opposite happens in the *antiferromagnetic* setting ($\beta < 0$). In more general variants of the model, one can allow different inverse temperatures β_e for each edge $e \in E$, as well as a vertex potential or external magnetic field. However, in this work, our emphasis will be on lower bounds for the identity testing problem, and hence we focus on the above mentioned simpler homogeneous setting (all $\beta_e = \beta$) with no external field.

The identity testing problem in the context of the Ising model is the following: given a graph $G = (V, E)$, a real number β and oracle access to independent random samples from an unknown Ising distribution μ_{G^*,β^*}, can we determine if $(G, \beta) = (G^*, \beta^*)$? If the models are distinct but their associated Gibbs distributions $\mu_{G,\beta}$ and μ_{G^*,β^*} are statistically close, an exponential (in $|V|$) number of samples may be required to determine that $(G, \beta) \neq (G^*, \beta^*)$. Hence, following a large body of work on identity testing (see, e.g., [3, 22, 21, 54, 23, 20, 14]), we study this problem in the property testing framework [48, 36]. That is, we are guaranteed that either $(G, \beta) = (G^*, \beta^*)$ or $\|\mu_{G,\beta} - \mu_{G^*,\beta^*}\| > \varepsilon$, for some standard distance $\|\cdot\|$ between distributions and $\varepsilon > 0$ fixed.

The most common distances for identity testing are total variation distance and Kullback-Leibler (KL) divergence, and it is known that a testing algorithm for the latter immediately provides one for the former [20]. Therefore, since our focus is on lower bounds, we work with total variation distance which we denote by $\|\cdot\|_{\text{TV}}$.

Identity testing for the Ising model is then formally defined as follows. For positive integers n and d let $\mathcal{M}(n, d)$ denote the family of all n-vertex graphs of maximum degree at most d.

> Given a graph $G \in \mathcal{M}(n, d)$, $\beta \in \mathbb{R}$ and sample access to a distribution μ_{G^*,β^*} for an unknown Ising model (G^*, β^*), where $G^* \in \mathcal{M}(n, d)$ and $\beta^* \in \mathbb{R}$, distinguish with probability at least $3/4$ between the cases:
>
> 1. $\mu_{G,\beta} = \mu_{G^*,\beta^*}$;
> 2. $\|\mu_{G,\beta} - \mu_{G^*,\beta^*}\|_{\text{TV}} > \frac{1}{3}$.

As usual in the property testing setting, the choice of $3/4$ for the probability of success is arbitrary, and it can be replaced by any constant in the interval $(\frac{1}{2}, 1)$ at the expense of a constant factor in the running time of the algorithm. The choice of $1/3$ for the accuracy parameter is also arbitrary: we shall see in our proofs that our lower bounds hold for any constant accuracy $\varepsilon \in (0, 1)$, provided n is sufficiently large.

Identity testing for the Ising model was studied first by Daskalakis, Dikkala and Kamath [20] who provided a polynomial time algorithm for the *ferromagnetic* Ising model (the $\beta > 0$ case). (We will discuss their results in more detail after further discussion.) In contrast, we present lower bounds for the *antiferromagnetic* Ising model ($\beta < 0$). Our lower bounds will be for the case when $\beta^* = \beta$, which means that they hold even under the additional promise that the hidden parameter β^* is equal to β. (For a discussion of the case $\beta^* \neq \beta$ see Remark 3.5).

The structure learning and parameter estimation problems, which, as discussed earlier, can be used to solve the identity testing problem, have been particularly well-studied in the context of the Ising model [6, 55, 38, 42]. Recently, Klivans and Meka [42] solved both of these problems for the Ising model with a nearly optimal algorithm. Their algorithm learns $G^* \in \mathcal{M}(n, d)$ and the parameter β^* in running time $\mathrm{e}^{O(|\beta^*|d)} \times O(n^2 \log n)$ and sample complexity $\mathrm{e}^{O(|\beta^*|d)} \times O(\log n)$. Consequently, when $|\beta^*|d = O(\log n)$ this method yields an identity testing algorithm with polynomial (in n) running time and sample complexity. In contrast, when $|\beta^*|d = \omega(\log n)$ (i.e.,

$|\beta^*|d/\log n \to \infty$), it is known that the structure learning problem cannot be solved in polynomial time [51], and so this approach for identity testing fails.

Our first result is that the identity testing problem for the antiferromagnetic Ising model is computationally hard in the same range of parameters. Specifically, we show that when $|\beta|d = \omega(\log n)$—or equivalently when $\beta = \beta^*$ and $|\beta^*|d = O(\log n)$—there is no polynomial running time identity testing algorithm for $\mathcal{M}(n,d)$ unless $RP = NP$; RP is the class of problems that can be solved in polynomial time by a randomized algorithm.

Theorem 1.1 *Suppose n, d are positive integers such that $3 \le d \le n^\theta$ for constant $\theta \in (0,1)$. If $RP \neq NP$, then for all real $\beta < 0$ satisfying $|\beta|d = \omega(\log n)$ and all n sufficiently large, there is no polynomial running time algorithm to solve the identity testing problem for the* antiferromagnetic *Ising model in $\mathcal{M}(n,d)$.*

In contrast to the above result, Daskalakis, Dikkala and Kamath [20] designed an identity testing algorithm for the Ising model with polynomial running time and sample complexity that works for arbitrary values of β (positive, negative or even non-homogeneous). This appears to contradict our lower bound in Theorem 1.1. However, the model in [20] assumes not only sampling access to the unknown distribution μ_{G^*,β^*}, but also that the covariances between the spins at every pair of vertices in the visible graph $G = (V,E)$ are given. More precisely, they assume that for every $u, v \in V$ the quantity $\mathrm{E}_{\mu_{G,\beta}}[X_u X_v]$ is known, where $X_u, X_v \in \{+1,-1\}$ are the random variables corresponding to the spins at u and v, respectively.

This is a reasonable assumption when these quantities can be computed (or approximated up to an additive error) efficiently. However, an immediate consequence of our results is that in the antiferromagnetic setting when $|\beta|d = \omega(\log n)$ there is no FPRAS[1] for estimating $\mathrm{E}_{\mu_{G,\beta}}[X_u X_v]$ unless $RP = NP$. In a related result, Goldberg and Jerrum [35] showed recently that there is no FPRAS for (multiplicatively) approximating the pairwise covariances for the antiferromagnetic Ising model unless $RP = \#P$. Further evidence for the hardness of this problem comes from the fact that sampling is hard in the antiferromagnetic setting [53, 30] and in the ferromagnetic model in the presence of inconsistent magnetic fields [33] (i.e., the vertex potential of distinct vertices may have different signs). In summary, the algorithmic results of [20] are most interesting for the ferromagnetic Ising model (with consistent fields), where there are known polynomial running time algorithms for estimating the pairwise covariances (see, e.g., [41, 45, 37, 17]).

In Theorem 1.1 we assume that $|\beta|d = \omega(\log n)$, but our main technical result (Theorem 2.1) is actually more general. We show that when $|\beta|d \ge c \ln n$, where $c > 0$ is a sufficiently large constant, if there is an identity testing algorithm with running time $T = T(n)$ and sample complexity $L = L(n)$ then there is also a randomized algorithm with running time $O(T + Ln)$ for computing the maximum cut of any graph with $N = n^{\Theta(1)}$ vertices. Theorem 1.1 then follows immediately from the fact that either T or L ought to be super-polynomial in n, as otherwise we obtain a randomized algorithm for the maximum cut problem with polynomial running time; this would imply that $RP = NP$.

Under a stronger (but also standard) computational theoretic assumption, namely that there is no randomized algorithm with sub-exponential running time for the 3-SAT problem, i.e., the *(randomized) exponential time hypothesis or $rETH$* [40, 13], our main theorem also implies a general lower bound for identity testing that holds for all β and d satisfying $|\beta|d \ge c \ln n$.

[1] A fully polynomial-time randomized approximation scheme (FPRAS) for an optimization problem with optimal solution Z produces an approximate solution $\hat{Z}$ such that, with probability at least $1-\delta$, $(1-\varepsilon)\hat{Z} \le Z \le (1+\varepsilon)\hat{Z}$ with running time polynomial in the instance size, ε^{-1} and $\log(\delta^{-1})$.

Theorem 1.2 *Suppose n, d are positive integers such that $3 \leq d \leq n^{\theta}$ for constant $\theta \in (0,1)$. Then, there exist constants $c = c(\theta) > 0$ and $\alpha = \alpha(\theta) \in (0,1)$ such that when $|\beta|d \geq c \ln n$, $rETH$ implies that the running time $T(n)$ of any algorithm that solves the identity testing problem for the* antiferromagnetic *Ising model in $\mathcal{M}(n,d)$ satisfies $T(n) \geq \min\left\{\exp(\Omega(n^{\alpha})), \frac{\exp(|\beta|d/c)}{30n}\right\}$.*

We remark that the bound in this theorem is comparable to the $\exp(\Omega(|\beta|d))$ lower bound for the sample complexity of structure learning [51], albeit requiring that $rETH$ is true.

The very high level idea of the proof of our main theorem for the Ising model (Theorem 2.1), from which Theorems 1.1 and 1.2 are derived as corollaries, is as follows: given a graph H and an integer k, we construct an identity testing instance Λ so that the output of the identity testing algorithm on Λ can be used to determine whether there is a cut in H of size at least k. A crucial component in our construction is a "degree reducing" gadget, which consists of a random bipartite graph and is inspired by similar random gadgets in seminal works on the hardness of approximate counting [52]. One of the main technical challenges in the paper is to establish precise bounds on the *edge expansion* of these random gadgets. A detailed overview of our proof is given in Section 2.

Lower bounds for proper q-colorings. The *proper q-colorings* of a graph $G = (V, E)$ constitute a canonical hard-constraint spin system, with multiple applications in statistical physics and theoretical computer science. In this model, the vertices of graph G are assigned spins (or colors) from $\{1, \ldots, q\}$, and the Gibbs distribution μ_G becomes the uniform distribution over the proper q-colorings of the graph G. The identity testing problem for this model in $\mathcal{M}(n,d)$ is defined as follows: given q, a graph $G \in \mathcal{M}(n,d)$ and sample access to random q-colorings of an unknown graph $G^* \in \mathcal{M}(n,d)$, distinguish with probability at least $3/4$ whether $\mu_G = \mu_{G^*}$ or $\|\mu_G - \mu_{G^*}\| > 1/3$.

We establish lower bounds for this problem, thus initiating the study of identity testing in the context of hard-constraint systems. While identity testing does not seem to have been studied for hard-constraint models before, the related structure learning problem has received some attention [8, 5]. For proper colorings, it is known that when $q \geq d + 1$ the hidden graph G can be learned from $\mathrm{poly}(n, d, q)$ samples, whereas when $q \leq d$ then the problem is non-identifiable, i.e., there are distinct graphs with the same collection of q-colorings [5]. Moreover, for $d \geq d_c(q) = q + \sqrt{q} + \Theta(1)$, it was also established in [5] that the easier *equivalent structure learning problem* (learning any graph with the same collection of q-colorings as the unknown graph) is computationally hard in the sense that the sample complexity is exponential in n. The threshold $d_c(q)$ is very close to the one for polynomial time/NP-completeness for the problem of determining if G is q-colorable [26, 44].

We prove here that the identity testing for proper q-colorings is also hard when $d \geq d_c(q)$, thus establishing another connection between the hardness of identity testing and structure learning. For this we use the complexity of #BIS, which is the problem of counting independent sets in bipartite graphs. #BIS is believed not to have an FPRAS and has gained considerable interest in approximate counting as a tool for proving relative complexity hardness; see [24, 34, 25, 11, 15, 12, 29].

Theorem 1.3 *Suppose n, d and q are positive integers such that $q \geq 3$ and $d \geq d_c(q)$. If* #BIS *does not admit an FPRAS, then there is no polynomial running time algorithm that solves the identity testing problem for proper q-colorings in $\mathcal{M}(n,d)$.*

In the proof of this theorem we reduce the #BIS-hard problem of counting 3-colorings in bipartite graphs to identity testing for q-colorings. The high level idea of our proof is as follows: given a bipartite graph H and an approximation $\hat{Z}$ for the number of 3-colorings $Z_3(H)$ of H, we

construct an identity testing instance that depends on both H and the value of $\hat{Z}$. We then show how to use an identity testing algorithm on this instance to check whether $\hat{Z}$ is an upper or lower bound for $Z_3(H)$. By adjusting $\hat{Z}$ and repeating this process we converge to a good approximation for $Z_3(H)$. A crucial element in our construction is again the design of a degree reducing gadget; in this case, our gadget is inspired by similar constructions in [26, 44, 5] for establishing the computational hardness of the decision and (equivalent) structure learning problems for $d \geq d_c(q)$. Finally, we mention that for 3-colorings, $d_c(3) = 4$ and thus our hardness result holds for all graphs with maximum degree at least 4.

An algorithm for the ferromagnetic Ising model. We provide an improved algorithm for the *ferromagnetic* Ising model. As mentioned, by combining the algorithm in [20] with previous results for sampling [41, 45, 37, 17], one obtains a polynomial running time algorithm for identity testing in the ferromagnetic setting. This algorithm works for symmetric-KL divergence which is a stronger notion of distance. We show that if one considers instead total variation distance, then there is a polynomial running time algorithm that solves the identity testing problem with sample complexity $\tilde{O}(n^2 d^2 \varepsilon^{-2})$. This is an improvement over the $\tilde{O}(n^2 d^2 \beta^2 \varepsilon^{-2})$ bound in [20], as there is no dependence on the inverse temperature β. A precise statement and proof of this result are provided in the full version of this paper [4].

The rest of the paper is organized as follows. In Section 2, we sketch the key ideas in the proof of our result for the Ising model (Theorem 1.1); the actual proof is given in Section 3. The main ideas in the proof of our testing lower bound for proper colorings (Theorem 1.3) are summarized in Section 4. The complete proof of this theorem is provided in [4].

2. Lower bounds for the Ising model: proof overview

To establish our lower bounds for the antiferromagnetic Ising model, we use the computational hardness of the maximum cut (MAXCUT) problem. Recall that in the decision problem, given a graph H and an integer $k > 0$, the goal is to determine whether there is a cut of size at least k in H. Our main technical result, from which Theorems 1.1 and 1.2 are derived, is the following.

Theorem 2.1 *Suppose n and d are positive integers such $3 \leq d \leq n^{1-\rho}$ for some constant $\rho \in (0,1)$. Then, for all n sufficiently large, there exist $c = c(\rho) > 0$ and an integer $N = \Theta(n^{\min\{\frac{\rho}{4}, \frac{1}{14}\}})$ such that when $|\beta| d \geq c \ln n$, any identity testing algorithm for $\mathcal{M}(n,d)$ for the* antiferromagnetic *Ising model with running time $T(n)$ and sample complexity $L(n) \leq \frac{\exp(|\beta| d/c)}{30n}$ provides a randomized algorithm for* MAXCUT *on any graph with N vertices. This algorithm outputs the correct answer with probability at least $11/20$ and has running time $O(T(n) + n \cdot L(n))$.*

In words, this theorem says that under some mild assumptions for d and $L(n)$, when $|\beta| d \geq c \ln n$, any identity testing algorithm with running time $T = T(n)$ and sample complexity $L = L(n)$ provides a randomized algorithm for MAXCUT on graphs of $\mathrm{poly}(n)$ size with running time $O(T + Ln)$. Hence, under the assumption that there is no polynomial running time randomized algorithm for MAXCUT (i.e., $RP \neq NP$), either T or L ought to be super-polynomial, and Theorem 1.1 from the introduction follows. Theorem 1.2 is also a direct corollary of Theorem 2.1. To see this, note that for $\alpha = \min\{\frac{\rho}{4}, \frac{1}{14}\}$, under the assumption that $rETH$ is true, there is no randomized algorithm for MAXCUT in graphs with $N = \Theta(n^\alpha)$ vertices with running time $\exp(o(n^\alpha))$. Thus, either $T = \exp(\Omega(n^\alpha))$ or the assumption $L \leq \frac{\exp(|\beta| d/c)}{30n}$ cannot hold, which implies $T \geq L > \frac{\exp(|\beta| d/c)}{30n}$.

Proof sketch for Theorem 2.1. To establish Theorem 2.1 we construct a class $\mathcal{N}$ of n-vertex graphs of maximum degree at most d and show how an algorithm that solves identity testing for $\mathcal{N} \subset \mathcal{M}(n,d)$ can be used to solve the MAXCUT problem on graphs with $N = \Theta(n^\alpha)$ vertices, where $\alpha \in (0,1)$ is a constant. (The exact value for α depends on d: if $d = O(1)$, then we can take $\alpha = 1/14$; otherwise, we set $\alpha = \rho/4$.)

Suppose we want to solve the MAXCUT problem for a graph $H = (V,E)$ and $k \in \mathbb{N}$. For this, we add two vertices s and t to H and connect both s and t to every vertex in V with $N = |V|$ edges (adding a total of $2N^2$ edges); we also add w edges between s and t. Let $\hat{H}_w$ be the resulting multigraph. (In our proofs we will convert $\hat{H}_w$ into a simple graph, but it is conceptually simpler to consider the multigraph for now.) The cut $(\{s,t\}, V)$ in $\hat{H}_w$ is of size $2N^2$. We consider the following variant of the MAXCUT problem.

Definition 2.2 *In the* TWOLARGECUTS *problem, given the graph H and $w \in \mathbb{N}$, the goal is to determine whether there are at least two cuts in $\hat{H}_w$ of size at least $2N^2$.*

MAXCUT can be reduced to TWOLARGECUTS by treating w, the number of edges between s and t, as a parameter. Observe that if $(S, V \setminus S)$ is a cut of size k in the original graph H, then $(S \cup \{s\}, (V \setminus S) \cup \{t\})$ is a cut of size

$$w + k + N|S| + N|V \setminus S| = w + k + N^2$$

in $\hat{H}_w$. Hence, $(\{s,t\}, V)$ is the unique large cut (i.e., cut of size $\geq 2N^2$) if and only if

$$w + \textsc{MaxCut}(H) + N^2 < 2N^2,$$

where $\textsc{MaxCut}(H)$ denotes the size of the maximum cut of H. Therefore, to solve MAXCUT for H and k, it is sufficient to solve the TWOLARGECUTS problem for $\hat{H}_w$ with $w = N^2 - k$. This yields that the TWOLARGECUTS problem is NP-complete and the following useful lemma. (We refer the reader to the full version of this paper [4] for detailed proofs of these facts.)

Lemma 2.3 *Let $H = (V,E)$ be an N-vertex graph and let $\delta \in (0, 1/2]$. Suppose there exists a randomized algorithm that solves the* TWOLARGECUTS *problem on inputs H and $w \leq N^2$ with probability at least $1/2 + \delta$ and running time R. Then, there exists a randomized algorithm to solve* MAXCUT *for H and $k \in \mathbb{N}$ with running time $R + O(N^2)$ and success probability at least $1/2 + \delta$.*

To determine if $(\{s,t\}, V)$ is the unique large cut in $\hat{H}_w$ we can use the antiferromagnetic Ising model on $\hat{H}_w$ as follows. Every Ising configuration of $\hat{H}_w$ determines a cut: all the "+" vertices belong to one side of the cut and the "−" vertices to the other (or vice versa). Observe that for every cut of $\hat{H}_w$ there are exactly two Ising configurations. The intuition is that the maximum cut of $\hat{H}_w$ corresponds to the two configurations of maximum likelihood in the Gibbs distribution. Indeed, when $|\beta|$ is sufficiently large, the distribution will be well-concentrated on the two configurations corresponding to the maximum cut. Therefore, a sample from the Gibbs distribution would reveal the maximum cut of $\hat{H}_w$ with high probability.

To simulate large magnitudes of β, we strengthen the interactions between neighboring vertices of $\hat{H}_w$ by replacing every edge by 2ℓ edges. However, sampling from the antiferromagnetic Ising distribution on the resulting multigraph $\hat{H}_{w,\ell}$ is also a hard problem, and we would need to provide a sampling procedure. For this, we use the identity testing algorithm as follows. We construct a simpler Ising model M^* with two key properties: (i) we can easily generate samples from M^*

and (ii) M^* is close in total variation distance to the Ising model $M = (\hat{H}_{w,\ell}, \beta)$ if and only if $(\{s,t\}, V)$ is the unique large cut of $\hat{H}_w$. Then, we give $\hat{H}_{w,\ell}$, the parameter β and samples from M^* as input to the tester. If the tester outputs YES, it means that it regarded the samples from M^* as samples from M and so $(\{s,t\}, V)$ must be the unique large cut of $\hat{H}_w$. Conversely, if the tester outputs NO, then the total variation distance between μ_M and μ_{M^*} must be large, in which case $(\{s,t\}, V)$ is not the unique large cut of $\hat{H}_w$.

In summary, this argument implies that an identity testing algorithm for n-vertex multigraphs gives a polynomial time randomized algorithm for MAXCUT on graphs with $n-2$ vertices. However, the maximum degree of $\hat{H}_{w,\ell}$ depends on ℓ, N and w and could be much larger than d. Hence, this argument does not apply for small values of d, even if we overlook the fact that we would be using identity testers for multigraphs instead of graphs. To extend the argument to *simple* graphs in $\mathcal{M}(n,d)$ for all $3 \leq d \leq n^{1-\rho}$, we introduce a "degree reducing" gadget, which is reminiscent of gadgets used in works concerning the hardness of approximate counting [52, 53].

Every vertex of $\hat{H}_{w,\ell}$ is replaced by a random bipartite graph $G = (L \cup R, E_G)$; see Section 3 for the precise random graph model. The graph G has maximum degree at most d, and some of its vertices, which we call *ports*, will have degree strictly less than d, so that they can be used for connecting the gadgets as indicated by the edges of $\hat{H}_{w,\ell}$. The resulting simple graph, which we denote by $\hat{H}^{\Gamma}_w$, will have maximum degree d. (Γ is the set of parameters of our random graph model; see Section 3 for the details.) In similar manner as described above for $\hat{H}_{w,\ell}$, the antiferromagnetic Ising model on $\hat{H}^{\Gamma}_w$ can be used to determine whether $(\{s,t\}, V)$ is the unique large cut of $\hat{H}_w$. This would involve sampling from the Gibbs distribution $\mu_{\hat{H}^{\Gamma}_w,\beta}$, which is hard but can be done using the identity testing algorithm. Since $\hat{H}^{\Gamma}_w$ has maximum degree at most d, Theorem 2.1 follows.

Finally, we mention that the main technical challenge in our approach is to establish that in every gadget, with high probability, either every vertex of L is assigned "+" and every vertex of R is assigned "−" or vice versa. To show this, we require very precise bounds on the *edge expansion* of the random bipartite graph G. When $d \to \infty$, these bounds can be derived in a fairly straightforward manner from the results in [10]. However, the case of $d = O(1)$ is more difficult, and it requires us to define the notion of edge expansion with respect to the ports of the gadget and extend some of the ideas in [39]. Our bounds for the edge expansion of random bipartite graphs may be of independent interest and are provided in the full version of this paper [4].

3. Proof of main result for the Ising model: Theorem 2.1

The Ising gadget. Suppose $m, p, d, d_{\text{IN}}, d_{\text{OUT}} \in \mathbb{N}^+$ are positive integers such that $m \geq p$, $d \geq 3$ and $d_{\text{IN}} + d_{\text{OUT}} = d$. Let $G = (V_G, E_G)$ be the random bipartite graph defined as follows:

1. Set $V_G = L \cup R$, where $|L| = |R| = m$ and $L \cap R = \emptyset$;
2. Choose $P \subset V_G$ uniformly at random among all the subsets such that $|P \cap L| = |P \cap R| = p$;
3. Let $M_1, \ldots, M_{d_{\text{IN}}}$ be d_{IN} random perfect matchings between L and R;
4. Let $M'_1, \ldots, M'_{d_{\text{OUT}}}$ be d_{OUT} random perfect matchings between $L \backslash P$ and $R \backslash P$;
5. Set $E_G = \left(\bigcup_{i=1}^{d_{\text{IN}}} M_i\right) \cup \left(\bigcup_{i=1}^{d_{\text{OUT}}} M'_i\right)$ and make G simple by removing repeated edges.

We use $\mathcal{G}(m, p, d_{\text{IN}}, d_{\text{OUT}})$ to denote the resulting distribution; i.e., $G \sim \mathcal{G}(m, p, d_{\text{IN}}, d_{\text{OUT}})$. Vertices in P are called *ports*. Every port has degree at most d_{IN} while every non-port has degree at most d.

In our proofs, we use instances of this random graph model with two different choices of parameters. For the case when d is such that $3 \leq d = O(1)$, we choose $p = \lfloor m^{1/4} \rfloor$, $d_{\text{IN}} = d - 1$ and $d_{\text{OUT}} = 1$; otherwise we take $p = m$ (i.e., every vertex is a port), $d_{\text{IN}} = \lfloor \theta d \rfloor$ and $d_{\text{OUT}} = d - \lfloor \theta d \rfloor$ for a suitable constant $\theta \in (0, 1)$. For both parameter choices we establish that the random graph G is a good expander with high probability. Using this, we can show that there are only two "typical" configurations for the Ising model on G, even in the presence of an external configuration (i.e., a boundary condition) exerting influence on the configuration of G via its ports.

We present some notation next that will allow us to formally state these facts. Let $\sigma^+(G)$ be the configuration of $G = (L \cup R, E_G)$ where every vertex in L is assigned "$+$" and every vertex in R is assigned "$-$"; similarly, define $\sigma^-(G)$ by interchanging "$+$" and "$-$". Suppose G is an induced subgraph of a larger graph $G' = (V_{G'}, E_{G'})$. Let $\partial P = V_{G'} \setminus V_G$. Assume that every vertex in $P \subseteq V_G$ is connected to up to d_{OUT} vertices in ∂P and that there are no edges between $V_G \setminus P$ and ∂P in G'. We use $\{\partial P = \tau\}$ for the event that the configuration in G' of ∂P is $\tau \in \{+, -\}^{\partial P}$. We can show that for any τ, with high probability over the choice of the random graph G, the Ising configuration of V_G on G' conditioned on $\{\partial P = \tau\}$ will likely be $\sigma^+(G)$ or $\sigma^-(G)$.

Theorem 3.1 *Suppose $\beta < 0$, $3 \leq d = O(1)$, $d_{\text{IN}} = d - 1$, $d_{\text{OUT}} = 1$ and $p = \lfloor m^\alpha \rfloor$, where $\alpha \in (0, \frac{1}{4}]$ is a constant independent of m. Then, there exists a constant $\delta > 0$ such that with probability $1 - o(1)$ over the choice of the random graph G the following holds for every configuration τ on ∂P:*

$$\mu_{G',\beta}(\{\sigma^+(G), \sigma^-(G)\} \mid \partial P = \tau) \geq 1 - \frac{2m}{\mathrm{e}^{\delta|\beta|d}}.$$

A similar theorem is also established for the case when $p = m$, $d_{\text{IN}} = \lfloor \theta d \rfloor$ and $d_{\text{OUT}} = d - \lfloor \theta d \rfloor$. The proofs of these theorems are provided in [4].

Testing instance construction. Let $H = (V, E)$ be a simple N-vertex graph and for integer $w \leq N^2$ let $\hat{H}_w$ be the multigraph defined in Section 2. We use an instance of the random bipartite graph $\mathcal{G}(m, p, d_{\text{IN}}, d_{\text{OUT}})$ as a gadget to define a simple graph $\hat{H}^\Gamma_w$, where $\Gamma = \{m, p, d_{\text{IN}}, d_{\text{OUT}}, \ell\}$; $\ell > 0$ is assumed to be an integer divisible by d_{OUT}. The graph $\hat{H}^\Gamma_w$ is constructed as follows:

1. Generate an instance $G = (L \cup R, E_G)$ of the random graph model $\mathcal{G}(m, p, d_{\text{IN}}, d_{\text{OUT}})$;
2. Replace every vertex v of $\hat{H}_w$ by a copy $G_v = (L_v \cup R_v, E_{G_v})$ of the generated instance G;
3. For every edge $\{v, u\} \in \hat{H}_w$, choose ℓ/d_{OUT} unused ports in L_v and ℓ/d_{OUT} unused ports in L_u and connect them with a simple bipartite d_{OUT}-regular graph;
4. Similarly, for every edge $\{v, u\} \in \hat{H}_w$, choose ℓ/d_{OUT} unused ports in R_v and ℓ/d_{OUT} unused ports in R_u and connect them with a simple bipartite d_{OUT}-regular graph.

Observe that our construction requires:

$$\text{(i) } d_{\text{IN}} + d_{\text{OUT}} = d \leq m \quad \text{(ii) } d_{\text{OUT}} \mid \ell \quad \text{(iii) } \ell(N^2 + w) \leq p \cdot d_{\text{OUT}} \quad \text{(iv) } d^2_{\text{OUT}} \leq \ell. \qquad (2)$$

To see that condition (2)(iii) is necessary, note that the maximum degree of $\hat{H}_w$ is $N^2 + w$ (this is the degree of vertices s and t), and so the total out-degree of the ports should be large enough to accommodate $\ell(N^2 + w)$ edges. Observe also that when condition (2)(iv) holds, there is always a simple bipartite d_{OUT}-regular graph with ℓ/d_{OUT} vertices on each side for steps 3 and 4.

The number of vertices in $\hat{H}^\Gamma_w$ is $2m(N + 2)$ and its maximum degree is $d = d_{\text{IN}} + d_{\text{OUT}}$; thus, $\hat{H}^\Gamma_w \in \mathcal{M}(2m(N+2), d)$. Let I be the empty graph with N vertices. By setting $H = I$ and $w = 0$,

we can analogously define the graphs $\hat{I}_0$ and $\hat{I}_0^\Gamma$ so that $\hat{I}_0^\Gamma \in \mathcal{M}(2m(N+2), d)$. Let M and M^* denote the Ising models $(\hat{H}_w^\Gamma, \beta)$ and $(\hat{I}_0^\Gamma, \beta)$, respectively. We show next that the models M and M^* are statistically close if and only if $(\{s,t\}, V)$ is the unique large cut of $\hat{H}_w$. To formally state this fact we require some additional notation.

For a configuration σ on $\hat{H}_w^\Gamma$, we say that the gadget $G_v = (L_v \cup R_v, E_{G_v})$ is in the plus (resp., minus) *phase* if all the vertices in L_v (resp., R_v) are assigned "+" in σ and all the vertices in R_v (resp., L_v) are assigned "−". Let Ω_{good} be the set of configurations of $\hat{H}_w^\Gamma$ where the gadget of every vertex is either in the plus or the minus phase. The set of Ising configurations of $\hat{H}_w^\Gamma$ and $\hat{I}_0^\Gamma$ is the same and is denoted by Ω. We use Z_M, Z_{M^*} for the partition functions of M, M^*, and $Z_M(\Lambda)$, $Z_{M^*}(\Lambda)$ for their restrictions to a subset of configurations $\Lambda \subseteq \Omega$. That is, $Z_M = \sum_{\sigma\in\Omega} w_M(\sigma)$ and $Z_M(\Lambda) = \sum_{\sigma\in\Lambda} w_M(\sigma)$ where $w_M(\sigma) := \mathrm{e}^{\beta A(\sigma)}$ is called the *weight* of the configuration σ in M; see (1). When $\beta < 0$, $w_M(\sigma) = \mathrm{e}^{-|\beta|A(\sigma)}$. The models M and M^* are related as follows:

Lemma 3.2 *Let $N \geq 1$, $w \geq 0$ be integers and let $\beta < 0$. Let $\Gamma = (m, p, d_{\text{IN}}, d_{\text{OUT}}, \ell)$ be such that $|\beta|(\ell - d) \geq N$ and the conditions in (2) are satisfied. If for the Ising models $M = (\hat{H}_w^\Gamma, \beta)$ and $M^* = (\hat{I}_0^\Gamma, \beta)$ we have $Z_M(\Omega_{\text{good}}) \geq (1-\varepsilon)Z_M$ and $Z_{M^*}(\Omega_{\text{good}}) \geq (1-\varepsilon)Z_{M^*}$ for some $\varepsilon \in (0,1)$, then with probability $1-o(1)$ over the choice of the random graph G the following holds:*

1. *If $(\{s,t\}, V)$ is the unique large cut of $\hat{H}_w$, then $\|\mu_M - \mu_{M^*}\|_{\text{TV}} \leq 2(\varepsilon + \mathrm{e}^{-2|\beta|d})$.*
2. *If $(\{s,t\}, V)$ is not the unique large cut of $\hat{H}_w$, then $\|\mu_M - \mu_{M^*}\|_{\text{TV}} > \frac{1}{2} - \varepsilon - \mathrm{e}^{-2|\beta|d}$.*

The next lemma shows that we can easily generate samples from the simpler model M^*.

Lemma 3.3 *Let $N \geq 1$ be an integer and let $\beta < 0$. Let $\Gamma = (m, p, d_{\text{IN}}, d_{\text{OUT}}, \ell)$ be such that $|\beta|(\ell N - d) \geq N$ and the conditions in (2) are satisfied. If for the Ising model $M^* = (\hat{I}_0^\Gamma, \beta)$ we have $Z_{M^*}(\Omega_{\text{good}}) \geq (1-\varepsilon)Z_{M^*}$ for some $\varepsilon \in (0,1)$, then there exists a sampling algorithm with running time $O(mN)$ such that with probability $1 - o(1)$ over the choice of the random graph G, the distribution μ_{ALG} of its output satisfies $\|\mu_{M^*} - \mu_{\text{ALG}}\|_{\text{TV}} \leq \varepsilon + \mathrm{e}^{-2|\beta|d}$.*

The proofs of Lemmas 3.2 and 3.3 are deferred to the full version of this paper [4]. We are now ready to prove Theorem 2.1.

Proof of Theorem 2.1 Let us assume first that $3 \leq d = O(1)$ and let $N = \lfloor n^{1/14} \rfloor - 2$ and $m = \lfloor \frac{n^{13/14}}{2} \rfloor$. If $\lfloor n^{1/14} \rfloor$ and $\lfloor \frac{n^{13/14}}{2} \rfloor$ are integers, then $n = 2m(N+2)$. For simplicity and without much loss of generality, we assume that this is indeed the case. For some brief remarks on how to extend the current proof to the case when $\lfloor n^{1/14} \rfloor$ or $\lfloor \frac{n^{13/14}}{2} \rfloor$ are not integers the reader is referred to the full version [4].

Let $H = (V, E)$ be an N-vertex graph. We show that an identity testing algorithm for $\mathcal{M}(n, d)$ with running time $T = T(n)$ and sample complexity $L = L(n) \leq \frac{\exp(|\beta|d/c)}{30n}$, henceforth called the TESTER, can be used to solve TWOLARGECUTS on inputs H and $w \leq N^2$ in $O(T + Ln)$ time.

We recall that in the TWOLARGECUTS problem the goal is to determine whether $(\{s,t\}, V)$ is the unique large cut of the graph $\hat{H}_w$; see Section 2 and Definition 2.2. For this, we construct the two Ising models $M = (\hat{H}_w^\Gamma, \beta)$ and $M^* = (\hat{I}_0^\Gamma, \beta)$ as described at the beginning of this section. When $3 \leq d = O(1)$, we choose $p = \lfloor m^{1/4} \rfloor$, $d_{\text{IN}} = d - 1$, $d_{\text{OUT}} = 1$ and $\ell = \Theta(n^{9/112})$. That is, $\Gamma = \{m, \lfloor m^{1/4} \rfloor, d-1, 1, \Theta(n^{9/112})\}$. Recall that ℓ is an integer divisible by d_{OUT} by assumption. Moreover, $d_{\text{IN}} + d_{\text{OUT}} = d$ and $\hat{H}_w^\Gamma$, $\hat{I}_0^\Gamma$ have exactly n vertices; hence, $\hat{H}_w^\Gamma, \hat{I}_0^\Gamma \in \mathcal{M}(n, d)$.

Suppose σ is sampled according to μ_M. Theorem 3.1 implies that with probability $1 - o(1)$ over the choice of the random gadget G, if the configuration in the gadget G_v for vertex v of $\hat{H}_w$ is re-sampled in σ, conditional on the configuration of σ outside of G_v, then the new configuration in G_v will be in either the plus or minus phase with probability at least $1 - \frac{2m}{\mathrm{e}^{\delta|\beta|d}}$, for a suitable constant $\delta > 0$. A union bound then implies that after re-sampling the configuration in every gadget one by one, the resulting configuration σ' is in the set Ω_{good} with probability $1 - \frac{2m(N+2)}{\mathrm{e}^{\delta|\beta|d}}$. Thus,

$$\mu_M(\Omega_{\text{good}}) = \frac{Z_M(\Omega_{\text{good}})}{Z_M} \geq 1 - \frac{2m(N+2)}{\mathrm{e}^{\delta|\beta|d}}. \tag{3}$$

The same is true if σ were sampled from μ_{M^*} instead, and so $\mu_{M^*}(\Omega_{\text{good}}) \geq 1 - \frac{2m(N+2)}{\mathrm{e}^{\delta|\beta|d}}$.

Let $\mu_M^{\otimes L}$, $\mu_{M^*}^{\otimes L}$ and $\mu_{\text{ALG}}^{\otimes L}$ be the product distributions corresponding to L independent samples from μ_M, μ_{M^*} and μ_{ALG} respectively. Let $\gamma = \min\{2, \delta\}$ and suppose $c > 1/\gamma$. Since in our reduction from MAXCUT to TWOLARGECUTS we only need consider w's such that $w \leq N^2$, our choices for N and Γ satisfy the conditions in (2), and $|\beta|(\ell - d) \geq N$ when $|\beta|d \geq c \ln n$. The following fact then follows from (3) and Lemmas 3.2 and 3.3; its proof is provided in [4].

Fact 3.4 *If $(\{s,t\}, V)$ is the unique large cut of $\hat{H}_w$, then $\|\mu_M^{\otimes L} - \mu_{\text{ALG}}^{\otimes L}\|_{\text{TV}} \leq \frac{1}{5}$.*

Our algorithm for TWOLARGECUTS inputs the Ising model M and L samples $\mathcal{S} = \{\sigma_1, \ldots, \sigma_L\}$ from μ_{ALG} to the TESTER (i.e., $\sigma_i \sim \mu_{\text{ALG}}$ and $\mathcal{S} \sim \mu_{\text{ALG}}^{\otimes L}$) and outputs the negation of the TESTER's output. By Fact 3.4, when $(\{s,t\}, V)$ is the unique large cut of $\hat{H}_w$ the set of samples given as input to the TESTER (i.e., $\mathcal{S}$) is distributed according to $\mu_M^{\otimes L}$ with probability at least $4/5$. Let $\mathcal{F}_{\mathcal{S}}$ be the event that this is indeed the case. Recall that the TESTER makes a mistake with probability at most $1/4$. Moreover, if $\mathcal{F}_{\mathcal{S}}$ occurs and the TESTER does not make a mistake, then the TESTER would output YES. Therefore, $\Pr[\text{TESTER outputs NO}] \leq \Pr[\neg\mathcal{F}_{\mathcal{S}}] + \Pr[\text{TESTER makes a mistake}] \leq 9/20$. Hence, the TESTER returns YES with probability at least $11/20$ in this case.

When $(\{s,t\}, V)$ is not the unique large cut of $\hat{H}_w$, (3) and part 2 of Lemma 3.2 imply that $\|\mu_M - \mu_{M^*}\|_{\text{TV}} > 1/3$ for sufficiently large n. Moreover, by Lemma 3.3, $\|\mu_{M^*}^{\otimes L} - \mu_{\text{ALG}}^{\otimes L}\|_{\text{TV}} \leq L\|\mu_{M^*} - \mu_{\text{ALG}}\|_{\text{TV}} \leq 1/15$. Thus, with probability at least $14/15$ the samples in $\mathcal{S}$ have distribution $\mu_{M^*}^{\otimes L}$. Let $\mathcal{F}_{\mathcal{S}}^*$ be the event that this is the case. Then, $\Pr[\text{TESTER outputs YES}] \leq \Pr[\neg\mathcal{F}_{\mathcal{S}}^*] + \Pr[\text{TESTER makes a mistake}] < 1/3$. Hence, the TESTER returns NO with probability at least $2/3$.

Therefore, our algorithm can solve the TWOLARGECUTS problem on $\hat{H}_w$ in $O(T + Ln)$ time with probability at least $11/20$. The result then follows from Lemma 2.3 and the fact that $|V| = N = \lfloor n^{1/14} \rfloor - 2 \geq \lfloor n^{\min\{\frac{\rho}{4}, \frac{1}{14}\}} \rfloor - 2$. The case when d is such that $d \leq n^{1-\rho}$ and $d \to \infty$ follows in similar fashion; see [4] for the details of this case. ∎

Remark 3.5 *Our hardness results for identity testing for the Ising model require $|\beta|d \geq c \ln n$ for a suitable constant $c > 0$. We further assume that $\beta^* = \beta$; namely, our lower bounds hold even under this additional promise. Our proof extends without any significant modification to the case where $\max\{|\beta|, |\beta^*|\} \cdot d \geq c \ln n$. As mentioned, there are polynomial running time algorithms for identity testing when either $|\beta^*|d = O(\log n)$ (via structure learning methods), or when $|\beta| = O(d^{-1})$ in the so-called "tree uniqueness region" where we can sample efficiently. Therefore, when $\beta \neq \beta^*$, β is in the tree non-uniqueness region (specifically, $\Omega(1) = |\beta|d < c \ln n$) and $|\beta^*|d = \omega(\log n)$, the computational complexity of identity testing is open, as there is no known polynomial running time algorithm, and our lower bound does not apply to this combination of parameter regimes.*

4. Lower bounds for proper colorings: proof overview

In Theorem 1.3 we establish that when $d \geq d_c(q) = q + \sqrt{q} + \Theta(1)$ there is no polynomial time identity testing algorithm for proper q-colorings in $\mathcal{M}(n, d)$ unless there is an FPRAS for #BIS. As mentioned, #BIS is the problem of counting independent sets in bipartite graphs; a standard complexity-theoretic assumption in approximate counting is that #BIS does not admit an FPRAS.

In our proof, we crucially use the hardness of #BIP-3-COL, the problem of counting proper 3-colorings in bipartite graphs. It is known that if there is an FPRAS for #BIP-3-COL, then there is also one for #BIS [24]. We show that when $d \geq d_c(q)$, an identity testing algorithm for proper q-colorings in $\mathcal{M}(n, d)$, with running time $T(n)$ and sample complexity $L(n)$, can be turned into a randomized algorithm for #BIP-3-COL on graphs of $\mathrm{poly}(n)$ size with running time $\mathrm{poly}(T(n), L(n))$. Theorem 1.3 follows from the fact that if $T(n)$ and $L(n)$ were both polynomials in n, then we would obtain an FPRAS for #BIP-3-COL.

To derive an algorithm for #BIP-3-COL we proceed as follows. Let H be an N-vertex connected bipartite graph, and suppose we want to compute an ε-approximation for the number of 3-colorings $Z_3(H)$ of H. Let B be the *complete* N-vertex bipartite graph with the same bipartition as H, and let $Z_3(B)$ denote the number of 3-colorings of B. Then, $Z_3(H) \in [Z_3(B), 3^N]$. We converge to an ε-approximation of $Z_3(H)$ via binary search in the interval $[Z_3(B), 3^N]$. Specifically, for $\hat{Z} \in [Z_3(B), 3^N]$ we construct a suitable identity testing instance and run the identity testing algorithm to determine whether we should consider larger or smaller values than $\hat{Z}$.

The testing instance is constructed as follows. For integers $k, \ell \geq 1$, we define the graph $\hat{H}_{k,\ell}$ that consists of k copies $H_1, \ldots, H_k$ of the original graph H and a complete $(q-3)$-partite graph J in which each cluster has ℓ vertices. In addition to the edges in J and in the k copies of H, $\hat{H}_{k,\ell}$ also contains edges between every vertex in J and every vertex in H_i for $i = 1, \ldots, k$. (Our definition of $\hat{H}_{k,\ell}$ requires $q \geq 4$; the case when $q = 3$ requires a slightly more complicated construction which is provided in [4].) For any $\hat{Z} \in [Z_3(B), 3^N]$, we choose k and ℓ in a way so that the output of the identity testing algorithm on $\hat{H}_{k,\ell}$ can be interpreted as feedback on whether or not $\hat{Z} > Z_3(H)$.

We set $k = \lceil N/\varepsilon \rceil$ where ε is the accuracy parameter. The choice of ℓ is more subtle. There are only two types of colorings for $\hat{H}_{k,\ell}$: (i) those where J uses $q-3$ colors and (ii) those where J uses $q-2$ colors. It can be easily checked that there are $|\Omega_1| = \Theta(Z_3(H)^k)$ colorings of the first type and $|\Omega_2| = \Theta(2^{\ell+k})$ of the second type. Hence, the choice of ℓ will determine which of these two types of colorings dominates in the uniform distribution $\mu_{k,\ell}$ over the proper colorings of $\hat{H}_{k,\ell}$.

To compare $\hat{Z}$ and $Z_3(H)$, we could set ℓ so that $\hat{Z}^k = |\Omega_2| = \Theta(2^{\ell+k})$ and draw a sample from $\mu_{k,\ell}$. If we get a coloring of the first kind, we may presume that $|\Omega_1| \gg |\Omega_2|$, or equivalently that $Z_3(H) > \hat{Z}$. Conversely, if the coloring is of the second kind, then it is likely that $|\Omega_1| \ll |\Omega_2|$ and $Z_3(H) < \hat{Z}$. Sampling from $\mu_{k,\ell}$ is hard, but we can emulate this approach with a testing algorithm.

Specifically, we construct a simpler graph $\hat{B}_{k,\ell}$ such that: (i) we can easily generate samples from $\hat{\mu}_{k,\ell}$, the uniform distribution over the proper q-colorings $\hat{B}_{k,\ell}$; and (ii) $\mu_{k,\ell}$ and $\hat{\mu}_{k,\ell}$ are close in total variation distance if and only if the dominant colorings in the Gibbs distributions are those of the second type. Then, we pass q, $\hat{H}_{k,\ell}$ and samples from $\hat{\mu}_{k,\ell}$ as input to the tester. Its output then reveals the dominant color class and hence whether $\hat{Z}$ is larger or smaller than $Z_3(H)$.

Our final obstacle is that the maximum degree of the graph $\hat{H}_{k,\ell}$ depends on N, k and ℓ, and could be much larger than d. To reduce the degree of $\hat{H}_{k,\ell}$ so that it belongs to $\mathcal{M}(n, d)$, we design a degree reducing gadget, which is inspired by the gadgets used to establish the hardness of the decision and structure learning problems [26, 44, 5]. The full proof of Theorem 1.3 is given in [4].

Acknowledgments

Research supported by NSF grants 1819546, CCF-1617306, CCF-1563838 and CCF-1563757.

References

[1] D.H. Ackley, G.E. Hinton, and T.J. Sejnowski. A Learning Algorithm for Boltzmann Machines. *Cognitive Science*, 9(1):147–169, 1985.

[2] A. Anandkumar, D.J. Hsu, F. Huang, and S.M. Kakade. Learning mixtures of tree graphical models. In *Advances in Neural Information Processing Systems (NeurIPS)*, pages 1052–1060, 2012.

[3] T. Batu, E. Fischer, L. Fortnow, R. Kumar, R. Rubinfeld, and P. White. Testing random variables for independence and identity. In *Proceedings of the 42nd Annual IEEE Symposium on Foundations of Computer Science (FOCS)*, pages 442–451, 2001.

[4] I. Bezáková, A. Blanca, Z. Chen, D. Štefankovič, and E. Vigoda. Lower bounds for testing graphical models: colorings and antiferromagnetic Ising models. *ArXiv preprint ArXiv:1901.07361*, 2019.

[5] A. Blanca, Z. Chen, D. Štefankovič, and E. Vigoda. Structure Learning of H-colorings. In *Proceedings of the 29th International Conference on Algorithmic Learning Theory (ALT)*, volume 83, pages 152–185, 2018.

[6] G. Bresler. Efficiently learning Ising models on arbitrary graphs. In *Proceedings of the 47th Annual ACM Symposium on Theory of Computing (STOC)*, pages 771–782, 2015.

[7] G. Bresler, E. Mossel, and A. Sly. Reconstruction of Markov random fields from samples: some observations and algorithms. *SIAM Journal on Computing*, 42(2):563–578, 2013.

[8] G. Bresler, D. Gamarnik, and D. Shah. Structure learning of antiferromagnetic Ising models. In *Advances in Neural Information Processing Systems (NeurIPS)*, pages 2852–2860, 2014.

[9] G. Bresler, D. Gamarnik, and D. Shah. Hardness of parameter estimation in graphical models. In *Advances in Neural Information Processing Systems (NeurIPS)*, pages 1062–1070, 2014.

[10] G. Brito, I. Dumitriu, and K.D. Harris. Spectral gap in random bipartite biregular graphs and its applications. *ArXiv preprint ArXiv:1804.07808*, 2018.

[11] A.A. Bulatov, M. Dyer, L.A. Goldberg, M. Jerrum, and C. McQuillan. The expressibility of functions on the Boolean domain, with applications to Counting CSPs. *Journal of the ACM (JACM)*, 60(5):32, 2013.

[12] J.-Y. Cai, A. Galanis, L.A. Goldberg, H. Guo, M. Jerrum, D. Štefankovič, and E. Vigoda. $\#BIS$-hardness for 2-spin systems on bipartite bounded degree graphs in the tree non-uniqueness region. *Journal of Computer and System Sciences*, 82(5):690–711, 2016.

[13] C. Calabro, R. Impagliazzo, V. Kabanets, and R. Paturi. The complexity of Unique k-SAT: An Isolation Lemma for k-CNFs. *Journal of Computer and System Sciences*, 74(3):386–393, 2008.

[14] C.L. Canonne, I. Diakonikolas, T. Gouleakis, and R. Rubinfeld. Testing Shape Restrictions of Discrete Distributions. *Theory of Computing Systems*, 62(1):4–62, 2018.

[15] X. Chen, M. Dyer, L.A. Goldberg, M. Jerrum, P. Lu, C. McQuillan, and D. Richerby. The complexity of approximating conservative counting CSPs. *Journal of Computer and System Sciences*, 81(1):311–329, 2015.

[16] C.K. Chow and C. Liu. Approximating discrete probability distributions with dependence trees. *IEEE Transactions on Information Theory*, 14(3):462–467, 1968.

[17] A. Collevecchio, T.M. Garoni, T. Hyndman, and D. Tokarev. The Worm process for the Ising model is rapidly mixing. *Journal of Statistical Physics*, 164(5):1082–1102, 2016.

[18] S. Dasgupta. Learning polytrees. In *Proceedings of the 15th Conference on Uncertainty in Artificial Intelligence (UAI)*, pages 134–141, 1999.

[19] C. Daskalakis, E. Mossel, and S. Roch. Evolutionary trees and the Ising model on the Bethe lattice: a proof of Steel's conjecture. *Probability Theory and Related Fields*, 149(1-2):149–189, 2011.

[20] C. Daskalakis, N. Dikkala, and G. Kamath. Testing Ising models. In *Proceedings of the 29th Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pages 1989–2007, 2018.

[21] I. Diakonikolas and D.M. Kane. A new approach for testing properties of discrete distributions. In *Proceedings of the 57th Annual IEEE Symposium on Foundations of Computer Science (FOCS)*, pages 685–694, 2016.

[22] I. Diakonikolas, D.M. Kane, and V. Nikishkin. Optimal algorithms and lower bounds for testing closeness of structured distributions. In *Proceedings of the 56th Annual IEEE Symposium on Foundations of Computer Science (FOCS)*, pages 1183–1202, 2015.

[23] I. Diakonikolas, T. Gouleakis, J. Peebles, and E. Price. Sample-Optimal Identity Testing with High Probability. In *Proceedings of the 45th International Colloquium on Automata, Languages, and Programming (ICALP)*, volume 1, pages 1–41, 2018.

[24] M. Dyer, L.A. Goldberg, C. Greenhill, and M. Jerrum. The relative complexity of approximate counting problems. *Algorithmica*, 38(3):471–500, 2004.

[25] M. Dyer, L.A. Goldberg, and M. Jerrum. An approximation trichotomy for Boolean #CSP. *Journal of Computer and System Sciences*, 76(3-4):267–277, 2010.

[26] T. Emden-Weinert, S. Hougardy, and B. Kreuter. Uniquely colourable graphs and the hardness of colouring graphs of large girth. *Combinatorics, Probability and Computing*, 7(4):375–386, 1998.

[27] J. Felsenstein. *Inferring phylogenies*, volume 2. Sinauer Associates, Inc., Sunderland, MA, 2004.

[28] S. Friedli and Y. Velenik. *Statistical mechanics of lattice systems: a concrete mathematical introduction*. Cambridge University Press, 2017.

[29] A. Galanis, L.A. Goldberg, and M. Jerrum. Approximately Counting H-Colourings is #BIS-Hard. *SIAM Journal on Computing*, 45(3):680–711, 2016.

[30] A. Galanis, D. Štefankovič, and E. Vigoda. Inapproximability of the partition function for the antiferromagnetic Ising and hard-core models. *Combinatorics, Probability and Computing*, 25(4):500–559, 2016.

[31] S. Geman and C. Graffigne. Markov random field image models and their applications to computer vision. In *Proceedings of the International Congress of Mathematicians*, volume 1, pages 1496–1517. Berkeley, CA, 1986.

[32] H.-O. Georgii. *Gibbs measures and phase transitions*, volume 9. Walter de Gruyter, 2011.

[33] L.A. Goldberg and M. Jerrum. The complexity of ferromagnetic Ising with local fields. *Combinatorics, Probability and Computing*, 16(1):43–61, 2007.

[34] L.A. Goldberg and M. Jerrum. Approximating the partition function of the ferromagnetic Potts model. *Journal of the ACM*, 59(5):25, 2012.

[35] L.A. Goldberg and M. Jerrum. Approximating pairwise correlations in the Ising Model. *ACM Transactions on Computation Theory.* To appear, 2019.

[36] O. Goldreich, S. Goldwasser, and D. Ron. Property testing and its connection to learning and approximation. *Journal of the ACM*, 45(4):653–750, 1998.

[37] H. Guo and M. Jerrum. Random cluster dynamics for the Ising model is rapidly mixing. In *Proceedings of the 28th Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pages 1818–1827, 2017.

[38] L. Hamilton, F. Koehler, and A. Moitra. Information theoretic properties of Markov random fields, and their algorithmic applications. In *Advances in Neural Information Processing Systems (NeurIPS)*, pages 2460–2469, 2017.

[39] S. Hoory, N. Linial, and A. Wigderson. Expander graphs and their applications. *Bulletin of the American Mathematical Society*, 43(4):439–561, 2006.

[40] R. Impagliazzo and R. Paturi. On the Complexity of k-SAT. *Journal of Computer and System Sciences*, 62(2):367–375, 2001.

[41] M. Jerrum and A. Sinclair. Polynomial-time approximation algorithms for the Ising model. *SIAM Journal on computing*, 22(5):1087–1116, 1993.

[42] A. Klivans and R. Meka. Learning graphical models using multiplicative weights. In *Proceedings of the 58th Annual Symposium on Foundations of Computer Science (FOCS)*, pages 343–354. IEEE, 2017.

[43] S.-I. Lee, V. Ganapathi, and D. Koller. Efficient Structure Learning of Markov Networks using L_1-Regularization. In *Advances in Neural Information Processing Systems (NeurIPS)*, pages 817–824, 2007.

[44] M. Molloy and B. Reed. Colouring graphs when the number of colours is nearly the maximum degree. In *Proceedings of the 33rd Annual ACM Symposium on Theory of Computing (STOC)*, pages 462–470, 2001.

[45] D. Randall and D. Wilson. Sampling spin configurations of an Ising system. In *Proceedings of the 10th Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pages 959–960, 1999.

[46] P. Ravikumar, M.J. Wainwright, and J.D. Lafferty. High-dimensional Ising model selection using ℓ_1-regularized logistic regression. *The Annals of Statistics*, 38(3):1287–1319, 2010.

[47] S. Roth and M.J. Black. Fields of experts: A framework for learning image priors. In *Proceedings of the 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR)*, volume 2, pages 860–867, 2005.

[48] R. Rubinfeld and M. Sudan. Robust characterizations of polynomials with applications to program testing. *SIAM Journal on Computing*, 25(2):252–271, 1996.

[49] R. Salakhutdinov and G. Hinton. An efficient learning procedure for Deep Boltzmann Machines. *Neural Computation*, 24(8):1967–2006, 2012.

[50] R. Salakhutdinov and H. Larochelle. Efficient learning of deep Boltzmann machines. In *Proceedings of the 13th International Conference on Artificial Intelligence and Statistics (AISTATS)*, pages 693–700, 2010.

[51] N.P. Santhanam and M.J. Wainwright. Information-theoretic limits of selecting binary graphical models in high dimensions. *IEEE Trans. Information Theory*, 58(7):4117–4134, 2012.

[52] A. Sly. Computational transition at the uniqueness threshold. In *Proceedings of the 51st Annual IEEE Symposium on Foundations of Computer Science (FOCS)*, pages 287–296, 2010.

[53] A. Sly and N. Sun. The computational hardness of counting in two-spin models on d-regular graphs. In *Proceedings of the 53rd Annual Symposium on Foundations of Computer Science (FOCS)*, pages 361–369, 2012.

[54] G. Valiant and P. Valiant. An automatic inequality prover and instance optimal identity testing. *SIAM Journal on Computing*, 46(1):429–455, 2017.

[55] M. Vuffray, S. Misra, A. Lokhov, and M. Chertkov. Interaction screening: Efficient and sample-optimal learning of Ising models. In *Advances in Neural Information Processing Systems (NeurIPS)*, pages 2595–2603, 2016.