Journal of Machine Learning Research 22 (2021) 1-56 Submitted 10/20; Revised 5/21; Published 6/21

Hardness of Identity Testing for Restricted Boltzmann Machines and Potts models

Antonio Blanca ABLANCA@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: Nicolas Vayatis

Abstract

We study the identity testing problem for restricted Boltzmann machines (RBMs), and more generally, for undirected graphical models. In this problem, given sample access to the Gibbs distribution corresponding to an unknown or hidden model M^* and given an explicit model M, the goal is to distinguish if either $M = M^*$ or if the models are (statistically) far apart.

We establish the computational hardness of identity testing for RBMs (i.e., mixed Ising models on bipartite graphs), even when there are no latent variables or an external field. Specifically, we show that unless RP = NP, there is no polynomial-time identity testing algorithm for RBMs when $\beta d = \omega(\log n)$, where d is the maximum degree of the visible graph and β is the largest edge weight (in absolute value); when $\beta d = O(\log n)$ there is an efficient identity testing algorithm that utilizes the structure learning algorithm of Klivans and Meka (2017). We prove similar lower bounds for purely ferromagnetic RBMs with inconsistent external fields and for the ferromagnetic Potts model. To prove our results, we introduce a novel methodology to reduce the corresponding approximate counting problem to testing utilizing the phase transition exhibited by these models.

Keywords: distribution testing, identity testing, graphical models, Restricted Boltzmann Machines, Potts model

1. Introduction

For graphical models, there are several fundamental computational tasks which are essential for utilizing these models. These computational problems can be broadly labeled as follows: sampling, counting, structure learning, and testing. Our big picture aim is to understand the relationship between these problems. The specific focus in this paper is on the computational complexity of the *identity testing* problem for *undirected* graphical models and its connections to the hardness of the counting problem.

Identity testing is a basic question in statistics for testing whether a given model fits a dataset. Roughly speaking, given data $\mathcal{D}$ sampled from the posterior or likelihood distribution of an unknown/hidden model M^* and given an explicit model M, can we distinguish whether $M = M^*$?

We study identity testing in the context of undirected graphical models (Murphy, 2012), which correspond to (pairwise) Markov random fields in probability theory and computer vision (Geman and Graffigne, 1986) and to spin systems in statistical physics (Georgii, 2011). We focus attention on examples of graphical models of particular interest: the Ising model, the Potts model, and Restricted Boltzmann Machines. The Ising model is the simplest example of an undirected graphical model, and, in fact, it is one of the most well-studied models in statistical physics where it is used to study phase transitions. The Potts model is the generalization of the Ising model from a two state system to an integer $q \geq 3$ state system. It is also well-studied in statistical physics as the nature of the phase transition changes as q increases (Duminil-Copin et al., 2016, 2017).

Restricted Boltzmann Machines (RBMs) are a simple class of undirected graphical models corresponding to the Ising model on bipartite graphs. Originally introduced by Smolensky in 1986 (Smolensky, 1986), they have played an important role in the history of computational learning theory. They have two layers of variables: one layer corresponding to the observed variables and another layer corresponding to the hidden/latent variables, and no intralayer connections so that the underlying graph is bipartite. Learning was shown to be practical in these restricted models (Hinton, 2002; Hinton et al., 2006) and henceforth played a seminal role in the development of deep learning (Salakhutdinov and Hinton, 2009; Osindero and Hinton, 2008; Salakhutdinov et al., 2007; Hinton and Salakhutdinov, 2009).

We define first the Potts model, as both the Ising model and RBMs may be viewed as special cases of this model. The Potts model is specified by a graph $G = (V, E)$, a set of vertex labels or spins $[q] = \{1, \ldots, q\}$, a set of edge weights defined by $\beta : E \to \mathbb{R}$ and a set of vertex weights $h : V \times [q] \to \mathbb{R}$. Configurations of the Potts model are the collection of vertex labelings $\Omega = \{1, \ldots, q\}^V$. The *Gibbs distribution* associated with the Potts model is a distribution over all configurations $\sigma \in \Omega$ such that:

$$\mu(\sigma) = \mu_{G,\beta,h}(\sigma) := \frac{1}{Z} \exp\left(\sum_{\{u,v\} \in E} \beta(\{u,v\}) \mathbb{1}(\sigma(u) = \sigma(v)) + \sum_{v \in V} h(v, \sigma(v)) \right),$$

where $Z = Z_{G,\beta,h}$ is the normalizing factor or *partition function* given by:

$$Z := \sum_{\sigma \in \Omega} \exp\left(\sum_{\{u,v\} \in E} \beta(\{u,v\}) \mathbb{1}(\sigma(u) = \sigma(v)) + \sum_{v \in V} h(v, \sigma(v)) \right).$$

When $\beta(e) > 0$ for every $e \in E$, the model is called *ferromagnetic* and neighboring vertices prefer to align to the same spin. Conversely, when $\beta(e) < 0$ for every $e \in E$ the model is called *antiferromagnetic*. Models where β is allowed to be both positive or negative for distinct edges are called *mixed* models.

The Ising model corresponds to the special case where there are only two spins; i.e., $q = 2$. RBMs are mixed Ising models restricted to bipartite graphs; that is, G is bipartite with bipartition $V = L \cup R$. Since the focus in this paper is on lower bounds, we often consider the case of no external field ($h = 0$) in order to obtain stronger hardness results.

Given a model specification, that is, a graph $G = (V, E)$, an edge weight function β and an external field h, the goal in the *sampling problem* is to generate samples from the Gibbs distribution $\mu = \mu_{G,\beta,h}$ (or from a distribution close to μ in total variation distance). The corresponding *counting problem* is to compute the partition function $Z = Z_{G,\beta,h}$. The (exact) counting problem is #P-hard (Valiant, 1979) even for restricted classes of graphs (Greenhill, 2000; Vadhan, 2001), and hence the focus on the approximate counting problem of obtaining an FPRAS (fully-polynomial randomized approximation scheme[1]) for Z. For a general class of models, the approximate counting and the approximate sampling problems are equivalent, i.e., there are polynomial-time reductions between them (Jerrum et al., 1986; Štefankovič et al., 2009; Kolmogorov, 2018). A seminal result of Jerrum and Sinclair (Jerrum and Sinclair, 1993) (see also (Randall and Wilson, 1999; Collevecchio et al., 2016; Guo and Jerrum, 2017)) presented an FPRAS for the partition function of the ferromagnetic Ising model.

Another two fundamental problems for undirected graphical models are *structure learning* and *identity testing*. The structure learning problem is as follows: given oracle access to samples from the Gibbs distribution μ_{M^*} for an unknown (i.e., "hidden") model $M^* = (G^*, \beta^*, h^*)$, can we learn G^* (i.e., the structure of the model) in polynomial-time with probability at least $2/3$? In the case of no latent variables (so the samples from the Gibbs distribution reveal the label of all vertices V of G) recent work of Klivans and Meka (Klivans and Meka, 2017) (see also (Bresler, 2015; Vuffray et al., 2019; Hamilton et al., 2017; Vuffray et al., 2016; Wu et al., 2019)) learns n-vertex graphs with $O(\log n) \times \exp(O(\beta d))$ samples and $O(n^2 \log n) \times \exp(O(\beta d))$ time where d is the maximum degree of G and $\beta := \max_{e \in E} |\beta(e)|$ is the maximum edge weight in absolute value; this bound has nearly-optimal sample complexity from an information-theory perspective (Santhanam and Wainwright, 2012).

For RBMs with latent variables (thus samples only reveal the labels for vertices on one side R), structure learning can be done in time $O(n^{d_L+1})$ where d_L is the maximum degree of the latent variables (Bresler et al., 2013; Klivans and Meka, 2017; Bresler et al., 2019). Recent work of Bresler, Koehler and Moitra (Bresler et al., 2019) proves that there is no algorithm with running time $n^{o(d_L)}$ assuming k-sparse noisy parity on n bits is hard to learn in time $n^{o(k)}$; they also show that for the special case of ferromagnetic RBMs with hidden variables there is a structure learning algorithm with $O(\log n) \times \exp(O(\beta d^2))$ sample complexity and $O(n^2 \log n) \times \exp(O(\beta d^2))$ running time; see also (Bresler and Buhai, 2020; Goel, 2020).

In the *identity testing* problem we are given oracle access to samples from the Gibbs distribution μ_{M^*} for an unknown model $M^* = (G^*, \beta^*, h^*)$ (as in structure learning) and we are also given an explicit model $M = (G, \beta, h)$. Our goal is to determine, with probability $\geq 2/3$, if either $M = M^*$ or if the models are $(1-\varepsilon)$-far apart; specifically, if the total variation distance between their Gibbs distributions is at least $1 - \varepsilon$ for a given $\varepsilon > 0$. (We note that previous works assumed separation $\geq \varepsilon$ in the later case, whereas we prove hardness even when we assume separation $\geq 1 - \varepsilon$.)

It is known that identity testing cannot be solved in polynomial time for general graphical models in the presence of hidden variables unless RP = NP (Bogdanov et al., 2008) . In this paper we assume there are no hidden variables and hence the samples from μ_{M^*} reveal the label of

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 $\log(\delta^{-1})$.

every vertex in the graph G; this setting is more interesting for hardness results. We explore a more refined picture of hardness of identity testing vs. polynomial-time algorithms.

It is known that identity testing can be reduced to sampling (Daskalakis et al., 2018) or structure learning (Bezáková et al., 2020): given an efficient algorithm for the associated sampling problem or an efficient algorithm for structure learning, then one can efficiently solve the identity testing problem. Hence, identity testing is (computationally) easier than sampling and structure learning. (To be precise, one needs to solve both the structure learning and the parameter estimation problems to solve identity testing; the algorithm of Klivans and Meka (Klivans and Meka, 2017) does in fact provide this.) This raises the question of whether identity testing can be efficiently solved in cases where sampling and structure learning are known to be hard. We prove (for the models studied here) that when sampling and structure learning are hard, then identity testing is also hard.

1.1 Our results

The ε-identity testing problem for the Ising and Potts models is formally defined as follows. For positive integers n and d, and positive real numbers β and h, let $\mathcal{M}_{\text{RBM}}(n, d, \beta, h)$ denote the family of RBMs on n-vertex bipartite graphs $G = (V, E)$ of maximum degree at most d, where the absolute value of all edge interactions is at most β and the field $|h(v, i)| \leq h$ for all $v \in V$ and $i \in [q]$; see Definition 4. We define $\mathcal{M}_{\text{POTTS}}(n, d, \beta, h)$ analogously for the family of Potts models, without the restriction of G being bipartite.

> Given an RBM $M \in \mathcal{M}_{\text{RBM}}(n, d, \beta, h)$, and sample access to a distribution μ_{M^*} for an unknown RBM $M^* \in \mathcal{M}_{\text{RBM}}(n, d, \beta, h)$, distinguish with probability at least $3/4$ between the cases:
>
> $$1.\ \mu_M = \mu_{M^*}; \qquad 2.\ \|\mu_M - \mu_{M^*}\|_{\text{TV}} \geq 1 - \varepsilon.$$

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 ε-identity testing problem for the Potts model is defined in the same manner, but assuming that both M and M^* belong to $\mathcal{M}_{\text{POTTS}}(n, d, \beta, h)$ instead.

Our first result concerns the identity testing problem on $\mathcal{M}_{\text{RBM}}(n, d, \beta, 0)$; that is, RBMs with both positive and negative edge weights (i.e., mixed RBMs) without external fields (i.e., $h(v, i) = 0$ for all $v \in V$, $i \in [q]$). We show that for RBMs the approach utilizing structure learning is essentially best possible. In particular we prove that when $\beta d = \omega(\log n)$ there is no poly-time identity testing algorithm, unless RP $=$ NP. Note that when $\beta d = O(\log n)$, the algorithm of Klivans and Meka (Klivans and Meka, 2017) for structure learning and parameter estimation provides an identity testing algorithm with $\text{poly}(n)$ sample complexity and running time.

Theorem 1 *Suppose n, d are positive integers such that $3 \leq d \leq n^\theta$ for constant $\theta \in (0, 1)$ and let $\varepsilon \in (0, 1)$. If RP $\neq$ NP, then for all real $\beta > 0$ satisfying $\beta d = \omega(\log n)$ there is no polynomial running time algorithm to solve the ε-identity testing problem for the class $\mathcal{M}_{\text{RBM}}(n, d, \beta, 0)$ of mixed RBMs without external fields.*

We note that the sample complexity of identity testing on $\mathcal{M}_{\text{RBM}}(n, d, \beta, 0)$, and more generally for any family of Ising models, is $\text{poly}(n, d, \beta)$ (Daskalakis et al., 2018); the above result

establishes the computational hardness of the problem on $\mathcal{M}_{\text{RBM}}(n, d, \beta, 0)$. Moreover, in contrast to Theorem 1, Daskalakis, Dikkala and Kamath (Daskalakis et al., 2018) provided a poly-time identity testing algorithm for all *ferromagnetic* Ising model with *consistent* fields (the external field is consistent if it only favors the same unique spin at every vertex; otherwise it is called inconsistent; see Definition 21). Their algorithm crucially utilizes the known poly-time sampling methods for the ferromagnetic Ising model (Jerrum and Sinclair, 1993; Randall and Wilson, 1999; Collevecchio et al., 2016; Guo and Jerrum, 2017). On the hardness side, super-polynomial lower bounds were recently established for identity testing for the *antiferromagnetic* Ising model on general (not necessarily bipartite) graphs when $\beta d = \omega(\log n)$ (Bezáková et al., 2020). This previous result utilizes the hardness of the maximum cut problem, since maximum cuts correspond to the "ground states" (maximum likelihood configurations) of the antiferromagnetic model; this is not the case for RBMs, and new insights are required (see Section 1.2 for a more detailed discussion). In particular we show a new approach to reduce from the counting problem.

Ferromagnetic and antiferromagnetic RBMs are equivalent models; that is, there is a one-to-one correspondence between configurations with the same weight. Therefore, the results established in (Daskalakis et al., 2018) solve the identity testing problem for both ferromagnetic and antiferromagnetic RBMs with no latent variables, even in the presence of a consistent external field. Moreover, Klivans and Meka's algorithm from (Klivans and Meka, 2017) together with the hardness results of Theorem 1 provides a fairly complete picture of the computational complexity of identity testing for (mixed) RBMs with no external field ($h = 0$).

Our next result concerns the hardness of identity testing for purely *ferromagnetic* RBMs with an *inconsistent* magnetic field; that is, a field that favors one spin for some of the vertices and the other spin for the rest; see Definition 21. 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 (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). Let $\mathcal{M}^{+}_{\text{RBM}}(n, d, \beta, h)$ be set of all ferromagnetic RBMs in $\mathcal{M}_{\text{RBM}}(n, d, \beta, h)$.

Theorem 2 *Suppose n, d are positive integers such that $3 \leq d \leq n^{\theta}$ for constant $\theta \in (0, 1)$ and let $\varepsilon \in (0, 1)$. If #BIS does not admit an* FPRAS*, there exists $h = O(1)$ such that when $\beta d = \omega(\log n)$ there is no polynomial running time algorithm that solves the ε-identity testing problem for the class $\mathcal{M}^{+}_{\text{RBM}}(n, d, \beta, h)$ of* ferrromagnetic *RBMs with* inconsistent *external fields.*

Given the efficient identity testing algorithm for ferromagnetic Ising models (Daskalakis et al., 2018; Jerrum and Sinclair, 1993), we may ask whether there are other (ferromagnetic) models that allow efficient testing algorithms. A prime candidate is the ferromagnetic Potts model. Both the ferromagnetic Ising and Potts models have a rich structure; for instance, their random-cluster representation (Grimmett, 2006) enables sophisticated (and widely-used) sampling algorithms such as the Swendsen-Wang algorithm (Swendsen and Wang, 1987). However, while there are efficient samplers for the ferromagnetic Ising model for all graphs G and all edge interactions β (Jerrum and Sinclair, 1993; Collevecchio et al., 2016; Guo and Jerrum, 2017), the case of the ferromagnetic Potts model (i.e., $q > 2$ spins) looks less promising. In fact, it is unlikely that there is an efficient sampling/counting algorithm for general ferromagnetic Potts models since this is a known #BIS-hard problem (Goldberg and Jerrum, 2012; Galanis et al., 2016b); this is due to a phenomena called *phase co-existence*, which we will also exploit; see Section 2.2.1. Given the weaker hardness of

sampling and approximate counting for the ferromagnetic Potts model, the hardness of the identity problem was less clear.

We prove that identity testing for the ferromagnetic Potts model is in fact hard in the same regime of parameters where sampling and structure learning are known to be hard. Specifically, we observe that the structure learning algorithm from (Klivans and Meka, 2017) applies to the Potts model, and hence implies a testing algorithm when $\beta d = O(\log n)$; we establish lower bounds when $\beta d = \omega(\log n)$ that hold even for the simpler case of models with no external field.

Theorem 3 *Suppose n, d, $q \geq 3$ are positive integers such that $3 \leq d \leq n^{\theta}$ for constant $\theta \in (0,1)$ and let $\varepsilon \in (0,1)$. If #BIS does not admit an* FPRAS*, then there is no polynomial running time algorithm that solves the ε-identity testing problem for the class $\mathcal{M}^{+}_{\text{Potts}}(n,d,\beta,0)$ of ferromagnetic q-state Potts models without an external field. Moreover, our lower bound applies restricted to the class of ferromagnetic Potts models on bipartite graphs in $\mathcal{M}^{+}_{\text{Potts}}(n,d,\beta,0)$.*

1.2 Our techniques

Our proof is a general approach that allows us to obtain hardness results for several models of interest. Specifically, we devise a methodology to reduce the problem of approximate counting (i.e., approximating partition functions) to identity testing. For this we consider a decision version of approximate counting and prove that this variant is as hard as the standard approximation problem; this first step of our reduction applies to many other models of interest (see Theorem 7 and Section 6).

In the second step of our reduction, given a hard counting instance, we use insights about the phase transition of the models to construct a testing instance whose output allows us to solve the decision version of approximate counting. The actual reduction is generic (see Theorem 14), but the insights about each model are needed to build a suitable testing instance; this construction is the only part of our proof that is model specific, whereas every other step in the proof applies to more general spin systems. Our approach is nicely illustrated in the context of the ferromagnetic Potts model; that is, in the proof of Theorem 3 in Section 2. There, we utilize the phase transition phenomenon in the associated mean-field Potts model which corresponds to the complete graph. In particular, there is a phase co-existence corresponding to a first-order phase transition which we utilize to approximate the partition function of the input graph; see Section 2.

In the third and final step of the reduction, we reduce the maximum degree of the graph in the testing instance by using random bipartite graphs as gadgets, as has been done in seminal hardness results for approximate counting (Sly, 2010; Sly and Sun, 2012), and more recently in (Bezáková et al., 2020) for the hardness of testing for the antiferromagnetic Ising model. This step is also generic and applies to a large class of models; see Section 5 and specifically Theorem 28. One interesting implication of our approach is that our gadget and reduction yields always bipartite graphs, and hence we immediately get hardness results for bipartite graphs for all of the models studied in this paper.

We pause to briefly contrast the above proof approach with that in (Bezáková et al., 2020), where it was established hardness of identity testing for the antiferromagnetic Ising model. As mentioned earlier, in the antiferromagnetic Ising model, the configurations with the highest weight or likelihood (i.e., the ground states) correspond to the maximum cuts of the original graph. Hence, it is natural to prove hardness of identity testing for the antiferromagnetic Ising model using a reduction from the maximum cut problem. The ground states of ferromagnetic systems, on the

other hand, correspond to the monochromatic configurations, so there is no hard optimization problem in the background to utilize in the reduction. (The similar obstacle for RBMs is that the maximum cut problem is trivial in bipartite graphs, so we cannot hope to use it to prove hardness.) We use the hardness of approximating the partition function instead, and consequently our reduction is of a completely different flavor to that in (Bezáková et al., 2020); we utilize the unique nature of the phase transition in these models in an essential way. We also mention that, using significantly different reductions, the hardness of approximating partition functions has also been employed for proving the hardness of the parameter estimation problem mentioned earlier (Montanari, 2015; Bresler et al., 2014).

To reduce the degree of the graphs in our construction we do utilize insights and certain technical lemmas from (Bezáková et al., 2020). Specifically, those concerning the expansion of random near-regular bipartite graphs. We note that the models we consider on these random graphs are different than those in (Bezáková et al., 2020); in particular, we consider mixed models and allowed external fields, whereas in (Bezáková et al., 2020) these gadgets are purely antiferromagnetic and there is no external field.

We present our proof approach in the context of the ferromagnetic Potts model first, specifically in Section 2 we prove Theorem 3. The proofs for RBMs, namely Theorems 1 and 2 which follow the same approach, are provided in Sections 3 and 4, respectively.

2. Testing ferromagnetic Potts models

In this section we prove Theorem 3, our lower bound for identity testing for the ferromagnetic Potts model. To prove this theorem, we introduce a new methodology to reduce approximate counting (i.e., the problem of finding an FPRAS for the partition function of a model), to identity testing. We later use this framework to establish our lower bounds for identity testing for RBMs (i.e., Theorems 1 and 2); we believe our methods could be used to establish the hardness of identity testing for other spin systems.

We introduce some useful notation next. Recall that in the introduction we define the families of models $\mathcal{M}_{\text{RBM}}$, $\mathcal{M}^{+}_{\text{RBM}}$, $\mathcal{M}_{\text{Potts}}$ and $\mathcal{M}^{+}_{\text{Potts}}$. We formalize and extend this notation as follows.

Definition 4 *For integers $n, d \geq 3$ and $\beta, h \in \mathbb{R}$, let $\mathcal{M}_{\text{Potts}}(n, d, \beta, h, q)$ denote the family of q-state Potts models on n-vertex graphs $G = (V_G, E_G)$ of maximum degree at most d with edge interactions and external field given by $\beta_G : E_G \to \mathbb{R}$ and $h_G : V_G \times [q] \to \mathbb{R}$, respectively, such that:*

(i) for every edge $\{u, v\} \in E_G$, $|\beta_G(\{u, v\})| \leq \beta$; and

(ii) for every vertex $v \in V_G$ and spin $i \in [q]$, $|h_G(v, i)| \leq h$.

Remark 5 *We omit q from the notation above as it is usually clear from context. For the special case of $q = 2$, i.e., the Ising model, we use $\mathcal{M}_{\text{Ising}}$; when $q = 2$ and the underlying graph is bipartite we use $\mathcal{M}_{\text{RBM}}$. In addition, we add "+" or "−" as a superscript to the notation to denote the corresponding ferromagnetic or antiferromagnetic subfamilies; e.g., $\mathcal{M}^{+}_{\text{Potts}}(n, d, \beta, h)$ denotes the subset of ferromagnetic Potts models in $\mathcal{M}_{\text{Potts}}(n, d, \beta, h)$. Finally, we add a circumflex, e.g., $\hat{\mathcal{M}}^{+}_{\text{Potts}}(n, d, \beta, h)$, for the subfamily of models where every edge weight is* exactly *equal to β.*

2.1 Step 1: Decision version of approximate counting

Our starting point is always a known hard approximate counting instance. For the ferromagnetic Potts model, we consider the problem of approximating its partition function on a graph G. As mentioned in the introduction, this problem is known to be #BIS-hard, even under the additional assumptions that all edges have the same interaction parameter $0 < \beta_G = \Theta(1)$ and that there is no external field (i.e., $h = 0$) (Goldberg and Jerrum, 2012; Galanis et al., 2016b). Our goal is to design an FPRAS for the partition function $Z_{G,\beta_G} := Z_{G,\beta_G,0}$ using a polynomial-time algorithm for identity testing, thus establishing the #BIS-hardness of this problem.

Our first step is to reduce the problem of approximating Z_{G,β_G} to a natural decision variant of the problem. This decision version will be more naturally solved by the testing algorithm and is more generally defined as follows:

Definition 6 (Decision r-approximate counting) *Given a Potts model (G,β_G,h_G), an approximation ratio $r > 1$ and an input $\hat{Z} \in \mathbb{R}$, distinguish with probability at least $5/8$ between the following two cases:*

$$\textit{(i)}\ Z_{G,\beta_G,h_G} \leq \frac{1}{r}\hat{Z} \qquad \textit{(ii)}\ Z_{G,\beta_G,h_G} \geq r\hat{Z}$$

We show that the decision version of approximate counting is as hard as the standard problem of approximating Z_{G,β_G,h_G}.

Theorem 7 *Let $n, d \geq 1$ be integers and let $\beta, h \geq 0$ be real numbers. Suppose that there is no* FPRAS *for the counting problem for a family of Potts models $\mathcal{M}$, where*

$$\mathcal{M} \in \{\hat{\mathcal{M}}^{+}_{\text{Potts}}(n, d, \beta, h), \hat{\mathcal{M}}^{-}_{\text{Ising}}(n, d, \beta, h), \hat{\mathcal{M}}^{+}_{\text{Ising}}(n, d, \beta, h)\}.$$

Then, for any $c > 0$ there is no polynomial-time algorithm for the decision version of n^c-approximate counting for $\mathcal{M}$.

Our proof of this theorem is provided in Section 6.

2.2 Step 2: Testing instance construction

We first construct a hard instance for the identity testing problem for the ferromagnetic Potts model on general graphs, with no restriction on the maximum degree and with a constant upper bound on the edge interactions. We prove first that identity testing is #BIS-hard in this setting.

Theorem 8 *Consider a ferromagnetic Potts model with no external field ($h = 0$) where the interaction on every edge is ferromagnetic and bounded from above by a constant $\beta_0 > 0$. Then, there is no polynomial-time identity testing algorithm for the model unless there is an* FPRAS *for #BIS.*

To establish this theorem, we construct an identity testing instance that allows us to solve the decision variant of approximate counting (see Definition 6). We note that this theorem does not immediately imply Theorem 3 from the introduction because we allow the degree to be unbounded;

specifically, Theorem 8 establishes hardness for $\mathcal{M}^{+}_{\text{Potts}}(n, n, \beta, 0)$. The next step of the proof uses this result and a degree-reducing gadget to establish Theorem 3 (see Section 2.3). Our main gadget in the proof of Theorem 8 will be a complete graph H on m vertices; this is known as the *mean-field* case in statistical physics.

2.2.1 The ferromagnetic mean-field q-state Potts model

Let $H = K_m$ be a complete graph on m vertices and let β_H be the interaction strength on the edges of H. By symmetry, the q-state Potts configurations on a complete graph can be described by their "signature"—by "signature" we mean the vector $(\sigma_1, \ldots, \sigma_q) \in \mathbb{Z}^q$ where $\sigma_i \geq 0$ is the number of vertices that have spin i; note that $\sum_{i=1}^{q} \sigma_i = m$.

In the complete graph, the ferromagnetic Potts model is known to undergo an "order-disorder" phase transition. Specifically, there exists a critical value $\beta_H = \mathfrak{B}_o/m$ such that when $\beta_H < \mathfrak{B}_o/m$, long-range correlations do not exist; the system is then said to be in a "disordered" state as the typical configurations have signature $\approx (m/q, \ldots, m/q)$ where each spin has roughly the same density (up to lower order terms). In contrast, when $\beta_H > \mathfrak{B}_o/m$, typical configurations have a dominant spin and the remaining spins are uniformly distributed. These configurations are thus referred to as "majority" configurations. More precisely there exists a constant $\alpha = \alpha(\beta_H) > 1/q$ and, with high probability, configurations from the Gibbs distribution have signature $\approx \left(\alpha m, \frac{(1-\alpha)m}{q-1}, \ldots, \frac{(1-\alpha)m}{q-1}\right)$ up to permutations and lower order terms.

When $q \geq 3$, the phase transition is known to be of *first-order*, which means that when $\beta_H = \mathfrak{B}_o/m$, a sample from the Potts distribution may have signature $\approx (m/q, \ldots, m/q)$ with constant probability, or signature $\approx \left(\alpha m, \frac{(1-\alpha)m}{q-1}, \ldots, \frac{(1-\alpha)m}{q-1}\right)$ (up to permutations) also with constant probability. This phenomena is referred to as *phase co-existence*, and it is known (or conjectured) to be present in a variety of graphs, being the root reason for the hardness of sampling and counting for the ferromagnetic Potts model. In contrast, in the Ising model (i.e., when $q = 2$), there is no phase co-existence; in this case, the majority density $\alpha(\mathfrak{B}_o/m)$ is $1/q$ and the two phases—disordered and majority—coincide at the critical point.

We now formalize the notion of the majority phase M, the disordered phase D, and the remaining configurations S with their corresponding partition functions Z_H^{M}, Z_H^{D}, and Z_H^{S}. The majority phase is defined with respect to a fixed constant $\hat{\alpha} = \hat{\alpha}(\mathfrak{B}_o)$ which is the density of the dominant color at the phase coexistence point $\mathfrak{B}_o/m$. Let Ω_H denote the set of Potts configurations on H and for $\sigma \in \Omega_H$, let $(\sigma_1, \ldots, \sigma_q) \in \mathbb{Z}^q$ denote its signature. Consider the following sets:

$$M := \left\{ \sigma \in \Omega_H \;\middle|\; \exists j \in [q] : \; |\sigma_j - \hat{\alpha} m| \leq m^{3/4} \text{ and } \left|\sigma_i - \frac{1-\hat{\alpha}}{q-1} m\right| \leq m^{3/4} \text{ for } i \in [q] \setminus \{j\} \right\},$$

$$D := \left\{ \sigma \in \Omega_H \;\middle|\; \forall i \in [q] : \; |\sigma_i - m/q| \leq m^{3/4} \right\},$$

and $S := \Omega_H \setminus (M \cup D)$.

For a configuration σ on the complete graph $H = (E_H, V_H)$, let

$$w_H^{\sigma}(\beta_H) = \exp\left(\sum_{\{u,v\} \in E(H)} \beta_H \mathbb{1}(\sigma(u) = \sigma(v)) \right)$$

denote the weight of σ in the mean-field model (H, β_H). Consider the contributions of each type of configuration to the partition function. That is,

$$Z_H^{\mathrm{M}}(\beta_H) := \sum_{\sigma \in M} w_H^{\sigma}(\beta_H), \qquad Z_H^{\mathrm{D}}(\beta_H) := \sum_{\sigma \in D} w_H^{\sigma}(\beta_H), \qquad Z_H^{\mathrm{S}}(\beta_H) := \sum_{\sigma \in S} w_H^{\sigma}(\beta_H).$$

Hence, the partition function of (H, β_H) is given by $Z_H(\beta_H) = Z_H^{\mathrm{M}}(\beta_H) + Z_H^{\mathrm{D}}(\beta_H) + Z_H^{\mathrm{S}}(\beta_H)$. We note that in our reduction later, we will choose a specific $\beta_H > 0$ depending on the instance of the approximate counting problem and the parameters of the identity testing algorithm; hence, to emphasize the effect of β_H, we parameterize Z_H^{M} (and other functions in this section) in terms of β_H.

The following two lemmas detail the relevant behavior of the mean-field Potts model at and around the critical point $\mathfrak{B}_o/m$. We note that as a consequence of the first-order phase transition, there is a critical window around $\mathfrak{B}_o/m$ where the non-dominant phase (i.e., disorder or majority) is still much more likely than any other type configurations; this phenomena is known as *metastability* and will also be crucial for us.

First we establish that in the critical window around $\mathfrak{B}_o/m$ the majority M and disordered D configurations are exponentially more likely than the remaining configurations S. Several variants of this result have been proved in some fashion before (see, e.g., Bollobás et al., 1996; Luczak and Łuczak, 2006; Goldberg and Jerrum, 2012; Cuff et al., 2012; Gheissari et al., 2018; Galanis et al., 2015; Blanca and Sinclair, 2015). However, the precise bound we require in our proofs does not seem to be available in the literature.

Lemma 9 *There exists constants $c, c' > 0$ such that for any β_H satisfying $|\beta_H - \mathfrak{B}_o/m| \le c' m^{-3/2}$ we have*

$$Z_H^{\mathrm{S}}(\beta_H) \le \min\{Z_H^{\mathrm{M}}(\beta_H), Z_H^{\mathrm{D}}(\beta_H)\} \exp(-c\sqrt{m}).$$

In addition, we show that we can find in $\mathrm{poly}(m)$ time a value for the parameter β_H in the critical window to achieve a specified ratio R of the majority partition function $Z_H^{\mathrm{M}}(\beta_H)$ to the disordered partition function $Z_H^{\mathrm{D}}(\beta_H)$.

Lemma 10 *There exist constants $c, c' > 0$ such that for any $R \in [e^{-c\sqrt{m}}, e^{c\sqrt{m}}]$ and any constant $\delta \in (0, 1)$, we can efficiently find $\beta_H > 0$ in $\mathrm{poly}(m)$ time such that $|\beta_H - \mathfrak{B}_o/m| \le c' m^{-3/2}$ and*

$$(1 - \delta)R \le \frac{Z_H^{\mathrm{M}}(\beta_H)}{Z_H^{\mathrm{D}}(\beta_H)} \le R. \tag{1}$$

The proof of these two lemmas are provided in Appendix A.

2.2.2 Identity testing reduction

Visible Model Construction. Let (G, β_G) be the instance of the ferromagnetic Potts model with no external field (i.e., $h = 0$) for which we are trying to approximate the partition function Z_{G,β_G}; we shall assume $G = (V_G, E_G)$ is an N-vertex graph and that every edge has interaction strength $0 < \beta_G = \Theta(1)$. Let $H = (V_H, E_H)$ be a complete graph on $m = N^{10}$ vertices. The graph $F = (V_F, E_F)$ is the result of connecting the vertices of H and G with a complete bipartite graph

$K_{m,N}$ with edges $E_{m,N}$; that is, $V_F = V_G \cup V_H$ and $E_F = E_H \cup E_G \cup E_{m,N}$. We consider the Potts model on the graph F with edge interactions $\beta_F : E_F \to \mathbb{R}$ given by:

$$\beta_F(e) = \begin{cases} \beta_H & \text{if } e \in E_H \\ \beta_G & \text{if } e \in E_G \\ \beta & \text{if } e \in E_{m,N}, \end{cases}$$

where $\beta_H, \beta > 0$ will be chosen later. We use $n := N + m$ for the number of vertices of F, and, with a slight abuse of notation, we use F for the Potts model (F, β_F) which will play the role of the visible model in our reduction; μ_F denotes the corresponding Gibbs distribution.

We study first the properties of "typical" configurations on G conditional on a configuration σ on the complete graph H. For this, we introduce some additional notation. Let Ω_F, Ω_H and Ω_G be the set of Potts configuration on the graph F, H and G respectively; note that $\Omega_F = \Omega_H \times \Omega_G$. For $\sigma \in \Omega_H$, define

$$Z_F^\sigma(\beta_H) := \sum_{\eta \in \Omega_F : \eta(V_H) = \sigma} w_F^\eta(\beta_H)$$

where the weight $w_F^\eta(\beta_H)$ of configuration η is given by

$$w_F^\eta(\beta_H) = \exp\left(\sum_{\{u,v\} \in E_F} \beta_F(\{u,v\}) \mathbb{1}(\eta(u) = \eta(v)) \right);$$

that is, $Z_F^\sigma(\beta_H)$ is the total contribution to the partition $Z_F(\beta_H)$ of F of the configurations that agree with σ on H.

If we fix a configuration σ on H and look at the configuration on G (under the Gibbs distribution on F conditional on σ) then σ will act as an external field on the vertices of G. We show that if σ is in the majority phase (i.e., in the set M), then the configuration on G will be monochromatic with high probability as these configurations will maximize the number of monochromatic edges between G and H. In contrast, when σ is in the disordered phase (i.e., in D), then every configuration on G will have (roughly) the same number of monochromatic edges between G and H; hence, the partition function $Z_F^\sigma(\beta_H)$ in this case will be proportional to Z_{G,β_G}.

To formalize this, we split the partition function of F into three parts depending on the signature on the complete graph H. Let

$$Z_F^{\mathrm{M}}(\beta_H) = \sum_{\sigma \in M} Z_F^\sigma(\beta_H), \quad Z_F^{\mathrm{D}}(\beta_H) = \sum_{\sigma \in D} Z_F^\sigma(\beta_H), \quad \text{and } Z_F^{\mathrm{S}}(\beta_H) = \sum_{\sigma \in S} Z_F^\sigma(\beta_H);$$

then, $Z_F(\beta_H) = Z_F^{\mathrm{M}}(\beta_H) + Z_F^{\mathrm{D}}(\beta_H) + Z_F^{\mathrm{S}}(\beta_H)$.

The following lemma details the above description of the properties of configurations on the original instance G conditional on the configuration on the complete graph H.

Lemma 11 *For any constants $\delta \in (0,1)$ and $c > 0$, and any β_H such that $|\beta_H - \mathfrak{B}_o/m| \le c m^{-3/2}$, there exists constants $c_1, c_2 > 0$ such that for any $\beta \in \left[\frac{c_1 N}{m}, \frac{c_2}{N m^{3/4}}\right]$:*

1. *When the configuration on H is in the majority phase, the configuration on G is likely to be monochromatic; more precisely,*

$$e^{-\delta} \cdot Z_H^{\mathrm{M}} \cdot \exp\left(\hat{\alpha}\beta N m + \beta_G |E_G|\right) \le Z_F^{\mathrm{M}}(\beta_H) \le e^{\delta} \cdot Z_H^{\mathrm{M}} \cdot \exp\left(\hat{\alpha}\beta N m + \beta_G |E_G|\right). \quad (2)$$

2. *When the configuration on H is in the disordered phase, the configuration on G will have very limited influence from the configuration on H; more precisely,*

$$e^{-\delta} \cdot Z_H^{\mathrm{D}} \cdot Z_G \cdot \exp\left(\beta N m/q\right) \leq Z_F^{\mathrm{D}}(\beta_H) \leq e^{\delta} \cdot Z_H^{\mathrm{D}} \cdot Z_G \cdot \exp\left(\beta N m/q\right). \tag{3}$$

3. *The remaining configurations on H have a small contribution to the partition function of the model F; more precisely,*

$$Z_F^{\mathrm{S}}(\beta_H) \leq Z_F(\beta_H) \exp\left(-\Omega(\sqrt{m})\right). \tag{4}$$

We remark that the factors $\exp(\hat{\alpha}\beta N m + \beta_G|E_G|)$ and $\exp(\beta N m/q)$ in (2) and (3), respectively, account for the contribution of all the monochromatic edges in G and between G and H in each case.

Proof of Lemma 11 We fix β_H and, for ease of notation, we drop the dependence on β_H throughout the proof; i.e., $Z_F^{\mathrm{M}}(\beta_H)$ becomes Z_F^{M}, $w_H^{\sigma}(\beta_H)$ becomes $w_H(\sigma)$ for $\sigma \in \Omega_H$ and $w_F^{\eta}(\beta_H)$ becomes $w_F(\eta)$ for $\eta \in \Omega_F$.

Let $\sigma \in \Omega_H$ and $\tau \in \Omega_G$. When computing weight for configuration $\sigma \cup \tau$ (i.e., the configuration of F that results from combining the spins assignment of σ and τ in H and G, respectively, it will be convenient to separate the interaction of edges in H (that captures the phase coexistence in the mean-field model) and the interaction in G and between H and G (that captures the effect of different phases on G). Thus, let

$$w_{F\setminus H}(\sigma \cup \tau) := \frac{w_F(\sigma \cup \tau)}{w_H(\sigma)}.$$

Then,

$$Z_F^{\mathrm{M}} = \sum_{\sigma \in M} \sum_{\tau \in \Omega_G} w_H(\sigma) w_{F\setminus H}(\sigma \cup \tau).$$

For $\sigma \in M$, let $(\sigma_1, \ldots, \sigma_q) \in \mathbb{Z}^q$ be its signature; suppose w.l.o.g. that σ_1 is such that $|\sigma_1 - \hat{\alpha} m| \leq m^{3/4}$ and $|\sigma_i - \frac{1-\hat{\alpha}}{q-1} m| \leq m^{3/4}$ for all $i \in \{2, \ldots, q\}$. Consider the configuration η_1 on G that assigns spin 1 to every vertex of G and let $a = \max_{i \in \{2,\ldots,q\}} \sigma_i$. For any other configuration $\tau \neq \eta_1$ on G with $t \geq 1$ vertices not assigned spin 1, we have that

$$w_{F\setminus H}(\sigma \cup \tau) \leq \exp\left(\beta_G|E_G| + \beta(\sigma_1(N-t) + at)\right) \leq \exp\left(\beta_G|E_G| + \beta(\sigma_1(N-1) + a)\right), \tag{5}$$

since there are at most $|E_G|$ monochromatic edges in G and at least one vertex in G has a vertex assigned a spin different from 1 (thus there are at most $\sigma_1(N-1) + a$ monochromatic edges between G and H). Hence, we get

$$\frac{w_{F\setminus H}(\sigma \cup \tau)}{w_{F\setminus H}(\sigma \cup \eta_1)} \leq \frac{\exp\left(\beta_G|E_G| + \beta(\sigma_1(N-1) + a)\right)}{\exp\left(\beta_G|E_G| + \sigma_1 N\beta\right)} \leq e^{(a-\sigma_1)\beta} \leq e^{\left(-\alpha' m + 2m^{3/4}\right)\beta} \leq e^{-\alpha'' m\beta},$$

where $\alpha' = \hat{\alpha} - (1-\hat{\alpha})/(q-1) > 0$ and the rightmost inequality is true for some $\alpha'' > 0$ and sufficiently large m. For $c_1 = (2\log q)/\alpha''$ we have for $\beta \geq c_1 N/m$

$$\frac{w_{F\setminus H}(\sigma \cup \tau)}{w_{F\setminus H}(\sigma \cup \eta_1)} \leq q^{-2N}.$$

Hence

$$\sum_{\tau \neq \eta_1 \in \Omega_G} w_{F\setminus H}(\sigma \cup \tau) \leq q^{-N} w_{F\setminus H}(\sigma \cup \eta_1). \tag{6}$$

Now,

$$\begin{aligned} w_{F\setminus H}(\sigma \cup \eta_1) &= \exp\left(\beta_G |E_G| + \sigma_1 N\beta\right) \\ &\leq \exp\left(\beta_G |E_G| + \hat{\alpha} m N\beta + m^{3/4} N\beta\right) \\ &\leq e^{\delta/2} \exp\left(\beta_G |E_G| + \hat{\alpha} m N\beta\right), \end{aligned} \tag{7}$$

where in the last equality we take $c_2 = \delta/2$ and use the fact that $\beta \leq c_2/(Nm^{3/4})$. Therefore, when $\sigma \in M$ is such that $|\sigma_1 - \hat{\alpha} m| < m^{3/4}$, we have

$$\sum_{\tau \in \Omega_G} w_{F\setminus H}(\sigma \cup \tau) \leq (1 + q^{-N}) w_{F\setminus H}(\sigma \cup \eta_1) \leq e^{\delta} \exp\left(\beta_G |E_G| + \hat{\alpha} m N\beta\right),$$

for N sufficiently large. By symmetry, we then get that

$$Z_F^{\mathrm{M}} \leq \sum_{\sigma \in M} w_H(\sigma) e^{\delta} \exp\left(\beta_G |E_G| + \hat{\alpha} m N\beta\right) = e^{\delta} Z_H^{\mathrm{M}} \exp\left(\beta_G |E_G| + \hat{\alpha} m N\beta\right).$$

The lower bound in (2) can be derived in similar fashion and part 1 of the lemma follows.

For part 2, suppose that $\sigma \in D$ and let $\tau \in \Omega_G$. Let τ_i be the number of vertices of G assigned spin i in τ and let $w_G(\tau)$ denote the weight of τ for the Potts model (G, β_G). Then,

$$\begin{aligned} w_{F\setminus H}(\sigma \cup \tau) &= w_G(\tau) \exp\left(\beta \sum_{i=1}^{q} \sigma_i \tau_i\right) \\ &\leq w_G(\tau) \exp\left(m^{3/4} N\beta + \beta m N/q\right) \\ &\leq e^{\delta} w_G(\tau) \exp\left(\beta m N/q\right), \end{aligned} \tag{8}$$

since recall we set $c_2 = \delta/2$. Hence,

$$Z_F^{\mathrm{D}} = \sum_{\sigma \in D} \sum_{\tau \in \Omega_G} w_H(\sigma) w_{F\setminus H}(\sigma \cup \tau) \leq e^{\delta} Z_H^{\mathrm{D}} Z_G \exp\left(\beta m N/q\right).$$

The lower bound for Z_F^{D} can be derived analogously and part 2 of the lemma follows.

Finally for part 3, note that

$$\begin{aligned} Z_F^{\mathrm{S}} = \sum_{\sigma \in S} \sum_{\tau \in \Omega_G} w_H(\sigma) w_{F\setminus H}(\sigma \cup \tau) &\leq q^N \exp\left(\beta_G N^2 + \beta N m\right) Z_H^{\mathrm{S}} \\ &\leq \min\{Z_H^{\mathrm{M}}, Z_H^{\mathrm{D}}\} \exp\left(-\Omega(\sqrt{m})\right), \end{aligned}$$

where the last inequality follows for sufficiently large N and m from Lemma 9 and the fact that $\beta < c_2/(Nm^{3/4})$. Then,

$$\frac{Z_F^{\mathrm{S}}}{Z_F} \leq \frac{Z_F^{\mathrm{S}}}{Z_F^{\mathrm{M}}} \leq \exp\left(-\Omega(\sqrt{m})\right),$$

and the result follows. ■

Hidden Model Construction. We now construct our hidden model and show that we can efficiently generate samples from its Gibbs distribution. Let F^* be the graph obtained by our construction above where we replace the graph G by a complete graph on N vertices. More precisely, let $K = K_N$ be a complete graph on N vertices and let F^* be the graph that results from connecting the vertices of K and H with a complete bipartite graph $K_{N,m}$.

The edges of K have parameter $\beta_K = \beta_G + 4\log q$, whereas the remaining edges have the same interaction strength as in F; that is, edges between K and H will have parameter β and those in H parameter β_H. This Potts model on F^*, which again with a slight abuse of notation we denote by F^*, will act as the hidden model. We choose $\beta_K = \beta_G + 4\log q$, so that K is more likely to be monochromatic than G. Let μ_{F^*} the corresponding Gibbs distribution on F^*. We show next that we can efficiently generate samples from μ_{F^*}.

Lemma 12 *There is an exact sampling algorithm for the distribution μ_{F^*} with running time* $\mathrm{poly}(n)$.

Proof Because of symmetry there are at most n^{2q} types of configurations—described by their signatures on H and K; recall that $n = m + N$. We can then enumerate every signature, explicitly compute its probability and sample from the resulting distribution. This involves computing multinomial coefficients, but they can each be expressed as product of q binomial coefficients which can be easily computed in $\mathrm{poly}(n)$ time. Once the signature is generated from the correct distribution, we can simply take a random permutation of the vertices to assign their spins. ■

Proof Overview. We provide the high-level idea of the reduction next. Recall that our goal is to provide a polynomial-time algorithm for the decision version of the r-approximate counting problem for the ferromagnetic Potts model (G, β_G). That is, for a real number $\hat{Z}$ we want to determine whether $Z_G \le \frac{1}{r}\hat{Z}$ or $Z_G \ge r\hat{Z}$, where $Z_G := Z_{G,\beta_G}$ is the partition function of the model (G, β_G) .

For any "reasonable" $\hat{Z} \in \mathbb{R}$ (i.e., $\hat{Z}$ that is not too small or too large, in which case the approximate counting problem becomes trivial), we can find a value of the parameter β_H for our construction such that

$$\frac{Z_F^{\mathrm{D}}(\beta_H)}{Z_F^{\mathrm{M}}(\beta_H)} \approx \frac{1}{\sqrt{\varepsilon L}} \frac{Z_G}{\hat{Z}},$$

where $L = L(n)$ and $\varepsilon = \varepsilon(n)$ are the sample complexity and accuracy parameter of the testing algorithm, respectively. This is possible because of the first-order phase transition of the ferromagnetic mean-field q-state Potts model for $q \ge 3$, and the associated phase coexistence and metastability phenomena discusses earlier; see Section 2.2.1. (Specifically, by Lemma 10 we can find β_H so that $Z_H^{\mathrm{M}}(\beta_H)/Z_H^{\mathrm{D}}(\beta_H) \approx R$ for any target R, and then we can use Lemma 11 to translate this value to a value for $Z_G \cdot Z_F^{\mathrm{M}}(\beta_H)/Z_F^{\mathrm{D}}(\beta_H)$.)

For this choice of β_H and setting $r \approx \sqrt{L/\varepsilon}$, note that if $Z_G \le \frac{1}{r}\hat{Z}$, then $Z_F^{\mathrm{D}}(\beta_H)/Z_F^{\mathrm{M}}(\beta_H)$ is small ($\lesssim 1/L$). Conversely, when $Z_G \ge r\hat{Z}$, the ratio is large ($\gtrsim 1/\varepsilon$). Therefore, to distinguish whether $Z_G \le \frac{1}{r}\hat{Z}$ or $Z_G \ge r\hat{Z}$ it is sufficient to determine whether the ratio $Z_F^{\mathrm{D}}(\beta_H)/Z_F^{\mathrm{M}}(\beta_H)$ is small or large. For this we can use the identity testing algorithm. In particular, when the ratio is small ($\lesssim 1/L$), the majority phase of H is dominant in F, and G will likely be monochromatic.

Since this is also the case in F^* (i.e., K is monochromatic with high probability), then the models F and F^* will be close in total variation distance ($\lesssim 1/L$), and the testing algorithm using only L samples would output Yes. Otherwise, when $Z_F^{\mathrm{D}}(\beta_H)/Z_F^{\mathrm{M}}(\beta_H)$ is large ($\gtrsim 1/\varepsilon$), the disorder phase is dominant, so F and F^* are likely to disagree on the spins of G and K; this implies that their total variation distance is large ($\gtrsim 1-\varepsilon$), and so the tester would output No. We proceed to flesh out the technical details next.

Lemma 13 *Let $\varepsilon \in (0,1)$ be a constant, $L = L(n) = \mathrm{poly}(n)$ and $r = 96\varepsilon^{-1}\sqrt{\varepsilon L + 1}$. Suppose $\hat{Z} \in \mathbb{R}$ is such that $rq\exp(\beta_G|E_G|) \le \hat{Z} \le \frac{1}{r}q^N \exp(\beta_G|E_G|)$. Then, there exists constants $c, c_1, c_2 > 0$ such that the following holds. For any $\beta \in \left[\frac{c_1 N}{m}, \frac{c_2}{Nm^{3/4}}\right]$, we can find $\beta_H > 0$ in the range $|\beta_H - \mathfrak{B}_o/m| \le cm^{-3/2}$ in $\mathrm{poly}(n)$ time such that all of the following holds:*

(i) $\frac{1}{4\sqrt{\varepsilon L+1}}\frac{Z_G}{\hat{Z}} \le \frac{Z_F^{\mathrm{D}}(\beta_H)}{Z_F^{\mathrm{M}}(\beta_H)} \le \frac{1}{\sqrt{\varepsilon L+1}}\frac{Z_G}{\hat{Z}}$, *and* $\frac{Z_F^{\mathrm{S}}(\beta_H)}{Z_F(\beta_H)} \le e^{-c_3\sqrt{m}}$;

(ii) $\frac{Z_{F^*}^{\mathrm{D}}(\beta_H)}{Z_{F^*}^{\mathrm{M}}(\beta_H)} \le \frac{2}{r\sqrt{\varepsilon L+1}}$, *and* $\frac{Z_{F^*}^{\mathrm{S}}(\beta_H)}{Z_{F^*}(\beta_H)} \le e^{-c_3\sqrt{m}}$;

(iii) If $Z_G \le \frac{1}{r}\hat{Z}$, then $\|\mu_F - \mu_{F^*}\|_{\mathrm{TV}} \le \frac{1}{16L}$;

(iv) If $Z_G \ge r\hat{Z}$, then $\|\mu_F - \mu_{F^*}\|_{\mathrm{TV}} \ge 1-\varepsilon$.

Proof Let $\alpha_0 = \hat{\alpha} - 1/q$. By parts 1 and 2 of Lemma 11, for any β_H such that $|\beta_H - \mathfrak{B}_o/m| \le cm^{-3/2}$ and any $\beta \in \left[\frac{c_1 N}{m}, \frac{c_2}{Nm^{3/4}}\right]$ for suitable constants $c_1, c_2 > 0$, we have

$$\frac{2}{3}\cdot\exp\left(-\alpha_0\beta Nm - \beta_G|E_G|\right)\cdot\frac{Z_H^{\mathrm{D}}}{Z_H^{\mathrm{M}}}\cdot Z_G \le \frac{Z_F^{\mathrm{D}}}{Z_F^{\mathrm{M}}} \le \frac{4}{3}\cdot\exp\left(-\alpha_0\beta Nm - \beta_G|E_G|\right)\cdot\frac{Z_H^{\mathrm{D}}}{Z_H^{\mathrm{M}}}\cdot Z_G,$$

where for ease of notation we dropped the dependence on β_H and set $Z_G = Z_{G,\beta_G}$. Moreover, part 3 of the same lemma implies that there exists a constant $c_3 > 0$ such that

$$\frac{Z_F^{\mathrm{S}}}{Z_F} \le e^{-c_3\sqrt{m}}. \tag{9}$$

Recall that $n = m + N$ and $m = N^{10}$. By Lemma 10, we can find $\beta_H > 0$ in $\mathrm{poly}(m)$ time such that $|\beta_H - \mathfrak{B}_o/m| \le cm^{-3/2}$ and

$$\frac{3}{8\sqrt{\varepsilon L+1}}\cdot\exp\left(\alpha_0\beta Nm + \beta_G|E_G|\right)\cdot\frac{1}{\hat{Z}} \le \frac{Z_H^{\mathrm{D}}}{Z_H^{\mathrm{M}}} \le \frac{3}{4\sqrt{\varepsilon L+1}}\cdot\exp\left(\alpha_0\beta Nm + \beta_G|E_G|\right)\cdot\frac{1}{\hat{Z}}; \tag{10}$$

note that the assumptions $rq\exp(\beta_G|E_G|) \le \hat{Z} \le \frac{1}{r}q^N\exp(\beta_G|E_G|)$ and $r = \mathrm{poly}(n)$ ensure that $Z_H^{\mathrm{D}}/Z_H^{\mathrm{M}}$ is in the desired range. Thus, for this choice of β_H we get

$$\frac{1}{4\sqrt{\varepsilon L+1}}\frac{Z_G}{\hat{Z}} \le \frac{Z_F^{\mathrm{D}}}{Z_F^{\mathrm{M}}} \le \frac{1}{\sqrt{\varepsilon L+1}}\frac{Z_G}{\hat{Z}}. \tag{11}$$

This establishes part (i) of the lemma.

For part (ii), we note that Lemma 11 holds for the hidden model F^* (with F and G replaced by F^* and K, respectively), without any change in the proof. Hence, we get

$$\frac{Z^{\mathrm{D}}_{F^*}}{Z^{\mathrm{M}}_{F^*}} \le \frac{4}{3} \cdot \exp\left(-\alpha_0 \beta N m - \beta_K |E_K|\right) \cdot \frac{Z^{\mathrm{D}}_H}{Z^{\mathrm{M}}_H} \cdot Z_K \tag{12}$$

and

$$\frac{Z^{\mathrm{S}}_{F^*}}{Z_{F^*}} \le e^{-c_3\sqrt{m}}. \tag{13}$$

Thus, for our choice of β_H we deduce from (10), (12) and (13) that

$$\frac{Z^{\mathrm{D}}_{F^*}}{Z^{\mathrm{M}}_{F^*}} \le \frac{1}{\sqrt{\varepsilon L+1}} \exp(\beta_G|E_G| - \beta_K|E_K|) \cdot \frac{Z_K}{\hat{Z}} \le \frac{2}{r\sqrt{\varepsilon L+1}},$$

where the last inequality follows from $q\exp(\beta_K|E_K|)/Z_K \ge 1/2$ when $\beta_K \ge 4\log q$ and the assumption that $\hat{Z} \ge rq\exp(\beta_G|E_G|)$.

We prove part (iii) next. Suppose that $Z_G \le \frac{1}{r}\hat{Z}$ and let ν_F be the conditional distribution of μ_F conditioned on the configuration on H being in the majority phase (i.e., in the set M). That is, for $\sigma \in \Omega_H$ and $\tau \in \Omega_G$,

$$\nu_F(\sigma \cup \tau) = \mathbb{1}(\sigma \in M)\frac{\mu_F(\sigma \cup \tau) Z_F}{Z^{\mathrm{M}}_F}.$$

From the definition of total variation distance we have

$$\|\mu_F - \nu_F\|_{\mathrm{TV}} = \sum_{\eta \in \Omega_F : \mu_F(\eta) \ge \nu_F(\eta)} \mu_F(\eta) - \nu_F(\eta) = \frac{Z^{\mathrm{D}}_F + Z^{\mathrm{S}}_F}{Z_F}.$$

From (9), (11) and the assumption that $Z_G \le \frac{1}{r}\hat{Z}$, we get

$$\|\mu_F - \nu_F\|_{\mathrm{TV}} \le \frac{Z^{\mathrm{D}}_F}{Z^{\mathrm{M}}_F} + \frac{Z^{\mathrm{S}}_F}{Z_F} \le \frac{1}{\sqrt{\varepsilon L+1}}\frac{Z_G}{\hat{Z}} + e^{-c_3\sqrt{m}} \le \frac{1}{r\sqrt{\varepsilon L+1}} + e^{-c_3\sqrt{m}}.$$

Since $r = 96\varepsilon^{-1}\sqrt{\varepsilon L+1}$, it follows that

$$\|\mu_F - \nu_F\|_{\mathrm{TV}} \le \frac{\varepsilon}{96(\varepsilon L+1)} + e^{-c_3\sqrt{m}} \le \frac{1}{96L} + e^{-c_3\sqrt{m}}. \tag{14}$$

Similarly, for the distribution μ_{F^*} and the conditional distribution ν_{F^*} of the majority phase, we also have

$$\|\mu_{F^*} - \nu_{F^*}\|_{\mathrm{TV}} \le \frac{Z^{\mathrm{D}}_{F^*}}{Z^{\mathrm{M}}_{F^*}} + \frac{Z^{\mathrm{S}}_{F^*}}{Z_{F^*}} \le \frac{2}{r\sqrt{\varepsilon L+1}} + e^{-c_3\sqrt{m}} \le \min\left\{\frac{1}{48L}, \frac{\varepsilon}{48}\right\} + e^{-c_3\sqrt{m}}. \tag{15}$$

Let $\mathcal{A}$ be the event that all vertices of G are assigned the same spin. By drawing a sample from ν_F and sequentially resampling the spin of each vertex of G, we deduce from a union bound and the fact $\hat{\alpha} > 1/q$ that

$$1 - \nu_F(\mathcal{A}) \le N \cdot \frac{\exp\left(\beta_G N + \beta(\frac{1-\hat{\alpha}}{q-1}m + m^{3/4})\right)}{\exp\left(\beta(\hat{\alpha}m - m^{3/4})\right)} \le e^{-\gamma\beta m}$$

for a suitable constant $\gamma > 0$; similarly

$$1 - \nu_{F^*}(\mathcal{A}) \le e^{-\gamma\beta m}.$$

Let $\rho = \nu_F(\cdot\,|\mathcal{A})$ denote the conditional distribution of ν_F given $\mathcal{A}$. Observe that ρ does not depend on the graph G, because we condition on the event that all vertices from G receive the same spin, and thus the structure of G does not affect the conditional distribution ρ. In particular, we have $\rho = \nu_F(\cdot\,|\mathcal{A}) = \nu_{F^*}(\cdot\,|\mathcal{A})$. Thus, we get

$$\|\nu_F - \nu_{F^*}\|_{\text{TV}} \le \|\nu_F - \rho\|_{\text{TV}} + \|\nu_{F^*} - \rho\|_{\text{TV}} = 1 - \nu_F(\mathcal{A}) + 1 - \nu_{F^*}(\mathcal{A}) \le 2e^{-\gamma\beta m}. \tag{16}$$

From (14), (15), (16) and the triangle inequality, we conclude that

$$\begin{aligned}\|\mu_F - \mu_{F^*}\|_{\text{TV}} &\le \|\mu_F - \nu_F\|_{\text{TV}} + \|\mu_{F^*} - \nu_{F^*}\|_{\text{TV}} + \|\nu_F - \nu_{F^*}\|_{\text{TV}} \\ &\le \frac{1}{32L} + 2e^{-c_3\sqrt{m}} + 2e^{-\gamma\beta m} \le \frac{1}{16L}.\end{aligned}$$

and part (i) follows.

Finally, for part (iv), suppose that $Z_G \ge r\hat{Z}$. Then,

$$\|\mu_F - \nu_F\|_{\text{TV}} = 1 - \frac{Z_F^{\text{M}}}{Z_F} \ge 1 - \frac{Z_F^{\text{M}}}{Z_F^{\text{D}}} \ge 1 - 4\sqrt{\varepsilon L + 1}\,\frac{\hat{Z}}{Z_G} \ge 1 - \frac{4}{r}\sqrt{\varepsilon L + 1} = 1 - \frac{\varepsilon}{24}. \tag{17}$$

Thus, equations (17), (15), (16) and the triangle inequality imply that

$$\begin{aligned}\|\mu_F - \mu_{F^*}\|_{\text{TV}} &\ge \|\mu_F - \nu_F\|_{\text{TV}} - \|\mu_{F^*} - \nu_{F^*}\|_{\text{TV}} - \|\nu_F - \nu_{F^*}\|_{\text{TV}} \\ &\ge 1 - \frac{\varepsilon}{16} - e^{-c_3\sqrt{m}} - 2e^{-\gamma\beta m} \ge 1 - \varepsilon,\end{aligned}$$

and the result follows. ■

2.2.3 A generic reduction from counting to testing

Theorem 8 will follow from Lemmas 12 and 13 using the following general reduction from the decision version of r-approximate counting to testing.

Theorem 14 *Let (G, β_G, h_G) be a Potts model on an N-vertex graph G with partition function Z_G and let $\hat{Z} \in \mathbb{R}$. Let $\varepsilon \in (0,1)$ be a constant, $n = \text{poly}(N)$ and suppose there exists an ε-identity testing algorithm for a family of Potts models $\mathcal{M}$ on n-vertex graphs with sample complexity $L = L(n) = \text{poly}(n)$ and $\text{poly}(n)$ running time. Suppose that given (G, β_G, h_G), $\hat{Z}$, ε and L, there exists $r = \text{poly}(L, \varepsilon^{-1})$ such that we can construct two models $F, F^* \in \mathcal{M}$ in $\text{poly}(n)$ time satisfying:*

- *(i) If $Z_G \le \frac{1}{r}\hat{Z}$, then $\|\mu_F - \mu_{F^*}\|_{\text{TV}} \le \frac{1}{16L}$;*
- *(ii) If $Z_G \ge r\hat{Z}$, then $\|\mu_F - \mu_{F^*}\|_{\text{TV}} \ge 1 - \varepsilon$; and*
- *(iii) We can generate samples from a distribution $\mu_{F^*}^{ALG}$ such that $\left\|\mu_{F^*} - \mu_{F^*}^{ALG}\right\|_{\text{TV}} \le \delta$ in time $\text{poly}(n, \delta^{-1})$.*

Then, there is a $\mathrm{poly}(N)$ *running time algorithm for the decision version of* r*-approximate counting for* (G, β_G, h_G) *that succeeds with probability at least* $5/8$*.*

Proof Recall that the input to the decision version of r-approximate counting is the model defined by (G, β_G, h_G) and a real number $\hat{Z} > 0$; the goal is to determine whether $Z_G \leq \frac{1}{r}\hat{Z}$ or $Z_G \geq r\hat{Z}$. The algorithm proceeds as follows:

1. Construct the Potts models F and F^* in $\mathcal{M}$.
2. Generate $L = L(n)$ δ-approximate samples $\mathcal{S} = \{\sigma_1, \ldots, \sigma_L\}$ from $\mu^{\text{ALG}}_{F^*}$, setting $\delta = \frac{1}{16L}$.
3. The input to the testing algorithm, henceforth called the Tester, is F, which plays the role of the visible model, and the samples $\mathcal{S}$.
4. If the Tester outputs Yes, then return $Z_G \leq \frac{1}{r}\hat{Z}$.
5. If the Tester outputs No, then return $Z_G \geq r\hat{Z}$.

We show next that our output for decision version of r-approximate counting is correct with probability at least $5/8$. Consider first the case when $Z_G \leq \frac{1}{r}\hat{Z}$. If this is the case, then by assumption we have $\|\mu_F - \mu_{F^*}\|_{\text{TV}} \leq \frac{1}{16L}$ and

$$\|\mu_{F^*} - \mu^{\text{ALG}}_{F^*}\|_{\text{TV}} \leq \frac{1}{16L}. \tag{18}$$

So, by the triangle inequality,

$$\|\mu_F - \mu^{\text{ALG}}_{F^*}\|_{\text{TV}} \leq \frac{1}{8L}.$$

Let $(\mu_F)^{\otimes L}$, $(\mu_{F^*})^{\otimes L}$ and $(\mu^{\text{ALG}}_{F^*})^{\otimes L}$ be the product distributions corresponding to L independent samples from μ_F, μ_{F^*} and $\mu^{\text{ALG}}_{F^*}$ respectively. We have

$$\left\|(\mu_F)^{\otimes L} - (\mu^{\text{ALG}}_{F^*})^{\otimes L}\right\|_{\text{TV}} \leq L\,\|\mu_F - \mu^{\text{ALG}}_{F^*}\|_{\text{TV}} \leq \frac{1}{8}.$$

Hence, if π is the optimal coupling of the distributions $(\mu^{\text{ALG}}_{F^*})^{\otimes L}$ and $(\mu_F)^{\otimes L}$, and $(\mathcal{S}, \mathcal{S}')$ is sampled from π, then $\mathcal{S} \sim (\mu^{\text{ALG}}_{F^*})^{\otimes L}$, $\mathcal{S}' \sim (\mu_F)^{\otimes L}$ and $\pi(S \neq S') \leq \frac{1}{8}$. Therefore,

$$\begin{aligned}
&\Pr[\text{Tester outputs No when given samples } \mathcal{S} \text{ where } \mathcal{S} \sim (\mu^{\text{ALG}}_{F^*})^{\otimes L}] \\
&\quad = \Pr[\text{Tester outputs No when given samples } \mathcal{S} \text{ where } (\mathcal{S}, \mathcal{S}') \sim \pi] \\
&\quad \leq \Pr[\text{Tester outputs No when given samples } \mathcal{S}' \text{ where } (\mathcal{S}, \mathcal{S}') \sim \pi] + \pi(S \neq S') \\
&\quad = \Pr[\text{Tester outputs No when given samples } \mathcal{S}' \text{ where } \mathcal{S}' \sim (\mu_F)^{\otimes L}] + \pi(S \neq S') \\
&\quad \leq \frac{1}{4} + \frac{1}{8} \leq \frac{3}{8}.
\end{aligned} \tag{19}$$

Hence, the Tester returns Yes (and our output is correct) with probability at least $5/8$.

Now, if $Z_G \geq r\hat{Z}$, then by assumption $\|\mu_F - \mu_{F^*}\|_{\text{TV}} > 1 - \varepsilon$. Moreover, by (18)

$$\left\|(\mu_{F^*})^{\otimes L} - (\mu^{\text{ALG}}_{F^*})^{\otimes L}\right\|_{\text{TV}} \leq L\,\|\mu_{F^*} - \mu^{\text{ALG}}_{F^*}\|_{\text{TV}} \leq \frac{1}{8}.$$

Thus, analogously to (19) (i.e., using the optimal coupling for $(\mu_{F^*}^{\text{ALG}})^{\otimes L}$ and $(\mu_{F^*})^{\otimes L}$), we get

$$\Pr\left[\text{Tester outputs Yes when given samples } \mathcal{S} \text{ where } \mathcal{S} \sim (\mu_{F^*}^{\text{ALG}})^{\otimes L}\right] \leq \frac{3}{8}.$$

Hence, the Tester returns No with probability at least $5/8$. Therefore, we can conclude that our algorithm for decision r-approximate counting succeeds with probability at least $5/8$. The result then follows from the fact that the running time of the algorithm is $\text{poly}(N)$, as each step of the algorithm takes at most $\text{poly}(N)$ time by our assumptions. ■

2.2.4 Proof of Theorem 8

We can now prove Theorem 8 which states hardness of identity testing for the ferromagnetic Potts model on general graphs.

Proof of Theorem 8 Consider the ferromagnetic Potts model on an N-vertex graph $G = (V_G, E_G)$ with constant edge weight β_G in every edge and no external field. Let $\hat{Z} > 0$ be a real number and let $n = N^{10}+N$. Suppose there is an ε-identity testing algorithm for $\mathcal{M}^{+}_{\text{Potts}}(n, n, \beta_G, 0)$ with sample complexity $L = L(n) = \text{poly}(n)$ and running time $\text{poly}(n)$. Let $r = 96\varepsilon^{-1}\sqrt{\varepsilon L + 1}$; our goal is to determine whether $Z_G \leq \frac{1}{r}\hat{Z}$ or $Z_G \geq r\hat{Z}$ where $Z_G := Z_{G,\beta_G}$.

We construct the Potts models F and F^* as describe in Section 2.2.2 with corresponding Gibbs distributions μ_F and μ_{F^*} using the values of β and β_H supplied by Lemma 13; hence the models F and F^* belong to $\mathcal{M}^{+}_{\text{Potts}}(n, n, \beta_G, 0)$, since $\beta_G > \max\{\beta, \beta_H\}$.

Lemmas 13 ensures that when

$$rqe^{\beta_G|E_G|} \leq \hat{Z} \leq \frac{q^N}{r}e^{\beta_G|E_G|}, \tag{20}$$

conditions (i) and (ii) in Theorem 14 are satisfied. Moreover, Lemma 12 gives condition (iii). Thus, Theorem 14 implies that we have an algorithm for the decision version of r-approximate counting for the Potts model on G when $\hat{Z}$ satisfies (20). Meanwhile, we can bound Z_G crudely by

$$qe^{\beta_G|E_G|} \leq Z_G \leq q^N e^{\beta_G|E_G|}.$$

Thus, if $\hat{Z} < rq\exp(\beta_G|E_G|) \leq rZ_G$, we can output $\hat{Z} \leq \frac{1}{r}Z_G$. Similarly, when

$$\hat{Z} > \frac{1}{r}q^N\exp(\beta_G|E_G|) \geq \frac{1}{r}Z_G$$

we can output $\hat{Z} \geq rZ_G$. Therefore, we have a $\text{poly}(N)$ algorithm for the decision version of r-approximate counting for $\hat{\mathcal{M}}^{+}_{\text{Potts}}(N, N, \beta_G, 0)$ where $N = \Theta(n^{1/10})$, $r = \text{poly}(N)$ and $\beta_G = \Theta(1)$. The result then follows from Theorem 7 and the fact that there is no FPRAS for $\hat{\mathcal{M}}^{+}_{\text{Potts}}(N, N, \beta_G, 0)$ unless there is one for #BIS (Goldberg and Jerrum, 2012; Galanis et al., 2016b). ■

2.3 Step 3: Degree reduction

The following result provides a reduction from identity testing in the family $\mathcal{M}_{\text{Potts}}(\hat{n}, d, \hat{\beta}, \hat{h})$ to identity testing in $\mathcal{M}_{\text{Potts}}(n, n, \beta, h)$, under some mild assumptions on the model parameters; this allows us to deduce the hardness of identity problem on graphs of bounded degree as stated in Theorem 3 using the main result Theorem 8 from the previous section.

Theorem 15 *Let $\hat{n}, d \in \mathbb{N}^+$ be such that $3 \leq d \leq \hat{n}^{1-\rho}$ for some constant $\rho \in (0,1)$. Suppose that for some constants $\beta, h \geq 0$ there is no $\mathrm{poly}(n)$ running time ε-identity testing algorithm for $\mathcal{M}_{\text{Potts}}(n, n, \beta, h)$. Then there exists a constant $c \in (0,1)$ such that, for any constant $\hat{\varepsilon} > \varepsilon$ there is no $\mathrm{poly}(\hat{n})$ running time $\hat{\varepsilon}$-identity testing algorithm for $\mathcal{M}_{\text{Potts}}(\hat{n}, d, \hat{\beta}, \hat{h})$ provided $\hat{\beta} d = \omega(\log \hat{n})$ and $\hat{h} \leq h\hat{n}^{-c}$.*

This theorem is a special case of our more general result in Theorem 28, which we prove in Section 5. We conclude with the proof of Theorem 3.

Proof of Theorem 3 Follows from Theorems 8 and 15. ■

3. Testing mixed RBMs with no external fields

In this section, we show that identity testing for RBMs with arbitrary edges interactions is computationally hard, even in the absence of an external field (i.e., $h = 0$); specifically, we prove Theorem 1 from the introduction. For this, we establish first the hardness of the identity testing problem for antiferromagnetic Ising models with bounded edge interactions. We then reduce this problem to identity testing for mixed RBMs using our degree reduction machinery (see Sections and 2.3 and 5) which conveniently also turns our instance into a bipartite graph.

We start by reducing the problem of approximating the partition function of the antiferromagnetic Ising models to identity testing. Hence, the following well-known result concerning the hardness of approximate counting in the antiferromagnetic setting plays an important role for us.

Theorem 16 (Sly and Sun, 2012; Galanis et al., 2016c) *Let $d \geq 3$ be an integer and let $\beta_0 > \beta_c(d) := \mathrm{arctanh}(1/(d-1))$ be a real number. Then, for a sufficiently large integer N, there is no* FPRAS *for the partition function of the antiferromagnetic Ising model on d-regular N-vertex graphs with interaction β_0 on every edge, unless RP = NP.*

The next step in our proof is a reduction from the decision version of approximate counting (see Definition 6) to identity testing.

Theorem 17 *Let $\varepsilon \in (0,1)$ be any constant. There exists $0 < \beta_0 = O(1)$ such that an ε-identity testing algorithm for $\mathcal{M}^{-}_{\text{Ising}}(n, n, \beta_0, 0)$ with $\mathrm{poly}(n)$ sample complexity and running time can be used to solve the decision r-approximate counting problem for $\hat{\mathcal{M}}^{-}_{\text{Ising}}(N, 3, -0.6, 0)$ in $\mathrm{poly}(N)$ time, where $N = \Theta(\sqrt{n})$ and $r = \mathrm{poly}(N)$.*

We can now provide the proof of Theorem 1.

Proof of Theorem 1 From Theorems 16 and 7, it follows that for any $c > 0$ there is no $\mathrm{poly}(N)$ running time algorithm for the decision version of N^c-approximate counting for the family $\hat{\mathcal{M}}^{-}_{\text{ISING}}(N, 3, -0.6, 0)$ unless RP $=$ NP. Theorem 17 then implies that, under the same assumption that RP $=$ NP, there is no ε-identity testing algorithm for $\mathcal{M}^{-}_{\text{ISING}}(n, n, \beta_0, 0)$ with $\mathrm{poly}(n)$ sample complexity and running time for constant $\varepsilon \in (0, 1)$, $N = \Theta(\sqrt{n})$ and a suitable constant $\beta_0 > 0$. The result then follows from Theorem 28. ■

We provide in the next section the missing proof of Theorem 17

3.1 Reducing counting to testing for the antiferromagnetic Ising model: proof of Theorem 17

Testing instance construction. Consider an antiferromagnetic Ising model on an N-vertex 3-regular graph $G = (V_G, E_G)$ with the same inverse temperature parameter $\beta_G = -0.6$ on every edge an no external field. We provide an algorithm for the decision version of r-approximate counting for $Z_G := Z_{G,\beta_G,0}$, using the presumed identity testing algorithm.

Define F to be a graph with the vertex set

$$V_F = V_G \cup \{s_1, s_2\} \cup \left\{u^{(i)}_{v,j} : v \in V_G, 1 \le i \le N, j \in \{1, 2\}\right\} \cup \left\{w^{(i)}_j : 1 \le i \le N^2, j \in \{1, 2\}\right\}$$

and the edge set

$$\begin{aligned} E_F = E_G &\cup \left\{\{u^{(i)}_{v,j}, v\}, \{u^{(i)}_{v,j}, s_j\} : v \in V_G, 1 \le i \le N, j \in \{1, 2\}\right\} \\ &\cup \left\{\{w^{(i)}_j, s_j\} : 1 \le i \le N^2, j \in \{1, 2\}\right\} \\ &\cup \left\{\{w^{(i)}_1, w^{(i)}_2\} : 1 \le i \le N^2\right\}; \end{aligned}$$

see Figure 1. Observe that F has $n = 4N^2 + N + 2$ vertices. Given two real numbers $\beta_1, \beta_2 > 0$, we then define an antiferromagnetic Ising model on the graph F as follows:

1. Every edge $\{u, v\} \in E_G$ has weight -0.6.
2. For every $v \in V_G$, $1 \le i \le N$ and $j = 1, 2$, the two edges $\{u^{(i)}_{v,j}, v\}$ and $\{u^{(i)}_{v,j}, s_j\}$ have weight $-\beta_1$;
3. For every $1 \le i \le N^2$, the edges $\{w^{(i)}_1, s_1\}$, $\{w^{(i)}_2, s_2\}$ and $\{w^{(i)}_1, w^{(i)}_2\}$ have weight $-\beta_2$.

We slight abuse of notation, we use F for the resulting Ising model on F and $\mu := \mu_F$ for the corresponding Gibbs distribution. F will be the visible model of our testing instance.

For the hidden model F^*, we consider the same construction above but replacing G with an independent set I_N on V_G. Let $\mu^* := \mu_{F^*}$ be the corresponding the Gibbs distribution. We note first that we can efficiently sample from μ^*.

Lemma 18 *There is an exact sampling algorithm for the distribution μ^* with running time* $\mathrm{poly}(n)$.

Proof Configurations in Ω_{F^*} can be classified by their type, which is given by the spins of s_1, s_2 and the number of spin 1's in the independent set I_N. There are $4(N + 1)$ types in total. Observe that configurations of each type have the same weight by symmetry, and this weight can be

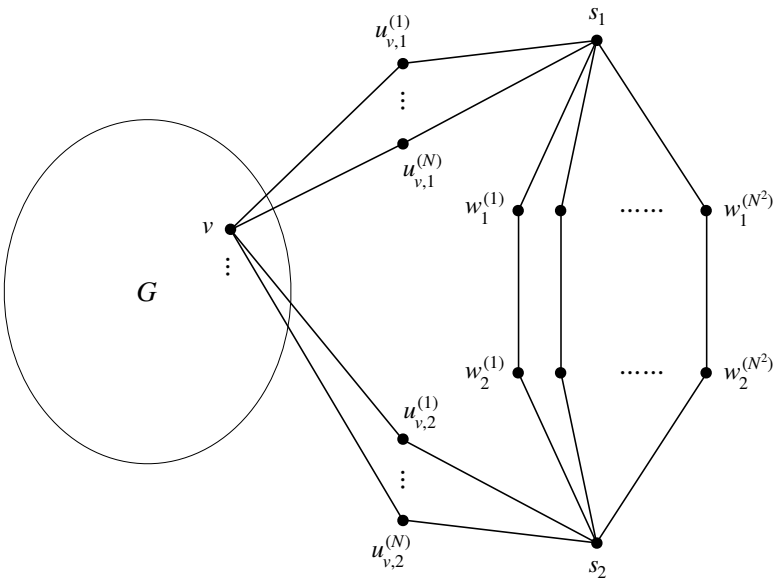

Figure 1: The graph F. For every vertex $v \in V_G$ and $j \in \{1,2\}$, v and s_j are connected by N disjoint paths of length 2. Also, s_1 and s_2 are connected by N^2 disjoint paths of length 3.

computed efficiently since given the spins of s_1, s_2 and I_N the remaining graph has only isolated vertices and edges. Also it is easy to get the number of configurations of each type. Thus, to sample from μ^*, we can first sample a type from the induced distribution on types, and then sample a configuration of the given type uniformly at random. ■

Our hidden and visible models F and F^* are related as follows.

Lemma 19 *Let $\varepsilon \in (0,1)$ be a constant, $L = L(n) = \mathrm{poly}(n)$ and $r = 96\varepsilon^{-1}\sqrt{\varepsilon L + 1}$. Suppose $\hat{Z} \in \mathbb{R}$ is such that $r2^N e^{-0.9N} \le \hat{Z} \le \frac{1}{r}2^N$. Then, for any $\beta_1 \ge 3$, we can find $0 < \beta_2 < \beta_1 + 2$ in $\mathrm{poly}(n)$ time such that all of the following holds:*

(i) $\frac{1}{4\sqrt{\varepsilon L+1}}\frac{Z_G}{\hat{Z}} \le \frac{Z_F^{\mathrm{D}}}{Z_F^{\mathrm{M}}} \le \frac{1}{\sqrt{\varepsilon L+1}}\frac{Z_G}{\hat{Z}}$;

(ii) $\frac{Z_{F^*}^{\mathrm{D}}}{Z_{F^*}^{\mathrm{M}}} \le \frac{1}{r\sqrt{\varepsilon L+1}}$;

(iii) If $Z_G \le \frac{1}{r}\hat{Z}$, then $\|\mu - \mu^*\|_{\mathrm{TV}} \le \frac{1}{16L}$;

(iv) If $Z_G \ge r\hat{Z}$, then $\|\mu - \mu^*\|_{\mathrm{TV}} \ge 1 - \varepsilon$.

The proof of Lemma 19 is provided in Section 3.2. We proceed first with the proof of Theorem 17 which follows along the lines of the proof of Theorem 8.

Proof of Theorem 17 Consider an antiferromagnetic Ising model on an N-vertex 3-regular graph $G = (V_G, E_G)$ with edge weight $\beta_G = -0.6$ on every edge and no external field; note that this model belongs to the family $\hat{\mathcal{M}}^{-}_{\text{ISING}}(N, 3, -0.6, 0)$. Let $\hat{Z} > 0$ be a real number, let $n = 4N^2 + N + 2$ and suppose there is an ε-identity testing algorithm for $\mathcal{M}^{-}_{\text{ISING}}(n, n, \beta_0, 0)$ with sample complexity $L = L(n) = \text{poly}(n)$ and running time $\text{poly}(n)$, where $\beta_0 > 0$ is a suitable constant we choose later. Let $r = 96\varepsilon^{-1}\sqrt{\varepsilon L + 1}$; we want to check whether $Z_G \leq \frac{1}{r}\hat{Z}$ or $Z_G \geq r\hat{Z}$ where $Z_G := Z_{G,\beta_G}$.

We construct the Ising models F and F^* with Gibbs distribution μ and μ^*, respectively as described above. We set $\beta_1 = 3$ and use the β_2 supplied by Lemma 19; hence the models F and F^* belong to $\mathcal{M}^{-}_{\text{ISING}}(n, n, \beta_0, 0)$, provided $\beta_0 \geq \max\{\beta_1, \beta_2\}$.

By Lemma 19 when $r2^N e^{-0.9N} \leq \hat{Z} \leq \frac{1}{r}2^N$, conditions (i) and (ii) from Theorem 14 are satisfied; condition (iii) is given by Lemma 18. Hence, we have an algorithm for the decision version of r-approximate counting for the Ising model on G for $\hat{Z}$ in this range. Otherwise, observe that the weight of every configuration is at least $e^{-0.9N}$, which corresponds to the weight of the monochromatic configuration, and at most 1. Thus, $2^N e^{-0.9N} \leq Z_G \leq 2^N$. If $\hat{Z} < r2^N e^{-0.9N} \leq rZ_G$, then we can output $\hat{Z} \leq \frac{1}{r}Z_G$. Similarly, $\hat{Z} > \frac{1}{r}2^N \geq \frac{1}{r}Z_G$ and we output $\hat{Z} \geq rZ_G$.

Thus, we have a $\text{poly}(N)$ running time algorithm for the decision version of r-approximate counting for $\hat{\mathcal{M}}^{-}_{\text{ISING}}(N, 3, -0.6, 0)$ where $N = \Theta(n^{1/2})$ and $r = \text{poly}(N)$, as desired. ■

3.2 Proof of Lemma 19

Our construction of the visible and hidden models is inspired by our construction in Section 2.2 for the ferromagnetic Potts model. In particular, the two vertices $\{s_1, s_2\}$ play the role of the complete graph H in our construction in Section 2.2.2. We partition $\Omega_F = \{+, -\}^{V_F}$ into two disjoint subsets $\Omega_F = \Omega_F^{\text{M}} \cup \Omega_F^{\text{D}}$, depending on whether $\sigma(s_1) = \sigma(s_2)$ (the *majority phase*) or $\sigma(s_1) \neq \sigma(s_2)$ (the *disordered phase*); more precisely, the set of *majority* configurations is given by

$$\Omega_F^{\text{M}} = \{\sigma \in \Omega_F : \sigma(s_1) = \sigma(s_2)\}$$

and the set of *disordered* configurations is

$$\Omega_F^{\text{D}} = \{\sigma \in \Omega_F : \sigma(s_1) \neq \sigma(s_2)\}.$$

The partition function for the majority phase is defined naturally as

$$Z_F^{\text{M}} = \sum_{\sigma \in \Omega_F^{\text{M}}} \exp\left(\sum_{\{u,v\} \in E_F} \beta_F(\{u, v\}) \mathbb{1}\{\sigma(u) = \sigma(v)\}\right),$$

and similarly for Z_F^{D}. Therefore, we have $Z_F = Z_F^{\text{M}} + Z_F^{\text{D}}$. In the same way, we also define the partition functions $Z_{F^*}^{\text{M}}$ and $Z_{F^*}^{\text{D}}$ for the hidden model on the graph F^* (notice that $\Omega_{F^*}^{\text{M}} = \Omega_F^{\text{M}}$ and $\Omega_{F^*}^{\text{D}} = \Omega_F^{\text{D}}$).

Proof of Lemma 19 Consider the following subset of configurations in Ω_F^{M} given by

$$\Omega_F^{\text{M}_0} = \left\{\sigma \in \Omega_F^{\text{M}} : \forall v \in V_G, \sigma(v) = \sigma(s_1)\right\}.$$

We also define the corresponding partition function $Z_F^{\mathrm{M}_0}$ and $Z_{F^*}^{\mathrm{M}_0}$ in the same way as above. We claim that $Z_F^{\mathrm{M}_0}$ (resp., $Z_{F^*}^{\mathrm{M}_0}$) is a good approximation (with only exponentially small error) of the partition function Z_F^{M} (resp., $Z_{F^*}^{\mathrm{M}}$) that we are interested in.

Claim 20 *If $\beta_1 \geq 3$, then $(1-e^{-2N})Z_F^{\mathrm{M}} \leq Z_F^{\mathrm{M}_0} \leq Z_F^{\mathrm{M}}$ and $(1-e^{-2N})Z_{F^*}^{\mathrm{M}} \leq Z_{F^*}^{\mathrm{M}_0} \leq Z_{F^*}^{\mathrm{M}}$.*

The proof of the following claim is postponed to the end of the section. We then derive explicit formula for Z_F^{D} and $Z_F^{\mathrm{M}_0}$. For configurations in Ω_F^{D}, every spin assignment to the vertices of G, s_1 and s_2 is multiplied by a $2e^{-\beta_1}(e^{-2\beta_1}+1)$ factor, corresponding to the weight of the edges $\{u_{v,j}^{(i)}, v\}$, $\{u_{v,j}^{(i)}, s_j\}$, $j \in \{1,2\}$ for every vertex $v \in V_G$ and every $1 \leq i \leq N$, and by a $3e^{-2\beta_2}+1$ factor for the edges $\{w_1^{(i)}, s_1\}$, $\{w_2^{(i)}, s_2\}$ and $\{w_1^{(i)}, w_2^{(i)}\}$ for every $1 \leq i \leq N^2$. For configurations in $\Omega_F^{\mathrm{M}_0}$, each monochromatic configuration on G is multiplied by a $(e^{-2\beta_1}+1)^2$ factor for the edges $\{u_{v,j}^{(i)}, v\}$, $\{u_{v,j}^{(i)}, s_j\}$, $j \in \{1,2\}$ for every vertex $v \in V_G$ and every $1 \leq i \leq N$, and by $e^{-3\beta_2}+3e^{-\beta_2}$ for the edges $\{w_1^{(i)}, s_1\}$, $\{w_2^{(i)}, s_2\}$ and $\{w_1^{(i)}, w_2^{(i)}\}$ for every $1 \leq i \leq N^2$. Thus, we obtain

$$Z_F^{\mathrm{D}} = 2\left(3e^{-2\beta_2}+1\right)^{N^2}\left(2e^{-\beta_1}\left(e^{-2\beta_1}+1\right)\right)^{N^2} Z_G;$$
$$Z_F^{\mathrm{M}_0} = 2\left(e^{-3\beta_2}+3e^{-\beta_2}\right)^{N^2}\left(e^{-2\beta_1}+1\right)^{2N^2} e^{-0.9N}.$$

Let $g(x) = (3e^{-2x}+1)/(e^{-3x}+3e^{-x})$ and recall that $\cosh x = \frac{1}{2}(e^x+e^{-x})$. We then deduce that

$$\frac{Z_F^{\mathrm{D}}}{Z_F^{\mathrm{M}_0}} = \left(\frac{g(\beta_2)}{\cosh\beta_1}\right)^{N^2} e^{0.9N} Z_G. \tag{21}$$

Now for $\beta_1 \geq 3$, we show that we can pick $\beta_2 > 0$ such that

$$\frac{1}{2\sqrt{\varepsilon L+1}}\frac{e^{-0.9N}}{\hat{Z}} \leq \left(\frac{g(\beta_2)}{\cosh\beta_1}\right)^{N^2} \leq \frac{1}{\sqrt{\varepsilon L+1}}\frac{e^{-0.9N}}{\hat{Z}}. \tag{22}$$

Such $\beta_2 > 0$ always exists and satisfies $\beta_2 < \beta_1 + 2$. To see this, we note that the function $g(x)$ is a continuous increasing function for $x \geq 0$ with $g(0) = 1$ and $g(\infty) = \infty$. Since $\hat{Z} \leq \frac{1}{r}2^N$, we get

$$\begin{aligned}\frac{1}{N^2}\log\left(\frac{1}{4\sqrt{\varepsilon L+1}}\frac{e^{-0.9N}}{\hat{Z}}\right) + \log(\cosh\beta_1) &\geq \frac{1}{N^2}\log\left(\frac{1}{4\sqrt{\varepsilon L+1}}\, r 2^{-N} e^{-0.9N}\right) + \beta_1 - 1\\ &\geq -\frac{2}{N} + 3 - 1 > 0,\end{aligned}$$

where the second inequality follows from $r/(4\sqrt{\varepsilon L+1}) = 6/\varepsilon \geq 1$. This shows that

$$\left(\frac{1}{2\sqrt{\varepsilon L+1}}\frac{e^{-0.9N}}{\hat{Z}}\right)^{\frac{1}{N^2}} \cosh\beta_1 \geq 1$$

and thus implies the existence of $\beta_2 > 0$. Meanwhile, since $\hat{Z} \geq r2^N e^{-0.9N}$ we have

$$\frac{1}{N^2}\log\left(\frac{1}{\sqrt{\varepsilon L+1}}\frac{e^{-0.9N}}{\hat{Z}}\right) + \log(\cosh\beta_1) \leq \frac{1}{N^2}\log\left(\frac{1}{\sqrt{\varepsilon L+1}}\frac{1}{r}2^{-N}\right) + \beta_1 < \beta_1,$$

where the second inequality follows from $r\sqrt{\varepsilon L+1} = 96\varepsilon^{-1}(\varepsilon L+1) \geq 1$. This shows that

$$e^{\beta_1} > g(\beta_2) = \frac{3e^{-2\beta_2}+1}{e^{-3\beta_2}+3e^{-\beta_2}} \geq \frac{1}{4e^{-\beta_2}} \geq e^{\beta_2-2}$$

and thus $\beta_2 < \beta_1 + 2$. Finally, we can compute a β_2 satisfying (22) in $\mathrm{poly}(n)$ time by, for example, the binary search algorithm.

Combining Claim 20 and equations (21) and (22), we deduce that

$$\frac{1}{4\sqrt{\varepsilon L+1}}\frac{Z_G}{\hat{Z}} \leq (1-e^{-2N})\frac{Z_F^{\mathrm{D}}}{Z_F^{\mathrm{M}_0}} \leq \frac{Z_F^{\mathrm{D}}}{Z_F^{\mathrm{M}}} \leq \frac{Z_F^{\mathrm{D}}}{Z_F^{\mathrm{M}_0}} \leq \frac{1}{\sqrt{\varepsilon L+1}}\frac{Z_G}{\hat{Z}}.$$

This shows the first part of the lemma. For part (ii), we can compute $Z_{F^*}^{\mathrm{D}}$ and $Z_{F^*}^{\mathrm{M}_0}$ in a similar fashion and obtain

$$\begin{aligned} Z_{F^*}^{\mathrm{D}} &= 2\left(3e^{-2\beta_2}+1\right)^{N^2}\left(2e^{-\beta_1}\left(e^{-2\beta_1}+1\right)\right)^{N^2}2^N; \\ Z_{F^*}^{\mathrm{M}_0} &= 2\left(e^{-3\beta_2}+3e^{-\beta_2}\right)^{N^2}\left(e^{-2\beta_1}+1\right)^{2N^2}. \end{aligned}$$

This gives

$$\frac{Z_{F^*}^{\mathrm{D}}}{Z_{F^*}^{\mathrm{M}_0}} = \left(\frac{g(\beta_2)}{\cosh\beta_1}\right)^{N^2}2^N. \tag{23}$$

Therefore, by equations (23) and (22) we obtain

$$\frac{Z_{F^*}^{\mathrm{D}}}{Z_{F^*}^{\mathrm{M}}} \leq \frac{Z_{F^*}^{\mathrm{D}}}{Z_{F^*}^{\mathrm{M}_0}} \leq \frac{1}{\sqrt{\varepsilon L+1}}\frac{e^{-0.9N}}{\hat{Z}}2^N \leq \frac{1}{r\sqrt{\varepsilon L+1}},$$

where the last inequality follows from the assumption $\hat{Z} \geq r2^N e^{-0.9N}$; thus, part (ii) follows.

Next, we derive parts (iii) and (iv). We define $\nu = \mu(\,\cdot\,|\Omega_F^{\mathrm{M}})$ to be the distribution conditioned on Ω_F^{M}, and similarly $\nu^* = \mu^*(\,\cdot\,|\Omega_{F^*}^{\mathrm{M}})$. By the definition of total variation distance we have

$$\|\mu-\nu\|_{\mathrm{TV}} = \left\|\mu-\mu(\,\cdot\,|\Omega_F^{\mathrm{M}})\right\|_{\mathrm{TV}} = \frac{Z_F^{\mathrm{D}}}{Z_F} = 1-\frac{Z_F^{\mathrm{M}}}{Z_F}.$$

For part (iii), if $Z_G \leq \frac{1}{r}\hat{Z}$, then we deduce from part (i) that

$$\|\mu-\nu\|_{\mathrm{TV}} \leq \frac{Z_F^{\mathrm{D}}}{Z_F^{\mathrm{M}}} \leq \frac{1}{\sqrt{\varepsilon L+1}}\frac{Z_G}{\hat{Z}} \leq \frac{1}{r\sqrt{\varepsilon L+1}} = \frac{\varepsilon}{96(\varepsilon L+1)} \leq \frac{1}{96L}.$$

Similarly, part (ii) implies

$$\|\mu^*-\nu^*\|_{\mathrm{TV}} \leq \frac{Z_{F^*}^{\mathrm{D}}}{Z_{F^*}^{\mathrm{M}}} \leq \frac{1}{r\sqrt{\varepsilon L+1}} = \frac{\varepsilon}{96(\varepsilon L+1)} \leq \min\left\{\frac{1}{96L}, \frac{\varepsilon}{96}\right\}.$$

Let $\rho = \nu(\,\cdot\,|\Omega_F^{\mathrm{M}_0})$ denote the conditional distribution of ν on $\Omega_F^{\mathrm{M}_0}$. Observe that ρ does not depend on the graph G, because we condition on the event that all vertices from G receive the

same spin, and thus the structure of G does not affect the conditional distribution ρ. In particular, we have $\rho = \nu(\,\cdot\,|\Omega_F^{\mathrm{M}_0}) = \nu^*(\,\cdot\,|\Omega_{F^*}^{\mathrm{M}_0})$. Then, Claim 20 implies that

$$\|\nu - \rho\|_{\mathrm{TV}} = 1 - \frac{Z_F^{\mathrm{M}_0}}{Z_F^{\mathrm{M}}} \leq e^{-2N}$$

and similarly $\|\nu^* - \rho\|_{\mathrm{TV}} \leq e^{-2N}$. Therefore, we obtain from the triangle inequality that

$$\|\nu - \nu^*\|_{\mathrm{TV}} \leq \|\nu - \rho\|_{\mathrm{TV}} + \|\nu^* - \rho\|_{\mathrm{TV}} \leq 2e^{-2N}.$$

We conclude again from the triangle inequality that

$$\|\mu - \mu^*\|_{\mathrm{TV}} \leq \|\mu - \nu\|_{\mathrm{TV}} + \|\mu^* - \nu^*\|_{\mathrm{TV}} + \|\nu - \nu^*\|_{\mathrm{TV}} \leq \frac{1}{96L} + \frac{1}{96L} + 2e^{-2N} \leq \frac{1}{16L}.$$

Finally, for part (iv), if $Z_G \geq r\hat{Z}$, then by part (ii) we have

$$\|\mu - \nu\|_{\mathrm{TV}} \geq 1 - \frac{Z_F^{\mathrm{M}}}{Z_F^{\mathrm{D}}} \geq 1 - 4\sqrt{\varepsilon L + 1}\,\frac{\hat{Z}}{Z_G} \geq 1 - \frac{4}{r}\sqrt{\varepsilon L + 1} = 1 - \frac{\varepsilon}{24}.$$

Hence,

$$\|\mu - \mu^*\|_{\mathrm{TV}} \geq \|\mu - \nu\|_{\mathrm{TV}} - \|\mu^* - \nu^*\|_{\mathrm{TV}} - \|\nu - \nu^*\|_{\mathrm{TV}} \geq 1 - \frac{\varepsilon}{24} - \frac{\varepsilon}{96} - 2e^{-2N} \geq 1 - \varepsilon,$$

as claimed. ■

Proof of Claim 20 For the first inequality, note that $Z_F^{\mathrm{M}_0} \leq Z_F^{\mathrm{M}}$. A union bound implies

$$\begin{aligned} 1 - \frac{Z_F^{\mathrm{M}_0}}{Z_F^{\mathrm{M}}} &= \Pr\left(\exists v \in V_G : \sigma(v) \neq \sigma(s_1) \middle| \sigma(s_1) = \sigma(s_2)\right) \\ &\leq \sum_{v \in V_G} \Pr(\sigma(v) \neq \sigma(s_1) | \sigma(s_1) = \sigma(s_2)). \end{aligned}$$

For every $\sigma \in \Omega_F^{\mathrm{M}}$ and $v \in V_G$, if $\sigma(v) \neq \sigma(s_1)$, then the total weight of edges incident to v is at most $(2e^{-\beta_1})^{2N}$; and if $\sigma(v) = \sigma(s_1)$, then it is at least $(e^{-2\beta_1} + 1)^{2N} \exp(\beta_G \deg_G(v)) \geq (e^{-2\beta_1} + 1)^{2N} e^{-1.8}$. Thus, we get

$$\begin{aligned} \Pr(\sigma(v) \neq \sigma(s_1) | \sigma(s_1) = \sigma(s_2)) &\leq \frac{(2e^{-\beta_1})^{2N}}{(2e^{-\beta_1})^{2N} + (e^{-2\beta_1} + 1)^{2N} e^{-1.8}} \\ &\leq e^{1.8} \left(\frac{2e^{-\beta_1}}{e^{-2\beta_1} + 1}\right)^{2N} \leq 10e^{-2(\beta_1 - 1)N} \leq 10e^{-4N}, \end{aligned}$$

where the last inequality follows from the assumption $\beta_1 \geq 3$. Therefore,

$$\frac{Z_F^{\mathrm{M}_0}}{Z_F^{\mathrm{M}}} \geq 1 - 10Ne^{-4N} \geq 1 - e^{-2N}.$$

The bound for F^* can be derived analogously. ■

4. Testing ferromagnetic RBMs with inconsistent fields

In this section, we establish our lower bound for identity testing for ferromagnetic RBMs with inconsistent fields; specifically, we prove Theorem 2 from the introduction. Let us formally define first the notions of consistent and inconsistent external fields.

Definition 21 *Consider an Ising model on a graph $G = (V_G, E_G)$ with external field $h_G : V_G \times \{1,2\} \to \mathbb{R}$. We say that the external field h_G is* consistent *if $\forall v \in V_G$, $h_G(v,1) \geq 0$ and $h_G(v,2) = 0$ or $\forall v \in V_G$, $h_G(v,2) \geq 0$ and $h_G(v,1) = 0$.*

We use once again our reduction strategy from r-approximate counting to testing. We start from the following well-known result.

Theorem 22 *(Goldberg and Jerrum, 2007) There is no* FPRAS *for the partition function of ferromagnetic Ising models with inconsistent fields, unless #BIS admits an* FPRAS.

The next step is the reduction from the decision version of approximate counting to identity testing.

Theorem 23 *Let $\varepsilon \in (0,1)$ be any constant. For every $\hat{\beta}, \hat{h} > 0$ there exist $\beta_0, h_0 > 0$ such that an ε-identity testing algorithm for $\mathcal{M}^{+}_{\text{ISING}}(n, n, \beta_0, h_0)$ with $\text{poly}(n)$ sample complexity and running time can be used to solve the decision r-approximate counting problem for $\hat{\mathcal{M}}^{+}_{\text{ISING}}(N, N, \hat{\beta}, \hat{h})$ in $\text{poly}(N)$ time, where $N = \Theta(\sqrt{n})$ and $r = \text{poly}(N)$.*

We can now provide the proof of Theorem 2.

Proof of Theorem 2 Follows from Theorems 22, 7, 23 and 28. ■

4.1 Reducing counting to testing for the ferromagnetic Ising model with an inconsistent field: proof of Theorem 23

Testing instance construction. Consider an instance (G, β_G, h_G) of ferromagnetic Ising models with an inconsistent field, where $G = (V_G, E_G)$ is the underlying graph with $N = |V_G|$, $\beta_G(e) = \hat{\beta} > 0$ for every $e \in E_G$, and at every vertex the external field is either $h_G = (\hat{h}, 0)$ or $h_G = (0, \hat{h})$ for $\hat{h} > 0$; that is, $\forall v \in V_G,\ h_G(v,j) = \mathbb{1}(j=1)\hat{h}$ for $j = \{1,2\}$ or $h_G(v,j) = \mathbb{1}(i=2)\hat{h}$ for $i = \{1,2\}$. Note that for consistency with the notation in the previous sections we use spins $\{1,2\}$ for the Ising model, instead of the usual "$+$" and "$-$" spins. Our goal is to give a r-approximate counting algorithm for the partition function $Z_G := Z_{G,\beta_G,h_G}$ for some $r = \text{poly}(N)$ using an identity testing algorithm.

Define F to be a graph with the vertex set

$$V_F = V_G \cup \{s_1, s_2\} \cup \left\{u^{(i)}_{v,j} : v \in V_G, 1 \leq i \leq N, j \in \{1,2\}\right\} \cup \left\{w^{(i)}_j : 1 \leq i \leq N^2, j \in \{1,2\}\right\}$$

and the edge set

$$\begin{aligned} E_F = E_G &\cup \left\{\{u^{(i)}_{v,j}, v\}, \{u^{(i)}_{v,j}, s_j\} : v \in V_G, 1 \leq i \leq N, j \in \{1,2\}\right\} \\ &\cup \left\{\{w^{(i)}_j, s_j\} : 1 \leq i \leq N^2, j \in \{1,2\}\right\}; \end{aligned}$$

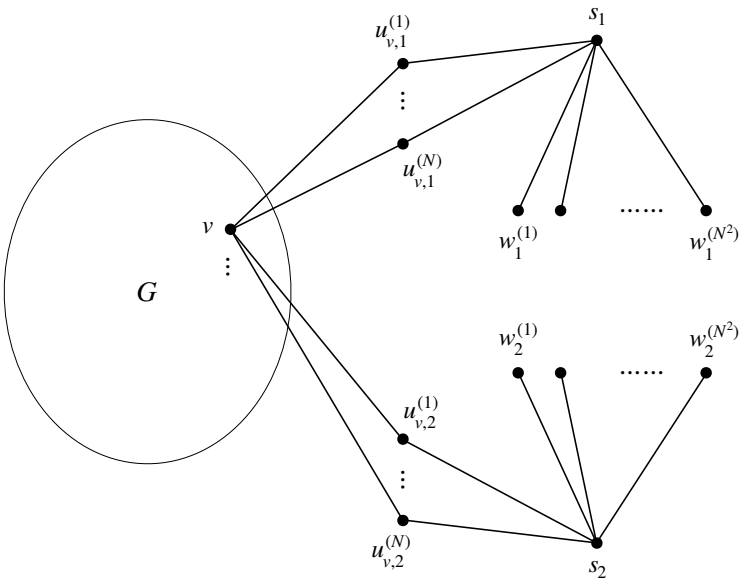

Figure 2: The graph F. For every vertex $v \in V_G$ and $j \in \{1,2\}$, v and s_j are connected by N disjoint paths of length 2. Also, each of s_1 and s_2 is adjacent to N^2 vertices with nonzero fields.

see Figure 2.

Given three real numbers $\beta_1, \beta_2, h > 0$, we then define a ferromagnetic Ising model on the graph F as follows:

1. Every edge $\{u, v\} \in E_G$ has weight $\hat{\beta}$ and every vertex $v \in V_G$ has external field given by h_G.

2. For every $v \in V_G$, $1 \le i \le N$ and $j \in \{1,2\}$, the two edges $\{u_{v,j}^{(i)}, v\}$ and $\{u_{v,j}^{(i)}, s_j\}$ have weight β_1;

3. For every $1 \le i \le N^2$ and $j \in \{1,2\}$, the edge $\{w_j^{(i)}, s_j\}$ has weight β_2;

4. For every $1 \le i \le N^2$, the vertex $w_1^{(i)}$ has external field $(h, 0)$ and the vertex $w_2^{(i)}$ has external field $(0, h)$; that is, $h_F(w_1^{(i)}, j) = \mathbb{1}(j = 1)h$ and $h_F(w_2^{(i)}, j) = \mathbb{1}(j = 2)h$.

Thus, F is a graph on $n = 4N^2 + N + 2$ vertices and the Ising model on F is ferromagnetic with an inconsistent external field. Let $\mu := \mu_F$ denote the corresponding Gibbs distribution.

For the hidden model F^*, we consider the same construction above but replacing G with a complete graph $K = K_N$ on N vertices where every edge has weight $\beta_K = \hat{\beta} + 4 \log 2 > 0$ and every vertex has the same field h_G as the Ising model on G. Let $\mu^* := \mu_{F^*}$ be the corresponding the Gibbs distribution. We note first that we can efficiently sample from μ^*.

Lemma 24 *There is an exact sampling algorithm for the distribution μ^* with running time* $\mathrm{poly}(n)$.

Proof Configurations in Ω_{F^*} are classified by their type, which is given by the spins of the vertices s_1 and s_2 and the number of vertices with spin 1 in the complete graph K_N. There are $4(N+1)$ types in total. Observe that configurations of each type have the same weight by symmetry, and this weight can be computed efficiently since given the spins of s_1, s_2 and K_N the remaining graph has only isolated vertices and edges. It is also straightforward to get the number of configurations of each type. Thus, to sample from μ^*, we first sample a type from the induced distribution on the types, and then sample a configuration of the given type uniformly at random. ■

Denote the sum of weights of two monochromatic configurations on G by

$$Z_G^{\mathrm{mo}} := \sum_{i\in\{1,2\}} \exp\left(\hat{\beta}|E_G| + \sum_{v\in V_G} h_G(v,i)\right).$$

The hidden and visible models F and F^* are related as follows.

Lemma 25 *Let $\varepsilon \in (0,1)$ be a constant, $L = L(n) = \mathrm{poly}(n)$ and $r = 96\varepsilon^{-1}\sqrt{\varepsilon L + 1}$. Suppose $\hat{Z} \in \mathbb{R}$ is such that $rZ_G^{\mathrm{mo}} \le \hat{Z} \le \frac{1}{r}\exp(\frac{1}{2}(\hat{\beta}+\hat{h}+1)N^2)$. Then, for any $\beta_1 \ge \frac{1}{2}(\hat{\beta}+\hat{h}+5)$, we can find $\beta_2 \in (0,\beta_1)$ in $\mathrm{poly}(n)$ time such that by setting $h = \beta_2$ al of the following holds:*

(i) $\frac{1}{4\sqrt{\varepsilon L+1}}\frac{Z_G}{\hat{Z}} \le \frac{Z_F^{\mathrm{D}}}{Z_F^{\mathrm{M}}} \le \frac{1}{\sqrt{\varepsilon L+1}}\frac{Z_G}{\hat{Z}}$;

(ii) $\frac{Z_{F^*}^{\mathrm{D}}}{Z_{F^*}^{\mathrm{M}}} \le \frac{2}{r\sqrt{\varepsilon L+1}}$;

(iii) If $Z_G \le \frac{1}{r}\hat{Z}$, then $\|\mu-\mu^*\|_{\mathrm{TV}} \le \frac{1}{16L}$;

(iv) If $Z_G \ge r\hat{Z}$, then $\|\mu-\mu^*\|_{\mathrm{TV}} \ge 1-\varepsilon$.

The proof of Lemma 25 is provided in Section 4.2. We provide next the proof of Theorem 23.

Proof of Theorem 23 Consider the ferromagnetic Ising model (G,β_G,h_G), where $G=(V_G,E_G)$ is an N-vertex graph, $\beta_G(e) = \hat{\beta}$ for all $e \in E_G$ and $h_G(v,j) = \mathbb{1}(j=1)\hat{h}$ for $j=\{1,2\}$ or $h_G(v,j) = \mathbb{1}(j=2)\hat{h}$ for $i=\{1,2\}$ for all $v \in V_G$; note that this model belongs to $\hat{\mathcal{M}}_{\mathrm{ISING}}^{+}(N,N,\hat{\beta},\hat{h})$. Let $\hat{Z}>0$ be a real number, let $n = 4N^2+N+2$ and suppose there is an ε-identity testing algorithm for $\mathcal{M}_{\mathrm{ISING}}^{+}(n,n,\beta_0,h_0)$ with sample complexity $L = L(n) = \mathrm{poly}(n)$ and running time $\mathrm{poly}(n)$, where $\beta_0, h_0 > 0$ are a suitable constants. Let $r = 96\varepsilon^{-1}\sqrt{\varepsilon L+1}$; we want to check whether $Z_G \le \frac{1}{r}\hat{Z}$ or $Z_G \ge r\hat{Z}$ where $Z_G := Z_{G,\beta_G,h_G}$.

We construct the Ising models F and F^* with Gibbs distribution μ and μ^*, respectively as described above, setting $\beta_1 = \frac{1}{2}(\hat{\beta}+\hat{h}+5)$, using the β_2 supplied by Lemma 25, and taking $h=\beta_2$; hence the models F and F^* belong to $\mathcal{M}_{\mathrm{ISING}}^{+}(n,n,\beta_0,h_0)$, provided $\beta_0 \ge \max\{\hat{\beta},\beta_K,\beta_1,\beta_2\}$ and $h_0 \ge \max\{\hat{h},h\}$.

By Lemma 25 when $rZ_G^{\mathrm{mo}} \le \hat{Z} \le \frac{1}{r}\exp(\frac{1}{2}(\hat{\beta}+\hat{h}+1)N^2)$, conditions (i) and (ii) of Theorem 14 are satisfied; condition (iii) is given by Lemma 24. Therefore, we have an algorithm for the decision version of r-approximate counting for the Ising model on G for $\hat{Z}$ in this range. When $\hat{Z}$ is not in this range, note that we have the following crude bounds on Z_G:

$$Z_G^{\mathrm{mo}} \le Z_G \le 2^N \cdot \exp\left(\hat{\beta}\frac{N^2}{2} + \hat{h}N\right) \le \exp\left(\frac{1}{2}(\hat{\beta}+\hat{h}+1)N^2\right).$$

Thus, if $\hat{Z} < rZ_G^{\mathrm{mo}} \le rZ_G$ we can output $\hat{Z} \le \frac{1}{r}Z_G$. Similarly, $\hat{Z} > \frac{1}{r}\exp(\frac{1}{2}(\hat{\beta}+\hat{h}+1)N^2) \ge \frac{1}{r}Z_G$ we output $\hat{Z} \ge rZ_G$.

Thus, we have a $\mathrm{poly}(N)$ running time algorithm for the decision version of r-approximate counting for $\hat{\mathcal{M}}^{+}_{\textsc{Ising}}(N, N, \hat{\beta}, \hat{h})$ where $N = \Theta(n^{1/2})$ and $r = \mathrm{poly}(N)$, as desired. ■

4.2 Proof of Lemma 25

We reuse the notation introduce in Section 3.2. Recall that $\Omega_F^{\mathrm{M}} = \{\sigma \in \Omega_F : \sigma(s_1) = \sigma(s_2)\}$ and $\Omega_F^{\mathrm{D}} = \{\sigma \in \Omega_F : \sigma(s_1) \neq \sigma(s_2)\}$. Also the partition function for the majority phase is given by

$$Z_F^{\mathrm{M}} = \sum_{\sigma \in \Omega_F^{\mathrm{M}}} \exp\left(\sum_{\{u,v\}\in E_F} \beta_F(\{u,v\})\mathbb{1}\{\sigma(u)=\sigma(v)\} + \sum_{v\in V_F} h_F(v,\sigma(v))\right)$$

and Z_F^{D} is defined similarly. The corresponding partition functions for the hidden model are denoted by $Z_{F^*}^{\mathrm{M}}$ and $Z_{F^*}^{\mathrm{D}}$.

Proof Let $\Omega_F^{\mathrm{M}_0} = \left\{\sigma \in \Omega_F^{\mathrm{M}} : \forall v \in V_G, \sigma(v) = \sigma(s_1)\right\}$ and consider restrictions of partition functions $Z_F^{\mathrm{M}_0}$ and $Z_{F^*}^{\mathrm{M}_0}$, as in the proof of Lemma 19. The following claim, whose proof is provided at the end of the section, has the same flavor as Claim 20.

Claim 26 *If* $\beta_1 \ge \frac{1}{2}(\hat{\beta}+\hat{h}+5)$, *then* $(1-e^{-2N})Z_F^{\mathrm{M}} \le Z_F^{\mathrm{M}_0} \le Z_F^{\mathrm{M}}$ *and* $(1-e^{-2N})Z_{F^*}^{\mathrm{M}} \le Z_{F^*}^{\mathrm{M}_0} \le Z_{F^*}^{\mathrm{M}}$.

We then derive explicit formulae for the two partition functions Z_F^{D} and $Z_F^{\mathrm{M}_0}$. For configurations $\sigma \in \Omega_F^{\mathrm{D}}$ with $\sigma(s_1) = 1$ and $\sigma(s_2) = 2$ (resp., $\sigma(s_1) = 2$ and $\sigma(s_2) = 1$), the weight of the configuration on G is multiply by a factor of $2e^{\beta_1}(e^{2\beta_1}+1)$ for each edge $\{u_{v,j}^{(i)}, v\}$, $\{u_{v,j}^{(i)}, s_j\}$, $j \in \{1,2\}$ for every $v \in V_G$ and every $1 \le i \le N$; it is also multiply by a $(e^{\beta_2+h}+1)^2$ (resp., $(e^{\beta_2}+e^h)^2$) factor for each edge $\{w_j^{(i)}, s_j\}$ and the vertex $w_j^{(i)}$, $j = \{1,2\}$ and $1 \le i \le N^2$. For configurations in $\Omega_F^{\mathrm{M}_0}$, both monochromatic configurations on G receive additional weight $(e^{2\beta_1}+1)^2$ for the edges $\{u_{v,j}^{(i)}, v\}$, $\{u_{v,j}^{(i)}, s_j\}$, $j = \{1,2\}$ for every vertex $v \in V_G$ and every $1 \le i \le N$, and a $(e^{\beta_2+h}+1)(e^{\beta_2}+e^h)$ factor for each edge $\{w_j^{(i)}, s_j\}$ and the vertex $w_j^{(i)}$, $j = \{1,2\}$ for every $1 \le i \le N^2$. Thus, we obtain that

$$Z_F^{\mathrm{D}} = \left[\left(e^{\beta_2+h}+1\right)^{2N^2} + \left(e^{\beta_2}+e^h\right)^{2N^2}\right]\left(2e^{\beta_1}\left(e^{2\beta_1}+1\right)\right)^{N^2} Z_G;$$
$$Z_F^{\mathrm{M}_0} = \left(e^{\beta_2+h}+1\right)^{N^2}\left(e^{\beta_2}+e^h\right)^{N^2}\left(e^{2\beta_1}+1\right)^{2N^2} Z_G^{\mathrm{mo}}.$$

Recall that $\cosh x = \frac{1}{2}(e^x + e^{-x})$. We then deduce that

$$\frac{Z_F^{\mathrm{D}}}{Z_F^{\mathrm{M}_0}} = \left[\left(\frac{\cosh(\frac{\beta_2+h}{2})}{\cosh(\frac{\beta_2-h}{2})}\right)^{N^2} + \left(\frac{\cosh(\frac{\beta_2-h}{2})}{\cosh(\frac{\beta_2+h}{2})}\right)^{N^2}\right]\left(\frac{1}{\cosh\beta_1}\right)^{N^2}\frac{Z_G}{Z_G^{\mathrm{mo}}}.$$

Since $\cosh x \geq 1$ for all $x \in \mathbb{R}$, let $h = \beta_2 > 0$ and then we get

$$\left(\frac{\cosh \beta_2}{\cosh \beta_1}\right)^{N^2} \frac{Z_G}{Z_G^{\mathrm{mo}}} \leq \frac{Z_F^{\mathrm{D}}}{Z_F^{\mathrm{M}_0}} \leq 2\left(\frac{\cosh \beta_2}{\cosh \beta_1}\right)^{N^2} \frac{Z_G}{Z_G^{\mathrm{mo}}}. \tag{24}$$

Now for $\beta_1 \geq \frac{1}{2}(\hat{\beta} + \hat{h} + 5)$, we pick $\beta_2 > 0$ such that

$$\frac{1}{3\sqrt{\varepsilon L + 1}} \frac{Z_G^{\mathrm{mo}}}{\hat{Z}} \leq \left(\frac{\cosh \beta_2}{\cosh \beta_1}\right)^{N^2} \leq \frac{1}{2\sqrt{\varepsilon L + 1}} \frac{Z_G^{\mathrm{mo}}}{\hat{Z}}, \tag{25}$$

Such $\beta_2 > 0$ always exists and satisfies $\beta_2 < \beta_1$. To see this, we note that since $\hat{Z} \leq \frac{1}{r}\exp(\frac{1}{2}(\hat{\beta} + \hat{h} + 1)N^2)$ and $Z_G^{\mathrm{mo}} \geq 2$, we have

$$\begin{aligned}\frac{1}{N^2}\log\left(\frac{1}{3\sqrt{\varepsilon L + 1}} \frac{Z_G^{\mathrm{mo}}}{\hat{Z}}\right) + \log(\cosh \beta_1) &\geq \frac{1}{N^2}\log\left(\frac{1}{3\sqrt{\varepsilon L + 1}} 2re^{-\frac{1}{2}(\hat{\beta}+\hat{h}+1)N^2}\right) + \beta_1 - 1 \\ &\geq -\frac{1}{2}(\hat{\beta} + \hat{h} + 1) + \frac{1}{2}(\hat{\beta} + \hat{h} + 5) - 1 = 1 > 0,\end{aligned}$$

where the second inequality follows from $2r/(3\sqrt{\varepsilon L + 1}) = 64/\varepsilon \geq 1$. This is equivalent to

$$\left(\frac{1}{3\sqrt{\varepsilon L + 1}} \frac{Z_G^{\mathrm{mo}}}{\hat{Z}}\right)^{\frac{1}{N^2}} \cosh \beta_1 \geq 1,$$

and hence $\beta_2 > 0$ satisfying (25) always exists and can be computed in $\mathrm{poly}(n)$ time. Note also that since $\hat{Z} \geq r Z_G^{\mathrm{mo}}$ we have

$$\frac{1}{2\sqrt{\varepsilon L + 1}} \frac{Z_G^{\mathrm{mo}}}{\hat{Z}} \leq \frac{1}{2r\sqrt{\varepsilon L + 1}} = \frac{\varepsilon}{192(\varepsilon L + 1)} < 1.$$

This shows that $\cosh \beta_2 / \cosh \beta_1 < 1$ and thus $\beta_2 < \beta_1$.

Combining Claim 26 and inequalities (24) and (25), we deduce that

$$\frac{1}{4\sqrt{\varepsilon L + 1}} \frac{Z_G}{\hat{Z}} \leq (1 - e^{-2N}) \frac{Z_F^{\mathrm{D}}}{Z_F^{\mathrm{M}_0}} \leq \frac{Z_F^{\mathrm{D}}}{Z_F^{\mathrm{M}}} \leq \frac{Z_F^{\mathrm{D}}}{Z_F^{\mathrm{M}_0}} \leq \frac{1}{\sqrt{\varepsilon L + 1}} \frac{Z_G}{\hat{Z}}.$$

This shows part (i). For part (ii), we can compute $Z_{F^*}^{\mathrm{D}}/Z_{F^*}^{\mathrm{M}_0}$ in a similar fashion and obtain

$$\left(\frac{\cosh \beta_2}{\cosh \beta_1}\right)^{N^2} \frac{Z_K}{Z_K^{\mathrm{mo}}} \leq \frac{Z_{F^*}^{\mathrm{D}}}{Z_{F^*}^{\mathrm{M}_0}} \leq 2\left(\frac{\cosh \beta_2}{\cosh \beta_1}\right)^{N^2} \frac{Z_K}{Z_K^{\mathrm{mo}}}. \tag{26}$$

Therefore, by inequalities (26) and (25) we obtain

$$\frac{Z_{F^*}^{\mathrm{D}}}{Z_{F^*}^{\mathrm{M}}} \leq \frac{Z_{F^*}^{\mathrm{D}}}{Z_{F^*}^{\mathrm{M}_0}} \leq \frac{1}{\sqrt{\varepsilon L + 1}} \frac{Z_G^{\mathrm{mo}}}{\hat{Z}} \frac{Z_K}{Z_K^{\mathrm{mo}}} \leq \frac{2}{r\sqrt{\varepsilon L + 1}},$$

where the last inequality follows from the assumption $\hat{Z} \geq r Z_G^{\mathrm{mo}}$ and the fact that $Z_K^{\mathrm{mo}}/Z_K \geq 1/2$ when $\beta_K \geq 4 \log 2$. Thus, part (ii) is established.

To establish part (iii), let us define $\nu = \mu(\,\cdot\,|\Omega_F^{\mathrm{M}})$ to be the distribution conditioned on Ω_F^{M}, and similarly $\nu^* = \mu^*(\,\cdot\,|\Omega_{F^*}^{\mathrm{M}})$. By the definition of total variation distance we have

$$\|\mu - \nu\|_{\mathrm{TV}} = \left\|\mu - \mu(\,\cdot\,|\Omega_F^{\mathrm{M}})\right\|_{\mathrm{TV}} = \frac{Z_F^{\mathrm{D}}}{Z_F} = 1 - \frac{Z_F^{\mathrm{M}}}{Z_F}.$$

For part (iii), if $Z_G \leq \frac{1}{r}\hat{Z}$, we deduce from part (i) that

$$\|\mu - \nu\|_{\mathrm{TV}} \leq \frac{Z_F^{\mathrm{D}}}{Z_F^{\mathrm{M}}} \leq \frac{1}{\sqrt{\varepsilon L + 1}} \frac{Z_G}{\hat{Z}} \leq \frac{1}{r\sqrt{\varepsilon L + 1}} = \frac{\varepsilon}{96(\varepsilon L + 1)} \leq \frac{1}{96L}.$$

Similarly, by part (ii) we have

$$\|\mu^* - \nu^*\|_{\mathrm{TV}} \leq \frac{Z_{F^*}^{\mathrm{D}}}{Z_{F^*}^{\mathrm{M}}} \leq \frac{2}{r\sqrt{\varepsilon L + 1}} = \frac{\varepsilon}{48(\varepsilon L + 1)} \leq \min\left\{\frac{1}{48L}, \frac{\varepsilon}{48}\right\}.$$

Let $\rho = \nu(\,\cdot\,|\Omega_F^{\mathrm{M}_0})$ denote the conditional distribution of ν on $\Omega_F^{\mathrm{M}_0}$. Observe that ρ does not depend on the graph G, because we condition on the event that all vertices from G receive the same spin, and thus the structure of G does not affect the conditional distribution ρ. In particular, we have $\rho = \nu(\,\cdot\,|\Omega_F^{\mathrm{M}_0}) = \nu^*(\,\cdot\,|\Omega_{F^*}^{\mathrm{M}_0})$. Then, Claim 26 implies that

$$\|\nu - \rho\|_{\mathrm{TV}} = 1 - \frac{Z_F^{\mathrm{M}_0}}{Z_F^{\mathrm{M}}} \leq e^{-2N}$$

and similarly $\|\nu^* - \rho\|_{\mathrm{TV}} \leq e^{-2N}$. Therefore, we obtain from the triangle inequality that

$$\|\nu - \nu^*\|_{\mathrm{TV}} \leq \|\nu - \rho\|_{\mathrm{TV}} + \|\nu^* - \rho\|_{\mathrm{TV}} \leq 2e^{-2N}.$$

We conclude again from the triangle inequality that

$$\|\mu - \mu^*\|_{\mathrm{TV}} \leq \|\mu - \nu\|_{\mathrm{TV}} + \|\mu^* - \nu^*\|_{\mathrm{TV}} + \|\nu - \nu^*\|_{\mathrm{TV}} \leq \frac{1}{96L} + \frac{1}{48L} + 2e^{-2N} \leq \frac{1}{16L}.$$

For part (iv), if $Z_G \geq r\hat{Z}$, then by part (i)

$$\|\mu - \nu\|_{\mathrm{TV}} \geq 1 - \frac{Z_F^{\mathrm{M}}}{Z_F^{\mathrm{D}}} \geq 1 - 4\sqrt{\varepsilon L + 1}\,\frac{\hat{Z}}{Z_G} \geq 1 - \frac{4}{r}\sqrt{\varepsilon L + 1} = 1 - \frac{\varepsilon}{24}.$$

Hence,

$$\|\mu - \mu^*\|_{\mathrm{TV}} \geq \|\mu - \nu\|_{\mathrm{TV}} - \|\mu^* - \nu^*\|_{\mathrm{TV}} - \|\nu - \nu^*\|_{\mathrm{TV}} \geq 1 - \frac{\varepsilon}{24} - \frac{\varepsilon}{48} - 2e^{-2N} \geq 1 - \varepsilon,$$

as claimed. ∎

Proof of Claim 26 Note first that $Z_F^{\mathrm{M}_0} \leq Z_F^{\mathrm{M}}$ and from a union bound we get

$$1 - \frac{Z_F^{\mathrm{M}_0}}{Z_F^{\mathrm{M}}} = \Pr\left(\exists v \in V_G : \sigma(v) \neq \sigma(s_1) \middle| \sigma(s_1) = \sigma(s_2)\right) \leq \sum_{v \in V_G} \Pr(\sigma(v) \neq \sigma(s_1) | \sigma(s_1) = \sigma(s_2)).$$

For every $\sigma \in \Omega_F^{\mathrm{M}}$ and $v \in V_G$, if $\sigma(v) \neq \sigma(s_1)$, then the total weight of edges incident to v is at most $(2e^{\beta_1})^{2N} \exp(\hat{\beta}(N-1) + \hat{h})$; and if $\sigma(v) = \sigma(s_1)$, then it is at least $(e^{2\beta_1} + 1)^{2N}$. Thus, we get

$$\begin{aligned}
\Pr(\sigma(v) \neq \sigma(s_1) | \sigma(s_1) = \sigma(s_2)) &\leq \frac{(2e^{\beta_1})^{2N} \exp(\hat{\beta}(N-1) + \hat{h})}{(2e^{\beta_1})^{2N} \exp(\hat{\beta}(N-1) + \hat{h}) + (e^{2\beta_1} + 1)^{2N}} \\
&\leq \left(\frac{2e^{\beta_1}}{e^{2\beta_1} + 1}\right)^{2N} \exp\left(\hat{\beta}(N-1) + \hat{h}\right) \\
&\leq \exp\left(-2(\beta_1 - 1)N\right) \cdot \exp\left((\hat{\beta} + \hat{h})N\right) \\
&\leq e^{-3N}.
\end{aligned}$$

where the last inequality follows from the assumption $\beta_1 \geq \frac{1}{2}(\hat{\beta} + \hat{h} + 5)$. Therefore, $\frac{Z_F^{\mathrm{M}_0}}{Z_F^{\mathrm{M}}} \geq 1 - Ne^{-3N} \geq 1 - e^{-2N}$. The bound for F^* is proved analogously. ■

5. Hardness of testing in bounded degree graphs

In this section, we provide a reduction from identity testing in bounded degree graphs to identity testing in general graphs. We introduce some convenient notation first. Recall that we use $\mathcal{M}_{\textsc{Potts}}(n, d, \beta, h)$ for the family of Potts models on n-vertex graphs with maximum degree at most d with the absolute value of the edge and vertex weights bounded by β and h, respectively; see Definition 4. We add "-Bip" to the subscript of this notation to denote the restriction to bipartite graphs; that is, $\mathcal{M}_{\textsc{Potts-Bip}}(n, d, \beta, h)$ denotes the set of models in $\mathcal{M}_{\textsc{Potts}}(n, d, \beta, h)$, where the underlying graphs is bipartite; note that $\mathcal{M}_{\textsc{Ising-Bip}} = \mathcal{M}_{\textsc{RBM}}$. Our reduction will also apply to Ising and Potts models with certain kinds of external fields, and so it is useful then to introduce the notion of *h-vertex-monochromatic* external fields.

Definition 27 *Consider a Potts model on a graph $G = (V_G, E_G)$ with external field $h_G : V_G \times [q] \to \mathbb{R}$. For $h \in \mathbb{R}$, we call h_G* h-vertex-monochromatic *if $|h_G(v, i)| \leq h$ for all $v \in V_G$, $i \in [q]$ and $|\{i \in [q] : h_G(v, i) \neq 0\}| \leq 1$ for all $v \in V_G$.*

In words, an h-vertex-monochromatic field is one that allows h_G to be non-zero (and at most h) for at most one spin at each vertex. We add "-Mono" to the subscript of $\mathcal{M}_{\textsc{Potts}}$ to denote the subfamily of models where the external field is h-vertex-monochromatic; namely, the models $\mathcal{M}_{\textsc{Potts-Mono}}(n, d, \beta, h)$ and $\mathcal{M}_{\textsc{Potts-Bip-Mono}}(n, d, \beta, h)$ respectively denote the subfamilies of models from $\mathcal{M}_{\textsc{Potts}}(n, d, \beta, h)$ and $\mathcal{M}_{\textsc{Potts-Bip}}(n, d, \beta, h)$ with h-vertex-monochromatic fields.

Theorem 28 *Let $\hat{n}, d \in \mathbb{N}^+$ be such that $3 \leq d \leq \hat{n}^{1-\rho}$ for some constant $\rho \in (0,1)$. Suppose that for some constants $\beta, h \geq 0$ there is no $\mathrm{poly}(n)$ running time ε-identity testing algorithm for $\mathcal{M}_{\text{Potts-Mono}}(n, n, \beta, h)$. Then there exists a constant $c \in (0,1)$ such that, for any constant $\hat{\varepsilon} > \varepsilon$ there is no $\mathrm{poly}(\hat{n})$ running time $\hat{\varepsilon}$-identity testing algorithm for $\mathcal{M}_{\text{Potts-Bip-Mono}}(\hat{n}, d, \hat{\beta}, \hat{h})$ provided $\hat{\beta} d = \omega(\log \hat{n})$ and $\hat{h} \leq h\hat{n}^{-c}$.*

Moreover, our reduction preserves ferromagnetism; that is, the statement remains true if we replace the family $\mathcal{M}_{\text{Potts-Mono}}$ by $\mathcal{M}^+_{\text{Potts-Mono}}$ and $\mathcal{M}_{\text{Potts-Bip-Mono}}$ by $\mathcal{M}^+_{\text{Potts-Bip-Mono}}$.

The proof of this theorem is fleshed out in the following sections. First in Section 5.1, we introduce our degree reducing gadget, which consists of a random bipartite graph of maximum degree d. In Section 5.2, we describe the construction of the testing instance (i.e., the reduction) and the actual proof of Theorem 28 is then finalized in Section 5.3.

5.1 A degree reducing gadget for the Potts model

Suppose $b, p, d, d_{\text{IN}}, d_{\text{OUT}}$ are positive integers such that $b \geq p$, $d \geq 3$ and $d_{\text{IN}} + d_{\text{OUT}} = d$. Let $B = (V_B, E_B)$ be the random bipartite graph defined as follows:

1. Set $V_B = L \cup R$, where $|L| = |R| = b$ and $L \cap R = \emptyset$;
2. Let P be subset of V_B 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 $E_B = \left(\bigcup_{i=1}^{d_{\text{IN}}} M_i\right) \cup \left(\bigcup_{i=1}^{d_{\text{OUT}}} M'_i\right)$;
6. Make the graph B simple by replacing multiple edges with single edges.

We use $\mathcal{G}(b, p, d_{\text{IN}}, d_{\text{OUT}})$ to denote the resulting distribution; that is, $B \sim \mathcal{G}(b, p, d_{\text{IN}}, d_{\text{OUT}})$. Vertices in P are called *ports*. Every port has degree at most d_{IN} while every non-port vertex has degree at most d. The set of ports P is chosen uniformly at random following (Bezáková et al., 2020), in order to use the expansion properties of $B \sim \mathcal{G}(b, p, d_{\text{IN}}, d_{\text{OUT}})$ proved there.

To capture the notion of an external configuration for the bipartite graph B, we assume that B is an induced subgraph of a larger graph $\mathbb{B} = (V_{\mathbb{B}}, E_{\mathbb{B}})$; i.e., $V_B \subset V_{\mathbb{B}}$ and $E_B \subset E_{\mathbb{B}}$. Let $\partial P = V_{\mathbb{B}} \setminus V_B$. We assume that every vertex in $P \subseteq V_B$ is connected to up to d_{OUT} vertices in ∂P and that there are no edges between $V_B \setminus P$ and ∂P in $\mathbb{B}$. Given a real number $\beta_B > 0$, we consider the Potts model on the graph $\mathbb{B}$ with:

1. edge interactions given by $\beta_{\mathbb{B}} : E_{\mathbb{B}} \to \mathbb{R}$, where $\max_{e \in E_{\mathbb{B}} \setminus E_B} |\beta_{\mathbb{B}}(e)| \leq \beta_B$ and $\beta_{\mathbb{B}}(e) = \beta_B$ for every $e \in E_B$;
2. an external field given by $h_{\mathbb{B}} : V_{\mathbb{B}} \times [q] \to \mathbb{R}$, where there exists $\kappa \in [q]$ and $h \in \mathbb{R}$ such that $h_{\mathbb{B}}(v, i) = h \cdot \mathbb{1}(i = \kappa) \cdot \mathbb{1}(v \in V_B)$.

We remark that the field $h_{\mathbb{B}}$ is h-vertex-monochromatic, but we also require that the spin for which the field is allowed to be not zero to be the same for all vertices.

Let $\sigma^i(B)$ be the configuration of $B = (V_B, E_B)$ where every vertex in V_B is assigned color $i \in [q]$. Let $\{\partial P = \tau\}$ denote the event that the configuration on ∂P is $\tau \in [q]^{\partial P}$. For certain choices of the random graph parameters we can show that for any τ, with high probability over B, the Potts configuration of V_B on $\mathbb{B}$ conditioned on $\{\partial P = \tau\}$ will likely be $\sigma^i(B)$ for some $i \in [q]$.

Theorem 29 *Suppose $3 \leq d = O_b(1)$, $d_{\text{IN}} = d - 1$, $d_{\text{OUT}} = 1$ and $p = \lfloor b^\alpha \rfloor$, where $\alpha \in (0, \frac{1}{4}]$ is a constant independent of b. Then, there exists a constant $\delta > 0$ such that with probability $1 - o(1)$ over the choice of the random graph B the following holds for every configuration τ on ∂P:*

$$\mu_{\mathbb{B}}\left(\bigcup_{i\in[q]}\{\sigma^i(B)\} \,\middle|\, \partial P = \tau\right) \geq \left(1 - \frac{q^2 e^{2h}}{e^{\delta\beta_B d}}\right)^{2b}.$$

Theorem 30 *Suppose $p = b$ and $4 + \frac{1200}{\rho} \leq d \leq b^{1-\rho}$ for some constant $\rho \in (0,1)$ independent of b. 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 B:*

$$\mu_{\mathbb{B}}\left(\bigcup_{i\in[q]}\{\sigma^i(B)\} \,\middle|\, \partial P = \tau\right) \geq \left(1 - \frac{q^2 e^{2h}}{e^{\delta\beta_B d}}\right)^{2b}.$$

These theorems are extensions of Theorems 4.1 and 4.2 in (Bezáková et al., 2020), where similar bounds are established for the case when every edge of $\mathbb{B}$ has the same weight $\beta < 0$; i.e., the antiferromagnetic setting. In this new setting, there is an external field, every edge in E_B has weight $\beta_B > 0$, and edges between P and ∂P are allowed to have either negative or positive weights bounded in absolute value by β_B.

To prove Theorems 29 and 30, we shall use the following facts about the expansion of the random graph $B \sim \mathcal{G}(b, p, d_{\text{IN}}, d_{\text{OUT}})$ proved in (Bezáková et al., 2020). For $S, T \subset V_B$ define

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

Theorem 31 (Theorem 17 (Bezáková et al., 2020)) *Suppose $p = b$ and $3 \leq d_{\text{IN}} \leq d \leq b^{1-\rho}$ where $\rho \in (0,1)$ is a constant independent of b. Then, with probability $1 - o(1)$ over the choice of the random graph B:*

$$\min_{\substack{S\subset V_B:\\ 0<|S|\leq b}} \frac{|E_B(S, V_B\backslash S)|}{|S|} \geq \frac{\rho d_{\text{IN}}}{300}.$$

Theorem 32 (Theorem 18 (Bezáková et al., 2020)) *Suppose $3 \leq d = O(1)$, $p = \lfloor b^\alpha \rfloor$ 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 b such that with probability $1 - o(1)$ over the choice of the random graph B:*

$$\min_{\substack{S\subset V_B:\\ 0<|S|\leq b}} \frac{|E_B(S, V_B\backslash S)|}{|S|} \geq \gamma d.$$

Theorem 33 (Theorem 19 (Bezáková et al., 2020)) *Suppose* $3 \le d = O(1)$, $p = \lfloor b^\alpha \rfloor$ *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* b *such that with probability* $1 - o(1)$ *over the choice of the random graph* B*:*

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

Proof of Theorems 29 and 30 Let $E(S,T)$ denote the set of edges between S and T in $E_{\mathbb{B}}$. For ease of notation, we set $\beta = \beta_B$. Let $P_i \subseteq \partial P$ be the set of vertices of ∂P that are assigned color $i \in [q]$ in τ. The weight of $\sigma^i(B)$ in $\mathbb{B}$ conditional on τ is then given by

$$w^i := w^\tau_{\mathbb{B}}(\sigma^i(B)) = \exp\left[\beta db + 2bh\mathbb{1}(i = \kappa) + \sum_{e \in E(P, P_i)} \beta_{\mathbb{B}}(e)\right]. \tag{27}$$

Let Ω_B be the set of Potts configurations of the graph B. For $\sigma \in \Omega_B$, let $S_\sigma(i) \subseteq V_B$ be the set of vertices that are assigned color $i \in [q]$ in σ. We let S_σ denote the set of maximum cardinality among $S_\sigma(1), \dots, S_\sigma(q)$. Let $\Omega^i_B \subseteq \Omega_B$ be the set of configurations σ such that $S_\sigma = S_\sigma(i)$. For $\sigma \in \Omega_B$, we use $w^\tau(\sigma)$ for the weight of the configuration on $\mathbb{B}$ that agrees with σ on V_B and with τ on $V_{\mathbb{B}} \setminus V_B$. By definition, the partition function $Z^\tau_{\mathbb{B}}$ for the conditional distribution $\mu_{\mathbb{B}}(\cdot \mid \partial P = \tau)$ satisfies

$$Z^\tau_{\mathbb{B}} = \sum_{\sigma \in \Omega_B} w^\tau(\sigma) = \sum_{\sigma \in \Omega_B : |S_\sigma| > b} w^\tau(\sigma) + \sum_{\sigma \in \Omega_B : |S_\sigma| \le b} w^\tau(\sigma). \tag{28}$$

We bound each term in the right-hand side of (28) separately. For $\sigma \in \Omega_B$, let $r(\sigma, i) = |S_\sigma(\kappa)|h - 2bh\mathbb{1}(i = \kappa)$. We will show that in the regimes of parameters in Theorems 29 and 30, with probability $1 - o(1)$ over the choice of the random graph $B \sim \mathcal{G}(b, p, d_{\text{IN}}, d_{\text{OUT}})$, there exists a constant $\delta > 0$ such that for every $\sigma \in \Omega_B$:

$$w^\tau(\sigma) \le w^i \cdot e^{-\delta\beta d|V_B \backslash S_\sigma(i)| + r(\sigma, i)} \text{ when } \sigma \in \Omega^i_B,\ |S_\sigma| > b; \text{ and} \tag{29}$$

$$w^\tau(\sigma) \le w^i \cdot e^{-\delta\beta db + r(\sigma, i)} \text{ when } \sigma \in \Omega^i_B,\ |S_\sigma| \le b. \tag{30}$$

Before proving these two bounds, we show how to use them to complete the proofs of the theorems. From (29), we get

$$\sum_{\sigma \in \Omega_B : |S_\sigma| > b} w^\tau(\sigma) = \sum_{i=1}^q \sum_{\sigma \in \Omega^i_B : |S_\sigma| > b} w^\tau(\sigma) \le \sum_{i=1}^q \sum_{\sigma \in \Omega^i_B : |S_\sigma| > b} w^i \cdot e^{-\delta\beta d|V_B \backslash S_\sigma(i)| + r(\sigma, i)}.$$

If $i = \kappa$,

$$\begin{aligned}\sum_{\sigma \in \Omega^\kappa_B : |S_\sigma| > b} e^{-\delta\beta d|V_B \backslash S_\sigma(\kappa)| + r(\sigma, \kappa)} &= \sum_{\sigma \in \Omega^\kappa_B : |S_\sigma| > b} e^{-\delta\beta d|V_B \backslash S_\sigma(\kappa)| - h|V_B \backslash S_\sigma(\kappa)|} \\ &= \sum_{x=0}^{b} \binom{2b}{x} (q-1)^x e^{-(\delta\beta d + h)x} \le \left(1 + \frac{q-1}{e^{\delta\beta d + h}}\right)^{2b}.\end{aligned}$$

If $i \neq \kappa$,

$$\begin{aligned}\sum_{\sigma\in\Omega_B^i:|S_\sigma|>b} e^{-\delta\beta d|V_B\setminus S_\sigma(i)|+r(\sigma,i)} &= \sum_{\sigma\in\Omega_B^i:|S_\sigma|>b} e^{-\delta\beta d|V_B\setminus S_\sigma(i)|+h|S_\sigma(\kappa)|}\\ &= \sum_{x=0}^{b}\sum_{y=0}^{x}\binom{2b}{x}\binom{x}{y}(q-2)^{x-y}e^{-\delta\beta dx+hy}\\ &\leq \sum_{x=0}^{b}\binom{2b}{x}(q-2)^x e^{-\delta\beta dx}\left(1+\frac{e^h}{q-2}\right)^x\\ &\leq \left(1+\frac{q-2+e^h}{e^{\delta\beta d}}\right)^{2b}.\end{aligned}$$

Hence, letting $W=\sum_{i=1}^q w^i$, we obtain

$$\sum_{\sigma\in\Omega_B:|S_\sigma|>b} w^\tau(\sigma) \leq w^\kappa\left(1+\frac{q-1}{e^{\delta\beta d+h}}\right)^{2b} + (W-w^\kappa)\left(1+\frac{q-2+e^h}{e^{\delta\beta d}}\right)^{2b}.$$

To bound the second summand from (28), note that from (30) we get

$$\begin{aligned}\sum_{\sigma:|S_\sigma|\leq b} w^\tau(\sigma) &= \sum_{i=1}^q \sum_{\sigma\in\Omega_B^i:|S_\sigma|\leq b} w^\tau(\sigma) \leq \sum_{i=1}^q \sum_{\sigma\in\Omega_B^i:|S_\sigma|\leq b} w^i\cdot e^{-\delta\beta db+r(\sigma,i)}\\ &\leq \sum_{i=1}^q w^i\cdot e^{-\delta\beta db}\sum_{\sigma\in\Omega_B^i:|S_\sigma|\leq b} e^{|S_\sigma(\kappa)|h}\\ &\leq W\cdot e^{-\delta\beta db}\sum_{x=0}^{2b}\binom{2b}{x}e^{xh}(q-1)^{2b-x}\\ &\leq W\cdot\left(\frac{(q-1+e^h)^2}{e^{\delta\beta d}}\right)^b.\end{aligned}$$

Thus,

$$\begin{aligned}Z_{\mathbb{B}}^\tau &\leq w^\kappa\left(1+\frac{q-1}{e^{\delta\beta d+h}}\right)^{2b} + (W-w^\kappa)\left(1+\frac{q-2+e^h}{e^{\delta\beta d}}\right)^{2b} + \left(\frac{(q-1+e^h)^2}{e^{\delta\beta d}}\right)^b W\\ &\leq W\left[\left(1+\frac{q-2+e^h}{e^{\delta\beta d}}\right)^{2b} + \left(\frac{(q-1+e^h)^2}{e^{\delta\beta d}}\right)^b\right].\end{aligned}$$

Setting $x=\frac{q-2+e^h}{e^{\delta\beta d}}$, $y=\frac{(q-1+e^h)^2}{e^{\delta\beta d}}$ and $z=\frac{q^2e^{2h}}{e^{\delta\beta d}}$

$$\mu_{\mathbb{B}}\left(\bigcup_{i\in[q]}\{\sigma^i(B)\}\mid \partial P=\tau\right)=\frac{W}{Z_{\mathbb{B}}^\tau}\geq\frac{1}{(1+x)^{2b}+y^b}\geq\frac{1}{(1+2z)^{2b}}\geq(1-2z)^{2b}$$

as claimed.

It remains for us to establish (29) and (30); we start with (29). For $S, T \subseteq V_B \cup \partial P$, let $[S, T]$ denote the number of edges between S and T in the graph $\mathbb{B}$. Then,

$$w^\tau(\sigma) = \exp\left[\beta \sum_{j=1}^{q} [S_\sigma(j), S_\sigma(j)] + \sum_{j=1}^{q} \sum_{e \in E(P_j, S_\sigma(j) \cap P)} \beta_{\mathbb{B}}(e) + h|S_\sigma(\kappa)|\right]. \tag{31}$$

Now, $\sum_{j=1}^{q} [S_\sigma(j), S_\sigma(j)] \leq db - [S_\sigma, V_B \setminus S_\sigma]$ and for any $i \in [q]$

$$\begin{aligned}\sum_{j=1}^{q} \sum_{e \in E(P_j, S_\sigma(j) \cap P)} \beta_{\mathbb{B}}(e) - \sum_{e \in E(P, P_i)} \beta_{\mathbb{B}}(e) &= \sum_{j \neq i} \sum_{e \in E(P_j, S_\sigma(j) \cap P)} \beta_{\mathbb{B}}(e) - \sum_{e \in E(P \setminus S_\sigma(i), P_i)} \beta_{\mathbb{B}}(e) \\ &\leq \beta \sum_{j \neq i} [S_\sigma(j) \cap P, P_j \cup P_i].\end{aligned}$$

Plugging these two bounds into (31) and using (27), we get for $\sigma \in \Omega_B^i$

$$\begin{aligned} w^\tau(\sigma) &\leq \exp\left[\beta(db - [S_\sigma, V_B \setminus S_\sigma]) + \beta \sum_{j \neq i} [S_\sigma(j) \cap P, P_j \cup P_i] + \sum_{e \in E(P, P_i)} \beta_{\mathbb{B}}(e) + h|S_\sigma(\kappa)|\right] \\ &= w^i \cdot \exp\left[-\beta [S_\sigma, V_B \setminus S_\sigma] + \beta \sum_{j \neq i} [S_\sigma(j) \cap P, P_j \cup P_i] + r(\sigma, i)\right] \\ &\leq w^i \cdot \exp\left[-\beta \left([S_\sigma, V_B \setminus S_\sigma] - [(V_B \setminus S_\sigma) \cap P, \partial P]\right) + r(\sigma, i)\right]. \end{aligned} \tag{32}$$

When $3 \leq d = O_b(1)$, $p = \lfloor b^\alpha \rfloor$ with $\alpha \in (0, \frac{1}{4}]$, $d_{\text{IN}} = d - 1$ and $d_{\text{OUT}} = 1$. Hence, $[V_B \setminus S_\sigma \cap P, \partial P] = |(V_B \setminus S_\sigma) \cap P|$. Theorems 32 and 33 imply that there exists a constant $\gamma > 0$ such that with probability $1 - o(1)$ over the choice of the random graph B we have

$$\begin{aligned} \frac{[S_\sigma, V_B \setminus S_\sigma]}{|V_B \setminus S_\sigma|} &\geq \gamma d, \text{ and} \\ \frac{[S_\sigma, V_B \setminus S_\sigma]}{|(V_B \setminus S_\sigma) \cap P|} &\geq 1 + \gamma. \end{aligned}$$

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

$$[S_\sigma, V_B \setminus S_\sigma] \geq |(V_B \setminus S_\sigma) \cap P| + \delta d |V_B \setminus S_\sigma|.$$

Plugging this bound into (32),

$$w^\tau(\sigma) \leq w^i \cdot \exp\left[-\delta \beta d |V_B \setminus S_\sigma| + r(\sigma, i)\right], \tag{33}$$

and we get (29), since $\sigma \in \Omega_B^i$ and so $S_\sigma = S_\sigma(i)$.

Under the assumptions in Theorem 30, we can also establish (33) as follows. When $b^{1-\rho} \geq d \geq d_{\text{IN}} = \lfloor \theta d \rfloor \geq 3$, Theorem 31 implies that

$$[S_\sigma, V_B \setminus S_\sigma] \geq \frac{\rho d_{\text{IN}}}{300} |V_B \setminus S_\sigma| = \frac{\rho \lfloor \theta d \rfloor}{300} |V_B \setminus S_\sigma|.$$

Moreover,

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

Hence, taking $\theta = \frac{300+0.75\rho}{300+\rho}$ we get that when $d \ge 4 + \frac{1200}{\rho}$:

$$\frac{\rho \lfloor \theta d \rfloor}{300} - (d - \lfloor \theta d \rfloor) \ge \frac{\rho d}{600}. \tag{34}$$

Together with (32) this implies

$$w^\tau(\sigma) \le w^i \cdot \exp\left[-\frac{\rho \beta d |V_B \setminus S_\sigma|}{600} + r(\sigma, i) \right],$$

which gives (33) for $\delta \le \rho/600$, and thus we again obtain (29). (Observe that our choice of θ guarantees $d - 1 \ge d_{\text{IN}} = \lfloor \theta d \rfloor \ge 3$ for all $d \ge 4$.)

We establish (30) next. Since

$$\sum_{j=1}^{q} [S_\sigma(j), S_\sigma(j)] = bd - \frac{1}{2} \sum_{j=1}^{q} [S_\sigma(j), V_B \setminus S_\sigma(j)],$$

and

$$\sum_{j=1}^{q} \sum_{e \in E(P_j, S_\sigma(j) \cap P)} \beta_{\mathbb{B}}(e) - \sum_{e \in E(P, P_i)} \beta_{\mathbb{B}}(e) \le \beta d_{\text{OUT}} |P|,$$

we get from (27) and (31) that for $\sigma \in \Omega_B^i$

$$\begin{aligned} w^\tau(\sigma) &\le w^i \cdot \exp\left[-\frac{\beta}{2} \sum_{j=1}^{q} [S_\sigma(j), V_B \setminus S_\sigma(j)] + \beta d_{\text{OUT}} |P| + r(\sigma, i) \right] \\ &\le w^i \cdot \exp\left[-\beta \left(\frac{1}{2} \sum_{j=1}^{q} [S_\sigma(j), V_B \setminus S_\sigma(j)] - d_{\text{OUT}} |P| \right) + r(\sigma, i) \right]. \end{aligned} \tag{35}$$

Since $|S_\sigma(j)| \le b$ for $j \in [q]$, our assumptions in Theorem 29 combined with Theorem 32 imply that there exists a constant $\gamma > 0$ such that with probability $1 - o(1)$ over the choice of the random graph B we have for all $j \in [q]$

$$\frac{[S_\sigma(j), V_B \setminus S_\sigma(j)]}{|S_\sigma(j)|} \ge \gamma d.$$

Plugging this bound into (35), and since $d_{\text{OUT}} = 1$ by assumption, we get

$$\begin{aligned} w^\tau(\sigma) &\le w^i \cdot \exp\left[-\beta \left(\frac{1}{2} \sum_{j=1}^{q} \gamma d |S_\sigma(j)| - |P| \right) + r(\sigma, i) \right] \\ &= w^i \cdot \exp\left[-\beta \left(\gamma d b - |P| \right) + r(\sigma, i) \right] \le w^i \cdot \exp\left[-\beta \delta d b + r(\sigma, i) \right], \end{aligned}$$

where the last inequality holds for a suitable constant $\delta > 0$ and b sufficiently large since $|P| \le \lfloor b^{1/4} \rfloor$.

Finally, the assumptions in Theorem 30 and Theorem 31 imply that

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

Hence, since $|P| = b$

$$\begin{aligned} w^\tau(\sigma) &\leq w^i \cdot \exp\left[-\beta\left(\frac{1}{2}\sum_{j=1}^{q}\frac{\rho\lfloor\theta d\rfloor}{300}|S_\sigma(j)| - (d - \lfloor\theta d\rfloor)|P|\right) + r(\sigma, i)\right] \\ &\leq w^i \cdot \exp\left[-\beta b\left(\frac{\rho\lfloor\theta d\rfloor}{300} - (d - \lfloor\theta d\rfloor)\right) + r(\sigma, i)\right] \\ &\leq w^i \cdot \exp\left[-\beta\delta d b + r(\sigma, i)\right], \end{aligned}$$

where the last inequality holds for a suitable constant $\delta > 0$ for θ satisfying (34). This completes the proofs of the theorem. ■

5.2 Testing instance construction

Consider a Potts model on an n-vertex graph $G = (V_G, E_G)$, with edge interactions $\beta_G : E_G \to \mathbb{R}$ and an h-vertex-monochromatic external field $h_G : V_G \times [q] \to \mathbb{R}$; see Definition 27. We show how to construct a Potts model on a larger graph of maximum degree at most d, with edge interactions bounded by $\hat{\beta}$ and an $\hat{h}$-vertex-monochromatic external field whose distribution captures that of the model (G, β_G, h_G). We can think of d, $\hat{\beta}$ and $\hat{h}$ as the parameters for our construction.

We use an instance of the random bipartite graph $\mathcal{G}(b, p, d_{\text{IN}}, d_{\text{OUT}})$ from Section 5.1 as a gadget to define a simple graph $G_\Gamma = (V_{G_\Gamma}, E_{G_\Gamma})$, where Γ denotes the set parameters $\{b, p, d_{\text{IN}}, d_{\text{OUT}}\}$. The graph G_Γ is constructed as follows:

1. Generate an instance $B = (V_B, E_B)$ of the random graph model $\mathcal{G}(b, p, d_{\text{IN}}, d_{\text{OUT}})$;
2. Replace every vertex v of G by a copy $B_v = (L_v \cup R_v, E_{B_v})$ of the generated instance B;
3. For every edge $e = \{v, u\} \in E_G$, let $\ell(e) = \lceil |\beta_G(e)|/\hat{\beta} \rceil$ and choose $d_{\text{OUT}} \cdot \lceil \ell(e)/d_{\text{OUT}}^2 \rceil$ unused ports in L_v, $d_{\text{OUT}} \cdot \lceil \ell(e)/d_{\text{OUT}}^2 \rceil$ unused ports in R_u and connect them with any simple bipartite graph of maximum degree at most d_{OUT} and exactly $\ell(e)$ edges;
4. Similarly, for every edge $e = \{v, u\} \in E_G$, choose $d_{\text{OUT}} \cdot \lceil \ell(e)/d_{\text{OUT}}^2 \rceil$ unused ports in R_v and $d_{\text{OUT}} \cdot \lceil \ell(e)/d_{\text{OUT}}^2 \rceil$ unused ports in L_u and connect them with any simple bipartite graph of maximum degree at most d_{OUT} and exactly $\ell(e)$ edges;

Let d_G be the maximum degree of the graph G. Our construction requires:

$$d_{\text{IN}} + d_{\text{OUT}} = d \leq b, \tag{36}$$

$$d_G \cdot \left(d_{\text{OUT}} \cdot \max_{e \in E_G}\left\lceil \frac{\ell(e)}{d_{\text{OUT}}^2} \right\rceil\right) \leq p. \tag{37}$$

Observe also that there is always a simple bipartite graph of maximum degree at most d_{OUT} and exactly $\ell(e)$ edges for steps 3 and 4; take, for example, $\lfloor \ell(e)/d_{\text{OUT}}^2 \rfloor$ disjoint copies of the complete

bipartite graph with d_{OUT} vertices on each side, and add one additional bipartite graph with d_{OUT} vertices on each side for the remaining edges when $\ell(e)/d_{\text{OUT}}^2$ is not an integer.

We consider the Potts model on the graph $(V_{G_\Gamma}, E_{G_\Gamma})$ with edge weights $\beta_{G_\Gamma} : E_{G_\Gamma} \to \mathbb{R}$ and external field $h_{G_\Gamma} : V_{G_\Gamma} \times [q] \to \mathbb{R}$ defined as follows:

1. each edge with both of its endpoints in the same gadget is assigned weight $\beta_B := \hat{\beta}$;
2. if the edge connects the gadgets corresponding to $u \neq v \in V_G$, then it is assigned weight $\frac{\beta_G(\{u,v\})}{2\ell(\{u,v\})}$.
3. for each vertex $v \in V_G$, every vertex u in the gadget B_v is assigned the field $h_{G_\Gamma}(u,i) := h_G(v,i)/2b$ for $i \in [q]$.

Note that if h_G is h-vertex-monochromatic, then h_{G_Γ} is $(h/2b)$-vertex-monochromatic, and that in the gadget of every vertex only one spin may receive a non-zero weight; in particular, if h_G is h-vertex-monochromatic, then the field in every gadget would satisfy the conditions Section 5.1.

For a configuration σ on G_Γ, we say that the gadget $B_v = (V_{B_v}, E_{B_v})$ is in the *i-th phase* if all the vertices in V_{B_v} are assigned spin $i \in \{1, \dots, q\}$. Let Ω_{good} be the set of configurations of G_Γ where the gadget of every vertex is in a phase (not necessarily the same). The set of all Potts configurations of G_Γ is denoted by Ω. We use Z_{G_Γ} for the partition function of the Potts model on G_Γ and $Z_{G_\Gamma}(\Lambda)$ for its restriction to a subset of configurations $\Lambda \subseteq \Omega$. That is, $Z_{G_\Gamma} = \sum_{\sigma\in\Omega} w_{G_\Gamma}(\sigma)$ and $Z_{G_\Gamma}(\Lambda) = \sum_{\sigma\in\Lambda} w_{G_\Gamma}(\sigma)$ where

$$w_{G_\Gamma}(\sigma) := \exp\left[\sum_{\{u,v\}\in E_{G_\Gamma}} \beta_{G_\Gamma}(\{u,v\}) \cdot \mathbb{1}(\sigma(u) = \sigma(v)) + \sum_{v\in V_{G_\Gamma}} h_{G_\Gamma}(v, \sigma(v))\right]$$

is the weight of the configuration σ.

For a configuration $\sigma \in \Omega_{\text{good}}$, let σ_G be the corresponding configuration on G where $\sigma_G(v)$ is set to the phase of gadget B_v in σ. Let μ_G and μ_{G_Γ} denote the Gibbs distribution for the Potts models we just defined on G and G_Γ. From our construction, we can deduce the following fact.

Lemma 34 *For any graph G, we have $\mu_{G_\Gamma}(\sigma \mid \sigma \in \Omega_{\text{good}}) = \mu_G(\sigma_G)$.*

Proof Let $Q \subseteq E_{G_\Gamma}$ be the edges of G_Γ that connect vertices between different gadgets. Then, for $\sigma \in \Omega_{\text{good}}$,

$$\sum_{\{u,v\}\in Q} \beta_{G_\Gamma}(\{u,v\})\mathbb{1}(\sigma(u) = \sigma(v)) = \sum_{\{u',v'\}\in E_G} \beta_G(\{u',v'\})\mathbb{1}(\sigma_G(u') = \sigma_G(v')),$$

$$\sum_{\{u,v\}\in E_{G_\Gamma}\setminus Q} \beta_{G_\Gamma}(\{u,v\})\mathbb{1}(\sigma(u) = \sigma(v)) = \exp\left(\beta_B d_{\text{IN}} bn\right), \text{ and}$$

$$\sum_{v\in V_{G_\Gamma}} h_{G_\Gamma}(v, \sigma(v)) = \sum_{v'\in V_G} h_G(v', \sigma_G(v')).$$

Thus, $w_{G_\Gamma}(\sigma) = w_G(\sigma_G)\exp\left(\beta_B d_{\text{IN}} bn\right)$, and

$$\mu_{G_\Gamma}(\sigma \mid \sigma \in \Omega_{\text{good}}) = \frac{w_{G_\Gamma}(\sigma)}{Z_{G_\Gamma}(\Omega_{\text{good}})} = \frac{w_G(\sigma_G)\exp\left(\beta_B d_{\text{IN}} bn\right)}{Z_G \exp\left(\beta_B d_{\text{IN}} bn\right)} = \mu_G(\sigma_G).$$

■

Lemma 35 *Let (G, β_G, h_G) and $(G^*, \beta_{G^*}, h_{G^*})$ be two Potts on the n-vertex graphs G and G^*, respectively. Let $\Gamma = (b, p, d_{\text{IN}}, d_{\text{OUT}})$ be such that conditions* (36) *and* (37) *are satisfied. Suppose that $\mu_{G_\Gamma}(\Omega_{\text{good}}) \geq 1 - \delta$ and $\mu_{G^*_\Gamma}(\Omega_{\text{good}}) \geq 1 - \delta$ for some $\delta \in (0, 1)$. Then,*

$$\|\mu_G - \mu_{G^*}\|_{\text{TV}} - 2\delta \leq \left\|\mu_{G_\Gamma} - \mu_{G^*_\Gamma}\right\|_{\text{TV}} \leq \|\mu_G - \mu_{G^*}\|_{\text{TV}} + 2\delta.$$

Proof From the assumptions that $\mu_{G_\Gamma}(\Omega_{\text{good}}) \geq 1 - \delta$ and $\mu_{G^*_\Gamma}(\Omega_{\text{good}}) \geq 1 - \delta$ we get

$$\begin{aligned}\|\mu_{G_\Gamma} - \mu_{G_\Gamma}(\cdot \,|\, \Omega_{\text{good}})\|_{\text{TV}} &= 1 - \mu_{G_\Gamma}(\Omega_{\text{good}}) \leq \delta, \text{ and}\\ \left\|\mu_{G^*_\Gamma} - \mu_{G^*_\Gamma}(\cdot \,|\, \Omega_{\text{good}})\right\|_{\text{TV}} &= 1 - \mu_{G^*_\Gamma}(\Omega_{\text{good}}) \leq \delta.\end{aligned}$$

Also, from Lemma 34 we have $\left\|\mu_{G_\Gamma}(\cdot \,|\, \Omega_{\text{good}}) - \mu_{G^*_\Gamma}(\cdot \,|\, \Omega_{\text{good}})\right\|_{\text{TV}} = \|\mu_G - \mu_{G^*}\|_{\text{TV}}$. Therefore, it follows from the triangle inequality that

$$\begin{aligned}\left\|\mu_{G_\Gamma} - \mu_{G^*_\Gamma}\right\|_{\text{TV}} &\leq \|\mu_{G_\Gamma} - \mu_{G_\Gamma}(\cdot \,|\, \Omega_{\text{good}})\|_{\text{TV}} + \|\mu_G - \mu_{G^*}\|_{\text{TV}} + \left\|\mu_{G^*_\Gamma} - \mu_{G^*_\Gamma}(\cdot \,|\, \Omega_{\text{good}})\right\|_{\text{TV}}\\ &\leq \|\mu_G - \mu_{G^*}\|_{\text{TV}} + 2\delta.\end{aligned}$$

The lower bound is derived in similar fashion:

$$\begin{aligned}\left\|\mu_{G_\Gamma} - \mu_{G^*_\Gamma}\right\|_{\text{TV}} &\geq \|\mu_G - \mu_{G^*}\|_{\text{TV}} - \|\mu_{G_\Gamma} - \mu_{G_\Gamma}(\cdot \,|\, \Omega_{\text{good}})\|_{\text{TV}} - \left\|\mu_{G^*_\Gamma} - \mu_{G^*_\Gamma}(\cdot \,|\, \Omega_{\text{good}})\right\|_{\text{TV}}\\ &\geq \|\mu_G - \mu_{G^*}\|_{\text{TV}} - 2\delta,\end{aligned}$$

as claimed. ■

We show next that if we have a sampling oracle for μ_G, then we can generate approximate samples from μ_{G_Γ} efficiently.

Lemma 36 *Consider the Potts model on an n-vertex graph G and let $\Gamma = (b, p, d_{\text{IN}}, d_{\text{OUT}})$ be such that conditions* (36) *and* (37) *are satisfied. Suppose that $\mu_{G_\Gamma}(\Omega_{\text{good}}) \geq 1 - \delta$ for some $\delta \in (0, 1)$. Then, given a sampling oracle for the distribution μ_G, there exists a sampling algorithm with running time $\text{poly}(n, b)$ such that the distribution $\mu^{\text{ALG}}_{G_\Gamma}$ of its output satisfies:*

$$\left\|\mu_{G_\Gamma} - \mu^{\text{ALG}}_{G_\Gamma}\right\|_{\text{TV}} \leq \delta.$$

Proof The algorithm first draws a sample σ_G from μ_G using the sampling oracle. It then constructs $\sigma \in \Omega_{G_\Gamma}$ by assigning the spin $\sigma_G(v)$ to every vertex in the gadget corresponding to v for each vertex v of G. This can be done in $O(bn)$ time. From Lemma 34 we see that the sampling algorithm in fact generates a sample from the distribution $\mu_{G_\Gamma}(\cdot \,|\, \Omega_{\text{good}})$, and so

$$\left\|\mu_{G_\Gamma} - \mu^{\text{ALG}}_{G_\Gamma}\right\|_{\text{TV}} = \|\mu_{G_\Gamma} - \mu_{G_\Gamma}(\cdot \,|\, \Omega_{\text{good}})\|_{\text{TV}} = 1 - \mu_{G_\Gamma}(\Omega_{\text{good}}) \leq \delta.$$

■

5.3 Proof of Theorem 28

We are now ready to prove Theorem 28.

Proof of Theorem 28 We show that if there is an identity testing algorithm for the family $\mathcal{M}_{\text{POTTS-BIP-MONO}}(\hat{n}, d, \hat{\beta}, \hat{h})$ with running time $T(\hat{n}) = \text{poly}(\hat{n})$ and sample complexity $L(\hat{n}) = \text{poly}(\hat{n})$, henceforth called the TESTER, then it can be used to solve the the identity testing problem for $\mathcal{M}_{\text{POTTS-MONO}}(n, n, \beta, h)$ in $\text{poly}(n)$ time; the parameters n, β and h depend on $\hat{n}$, d, $\hat{\beta}$ and $\hat{h}$ and will be specified next.

Let us consider first the case when $3 \leq d = O(1)$. In this case, we choose n such that $\hat{n} = 2n^6$, $\beta = \hat{\beta}$ and $h = 2b\hat{h}$. Our identity testing algorithm for the family $\mathcal{M}_{\text{POTTS-MONO}}(n, n, \beta, h)$ constructs the graph G_Γ and the Potts model on G_Γ from Section 5.2 using $\Gamma = (n^5, \lfloor n^{5/4} \rfloor, d-1, 1)$ as the parameters for the random bipartite graph. This choice of parameters ensures that conditions (36) and (37) are satisfied. Note also G_Γ is bipartite by construction and that $|h_{G_\Gamma}(u, i)| \leq h/2b = O(\log n)$ for all $u \in V_{G_\Gamma}$ and $i \in [q]$.

Let (G, β_G, h_G) be a Potts model from $\mathcal{M}_{\text{POTTS-MONO}}(n, n, \beta, h)$, and suppose that there is a hidden model $(G^*, \beta_{G^*}, h_{G^*})$ from $\mathcal{M}_{\text{POTTS-MONO}}(n, n, \beta, h)$ from which we are given samples. We want to use the TESTER to distinguish with probability at least $3/4$ between the cases $\mu_G = \mu_{G^*}$ and $\|\mu_G - \mu_{G^*}\|_{\text{TV}} > 1 - \varepsilon$.

Suppose that σ is sampled from μ_{G_Γ}. Since the field h_G is h-vertex-monochromatic by assumption, it follows from our construction that for each gadget there exists $\kappa \in [q]$ such that for each vertex v in the gadget $h_G(v, j) = \hat{h} \cdot \mathbb{1}(j = \kappa)$. Hence, Theorem 29 implies that with probability $1 - o(1)$ over the choice of the random gadget B, if the configuration in the gadget B_v for a vertex $v \in V_G$ is re-sampled, conditional on the configuration of σ outside of B_v, then the new configuration in B_v will be in a phase with probability at least

$$\left(1 - \frac{q^2 e^{2\hat{h}}}{e^{\delta' \beta_B d}}\right)^{2b} \geq 1 - \frac{2q^2 b}{e^{\delta \beta_B d}}$$

for suitable constants $\delta, \delta' > 0$, since $\hat{h} = O(\log n)$ and $\beta_B d = \omega(\log n)$. 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 $1 - \frac{2q^2 bn}{e^{\delta \beta_B d}}$. Thus,

$$\mu_{G_\Gamma}(\Omega_{\text{good}}) \geq 1 - \frac{q^2 \hat{n}}{e^{\delta \beta_B d}}. \tag{38}$$

We also consider the Potts model on G^*_Γ, obtained from G^* using the same random bipartite graph B. Note that we can not actually construct G^*_Γ, since we only have sample access to $(G^*, \beta_{G^*}, h_{G^*})$, but we can similarly deduce that

$$\mu_{G^*_\Gamma}(\Omega_{\text{good}}) \geq 1 - \frac{q^2 \hat{n}}{e^{\delta \beta_B d}}. \tag{39}$$

Since we are given samples from μ_{G^*}, (39) and Lemma 36 imply that we can generate L samples $\mathcal{S} = \{\sigma_1, \ldots, \sigma_L\}$ from a distribution $\mu^{\text{ALG}}_{G^*_\Gamma}$ in $\text{poly}(n)$ time such that

$$\left\| \mu_{G^*_\Gamma} - \mu^{\text{ALG}}_{G^*_\Gamma} \right\|_{\text{TV}} \leq \frac{q^2 \hat{n}}{e^{\delta \beta_B d}}. \tag{40}$$

Our testing algorithm inputs the Potts model on G_Γ and the L samples $\mathcal{S}$ to the Tester and outputs the Tester's output. Recall that the Tester returns Yes if it regards the samples in $\mathcal{S}$ as samples from μ_{G_Γ}; it returns No if it regards them to be from some other distribution ν such that $\|\mu_{G_\Gamma} - \nu\|_{\text{TV}} > 1 - \varepsilon$.

If $\mu_G = \mu_{G^*}$, then $\mu_{G_\Gamma} = \mu_{G^*_\Gamma}$. Hence, (40) implies that:

$$\left\|\mu_{G_\Gamma} - \mu^{\text{ALG}}_{G^*_\Gamma}\right\|_{\text{TV}} = \left\|\mu_{G^*_\Gamma} - \mu^{\text{ALG}}_{G^*_\Gamma}\right\|_{\text{TV}} \leq \frac{q^2\hat{n}}{e^{\delta\beta_B d}}.$$

Let $(\mu_{G_\Gamma})^{\otimes L}$, $(\mu_{G^*_\Gamma})^{\otimes L}$ and $(\mu^{\text{ALG}}_{G^*_\Gamma})^{\otimes L}$ be the product distributions corresponding to L independent samples from μ_{G_Γ}, $\mu_{G^*_\Gamma}$ and $\mu^{\text{ALG}}_{G^*_\Gamma}$ respectively. We have

$$\left\|(\mu_{G_\Gamma})^{\otimes L} - (\mu^{\text{ALG}}_{G^*_\Gamma})^{\otimes L}\right\|_{\text{TV}} \leq L\left\|\mu_{G_\Gamma} - \mu^{\text{ALG}}_{G^*_\Gamma}\right\|_{\text{TV}} \leq \frac{q^2\hat{n}L}{e^{\delta\beta_B d}} = o_{\hat{n}}(1),$$

since $L = \text{poly}(\hat{n})$ and $\beta_B d = \hat{\beta} d = \omega(\log \hat{n})$. Hence, using the optimal coupling of the distributions $(\mu^{\text{ALG}}_{G^*_\Gamma})^{\otimes L}$ and $(\mu_{G_\Gamma})^{\otimes L}$ as in (19), we obtain

$$\Pr[\text{Tester outputs No given samples } \mathcal{S} \text{ where } \mathcal{S} \sim (\mu^{\text{ALG}}_{G^*_\Gamma})^{\otimes L}] \leq \frac{1}{4} + o_{\hat{n}}(1) < \frac{1}{3}.$$

Hence, the Tester returns Yes with probability at least $2/3$ in this case.

If $\|\mu_G - \mu_{G^*}\|_{\text{TV}} \geq 1 - \varepsilon$, (38), (39) and Lemma 35 imply

$$\left\|\mu_{G_\Gamma} - \mu_{G^*_\Gamma}\right\|_{\text{TV}} \geq 1 - \varepsilon - \frac{2q^2\hat{n}}{e^{\delta\beta_B d}} = 1 - \varepsilon - o_{\hat{n}}(1), \tag{41}$$

because $\beta_B d = \hat{\beta} d = \omega(\log \hat{n})$. Moreover, from (40) we get

$$\left\|(\mu_{G^*_\Gamma})^{\otimes L} - (\mu^{\text{ALG}}_{G^*_\Gamma})^{\otimes L}\right\|_{\text{TV}} \leq L\left\|\mu_{G^*_\Gamma} - \mu^{\text{ALG}}_{G^*_\Gamma}\right\|_{\text{TV}} \leq \frac{q^2\hat{n}L}{e^{\delta\beta_B d}} = o_{\hat{n}}(1).$$

Thus, analogously to (19) (i.e., using the optimal coupling between $(\mu^{\text{ALG}}_{G^*_\Gamma})^{\otimes L}$ and $(\mu_{G^*_\Gamma})^{\otimes L}$), we get

$$\Pr\left[\text{Tester outputs Yes given samples } \mathcal{S} \text{ where } \mathcal{S} \sim (\mu^{\text{ALG}}_{G^*_\Gamma})^{\otimes L}\right] \leq \frac{1}{3}.$$

Hence, the Tester returns No with probability at least $2/3$.

The case when d is such that $d \leq \hat{n}^{1-\rho}$ but $d = d(\hat{n}) \to \infty$ follows in similar fashion. In particular, we can take $b = \left\lfloor n^{4/\rho - 1} \right\rfloor$ and $\Gamma = \{b, b, \lfloor \theta d \rfloor, d - \lfloor \theta d \rfloor\}$, where $\theta = \theta(\rho)$ is a suitable constant. That is, $p = b$, $d_{\text{IN}} = \lfloor \theta d \rfloor$, $d_{\text{OUT}} = d - \lfloor \theta d \rfloor$ and $\hat{n} = \Theta(n^{4/\rho})$. This choice parameters also satisfies conditions (36) and (37). Hence, (38) and (39) can be deduced similarly using Theorem 30 instead. The rest of the proof remains unchanged for this case. ■

6. Hardness of the decision version of approximate counting

In this section we give a general reduction from the approximate counting problem to the decision version of the problem. In particular, we prove Theorem 7 from Section 2.1. We state our results for the models of interest in this paper, but they extend straightforwardly to other spin systems.

We restate first the definition of the decision version of r-approximate counting.

Definition 37 (Decision r-approximate counting) *Given a Potts model (G,β_G,h_G), an approximation ratio $r > 1$ and an input $\hat{Z} \in \mathbb{R}$, distinguish with probability at least $5/8$ between the following two cases:*

$$\text{(i) } Z_{G,\beta_G,h_G} \leq \frac{1}{r}\hat{Z} \qquad \text{(ii) } Z_{G,\beta_G,h_G} \geq r\hat{Z}$$

Recall also that a *fully polynomial-time randomized approximation scheme (*FPRAS*)* for an optimization problem with solutions OPT is a randomized algorithm that for any $\rho > 0$ outputs a solution $\hat{Z}$ satisfying $e^{-\rho}\,\mathrm{OPT} \leq \hat{Z} \leq e^{\rho}\,\mathrm{OPT}$ with probability at least $3/4$ and has running time $\mathrm{poly}(n, 1/\rho)$ where n is the size of the input. To prove Theorem 7, we introduce an intermediate problem referred as r-approximate counting.

Definition 38 (r-approximate counting) *Given a Potts model (G,β_G,h) and an approximation ratio $r > 1$, output a real number $\hat{Z}$ satisfying the following with probability at least $3/4$:*

$$\frac{1}{r}\, Z_{G,\beta_G,h} < \hat{Z} < r\, Z_{G,\beta_G,h}.$$

Notice that an FPRAS for the counting problem is equivalent to an algorithm for the e^{ρ}-approximate counting problem with running time $\mathrm{poly}(n, 1/\rho)$ for all $\rho > 0$. We first show the equivalence of r-approximate counting and its decision version.

Lemma 39 *Let $n, d \geq 1$ be integers and let $\beta, h \geq 0$ be real numbers. Assume that $r = r(n) > 1$ is the approximation ratio. Then, given a polynomial-time algorithm for the decision version of r-approximate counting for a family of Potts models $\mathcal{M}$, where*

$$\mathcal{M} \in \{\hat{\mathcal{M}}^{+}_{\textsc{Potts}}(n, d, \beta, h), \hat{\mathcal{M}}^{-}_{\textsc{Ising}}(n, d, \beta, h), \hat{\mathcal{M}}^{+}_{\textsc{Ising}}(n, d, \beta, h)\},$$

there is also a polynomial-time algorithm for $2r$-approximate counting for $\mathcal{M}$.

Proof Consider a Potts model from $\mathcal{M}$ with the underlying graph G. We note first that using a standard argument we can boost the success probability of the algorithm for the decision version of r-approximate counting in polynomial time. More precisely, for a given $\hat{Z} > 0$ we run the algorithm for

$$k = 80 \left\lceil \log(8 \log(4c_1 n^2 + 4 \log r)) \right\rceil + 1$$

times and output the majority answer. Let X_i be the indicator random variable of the event that the i-th answer is correct and let $X = \sum_{i=1}^{k} X_i$. Then by our assumption we have $\mathbb{E}[X] \geq \frac{5}{8}k$. The Chernoff bound then implies that the majority answer is incorrect with probability at most

$$\Pr\left(X \leq \frac{k}{2}\right) \leq \Pr\left(X \leq \frac{4}{5}\mathbb{E}[X]\right) \leq \exp\left(-\frac{\mathbb{E}[X]}{50}\right) \leq \exp\left(-\frac{k}{80}\right) \leq \frac{1}{8\log(4c_1n^2 + 4\log r)}.$$

Using the boosted version of the decision r-approximate counting algorithm, henceforth call BoostedDecider, we use binary search procedure to give an r-approximate counting algorithm. First note that there exists a constant $c_1 := c_1(q, \beta, h) > 0$ such that

$$\exp\left(-c_1n^2\right) \leq Z_G \leq \exp\left(c_1n^2\right).$$

Then, let $\ell_0 = \frac{1}{r}\exp(-c_1n^2)$ and $u_0 = r\exp(c_1n^2)$. For $i \geq 1$, let $c_i = \sqrt{\ell_{i-1}u_{i-1}}$ and run the testing algorithm with $\hat{Z} = c_i$. If BoostedDecider outputs $Z_G \leq \frac{1}{r}\hat{Z}$ then we let $(\ell_i, u_i) = (\ell_{i-1}, c_i)$, and if BoostedDecider outputs $Z_G \geq r\hat{Z}$ then we let $(\ell_i, u_i) = (c_i, u_{i-1})$. We repeat this process until $u_i/\ell_i \leq 2$, and then output $\hat{Z} = \ell_i$. Observe that $\log u_i - \log \ell_i$ decreases by a factor 2 in each iteration. Thus, the number of times that outputs is called is at most

$$\log_2\left(\frac{\log u_0 - \log \ell_0}{\log 2}\right) = \log_2\left(\frac{2c_1n^2 + 2\log r}{\log 2}\right) \leq 2\log(4c_1n^2 + 4\log r).$$

Assume that BoostedDecider never makes a mistake in all these calls; this happens with probability at least $3/4$ by a union bound. Then, for each $j \geq 0$, the algorithm outputs $Z_G \leq \frac{1}{r}\hat{Z}$ for $\hat{Z} = u_j$ and $Z_G \geq r\hat{Z}$ for $\hat{Z} = \ell_j$. This implies that

$$\frac{1}{r}\ell_j < Z_G < ru_j$$

for all $j \geq 0$. Hence, the final output satisfies

$$\frac{1}{r}\ell_i < Z_G < ru_i \leq 2r\ell_i$$

with probability at least $3/4$. The running time of the algorithm is polynomial in n, assuming that $r \leq \exp(c_1n^2)$. If we have $r > \exp(c_1n^2)$ instead, then the algorithm can just output 1, which is already a r-approximation of Z_G. ■

We show next that a polynomial-time n^c-approximate counting algorithm for a family of Potts models on n-vertex graphs can be turned into an FPRAS.

Lemma 40 *Let $n, d \geq 1$ be integers and let $\beta, h \geq 0$ be real numbers. For any $c > 0$, given a polynomial-time n^c-approximate counting algorithm for a family of Potts models $\mathcal{M}$, where*

$$\mathcal{M} \in \{\hat{\mathcal{M}}^{+}_{\text{Potts}}(n, d, \beta, h), \hat{\mathcal{M}}^{-}_{\text{Ising}}(n, d, \beta, h), \hat{\mathcal{M}}^{+}_{\text{Ising}}(n, d, \beta, h)\},$$

there is an FPRAS *for the counting problem for $\mathcal{M}$.*

Proof Suppose that there is a polynomial-time n^c-approximate counting algorithm for $\mathcal{M}$ where $c > 0$ is a constant. Consider a Potts model from $\mathcal{M}$ defined on a graph G of n vertices. We will give an FPRAS for its partition function. For an arbitrary $\rho > 0$, let k be the smallest integer such that $k \geq (c \log(kn))/\rho$. Notice that $k \leq \mathrm{poly}(\log n, 1/\rho)$. Define a Potts model that is a disjoint union of k copies of the Potts model on G. That is, the underlying graph G' consists of k copies of G, and the weights for each copy are the same as the original model. It follows immediately that $Z_{G'} = (Z_G)^k$. We run the $(kn)^c$-approximate counting algorithm for the Potts model on G' and assume the output is $\hat{Z}$. Then with probability at least $3/4$ we have

$$(kn)^{-c} Z_{G'} < \hat{Z} < (kn)^c Z_{G'}.$$

Assuming this holds, then we get

$$e^{-\rho} Z_G \leq (kn)^{-c/k} Z_G < \hat{Z}^{1/k} < (kn)^{c/k} Z_G \leq e^{\rho} Z_G$$

Thus, $\hat{Z}^{1/k}$ is a e^{ρ}-approximation of Z_G with probability at least $3/4$ and can be computed in $\mathrm{poly}(kn) = \mathrm{poly}(n, 1/\rho)$ time. ■

Proof of Theorem 7 Follows immediately from Lemmas 39 and 40. ■

7. Concluding remarks

We have presented a fairly general method to establish the hardness of identity testing from the hardness of approximate counting. Our technology, however, currently requires insights about each specific model. We conjecture that this is not necessary, and that in fact when approximate counting and structure learning are both hard, the identity testing problem is also hard.

Acknowledgments

The authors thank the anonymous referees for detailed comments on the manuscript. The research of A.B. was supported in part by NSF grant CCF-1850443, Z.C. and E.V.'s by NSF grant CCF-2007022, and D.Š.'s by NSF grants CCF-2007287 and CCF-1934962.

References

I. Bezáková, A. Blanca, Z. Chen, D. Štefankovič, and E. Vigoda. Lower bounds for testing graphical models: Colorings and antiferromagnetic Ising models. *Journal of Machine Learning Research*, 21(25):1–62, 2020.

A. Blanca and A. Sinclair. Dynamics for the mean-field random-cluster model. *Proceedings of the 19th International Workshop on Randomization and Computation*, pages 528–543, 2015.

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.

B. Bollobás, G. Grimmett, and S. Janson. The random-cluster model on the complete graph. *Probability Theory and Related Fields*, 104(3):283–317, 1996. ISSN 0178-8051. doi: 10.1007/BF01213683.

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 and R.D. Buhai. Learning Restricted Boltzmann Machines with few latent variables. In *Advances in Neural Information Processing Systems (NeurIPS)*, volume 33, pages 7020–7030, 2020.

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, 2014.

G. Bresler, F. Koehler, and A. Moitra. Learning restricted Boltzmann machines via influence maximization. In *Proceedings of the 51st Annual ACM Symposium on Theory of Computing (STOC)*, pages 828–839, 2019.

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.

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.

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.

I. Csiszár and P.C. Shields. Information theory and statistics: A tutorial. *Foundations and Trends® in Communications and Information Theory*, 1(4):417–528, 2004.

P. Cuff, J. Ding, O. Louidor, E. Lubetzky, Y. Peres, and A. Sly. Glauber dynamics for the mean-field Potts model. *Journal of Statistical Physics*, 149(3):432–477, 2012. ISSN 0022-4715. doi: 10.1007/s10955-012-0599-2.

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.

H. Duminil-Copin, M. Gagnebin, M. Harel, I. Manolescu, and V. Tassion. Discontinuity of the phase transition for the planar random-cluster and Potts models with $q > 4$. *arXiv preprint arXiv:1611.09877*, 2016.

H. Duminil-Copin, V. Sidoravicius, and V. Tassion. Continuity of the Phase Transition for Planar Random-Cluster and Potts Models with $1 \leq q \leq 4$. *Communications in Mathematical Physics*, 349 (1):47–107, 2017.

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.

A. Galanis, D. Štefankovič, and E. Vigoda. Swendsen-Wang algorithm on the mean-field Potts model. *Proceedings of the 19th International Workshop on Randomization and Computation*, pages 815–828, 2015.

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, E. Vigoda, and L. Yang. Ferromagnetic Potts model: Refined #BIS-hardness and related results. *SIAM Journal on Computing*, 45(6):2004–2065, 2016b.

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, 2016c.

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.

R. Gheissari, E. Lubetzky, and Y. Peres. Exponentially slow mixing in the mean-field Swendsen-Wang dynamics. In *Proceedings of the 29th Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pages 1981–1988. SIAM, 2018.

Surbhi Goel. Learning Ising and Potts models with latent variables. In *International Conference on Artificial Intelligence and Statistics*, pages 3557–3566. PMLR, 2020.

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.

C. Greenhill. The complexity of counting colourings and independent sets in sparse graphs and hypergraphs. *Computational Complexity*, 9(1):52–72, 2000.

G.R. Grimmett. *The Random-Cluster Model*, volume 333 of *Grundlehren der mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*. Springer-Verlag, Berlin, 2006.

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.

G.E. Hinton. Training products of experts by minimizing contrastive divergence. *Neural computation*, 14(8):1771–1800, 2002.

G.E. Hinton and R.R. Salakhutdinov. Replicated softmax: an undirected topic model. In *Advances in neural information processing systems*, pages 1607–1614, 2009.

G.E. Hinton, S. Osindero, and Y-W. Teh. A fast learning algorithm for deep belief nets. *Neural computation*, 18(7):1527–1554, 2006.

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

M.R. Jerrum, L.G. Valiant, and V.V. Vazirani. Random generation of combinatorial structures from a uniform distribution. *Theoretical Computer Science*, 43:169–188, 1986.

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.

V. Kolmogorov. A faster approximation algorithm for the Gibbs partition function. In *Proceedings of the 31st Conference On Learning Theory*, volume 75 of *Proceedings of Machine Learning Research*, pages 228–249. PMLR, 06–09 Jul 2018.

M.J. Luczak and T. Łuczak. The phase transition in the cluster-scaled model of a random graph. *Random Structures & Algorithms*, 28(2):215–246, 2006.

A. Montanari. Computational implications of reducing data to sufficient statistics. *Electronic Journal of Statistics*, 9(2):2370–2390, 2015.

K.P. Murphy. *Machine learning: a probabilistic perspective*. MIT press, 2012.

S. Osindero and G.E. Hinton. Modeling image patches with a directed hierarchy of Markov random fields. In *Advances in neural information processing systems*, pages 1121–1128, 2008.

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)*, volume 17, pages 959–960, 1999.

R. Salakhutdinov and G.E. Hinton. Deep Boltzmann machines. In *Artificial intelligence and statistics*, pages 448–455, 2009.

R. Salakhutdinov, A. Mnih, and G.E. Hinton. Restricted Boltzmann machines for collaborative filtering. In *Proceedings of the 24th international conference on Machine learning*, pages 791–798, 2007.

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.

P. Smolensky. Information processing in dynamical systems: Foundations of harmony theory. Technical report, Colorado Univ at Boulder Dept of Computer Science, 1986.

D. Štefankovič, S. Vempala, and E. Vigoda. Adaptive simulated annealing: A near-optimal connection between sampling and counting. *Journal of the ACM (JACM)*, 56(3):1–36, 2009.

R.H. Swendsen and J.S. Wang. Nonuniversal critical dynamics in Monte Carlo simulations. *Physical Review Letters*, 58:86–88, 1987.

S.P Vadhan. The complexity of counting in sparse, regular, and planar graphs. *SIAM Journal on Computing*, 31(2):398–427, 2001.

L.G. Valiant. The complexity of enumeration and reliability problems. *SIAM Journal on Computing*, 8(3):410–421, 1979.

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*, pages 8069–8079, 2019.

Appendix A. The ferromagnetic mean-field Potts model: proofs

In this appendix we prove our detailed results concerning the phase transitions of the ferromagnetic mean-field Potts models (i.e., Lemmas 9 and 10). As mentioned, several variants of these results have appeared before, e.g., (Bollobás et al., 1996; Luczak and Łuczak, 2006; Goldberg and Jerrum, 2012; Cuff et al., 2012; Gheissari et al., 2018; Galanis et al., 2015; Blanca and Sinclair, 2015), but we need slightly more precise results.

Proof of Lemma 9 Let us introduce some convenient notation first. For an integer $m \geq 1$, let

$$\hat{A} = \left\{ (\alpha_1, \ldots, \alpha_q) \in \mathbb{R}^q : \alpha_i \geq 0,\ \sum_{i=1}^{q} \alpha_i = 1,\ \alpha_i m \in \mathbb{N} \right\},$$

$$\hat{D} = \mathrm{Ball}_\infty(u, m^{-1/4}) = \left\{ \alpha \in \hat{A} : \|\alpha - u\|_\infty \leq m^{-1/4} \right\},$$

$$\hat{M} = \bigcup_{i=1}^{q} \mathrm{Ball}_\infty(\alpha^{*,i}, m^{-1/4}) = \bigcup_{i=1}^{q} \left\{ \alpha \in \hat{A} : \left\|\alpha - \alpha^{*,i}\right\|_\infty \leq m^{-1/4} \right\},$$

and $\hat{S} = \hat{A} \backslash (\hat{D} \cup \hat{M})$. Setting $\hat{\beta}_H = \beta_H \cdot m$, we have

$$Z_H^{\mathrm{D}}(\beta_H) = \sum_{\alpha \in \hat{D}} \binom{m}{\alpha_1 m \ \cdots \ \alpha_q m} \exp\left(\frac{\hat{\beta}_H}{m} \sum_{i=1}^{q} \binom{\alpha_i m}{2} \right), \tag{42}$$

and similarly for $Z_H^{\mathrm{M}}(\beta_H)$ and $Z_H^{\mathrm{S}}(\beta_H)$ with the summation over $\hat{M}$ and $\hat{S}$ respectively.

Using standard bounds for the multinomial coefficient (see, e.g., Lemma 2.2 in Csiszár and Shields, 2004), we have for every $\alpha \in \hat{A}$

$$\frac{1}{|\hat{A}|} e^{H(\alpha)m} \leq \binom{m}{\alpha_1 m \ \cdots \ \alpha_q m} \leq e^{H(\alpha)m}, \tag{43}$$

where $H(\alpha) = \sum_{i=1}^{q} -\alpha_i \ln \alpha_i$. Hence, for $\beta \in \mathbb{R}$ and $\alpha \in \mathbb{R}^q$, we introduce:

$$\Phi_\beta(\alpha) = H(\alpha) + \frac{\beta}{2} \|\alpha\|_2^2 .$$

The function Φ_β have the following properties, which we prove later and will be useful throughout the proof.

Fact 41 *(i) For $\alpha \in \hat{A}$ and $\beta_1, \beta_2 > 0$, we have $|\Phi_{\beta_1}(\alpha) - \Phi_{\beta_2}(\alpha)| \leq \frac{1}{2}|\beta_1 - \beta_2|$.*

(ii) When $\hat{\beta}_H = \mathfrak{B}_o$, the function $\Phi_{\mathfrak{B}_o}$ has exactly $q+1$ global maxima in $\hat{A}$ consisting of one disordered phase $u = (1/q, \ldots, 1/q)$ and q majority phases $\alpha^{*,i}$ with $i \in [q]$, where the i-th coordinate of $\alpha^{*,i}$ is strictly larger than $1/q$.

(iii) There exist constants $\varepsilon, c > 0$ such that $\Phi_{\mathfrak{B}_o}(\alpha)$ is c-strongly concave in the balls $\mathrm{Ball}_\infty(u, \varepsilon)$ and $\mathrm{Ball}_\infty(\alpha^{*,i}, \varepsilon)$ for $i \in [q]$. That is, $\forall \alpha \in \hat{A}$ such that $\|\alpha - u\|_\infty \leq \varepsilon$ or $\left\|\alpha - \alpha^{*,i}\right\|_\infty \leq \varepsilon$ for some $i \in [q]$, we have $\nabla^2 \Phi_{\mathfrak{B}_o}(\alpha) \preceq -c \cdot I$, where I is the $q \times q$ identity matrix.

Hence, (43) and part (ii) of this fact imply

$$\begin{aligned} Z_H^{\mathrm{D}}(\beta_H) &\geq \sum_{\alpha \in \hat{D}} \frac{e^{-\hat{\beta}_H/2}}{|\hat{A}|} \exp\left[\Phi_{\hat{\beta}_H}(\alpha) m \right] \\ &\geq \frac{e^{-\hat{\beta}_H/2}}{|\hat{A}|} \exp\left(-\frac{1}{2} |\hat{\beta}_H - \mathfrak{B}_o| m \right) \sum_{\alpha \in \hat{D}} \exp\left[\Phi_{\mathfrak{B}_o}(\alpha) m \right] \\ &\geq \frac{e^{-\hat{\beta}_H/2}}{|\hat{A}|} \exp\left(-\frac{c'}{2} \sqrt{m} \right) \exp\left[\Phi_{\mathfrak{B}_o}(u) m \right]. \end{aligned} \tag{44}$$

Similarly, we deduce that

$$\begin{aligned} Z_H^{\mathrm{M}}(\beta_H) &\geq \frac{e^{-\hat{\beta}_H/2}}{|\hat{A}|} \exp\left(-\frac{c'}{2} \sqrt{m} \right) \sum_{i=1}^{q} \exp\left[\Phi_{\mathfrak{B}_o}(\alpha^{*,i}) m \right] \\ &= \frac{q e^{-\hat{\beta}_H/2}}{|\hat{A}|} \exp\left(-\frac{c'}{2} \sqrt{m} \right) \exp\left[\Phi_{\mathfrak{B}_o}(u) m \right], \end{aligned} \tag{45}$$

and

$$Z_H^{\mathrm{S}}(\beta_H) \le e^{-\hat{\beta}_H/2} \exp\left(\frac{1}{2}|\hat{\beta}_H - \mathfrak{B}_o|m\right) \sum_{\alpha\in\hat{S}} \exp\left[\Phi_{\mathfrak{B}_o}(\alpha)m\right]$$
$$\le e^{-\hat{\beta}_H/2}|\hat{A}| \exp\left(\frac{c'}{2}\sqrt{m}\right) \exp\left[\left(\max_{\alpha\in\hat{S}} \Phi_{\mathfrak{B}_o}(\alpha)\right) m\right]. \tag{46}$$

Let $\hat{S} = \hat{S}_1 \cup \hat{S}_2$ where

$$\hat{S}_1 = \hat{A} \setminus \left(\mathrm{Ball}_\infty(u,\varepsilon) \cup \bigcup_{i=1}^{q} \mathrm{Ball}_\infty(\alpha^{*,i},\varepsilon)\right)$$

and

$$\hat{S}_2 = \left(\mathrm{Ball}_\infty(u,\varepsilon) \cup \bigcup_{i=1}^{q} \mathrm{Ball}_\infty(\alpha^{*,i},\varepsilon)\right) \setminus (\hat{D} \cup \hat{M}).$$

Since the function $\Phi_{\mathfrak{B}_o}$ is continuous, and $u, \alpha^{*,1}, \dots, \alpha^{*,q}$ are its only global maxima, for constant $\varepsilon > 0$ there exists constant $\delta = \delta(\varepsilon) > 0$ such that for all $\alpha \in \hat{S}_1$ we have

$$\Phi_{\mathfrak{B}_o}(\alpha) \le \Phi_{\mathfrak{B}_o}(u) - \delta.$$

By part (iii) of Fact 41, $\Phi_{\mathfrak{B}_o}(\alpha)$ is c-strongly concave in $\hat{S}_2$; thus, for all $\alpha \in \mathrm{Ball}_\infty(u,\varepsilon)\setminus\hat{D}$ we have

$$\Phi_{\mathfrak{B}_o}(\alpha) \le \Phi_{\mathfrak{B}_o}(u) + \nabla\Phi_{\mathfrak{B}_o}(u)(\alpha - u) - c\|\alpha - u\|^2$$
$$= \Phi_{\mathfrak{B}_o}(u) - c\|\alpha - u\|^2$$
$$\le \Phi_{\mathfrak{B}_o}(u) - cm^{-1/2},$$

and similarly for all $\alpha \in \mathrm{Ball}_\infty(\alpha^{*,i},\varepsilon)\setminus\hat{M}$ we have

$$\Phi_{\mathfrak{B}_o}(\alpha) \le \Phi_{\mathfrak{B}_o}(\alpha^{*,i}) + \nabla\Phi_{\mathfrak{B}_o}(\alpha^{*,i})(\alpha - u) - c\left\|\alpha - \alpha^{*,i}\right\|^2$$
$$= \Phi_{\mathfrak{B}_o}(u) - c\left\|\alpha - \alpha^{*,i}\right\|^2$$
$$\le \Phi_{\mathfrak{B}_o}(u) - cm^{-1/2}.$$

Therefore,

$$\max_{\alpha\in\hat{S}} \Phi_{\mathfrak{B}_o}(\alpha) \le \Phi_{\mathfrak{B}_o}(u) - cm^{-1/2}.$$

Plugging this bound into (46) and combining it with (44), we get

$$Z_H^{\mathrm{S}}(\beta_H) \le e^{-\hat{\beta}_H/2}|\hat{A}| \exp\left(\frac{c'}{2}\sqrt{m}\right) \exp\left(-c\sqrt{m}\right) \exp\left[\Phi_{\mathfrak{B}_o}(u)m\right]$$
$$\le |\hat{A}|^2 \exp\left(-(c-c')\sqrt{m}\right) Z_H^{\mathrm{D}}(\beta_H).$$

Combining with (45) instead we obtain

$$Z_H^{\mathrm{S}}(\beta_H) \le \frac{|\hat{A}|^2}{q} \exp\left(-(c-c')\sqrt{m}\right) Z_H^{\mathrm{M}}(\beta_H).$$

The results then follows by picking $c' = c/2$. ■

We wrap up the proof of Lemma 9 by establishing the facts used of the function in $\Phi_{\mathfrak{B}_o}$.

Proof of Fact 41 Part (i) follows from the definition of the function Φ_β, since when $\alpha \in \hat{A}$, $\|\alpha\|_1 = 1$, and so $\|\alpha\|_2 \leq 1$.

For part (ii), suppose $\alpha = (\alpha_1, \ldots, \alpha_q)$ is a local maxima for $\Phi_{\mathfrak{B}_o}(\alpha)$. Using the method of Lagrange multipliers, we obtain that α must satisfy:

$$\mathfrak{B}_o \alpha_i - \log(\alpha_i) = 1 - \lambda,\ i \in [q].$$

The function $\mathfrak{B}_o x - \log x$ is decreasing for $x < 1/\mathfrak{B}_o$ and increasing for $x > 1/\mathfrak{B}_o$. This implies that for any λ there are at most 2 solutions to $\mathfrak{B}_o x - \log x = 1 - \lambda$ and hence there are at most two different values of α_i. If there is only one value of α_i then $\alpha_i = 1/q$ for $i \in [q]$. If there are two values of α_i then one of them is in $(0, 1/\mathfrak{B}_o)$ and one of them is in $(1/\mathfrak{B}_o, 1)$.

Now the Hessian of $\Phi_{\mathfrak{B}_o}$ is

$$\nabla^2 \Phi_{\mathfrak{B}_o}(\alpha) = -\mathrm{diag}(\alpha_1^{-1}, \ldots, \alpha_q^{-1}) + \mathfrak{B}_o I, \tag{47}$$

and since α is a maxima for $\Phi_{\mathfrak{B}_o}$, then $\nabla^2 \Phi_{\mathfrak{B}_o}(\alpha)$ must be negative definite in the subspace of vectors perpendicular to 1 (since the sum of α_i is constrained to be 1 the perturbations must maintain this constraint). If there were at least two indexes (w.l.o.g., make the indexes 1 and 2) such that $\alpha_1 = \alpha_2 > 1/\mathfrak{B}_o$ then the Hessian is not negative definite in the subspace of vectors perpendicular to 1 (e.g., take the vector $x = (1, -1, 0, \ldots, 0)$; then $x^T \nabla^2 \Phi_{\mathfrak{B}_o}(\alpha) x = 2(\mathfrak{B}_o - 1/\alpha_1) > 0$). Thus a (constrained) maxima α of $\Phi_{\mathfrak{B}_o}$ will either have all α_i equal to $1/q$, or exactly $q - 1$ of the α_i's will be the same.

Hence, the maxima of $\Phi_{\mathfrak{B}_o}$ will coincide with those of a one-dimensional version of it , denoted by Ψ_1, previously studied in (Galanis et al., 2015). The function $\Psi_1 : [0, 1] \to \mathbb{R}$ is define as $\Psi_1(x) = \Phi_{\mathfrak{B}_o}(x, y, \ldots, y)$, where $y = \frac{1-x}{q-1}$. The function Ψ_1 has 2 global maxima (see Lemma 2 in (Galanis et al., 2015)) and hence $\Phi_{\mathfrak{B}_o}$ has exactly $q + 1$ global maxima (one of the maxima of Ψ_1 corresponds to q maxima of $\Phi_{\mathfrak{B}_o}$). Finally, observe that $\mathfrak{B}_o < q$, and so the coordinate of the maxima of $\Phi_{\mathfrak{B}_o}$ in $(1/\mathfrak{B}_o, 1)$ is greater than $1/q$.

For part (iii), note that the Hessian in equation (47) is continuous around $(\alpha_1, \ldots, \alpha_q)$ and hence it is negative definite in a sufficiently small ball around u and $\alpha^{*,i}$. ■

We will provide next the proof of Lemma 10, in which we will use the following bound on the ratio $\frac{Z_H^{\mathrm{M}}(\mathfrak{B}_o)}{Z_H^{\mathrm{D}}(\mathfrak{B}_o)}$, which is derived similarly to Lemma 9.

Fact 42 $\frac{1}{q|\hat{A}|} \leq \frac{Z_H^{\mathrm{M}}(\mathfrak{B}_o/m)}{Z_H^{\mathrm{D}}(\mathfrak{B}_o/m)} \leq q|\hat{A}|^2$.

Proof From (42) and (43), we obtain

$$\begin{aligned} Z_H^{\mathrm{D}}(\mathfrak{B}_o/m) &= \sum_{\alpha \in \hat{\mathrm{D}}} \binom{m}{\alpha_1 m\ \cdots\ \alpha_q m} \exp\left(\frac{\mathfrak{B}_o}{m} \sum_{i=1}^{q} \binom{\alpha_i m}{2}\right) \\ &\leq \sum_{\alpha \in \hat{\mathrm{D}}} e^{-\mathfrak{B}_o/2} \exp\left[\Phi_{\mathfrak{B}_o}(\alpha) m\right] \\ &\leq |\hat{A}| e^{-\mathfrak{B}_o/2} \exp\left[\Phi_{\mathfrak{B}_o}(u) m\right]. \end{aligned}$$

Similarly we have

$$Z_H^{\mathrm{M}}(\mathfrak{B}_o/m) \le |\hat{A}|e^{-\mathfrak{B}_o/2}\sum_{i=1}^{q}\exp\left[\Phi_{\mathfrak{B}_o}(\alpha^{*,i})m\right] \le q|\hat{A}|e^{-\mathfrak{B}_o/2}\exp\left[\Phi_{\mathfrak{B}_o}(u)m\right].$$

Combining our upper and lower bounds on $Z_H^{\mathrm{D}}(\mathfrak{B}_o/m)$ and $Z_H^{\mathrm{D}}(\mathfrak{B}_o/m)$ we obtain the result. ■

We are now ready to proof Lemma 10.

Proof of Lemma 10 For ease of notation let $f(\beta) = \frac{Z_H^{\mathrm{M}}(\beta)}{Z_H^{\mathrm{D}}(\beta)}$. We show that for suitable constants $c, c' > 0$, for $\beta_L = \mathfrak{B}_o/m - c'm^{-3/2}$ we have

$$f(\beta_L) \le \exp(-c\sqrt{m}), \tag{48}$$

and for $\beta_U = \mathfrak{B}_o/m + c'm^{-3/2}$ we have

$$f(\beta_U) \ge \exp(c\sqrt{m}). \tag{49}$$

Since $|\hat{M}| = O(m^q)$ and $|\hat{D}| = O(m^q)$, we can compute $Z_H^{\mathrm{M}}(\beta)$ and $Z_H^{\mathrm{D}}(\beta)$ for any $\beta \in [\beta_L, \beta_U]$ in $\mathrm{poly}(m)$ time by enumerating over elements of $\hat{M}$ and $\hat{D}$, respectively. (Note that this involves computing multinomial coefficients, which can be done for example by expressing them as product of q binomial coefficients; see (42).) Then, given (48) and (49), for any $R \in [\exp(-c\sqrt{m}), \exp(c\sqrt{m})]$ and small enough $\xi > 0$, we can use the bisection method with $[\beta_L, \beta_U]$ as the starting interval to find a $\beta \in [\beta_L, \beta_U]$ such that

$$f(\beta) \le R \le f(\beta + \xi) \le f(x) + \xi \cdot \max_{\beta_0 \in [\beta_L, \beta_U]} f'(\beta_0)$$

in time polynomial in m and $\log \xi^{-1}$. Since $f'(\beta_0) = \exp(O(m))$ for $\beta_0 \in [\beta_L, \beta_U]$, we can choose $\xi = \exp(-\Theta(m))$ so that $f(\beta) \le R \le f(\beta) + \delta R$ as desired.

To establish (48) and (49) we consider the function

$$g(\beta) = \log Z_H^{\mathrm{M}}(\beta) - \log Z_H^{\mathrm{D}}(\beta).$$

Note that

$$\frac{\partial}{\partial\beta}g(\beta) = \frac{\frac{\partial}{\partial\beta}Z_H^{\mathrm{M}}(\beta)}{Z_H^{\mathrm{M}}(\beta)} - \frac{\frac{\partial}{\partial\beta}Z_H^{\mathrm{D}}(\beta)}{Z_H^{\mathrm{D}}(\beta)}. \tag{50}$$

By a direct (and standard) calculation, we can check that the first term in the right-hand-side expression in (50) corresponds to the expected number of monochromatic edges in a random configuration σ of the model conditioned on σ being in the set M. Therefore,

$$\frac{\frac{\partial}{\partial\beta}Z_H^{\mathrm{M}}(\beta)}{Z_H^{\mathrm{M}}(\beta)} \ge \binom{\hat{\alpha}m - m^{3/4}}{2} + (q-1)\binom{\frac{(1-\hat{\alpha})m}{q-1} - m^{3/4}}{2}, \tag{51}$$

where $\hat{\alpha}$ is the constant in the definition of the set M. Similarly, the second term in the right-hand-side of (50) is the expected number of monochromatic edges in a random configuration σ of the model conditioned on σ being in the set D and so

$$\frac{\frac{\partial}{\partial\beta} Z_H^{\mathrm{D}}(\beta)}{Z_H^{\mathrm{D}}(\beta)} \leq q\binom{m/q + m^{3/4}}{2}. \tag{52}$$

Combining (51) and (52) and using the fact that $\hat{\alpha} > 1/q$, we obtain for a suitable constant $\rho > 0$ and sufficiently large m that for any $\beta \in [\beta_L, \beta_U]$

$$\frac{\partial}{\partial\beta} g(\beta) \geq \rho m^2. \tag{53}$$

Since $|\hat{A}| = \Theta(m^q)$, Fact 42 implies that $|g(\mathfrak{B}_o/m)| = \Theta(\log m)$. Hence, by the mean value theorem

$$g(\beta_L) \leq g(\mathfrak{B}_o/m) - \rho m^2 |\mathfrak{B}_o/m - \beta_L| \leq -c\sqrt{m}$$

and similarly $g(\beta_U) \geq c\sqrt{m}$ for a suitable constant $c > 0$. Since $g = \log f$, (48) and (49) follow and the proof is complete. ■