Journal of Machine Learning Research 21 (2020) 1-62 Submitted 7/19; Revised 1/20; Published 2/20

Lower Bounds for Testing Graphical Models: Colorings and Antiferromagnetic Ising Models

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

Antonio Blanca ABLANCA@CSE.PSU.EDU
Pennsylvania State University
University Park, PA, USA

Zongchen Chen CHENZONGCHEN@GATECH.EDU
Georgia Institute of Technology
Atlanta, GA, USA

Daniel Štefankovič STEFANKO@CS.ROCHESTER.EDU
University of Rochester
Rochester, NY, USA

Eric Vigoda VIGODA@GATECH.EDU
Georgia Institute of Technology
Atlanta, GA, USA

Editor: Gabor Lugosi

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 any 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 polynomial running time identity testing algorithm unless $RP = NP$. 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.

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

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 $\mathcal{H}$ 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., Chow and Liu, 1968; Dasgupta, 1999; Lee et al., 2007; Anandkumar et al., 2012; Ravikumar et al., 2010; Bresler et al., 2013, 2014b; Bresler, 2015; Vuffray et al., 2016; Hamilton et al., 2017; Klivans and Meka, 2017). Once the graph G^* is known, it is often a simpler task to estimate $\mathcal{H}^*$ (Bresler, 2015); 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 settings where these two problems are hard in both the Ising model and proper colorings coincide.

1.1. 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 (Felsenstein, 2004; Daskalakis et al., 2011), computer vision (Geman and Graffigne, 1986; Roth and Black, 2005), statistical mechanics (Georgii, 2011; Friedli and Velenik, 2017) and deep learning, where it appears under the guise of Boltzmann machines (Ackley et al., 1985; Salakhutdinov and Larochelle, 2010; Salakhutdinov and Hinton, 2012). 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 exponentially close to each other (measured by some notion of statistical distance between distributions as a function of $|V|$), 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., Batu et al., 2001; Diakonikolas et al., 2015; Diakonikolas and Kane, 2016; Valiant and Valiant, 2017; Diakonikolas et al., 2018; Daskalakis et al., 2018; Canonne et al., 2018), we study this problem in the property testing framework (Rubinfeld and Sudan, 1996; Goldreich et al., 1998). 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 (Daskalakis et al., 2018). Therefore, since our focus is on lower bounds, we work with total variation distance, which we denote by $\|\cdot\|_{\mathrm{TV}}$.

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. $\|\mu_{G,\beta} - \mu_{G^*,\beta^*}\|_{\mathrm{TV}} > \frac{1}{3}$.

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 $\varepsilon \in (0,1)$, provided n is sufficiently large; see also Remark 16.

Identity testing for the Ising model was studied by Daskalakis, Dikkala and Kamath (2018) 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 $\beta^* \neq \beta$, as well as for some related open problems, see Section 8.) We mention that the hardness of identity testing for general spin systems was previously considered by Bogdanov, Mossel and Vadhan (2008), but in the harder setup where there are hidden variables (some of the spins are not observed).

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. Bresler (2015) presented a structure learning algorithm for the Ising model that can learn $G^* \in \mathcal{M}(n,d)$ in time $\mathrm{e}^{\exp(O(|\beta^*|d^2))} \times O(n^2 \log n)$ and sample complexity $\mathrm{e}^{\exp(O(|\beta^*|d^2))} \times O(\log n)$. The dependency on the parameters β^* and d was improved by Vuffray et al. (2016), who provide an algorithm to learn both G^* and β^* with running time $\mathrm{e}^{O(|\beta^*|d)} \times O(n^4 \log n)$ and sample complexity $\mathrm{e}^{O(|\beta^*|d)} \times O(\log n)$. More recently, Klivans and Meka (2017) provided a nearly optimal algorithm that learns G^* and β^* in time $\mathrm{e}^{O(|\beta^*|d)} \times O(n^2 \log n)$ using $\mathrm{e}^{O(|\beta^*|d)} \times O(\log n)$ samples; for other recent related work on the structure learning and parameter estimation problems for the Ising model see (Hamilton et al., 2017; Vuffray et al., 2019; Wu et al., 2019).

Consequently, when $|\beta^*|d = O(\log n)$, or when $\beta = \beta^*$ and $|\beta|d = O(\log n)$, structure learning algorithms provide 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 (Santhanam and Wainwright, 2012), and this approach to 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 *Suppose n, d are positive integers such that $3 \leq d \leq n^\theta$ 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 (2018) 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. However, the model in (Daskalakis et al., 2018) 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 (Goldberg and Jerrum, 2019) 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 (Sly and Sun, 2012; Galanis et al., 2016b) and in the ferromagnetic model in the presence of inconsistent magnetic fields (Goldberg and Jerrum, 2007) (i.e., the vertex potential of distinct vertices may have different signs). In summary, the algorithmic results of (Daskalakis et al., 2018) 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., Jerrum and Sinclair, 1993; Randall and Wilson, 1999; Guo and Jerrum, 2017; Collevecchio et al., 2016).

In Theorem 1 we assume that $|\beta|d = \omega(\log n)$, but our main technical result (Theorem 4) is actually more general. We show that when $|\beta|d \geq c \ln n$, 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 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* (Impagliazzo and Paturi, 2001; Calabro et al., 2008), our main theorem also implies a general lower bound for identity testing that holds for all β and d satisfying $|\beta|d \geq c \ln n$.

Theorem 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 $T(n) \geq \min\left\{\exp(\Omega(n^{\alpha})), \frac{\exp(\Omega(|\beta|d))}{n}\right\}$.*

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 (Santhanam and Wainwright, 2012), albeit requiring that *rETH* is true. We also mention that the lower bounds in Theorems 1 and 2 are both much stronger than the $\Omega(\sqrt{n})$ lower bound in (Daskalakis et al., 2018).

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 $1 - \delta$, $(1 - \varepsilon)\hat{Z} \leq Z \leq (1 + \varepsilon)\hat{Z}$ with running time polynomial in the instance size, ε^{-1} and $\log(\delta^{-1})$.

The very high level idea of the proof of our main theorem for the Ising model (Theorem 4), from which Theorems 1 and 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 (Sly, 2010). 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 2.1.

1.2. 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 spin systems. While identity testing does not seem to have been studied in this context before, the related structure learning problem has received some attention (Bresler et al., 2014a; Blanca et al., 2018). For proper colorings, it is known that when $q \geq d + 1$ the hidden graph G can be learned from $\text{poly}(n, d, q)$ 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 (Blanca et al., 2018). Moreover, for $d \geq d_c(q) = q + \lceil \sqrt{q} \rceil - 1$, or equivalently $q \leq d - \sqrt{d} + \Theta(1)$, it was also established in (Blanca et al., 2018) 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 $d_c(q)$ coincides exactly with the one for polynomial time/NP-completeness for the problem of determining if G is q-colorable (Emden-Weinert et al., 1998; Molloy and Reed, 2001); see (22) for the precise definition of $d_c(q)$.

We prove here that the identity testing problem 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 utilize the complexity of #BIS, which is the problem of counting the independent sets in a bipartite graph. #BIS is believed not to have an FPRAS, and it has achieved considerable interest in approximate counting as a tool for proving relative complexity hardness (see, e.g., Dyer et al., 2004; Goldberg and Jerrum, 2012; Dyer et al., 2010; Bulatov et al., 2013; Chen et al., 2015; Cai et al., 2016; Galanis et al., 2016a).

Theorem 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 (Emden-Weinert et al., 1998; Molloy and Reed, 2001; Blanca et al., 2018) 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.

1.3. An Algorithm for the Ferromagnetic Ising Model

We provide an improved algorithm for the *ferromagnetic* Ising model. As mentioned, by combining the algorithm in (Daskalakis et al., 2018) with previous results for sampling (see Jerrum and Sinclair, 1993; Randall and Wilson, 1999; Guo and Jerrum, 2017; Collevecchio et al., 2016), one obtains a polynomial running time algorithm for identity testing in the ferromagnetic setting. The algorithm in (Daskalakis et al., 2018) 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 (Daskalakis et al., 2018), as there is no dependence on the inverse temperature β. See Theorem 30 in Section 6 for a precise statement of this result.

The rest of the paper is organized as follows. In Section 2 we state our main technical theorem (Theorem 4), and we derive Theorems 1 and 2 as corollaries. In Section 2.1, we sketch the key ideas in the proof our main result. The actual proof of Theorem 4 is fleshed out in Section 4. Before that, we introduce a useful variant of the maximum cut problem and study its complexity in Section 3. In Section 5 we provide bounds for the edge expansion of random bipartite graphs which are crucially used in our proofs and could be of independent interest. Our algorithm for the ferromagnetic Ising model is analyzed in Section 6, and our results for proper q-colorings (Theorem 3) are derived in Section 7; specifically, the reduction for the $q \geq 4$ case is presented in Section 7.4, and the more elaborate construction for $q = 3$ is given in Section 7.5.

2. Lower Bounds for the Ising Model

To establish our lower bounds in Theorems 1 and 2 we use the computational hardness of the maximum cut (MaxCut) problem. Recall that in the search variant of this 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 and 2 are derived, is the following.

Theorem 4 *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, when $|\beta|d \geq c \ln n$, any identity testing algorithm with running time $T(n)$ and sample complexity $L(n)$ provides a randomized algorithm for MaxCut on graphs of poly(n) size with running time $O(T(n) + n \cdot L(n))$. The high level ideas in the proof of this theorem are described next in Section 2.1; its actual proof is fleshed out in Section 4. Several important consequences of this result, including Theorems 1 and 2 from the introduction, are derived in Section 2.2.

2.1. Main Result for the Ising Model: Proof Overview

To establish Theorem 4 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 $k \in \mathbb{N}$. At a high level, we use H and k to construct an identity testing instance consisting of an antiferromagnetic Ising model M, which plays the role of the visible model, and samples from a simpler model M^*. Our construction of these two models ensures that sampling from M^* is easy, and that the output of the testing algorithm reveals information about the maximum cuts of H. To implement this approach, we consider a variant of the MaxCut problem which we call the TwoLargeCuts problem. We shall see that this problem is also NP-hard (by a reduction from MaxCut).

An instance of the TwoLargeCuts problem is constructed as follows. Given $H = (V, E)$ and $k \in \mathbb{N}$, 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 $2N^2$.

In the TwoLargeCuts problem the goal is to determine whether there are at least two cuts in $\hat{H}_w$ of size at least $2N^2$ (see Definition 8). MaxCut can be reduced to TwoLargeCuts by treating w, the number of edges between s and t, as a parameter. Specifically, 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

$w = N^2 - k$; see Section 3 for the proof of this fact. This yields that the TwoLargeCuts problem is NP-complete and the following useful lemma.

Lemma 5 *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 solve the TwoLargeCuts problem, we can use the antiferromagnetic Ising model on $\hat{H}_w$ to determine if $(\{s, t\}, V)$ is the maximum cut of $\hat{H}_w$ 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 corresponding Ising configurations. The intuition is that the maximum cuts of $\hat{H}_w$ correspond to the configurations of maximum likelihood in the Gibbs distribution. Indeed, when $|\beta|$ is sufficiently large, the distribution will be well-concentrated on the configurations that correspond to the maximum cuts. Therefore, a sample from the Gibbs distribution would reveal with high probability whether or not $(\{s, t\}, V)$ is the maximum cut of $\hat{H}_w$ and thus solve the TwoLargeCuts problem.

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 a hard problem, and we would need to provide a sampling procedure. Instead, 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 $\hat{H}_w$. Our construction of M^* is facilitated by our construction of the TwoLargeCuts instance $\hat{H}_w$ where there is a trivial (easy to find) large cut; i.e., the cut $(\{s, t\}, V)$. (The precise construction of M and M^* is described in Section 4, and their desired properties are established in Sections 4.2 and 4.3.) 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 (Sly, 2010; Sly and Sun, 2012).

Every vertex of $\hat{H}_{w,\ell}$ is replaced by a random bipartite graph $G = (L \cup R, E_G)$; see Section 4 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 4 for the details.) In similar manner as described above for $\hat{H}_{w,\ell}$, an identity testing algorithm for the antiferromagnetic Ising model on $\hat{H}^{\Gamma}_{w}$ can be used to determine whether $(\{s,t\},V)$ is the unique large cut of $\hat{H}_{w}$; see Lemma 12. Since $\hat{H}^{\Gamma}_{w}$ has maximum degree at most d, Theorem 4 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; see Theorems 10 and 11. 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 (Brito et al., 2018). However, the case of $d = O(1)$ is more difficult, and it requires for us to define the notion of edge expansion with respect to the ports of the gadget and extending some of the ideas in (Hoory et al., 2006) (see Theorem 19). Our bounds for the edge expansion of random bipartite graphs may be of independent interest; see Section 5.

2.2. Consequences of Main Result for the Ising Model: Proofs of Theorems 1 and 2

In this section we show how to derive Theorems 1 and 2 from Theorem 4. For Theorem 1 we also use the fact that there is no randomized algorithm for MaxCut with polynomial running time unless $RP = NP$. (We recall that RP is the class of problems that can be solved in polynomial time by a randomized algorithm.) For Theorem 2 we use a stronger assumption, namely the (randomized) exponential time hypothesis (or $rETH$) (Impagliazzo and Paturi, 2001; Calabro et al., 2008).

Proof of Theorem 1 Suppose there is an identity testing algorithm for $\mathcal{M}(n,d)$ with $\mathrm{poly}(n)$ running time and sample complexity; that is, $L \le T = \mathrm{poly}(n)$. Since $|\beta|d = \omega(\ln n)$,

$$L \le \frac{\exp\left(|\beta|d/c\right)}{30n}.$$

Hence, Theorem 4 implies there is a randomized algorithm for MaxCut on graphs of size $N = \Theta(n^{\min\{\frac{\rho}{4},\frac{1}{14}\}})$ that succeeds with probability at least $11/20$ and has running time $O(T + Ln) = \mathrm{poly}(n)$. This implies that MaxCut is in BPP. (BPP is the class of all decision problems solvable in polynomial time with success probability greater than $1/2$ on both "yes" and "no" instances; in contrast, RP only allows errors on "no" instances.) Since MaxCut is NP-complete, then $NP \subseteq BPP$, and the result follows from the standard fact that if $NP \subseteq BPP$, then $RP = NP$; see, e.g., (Ko, 1982). ■

Proof of Theorem 2 Suppose there exists an identity testing algorithm with running time T and sample complexity L. If $L > \frac{\exp(|\beta|d/c)}{30n}$, then

$$T \ge L > \frac{\exp\left(|\beta|d/c\right)}{30n}$$

and the result follows. Otherwise, when $|\beta|d \ge c \ln n$ for a suitable constant $c = c(\rho) > 0$, Theorem 4 implies that there exists a randomized algorithm for MaxCut on graphs with

$N = \Theta(n^{\min\{\frac{\rho}{4}, \frac{1}{14}\}})$ vertices with running time at most $O(T + Ln)$ and success probability at least 11/20. However, under the assumption that $rETH$ is true, there is no randomized algorithm for MaxCut in such graphs with running time $\mathrm{e}^{o(n^\alpha)}$, where $\alpha = \min\{\frac{\rho}{4}, \frac{1}{14}\}$. Thus, there exist constants $\delta, \gamma > 0$ such that

$$\delta(T + Ln) \geq \mathrm{e}^{\gamma n^\alpha}.$$

Consequently, if $L \leq \frac{\mathrm{e}^{\gamma n^\alpha}}{2\delta n}$, then $T \geq \frac{\mathrm{e}^{\gamma n^\alpha}}{2\delta}$; otherwise $T \geq L \geq \frac{\mathrm{e}^{\gamma n^\alpha}}{2\delta n}$. Putting these bounds together we get

$$T \geq \min\left\{\frac{\exp(|\beta|d/c)}{30n}, \frac{\exp(\gamma n^\alpha)}{2\delta n}\right\},$$

and the result follows. ■

Finally, we note that Theorem 4 also implies a polynomial (in n) lower bound for the running time of any identity testing algorithm when $|\beta|d = \Theta(\log n)$. This regime is not covered by Theorem 1, where the assumption is that $|\beta|d = \omega(\log n)$, and Theorem 2 applies to this setting, but under the stronger $rETH$ assumption. Our next theorem shows that the weaker complexity theoretic assumption $RP \neq NP$ suffices.

Theorem 6 *Suppose n, d are positive integers such $3 \leq d \leq n^{1-\rho}$ for some constant $\rho \in (0, 1)$ and $\beta < 0$ is such that $|\beta|d > c \ln n$, where $c = c(\rho)$ is the constant from Theorem 4. If $RP \neq NP$, then, for all n sufficiently large, any algorithm that solves the identity testing problem for $\mathcal{M}(n, d)$ for the antiferromagnetic Ising model has running time $T = \Omega(n^\Delta)$, where $\Delta = \frac{|\beta|d}{c \ln n} - 1$.*

Proof Suppose there is an identity testing algorithm for $\mathcal{M}(n, d)$ with running time T and sample complexity L. We consider two cases. First, if $L \leq \frac{\exp(|\beta|d/c)}{30n}$, then Theorem 4 implies that there is a randomized algorithm for MaxCut on graphs with $N = \Theta(n^{\min\{\frac{\rho}{4}, \frac{1}{14}\}})$ vertices that has running time $O(T + Ln)$ and success probability 11/20. Therefore, $T + Ln = n^{\omega(1)}$ since otherwise $NP \subseteq BPP$ and thus $RP = NP$ (Ko, 1982). Hence, if $Ln = O(\mathrm{poly}(n))$, then $T = \Omega(n^{\omega(1)})$; otherwise $T \geq L = \Omega(n^{\omega(1)})$. For the second case, when $L > \frac{\exp(|\beta|d/c)}{30n}$, we have

$$T \geq L > \frac{\exp\left(|\beta|d/c\right)}{30n} = \frac{n^\Delta}{30},$$

and the result follows. ■

3. Hardness of the TwoLargeCuts Problem

In this section we prove Lemma 5, where the hardness of the TwoLargeCuts problem is established. Recall that the TwoLargeCuts problem is a variant of the MaxCut problem which we will reduce to the identity testing problem. We formally define the TwoLargeCuts problem next.

Definition 7 *Let $H = (V, E)$ be a graph and let $w \in \mathbb{N}$. Let $\hat{H}_w = (\hat{V}, \hat{E})$ be the* multigraph *defined as follows:*

1. *$\hat{V}$ contains all vertices in V and two new vertices s and t; i.e., $\hat{V} = V \cup \{s, t\}$;*
2. *$\hat{E}$ contains all edges in E, N copies of edges $\{s, v\}$ and $\{t, v\}$ for each $v \in V$, and w copies of the edge $\{s, t\}$.*

Observe that the cut $(\{s, t\}, V)$ contains exactly $2N^2$ edges.

Definition 8 *In the* TwoLargeCuts *problem, given a 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 $2N^2$.*

Lemma 5 is a direct corollary of the following lemma, which implies that MaxCut can be reduced to TwoLargeCuts.

Lemma 9 *Let $H = (V, E)$ be an N-vertex graph and let $w \in \mathbb{N}$. The cut $(\{s, t\}, V)$ is the* unique *maximum cut of $\hat{H}_w$ if and only if* MaxCut$(H) < N^2 - w$.

Consequently, to solve MaxCut on inputs H and k, it is sufficient to solve the TwoLargeCuts problem for $\hat{H}_w$ with $w = N^2 - k$. Hence, Lemma 5 is a direct corollary of Lemma 9. (Note that Lemma 9 also implies that the TwoLargeCuts problem is NP-complete.)

Proof of Lemma 9 Let (S, T) be a cut of $\hat{H}_w$ (i.e., $S \cup T = \hat{V}$ and $S \cap T = \emptyset$) and let $E_{\hat{H}_w}(S, T) \subseteq \hat{E}$ be the set of edges between S and T in $\hat{H}_w$. Similarly, for $S', T' \subseteq V$, let $E_H(S', T') \subseteq E$ be the set of edges between S' and T' in H.

Let us consider first the cuts (S, T) where s and t belong to the same set. Without loss of generality assume $s, t \in S$, and let $S_0 = S \setminus \{s, t\}$. Then, (S_0, T) is a cut of H and so

$$\left|E_{\hat{H}_w}(S, T)\right| = \left|E_{\hat{H}_w}(S_0, T)\right| + \left|E_{\hat{H}_w}(\{s, t\}, T)\right| = |E_H(S_0, T)| + 2N|T| \leq (N - |T|)|T| + 2N|T|.$$

The quadratic function $f(x) = (N - x)x + 2Nx$ is maximized at $x = N$ for $0 \leq x \leq N$ and $f(N) = 2N^2$. Thus, $\left|E_{\hat{H}_w}(S, T)\right| \leq 2N^2$, and the maximum value $2N^2$ can be attained only when $|T| = N$; i.e., $S_0 = \emptyset$ and $(S, T) = (\{s, t\}, V)$.

Now, for the cuts where s and t belong to distinct sets of the cut, let us assume without loss of generality that $s \in S$ and $t \in T$. Let $S_0 = S \setminus \{s\}$ and $T_0 = T \setminus \{t\}$. Then, (S_0, T_0) is a cut of H, and

$$\begin{aligned}\left|E_{\hat{H}_w}(S, T)\right| &= \left|E_{\hat{H}_w}(S_0, T_0)\right| + \left|E_{\hat{H}_w}(S_0, \{t\})\right| + \left|E_{\hat{H}_w}(\{s\}, T_0)\right| + \left|E_{\hat{H}_w}(\{s\}, \{t\})\right| \\ &= |E_H(S_0, T_0)| + N^2 + w.\end{aligned}$$

Hence, the maximum cut of this class corresponds to the case when (S_0, T_0) is a maximum cut of H, and

$$\left|E_{\hat{H}_w}(S, T)\right| = \text{MaxCut}(H) + N^2 + w.$$

Combining the above two cases, we conclude that $(\{s, t\}, V)$ is the unique maximum cut of $\hat{H}_w$ if and only if $2N^2 >$ MaxCut$(H) + N^2 + w$, and the result follows. ■

4. Proof of Main Result for the Ising Model: Theorem 4

We describe our gadget for the Ising model first. 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. Let P be subset of V_G chosen 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 $E_G = \left(\bigcup_{i=1}^{d_{\text{IN}}} M_i\right) \cup \left(\bigcup_{i=1}^{d_{\text{OUT}}} M'_i\right)$;
6. Make the graph G simple by replacing multiple edges with single edges.

We use $\mathcal{G}(m, p, d_{\text{IN}}, d_{\text{OUT}})$ to denote the resulting distribution; that is, $G \sim \mathcal{G}(m, p, d_{\text{IN}}, d_{\text{OUT}})$. Vertices in P are called *ports*. Every port has degree at most d_{IN} while every non-port vertex 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; see Section 5. 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 "−". To capture the notion of an external configuration for the bipartite graph G, we assume that 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 10 *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:*

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

Theorem 11 *Suppose $\beta < 0$, $p = m$ and $4 + \frac{1200}{\rho} \le d \le m^{1-\rho}$ for some constant $\rho \in (0,1)$ independent of m. Then, there exist constants $\delta = \delta(\rho) > 0$ and $\theta = \theta(\rho) \in (0,1)$ such that when $d_{\text{IN}} = \lfloor \theta d \rfloor$ and $d_{\text{OUT}} = d - \lfloor \theta d \rfloor$ the following holds for every configuration τ on ∂P with probability $1 - o(1)$ over the choice of the random graph G:*

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

The proofs of these theorems are given in Section 5.

4.1. Testing Instance Construction

Let $H = (V, E)$ be a simple N-vertex graph and for $w \le N^2$ let $\hat{H}_w$ be the multigraph from Definition 7. 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 $\hat{H}_w^\Gamma$, where Γ denotes the set parameters $\{m, p, d_{\text{IN}}, d_{\text{OUT}}, \ell\}$; $\ell > 0$ is assumed to be an integer divisible by d_{OUT}. The graph $\hat{H}_w^\Gamma$ 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 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:

$$d_{\text{IN}} + d_{\text{OUT}} = d \le m, \tag{2}$$
$$d_{\text{OUT}} \mid \ell, \tag{3}$$
$$\ell(N^2 + w) \le p \cdot d_{\text{OUT}}, \tag{4}$$
$$d_{\text{OUT}}^2 \le \ell. \tag{5}$$

To see that (4) 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 (5) 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 $\hat{H}_w^\Gamma$ is $2m(N+2)$ and its maximum degree is $d = d_{\text{IN}} + d_{\text{OUT}}$; thus, $\hat{H}_w^\Gamma \in \mathcal{M}(2m(N+2), d)$. Let I be an independent set 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. Our testing instance will consist of the model M and (approximate) samples from M^*. 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 $\hat{H}_w$.

4.2. Relating the Ising Models M and M^*

To formally study the relationship between the models M and M^* 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 Ising models M and M^* are related as follows.

Lemma 12 *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 conditions (2)—(5) are satisfied. If for the Ising model $M = (\hat{H}_w^\Gamma, \beta)$ we have $Z_M(\Omega_{\text{good}}) \geq (1-\varepsilon)Z_M$ 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 maximum 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 maximum cut of $\hat{H}_w$, then*

$$\|\mu_M - \mu_{M^*}\|_{\text{TV}} > \frac{1}{2} - \varepsilon - \mathrm{e}^{-2|\beta|d}.$$

3. *If there is a cut in $\hat{H}_w$ with strictly more edges than $(\{s,t\}, V)$, then*

$$\|\mu_M - \mu_{M^*}\|_{\text{TV}} \geq 1 - \varepsilon - 2\mathrm{e}^{-2|\beta|d}.$$

Let $\sigma^+ = \sigma^+(\hat{H}_w^\Gamma)$ be the configuration of $\hat{H}_w^\Gamma$ such that the gadgets for s and t are in the plus phase and every other gadget is in the minus phase; define $\sigma^- = \sigma^-(\hat{H}_w^\Gamma)$ in similar manner but interchanging "+" and "−" everywhere. Let $\Omega^0 = \Omega^0(\hat{H}_w^\Gamma) = \{\sigma^+, \sigma^-\}$. We will use the following fact to prove Lemma 12.

Fact 13 *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 conditions (2)—(5) 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 $\mu_{M^*}(\Omega^0) \geq 1 - \varepsilon - \mathrm{e}^{-2|\beta|d}$.*

Proof The weight of the configurations σ^+, σ^- satisfy: $w_{M^*}(\sigma^+) = w_{M^*}(\sigma^-) = 1$. If $\sigma \in \Omega_{\text{good}} \setminus \Omega^0$, then the gadget for either s or t is connected to the gadget of at least one other vertex in the same phase by $2\ell N$ edges. Hence, $w_{M^*}(\sigma) \leq \mathrm{e}^{2\beta\ell N} = \mathrm{e}^{-2|\beta|\ell N}$ and

$$Z_{M^*}(\Omega_{\text{good}} \setminus \Omega^0) = \sum_{\sigma\in\Omega_{\text{good}}\setminus\Omega_0} w_{M^*}(\sigma) \leq |\Omega_{\text{good}}| \cdot \mathrm{e}^{-2|\beta|\ell N} = 2^{N+2} \cdot \mathrm{e}^{-2|\beta|\ell N} \leq \mathrm{e}^{-2|\beta|d},$$

where in the last inequality we used the fact that $|\beta|(\ell N - d) \geq N$ by assumption. Then,

$$Z_{M^*}(\Omega_{\text{good}}) \leq 2 + \mathrm{e}^{-2|\beta|d}$$

and so

$$\mu_{M^*}(\Omega^0) = \frac{2}{Z_{M^*}(\Omega_{\text{good}})} \cdot \frac{Z_{M^*}(\Omega_{\text{good}})}{Z_{M^*}} \geq \left(1 - \mathrm{e}^{-2|\beta|d}\right) \frac{Z_{M^*}(\Omega_{\text{good}})}{Z_{M^*}} \geq 1 - \mathrm{e}^{-2|\beta|d} - \varepsilon,$$

as claimed. ■

We are now ready to prove Lemma 12.

Proof of Lemma 12 We show that when $(\{s,t\}, V)$ is the unique maximum cut of $\hat{H}_w$, then

$$\mu_M(\Omega^0) \geq 1 - \varepsilon - \mathrm{e}^{-2|\beta|d}. \tag{6}$$

Since by symmetry $\mu_M(\sigma^+) = \mu_M(\sigma^-)$ and $\mu_{M^*}(\sigma^+) = \mu_{M^*}(\sigma^-)$, Fact 13 implies

$$\|\mu_M - \mu_{M^*}\|_{\text{TV}} \leq \left|\mu_M(\sigma^+) - \mu_{M^*}(\sigma^+)\right| + \frac{\mu_M(\Omega \setminus \Omega^0) + \mu_{M^*}(\Omega \setminus \Omega^0)}{2} \leq 2(\varepsilon + \mathrm{e}^{-2|\beta|d})$$

and part 1 follows. (Recall that Ω is the set of Ising configurations of the graphs $\hat{H}_w^\Gamma$ and $\hat{I}_0^\Gamma$.)

To establish (6), observe that

$$\mu_M(\Omega^0) = \frac{Z_M(\Omega^0)}{Z_M(\Omega_{\text{good}})} \cdot \frac{Z_M(\Omega_{\text{good}})}{Z_M} \geq \frac{(1-\varepsilon) Z_M(\Omega^0)}{Z_M(\Omega_{\text{good}})}, \tag{7}$$

where the last inequality follows from the assumption that $Z_M(\Omega_{\text{good}}) \geq (1-\varepsilon) Z_M$.

For $\sigma \in \Omega_{\text{good}}$, let $\mathcal{I}(\sigma)$ be the number of edges $\{u,v\}$ of $\hat{H}_w$ such that the gadgets corresponding to vertices u and v in $\hat{H}_w^\Gamma$ are in the same phase in σ. Since every edge of $\hat{H}_w$ correspond to exactly 2ℓ edges in $\hat{H}_w^\Gamma$, we have $w_M(\sigma) = \mathrm{e}^{2\beta\ell\mathcal{I}(\sigma)} = \mathrm{e}^{-2|\beta|\ell\mathcal{I}(\sigma)}$. Moreover, $\mathcal{I}(\sigma^+) = \mathcal{I}(\sigma^-) = w + |E|$, where E is the set of edges of the graph H. When $(\{s,t\}, V)$ is the unique maximum cut of $\hat{H}_w$, $\mathcal{I}(\sigma) \geq w + |E| + 1$ for all $\sigma \in \Omega_{\text{good}} \setminus \Omega_0$. Therefore, $Z_M(\Omega^0) = 2\mathrm{e}^{-2|\beta|\ell(w+|E|)}$ and for $\sigma \in \Omega_{\text{good}} \setminus \Omega_0$

$$w_M(\sigma) \leq \mathrm{e}^{-2|\beta|\ell(w+|E|+1)} = \frac{Z_M(\Omega^0)}{2\mathrm{e}^{2|\beta|\ell}}.$$

Then,

$$Z_M(\Omega_{\text{good}}) = Z_M(\Omega^0) + \sum_{\sigma \in \Omega_{\text{good}} \setminus \Omega_0} w_M(\sigma) \leq Z_M(\Omega^0) + |\Omega_{\text{good}}| \cdot \frac{Z_M(\Omega^0)}{2\mathrm{e}^{2|\beta|\ell}} = Z_M(\Omega^0) \left(1 + \frac{2^{N+1}}{\mathrm{e}^{2|\beta|\ell}}\right).$$

By assumption $|\beta|(\ell - d) \geq N$, so $Z_M(\Omega_{\text{good}}) \leq Z_M(\Omega^0)\left(1 + \mathrm{e}^{-2|\beta|d}\right)$. Thus, we deduce that

$$\frac{Z_M(\Omega^0)}{Z_M(\Omega_{\text{good}})} \geq \frac{1}{1 + \mathrm{e}^{-2|\beta|d}} \geq 1 - \mathrm{e}^{-2|\beta|d}.$$

Plugging this bound into (7) gives (6) and the proof of part 1 of the lemma is complete.

For the second part we show that when $(\{s,t\},V)$ is not the unique maximum cut of $\hat{H}_w$, then

$$\mu_M(\Omega^0) \le \frac{1}{2}. \tag{8}$$

By Fact 13, $\mu_{M^*}(\Omega^0) \ge 1-\varepsilon-\mathrm{e}^{-2|\beta|d}$; hence,

$$\|\mu_M - \mu_{M^*}\|_{\mathrm{TV}} \ge \left|\mu_{M^*}(\Omega^0) - \mu_M(\Omega^0)\right| \ge \frac{1}{2} - \varepsilon - \mathrm{e}^{-2|\beta|d}$$

and part 2 follows.

To establish (8), let $(S,\hat{V}\setminus S) \neq (\{s,t\},V)$ be a maximum cut of the graph $\hat{H}_w$. Let σ_*^+ (resp., σ_*^-) be the Ising configuration of the graph $\hat{H}_w^\Gamma$ where the gadgets corresponding to vertices in S are in the plus phase (resp., minus phase), and the remaining gadgets are in the minus phase (resp., plus phase). Since $(S,\hat{V}\setminus S)$ is a maximum cut of $\hat{H}_w$, $\mathcal{I}(\sigma_*^+) = \mathcal{I}(\sigma_*^-) \le \mathcal{I}(\sigma^+) = \mathcal{I}(\sigma^-)$ and so $w_M(\{\sigma_*^+,\sigma_*^-\}) \ge w_M(\Omega^0)$. It follows that

$$Z_M \ge w_M(\Omega^0) + w_M(\sigma_*^+,\sigma_*^-) \ge 2w_M(\Omega^0)$$

and $\mu_M(\Omega^0) = w_M(\Omega^0)/Z_M \le 1/2$; this gives (8) and part 2 follows.

Part 3 follows in similar fashion. Let $(S,\hat{V}\setminus S)$ be a cut of $\hat{H}_w$ with strictly more edges than the cut $(\{s,t\},V)$. Let σ_*^+ (resp., σ_*^-) be Ising configuration of $\hat{H}_w^\Gamma$ determined by $(S,\hat{V}\setminus S)$ as in the proof of part 2. Then, $\mathcal{I}(\sigma_*^+) = \mathcal{I}(\sigma_*^-) < \mathcal{I}(\sigma^+) = \mathcal{I}(\sigma^-)$ and $w_M(\{\sigma_*^+,\sigma_*^-\}) \ge \mathrm{e}^{2|\beta|\ell} w_M(\Omega^0)$. It follows that

$$Z_M \ge w_M(\Omega^0) + w_M(\sigma_*^+,\sigma_*^-) \ge (1+\mathrm{e}^{2|\beta|\ell}) w_M(\Omega^0)$$

and

$$\mu_M(\Omega^0) = \frac{w_M(\Omega^0)}{Z_M} \le \frac{1}{1+\mathrm{e}^{2|\beta|\ell}} \le \mathrm{e}^{-2|\beta|\ell} \le \mathrm{e}^{-2|\beta|d},$$

where in the last inequality we use the assumption that $|\beta|(\ell-d) \ge N$ and so $\ell \ge d$. This bound and Fact 13 imply

$$\|\mu_M - \mu_{M^*}\|_{\mathrm{TV}} \ge \left|\mu_{M^*}(\Omega^0) - \mu_M(\Omega^0)\right| \ge 1 - \varepsilon - 2\mathrm{e}^{-2|\beta|d},$$

as claimed. ■

4.3. Proof of Theorem 4

In Theorem 4 we show that, under some mild assumptions, any identity testing algorithm with running time $T(n)$ and sample complexity $L(n)$ provides a randomized algorithm for MaxCut on graphs of poly(n) size with running time $O(T(n)+n\cdot L(n))$ when $|\beta|d \ge c\ln n$. For convenience, we restate Theorem 4 here.

Theorem 4 *Suppose n and d are positive integers such $3 \le d \le n^{1-\rho}$ 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 $|\beta|d \ge c\ln n$, 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* $O(T(n) + n \cdot L(n))$.

The testing instance in our reduction will consist of the model M and (approximate) samples from M^*. Hence, in our proof of Theorem 4, it will be crucial that we can easily generate samples from the simpler model M^*. This is established next.

Lemma 14 *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 conditions (2)—(5) 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:*

$$\|\mu_{M^*} - \mu_{\text{ALG}}\|_{\text{TV}} \leq \varepsilon + \mathrm{e}^{-2|\beta|d}.$$

Proof By Fact 13, $\mu_{M^*}(\Omega^0) \geq 1 - \varepsilon - \mathrm{e}^{-2|\beta|d}$. Also, $\mu_{M^*}(\sigma^+) = \mu_{M^*}(\sigma^-) = \mu_{M^*}(\Omega^0)/2$. Hence, if μ_{ALG} is the uniform distribution over $\{\sigma^+, \sigma^-\}$, we have

$$\|\mu_{M^*} - \mu_{\text{ALG}}\|_{\text{TV}} = \left|\mu_{M^*}(\sigma^+) - \frac{1}{2}\right| + \frac{1 - \mu_{M^*}(\Omega^0)}{2} \leq \varepsilon + \mathrm{e}^{-2|\beta|d}.$$

The results follows from the fact that a sample from μ_{ALG} can be generated in $O(mN)$ time. ■

We are now ready to prove Theorem 4.

Proof of Theorem 4 Let us assume first that $3 \leq d = O(1)$. In this case, we take

$$N = \left\lfloor n^{1/14} \right\rfloor - 2, \quad \text{and} \quad m = \left\lfloor \frac{n^{13/14}}{2} \right\rfloor.$$

If $n^{1/4}$ and $\frac{n^{13/14}}{2}$ are both integers, then $n = 2m(N+2)$. For simplicity and without much loss of generality, we assume that this is indeed the case. See Remark 15 for a brief explanation on how to extend the current proof to the case when $n^{1/4}$ or $\frac{n^{13/14}}{2}$ are not integers.

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

We recall that in the TwoLargeCuts problem the goal is to determine whether $(\{s,t\}, V)$ is the unique maximum cut of the graph $\hat{H}_w$; see Definitions 7 and 8. 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 from μ_M. Theorem 10 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 $v \in \hat{V}$ 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 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 $1 - \frac{2m(N+2)}{\mathrm{e}^{\delta|\beta|d}}$. The same is true if σ were sampled from μ_{M^*} instead. Thus,

$$\mu_M(\Omega_{\text{good}}) = \frac{Z_M(\Omega_{\text{good}})}{Z_M} \geq 1 - \frac{2m(N+2)}{\mathrm{e}^{\delta|\beta|d}}, \text{ and} \tag{9}$$

$$\mu_{M^*}(\Omega_{\text{good}}) = \frac{Z_{M^*}(\Omega_{\text{good}})}{Z_{M^*}} \geq 1 - \frac{2m(N+2)}{\mathrm{e}^{\delta|\beta|d}}. \tag{10}$$

Our choices for N and Γ satisfy conditions (2)—(5). It can also be checked that $|\beta|(\ell N - d) \geq N$ when $|\beta|d \geq c \ln n$. Then, (10) and Lemma 14 imply that we can generate L samples $\mathcal{S} = \{\sigma_1, \ldots, \sigma_L\}$ from a distribution μ_{ALG} in $O(nL)$ time such that

$$\|\mu_{M^*} - \mu_{\text{ALG}}\|_{\text{TV}} \leq \frac{2m(N+2)}{\mathrm{e}^{\delta|\beta|d}} + \frac{1}{\mathrm{e}^{2|\beta|d}} \leq \frac{2n}{\mathrm{e}^{\gamma|\beta|d}}, \tag{11}$$

where $\gamma = \min\{2, \delta\}$.

Our algorithm for TwoLargeCuts inputs the Ising model M and the L samples $\mathcal{S}$ to the Tester and outputs the negation of the Tester's output. Recall that the Tester returns Yes if it regards the samples in $\mathcal{S}$ as samples from μ_M; it returns No if it regards them to be from some other distribution ν such that $\|\mu_M - \nu\|_{\text{TV}} > 1/3$.

If $(\{s, t\}, V)$ is the unique maximum cut of $\hat{H}_w$, then (9) and part 2 of Lemma 12 imply:

$$\|\mu_M - \mu_{M^*}\|_{\text{TV}} \leq 2\left(\frac{2m(N+2)}{\mathrm{e}^{\delta|\beta|d}} + \frac{1}{\mathrm{e}^{2|\beta|d}}\right) \leq \frac{4n}{\mathrm{e}^{\gamma|\beta|d}}.$$

The triangle inequality and (11) imply:

$$\|\mu_M - \mu_{\text{ALG}}\|_{\text{TV}} \leq \|\mu_M - \mu_{M^*}\|_{\text{TV}} + \|\mu_{M^*} - \mu_{\text{ALG}}\|_{\text{TV}} \leq \frac{6n}{\mathrm{e}^{\gamma|\beta|d}}.$$

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. When $c > 1/\gamma$, we have

$$\left\|\mu_M^{\otimes L} - \mu_{\text{ALG}}^{\otimes L}\right\|_{\text{TV}} \leq L\,\|\mu_M - \mu_{\text{ALG}}\|_{\text{TV}} \leq \frac{\mathrm{e}^{|\beta|d/c}}{30n} \cdot \frac{6n}{\mathrm{e}^{\gamma|\beta|d}} \leq \frac{1}{5}.$$

Note that if $\pi^{\otimes L}$ is the optimal coupling of the distributions $\mu_{\text{ALG}}^{\otimes L}$ and $\mu_M^{\otimes L}$, and $(\mathcal{S}, \mathcal{S}')$ is sampled from $\pi^{\otimes L}$, then $\mathcal{S}' = \mathcal{S}$ with probability at least $4/5$, $\mathcal{S} \sim \mu_{\text{ALG}}^{\otimes L}$ and $\mathcal{S}' \sim \mu_M^{\otimes L}$.

Therefore,

$$\begin{aligned}
&\Pr[\textsc{Tester} \text{ outputs } \textsc{No} \text{ when given samples } \mathcal{S} \text{ where } \mathcal{S} \sim \mu_{\text{ALG}}^{\otimes L}] \\
&\quad = \Pr[\textsc{Tester} \text{ outputs } \textsc{No} \text{ when given samples } \mathcal{S} \text{ where } (\mathcal{S}, \mathcal{S}') \sim \pi^{\otimes L}] \\
&\quad \le \Pr[\textsc{Tester} \text{ outputs } \textsc{No} \text{ when given samples } \mathcal{S}' \text{ where } (\mathcal{S}, \mathcal{S}') \sim \pi^{\otimes L}] + \pi^{\otimes L}(S \neq S') \\
&\quad = \Pr[\textsc{Tester} \text{ outputs } \textsc{No} \text{ when given samples } \mathcal{S}' \text{ where } \mathcal{S}' \sim \mu_M^{\otimes L}] + \pi^{\otimes L}(S \neq S') \\
&\quad \le \frac{1}{4} + \frac{1}{5} = \frac{9}{20}. \qquad (12)
\end{aligned}$$

Hence, the Tester returns Yes with probability at least 11/20 in this case.

When $(\{s,t\}, V)$ is not the unique maximum cut of $\hat{H}_w$, (9) and the second part of Lemma 12 imply

$$\|\mu_M - \mu_{M^*}\|_{\text{TV}} > \frac{1}{2} - \frac{n}{\mathrm{e}^{\delta|\beta|d}} - \frac{1}{\mathrm{e}^{2|\beta|d}} > \frac{1}{3}, \qquad (13)$$

where the last inequality holds for n large enough, since by assumption that $|\beta|d \ge c \ln n$ and we chose $c > 1/\gamma$. Moreover, from (11) we get

$$\left\|\mu_{M^*}^{\otimes L} - \mu_{\text{ALG}}^{\otimes L}\right\|_{\text{TV}} \le L \left\|\mu_{M^*} - \mu_{\text{ALG}}\right\|_{\text{TV}} \le \frac{1}{15}.$$

Thus, analogously to (12) (i.e., using the optimal coupling between $\mu_{\text{ALG}}^{\otimes L}$ and $\mu_{M^*}^{\otimes L}$), we deduce that

$$\Pr\left[\textsc{Tester} \text{ outputs } \textsc{Yes} \text{ when given samples } \mathcal{S} \text{ where } \mathcal{S} \sim \mu_{\text{ALG}}^{\otimes L}\right] \le \frac{1}{4} + \frac{14}{15} \le \frac{19}{60}.$$

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 results for the case when $3 \le d = O(1)$ then follows from Corollary 5 and the fact that $|V| = N = \lfloor n^{1/14} \rfloor - 2 \ge \lfloor n^{\min\{\frac{\rho}{4}, \frac{1}{14}\}} \rfloor - 2$.

Now, for d such that $d \le n^{1-\rho}$ but $d \to \infty$, we take

$$N = \left\lfloor n^{\rho/4} \right\rfloor - 2, \quad m = \left\lfloor \frac{n^{1-\rho/4}}{2} \right\rfloor, \quad \text{and} \quad \Gamma = \{m, m, \lfloor \theta d \rfloor, d - \lfloor \theta d \rfloor, \Theta(n^{1-\frac{3\rho}{4}})\},$$

where $\theta = \theta(\rho)$ is a suitable constant. That is, $p = m$, $d_{\text{IN}} = \lfloor \theta d \rfloor$, $d_{\text{OUT}} = d - \lfloor \theta d \rfloor$ and $\ell = \Theta(n^{1-\frac{3\rho}{4}})$. These choices for N, m and Γ satisfy conditions (2)—(5). Hence, (9) and (10) can be deduced in similar fashion using Theorem 11 instead. The rest of the proof remains unchanged for this case. Note that for this choice of parameters, $|V| = N = \lfloor n^{\rho/4} \rfloor - 2 \ge \lfloor n^{\min\{\frac{\rho}{4}, \frac{1}{14}\}} \rfloor - 2$. ■

Remark 15 *When either $n^{1/4}$ or $\frac{n^{13/14}}{2}$ is not an integer, then $2(m+1)(N+3) \ge n > 2m(N+2)$, and an identity testing algorithm for $\mathcal{M}(n,d)$ with running time T and sample complexity L can be used to solve the same problem for $\mathcal{M}(2m(N+2), d)$ by simply "padding" the graph from $\mathcal{M}(2m(N+2), d)$ with $n - 2m(N+2)$ isolated vertices. The samples from the hidden distribution can be extended by adding isolated vertices and independently assigning*

"+" or "−" with probability $1/2$ *to each of them. Hence, the resulting algorithm for identity testing in* $\mathcal{M}(2m(N+2), d)$ *has running time* $T' = T(n)+O(n)$ *and sample complexity* $L(n)$*. In this case, the proof of Theorem 4 gives that there is an algorithm for the* TwoLargeCuts *problem on graphs with* N *vertices with running time* $O(T'+Nm\cdot L(n)) = O(T(n)+n\cdot L(n))$*. Hence, Theorem 4 holds for all sufficiently large* n*.*

Remark 16 *From* (13) *we see that our proof works when* $\|\mu_{G,\beta} - \mu_{G^*,\beta^*}\|_{\text{TV}} > \varepsilon$ *for any constant* $\varepsilon \in (0, 1/2)$*. With a minor modification to the proof, we can extend our result to all constant values of* $\varepsilon \in (0, 1)$*, provided* n *is sufficiently large. Specifically, if we assume that the starting graph* H *has a maximum cut of* odd *size, then it is straightforward to verify that either* $(\{s, t\}, V)$ *is the unique maximum cut of* $\hat{H}_w$ *or there is some other cut with* strictly *more edges. If this is the case, then we can use part 3 of Lemma 12 (instead of part 2) and deduce that the bound in* (13) *becomes* $\|\mu_M - \mu_{M^*}\|_{\text{TV}} > 1-\varepsilon$ *for any desired constant* $\varepsilon \in (0, 1)$*. It can be easily checked that the* TwoLargeCuts *problem restricted to graphs with odd maximum cuts is still hard; in particular, any algorithm for* TwoLargeCuts *that works for this type of input can be used to solve the* MaxCut *problem efficiently.*

5. Properties of the Ising Gadget

In this section we prove the key properties of the random bipartite graph $\mathcal{G}(m, p, d_{\text{IN}}, d_{\text{OUT}})$ that were used in Section 4 to establish our main result. In particular, we establish Theorems 10 and 11. Throughout this section we let $G = (V_G = L \cup R, E_G)$ be an instance of $\mathcal{G}(m, p, d_{\text{IN}}, d_{\text{OUT}})$ as defined in Section 4. Recall that $|L| = |R| = m$ and that there is a set P of ports such that $|P \cap L| = |P \cap R| = p$. For $S, T \subset V_G$ define

$$E(S, T) = \{\{u, v\} \in E_G : u \in S, v \in T\}.$$

In the proof of Theorems 10 and 11 we crucially use the following facts about the edge expansion of the random graph G.

Theorem 17 *Suppose* $p = m$ *and* $3 \le d_{\text{IN}} \le d \le m^{1-\rho}$ *where* $\rho \in (0, 1)$ *is a constant independent of* m*. Then, with probability* $1 - o(1)$ *over the choice of the random graph* G*:*

$$\min_{\substack{S\subset V_G:\\ 0<|S|\le m}} \frac{|E(S, V_G\backslash S)|}{|S|} \ge \frac{\rho d_{\text{IN}}}{300}.$$

Theorem 18 *Suppose* $3 \le d = O(1)$*,* $p = \lfloor m^\alpha \rfloor$ *with* $\alpha \in (0, \frac{1}{4}]$*,* $d_{\text{IN}} = d - 1$ *and* $d_{\text{OUT}} = 1$*. Then, there exists a constant* $\gamma > 0$ *independent of* m *such that with probability* $1 - o(1)$ *over the choice of the random graph* G*:*

$$\min_{\substack{S\subset V_G:\\ 0<|S|\le m}} \frac{|E(S, V_G\backslash S)|}{|S|} \ge \gamma d.$$

Theorem 19 *Suppose* $3 \le d = O(1)$*,* $p = \lfloor m^\alpha \rfloor$ *with* $\alpha \in (0, \frac{1}{4}]$*,* $d_{\text{IN}} = d - 1$ *and* $d_{\text{OUT}} = 1$*. Then, there exists a constant* $\gamma > 0$ *independent of* m *such that with probability* $1 - o(1)$ *over the choice of the random graph* G*:*

$$\min_{\substack{S\subset V_G:\\ 0<|P\cap S|\le|S|\le m}} \frac{|E(S, V_G\backslash S)|}{|P \cap S|} > 1 + \gamma.$$

Theorem 17 is proved in Section 5.1 and Theorems 18 and 19 in Section 5.2. Before that, we prove Theorems 10 and 11 which are restated below for sake of clarity.

Theorem 10 *Suppose $\beta < 0$, $3 \le d = O(1)$, $d_{\text{IN}} = d-1$, $d_{\text{OUT}} = 1$ 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:*

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

Theorem 11 *Suppose $\beta < 0$, $p = m$ and $4 + \frac{1200}{\rho} \le d \le m^{1-\rho}$ for some constant $\rho \in (0,1)$ independent of m. Then, there exist constants $\delta = \delta(\rho) > 0$ and $\theta = \theta(\rho) \in (0,1)$ such that when $d_{\text{IN}} = \lfloor \theta d \rfloor$ and $d_{\text{OUT}} = d - \lfloor \theta d \rfloor$ the following holds for every configuration τ on ∂P with probability $1-o(1)$ over the choice of the random graph G:*

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

Proof of Theorems 10 and 11 Let σ and τ be Ising configurations on V_G and ∂P respectively. Let $P^+ \subseteq \partial P$ be the set of vertices of ∂P that are assigned "+" by τ and let P^- be those that are assigned "−". Let $L^+_\sigma \subseteq L$ and $L^-_\sigma \subseteq L$ be the set of vertices of L that are assigned "+" and "−", respectively, in σ and define $R^+_\sigma, R^-_\sigma \subseteq R$ similarly. Let S_σ denote the set of smaller cardinality between $L^+_\sigma \cup R^-_\sigma$ and $L^-_\sigma \cup R^+_\sigma$; hence $S_\sigma \le m$. Suppose $|S_\sigma| > 0$; i.e., $\sigma \ne \sigma^+(G)$ and $\sigma \ne \sigma^-(G)$, where $\sigma^+(G)$ (resp., $\sigma^-(G)$) is the configuration of G in which every vertex in L is assigned "+" (resp., "−") and every vertex in R is assigned "−" (resp., "+").

For $S, T \subseteq V_G \cup \partial P$, we use $[S,T]$ for the number of edges between S and T in the graph $G' = (V_G \cup \partial P, E_G \cup E(P, \partial P))$. Observe that the weights of $\sigma^+(G)$ and $\sigma^-(G)$ in G' conditional on τ are:

$$w^+ := w^\tau_{G',\beta}(\sigma^+(G)) = \mathrm{e}^{-|\beta|([L,P^+]+[R,P^-])}, \text{ and}$$
$$w^- := w^\tau_{G',\beta}(\sigma^-(G)) = \mathrm{e}^{-|\beta|([L,P^-]+[R,P^+])}.$$

Henceforth we use $w^\tau(\cdot)$ for $w^\tau_{G',\beta}(\cdot)$.

We consider first the case when $S_\sigma = L^+_\sigma \cup R^-_\sigma$. Then,

$$\begin{aligned} w^\tau(\sigma) &= \exp\left[-|\beta|(|E(S_\sigma, V_G \setminus S_\sigma)| + [L^+_\sigma, P^+] + [L^-_\sigma, P^-] + [R^+_\sigma, P^+] + [R^-_\sigma, P^-])\right] \\ &= w^- \cdot \exp\left[-|\beta|(|E(S_\sigma, V_G \setminus S_\sigma)| + [L^+_\sigma, P^+] + [R^-_\sigma, P^-] - [L^+_\sigma, P^-] - [R^-_\sigma, P^+])\right] \\ &\le w^- \cdot \exp\left[-|\beta|(|E(S_\sigma, V_G \setminus S_\sigma)| - [L^+_\sigma, \partial P] - [R^-_\sigma, \partial P])\right] \\ &\le w^- \cdot \exp\left[-|\beta|(|E(S_\sigma, V_G \setminus S_\sigma)| - [S_\sigma, \partial P])\right]. \end{aligned} \tag{14}$$

where the first inequality follows from $[L^+_\sigma, P^-] - [L^+_\sigma, P^+] \le [L^+_\sigma, \partial P]$ and $[R^-_\sigma, P^+] - [R^-_\sigma, P^-] \le [R^-_\sigma, \partial P]$.

In Theorem 10 we assume that $3 \le d = O(1)$, $p = \lfloor m^\alpha \rfloor$ with $\alpha \in (0, \frac{1}{4}]$, $d_{\text{IN}} = d-1$ and $d_{\text{OUT}} = 1$. Hence, $[S_\sigma, \partial P] = |S_\sigma \cap P|$ and since $0 < |S_\sigma| \le m$, Theorems 18 and 19

imply that exists a constant $\gamma > 0$ such that with probability $1 - o(1)$ over the choice of the random graph G we have

$$\frac{|E(S_\sigma, V_G \setminus S_\sigma)|}{|S_\sigma|} \geq \gamma d, \text{ and}$$
$$\frac{|E(S_\sigma, V_G \setminus S_\sigma)|}{|S_\sigma \cap P|} \geq 1 + \gamma.$$

Combining these two inequalities we get for $\delta = \frac{\gamma^2}{1+\gamma}$ that

$$|E(S_\sigma, V_G \setminus S_\sigma)| \geq |S_\sigma \cap P| + \delta d |S_\sigma| = [S_\sigma, \partial P] + \delta d |S_\sigma|.$$

Plugging this bound into (14),

$$w^\tau(\sigma) \leq w^- \cdot \exp\left[-\delta|\beta| d |S_\sigma|\right]. \tag{15}$$

Under the assumptions in Theorem 11, we can also establish (15) as follows. When $m^{1-\rho} \geq d \geq d_{\text{IN}} = \lfloor \theta d \rfloor \geq 3$, Theorem 17 implies that

$$|E(S_\sigma, V_G \setminus S_\sigma)| \geq \frac{\rho d_{\text{IN}}}{300} |S_\sigma| = \frac{\rho \lfloor \theta d \rfloor}{300} |S_\sigma|.$$

Moreover,

$$[S_\sigma, \partial P] \leq d_{\text{OUT}} |S_\sigma| = (d - \lfloor \theta d \rfloor) |S_\sigma|.$$

Hence, taking

$$\theta = \frac{300 + 0.75\rho}{300 + \rho}$$

we get that when $d \geq 4 + \frac{1200}{\rho}$:

$$\frac{\rho \lfloor \theta d \rfloor}{300} - (d - \lfloor \theta d \rfloor) \geq \frac{\rho d}{600}.$$

Together with (14) this implies

$$w^\tau(\sigma) \leq w^- \cdot \exp\left[-\frac{\rho|\beta| d |S_\sigma|}{600}\right],$$

which gives (15) for $\delta \leq \rho/600$. Observe that our choice of θ guarantees $d - 1 \geq d_{\text{IN}} = \lfloor \theta d \rfloor \geq 3$ for all $d \geq 4$.

For the case when $S_\sigma = L_\sigma^- \cup R_\sigma^+$ we deduce analogously that for a suitable $\delta > 0$

$$w^\tau(\sigma) \leq w^+ \cdot \exp\left[-\delta|\beta| d |S_\sigma|\right]. \tag{16}$$

Let Ω_G be the set of Ising configurations of the graph G. By definition, the partition function $Z_{G',\beta,\tau}$ for the conditional distribution $\mu_{G',\beta}(\cdot \mid \partial P = \tau)$ satisfies:

$$Z_{G',\beta,\tau} = \sum_{\sigma \in \Omega_G} w^\tau(\sigma) \leq \sum_{\sigma:\, 0 \leq |L_\sigma^+ \cup R_\sigma^-| \leq m} w^\tau(\sigma) + \sum_{\sigma:\, 0 \leq |L_\sigma^- \cup R_\sigma^+| \leq m} w^\tau(\sigma).$$

From (15) we get

$$\sum_{\sigma\colon 0\le|L_\sigma^+\cup R_\sigma^-|\le m} w^\tau(\sigma) \le \sum_{\sigma\colon 0\le|L_\sigma^+\cup R_\sigma^-|\le m} w^-\cdot \mathrm{e}^{-\delta|\beta|d|L_\sigma^+\cup R_\sigma^-|}$$
$$= w^- \sum_{k=0}^{m} \binom{2m}{k} \mathrm{e}^{-\delta|\beta|dk} \le w^- \left(1+\mathrm{e}^{-\delta|\beta|d}\right)^{2m}.$$

Similarly, we deduce from (16)

$$\sum_{\sigma\colon 0\le|L_\sigma^-\cup R_\sigma^+|\le m} w^\tau(\sigma) \le w^+ \left(1+\mathrm{e}^{-\delta|\beta|d}\right)^{2m}.$$

Hence,

$$Z_{G',\beta,\tau} \le \left(w^- + w^+\right)\left(1+\mathrm{e}^{-\delta|\beta|d}\right)^{2m},$$

and

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

where in the last inequality we use $(\frac{1}{1+x})^{2m} \ge (1-x)^{2m} \ge 1-2mx$ for all $x\in[0,1)$. ■

5.1. Gadget Expansion when $d \le m^{1-\rho}$: Proof of Theorem 17

We derive Theorem 17 as a consequence of the following two properties of the random bipartite multigraphs obtained as the union of perfect matchings.

Lemma 20 *(Brito et al., 2018, Theorem 4). Let $\hat{G}=(L_{\hat{G}}\cup R_{\hat{G}}, E_{\hat{G}})$ be a random bipartite graph obtained as the union of d random perfect matchings between $L_{\hat{G}}$ and $R_{\hat{G}}$. Let $\lambda_2(\hat{G})$ denote the second largest eigenvalue of its adjacency matrix. Then, for $3\le d=O(1)$ and any constant $\delta>0$ (independent of m), with probability $1-o(1)$, the following holds:*

$$\lambda_2(\hat{G}) < 2\sqrt{d-1}+\delta.$$

Lemma 21 *Let $\hat{G}=(L_{\hat{G}}\cup R_{\hat{G}}, E_{\hat{G}})$ be a random bipartite graph obtained as the union of d random perfect matchings between $L_{\hat{G}}$ and $R_{\hat{G}}$. Suppose $|L_{\hat{G}}|=|R_{\hat{G}}|=m$ and that $d\le m^{1-\rho}$ for some constant $\rho\in(0,1)$ independent of m. Then, the probability that an edge between $L_{\hat{G}}$ and $R_{\hat{G}}$ is chosen by more than $\lceil 3/\rho\rceil$ random perfect matchings is $O(m^{-1})$.*

Proof Let $L_{\hat{G}}=\{v_1,\dots,v_m\}$ and $R_{\hat{G}}=\{u_1,\dots,u_m\}$. Let X_{ij} be the random variable corresponding to the number of perfect matchings that use the edge $\{v_i,u_j\}$ and let $\kappa=\lceil 3/\rho\rceil$. Then,

$$\Pr[X_{ij}\ge\kappa] = \sum_{a=\kappa}^{d}\binom{d}{a}\frac{1}{m^a}\left(1-\frac{1}{m}\right)^{d-a} \le \sum_{a=\kappa}^{d}\left(\frac{\mathrm{e}d}{am}\right)^a \le \left(\frac{\mathrm{e}}{\kappa m^\rho}\right)^\kappa \sum_{a=0}^{d-\kappa}\left(\frac{\mathrm{e}}{\kappa m^\rho}\right)^a \le O(m^{-3}).$$

The result follows by a union bound over the pairs $\{v_i,u_j\}$. ■

We are now ready to prove Theorem 17, which for convenience we restate first.

Theorem 17 *Suppose $p = m$ and $3 \le d_{\text{IN}} \le d \le m^{1-\rho}$ where $\rho \in (0,1)$ is a constant independent of m. Then, with probability $1 - o(1)$ over the choice of the random graph G:*

$$\min_{\substack{S \subset V_G: \\ 0<|S|\le m}} \frac{|E(S, V_G \backslash S)|}{|S|} \ge \frac{\rho d_{\text{IN}}}{300}.$$

Proof Let $G = (V_G = L \cup R, E_G)$ be a random bipartite graph sampled from $\mathcal{G}(m, p, d_{\text{IN}}, d_{\text{OUT}})$. For $k \ge 3$ let $F_k = (L \cup R, E(F_k))$ be a random bipartite graph obtained as the union of k random perfect matchings between L and R. For $S \subseteq V_G$ let $E_{F_k}(S, V_G \setminus S) \subseteq E(F_k)$ be set edges between S and $V_G \setminus S$ in F_k. Lemma 21 implies that for $\kappa = \lceil 3/\rho \rceil$, with probability $1 - o(1)$, we have

$$\min_{\substack{S \subset V_G: \\ 0<|S|\le m}} \frac{|E(S, V_G \backslash S)|}{|S|} \ge \min_{\substack{S \subset V_G: \\ 0<|S|\le m}} \frac{|E_{F_{d_{\text{IN}}}}(S, V_G \setminus S)|}{\kappa |S|}. \tag{17}$$

Let $d' \ge 3$ be the unique integer divisible by 3 such that $d_{\text{IN}} \ge d' \ge d_{\text{IN}} - 2$. Then,

$$\min_{\substack{S \subset V_G: \\ 0<|S|\le m}} \frac{|E_{F_{d_{\text{IN}}}}(S, V_G \setminus S)|}{\kappa |S|} \ge \min_{\substack{S \subset V_G: \\ 0<|S|\le m}} \frac{|E_{F_{d'}}(S, V_G \setminus S)|}{\kappa |S|}. \tag{18}$$

The random graph $F_{d'}$ can also be obtained as the union of $d'/3$ independent instances of the random graph F_3; let $F_3^{(1)}, \ldots, F_3^{(d'/3)}$ be these instances. For each $i \in \{1, \ldots, d'/3\}$, Cheeger's inequality and Lemma 20 (with $\delta = 0.01$) imply that with probability $r = 1 - o(1)$:

$$\min_{\substack{S \subset V_G: \\ 0<|S|\le m}} \frac{\left|E_{F_3^{(i)}}(S, V_G \setminus S)\right|}{|S|} \ge \frac{3 - \lambda_2(F_3^{(i)})}{2} \ge 0.08.$$

Let Z be number of $F_3^{(i)}$'s that satisfy this property. We have $\mathrm{E}[Z] = rd'/3$, $\mathrm{Var}(Z) = r(1-r)d'/3$ and by Chebyshev's inequality for sufficiently large m

$$\Pr\left[Z \le \frac{3d'}{10}\right] \le \Pr\left[\left|Z - \mathrm{E}[Z]\right| \ge \left(\frac{r}{3} - \frac{3}{10}\right) d'\right] \le \frac{r(1-r)}{3\left(\frac{r}{3} - \frac{3}{10}\right)^2 d'} = o(1).$$

Therefore, with probability $1 - o(1)$ we have

$$\min_{\substack{S \subset V_G: \\ 0<|S|\le m}} \frac{|E_{F_{d'}}(S, V_G \setminus S)|}{|S|} = \min_{\substack{S \subset V_G: \\ 0<|S|\le m}} \sum_{i=1}^{\frac{d'}{3}} \frac{\left|E_{F_3^{(i)}}(S, V_G \setminus S)\right|}{|S|} \ge 0.08Z \ge 0.024d'.$$

This bound, combined with (17) and (18), implies that with probability $1 - o(1)$:

$$\min_{\substack{S \subset V_G: \\ 0<|S|\le m}} \frac{|E(S, V_G \backslash S)|}{|S|} \ge \frac{0.024d'}{\kappa} \ge \frac{\rho d_{\text{IN}}}{300},$$

where in the last inequality we use $\kappa \le 3/\rho + 1 \le 4/\rho$ and $d' \ge \frac{3}{5} d_{\text{IN}}$ for $d_{\text{IN}} \ge 3$. ■

5.2. Gadget Expansion when $d = O(1)$: Proof of Theorems 18 and 19

Let m, p and d be positive integers such that $3 \leq d = O(1)$ and $p = \lfloor m^\alpha \rfloor$ for some constant $\alpha \in (0,1)$. Throughout this section we let $G = (V_G = L \cup R, E_G)$ be a random bipartite graph distributed according to $\mathcal{G}(m,p,d-1,1)$; that is, $d_{\text{IN}} = d-1$ and $d_{\text{OUT}} = 1$. The random graph G can be equivalently generated in the following two ways.

Lemma 22 *Let $m,p,d \in \mathbb{N}^+$ be positive integers such that $m \geq p$ and $d \geq 3$. Let $G' = (V_G = L \cup R, E_{G'})$ be the random bipartite graph generated as follows:*

1. *Let $M_1, M_2, \ldots, M_d$ be d random perfect matchings between L and R;*
2. *Let P_1 be a subset of L chosen uniformly at random among all the subsets of L such that $|P_1| = p$;*
3. *Let $P_2 \subset R$ be the set of vertices in R that are matched to P_1 in M_d, and let $A \subset M_d$ be the set of edges between P_1 and P_2;*
4. *Let $P = P_1 \cup P_2$ and $E_{G'} = \bigcup_{i=1}^{d} M_i \setminus A$.*

Then G' has distribution $\mathcal{G}(m,p,d-1,1)$.

Lemma 23 *Let $m,p,d \in \mathbb{N}^+$ be positive integers such that $m \geq p$ and $d \geq 3$. Let $G'' = (V_G = L \cup R, E_G)$ be the random bipartite graph generated as follows:*

1. *Let $M_1, M_2, \ldots, M_{d-1}$ be $d-1$ random perfect matchings between L and R;*
2. *Let P_1 be a subset of L chosen uniformly at random among all the subsets of L such that $|P_1| = p$;*
3. *Let M_1' be a random complete matching between $L \setminus P_1$ and R, and let $P_2 \subset R$ be the set of unmatched vertices in R; hence $|P_2| = p$.*
4. *Let $P = P_1 \cup P_2$ and $E_{G''} = \left(\bigcup_{i=1}^{d-1} M_i\right) \bigcup M_1'$.*

Then G'' has distribution $\mathcal{G}(m,p,d-1,1)$.

Lemmas 22 and 23 are both proved in Section 5.4.

The edge expansion of the random graph G satisfies the following two bounds, which we will crucially use in our proofs of Theorems 18 and 19.

Lemma 24 *For $3 \leq d = O(1)$, $\alpha \in (0,1)$ and $\delta > 0$, there exists $\varepsilon > 0$ such that with probability $1 - o(1)$:*

$$\min_{\substack{S \subset V_G: \\ 0 < |S| \leq \varepsilon m}} \frac{|E(S, V_G \setminus S)|}{|S|} > d - 2 - \alpha - \delta.$$

Lemma 25 *For $3 \leq d = O(1)$, $\alpha \in (0, \frac{1}{4}]$, $\delta > 0$ and $\xi \in (0,1)$, it holds with probability $1 - o(1)$:*

$$\min_{\substack{S \subset V_G: \\ 0 < \xi|S| \leq |P \cap S|}} \frac{|E(S, V_G \setminus S)|}{|S|} > \xi(d - \alpha - \delta) - 1.$$

We are now ready to prove Theorems 18 and 19; in both cases, we restate the corresponding theorem first.

Theorem 18 *Suppose* $3 \leq d = O(1)$, $p = \lfloor m^{\alpha} \rfloor$ *with* $\alpha \in (0, \frac{1}{4}]$, $d_{\text{IN}} = d - 1$ *and* $d_{\text{OUT}} = 1$. *Then, there exists a constant* $\gamma > 0$ *independent of* m *such that with probability* $1 - o(1)$ *over the choice of the random graph* G*:*

$$\min_{\substack{S \subset V_G: \\ 0 < |S| \leq m}} \frac{|E(S, V_G \backslash S)|}{|S|} \geq \gamma d.$$

Proof By Lemma 24, for $\alpha \in (0, \frac{1}{4}]$ and $\delta \leq \frac{1}{4}$ there exists $\varepsilon > 0$ such that with probability $1 - o(1)$:

$$\min_{\substack{S \subset V_G: \\ 0 < |S| \leq \varepsilon m}} \frac{|E(S, V_G \backslash S)|}{|S|} > d - \frac{5}{2} \geq \frac{d}{6}.$$

For $S \subset V_G$ with $\varepsilon m < |S| \leq m$, we consider the random bipartite graph $\hat{G}$ obtained as the union of d random perfect matchings $M_1, \ldots, M_d$ between L and R. Let $\lambda_2(\hat{G})$ denote the second largest eigenvalue of the adjacency matrix of $\hat{G}$. Lemma 20 implies that for any constant $\delta > 0$ (independent of m), with probability $1 - o(1)$, the following holds:

$$\lambda_2(\hat{G}) < 2\sqrt{d-1} + \delta.$$

Now, let $\hat{G} = (V_G, E_{\hat{G}})$ and let $\hat{E}(S, V_G \setminus S)$ be the set of edges between S and $V_G \setminus S$ in $\hat{G}$. Cheeger's inequality implies

$$\min_{\substack{S \subset V_G: \\ 0 < |S| \leq m}} \frac{|\hat{E}(S, V_G \backslash S)|}{|S|} \geq \frac{d - \lambda_2(\hat{G})}{2} > \frac{d}{2} - \sqrt{d-1} - \frac{\delta}{2} \geq \frac{d}{40},$$

where the last inequality holds for $\delta \leq 0.01$.

We use this bound on the edge expansion of $\hat{G}$ to deduce a bound for the edge expansion of G. First note that by Lemma 22, G can be obtained from $\hat{G}$ as follows:

1. Choose $P_1 \subset L$ uniformly at random among all the subsets of L of size p;
2. Let $P_2 \subset R$ be the set of vertices matched to P_1 in M_d, and let $A \subset M_d$ be the set of edges between P_1 and P_2;
3. Set $P = P_1 \cup P_2$ and $E_G = \bigcup_{i=1}^{d} M_i \setminus A$;
4. Replace all the multiedges in E_G by single edges.

Moreover, since $3 \leq d = O(1)$, Lemma 21 implies there exists a constant κ such that with probability $1 - O(m^{-1})$ the multiplicity of every edge in $\hat{G}$ is at most κ. Hence, in order to obtain G from $\hat{G}$, p edges are removed and the multiplicity of every edge may decrease by a factor of at most κ. Therefore, for every $S \subset V_G$

$$E(S, V_G \setminus S) \geq \frac{\hat{E}(S, V_G \setminus S) - p}{\kappa},$$

and

$$\min_{\substack{S\subset V_G:\\ \varepsilon m<|S|\le m}} \frac{|E(S,V_G\backslash S)|}{|S|} \ge \min_{\substack{S\subset V_G:\\ \varepsilon m<|S|\le m}} \frac{|\hat{E}(S,V_G\setminus S)|-p}{\kappa|S|} \ge \frac{d}{40\kappa} - \frac{1}{\varepsilon\kappa m^{3/4}},$$

and the result is established. ■

Theorem 19 *Suppose $3 \le d = O(1)$, $p = \lfloor m^\alpha \rfloor$ with $\alpha \in (0, \frac{1}{4}]$, $d_{\text{IN}} = d-1$ and $d_{\text{OUT}} = 1$. Then, there exists a constant $\gamma > 0$ independent of m such that with probability $1-o(1)$ over the choice of the random graph G:*

$$\min_{\substack{S\subset V_G:\\ 0<|P\cap S|\le|S|\le m}} \frac{|E(S,V_G\backslash S)|}{|P\cap S|} > 1+\gamma.$$

Proof Let $S \subset V_G$ such that $0 < |S| \le m$. Let $\hat{d} = d-\alpha-\delta$, $\delta = 0.01$ and let $\varepsilon > 0$ be the constant from Lemma 24. We consider three cases, depending on the sizes of S and $S \cap P$. Suppose first that $|S| > \varepsilon m$. Then, Theorem 18 implies that with probability $1-o(1)$, there exists $\gamma' > 0$ such that

$$|E(S,V\backslash S)| > \gamma'|S| > \gamma'\varepsilon m \ge (1+\gamma)m^\alpha > (1+\gamma)|P\cap S|,$$

where the second to last inequality holds for sufficiently large m and a suitable constant $\gamma > 0$.

For the second case, suppose that $|S| \le \varepsilon m$ and $|P \cap S| < \xi|S|$, where

$$\xi = \frac{\sqrt{4\hat{d}(\hat{d}-2)+1}+1}{2\hat{d}}.$$

Then, by Lemma 24, with probability $1-o(1)$:

$$|E(S,V\backslash S)| > (\hat{d}-2)|S| > \frac{\hat{d}-2}{\xi}|P\cap S| = \frac{\sqrt{4\hat{d}(\hat{d}-2)+1}-1}{2}|P\cap S| \ge (1+\gamma)|P\cap S|,$$

where the last inequality holds for sufficiently small γ since $\hat{d} \ge 2.74$.

Finally, suppose $|S| \le \varepsilon m$ and $|P \cap S| \ge \xi|S|$. In this case, Lemma 25 also implies that with probability $1-o(1)$:

$$|E(S,V\backslash S)| > \left(\xi\hat{d}-1\right)|S| = \frac{\sqrt{4\hat{d}(\hat{d}-2)+1}-1}{2}|P\cap S| \ge (1+\gamma)|P\cap S|,$$

and the result follows. ■

5.3. Gadget Expansion when $d = O(1)$: Proof of Auxiliary Lemmas 24 and 25

In this section, we give the proof of our two key bounds on the edge expansion of the random graph $\mathcal{G}(m, p, d-1, 1)$. In particular, we prove Lemmas 24 and 25. Suppose $3 \leq d = O(1)$, $p = \lfloor m^{\alpha} \rfloor$ for some constant $\alpha \in (0,1)$, and let $G = (V_G = L \cup R, E_G)$ be a random bipartite graph distributed according to $\mathcal{G}(m, p, d-1, 1)$. Our proofs of Lemmas 24 and 25 will be based on the following bound for the *vertex expansion* of small subsets of L (or R). Its proof has similar flavor to that of Theorem 4.16 in (Hoory et al., 2006) for random regular bipartite graphs. Recall that the vertex expansion is defined as:

$$\partial S = \{v \in V_G \backslash S : \exists u \in S, \{u, v\} \in E_G\}.$$

Lemma 26 *For $3 \leq d = O(1)$, $\alpha \in (0,1)$ and $\delta > 0$, there exists $\varepsilon > 0$ such that with probability $1 - o(1)$ the following holds:*

$$\min_{\substack{S \subset L: \\ 0 < |S| \leq \varepsilon m}} \frac{|\partial S|}{|S|} > d - 1 - \alpha - \delta.$$

Proof Let

$$\eta = d - 1 - \alpha - \delta.$$

For $S \subset L$ and $T \subset R$, let $X_{S,T}$ be an indicator random variable for the event $\partial S \subseteq T$. For some $\varepsilon \in (0, 1/\eta)$ to be chosen later, let

$$X = \sum_{\substack{S \subset L: \\ 0 < |S| \leq \varepsilon m}} \sum_{\substack{T \subset R: \\ |T| = \lfloor \eta |S| \rfloor}} X_{S,T}.$$

Since every set $T \subset R$ of size less than $\lfloor \eta |S| \rfloor$ is included in some subset of R of size exactly $\lfloor \eta |S| \rfloor$, it suffices to show that

$$\Pr[X > 0] \leq O\left(\frac{(\ln m)^{\delta}}{m^{\delta}}\right).$$

Write $|S| = s$, $|T| = t$ and $|P \cap S| = r$. Then, Lemma 23 implies

$$\Pr\left[X_{S,T} = 1\right] = \left(\frac{t(t-1)\dots(t-s+1)}{m(m-1)\dots(m-s+1)}\right)^{d-1} \left(\frac{t(t-1)\dots(t-(s-r)+1)}{m(m-1)\dots(m-(s-r)+1)}\right) \leq \left(\frac{t}{m}\right)^{ds-r}.$$

Using Markov's inequality, we get

$$\Pr\left[X > 0\right] = \Pr\left[X \geq 1\right] \leq \mathrm{E}[X],$$

and so

$$\begin{aligned} \Pr\left[X > 0\right] &\leq \sum_{\substack{S \subset L: \\ 0 < |S| \leq \varepsilon m}} \sum_{\substack{T \subset R: \\ |T| = \lfloor \eta |S| \rfloor}} \mathrm{E}[X_{S,T}] \\ &\leq \sum_{s=1}^{\lfloor \varepsilon m \rfloor} \sum_{r=0}^{s} \binom{p}{r} \binom{m-p}{s-r} \binom{m}{t} \left(\frac{t}{m}\right)^{ds-r}. \end{aligned}$$

From the inequality $\binom{m}{k} \le \left(\frac{\mathrm{e}m}{k}\right)^k$, we get

$$\begin{aligned}\Pr[X>0] &\le \sum_{s=1}^{\lfloor \varepsilon m\rfloor}\sum_{r=0}^{s}\left(\frac{\mathrm{e}p}{r}\right)^r\left(\frac{\mathrm{e}(m-p)}{s-r}\right)^{s-r}\left(\frac{\mathrm{e}m}{t}\right)^t\left(\frac{t}{m}\right)^{ds-r}\\ &= \sum_{s=1}^{\lfloor \varepsilon m\rfloor}\sum_{r=0}^{s}\left[(\mathrm{e}p)^r(\mathrm{e}(m-p))^{s-r}\right]\cdot\left[\left(\frac{1}{r}\right)^r\left(\frac{1}{s-r}\right)^{s-r}\right]\cdot\left[\left(\frac{\mathrm{e}m}{t}\right)^t\left(\frac{t}{m}\right)^{ds-r}\right].\end{aligned}$$

Since $p \le m^\alpha$, the first term is be bounded by

$$(\mathrm{e}p)^r(\mathrm{e}(m-p))^{s-r} \le \mathrm{e}^s p^r m^{s-r} = \left(\mathrm{e}m^{1-\frac{(1-\alpha)r}{s}}\right)^s. \tag{19}$$

Also, the AM-GM inequality yields

$$\left(\frac{1}{r}\right)^r\left(\frac{1}{s-r}\right)^{s-r} \le \left(\frac{r\cdot\frac{1}{r}+(s-r)\cdot\frac{1}{s-r}}{s}\right)^s = \left(\frac{2}{s}\right)^s. \tag{20}$$

Finally, since for $\varepsilon \in (0,1/\eta)$, $t \le \eta s \le \eta\varepsilon m < m$ we have

$$\left(\frac{\mathrm{e}m}{t}\right)^t\left(\frac{t}{m}\right)^{ds-r} \le \left(\frac{\mathrm{e}m}{t}\right)^{\eta s}\left(\frac{t}{m}\right)^{ds-r} = \left[\mathrm{e}^\eta\left(\frac{t}{m}\right)^{d-\frac{r}{s}-\eta}\right]^s \le \left[\mathrm{e}^\eta\left(\frac{\eta s}{m}\right)^{d-\frac{r}{s}-\eta}\right]^s. \tag{21}$$

Combining the inequalities (19), (20) and (21) we deduce

$$\begin{aligned}\Pr[X>0] &\le \sum_{s=1}^{\lfloor \varepsilon m\rfloor}\sum_{r=0}^{s}\left[\mathrm{e}m^{1-(1-\alpha)\frac{r}{s}}\cdot\frac{2}{s}\cdot\mathrm{e}^\eta\left(\frac{\eta s}{m}\right)^{d-\frac{r}{s}-\eta}\right]^s\\ &= \sum_{s=1}^{\lfloor \varepsilon m\rfloor}\sum_{r=0}^{s}\left[2\mathrm{e}^{\eta+1}\eta^{d-\frac{r}{s}-\eta}\cdot\left(\frac{s}{m}\right)^{d-1-\alpha\frac{r}{s}-\eta}\cdot s^{-(1-\alpha)\frac{r}{s}}\right]^s.\end{aligned}$$

We recall that $\eta = d-1-\alpha-\delta$, so for $0 \le r \le s$ we have

$$\eta^{d-\frac{r}{s}-\eta} \le (d-1)^{d-\eta}, \qquad \left(\frac{s}{m}\right)^{d-1-\alpha\frac{r}{s}-\eta} \le \left(\frac{s}{m}\right)^{\delta} \quad \text{and} \quad s^{-(1-\alpha)\frac{r}{s}} \le 1.$$

Thus, taking $c = c(d,\alpha,\delta) = 2\mathrm{e}^{\eta+2}(d-1)^{d-\eta}$ we obtain

$$\begin{aligned}\Pr[X>0] &\le \sum_{s=1}^{\lfloor \varepsilon m\rfloor}\sum_{r=0}^{s}\left[2\mathrm{e}^{\eta+1}(d-1)^{d-\eta}\cdot\left(\frac{s}{m}\right)^{\delta}\right]^s\\ &\le \sum_{s=1}^{\lfloor \varepsilon m\rfloor}(s+1)\left[\frac{c}{\mathrm{e}}\cdot\left(\frac{s}{m}\right)^{\delta}\right]^s\\ &\le \sum_{s=1}^{\lfloor \varepsilon m\rfloor}\left[c\left(\frac{s}{m}\right)^{\delta}\right]^s,\end{aligned}$$

where in the last inequality we use the fact that $(s+1)^{1/s} \leq \mathrm{e}$ for all $s > 0$.

Now given $\delta > 0$, since $c \leq 2d^{2+\delta}\mathrm{e}^{d+1}$, we can choose $\varepsilon \in (0, 1/\eta)$ such that $c\varepsilon^{\delta} < \mathrm{e}^{-1}$. For this choice of ε we have

$$\sum_{s=\lfloor \delta \ln m \rfloor + 1}^{\lfloor \varepsilon m \rfloor} \left[c \left(\frac{s}{m} \right)^{\delta} \right]^s \leq \sum_{s=\lfloor \delta \ln m \rfloor + 1}^{\lfloor \varepsilon m \rfloor} \left[c\varepsilon^{\delta} \right]^s \leq \sum_{s=\lfloor \delta \ln m \rfloor + 1}^{\lfloor \varepsilon m \rfloor} \frac{1}{\mathrm{e}^s} = O\left(\frac{1}{m^{\delta}} \right)$$

and

$$\sum_{s=1}^{\lfloor \delta \ln m \rfloor} \left[c \left(\frac{s}{m} \right)^{\delta} \right]^s \leq \sum_{s=1}^{\lfloor \delta \ln m \rfloor} \left[c \left(\frac{\delta \ln m}{m} \right)^{\delta} \right]^s = O\left(\frac{(\ln m)^{\delta}}{m^{\delta}} \right).$$

Hence,

$$\Pr\left[X > 0 \right] \leq O\left(\frac{(\ln m)^{\delta}}{m^{\delta}} \right)$$

and the lemma follows. ■

We are now ready to prove Lemma 24, which as usual we restate first.

Lemma 24 *For $3 \leq d = O(1)$, $\alpha \in (0,1)$ and $\delta > 0$, there exists $\varepsilon > 0$ such that with probability $1 - o(1)$:*

$$\min_{\substack{S \subset V_G: \\ 0 < |S| \leq \varepsilon m}} \frac{|E(S, V_G \backslash S)|}{|S|} > d - 2 - \alpha - \delta.$$

Proof By Lemma 26, for $3 \leq d = O(1)$, $\alpha \in (0,1)$ and $\delta > 0$ there exists $\varepsilon > 0$ such that for all $T \subset V_G$ with $0 < |T| \leq \varepsilon m$ and either $T \subset L$ or $T \subset R$, with probability $1 - o(1)$ we have

$$|\partial T| > (d - 1 - \alpha - \delta)|T|.$$

Suppose this holds for every such T. Let $S \subset V_G$ such that $0 < |S| \leq \varepsilon m$ and let $S_L = S \cap L$ and $S_R = S \cap R$. Then, $S_L \subset L$, $S_R \subset R$ and $\max\{|S_L|, |S_R|\} \leq |S| \leq \varepsilon m$. Hence,

$$\begin{aligned}
|E(S, V \backslash S)| &\geq |\partial S| \\
&= |\partial S_L \backslash S_R| + |\partial S_R \backslash S_L| \\
&\geq |\partial S_L| - |S_R| + |\partial S_R| - |S_L| \\
&> (d - 1 - \alpha - \delta)|S_L| - |S_R| + (d - 1 - \alpha - \delta)|S_R| - |S_L| \\
&= (d - 2 - \alpha - \delta)|S|.
\end{aligned}$$

Therefore, this holds for all such S with probability $1 - o(1)$ and the result follows. ■

To prove Lemma 25 we also need the following fact.

Fact 27 *For $3 \leq d = O(1)$ and $\alpha \in (0, \frac{1}{4}]$, the event $E(L \cap P, R \cap P) = \emptyset$ occurs with probability $1 - O(m^{-1/2})$.*

Proof By definition, no edges are added between $L \cap P$ and $R \cap P$ in M_1'. Thus, it is sufficient to show that $E(L \cap P, R \cap P) = \emptyset$ after adding the $d-1$ random perfect matchings $M_1, \ldots, M_{d-1}$. We may generate a random matching by choosing for each vertex of L, in any order, a uniformly random unmatched vertex in R. Say $M_1, \ldots, M_{d-1}$ are generated in this manner always matching the vertices in $L \cap P$ first. Then, the probability of the event $E(L \cap P, R \cap P) = \emptyset$ is:

$$\left[\prod_{i=0}^{p-1}\left(1 - \frac{p}{m-i}\right)\right]^{d-1} \geq \left(1 - \frac{p}{m-p+1}\right)^{p(d-1)} \geq 1 - O(m^{-1/2}). \qquad \blacksquare$$

We now have all the ingredients for the proof of Lemma 25.

Lemma 25 *For $3 \leq d = O(1)$, $\alpha \in (0, \frac{1}{4}]$, $\delta > 0$ and $\xi \in (0,1)$, it holds with probability $1 - o(1)$:*

$$\min_{\substack{S \subset V_G: \\ 0 < \xi|S| \leq |P \cap S|}} \frac{|E(S, V_G \backslash S)|}{|S|} > \xi(d - \alpha - \delta) - 1.$$

Proof By Lemma 26, for $3 \leq d = O(1)$ and $\delta > 0$ there exists $\varepsilon > 0$ such that, with probability $1 - o(1)$, for all $S \subset V$ with $0 < |S| \leq \varepsilon m$ and either $S \subset L$ or $S \subset R$,

$$|\partial S| > (d - 1 - \alpha - \delta)\,|S|.$$

Also, by Fact 27, P is an independent set with probability $1 - O(m^{-1/2})$. Hence, by a union bound, both of these events occur with probability $1 - o(1)$. Suppose this is the case.

For $\xi \in (0,1)$, let $S \subset V$ such that $0 < \xi|S| \leq |P \cap S|$. Then, for sufficiently large m

$$|S| \leq \frac{|P \cap S|}{\xi} \leq \frac{2p}{\xi} \leq \varepsilon m.$$

Let $S_L = S \cap L$ and $S_R = S \cap R$. Since there is no edge between any pair of vertices in P, we have

$$\partial S_L \backslash S_R \supset \partial(P \cap S_L) \backslash S_R = \partial(P \cap S_L) \backslash (S_R \backslash P),$$

and similarly $\partial S_R \backslash S_L \supset \partial(P \cap S_R) \backslash (S_L \backslash P)$. Moreover, $S_L \subset L$, $S_R \subset R$ and $\max\{|S_L|, |S_R|\} \leq |S| \leq \varepsilon m$. It follows that

$$\begin{aligned}
|E(S, V \backslash S)| &\geq |\partial S| \\
&= |\partial S_L \backslash S_R| + |\partial S_R \backslash S_L| \\
&\geq |\partial(P \cap S_L) \backslash (S_R \backslash P)| + |\partial(P \cap S_R) \backslash (S_L \backslash P)| \\
&\geq |\partial(P \cap S_L)| - |S_R \backslash P| + |\partial(P \cap S_R)| - |S_L \backslash P| \\
&> (d - 1 - \alpha - \delta)|P \cap S_L| + (d - 1 - \alpha - \delta)|P \cap S_R| - |S \backslash P| \\
&= (d - \alpha - \delta)|P \cap S| - |S| \\
&\geq (\xi(d - \alpha - \delta) - 1)\,|S|.
\end{aligned}$$

Thus, this holds for all $S \subset V$ such that $0 < \xi|S| \leq |P \cap S|$ with probability $1 - o(1)$ as claimed. $\blacksquare$

5.4. Ising Gadget: Equivalent Generation

We previously stated, in Lemmas 22 and 23, two alternative procedures to generate a random graph G with distribution $\mathcal{G}(m, p, d-1, 1)$. We conclude this section with a proof of these facts.

Proof of Lemmas 22 and 23 Denote the random bipartite graph defined in Lemma 22 by G' and the one in Lemma 23 by G''. We need to show that both G' and G'' have distribution $\mathcal{G}(m, p, d-1, 1)$. Recall that $P_1 = P \cap L$, $P_2 = P \cap R$ and M_1' is the perfect matching between $L \backslash P_1$ and $R \backslash P_2$. By the definitions of $\mathcal{G}(m, p, d-1, 1)$, G' and G'', it suffices to show that the joint distributions of (P_2, M_1') in these three models are the same. We recall that:

- In $\mathcal{G}(m, p, d-1, 1)$, the joint distribution ρ of (P_2, M_1') is:
  1. P_2 is a subset of R chosen uniformly at random among all the subsets of R such that $|P_2| = p$;
  2. M_1' is a random perfect matching between $L \backslash P_1$ and $R \backslash P_2$.
- In G', the joint distribution ρ' of (P_2, M_1') is:
  1. M_d is a random perfect matching between L and R;
  2. $P_2 \subset R$ is the set of vertices in R that are matched to P_1;
  3. $A \subset M_d$ is the set of edges between P_1 and P_2, and let $M_1' = M_d \backslash A$.
- In G'', the joint distribution ρ'' of (P_2, M_1') is:
  1. M_1' is a random complete matching between $L \backslash P_1$ and R;
  2. $P_2 \subset R$ is the set of unmatched vertices in R.

We first show that $\rho'' = \rho$. In ρ'', the set P_2 of unmatched vertices in R is a uniformly random subset of R over all subsets such that $|P_2| = p$. Also, given $P_2 \subset R$, a random complete matching between $L \backslash P_1$ and R conditioned on that vertices in P_2 are unmatched is a random perfect matching between $L \backslash P_1$ and $R \backslash P_2$. This gives $\rho'' = \rho$. To see that $\rho' = \rho''$, we observe that a random complete matching between $L \backslash P_1$ and R can be obtained by first drawing a random perfect matching between L and R and then removing all edges incident to P_1. ■

6. Identity Testing Algorithm for the Ferromagnetic Ising Model

Let $G = (V, E) \in \mathcal{M}(n, d)$ be an n-vertex graph of maximum degree at most d. In this section, we focus on the ferromagnetic (attractive) setting. We will allow each edge to have distinct but *positive* interaction parameter which may depend on n. In setting, $\beta = \{\beta(v, w)\}_{\{v,w\} \in E}$ with $\beta(v, w) > 0$, and the Gibbs distribution becomes

$$\mu_{G,\beta}(\sigma) = \frac{1}{Z_{G,\beta}} \exp\left(\sum_{\{v,w\} \in E} \beta(v, w) \mathbb{1}\{\sigma(v) = \sigma(w)\} \right),$$

for every $\sigma \in \{+,-\}^V$; cf., (1). With slight abuse of notation, we also use β for the largest $\beta(v,w)$; i.e., $\beta = \max_{\{v,w\}\in E} \beta(v,w)$. We remark that the Ising model is also well-defined when G is a multigraph. Indeed, an Ising model on a multigraph can be transformed into an equivalent model on a simple graph by collapsing all parallel edges and setting $\beta(v,w)$ to be the sum of the weights of all the edges between $v, w \in V$. We restrict attention again in this section to the simpler case where there is no external magnetic field. This simplification is not actually necessary and is done only for the sake of clarity in our proofs; for a discussion about how our algorithmic results extend to models with external field see Remark 31.

In (Daskalakis et al., 2018), the authors give an algorithm for identity testing for $\mathcal{M}(n,d)$ (see Algorithm 2 in (Daskalakis et al., 2018)). We call this algorithm the *DDK algorithm*. As discussed in the introduction, the running time and sample complexity of this algorithm for the ferromagnetic Ising model, where we can sample in polynomial time, is $\mathrm{poly}(n, d, \beta, \varepsilon^{-1})$. We provide here an algorithm whose running time and sample complexity is polynomial in n, d and ε^{-1} but independent of β.

Our algorithm will use as a subroutine the DDK algorithm for multigraphs, which extends straightforwardly to this more general setting. We will also use the fact that we can generate samples from the ferromagnetic Ising distribution in polynomial time (see Jerrum and Sinclair, 1993; Randall and Wilson, 1999; Guo and Jerrum, 2017; Collevecchio et al., 2016) for various methods. These two facts are rigorously stated in the following theorems. For positive integers n and m, let $\mathcal{M}_{\mathrm{multi}}(n,m)$ denote the family of all n-vertex multigraphs with at most m edges.

Theorem 28 *Let $H \in \mathcal{M}_{\mathrm{multi}}(n,m)$ where n, m are positive integers. Then, for all $\beta, \delta > 0$, there exists an algorithm that generates a sample from a distribution μ_{ALG} satisfying:*

$$\|\mu_{H,\beta} - \mu_{\mathrm{ALG}}\|_{\mathrm{TV}} \le \delta,$$

with running time $\mathrm{poly}(m, \log(1/\delta))$.

Theorem 29 *(Daskalakis et al., 2018). The DDK algorithm for the identity testing problem in $\mathcal{M}_{\mathrm{multi}}(n,m)$ has sample complexity $O(m^2\beta^2\varepsilon^{-2}\log n)$, running time* $\mathrm{poly}(m, \beta, \varepsilon^{-1})$ *and success probability at least* $4/5$.

Before presenting our algorithm, we first introduce some necessary notations and definitions. For a set of vertices $A \subset V$, let $\binom{A}{2}$ denote the collection of all pairs of vertices in A; i.e.,

$$\binom{A}{2} = \{\{v,w\} : v, w \in A, v \neq w\}.$$

Suppose $P = \{C_1, \ldots, C_k\}$ is a partition of V; that is, $\cup_{i=1}^k C_i = V$ and $C_i \cap C_j = \emptyset$ for $1 \le i < j \le k$. Let

$$E(P) = \bigcup_{i=1}^{k} \binom{C_i}{2}.$$

The *quotient graph* $G_{\mathrm{P}} = (V(G_{\mathrm{P}}), E(G_{\mathrm{P}}))$ is a multigraph defined as follows:

1. Every vertex of G_{P} is a partition class C_i from P where $1 \le i \le k$; i.e., $V(G_{\mathrm{P}}) = P$.

2. Every edge between C_i and C_j of G_{P} corresponds to an edge $\{v, w\}$ of G where $v \in C_i$ and $w \in C_j$, allowing parallel edges. The number of edges between C_i and C_j in G_{P} is equal to the size of the set $\{\{v, w\} \in E : v \in C_i,\, w \in C_j\}$.

Observe also that there is a one-to-one correspondence between the edge sets $E(G_{\text{P}})$ and $E \setminus E(P)$, which we represent by the bijective map $\varphi : E(G_{\text{P}}) \to E \setminus E(P)$. Using φ, we can define an Ising model $(G_{\text{P}}, \beta_{\text{P}})$ on the quotient graph G_{P}; the parameter β_{P} is given by

$$\beta_{\text{P}}(e) = \beta(\varphi(e))$$

for every $e \in E(G_{\text{P}})$. We call $(G_{\text{P}}, \beta_{\text{P}})$ the *quotient model*.

Suppose $\tau \in \{+, -\}^V$ is an Ising configuration of the original model (G, β) satisfying $\tau(v) = \tau(w)$ for all $\{v, w\} \in E(P)$; that is, every vertex in the same partition class C_i shares the same spin. Then, we can define a corresponding configuration $\tau_{\text{P}} \in \{+, -\}^P$ of the quotient model $(G_{\text{P}}, \beta_{\text{P}})$ as follows. For each $C_i \in P$ and $v \in C_i$

$$\tau_{\text{P}}(C_i) = \tau(v);$$

our assumption on τ guarantees that the configuration τ_{P} is well-defined.

Recall that in the identity testing problem for $\mathcal{M}(n, d)$, we are given a graph $G = (V, E) \in \mathcal{M}(n, d)$, the parameter β and sample access to an unknown Ising distribution on a graph $G^* \in \mathcal{M}(n, d)$. We will reduce this problem to identity testing for the corresponding quotient models.

For ease of notation, we set $\mu = \mu_{G,\beta}$, $\mu_{\text{P}} = \mu_{G_{\text{P}},\beta_{\text{P}}}$, $\mu^* = \mu_{G^*,\beta^*}$ and $\mu^*_{\text{P}} = \mu_{G^*_{\text{P}},\beta^*_{\text{P}}}$. For a partition P of V, define $\mathcal{P}$ to be the event that vertices from the same partition class of P receive the same spin; i.e.,

$$\mathcal{P} = \{X_v = X_w,\, \forall \{v, w\} \in E(P)\},$$

where recall that $X_v, X_w \{+1, -1\}$ are the random variables for the spins at vertices v and w respectively. Interchangeably, we also use $\mathcal{P}$ for the set

$$\mathcal{P} = \left\{\sigma \in \{+, -\}^V : \sigma(v) = \sigma(w),\, \forall \{v, w\} \in E(P)\right\}.$$

We remark that the conditional distributions $\mu(\cdot|\mathcal{P})$ and $\mu^*(\cdot|\mathcal{P})$ are equivalent to the Gibbs distributions μ_{P} and μ^*_{P}, respectively, for the quotient models.

Given L samples $\{\sigma_1, \ldots, \sigma_L\}$ from μ, we define μ_{EMP} to be the empirical distribution of these samples; in particular,

$$\mu_{\text{EMP}}(\mathcal{P}) = \frac{1}{L} \sum_{i=1}^{L} \mathbb{1}\{\sigma_i \in \mathcal{P}\}$$

Observe that $\mu_{\text{EMP}}(\mathcal{P}) = 1$ if and only if in all the L samples every vertex from the same partition class of P has the same spin. Similarly, given L samples $\{\tau_1, \ldots, \tau_L\}$ from μ^*, we also define the empirical distribution μ^*_{EMP} and the empirical probability $\mu^*_{\text{EMP}}(\mathcal{P})$.

Suppose we are given a known Ising model (G, β), L samples $\{\tau_1, \ldots, \tau_L\}$ from an unknown Ising model (G^*, β^*) and a parameter $\varepsilon > 0$, where $G, G^* \in \mathcal{M}(n, d)$ and $\beta, \beta^* > 0$.

Algorithm 1: Identity testing for ferromagnetic Ising models

input : An Ising model (G, β), L samples $\{\tau_i\}_{i=1}^L$ from an unknown Ising model (G^*, β^*) and a parameter $\varepsilon > 0$.

output: Yes if it regards $\{\tau_i\}_{i=1}^L$ as samples from $\mu_{G,\beta}$;
No if it regards $\{\tau_i\}_{i=1}^L$ as samples from μ_{G^*,β^*} such that $\|\mu_{G,\beta} - \mu_{G^*,\beta^*}\|_{\text{TV}} > \varepsilon$.

1 $P = \{V\}$;
2 **for** $i \leftarrow 1$ **to** L **do**
3 **foreach** $C \in P$ **do**
4 $C_+ = \{v \in C : \tau_i(v) = +\}$;
5 $C_- = \{v \in C : \tau_i(v) = -\}$;
6 $P \leftarrow P \backslash \{C\} \cup \{C_+, C_-\}$;
7 Generate L independent $(\varepsilon/16)$-approximate samples $\{\sigma_i\}_{i=1}^L$ from $\mu_{G,\beta}$;
8 **if** $\mu_{\text{EMP}}(P) \leq 1 - \varepsilon/4$ **then**
9 return No;
10 **foreach** $\{v, w\} \in E \backslash E(P)$ **do**
11 **if** $\beta(v, w) \geq \ln(20n^2L)$ **then**
12 return No;
13 Run the DDK algorithm on the quotient model $(G_{\text{P}}, \beta_{\text{P}})$, with samples $\{(\tau_i)_{\text{P}}\}_{i=1}^L$ and parameter $\varepsilon' = \varepsilon/2$.
14 Return the output of the DDK algorithm.

Our algorithm (see Algorithm 1) tests whether $(G, \beta) = (G^*, \beta^*)$ or $\|\mu_{G,\beta} - \mu_{G^*,\beta^*}\|_{\text{TV}} > \varepsilon$. For this, it first identifies "heavy" edges. These are the edges $\{v, w\}$ whose interaction parameter $\beta(v, w)$ is very large, and thus its endpoints v and w are likely to have the same spin. To identify the heavy edges, the algorithm looks for the coarsest partition P of V such that for every partition class C of P, all vertices from C have the same spin in each of the L samples $\{\tau_1, \ldots, \tau_L\}$; i.e., $\mu^*_{\text{EMP}}(\mathcal{P}) = 1$. The algorithm will regard the edges in $E(P)$ as the heavy ones.

The algorithm then generates L samples from the given Ising model (G, β) and computes $\mu_{\text{EMP}}(P)$ from these samples. If $\mu_{\text{EMP}}(P)$ is small, then it means that there is substantial disagreement between the sets of heavy edges in the known and hidden models, so the algorithm will output No. Otherwise, the algorithm has found a partition P such that vertices from the same partition class are very likely to receive the same spin in both (G, β) and (G^*, β^*). The algorithm outputs No if any heavy edge of G is not included in $E(P)$.

As a result, after Steps 10-12, the weights of the edges in $E(G_{\text{P}})$ are guaranteed to be $O(\log n)$ and those in $E(G^*_{\text{P}})$ are also $O(\log n)$ with high probability. Consequently, the original identity testing problem for (G, β) and (G^*, β^*) reduces to the same problem for the quotient models $(G_{\text{P}}, \beta_{\text{P}})$ and $(G^*_{\text{P}}, \beta^*_{\text{P}})$ where $\beta_{\text{P}}, \beta^*_{\text{P}} = O(\log n)$. Hence, when we run the DDK algorithm on the quotient models, the dependence on β can be replaced by a

poly($\log n$) term. The precise sample complexity of Algorithm 1 is given in the following theorem.

Theorem 30 *Suppose $G, G^* \in \mathcal{M}(n,d)$ and $\beta, \beta^* > 0$. For all sufficiently large n, Algorithm 1 outputs the correct answer for the identity testing problem with probability at least $3/4$ and has sample complexity $L = O(n^2 d^2 \varepsilon^{-2} \log^3 n)$. Moreover, the running time of Algorithm 1 is* $\mathrm{poly}(n, d, \varepsilon^{-1})$.

Remark 31 *Our guarantees for Algorithm 1 in Theorem 30 extend without significant modification to the case where the Gibbs distribution $\mu_{G,\beta}$ includes a* consistent *magnetic field or vertex potential. (Recall that a magnetic field is consistent if it has the same sign in every vertex.) We believe that identity testing for the ferromagnetic Ising model with inconsistent fields is actually hard since sampling is already known to be* #BIS-HARD *in this setting (Goldberg and Jerrum, 2007). Observe that the DDK algorithm is not guaranteed to run in polynomial time with inconsistent fields since in this case we do not know how to compute the pairwise covariances efficiently.*

Before proving Theorem 30, we first show that, with high probability, the empirical distributions μ_{EMP}, μ^*_{EMP} are close to the corresponding Gibbs distributions μ, μ^*. Define $\mathcal{F}$ to be the event that the following two events occur:

1. For every partition P of V, $|\mu_{\text{EMP}}(\mathcal{P}) - \mu(\mathcal{P})| < \frac{\varepsilon}{8}$ and $|\mu^*_{\text{EMP}}(\mathcal{P}) - \mu^*(\mathcal{P})| < \frac{\varepsilon}{8}$;
2. For every $v, w \in V$, if $\mu^*(X_v = X_w) \geq 1 - \frac{1}{20n^2L}$, then $\tau_i(v) = \tau_i(w)$ for all $1 \leq i \leq L$.

The probability space associated with the event $\mathcal{F}$ is determined by the random samples $\{\sigma_1, \ldots, \sigma_L\}$ and $\{\tau_1, \ldots, \tau_L\}$. If $\mathcal{F}$ occurs, then the empirical and true distributions are close to each other. We can prove that the event $\mathcal{F}$ occurs with probability at least $19/20$.

Lemma 32 *Suppose $L \geq 800n^2\varepsilon^{-2}$. Then for n sufficiently large we have* $\Pr[\mathcal{F}] \geq \frac{19}{20}$.

We justify next Steps 8-12 of Algorithm 1. Let P be the partition of V we get after Step 6 of the algorithm. Let $\mathcal{E}$ be the event that the following two events occur:

1. $\mu_{\text{EMP}}(\mathcal{P}) > 1 - \frac{\varepsilon}{4}$;
2. For every edge $\{v, w\} \in E \backslash E(P)$, $\beta(v, w) < \ln(20n^2L)$.

Like $\mathcal{F}$, the probability space associated with $\mathcal{E}$ is determined by the random samples $\{\sigma_1, \ldots, \sigma_L\}$ and $\{\tau_1, \ldots, \tau_L\}$. Observe that if the event $\mathcal{E}$ does not occur, then Algorithm 1 will output No. The following lemma justifies this.

Lemma 33 *Assume the event $\mathcal{F}$ occurs. If $(G, \beta) = (G^*, \beta^*)$, then the event $\mathcal{E}$ always occurs.*

Now, suppose $\mathcal{F}$ and $\mathcal{E}$ occur; then, the algorithm outputs the answer by running the DDK algorithm on the quotient models. The next two lemmas, in which we establish several useful properties of the quotient models $(G_{\text{P}}, \beta_{\text{P}})$ and $(G^*_{\text{P}}, \beta^*_{\text{P}})$, will be used to guarantee the correctness of Algorithm 1 in this case.

Lemma 34 *Assume both of the events $\mathcal{F}$ and $\mathcal{E}$ occur. Let P be the partition of V we get after Step 6 of Algorithm 1. Then the following events always occur:*

1. *The original Gibbs distribution μ and the conditional Gibbs distribution $\mu(\cdot|\mathcal{P})$ are close:*
$$\|\mu - \mu(\cdot|\mathcal{P})\|_{\text{TV}} < \frac{3\varepsilon}{8};$$
2. *The original Gibbs distribution μ^* and the conditional Gibbs distribution $\mu^*(\cdot|\mathcal{P})$ are close:*
$$\|\mu^* - \mu^*(\cdot|\mathcal{P})\|_{\text{TV}} < \frac{\varepsilon}{8}.$$

Lemma 35 *Assume both of the events $\mathcal{F}$ and $\mathcal{E}$ occur. Let P be the partition of V we get after Step 6 of Algorithm 1. Then the following events always occur:*

1. *For every edge $\{v, w\} \in E \backslash E(P)$, $\beta(v, w) < \ln(20n^2L)$;*
2. *For every edge $\{v, w\} \in E^* \backslash E(P)$, $\beta^*(v, w) < \ln(20n^2L)$.*

The proof of Lemmas 32, 33, 34 and 35 are provided in Section 6.1. We are now ready to prove Theorem 30.

Proof of Theorem 30 Assume the event $\mathcal{F}$ occurs. If the event $\mathcal{E}$ does not occur, then by Lemma 33 we have $(G, \beta) \neq (G^*, \beta^*)$, and Algorithm 1 will accordingly return No. Let us assume that the event $\mathcal{E}$ occurs. Recall that we denote the Gibbs distributions of the quotient models $(G_{\text{P}}, \beta_{\text{P}})$ and $(G^*_{\text{P}}, \beta^*_{\text{P}})$ by μ_{P} and μ^*_{P} respectively. If $\mu = \mu^*$, then $\mu(\cdot|\mathcal{P}) = \mu^*(\cdot|\mathcal{P})$ and therefore $\mu_{\text{P}} = \mu^*_{\text{P}}$. Otherwise, if $\|\mu - \mu^*\|_{\text{TV}} > \varepsilon$, then we deduce from Lemma 34 that

$$\begin{aligned}\|\mu_{\text{P}} - \mu^*_{\text{P}}\|_{\text{TV}} &= \|\mu(\cdot|\mathcal{P}) - \mu^*(\cdot|\mathcal{P})\|_{\text{TV}} \\ &\geq \|\mu - \mu^*\|_{\text{TV}} - \|\mu - \mu(\cdot|\mathcal{P})\|_{\text{TV}} - \|\mu^* - \mu^*(\cdot|\mathcal{P})\|_{\text{TV}} \\ &> \varepsilon - \frac{3\varepsilon}{8} - \frac{\varepsilon}{8} = \frac{\varepsilon}{2}.\end{aligned}$$

Since in every sample τ_i, vertices from the same partition class of P always receive the same spin, we can regard $\{\tau_i\}_{i=1}^L$ as independent samples from the conditional distribution $\mu^*(\cdot|\mathcal{P})$. Therefore, $\{(\tau_i)_{\text{P}}\}_{i=1}^L$ are independent samples from the Gibbs distribution μ^*_{P} of the quotient model $(G^*_{\text{P}}, \beta^*_{\text{P}})$. Thus, we can run the DDK algorithm on inputs $(G_{\text{P}}, \beta_{\text{P}})$, $\{(\tau_i)_{\text{P}}\}_{i=1}^L$ and $\varepsilon/2$. By Theorem 29 and Lemma 35, the number of samples needed is $L = O(n^2 d^2 \varepsilon^{-2} \log^3 n)$.

Consequently, Algorithm 1 fails only if the event $\mathcal{F}$ does not occur or the DDK algorithm makes a mistake. By Theorem 29, the DDK algorithm has success probability 4/5. Thus, Lemma 32 implies that the failure probability of Algorithm 1 is at most $1/20 + 1/5 = 1/4$, provided $L \geq 800n^2\varepsilon^{-2}$. Finally, Theorems 28 and 29 imply that the overall running time of Algorithm 1 is $\text{poly}(n, d, \varepsilon^{-1})$ as claimed. ■

6.1. Proofs of Auxiliary Lemmas

In this section we provide the missing proofs of Lemmas 32, 33, 34 and 35.

Proof of Lemma 32 Let P be a partition of V. Since the samples $\{\sigma_1, \ldots, \sigma_L\}$ from (G, β) are $(\varepsilon/16)$-approximate, we have

$$|\mu_{\text{ALG}}(\mathcal{P}) - \mu(\mathcal{P})| \leq \|\mu_{\text{ALG}} - \mu\|_{\text{TV}} \leq \frac{\varepsilon}{16}.$$

Then, by the triangle inequality

$$|\mu_{\text{EMP}}(\mathcal{P}) - \mu(\mathcal{P})| \leq |\mu_{\text{EMP}}(\mathcal{P}) - \mu_{\text{ALG}}(\mathcal{P})| + |\mu_{\text{ALG}}(\mathcal{P}) - \mu(\mathcal{P})| \leq |\mu_{\text{EMP}}(\mathcal{P}) - \mu_{\text{ALG}}(\mathcal{P})| + \frac{\varepsilon}{16}$$

A Chernoff bound then implies

$$\begin{aligned}
\Pr\left[|\mu_{\text{EMP}}(\mathcal{P}) - \mu(\mathcal{P})| \geq \frac{\varepsilon}{8}\right] &\leq \Pr\left[|\mu_{\text{EMP}}(\mathcal{P}) - \mu_{\text{ALG}}(\mathcal{P})| \geq \frac{\varepsilon}{16}\right] \\
&= \Pr\left[\left|\sum_{i=1}^{L} \mathbb{1}\{\sigma_i \in \mathcal{P}\} - L\mu_{\text{ALG}}(\mathcal{P})\right| \geq \frac{\varepsilon L}{16}\right] \\
&\leq 2\exp\left(-\frac{\varepsilon^2 L}{768\mu_{\text{ALG}}(\mathcal{P})}\right) \\
&\leq 2\exp\left(-\frac{\varepsilon^2 L}{768}\right) \leq 2\mathrm{e}^{-n^2},
\end{aligned}$$

where the last inequality holds when $\varepsilon^2 L \geq 800n^2$. The total number of partitions of V is at most n^n. It then follows from the union bound that

$$\Pr\left[\exists \text{ a partition } P \text{ of } V : |\mu_{\text{EMP}}(\mathcal{P}) - \mu(\mathcal{P})| \geq \frac{\varepsilon}{8}\right] \leq n^n \cdot 2\mathrm{e}^{-n^2} = 2\mathrm{e}^{n\ln n - n^2} \leq \frac{1}{80},$$

for large enough n.

In similar fashion, we deduce that the same holds for $|\mu^*_{\text{EMP}}(\mathcal{P}) - \mu^*(\mathcal{P})|$. Namely,

$$\Pr\left[\exists \text{ a partition } P \text{ of } V : |\mu^*_{\text{EMP}}(\mathcal{P}) - \mu^*(\mathcal{P})| \geq \frac{\varepsilon}{8}\right] \leq \frac{1}{80}.$$

Finally, for each $v, w \in V$ such that $\mu^*(X_v = X_w) \geq 1 - (20n^2L)^{-1}$, we obtain from a union bound over the samples that

$$\Pr\left[\exists i,\, 1 \leq i \leq L : \tau_i(v) \neq \tau_i(w)\right] \leq L \cdot \mu^*(X_v \neq X_w) \leq \frac{1}{20n^2}.$$

Another union bound, this time over the pairs of vertices, implies

$$\Pr\left[\exists v, w \in V,\, \hat{\mu}(X_v = X_w) \geq 1 - (20n^2L)^{-1},\, \exists i,\, 1 \leq i \leq L : \tau_i(v) \neq \tau_i(w)\right] \leq \frac{n^2}{2} \cdot \frac{1}{20n^2} = \frac{1}{40}.$$

Combining all bounds above, we obtain from another union bound that

$$\Pr[\neg\mathcal{F}] \leq \frac{1}{80} + \frac{1}{80} + \frac{1}{40} = \frac{1}{20},$$

as desired. ■

Proof of Lemma 33 Assume $(G, \beta) = (G^*, \beta^*)$. Since $\mathcal{F}$ occurs, we have

$$\mu_{\text{EMP}}(\mathcal{P}) > \mu(\mathcal{P}) - \frac{\varepsilon}{8} = \mu^*(\mathcal{P}) - \frac{\varepsilon}{8} > \mu^*_{\text{EMP}}(\mathcal{P}) - \frac{\varepsilon}{4} = 1 - \frac{\varepsilon}{4}.$$

Suppose next that there exists some $\{v, w\} \in E \backslash E(P)$ such that $\beta(v, w) \geq \ln(20n^2L)$. Then, $\beta^*(v, w) = \beta(v, w) \geq \ln(20n^2L)$. We deduce from Lemma 11 in (Daskalakis et al., 2018) that

$$\mu^*(X_v \neq X_w) \leq \frac{1}{\mathrm{e}^{\beta^*(v,w)} + 1} \leq \mathrm{e}^{-\beta^*(v,w)} \leq \frac{1}{20n^2L}.$$

The event $\mathcal{F}$ implies that $\tau_i(v) = \tau_i(w)$ for all $1 \leq i \leq L$. It follows that v and w must belong to the same partition class of P; i.e., $\{u, v\} \in E(P)$, which leads to a contradiction. Hence, the event $\mathcal{E}$ always occurs when $\mathcal{F}$ occurs. ■

Proof of Lemma 34 Since both of the events $\mathcal{F}$ and $\mathcal{E}$ occur, we have

$$\mu(\mathcal{P}) > \mu_{\text{EMP}}(\mathcal{P}) - \frac{\varepsilon}{8} > 1 - \frac{3\varepsilon}{8}.$$

It follows immediately that $\|\mu - \mu(\cdot|\mathcal{P})\|_{\text{TV}} = 1 - \mu(\mathcal{P}) < 3\varepsilon/8$. Similarly, since the event $\mathcal{F}$ occurs, we have

$$\mu^*(\mathcal{P}) > \mu^*_{\text{EMP}}(\mathcal{P}) - \frac{\varepsilon}{8} = 1 - \frac{\varepsilon}{8},$$

and $\|\mu^* - \mu^*(\cdot|\mathcal{P})\|_{\text{TV}} = 1 - \mu^*(\mathcal{P}) < 3\varepsilon/8$. ■

Proof of Lemma 35 When $\mathcal{E}$ occurs, for all $\{v, w\} \in E \backslash E(P)$, we have $\beta(v, w) < \ln(20n^2L)$. Suppose $\{v, w\} \in E^* \backslash E(P)$ and $\beta^*(v, w) \geq \ln(20n^2L)$. By Lemma 11 in (Daskalakis et al., 2018) we have

$$\mu^*(X_v \neq X_w) \leq \frac{1}{\mathrm{e}^{\beta^*(v,w)} + 1} \leq \mathrm{e}^{-\beta^*(v,w)} \leq \frac{1}{20n^2L}.$$

When $\mathcal{F}$ occurs, $\tau_i(v) = \tau_i(w)$ for all $1 \leq i \leq L$ and thus v and w belong to the same partition class of P. It follows that $\{v, w\} \in E(P)$ which is a contradiction. ■

7. Lower Bounds for Proper q-colorings

Let d and q be positive integers and let $G = (V, E) \in \mathcal{M}(n, d)$, where, as in the previous sections, $\mathcal{M}(n, d)$ denotes the family of all n-vertex graphs of maximum degree at most d. We use Ω_G for the set of all proper q-colorings of G and μ_G for the uniform distribution on Ω_G. We recall that a coloring of the vertices of G using colors $\{1, \ldots, q\}$ is proper if the endpoints of every edge in G are assigned different colors. The proper q-colorings model is one of the easiest combinatorial examples of a hard-constraint model.

The identity testing problem for proper q-colorings in $\mathcal{M}(n, d)$ is described 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 spin systems.

Our lower bounds will crucially use the presumed hardness of the #BIS problem. This is the problem of counting independent sets in bipartite graphs. #BIS is believed not to have an FPRAS, and it is widely used in the study of the complexity of approximate counting problems (see, e.g. Dyer et al., 2004; Goldberg and Jerrum, 2012; Dyer et al., 2010; Bulatov et al., 2013; Chen et al., 2015; Cai et al., 2016; Galanis et al., 2016a). Specifically, we utilize the hardness of the problem of counting proper 3-colorings in bipartite graphs, which we denote by #BIP-3-COL and is known to be no easier than #BIS.

Theorem 36 *(Dyer et al., 2004). If* #BIP-3-COL *admits an FPRAS, then* #BIS *admits an FPRAS.*

We show that when $d \geq q+\lceil\sqrt{q}\rceil-1$, any identity testing algorithm for proper q-colorings for $\mathcal{M}(n,d)$ with running time $T(n)$ and sample complexity $L(n)$ provides a randomized algorithm for #BIP-3-COL on graphs of poly(n) size with running time poly$(T(n),L(n))$. This will allow us to establish Theorem 3 from the introduction, since if $T(n)$ and $L(n)$ were polynomials in n, then one would obtain an FPRAS for #BIP-3-COL and for #BIS by Theorem 36. For $q \in \mathbb{N}^+$, let

$$d_c(q) = q + \lceil\sqrt{q}\rceil - 1. \tag{22}$$

Theorem 37 *Let d and q be positive integers such that $q \geq 3$ and $d \geq d_c(q)$. Suppose that, for all sufficiently large n, there is an identity testing algorithm for proper q-colorings in $\mathcal{M}(n,d)$ with running time $T(n)$ and sample complexity $L(n)$. Then, for every integer N sufficiently large, $\delta \in (0,1)$ and $\varepsilon \in (0,1)$, there exists an integer $n = \Theta(\varepsilon^{-2}N^4)$ such that if $L(n) \leq 2^{N-4}$, then there is an algorithm that with probability at least $1-\delta$ computes an ε-approximation for* #BIP-3-COL *on bipartite graphs with N vertices. The running time of this algorithm is*

$$O\left([nL(n)+T(n)]N\ln(N/\delta)+\varepsilon^{-8}\right).$$

Theorem 3 is a direct corollary of this result.

Proof of Theorem 3 Suppose there is an identity testing algorithm for q-colorings in $\mathcal{M}(n,d)$ with poly(n) running time and sample complexity; i.e., $L(n) \leq T(n) = \text{poly}(n)$. Then, by Theorem 37, for any $\varepsilon,\delta \in (0,1)$ there is an algorithm for #BIP-3-COL on an N-vertex bipartite graph that outputs an ε-approximation solution with probability at least $1-\delta$ in time poly$(N,\varepsilon^{-1},\ln(\delta^{-1}))$. That is, there is an FPRAS for #BIP-3-COL and thus also one for #BIS by Theorem 36. This leads to a contradiction and the result follows. ■

The proof of Theorem 37 is divide into two cases: $q \geq 4$ and $q = 3$. Conceptually, these two cases are proved in the same manner but in the $q = 3$ case the construction of the testing instance requires some additional ideas. The proof for $q \geq 4$ is provided in Section 7.4. The $q = 3$ case is considered in in Section 7.5. Before that, we provide a proof sketch containing the high level ideas of our proof in Section 7.1, we introduce our gadget $G(m,q,t)$ in Section 7.2, and we describe the construction of the coloring instance in Section 7.3.

7.1. Lower Bounds for Proper Colorings: Proof Overview

As mentioned, we crucially use in our proof the hardness of #BIP-3-COL, the problem of counting proper 3-colorings in bipartite graphs. 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 poly(n) size with running time poly$(T(n), L(n))$; see Theorem 36. Theorem 37 follows from the fact that if $T(n)$ and $L(n)$ were both polynomials in n, then we would obtain an algorithm that computes an ε-approximation for #BIP-3-COL in polynomial time.

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 Section 7.5.) 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 (Emden-Weinert et al., 1998; Molloy and Reed, 2001; Blanca et al., 2018).

7.2. The Colorings Gadget

In this section, we present our construction of the coloring gadget, which is inspired by similar constructions in (Emden-Weinert et al., 1998; Molloy and Reed, 2001; Blanca et al., 2018) for establishing the computational hardness of the decision and (equivalent) structure learning problems for proper q-colorings.

For $m, q, t \in \mathbb{N}^+$ with $t < q$, the graph $G(m, q, t) = (V(m, q, t), E(m, q, t))$ is defined as follows. Let $C_1, \ldots, C_m$ be cliques of size $q - 1$ and let $I_1, \ldots, I_m$ be independent sets of size t. Then, set

$$V(m, q, t) = \bigcup_{i=1}^{m} \big(V(C_i) \cup V(I_i)\big)$$

where $V(C_i)$ and $V(I_i)$ are the vertex sets of C_i and I_i respectively for $1 \leq i \leq m$. The cliques C_i's and the independent sets I_i's are connected in the following way:

1. For $1 \leq i \leq m$, there is a complete bipartite graph between C_i and I_i. That is, for $u \in C_i$ and $v \in I_i$, $\{u, v\} \in E(m, q, t)$.

2. For $2 \leq i \leq m$, each C_i is partitioned into t almost-equally-sized disjoint subsets $C_{i,1}, \ldots, C_{i,t}$ of size either $\lfloor (q-1)/t \rfloor$ or $\lceil (q-1)/t \rceil$. Then, the j-th vertex of I_{i-1} is connected to every vertex in $C_{i,j}$.

Together with the edges in the cliques C_i for $1 \leq i \leq m$, these edges constitute the edge set $E(m, q, t)$. Furthermore, a vertex is said to be a *port* of the graph $G(m, q, t)$ if it is either in I_m or in some I_{i-1} for $2 \leq i \leq m$ and adjacent to some $C_{i,j}$ of size exactly $\lfloor (q-1)/t \rfloor$. Note that every independent set I_i contains at least one port, and thus there are at least m ports in the graph $G(m, q, t)$. See Figure 1 for an illustration of the graph $G(m, q, t)$ and Figure 2 for $G(3, 3, 2)$ as an example.

The following key fact of the gadget $G(m, q, t)$ follows from its definition.

Lemma 38 *Let $m, q, t \in \mathbb{N}^+$ with $t < q$. Then in every proper q-coloring of $G(m, q, t)$, all vertices in the independent sets $I_1, \ldots, I_m$ have the same color and all vertices in the cliques $C_1, \ldots, C_m$ are assigned the remaining $q - 1$ colors.*

Proof Consider a proper q-coloring σ of $G(m, q, t)$. Since each C_i is a clique of size $q - 1$ for $1 \leq i \leq m$, it receives $q - 1$ colors in σ. For $1 \leq i \leq m$, for each $v \in I_i$, v is adjacent to all vertices in C_i; hence, v receives the only color that is not used by C_i in σ. That means, for each i all vertices in I_i have the same color which does not appear in C_i. Next, for $2 \leq i \leq m$, each vertex in C_i is adjacent to some vertex in I_{i-1}. Since all vertices in I_{i-1} have the same color, we deduce that I_{i-1} receives the color which does not appear in C_i. It follows immediately that all independent sets $I_1, \ldots, I_n$ have the same color and the cliques $C_1, \ldots, C_m$ use the remaining $q - 1$ colors. ■

The following lemma shows that when $d \geq d_c(q)$ the maximum degree of the gadget $G(m, q, t)$ is at most d for a certain choice of t; moreover, every port has degree at most $d - 1$. See Figure 2 for an example.

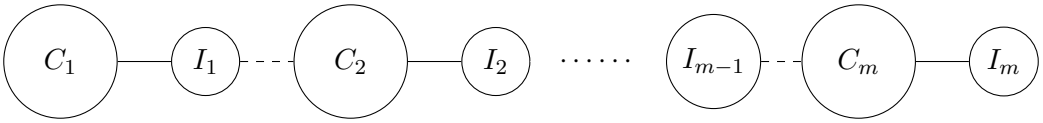

Figure 1: The graph $G(m, q, t)$. Each of $C_1, \ldots, C_m$ is a clique of size $q - 1$ and each of $I_1, \ldots, I_m$ is an independent set of size $t < q$. Solid lines between C_i and I_i mean that every vertex in C_i is adjacent to every vertex in I_i. Dashed lines between I_{i-1} and C_i mean that every vertex in I_{i-1} is adjacent to roughly $(q-1)/t$ vertices in C_i with no two vertices in I_{i-1} sharing a common neighbor in C_i.

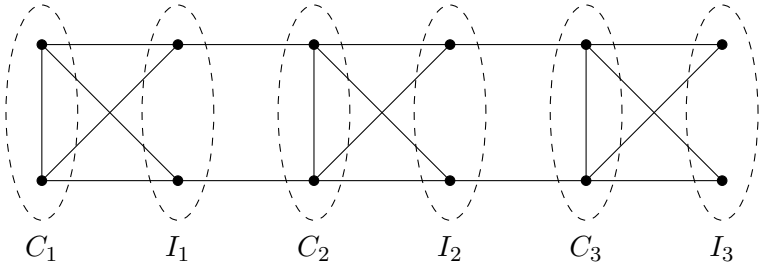

Figure 2: The graph $G(3, 3, 2)$ for $m = 3$, $q = 3$ and $t = \lceil \sqrt{q} \rceil = 2$. All vertices in $I_1 \cup I_2 \cup I_3$ are ports. Every port has degree at most 3. Every non-port has degree at most 4. Recall that $d_c(3) = 3 + \lceil \sqrt{3} \rceil - 1 = 4$.

Lemma 39 *Suppose $q \geq 3$ and $d \geq d_c(q)$. If $t = \lceil \sqrt{q} \rceil$, then every port of the graph $G(m, q, t)$ has degree at most $d - 1$ and every non-port of $G(m, q, t)$ has degree at most d.*

Proof The degree of a port in $G(m, q, t)$ is bounded by

$$(q-1) + \left\lfloor \frac{q-1}{t} \right\rfloor \overset{\text{(i)}}{=} (q-1) + \left\lceil \frac{q}{t} \right\rceil - 1 \overset{\text{(ii)}}{\leq} (q-1) + \lceil \sqrt{q} \rceil - 1 \leq d - 1,$$

where (i) follows from the fact that $\lfloor (a-1)/b \rfloor + 1 = \lceil a/b \rceil$ for all $a, b \in \mathbb{N}^+$ and (ii) follows from $t \geq \sqrt{q}$. Meanwhile, the degree of a non-port in an independent set I_i is at most

$$(q-1) + \left\lceil \frac{q-1}{t} \right\rceil \leq (q-1) + \left\lceil \frac{q}{t} \right\rceil \leq (q-1) + \lceil \sqrt{q} \rceil \leq d,$$

and the degree of a non-port in a clique C_i is at most

$$(q-2) + t + 1 = (q-1) + \lceil \sqrt{q} \rceil \leq d. \qquad \blacksquare$$

We define the *phase* of a proper q-coloring of $G(m, q, t)$ to be the color of its ports. In the following lemma we bound the number of q-colorings with a given phase, which is used later in the proof of Theorem 37.

Lemma 40 *Let $m, q, t \in \mathbb{N}^+$ with $t < q$. Then, the number of proper q-colorings of $G(m, q, t)$ with a given phase is $[(q-1)!]^m$.*

Proof By Lemma 38, in every proper q-colorings the vertices in the independent sets $I_1, \ldots, I_m$ are assigned the same color, which is given by its phase. With the coloring of these vertices fixed, the number of q-colorings of each clique C_i is $(q-1)!$. Since the cliques are disjoint the total number of q-colorings is $[(q-1)!]^m$. ■

7.3. Testing Instance Construction: the $q \geq 4$ Case

Let $H = (V, E)$ be a bipartite connected graph on N vertices and suppose we want to approximately count the number of 3-colorings of H. In this section we show how to construct the testing instance from H when $q \geq 4$. Our construction uses the gadget from Section 7.2.

For integers $k, \ell \geq 1$, define the simple graph $\hat{H}_{k,\ell} = (\hat{V}, \hat{E})$ as follows:

1. Let $H_1 = (V(H_1), E(H_1)), \ldots, H_k = (V(H_k), E(H_k))$ be k copies of the graph H;
2. Let J be a complete $(q-3)$-partite graph in which each cluster has ℓ vertices;
3. Set $\hat{V} = \left(\bigcup_{i=1}^{k} V(H_i)\right) \cup V(J)$;
4. In addition to the edges in J and those in H_i for $1 \leq i \leq k$, $\hat{E}$ also contains edges between every vertex in H_i for $1 \leq i \leq k$ and every vertex in J; i.e., for $u \in H_i$ and $v \in J$, we have $\{u, v\} \in \hat{E}$.

We remark that our definition of $\hat{H}_{k,\ell}$ requires $q \geq 4$; when $q = 4$, J is simply an independent set with ℓ vertices. Next, we use the graph $G(m, q, t)$ from Section 7.2 as a gadget to construct a simple graph $\hat{H}^{\Gamma}_{k,\ell}$ based on $\hat{H}_{k,\ell}$ where $\Gamma = \{m, q, t\}$. We proceed as follows:

1. Replace every vertex v of $\hat{H}_{k,\ell}$ by a copy G_v of $G(m, q, t)$;
2. For every edge $\{u, v\} \in \hat{E}$, pick an unused port in G_u and an unused port in G_v and connect them; in this way, every port is connected with at most one port from another gadget.

The number of ports in a gadget $G(m, q, t)$ is at least m, and the total number of vertices in $\hat{H}_{k,\ell}$ is $kN + \ell(q-3)$. The graph $\hat{H}^{\Gamma}_{k,\ell}$ is well-defined only if we have enough ports in every gadget G_v to connect them with ports from other gadgets. For this, it suffices that

$$m \geq kN + \ell(q-3),$$

and so we set

$$m = kN + \ell(q-3) \qquad \text{and} \qquad t = \lceil \sqrt{q} \rceil \geq 1. \tag{23}$$

Let B be a complete bipartite graph with the same vertex bipartition as H. By setting $H = B$, we can define the graphs $\hat{B}_{k,\ell}$ and $\hat{B}^{\Gamma}_{k,\ell}$. Given $k, \ell \in \mathbb{N}^+$, we write $G = \hat{H}^{\Gamma}_{k,\ell}$ and $G^* = \hat{B}^{\Gamma}_{k,\ell}$ for our choice of m and t. Suppose $q \geq 3$ and $d \geq d_c(q)$. Then, Lemma 39 implies that $G, G^* \in \mathcal{M}(n, d)$ for

$$n = [kN + \ell(q-3)] \cdot [m(q-1+t)] = m^2(q-1+t). \tag{24}$$

Let $Z_3(H)$ and $Z_3(B)$ denote the number of 3-colorings of H and B, respectively. The two uniform distributions μ_G and μ_{G^*} over the q-colorings of G and G^* are related as follows.

Lemma 41 *Let $k, \ell \in \mathbb{N}^+$ with $\ell \geq 2$. Define $\psi(k,\ell) = (q-3)^{1/k}\, 2^{1+\ell/k}$. Then the following holds:*

1. *If $Z_3(H) < \psi(k,\ell)$, then*

$$\|\mu_G - \mu_{G^*}\|_{\text{TV}} \leq \frac{4}{3}\left(\frac{Z_3(H)}{\psi(k,\ell)}\right)^k.$$

2. *If $Z_3(H) \geq \psi(k,\ell)$, then*

$$\|\mu_G - \mu_{G^*}\|_{\text{TV}} \geq \frac{2}{5}\left(1 - \left(\frac{Z_3(B)}{Z_3(H)}\right)^k\right).$$

Finally, we note that we can generate random q-colorings of G^* in polynomial time.

Lemma 42 *There exists an algorithm with running time $O(n)$ that generates a sample from the distribution μ_{G^*}.*

The proof of both of these lemmas are provided in Section 7.4.1.

7.4. Proof of Theorem 37: the $q \geq 4$ Case

In this section, we prove Theorem 37 for the case when $q \geq 4$. Our proof relies on Lemmas 41 and 42. We converge to a good approximation for the number of 3-colorings $Z_3(H)$ of a bipartite graph H using the presumed algorithm for the identity testing problem. In each round, we choose k, ℓ and generate the graph $G = \hat{H}^{\Gamma}_{k,\ell}$ as described in Section 7.3; the size of the graph G depends on k, ℓ and thus it varies in each round. We then generate samples from μ_{G^*} in polynomial time by Lemma 42 where $G^* = \hat{B}^{\Gamma}_{k,\ell}$. These samples and the graph G are passed as input to the identity testing algorithm. If $Z_3(H) < \psi(k,\ell)$, then μ_G and μ_{G^*} are close in total variation distance (see Lemma 41), and the tester would return Yes. Otherwise, if $Z_3(H) \geq \psi(k,\ell)$, then μ_G and μ_{G^*} are statistically far from each other, and the tester would return No. Thus, using binary search over k, ℓ we can obtain a good approximation for $Z_3(H)$.

Proof of Theorem 37 for $q \geq 4$ Let $H = (V(H), E(H))$ be an N-vertex connected bipartite graph with $N \geq 5$. Suppose we want to approximately count the number of 3-colorings of H. Recall that B is the complete bipartite graph with the same vertex bipartition as H. Then, $Z_3(B) \leq Z_3(H) \leq 3^N$ where the upper bound corresponds to the independent set on N vertices.

Fix $\varepsilon, \delta \in (0,1)$. Our goal is to find an integer $\hat{Z} \in [Z_3(B), 3^N]$ such that with probability at least $1-\delta$

$$(1-\varepsilon)\hat{Z} \leq Z_3(H) \leq (1+\varepsilon)\hat{Z}. \tag{25}$$

We assume first that $\varepsilon \geq 2^{-N/4}$. The case when $\varepsilon < 2^{-N/4}$ is much simpler and will be considered at the end of the proof. We give an algorithm that with probability at least $1-\delta$ outputs an integer $\hat{Z} \in [Z_3(B), 3^N]$ such that

$$2^{-\varepsilon}(\hat{Z}-1) \leq Z_3(H) \leq 2^{\varepsilon}\hat{Z}. \tag{26}$$

Then, (25) follows from the following fact.

Fact 43 *For all $\varepsilon \in [2^{-N/4}, 1)$, if $\hat{Z}$ is such that $2^{-\varepsilon}(\hat{Z}-1) \le Z_3(H) \le 2^{\varepsilon}\hat{Z}$, then $(1-\varepsilon)\hat{Z} \le Z_3(H) \le (1+\varepsilon)\hat{Z}$.*

Let $k = \lceil N/\varepsilon \rceil$. Recall that $\psi(k,\ell) = (q-3)^{1/k}\ 2^{1+\ell/k}$. For any $\hat{Z} \in \mathbb{N}^+$ for which we would like to test if (26) hold, we choose an integer ℓ satisfying

$$2^{-\varepsilon}\psi(k,\ell) \le \hat{Z} \le \psi(k,\ell).$$

Such an ℓ would exist if and only if it satisfies

$$k\log_2 \hat{Z} - \log_2(q-3) - k \le \ell \le k\log_2 \hat{Z} - \log_2(q-3) - k + k\varepsilon.$$

Since the difference between the upper and lower bounds is $k\varepsilon \ge N \ge 1$, there is always at least one possible value for ℓ. Note also that $\ell \le k\log_2(3^N) \le 2kN$ as $\hat{Z} \le 3^N$.

After choosing k and ℓ, which depend on N, q, ε and $\hat{Z}$, we construct the graphs $G = \hat{H}^{\Gamma}_{k,\ell}$ and $G^* = \hat{B}^{\Gamma}_{k,\ell}$ as defined in Section 7.3. Then, the graphs G and G^* belong to $\mathcal{M}(n_{k,\ell}, d)$, where given our choices for m, t, k and ℓ, we have:

$$m = kN + \ell(q-3) \le \frac{4qN^2}{\varepsilon} \quad \text{and} \quad n_{k,\ell} = m^2(q-1+t) \le \frac{32q^3N^4}{\varepsilon^2};$$

see (23) and (24). Given $\hat{Z}$, our input to the identity testing algorithm (henceforth called the Tester) is the graph G and $L = L(n_{k,\ell})$ random q-colorings of G^*. By Lemma 42, we can generate one sample from μ_{G^*} in $O(n_{k,\ell})$ time. Thus, the total running time for one call of the Tester (including the generation of the samples) is $O(nL(n) + T(n))$ for $n = \lceil 32q^3\varepsilon^{-2}N^4 \rceil$. The following claim which is proved later follows from Lemma 41.

Claim 44 *Suppose $Z_3(B) \le \hat{Z} \le 3^N$ and $L \le 2^{N-4}$.*

1. *If $Z_3(H) < 2^{-\varepsilon}\hat{Z}$, then the* Tester *outputs* Yes *with probability at least $2/3$;*
2. *If $Z_3(H) > 2^{\varepsilon}\hat{Z}$, then the* Tester *outputs* No *with probability at least $2/3$.*

We test whether $\hat{Z}$ provides a bound for $Z_3(H)$ using the following algorithm. For $R \ge 1$ odd, we construct the corresponding graph $G = \hat{H}^{\Gamma}_{k,\ell}$, generate $L \cdot R$ random colorings of $G^* = \hat{B}^{\Gamma}_{k,\ell}$, and run the Tester R times using L samples each time (every sample is used only once). The output of this algorithm would be the majority answer in the R rounds. We call this algorithm the *R-round-*Tester for $\hat{Z}$. The following claim, which follows directly from a Chernoff bound and is provided later, establishes the guarantee for the accuracy of the R-round-Tester.

Claim 45 *Let $R = 48\lceil \ln(2N/\delta) \rceil + 1$.*

1. *If $Z_3(H) < 2^{-\varepsilon}\hat{Z}$, then R-round-*Tester *for $\hat{Z}$ outputs* Yes *with probability at least $1 - \frac{\delta}{2N}$;*
2. *If $Z_3(H) > 2^{\varepsilon}\hat{Z}$, then R-round-*Tester *for $\hat{Z}$ outputs* No *with probability at least $1 - \frac{\delta}{2N}$.*

The algorithm for counting 3-colorings in H is based on binary search over the interval $[Z_3(B), 3^N]$; in each iteration it uses the R-round-Tester to determine the interval for the next iteration. We proceed as follows.

1. Run the R-round-Tester for $\hat{Z} = Z_3(B)$. If the R-round-Tester outputs Yes, then return $\hat{Z} = Z_3(B)$;
2. Run the R-round-Tester for $\hat{Z} = 3^N$. If the R-round-Tester outputs No, then return $\hat{Z} = 3^N$;
3. Let $(L_0, U_0) = (Z_3(B), 3^N)$. For $i \geq 1$:
   (a) Let $C_i = \lfloor (L_{i-1} + U_{i-1})/2 \rfloor$;
   (b) Run the R-round-Tester for $\hat{Z} = C_i$;
   (c) If the R-round-Tester outputs Yes, then set $(L_i, U_i) = (L_{i-1}, C_i)$;
   (d) If the R-round-Tester outputs No, then set $(L_i, U_i) = (C_i, U_{i-1})$;
   (e) If $U_i - L_i = 1$, return $\hat{Z} = U_i$; otherwise, set $i := i + 1$ and repeat.

Observe that $U_i - L_i - 1$ decreases by a factor 2 in each iteration. Thus, the R-round-Tester is called at most $2 + \log_2(3^N) \leq 2N$ times for $N \geq 5$.

Now, let $\mathcal{F}$ be the event that in a single run of the binary search algorithm the following two conditions are maintained:

(i) If $Z_3(H) < 2^{-\varepsilon}\hat{Z}$, then the R-round-Tester outputs Yes for $\hat{Z}$;

(ii) If $Z_3(H) > 2^{\varepsilon}\hat{Z}$, then the R-round-Tester outputs No for $\hat{Z}$.

Claim 45 and a union bound imply that

$$\Pr[\neg\mathcal{F}] \leq \frac{\delta}{2N} \cdot 2N = \delta. \tag{27}$$

We claim that when $\mathcal{F}$ occurs, the output of the binary search algorithm satisfies (26). For this we consider three cases. First, if the algorithm stops in step 1, then $\hat{Z} = Z_3(B)$; that is, the R-round-Tester outputs Yes for $\hat{Z} = Z_3(B)$. Therefore,

$$2^{-\varepsilon}(\hat{Z} - 1) \leq Z_3(B) \leq Z_3(H) \leq 2^{\varepsilon}\hat{Z},$$

where the last inequality follows from condition (ii) in the definition of the event $\mathcal{F}$. Similarly, if the algorithm stops in step 2, then $\hat{Z} = 3^N$. Namely, the R-round-Tester outputs No for $\hat{Z} = 3^N$, and so

$$2^{-\varepsilon}(\hat{Z} - 1) \leq 2^{-\varepsilon}\hat{Z} \leq Z_3(H) \leq 3^N \leq 2^{\varepsilon}\hat{Z},$$

where the second inequality follows from condition (i) in the definition of $\mathcal{F}$.

Finally, suppose that the binary search algorithm stops in step 3 and $\hat{Z} = U_i$ for some $i \geq 1$. Observe that $L_i < U_i$ for all $i \geq 1$. Moreover, for each $i \geq 1$ the R-round-Tester

outputs No for $\hat{Z} = L_i$ and Yes for $\hat{Z} = U_i$. The algorithm stops when $U_i - L_i = 1$ for some i. It follows from the definition of $\mathcal{F}$ that

$$2^{-\varepsilon}(\hat{Z} - 1) = 2^{-\varepsilon} L_i \leq Z_3(H) \leq 2^{\varepsilon} U_i = 2^{\varepsilon} \hat{Z}.$$

Therefore, the output of the binary search algorithm satisfies (26) whenever $\mathcal{F}$ occurs. From (27), it follows that we obtain an ε-approximation for $Z_3(H)$ with probability at least $1 - \delta$ as desired.

It remains for us to consider the overall running time of the binary search procedure. As mentioned, the R-round-Tester algorithm is called at most $2N$ times, and the running time of each call is $O(nL(n) + T(n))$ where $n = \lceil 32q^3\varepsilon^{-2}N^4 \rceil$. Hence, the overall running time of the algorithm is $O\left((nL(n) + T(n))N \ln(N/\delta)\right)$.

Finally, we mention that for the trivial case when $\varepsilon < 2^{-N/4}$, we can simply enumerate every 3-labeling $\sigma : V(H) \to \{1, 2, 3\}$ of H and count the number of proper 3-colorings. The running time of this process is $O(3^N) \leq O(\varepsilon^{-8})$. ■

We finalize the proof of Theorem 37 for $q \geq 4$ by providing the missing proofs of Fact 43 and Claims 44 and 45 .

Proof of Fact 43 For $\varepsilon \in (0, 1)$, we have $2^{\varepsilon}\hat{Z} \leq (1 + \varepsilon)\hat{Z}$. Moreover, for $\varepsilon \in [2^{-N/4}, 1)$ we have

$$\frac{1}{1 - 2^{\varepsilon}(1 - \varepsilon)} \leq \frac{1}{1 - (1 + \varepsilon)(1 - \varepsilon)} = \frac{1}{\varepsilon^2} \leq 2^{N/2} \leq Z_3(B) \leq \hat{Z}.$$

This implies that $2^{-\varepsilon}(\hat{Z} - 1) \geq (1 - \varepsilon)\hat{Z}$ and the theorem follows. ■

Proof of Claim 44 Recall that we choose ℓ such that $2^{-\varepsilon}\psi(k, \ell) \leq \hat{Z} \leq \psi(k, \ell)$, where $\psi(k, \ell) = (q - 3)^{1/k}\, 2^{1+\ell/k}$. Hence, when $Z_3(H) < 2^{-\varepsilon}\hat{Z}$ we have

$$Z_3(H) < 2^{-\varepsilon}\hat{Z} \leq 2^{-\varepsilon}\psi(k, \ell) < \psi(k, \ell).$$

Part 1 of Lemma 41 implies

$$\|\mu_G - \mu_{G^*}\|_{\mathrm{TV}} \leq \frac{4}{3}\left(\frac{Z_3(H)}{\psi(k, \ell)}\right)^k \leq \frac{4}{3} \cdot 2^{-k\varepsilon} \leq \frac{4}{3} \cdot 2^{-N},$$

where the last inequality follows from the fact that $k = \lceil N/\varepsilon \rceil$. Let $\mu_G^{\otimes L}$ (resp., $\mu_{G^*}^{\otimes L}$) be the product distribution corresponding to L independent samples from μ_G (resp., μ_{G^*}). Recall that $L \leq 2^{N-4}$ by assumption. Then we get

$$\left\|\mu_G^{\otimes L} - \mu_{G^*}^{\otimes L}\right\|_{\mathrm{TV}} \leq L \left\|\mu_G - \mu_{G^*}\right\|_{\mathrm{TV}} \leq L \cdot \frac{4}{3} \cdot 2^{-N} \leq 2^{N-4} \cdot \frac{4}{3} \cdot 2^{-N} = \frac{1}{12}.$$

Consider the optimal coupling $\pi^{\otimes L}$ of the distributions $\mu_G^{\otimes L}$ and $\mu_{G^*}^{\otimes L}$. In a sample $(\mathcal{S}, \mathcal{S}')$ from $\pi^{\otimes L}$, the colorings from G and G^* are equal with probability at least $11/12$. Hence, following (12), we obtain

$$\Pr[\text{Tester outputs No when given samples } \mathcal{S} \sim \mu_{G^*}^{\otimes L}] \leq \frac{1}{4} + \frac{1}{12} = \frac{1}{3},$$

which establishes part 1 of the claim.

If $Z_3(H) > 2^{\varepsilon}\hat{Z}$, then $Z_3(H) > 2^{\varepsilon}\hat{Z} \geq \psi(k,\ell)$ and $Z_3(B) \leq \hat{Z} < 2^{-\varepsilon}Z_3(H)$. Part 2 of Lemma 41 implies that for $N \geq 5$

$$\|\mu_G - \mu_{G^*}\|_{\text{TV}} \geq \frac{2}{5}\left(1 - \left(\frac{Z_3(B)}{Z_3(H)}\right)^k\right) \geq \frac{2}{5}\left(1 - 2^{-k\varepsilon}\right) \geq \frac{2}{5}\left(1 - 2^{-N}\right) > \frac{1}{3}.$$

It follows that

$$\Pr[\textsc{Tester} \text{ outputs } \textsc{Yes}] = \Pr[\textsc{Tester} \text{ makes a mistake}] \leq \frac{1}{4} < \frac{1}{3}. \qquad \blacksquare$$

Proof of Claim 45 For $i = 1, \ldots, R$, let X_i be the indicator of the event that in the i-th round the Tester outputs Yes. Let $X = \sum_{i=1}^{R} X_i$. If $Z_3(H) < 2^{-\varepsilon}\hat{Z}$, then Claim 44 implies that $\mathrm{E}[X] \geq \frac{2}{3}R$. The Chernoff bound then implies that the probability that the R-round-Tester outputs No is

$$\Pr\left[X \leq \frac{R}{2}\right] \leq \Pr\left[X \leq \frac{3}{4}\mathrm{E}[X]\right] \leq \exp\left(-\frac{\mathrm{E}[X]}{32}\right) \leq \exp\left(-\frac{R}{48}\right) \leq \frac{\delta}{2N}.$$

The case when $Z_3(H) > 2^{\varepsilon}\hat{Z}$ (part 2) can be derived analogously. $\blacksquare$

7.4.1. Colorings of G and G^*: Proof of Lemmas 41 and 42

In this section we establish first several facts about the q-colorings of $G = \hat{H}^{\Gamma}_{k,\ell}$ and $G^* = \hat{B}^{\Gamma}_{k,\ell}$. We then use these facts to bound $\|\mu_G - \mu_{G^*}\|_{\text{TV}}$ (Lemma 41) and to design an algorithm for sampling the q-colorings of G^* (Lemma 42).

For $r \in \{2,3\}$, let $Z_r(H)$ denote the number of r-colorings of H. Since H is a connected bipartite graph, we have $Z_2(H) = 2$. The following lemma establishes a useful partition of the q-colorings of $\hat{H}_{k,\ell}$.

Lemma 46 *Let $k, \ell \in \mathbb{N}^+$. Let Ω^a and Ω^b be the set of q-colorings of $\hat{H}_{k,\ell}$ in which J is colored by exactly $q-3$ and $q-2$ colors respectively. Then $\{\Omega^a, \Omega^b\}$ is a partition for the set of q-colorings of $\hat{H}_{k,\ell}$; moreover,*

$$|\Omega^a| = \frac{1}{6}\,q!\,Z_3(H)^k \qquad \text{and} \qquad |\Omega^b| = \frac{1}{4}(q-3)\,q!\,(2^{\ell}-2)2^k.$$

Observe that in the colorings from Ω^a, the H_i's are assigned the remaining 3 colors, and in those from Ω^b they are colored with 2 colors. We provide the proof of this lemma next.

Proof of Lemma 46 Observe that J is a complete $(q-3)$-partite graph, so it requires at least $q-3$ colors in every proper q-coloring of $\hat{H}_{k,\ell}$. Moreover, since each H_i is a connected bipartite graph, it requires at least 2 colors in every q-coloring. Also, every vertex in H_i for $1 \leq i \leq k$ is adjacent to every vertex in J. Thus, the H_i's do not receive the colors that are used to color J. It then follows that $\{\Omega^a, \Omega^b\}$ is a partition for the set of q-colorings of $\hat{H}_{k,\ell}$.

We count next the number of q-colorings of each type. For colorings in Ω^a, there are $q!/3!$ ways to color J, and given the colors of J, there are $Z_3(H)$ colorings of each H_i that use the remaining 3 colors. This gives

$$|\Omega^a| = \frac{q!}{3!} \cdot Z_3(H)^k = \frac{1}{6}\, q!\, Z_3(H)^k.$$

For colorings in Ω^b, the complete $(q-3)$-partite graph J receives exactly $q-2$ colors. Hence, there is one cluster of J that is assigned 2 colors, and every other cluster of J is colored by one color of its own. There are $q-3$ ways of selecting the bichromatic cluster, $q!/4!$ choices for the colors of the $q-4$ monochromatic clusters and $\binom{4}{2}$ choices for the colors of the bichromatic cluster. Also, there are $2^\ell - 2$ colorings of the bichromatic cluster using exactly 2 colors. Finally, given the colors of J, we have $Z_2(H) = 2$ colorings for each H_i using the remaining 2 colors. Combining these, we get

$$|\Omega^b| = (q-3) \cdot \frac{q!}{4!} \cdot \binom{4}{2} \cdot (2^\ell - 2) \cdot Z_2(H)^k = \frac{1}{4}(q-3)\, q!\, (2^\ell - 2) 2^k. \qquad \blacksquare$$

Recall that the phase of a q-coloring of a gadget $G(m, q, t)$ is the color of its ports. Let σ be a q-coloring of $\hat{H}^\Gamma_{k,\ell}$. The *phase vector* of σ is a mapping $\tau : V(\hat{H}_{k,\ell}) \to \{1, \ldots, q\}$ defined as follows: for every vertex v of $\hat{H}_{k,\ell}$, $\tau(v)$ is the phase of the coloring σ in the gadget G_v for the vertex v. We show next that the phase vector of a q-coloring of $\hat{H}^\Gamma_{k,\ell}$ determines a q-coloring of $\hat{H}_{k,\ell}$.

Lemma 47 *Let σ be a q-coloring of $\hat{H}^\Gamma_{k,\ell}$ and τ be the phase vector of σ. Then, τ is a q-coloring of $\hat{H}_{k,\ell}$. Moreover, if τ is a q-coloring of $\hat{H}_{k,\ell}$, then there are $((q-1)!)^{m^2}$ q-colorings of $\hat{H}^\Gamma_{k,\ell}$ whose phase vector is τ.*

Proof In our construction, for every edge $\{u, v\}$ of $\hat{H}_{k,\ell}$ we connect one port of the gadget G_u with one port of G_v. Thus, the phase of G_u and the phase of G_v are distinct. This gives $\tau(u) \neq \tau(v)$ for every edge $\{u, v\}$ of $\hat{H}_{k,\ell}$. Hence, τ is a q-coloring of $\hat{H}_{k,\ell}$.

Given the phase vector τ of a q-coloring of $\hat{H}^\Gamma_{k,\ell}$, the number of ways to color each gadget is $((q-1)!)^m$ by Lemma 40. Since gadgets are connected to each other only by edges between ports, we deduce that given the phase vector τ (namely, the colors of all the ports in all the gadgets) the number of q-colorings of $\hat{H}^\Gamma_{k,\ell}$ is

$$[((q-1)!)^m]^{kN+\ell(q-3)} = ((q-1)!)^{m^2}$$

where we recall that the number of vertices of $\hat{H}_{k,\ell}$ is $kN + \ell(q-3)$ and we set $m = kN + \ell(q-3)$. $\blacksquare$

Combining Lemmas 46 and 47, we can also partition the q-colorings of $\hat{H}^\Gamma_{k,\ell}$ into two types.

Lemma 48 *Let $k, \ell \in \mathbb{N}^+$. Let Ω^A and Ω^B be the set of q-colorings of $\hat{H}^\Gamma_{k,\ell}$ whose phase vector is a q-coloring of $\hat{H}_{k,\ell}$ that belongs to Ω^a and Ω^b respectively. Then $\{\Omega^A, \Omega^B\}$ is a*

partition for the set of q-colorings of $\hat{H}^{\Gamma}_{k,\ell}$; moreover,

$$|\Omega^A| = |\Omega^a| \cdot [(q-1)!]^{m^2} = \frac{1}{6}\, q!\, Z_3(H)^k ((q-1)!)^{m^2},\ \textit{and}$$
$$|\Omega^B| = |\Omega^b| \cdot [(q-1)!]^{m^2} = \frac{1}{4}(q-3)\, q!\, (2^\ell - 2) 2^k ((q-1)!)^{m^2}.$$

Proof Follows immediately from Lemmas 46 and 47. ■

We are now ready to prove Lemmas 41 and 42.

Proof of Lemma 41 Let Ω_G, Ω^A_G and Ω^B_G denote the set of all q-colorings, q-colorings from Ω^A and q-colorings from Ω^B of the graph $G = \hat{H}^{\Gamma}_{k,\ell}$ respectively. Define Ω_{G^*}, $\Omega^A_{G^*}$ and $\Omega^B_{G^*}$ similarly for $G^* = \hat{B}^{\Gamma}_{k,\ell}$. By Lemma 48, we have $|\Omega_G| = |\Omega^A_G| + |\Omega^B_G|$, $|\Omega_{G^*}| = |\Omega^A_{G^*}| + |\Omega^B_{G^*}|$ and $|\Omega^B_G| = |\Omega^B_{G^*}|$. Since H is a subgraph of B, we deduce that $\hat{H}_{k,\ell}$ is a subgraph of $\hat{B}_{k,\ell}$ and also G is a subgraph of G^* (by selecting the same ports when constructing G and G^*). Therefore, $\Omega_G \supset \Omega_{G^*}$. It follows that

$$\begin{aligned}\|\mu_G - \mu_{G^*}\|_{\text{TV}} &= \sum_{\sigma:\, \mu_G(\sigma) > \mu_{G^*}(\sigma)} \mu_G(\sigma) - \mu_{G^*}(\sigma) = \sum_{\sigma \in \Omega_G \setminus \Omega_{G^*}} \frac{1}{|\Omega_G|} \\ &= 1 - \frac{|\Omega_{G^*}|}{|\Omega_G|} = 1 - \frac{|\Omega^A_{G^*}| + |\Omega^B_{G^*}|}{|\Omega^A_G| + |\Omega^B_G|} = \frac{|\Omega^A_G| - |\Omega^A_{G^*}|}{|\Omega^A_G| + |\Omega^B_G|}. \end{aligned} \tag{28}$$

If $Z_3(H) < \psi(k,\ell) = (q-3)^{1/k}\, 2^{1+\ell/k}$, then we deduce from Lemma 48 that

$$\|\mu_G - \mu_{G^*}\|_{\text{TV}} \le \frac{|\Omega^A_G|}{|\Omega^B_G|} = \frac{\frac{1}{6}\, q!\, Z_3(H)^k ((q-1)!)^{m^2}}{\frac{1}{4}(q-3)\, q!\, (2^\ell - 2) 2^k ((q-1)!)^{m^2}} \le \frac{2 Z_3(H)^k}{3(q-3) 2^{\ell-1+k}} = \frac{4}{3}\left(\frac{Z_3(H)}{\psi(k,\ell)}\right)^k,$$

where the second inequality uses the fact that $2^\ell - 2 \ge 2^{\ell-1}$ for $\ell \ge 2$. This establishes part 1 of the lemma.

For part 2, if $Z_3(H) \ge \psi(k,\ell)$, then by Lemma 48

$$|\Omega^A_G| = \frac{1}{6}\, q!\, Z_3(H)^k ((q-1)!)^{m^2} \ge \frac{1}{6}(q-3)\, q!\, 2^{\ell+k} ((q-1)!)^{m^2} \ge \frac{2}{3}|\Omega^B_G|.$$

We deduce that

$$\|\mu_G - \mu_{G^*}\|_{\text{TV}} \ge \frac{|\Omega^A_G| - |\Omega^A_{G^*}|}{|\Omega^A_G| + \frac{3}{2}|\Omega^A_G|} = \frac{2}{5}\left(1 - \frac{|\Omega^A_{G^*}|}{|\Omega^A_G|}\right) = \frac{2}{5}\left(1 - \left(\frac{Z_3(B)}{Z_3(H)}\right)^k\right). \quad ■$$

Proof of Lemma 42 By Lemma 47, the number of q-colorings of $G^* = \hat{B}^{\Gamma}_{k,\ell}$ given a phase vector τ is $((q-1)!)^{m^2}$, which is independent of τ. Thus, the phase vector τ of a uniformly random q-coloring of G^* is a uniformly random q-coloring of $\hat{B}_{k,\ell}$. Our algorithm for sampling from the distribution μ_{G^*} then works as follows:

1. Generate a random q-coloring τ of $\hat{B}_{k,\ell}$;

2. For each $v \in \hat{V} = V(\hat{B}_{k,\ell})$, color all ports of the gadget G_v in $\hat{B}^{\Gamma}_{k,\ell}$ with $\tau(v)$, and then color all non-ports of G_v, which are disjoint cliques of size $q-1$, with a random $(q-1)$-coloring using all colors but $\tau(v)$.

To generate a q-coloring of $\hat{B}_{k,\ell}$ uniformly at random, we can proceed as follows:

1. Compute $|\Omega^a|$, $|\Omega^b|$ and $|\Omega| = |\Omega^a| + |\Omega^b|$;
2. With probability $|\Omega^a|/|\Omega|$ generate a random q-coloring from Ω^a;
3. With probability $|\Omega^b|/|\Omega|$ generate a random q-coloring from Ω^b.

To compute $|\Omega^a|$ and $|\Omega^b|$, assume (U, W) is the bipartition of the vertex set of the complete bipartite graph B. Suppose $|U| = N_1$ and $|W| = N_2$. Then we have

$$Z_3(B) = 3 \cdot 2 + 3 \cdot (2^{N_1} - 2) + 3 \cdot (2^{N_2} - 2) = 3\left(2^{N_1} + 2^{N_2} - 2\right).$$

Lemma 46 implies that

$$|\Omega^a| = \frac{1}{6}\, q!\, 3^k \left(2^{N_1} + 2^{N_2} - 2\right)^k \qquad \text{and} \qquad |\Omega^b| = \frac{1}{4}(q-3)\, q!\, (2^\ell - 2)2^k.$$

To generate a coloring from Ω^a, first choose $q-3$ random colors for the complete $(q-3)$-partite graph J and randomly assign one of these colors to each cluster of J. The k copies of B are colored with the remaining 3 colors. Since B is a complete bipartite graph, it is straightforward to generate a random 3-coloring in linear time.

In similar manner, to generate a coloring from Ω^b, first choose $q-2$ colors and color J with these $q-2$ colors. This can be done by first picking a random cluster of J and coloring it with 2 different random colors, and then coloring the other $q-4$ clusters with the remaining $q-4$ colors. Finally color the k copies of B with the 2 colors not used in J.

Since each step of the sampling procedure for $\hat{B}_{k,\ell}$ takes at most linear time, the running time of generating a random q-coloring of $\hat{B}_{k,\ell}$ is $O(kN + \ell(q-3))$. Therefore, the running time of sampling from μ_{G^*} is $O(n)$. ■

7.5. Proof of Theorem 37: the $q = 3$ case

In this section we provide the proof of Theorem 37 for $q = 3$. The proof of this case is very similar to that of $q \geq 4$, but we are required to modify the construction of the testing instance slightly and rederive the results in Lemmas 41 and 42.

Let $H = (V, E)$ be a connected bipartite graph on N vertices for which we want to count the number of 3-colorings. Recall that for $k, \ell \in \mathbb{N}^+$, we define $\hat{H}_{k,\ell}$ to be the graph that contains k copies of H, a complete $(q-3)$-partite graph J with $(q-3)\ell$ vertices, and a complete bipartite graph connecting J and all copies of H. If J is colored by $q-2$ colors, then every copy of H is assigned the remaining 2 colors; on the other hand, if J is colored by $q-3$ colors, then the copies of H are colored with the remaining 3 colors. By checking which of the two types of q-colorings dominates using the Tester, we can obtain a bound on $Z_3(H)$. This approach works only for $q \geq 4$ as the construction of $\hat{H}_{k,\ell}$ (in particular, the complete $(q-3)$-partite graph J) requires $q \geq 4$.

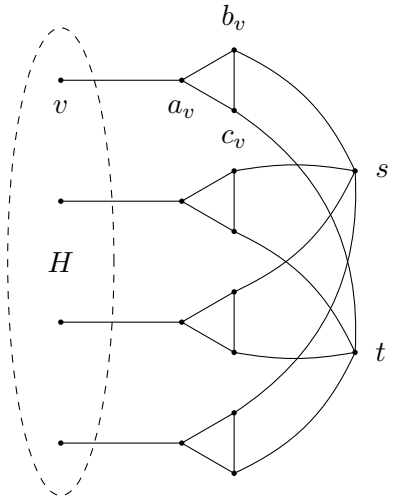

Figure 3: The graph $\tilde{H}$.

For $q = 3$, we need one additional idea. We construct first a graph $\tilde{H}$ which consists of the original graph H, two additional vertices $\{s, t\}$, and several intermediate vertices connecting H and $\{s, t\}$ (see Figure 3). The graph $\tilde{H}$ is constructed in a way such that in every 3-coloring: if s and t receive the same color, then H is colored by exactly two colors; and, if s and t receive two distinct colors, then H can be colored by any proper 3-coloring with equal probability. The problem then reduces to counting the 3-colorings of $\tilde{H}$. We can define a graph $\tilde{H}_{k,\ell}$ using the similar construction as $\hat{H}_{k,\ell}$ for $q \geq 4$, but with two modifications: Firstly, we define J to be an independent set instead of a complete $(q-3)$-partite graph; Secondly, we connect every vertex of J with only the vertices s's and t's in all copies of $\tilde{H}$ instead of all vertices. After constructing the testing instance $\tilde{H}_{k,\ell}$ and $\tilde{H}^{\Gamma}_{k,\ell}$, the proof of Theorem 37 for $q = 3$ follows in the same manner as for $q \geq 4$.

We define next the graph $\tilde{H} = (\tilde{V}, \tilde{E})$, which unlike H is not a bipartite graph.

1. Let s, t be two vertices called *interfaces*;
2. For each $v \in V$, let T_v be a triangle on $\{a_v, b_v, c_v\}$ (clique on 3 vertices);
3. Set $\tilde{V} = V \cup \left(\bigcup_{v \in V} V(T_v)\right) \cup \{s, t\}$;
4. Set $\tilde{E} = E \cup \left(\bigcup_{v \in V} E(T_v)\right) \cup \big\{\{v, a_v\}, \{s, b_v\}, \{t, c_v\} : v \in V\big\}$;

see Figure 3 for an illustration of the graph $\tilde{H}$. Observe that $\tilde{H}$ has $\tilde{N} = 4N + 2$ vertices.

Let $I(\tilde{H}) = \{s, t\}$. For $k, \ell \in \mathbb{N}^+$, we also define the graph $\tilde{H}_{k,\ell} = (V(\tilde{H}_{k,\ell}), E(\tilde{H}_{k,\ell}))$ as follows:

1. Let $\tilde{H}_1, \ldots, \tilde{H}_k$ be k copies of the graph $\tilde{H}$;
2. Let J be an independent set on ℓ vertices;
3. Set $V(\tilde{H}_{k,\ell}) = \left(\bigcup_{i=1}^{k} V(\tilde{H}_i)\right) \cup V(J)$;

4. In addition to the edges in $\tilde{H}_i$ for $1 \le i \le k$, $E(\tilde{H}_{k,\ell})$ also contains edges between the interfaces of $\tilde{H}_i$ for $1 \le i \le k$ and every vertex in J; i.e., for $I(\tilde{H}_i) = \{s_i, t_i\}$ and $v \in J$, we have $\{s_i, v\}, \{t_i, v\} \in E(\tilde{H}_{k,\ell})$.

Finally, we define the graph $\tilde{H}^{\Gamma}_{k,\ell}$ where $\Gamma = \{m, 3, t\}$ in the same way as for $q \ge 4$; namely, we replace every vertex of $\tilde{H}_{k,\ell}$ by a copy of the graph $G(m, 3, t)$ and every edge by an edge between two (unused) ports of the corresponding two gadgets. Furthermore, to make the graph $\tilde{H}^{\Gamma}_{k,\ell}$ well-defined, we set

$$m = k\tilde{N} + \ell = k(4N + 2) + \ell \qquad \text{and} \qquad t = \lceil \sqrt{q} \rceil = 2.$$

Let B be a complete bipartite graph with the same vertex bipartition as H. By setting $H = B$, we also define the graphs $\tilde{B}_{k,\ell}$ and $\tilde{B}^{\Gamma}_{k,\ell}$. Given $k, \ell \in \mathbb{N}^+$, let $G = \tilde{H}^{\Gamma}_{k,\ell}$ and $G^* = \tilde{B}^{\Gamma}_{k,\ell}$. Suppose $d \ge d_c(3) = 4$. Then, Lemma 39 implies that $G, G^* \in \mathcal{M}(n, d)$ for

$$n = (k\tilde{N} + \ell) \cdot 4m = 4m^2.$$

The next two lemmas will play the role of Lemmas 41 and 42 in the proof of Theorem 37 for the case $q = 3$.

Lemma 49 *Let $k, \ell \in \mathbb{N}^+$ with $\ell \ge 2$. Then the following holds:*

1. *If $Z_3(H) < 2^{\ell/k} - 2$, then*

$$\|\mu_G - \mu_{G^*}\|_{\mathrm{TV}} \le 2 \left(\frac{Z_3(H) + 2}{2^{\ell/k}} \right)^k.$$

2. *If $Z_3(H) \ge 2^{\ell/k} - 2$, then*

$$\|\mu_G - \mu_{G^*}\|_{\mathrm{TV}} \ge \frac{1}{2} \left(1 - \left(\frac{Z_3(B) + 2}{Z_3(H) + 2} \right)^k \right).$$

Lemma 50 *There exists an algorithm with running time $O(n)$ that generates a sample from the distribution μ_{G^*}.*

With these two lemmas in hand, the proof of Theorem 37 for $q = 3$ is then identical to that for the $q \ge 4$ case and is thus omitted.

7.5.1. Colorings of G and G^*: Proof of Lemmas 49 and 50

It remains for us to prove Lemmas 49 and 50. First, we establish several facts about the 3-colorings of $G = \tilde{H}^{\Gamma}_{k,\ell}$ and $G^* = \tilde{B}^{\Gamma}_{k,\ell}$. We then use these facts as basis to bound $\|\mu_G - \mu_{G^*}\|_{\mathrm{TV}}$ (Lemma 49) and give a sampling algorithm for μ_{G^*} (Lemma 50). Some of these facts are counterparts of those established in Section 7.4.1 for $q \ge 4$.

Let $Z_3(H)$ denote the number of 3-colorings of H. For $i, j \in \{1, 2, 3\}$, let $Z_3^{i,j}(\tilde{H})$ be the number of 3-colorings of $\tilde{H}$ such that s receives color i and t receives color j.

Lemma 51 *For $i, j \in \{1, 2, 3\}$, $Z_3^{i,i}(\tilde{H}) = 2^{N+1}$ and $Z_3^{i,j}(\tilde{H}) = 2^N Z_3(H)$ when $i \neq j$.*

Proof We first compute $Z_3^{1,1}(\tilde{H})$. Suppose that both s and t are colored with color 1. Then, for each $v \in V$, colors 2 and 3 are both required to color b_v and c_v. As a result, a_v receives color 1 and v can not be colored by 1 for all $v \in V$. Hence, the vertices of H in $\tilde{H}$ can only be assigned colors 2 or 3. There are only two ways to color the connected bipartite graph H with two colors. This gives

$$Z_3^{1,1}(\tilde{H}) = 2 \cdot 2^N = 2^{N+1},$$

and by symmetry $Z_3^{2,2}(\tilde{H}) = Z_3^{3,3}(\tilde{H}) = 2^{N+1}$.

We compute next $Z_3^{1,2}(\tilde{H})$. Let σ be any 3-coloring of H. We claim that there are 2^N colorings of $\tilde{H}$ in which s is assigned color 1, t is assigned color 2, and σ is the coloring in H. From this, it follows immediately that

$$Z_3^{1,2}(\tilde{H}) = 2^N Z_3(H).$$

Let $v \in V$ and consider 3-colorings of the triangle T_v. If $\sigma(v) = 1$, then the only 3-colorings of the triangle (a_v, b_v, c_v) are $(2, 3, 1)$ and $(3, 2, 1)$ since $\{v, a_v\}$, $\{s, b_v\}$ and $\{t, c_v\}$ are all edges of $\tilde{H}$. Similarly, when $\sigma(v) = 2$ or $\sigma(v) = 3$, there are also two possible colorings for (a_v, b_v, c_v) in each case. Therefore, if s and t are assigned colors 1 and 2 respectively, and σ is the coloring in H, there are exactly two proper 3-colorings of T_v for each $v \in V$. As the triangles T_v are disjoint for $v \in V$, once the colors of s, t and H are assigned, there are 2^N proper 3-colorings of $\tilde{H}$. This proves our claim, and by symmetry, for any $i, j \in \{1, 2, 3\}$ with $i \neq j$, we obtain $Z_3^{i,j}(\tilde{H}) = 2^N Z_3(H)$. ■

As in Lemma 46, we can partition the 3-colorings of $\tilde{H}_{k,\ell}$ into two categories.

Lemma 52 *Let $k, \ell \in \mathbb{N}^+$. Let Ω^a and Ω^b be the set of 3-colorings of $\tilde{H}_{k,\ell}$ in which J is colored by exactly 1 and 2 colors respectively. Then $\{\Omega^a, \Omega^b\}$ is a partition for the set of 3-colorings of $\tilde{H}_{k,\ell}$; moreover,*

$$|\Omega^a| = 3 \cdot 2^{k(N+1)}(Z_3(H) + 2)^k \qquad \text{and} \qquad |\Omega^b| = 3 \cdot (2^\ell - 2)2^{k(N+1)}.$$

Proof Observe that in every 3-coloring of $\tilde{H}_{k,\ell}$ the number of colors we can assign to the independent set J is at least one and at most two, since all the vertices of J have at least one common neighbor. It follows immediately that $\{\Omega^a, \Omega^b\}$ is a partition for the set of 3-colorings of $\tilde{H}_{k,\ell}$. For the 3-colorings in Ω^a, we first assign a color to J, say color 1. Then, we count the number of 3-colorings of each $\tilde{H}_i$ whose interfaces $\{s_i, t_i\}$ cannot be assigned color 1. Lemma 51 and symmetry imply

$$|\Omega^a| = 3\left(Z_3^{2,2}(\tilde{H}) + Z_3^{2,3}(\tilde{H}) + Z_3^{3,2}(\tilde{H}) + Z_3^{3,3}(\tilde{H})\right)^k = 3 \cdot 2^{k(N+1)}(Z_3(H) + 2)^k.$$

For 3-colorings in Ω^b, we pick the two colors that color J, say color 1 and 2. Then, the interfaces of $\tilde{H}_i$ have to be assigned color 3 for each i. The number of ways to color J with both colors 1 and 2 is $2^\ell - 2$. Then, by Lemma 51 and symmetry, we get

$$|\Omega^b| = 3 \cdot (2^\ell - 2)Z_3^{3,3}(\tilde{H})^k = 3 \cdot (2^\ell - 2)2^{k(N+1)}.$$ ■

Lemma 53 *Let σ be a 3-coloring of $\tilde{H}^{\Gamma}_{k,\ell}$ and τ be the phase vector of σ. Then, τ is a 3-coloring of $\tilde{H}_{k,\ell}$. Moreover, if τ is a 3-coloring of $\tilde{H}_{k,\ell}$, then there are 2^{m^2} 3-colorings of $\tilde{H}^{\Gamma}_{k,\ell}$ whose phase vector is τ.*

Proof The proof is analogous to that of Lemma 47. ■

Lemma 54 *Let $k, \ell \in \mathbb{N}^+$. Let Ω^A and Ω^B be the set of 3-colorings of $\tilde{H}^{\Gamma}_{k,\ell}$ whose phase vector is a 3-coloring of $\tilde{H}_{k,\ell}$ that belongs to Ω^a and Ω^b respectively. Then $\{\Omega^A, \Omega^B\}$ is a partition for the set of 3-colorings of $\tilde{H}^{\Gamma}_{k,\ell}$; moreover,*

$$|\Omega^A| = |\Omega^a| \cdot 2^{m^2} = 3 \cdot 2^{k(N+1)}(Z_3(H)+2)^k \cdot 2^{m^2}, \text{ and}$$
$$|\Omega^B| = |\Omega^b| \cdot 2^{m^2} = 3 \cdot (2^\ell - 2)2^{k(N+1)} \cdot 2^{m^2}.$$

Proof Follows immediately from Lemmas 52 and 53. ■

Proof of Lemma 49 We use the notation from the proof of Lemma 41. Following the derivation of (28), we get

$$\|\mu_G - \mu_{G^*}\|_{\text{TV}} = \frac{|\Omega^A_G| - |\Omega^A_{G^*}|}{|\Omega^A_G| + |\Omega^B_G|}.$$

If $Z_3(H) < 2^{\ell/k} - 2$, then we deduce from Lemma 54 that

$$\begin{aligned}\|\mu_G - \mu_{G^*}\|_{\text{TV}} &\le \frac{|\Omega^A_G|}{|\Omega^B_G|} = \frac{3 \cdot 2^{k(N+1)}(Z_3(H)+2)^k \cdot 2^{m^2}}{3 \cdot (2^\ell - 2)2^{k(N+1)} \cdot 2^{m^2}} \\ &\le \frac{(Z_3(H)+2)^k}{2^{\ell-1}} = 2\left(\frac{Z_3(H)+2}{2^{\ell/k}}\right)^k.\end{aligned}$$

If $Z_3(H) \ge 2^{\ell/k} - 2$, then by Lemma 54

$$|\Omega^A_G| = 3 \cdot 2^{k(N+1)}(Z_3(H)+2)^k \cdot 2^{m^2} \ge 3 \cdot 2^{k(N+1)}2^\ell \cdot 2^{m^2} \ge |\Omega^B_G|.$$

Thus, we get

$$\|\mu_G - \mu_{G^*}\|_{\text{TV}} \ge \frac{|\Omega^A_G| - |\Omega^A_{G^*}|}{2|\Omega^A_G|} = \frac{1}{2}\left(1 - \frac{|\Omega^A_{G^*}|}{|\Omega^A_G|}\right) = \frac{1}{2}\left(1 - \left(\frac{Z_3(B)+2}{Z_3(H)+2}\right)^k\right).$$ ■

Proof of Lemma 50 This can be done in the same way as the proof of Lemma 42. It suffices to first generate a random 3-coloring τ of $\tilde{B}_{k,\ell}$ and then sample from μ_{G^*} given τ as the phase vector where $G^* = \tilde{B}^{\Gamma}_{k,\ell}$. To sample a random 3-coloring of $\tilde{B}_{k,\ell}$, we do the following:

1. Compute $|\Omega^a|$, $|\Omega^b|$ and $|\Omega| = |\Omega^a| + |\Omega^b|$;
2. With probability $|\Omega^a|/|\Omega|$ generate a random 3-coloring from Ω^a;

3. With probability $|\Omega^b|/|\Omega|$ generate a random 3-coloring from Ω^b.

We can compute $|\Omega^a|$ and $|\Omega^b|$ by Lemma 52. To sample from Ω^a, we first pick one color for J and then color each copy of $\tilde{B}$. Notice that since B is a complete bipartite graph, we can sample a random 3-coloring of B, and consequently $\tilde{B}$, in linear time. To sample from Ω^b, we pick two colors to color J; then in every copy of $\tilde{B}$, the complete bipartite graph B will receive only two colors. The total running time for sampling a random 3-coloring of $\tilde{B}_{k,\ell}$ is $O(k\tilde{N} + \ell)$ and the running time for sampling from μ_{G^*} is $O(n)$. ■

8. Discussion

Our hardness results for identity testing for the Ising model require $|\beta|d \geq c \log n$ 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 $\max\{|\beta|, |\beta^*|\} \cdot d \geq c \log n$. As mentioned, there are polynomial running time algorithms for identity testing when either $|\beta^*|d = O(\log n)$, in which case we can use structure learning methods, or when $|\beta| = O(d^{-1})$ is in the tree uniqueness region, and known sampling methods can be combined with the testing algorithm in (Daskalakis et al., 2018). Therefore, when β is the non-uniqueness region ($|\beta|d < c \log n$) 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 regime of parameters.

Acknowledgments

Research supported in part by NSF grants 1819546, CCF-1850443, CCF-1617306, CCF-1563838 and CCF-1563757.

References

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

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.

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.

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.

A. Bogdanov, E. Mossel, and S. Vadhan. The Complexity of Distinguishing Markov Random Fields. In A. Goel, K. Jansen, J.D.P Rolim, and R. Rubinfeld, editors, *Approximation,*

Randomization and Combinatorial Optimization. Algorithms and Techniques, pages 331–342, Berlin, Heidelberg, 2008. Springer Berlin Heidelberg. ISBN 978-3-540-85363-3.

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.

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.

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

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, 2014b.

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

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.

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.

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.

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.

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.

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

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.

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

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.

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.

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.

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.

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.

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

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.

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.

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

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

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

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, 2016b.

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.

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

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

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

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

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.

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.

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.

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

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

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

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.

K.-I. Ko. Some observations on the probabilistic algorithms and NP-hard problems. *Information Processing Letters*, 14(1):39–43, 1982.

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.

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.

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.

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.

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.

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

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

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.

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.

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.

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.

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

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.

M. Vuffray, S. Misra, and A.Y. Lokhov. Efficient Learning of Discrete Graphical Models. *ArXiv preprint ArXiv:1902.00600*, 2019.

S. Wu, S. Sanghavi, and A.G. Dimakis. Sparse logistic regression learns all discrete pairwise graphical models. In *Advances in Neural Information Processing Systems (NeurIPS)*, pages 8069–8079, 2019.