Spectral Bounds for the Ising Ferromagnet on an Arbitrary Given Graph

Alaa Saade, Florent Krzakala, Lenka Zdeborová

Spectral Bounds for the Ising Ferromagnet
on an Arbitrary Given Graph

Alaa Saade[1]   Florent Krzakala[1,2,3]   Lenka Zdeborová[3,4]

Abstract

We revisit classical bounds of M. E. Fisher on the ferromagnetic Ising model [1], and show how to efficiently use them on an arbitrary given graph to rigorously upper-bound the partition function, magnetizations, and correlations. The results are valid on any finite graph, with arbitrary topology and arbitrary positive couplings and fields. Our results are based on high temperature expansions of the aforementioned quantities, and are expressed in terms of two related linear operators: the non-backtracking operator and the Bethe Hessian. As a by-product, we show that in a well-defined high-temperature region, the susceptibility propagation algorithm [2] converges and provides an upper bound on the true spin-spin correlations.

1 Introduction

Undirected graphical models, or Markov random fields, are a powerful paradigm for multivariate statistical modeling. They allow to encode information about the conditional dependencies of a large number of interacting variables in a compact way, and provide a unified view of inference and learning problems in areas as diverse as statistical physics, computer vision, coding theory or machine learning (see e.g. [3, 4] for examples of applications). Arguably one of the oldest and most prototypical example of a (pairwise) Markov random field is the Ising model, defined by the joint probability distribution

$$\mathbb{P}(s) = \frac{1}{\mathcal{Z}} \exp\left[\sum_{(ij)\in E} J_{ij} s_i s_j + \sum_{i\in V} h_i s_i \right], \tag{1}$$

where $G = (V = [n], E)$ is an arbitrary undirected graph, with n vertices and m edges, $s \in \{\pm 1\}^n$ is a collection of binary spins, and the *partition function* $\mathcal{Z}$ is defined by

$$\mathcal{Z} = \sum_{s\in\{\pm 1\}^n} \prod_{(ij)\in E} \exp\left(J_{ij} s_i s_j \right) \prod_{i\in V} \exp\left(h_i s_i \right). \tag{2}$$

We will call this model *ferromagnetic* if the couplings $(J_{ij})_{(ij)\in E}$ and the fields $(h_i)_{i\in[n]}$ are all positive. It can be shown ([3]) that the couplings and the fields of the model (1) are in one-to-one

---

[1]Laboratoire de Physique Statistique (CNRS UMR-8550), PSL Universités & École Normale Supérieure, 75005 Paris

[2]Sorbonne Universités, UPMC Univ. Paris 06

[3]Simons Institute for the Theory of Computing, University of California, Berkeley, Berkeley, CA, 94720

[4]Institut de Physique Théorique, CNRS, CEA, Université Paris-Saclay, 91191, Gif-sur-Yvette, France.

correspondence with its *magnetizations* $(m_a)_{a \in [n]}$ and *susceptibility matrix* $\chi \in \mathbb{R}^{n \times n}$, defined by

$$m_a = \mathbb{E}(s_a) = \sum_{s \in \{\pm 1\}^n} s_a \, \mathbb{P}(s) \qquad \text{for } a \in [n] \,, \tag{3}$$

$$\chi_{ab} = \mathbb{E}(s_a s_b) = \sum_{s \in \{\pm 1\}^n} s_a s_b \, \mathbb{P}(s) \qquad \text{for } a, b \in [n] \,. \tag{4}$$

The task of computing the quantities (2,3,4) knowing the joint probability distribution (1) is sometimes referred to as *inference*, or direct problem. Its applications include e.g. reconstructing partially observed binary images in computer vision [5], or similarity-based clustering into 2 groups [6]. Conversely, the task of computing the joint probability distribution (1) knowing the magnetizations (3) and correlations (4) is usually called *learning*, or inverse problem. The practical importance of this inverse task stems from the fact that the model (1) is the maximum entropy model with constrained means and correlations. In particular, it has found numerous applications in biology [7, 8], from predicting the three-dimensional folding of proteins, to identifying neural activity patterns. In machine learning, a variant of this inverse problem is usually referred to as Boltzmann Machine learning [9], and is a prototypical unsupervised learning problem. In practice, learning is done through a local optimization of the likelihood function, which involves the partition function (2), and which gradients involve the magnetizations (3) and correlations (4).

Despite their considerable practical importance, the estimation of these three quantities in general is a notoriously intractable problem, except in particular cases, notably when the graph G is planar and the external fields $(h_i)_{i \in [n]}$ vanish. In the latter case, explicit expressions exist, originating in the work of Kac and Ward [10, 11, 12], and allowing to devise polynomial-time inference algorithms [13]. Recently, [14] proposed a greedy algorithm to approximate an arbitrary graph by a planar one, thus making inference and learning tractable in a more general setting. Their approach, however, does not provide bounds on their estimates.

For the case of the ferromagnetic Ising model (1) with positive couplings $(J_{ij})_{(ij) \in E}$ and fields $(h_i)_{i \in [n]}$, it is well known [15] that the ground-state, i.e. the configuration of the spins with minimum energy, can be found in polynomial time using graph cuts. This method has been used successfully in a number of computer vision problems, (see e.g. [16, 17, 5]), where a ferromagnetic Ising model is used to denoise a partially observed image. Under the same ferromagnetic assumption, sometimes called attractive or log-supermodular in the machine learning community, [18, 19] showed that the stationary points of the Bethe free energy allow to lower bound the partition function (2). The proof of [19] relies on a loop series expansion of the partition function first derived by [20]. In this paper, we show how to *upper* bound the marginals, correlations and partition function of the ferromagnetic Ising model. Interestingly, our results also have a strong connection with the Bethe approximation (see section 2.3).

The difficulty of inference and learning in the model (1) has prompted the development of a wealth of numerical methods to approximate the quantities (2,3,4). In the ferromagnetic case, a particularly important breakthrough originated in the so-called cluster Monte Carlo methods of [21, 22]. By updating a whole cluster of spins instead of just one, these algorithms can make non-local moves in the configuration space while still verifying the detailed balance condition. Therefore, they provide a substantial speed-up over conventional Markov Chain Monte Carlo methods, especially when the model (1) is near criticality. Other numerical methods are typically based on a low or high temperature expansion of the quantities of interest, and an exhaustive enumeration of an increasing number of the terms contributing to this expansion [23]. On the learning side of the problem, [24] introduced a principled and accurate approximation scheme based on the identification of the clusters of spins that contribute the most to the entropy of the model. While potentially allowing to reach an

arbitrary accuracy in the determination of the quantities (2,3,4), these numerical approaches do not, by nature, admit a closed-form solution, and do not in general provide bounds on their estimates.

In this work, we prove simple and explicit upper bounds on the three quantities of interest (2,3,4) for *arbitrary* graphs. These bounds are valid under two assumptions. First, we consider the particular case where the couplings $(J_{ij})_{(ij)\in E}$ and the fields $(h_i)_{i\in[n]}$ are positive. Second, we require the model (1) to be in a well-defined high temperature region, specified by a condition on the spectral radius of a linear operator associated with the model (1). Our results use the same starting point as some the intensive numerical methods described previously, but are much simpler in nature, and provide an efficient closed-form bound, valid on any finite graph, as long as the couplings and the fields are positive. We therefore expect our results to find natural applications, e.g. in the inference and learning examples listed above, where the computation of quantities such as the partition function (2) and the marginals (3,4) play a central role.

Our approach is based on the high temperature expansion of the Ising model, reviewed in section 3.1. For ferromagnetic models, this expansion is composed of positive contributions corresponding to certain paths on the graph G. Our upper bounds are obtained by noting that the set of such paths is included in a tractable set of more general walks, for which an analytical expression can be derived. In this sense, our methodology is reminiscent of the one used in [1] who derived similar upper bounds. The crucial difference between our work and [1] is that whereas [1] derives formulas for specific regular lattices in the thermodynamic limit, our results are applicable to arbitrary finite graphs and arbitrary positive couplings and are therefore not only of theoretical interest, but can be readily incorporated into data processing algorithms.

1.1 Notations and definitions

We will denote the set of edges of a graph G as $E(G)$. We will call n the number of vertices in G and m the number of edges. For any integer $k \geq 1$, we denote by $[k]$ the set of integers greater or equal to 1 and lower or equal to k. For any vertex $i \in [n]$, ∂i will denote the set of neighbors of i in the graph G. Recall that the spectral radius $\rho(A)$ of a linear operator A is defined as

$$\rho(A) = \max_{\lambda \in \text{Sp}(A)} |\lambda|, \tag{5}$$

where $\text{Sp}(A)$ denotes the eigenvalue spectrum of A. For the Ising model in vanishing fields (8) on a finite graph G, we define the non-backtracking operator $\text{B} \in \mathbb{R}^{2m \times 2m}$, acting on the directed edges of the graph, by its elements

$$\text{B}_{(i \to j),(k \to l)} = \tanh(J_{kl})\mathbf{1}(i = l)\mathbf{1}(j \neq k). \tag{6}$$

Finally, we define the Bethe Hessian operator $\text{H} \in \mathbb{R}^{n \times n}$ with elements

$$\text{H}_{ij} = \mathbf{1}(i = j)\left(1 + \sum_{k \in \partial i} \frac{\tanh(J_{ik})^2}{1 - \tanh(J_{ik})^2}\right) - \mathbf{1}(j \in \partial i)\frac{\tanh(J_{ij})}{1 - \tanh(J_{ij})^2}. \tag{7}$$

The non-backtracking operator, known as the Hashimoto matrix in graph theory [25], was first introduced in the context of inference by [26], who used it as the basis of a spectral community detection method. Other applications of the non-backtracking operator to retrieval in the Hopfield model and to similarity-based clustering are studied in [27, 28]. The use of the Bethe Hessian in inference was pioneered in [29], where its connection with the non-backtracking operator is highlighted in the context of community detection. Other applications to matrix completion and similarity-based clustering are considered in [30, 28]. Note that the Bethe Hessian corresponds to the Hessian of the Bethe free energy of the Ising model (8), i.e. to the inverse susceptibility matrix in the Bethe approximation [30].

3

1.2 Assumptions

Our results hold for arbitrary finite graphs G, provided the two following conditions hold.

- First, we assume that all the couplings $(J_{ij})_{(ij)\in E}$ as well as the external fields $(h_i)_{i\in[n]}$ are non-negative.

- Second, we assume that the model (1) is in a high temperature region, specified by the condition on the spectral radius of the non-backtracking matrix $\rho(\mathrm{B}) < 1$.

It will prove convenient to restrict our analysis to the Ising model in vanishing external fields, defined by

$$\mathrm{P}(s) = \frac{1}{\mathcal{Z}} \exp \sum_{(ij)\in E} (J_{ij} s_i s_j). \tag{8}$$

This is in fact not a restriction, as any n-spin Ising model with fields can be expressed as a $n+1$-spin model without fields, so that model (8) is as expressive as model (1) [1]. To make this claim precise, we will use proposition 1 in [14], which we recall here for completeness.

Proposition 1.1. *(proposition 1 in [14]) Consider the Ising model (1) on the graph $G = ([n], E(G))$, with couplings $(J_{ij})_{(ij)\in E(G)}$ and fields $(h_i)_{i\in[n]}$, and corresponding partition function $\mathcal{Z}$, magnetizations $(m_a)_{a\in[n]}$, and correlations $(\chi_{ab})_{(a,b)\in[n]^2}$. Define another Ising model on the graph $\hat{G} = ([n+1], E(G) \cup \{(i, n+1), \forall i \in [n]\})$ with vanishing fields, and couplings*

$$\hat{J}_{ij} = \begin{cases} J_{ij} & \text{if } j < n+1 \\ h_i & \text{if } j = n+1 \end{cases}. \tag{9}$$

Call $\hat{\mathcal{Z}}$ its partition function, and $(\hat{\chi}_{ab})_{(a,b)\in[n+1]^2}$ its correlations. Then $\hat{\mathcal{Z}} = 2\mathcal{Z}$ and

$$\hat{\chi}_{ab} = \begin{cases} \chi_{ab} & \text{if } a < n+1 \text{ and } b < n+1 \\ m_a & \text{if } a < n+1 \text{ and } b = n+1 \end{cases}. \tag{10}$$

We will therefore consider model (8) in the following, and bound its partition function and correlations, which will yield a bound on the partition function, magnetizations and correlations of model (1). Note also that from proposition 1.1, the new couplings $(\hat{J}_{ij})_{(ij)\in E(\hat{G})}$ are positive if and only if both the original couplings $(J_{ij})_{(ij)\in E(G)}$ and fields $(h_i)_{i\in[n]}$ are positive.

1.3 Algorithms

Before stating our rigorous results in the next section, we describe here the corresponding procedures to upper bound the partition function, magnetizations and correlations of the Ising model on a given graph. As will be apparent from the upcoming results, our bounds can be expressed either using the non-backtracking operator or the Bethe Hessian. Compared to the former, the latter is a smaller, real and symmetric matrix. We therefore present here algorithms relying on the Bethe Hessian in the interest of numerical efficiency.

As stated in the previous section, when given a general Ising model with finite fields of the form (1), our first step, common to both the two upcoming procedures, is to define an associated Ising model in vanishing fields of the form (8). By bounding the partition function and correlations of the resulting model in vanishing fields, we obtain bounds on the partition function, magnetizations, and correlations of the original model using the correspondence recalled in proposition 1.1.

4

Algorithm 1 Bound on the log partition function of the Ising model in vanishing fields
---
Input: Graph $G = (n, E(G))$, couplings $(J_{ij})_{(ij) \in E(G)}$
Output: Upper bound $\mathcal{L}$ on the log partition function $\log \mathcal{Z}$
  1: Build the Bethe Hessian of equation (7)
  2: Output $\mathcal{L} = n \log 2 - \frac{1}{2} \log \det(\mathrm{H}) + 2 \sum_{(ij) \in E(G)} \log \cosh(J_{ij})$

Algorithm 1 describes our procedure to estimate the logarithm of the partition function. We will show in the following that under the assumptions of section 1.2, this simple algorithm yields an upper bound that we expect to be most accurate on graphs containing few (or large) loops. We note that on sparse graphs, the logarithm of the determinant of the Bethe Hessian can be computed efficiently by first performing a Cholesky decomposition of the sparse matrix H.

Algorithm 2 Bounds on the correlations of the Ising model in vanishing fields
---
Input: Graph $G = (n, E(G))$, couplings $(J_{ij})_{(ij) \in E(G)}$
Output: Upper bounds c_{ab} on the correlations χ_{ab}, for $a, b \in [n]$
  1: Build the Bethe Hessian of equation (7)
  2: Output $c_{ab} = \left(\mathrm{H}^{-1} \right)_{ab}$ for $a, b \in [n]$

Algorithm 2 yields an estimate of the correlations in the Ising model with vanishing fields which we will show to be an upper bound under the conditions of section 1.2. Note again that this algorithm can be used to compute an upper bound on the magnetizations of an Ising model with finite fields using proposition 1.1. Once more, for sparse models, this procedure can be made numerically efficient by using sparse linear solvers instead of directly inverting the Bethe Hessian.

2 Main results

We now state our main results and some of their consequences, leaving the proofs to the next section.

2.1 Bound on the partition function

Our first result is an upper bound on the partition function (2).

Theorem 2.1. *Consider model (8) with positive couplings $J_{ij} > 0$ for $(ij) \in E(G)$. Assume that the spectral radius $\rho(\mathrm{B}) < 1$, where B is the non-backtracking matrix defined in (6), and let H be the Bethe Hessian defined in (7). Then*

$$\mathcal{Z} \leq 2^n \det \left(\mathrm{I} - \mathrm{B} \right)^{-1/2} \prod_{(ij) \in E(G)} \cosh(J_{ij}) = 2^n \det(\mathrm{H})^{-1/2} \prod_{(ij) \in E(G)} \cosh(J_{ij})^2 . \tag{11}$$

Note that the non-backtracking matrix is a large $(2m \times 2m)$, non-symmetric object, non-trivial to build or manipulate. The equality in (11) allows to compute this bound without having to build the non-backtracking matrix, by using instead the smaller $(n \times n)$ and symmetric Bethe Hessian. This last equality is a simple consequence of the Ihara-Bass formula, which, in its multivariate version ([31], theorem 2) states that

$$\det(\mathrm{I} - \mathrm{B}) = \det(\mathrm{H}) \prod_{(ij) \in E(G)} \cosh(J_{ij})^{-2} . \tag{12}$$

5

In particular, this formula implies that on any *finite* graph, H is non-singular if $\rho(\mathrm{B}) < 1$. In the thermodynamic limit $n \to \infty$, some subtleties arise, as discussed in the next section.

As will be apparent from the proof, this bound is based on an over-counting of the subgraphs of G contributing to the high temperature expansion of the partition function (see section 3.1). These subgraphs correspond to closed paths (i.e. closed self-avoiding walks) that are notoriously hard to count [32]. To derive an analytical upper bound on the partition function, the set of contributing subgraphs must be included in a larger set of walks on the graph G which contribution can be computed analytically. A simple bound can be obtained by including the set of closed paths in the set of all closed walks on the graph G. This approach yields an analytical bound similar to Theorem 2.1 expressed in terms of the adjacency matrix of the graph, namely

$$\mathcal{Z} \leq 2^n \det\left(\mathrm{I} - A\right)^{-1} \prod_{(ij) \in E(G)} \cosh(J_{ij}), \tag{13}$$

whenever $\rho(A) < 1$, where A is the weighted adjacency matrix of the graph G, with entries $A_{ij} = \tanh(J_{ij})\mathbf{1}(i \in \partial j)$ (see section 3.2 for details). This bound is, however, too loose, because it counts many spurious contributions, in particular walks that are allowed to go back and forth on the same edge. We improve this bound by forbidding that the walk immediately backtracks to the previous edge. This is achieved by replacing the adjacency matrix with the non-backtracking operator. From the proof of section 3.2, it is straightforward to see that the bound (11) is less tight if the graph G contains many loops. In practice, however, figure 1 shows that using the non-backtracking operator instead of the adjacency matrix leads to a dramatic improvement, even in the case of a 3D lattice Ising model, which has many loops.

2.2 Bound on the susceptibility

We now state our upper bound on the correlations, encoded in the susceptibility matrix χ of equation (4).

Theorem 2.2. *Consider model (8) with positive couplings $J_{ij} > 0$ for $(ij) \in E(G)$. Assume that $\rho(\mathrm{B}) < 1$, where B is the non-backtracking matrix defined in (6). Define the matrices $\mathrm{P}, \mathrm{Q} \in \mathbb{R}^{n \times 2m}$ by their elements, for $a \in [n], (ij) \in E(G)$:*

$$\mathrm{P}_{a,(i \to j)} = \tanh(J_{ij})\mathbf{1}(a = j), \qquad \mathrm{Q}_{a,(i \to j)} = \mathbf{1}(a = i). \tag{14}$$

Then

$$\chi \leq \mathrm{P}(\mathrm{I} - \mathrm{B})^{-1}\mathrm{Q}^{\mathsf{T}} + \mathrm{I}_{n \times n}, \tag{15}$$

where the inequality holds element-wise.

Once more, it is possible to express this last result in terms of the Bethe Hessian rather than the non-backtracking operator, yielding a surprisingly simple result.

Corollary 2.3. *Let H be the matrix defined in (7). Under the same assumptions as theorem 2.2, it holds that H is invertible, and*

$$\chi \leq \mathrm{H}^{-1}, \tag{16}$$

where the inequality holds element-wise.

As for the partition function, these bounds rely on the inclusion of the set of paths between two spins in a larger, tractable set of walks on the graph G (see section 3.3). More precisely, expression (15) follows from the inclusion of the set of paths between two fixed spins $a, b \in [n]$ in the set of

non-backtracking walks starting at a and ending at b. If we allow these walks to backtrack, i.e. if we include the set of paths in the set of all walks, we get the looser bound

$$\chi \leq (\mathrm{I}_n - A)^{-1},\tag{17}$$

whenever $\rho(A) < 1$ where A is the weighted adjacency matrix with elements $A_{ij} = \tanh(J_{ij})\mathbf{1}(i \in \partial j)$. Once more, forbidding backtracking allows to dramatically improve the bound on the susceptibility, as shown in figure 1.

While valid only on finite graphs, these results allow us, in certain cases, to derive a bound on the paramagnetic to ferromagnetic phase transition. More precisely, for a sequence of Ising models, the previous results allow to bound the scalar susceptibility, defined as

$$\bar{\chi} := \frac{1}{n} \sum_{x,y=1}^{n} \chi_{xy}.\tag{18}$$

Corollary 2.4. *Consider a sequence $(I_p)_{p \in \mathbb{N}}$ of Ising models of the form (8), each of them defined on a graph G_p, with positive couplings $J_{ij}^{(p)} > 0$, for $(ij) \in E(G_p)$, and scalar susceptibility (18) denoted $\bar{\chi}_p$. Define B_p to be the non-backtracking operator defined in (6) for the Ising model I_p, and assume that $\forall p \in \mathbb{N}$, $\rho(\mathrm{B}_p) < 1$. Define H_p to be the Bethe Hessian defined in (7) for the Ising model I_p, and let $\lambda_{\min}(\mathrm{H}_p)$ denote its smallest eigenvalue. Assume that there exists $\epsilon > 0$ such that $\forall p \in \mathbb{N}$, $\lambda_{\min}(\mathrm{H}_p) > \epsilon$. Then $\forall p \in \mathbb{N}$,*

$$\bar{\chi}_p \leq \frac{1}{\epsilon}\tag{19}$$

Proof: For any fixed $p \in \mathbb{N}$, let n_p be the number of vertices of the graph G_p. Defining $\mathrm{U}_p \in \mathbb{R}^{n_p}$ to be the vector with all its entries equal to 1, and denoting by $||.||_p$ the Euclidean norm, we have from Corollary 2.3

$$\bar{\chi}_p \leq \frac{1}{n_p} \sum_{x,y=1}^{n_p} (\mathrm{H}_p)_{xy}^{-1} = \frac{\mathrm{U}_p^\intercal \mathrm{H}_p^{-1} \mathrm{U}_p}{||\mathrm{U}_p||_p^2} \leq \rho(\mathrm{H}_p^{-1}) = \frac{1}{\lambda_{\min}(\mathrm{H}_p)} \leq \frac{1}{\epsilon},\tag{20}$$

where we have used that the matrix H_p is symmetric.

For sequences of graphs that admit a thermodynamic limit, Corollary 2.4 implies a condition under which the infinite Ising model is in the paramagnetic phase, therefore yielding a bound on the paramagnetic to ferromagnetic transition. As an explicit example, let us consider the case of a sequence $(G_p)_{p \in \mathbb{N}}$ of d-regular graphs with uniform couplings $J_{ij} = \beta$. We assume that the number of vertices in G_p goes to infinity as $p \to \infty$. It is straightforward to check that, for any $p \in \mathbb{N}$,

$$\rho(\mathrm{B}_p) = (d-1)\tanh(\beta) \qquad \text{and} \qquad \lambda_{\min}(\mathrm{H}_p) = 1 - \frac{d\tanh(\beta)}{1+\tanh(\beta)}.\tag{21}$$

By application of Corollary 2.4, it follows that if

$$(d-1)\tanh(\beta) < 1,\tag{22}$$

the scalar susceptibility $\bar{\chi}_p$ remains bounded. Equivalently, if the sequence admits a thermodynamic limit with a phase transition at β_c, then

$$\beta_c \geq \operatorname{atanh} \frac{1}{d-1}.\tag{23}$$

7

Interestingly, the right hand side of the last equation corresponds to the critical inverse temperature in the Bethe approximation, which is already known to be a lower bound on β_c for certain ferromagnetic models on hypercubical lattices [1]. Our contribution generalizes such previous results, giving a simple algorithm to derive a lower bound on arbitrary graphs, with arbitrary (positive) couplings. It is worth noting that, perhaps unsurprisingly, this bound is tight on sparse random d-regular graphs with uniform couplings $\forall (ij) \in E(G), J_{ij} = \beta$. Indeed, on such locally tree-like graphs, the existence of the thermodynamic limit has been proved, and, the critical temperature has been shown to verify $(d-1)\tanh(\beta) = 1$ [33, 34]. More generally, we expect the implicit bound provided by Corollary 2.4 on the transition of the Ising model to be tight on any sparse random graph ensemble whose degree distribution has a finite second moment.

From the Ihara-Bass formula (12), it holds that on any *finite* graph, $\lambda_{\min}(H) > 0$ as long as $\rho(B) < 1$. Therefore one may expect that the condition on the smallest eigenvalue of H_p in Corollary 2.4 could be relaxed, e.g. by assuming instead that $\rho(B) < 1 - \epsilon$ for some $\epsilon > 0$. This relaxation turns out to be not possible, because it may happen that the spectral radius of B is bounded away from 1, while the smallest eigenvalue of H tends to 0, the limit $p \to \infty$. As an example, take G_p to be the star graph with $n_p = p + 1$ spins and edges $(i, n+1)$ for $i \in [n]$, and uniform couplings $J_{ij} = \beta$. Since G_p is a tree, and the non-backtracking operator is nilpotent on trees, it holds that $\rho(B_p) = 0$ for all p. On the other hand, the spectrum of H_p can be computed explicitly, and its smallest eigenvalue shown to verify

$$\lambda_{\min}(H_p) \underset{p \to \infty}{\sim} \frac{1}{p \tanh(\beta)^2} \underset{p \to \infty}{\longrightarrow} 0. \tag{24}$$

2.3 Relation to belief propagation and susceptibility propagation

A standard approximation to compute the magnetizations of model (8) is the belief propagation algorithm, or cavity method in statistical physics [35], which solutions can be shown to yield stationary points of the Bethe free energy [36]. Belief propagation is a recursive message-passing algorithm that approximates the magnetizations $m_i = \mathbb{E}(s_i)$ as

$$m_i \approx \tanh\left(\sum_{l \in \partial i} \operatorname{atanh}\left(m_{l \to i} \tanh J_{il} \right) \right), \tag{25}$$

where the so-called cavity fields $(m_{i \to j})_{(i \to j) \in \vec{E}(G)}$ are defined on the set $\vec{E}(G)$ of directed edges of G, and verify the fixed point equation

$$m_{i \to j} = \tanh\left(\sum_{l \in \partial i \backslash j} \operatorname{atanh}\left(m_{l \to i} \tanh J_{il} \right) \right). \tag{26}$$

In practice, starting from a random initial condition, one iterates equation (26) until convergence, and outputs the result of (25). Belief propagation is known to be exact on trees, and widely believed to yield asymptotically accurate results for sparse, locally-tree like graphs, as well as models with small couplings [35].

The message-passing approach can be extended to allow the computation of the correlations χ_{ij} for any $i, j \in [n]$ using the fluctuation dissipation theorem. The resulting algorithm is called susceptibility propagation [37, 2], and approximates χ_{ij} as

$$\chi_{ij} \approx (1 - m_i^2) \left(\mathbf{1}(i = j) + \sum_{l \in \partial i} \frac{\chi_{l \to i, j} \tanh J_{il}}{1 - m_{l \to i}^2 \tanh^2 J_{il}} \right), \tag{27}$$

8

where the $(m_{i \to j})_{(i \to j) \in \vec{E}(G)}$ are the fixed point of (26), the $(m_i)_{i \in [n]}$ are the belief propagation estimates of the magnetization derived from (25) and the $(\chi_{i \to j,k})_{(i \to j) \in \vec{E}(G), k \in [n]}$ verify the fixed point equation

$$\chi_{i \to j,k} = (1 - m_{i \to j}^2) \left(\mathbf{1}(i = k) + \sum_{l \in \partial i \backslash j} \frac{\chi_{l \to i,k} \tanh J_{il}}{1 - m_{l \to i}^2 \tanh^2 J_{il}} \right), \tag{28}$$

Similarly to belief propagation, starting from a random initial condition, equation (28) is first iterated until convergence, and the correlations are then estimated from equation (27). Note that it is possible to invert the relation between the susceptibilities and the couplings, resulting in an inference algorithm for solving the inverse Ising model. This algorithm has been shown to yield better results than other mean field approaches on certain problems ([37, 38]).

While only exact on trees, both belief propagation and susceptibility propagation have been observed to yield good approximate results on more general topologies, when they converge. However, these algorithms are based on the Bethe approximation [36], which, unlike the naive mean-field approach, does not provide bounds on the actual partition function [3]. The following result might therefore appear surprising.

Corollary 2.5. *Consider model (8) with positive couplings $J_{ij} > 0$ for $(ij) \in E(G)$. Assume that $\rho(\mathrm{B}) < 1$, where B is the non-backtracking matrix defined in (6). Then $(m_{i \to j} = 0)_{(i \to j) \in \vec{E}(G)}$ is a* stable *fixed point of the belief propagation recursion (26). Additionally, the corresponding susceptibility propagation algorithm converges, and yields an upper bound on the true correlations, regardless of the topology of the graph.*

Proof: The fact that $(m_{i \to j} = 0)_{(i \to j) \in \vec{E}(G)}$ is a fixed point of belief propagation is readily checked on equation (26). To see that it is stable, starting from a small perturbation $\delta m_{i \to j}^0$, we have at first order at iteration $t \geq 1$

$$\delta m_{i \to j}^t = \sum_{l \in \partial i \backslash j} \delta m_{l \to i}^{t-1} \tanh J_{il}, \tag{29}$$

which in matrix form can be written $\delta m^t = \mathrm{B}\, \delta m^{t-1}$. The stability of this fixed point follows from the assumption $\rho(\mathrm{B}) < 1$. The corresponding belief propagation solution is $m_i = 0, \forall i \in [n]$. The corresponding susceptibility propagation recursion reads

$$\chi_{i \to j,k} = \mathbf{1}(i = k) + \sum_{l \in \partial i \backslash j} \chi_{l \to i,k} \tanh J_{il}. \tag{30}$$

Define a vector $\chi_k \in \mathbb{R}^{2m}$ with elements $\chi_{i \to j,k}$, for $(i \to j) \in \vec{E}(G)$, and call Q_k the k-th line of the matrix Q defined in Theorem 2.2. Then equation (30) can be rewritten in matrix form as

$$\chi_k = \mathrm{Q}_k{}^\intercal + \mathrm{B}\chi_k. \tag{31}$$

which solution $\chi_k = (\mathrm{I} - \mathrm{B})^{-1}\mathrm{Q}_k{}^\intercal$ exists and is unique, since $\rho(\mathrm{B}) < 1$. Iterating equation (30) starting from the initial condition χ_k^0, we get at iteration $t \geq 1$

$$\chi_k^t - \chi_k = \mathrm{B}\left(\chi_k^{t-1} - \chi_k \right), \tag{32}$$

so that $\chi_k^t \to \chi_k$ as $t \to \infty$, using again that $\rho(B) < 1$. Finally, using equation (27), it is straightforward to check that the correlations output by susceptibility propagation are given in

9

matrix form by $P(I - B)^{-1}Q^\intercal + I_{n \times n}$, which is, by Theorem 2.2, an upper bound on the true correlations.
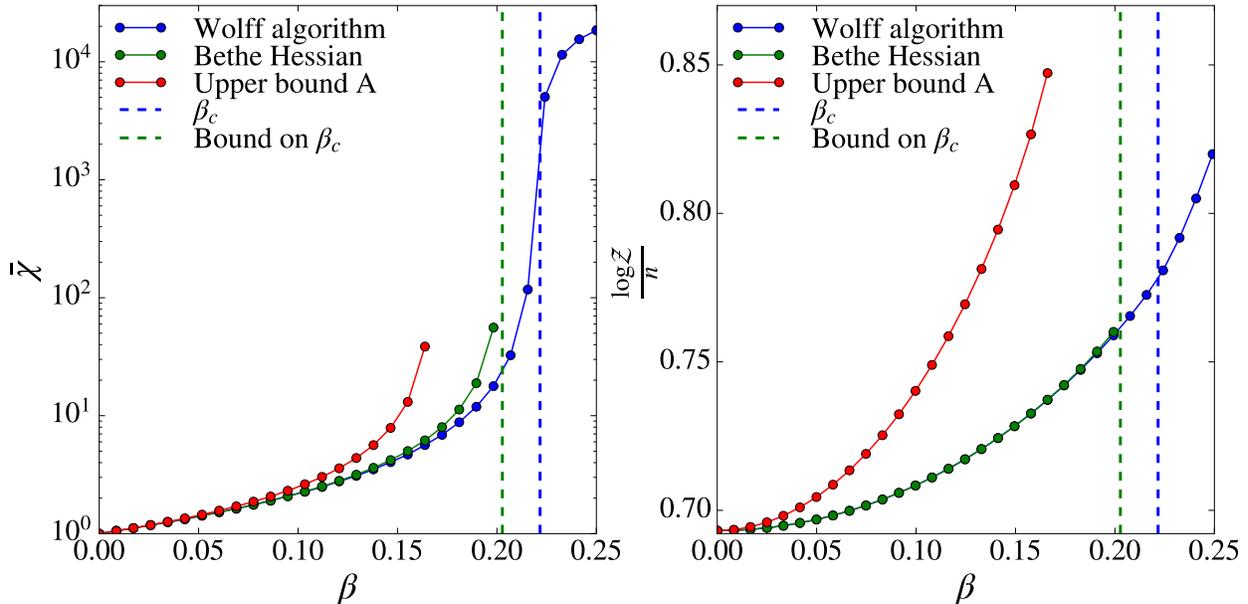

Figure 1: Numerical simulation of the 3D lattice Ising model with $n = 32^3$ spins, with periodic boundary conditions, and uniform couplings $J_{ij} = \beta$. The left panel is the scalar susceptibility $\bar{\chi}$ of equation (18), and the right panel is the logarithm of the partition function defined in equation (2). For both quantities we represented the exact value obtained via a computationally expensive Monte Carlo simulation (using the Wolff algorithm [22]), and the upper bounds (11) and (16) expressed in terms of the Bethe Hessian operator. The dashed lines signal the critical temperature β_c of the 3D Ising model as computed numerically in [39], and the lower bound on this critical temperature provided by the Bethe Hessian (23). Finally, we included for comparison the upper bound obtained by using the adjacency matrix A instead of the non-backtracking matrix (equations (13) and (17)).

3   Proofs

3.1   High temperature expansion

Central to our results is the high temperature expansion of the partition function, which we quickly rederive here. We use the following rewriting of the Boltzmann weight, relying on the fact that the spins are binary variables equal to ± 1

$$\exp\left(J_{ij}s_i s_j\right) = a_{ij}(1 + b_{ij}s_i s_j),\tag{33}$$

with $a_{ij} = \cosh(J_{ij})$, $b_{ij} = \tanh(J_{ij})$. The partition function of model (8) is given by

$$\mathcal{Z} = \sum_{s \in \{\pm 1\}^n} \prod_{(ij) \in E(G)} \exp\left(J_{ij} s_i s_j\right) \tag{34}$$

$$= \left(\prod_{(ij) \in E(G)} a_{ij}\right) \sum_{s \in \{\pm 1\}^n} \prod_{(ij) \in E(G)} (1 + b_{ij} s_i s_j) \tag{35}$$

$$= \left(\prod_{(ij) \in E(G)} a_{ij}\right) \sum_{s \in \{\pm 1\}^n} \left[1 + \sum_{(ij) \in E(G)} b_{ij} s_i s_j + \sum_{(ij),(kl) \in E(G)} b_{ij} b_{kl} s_i s_j s_k s_l + \cdots\right]. \tag{36}$$

After summing on the spin configurations $s \in \{\pm 1\}^n$, the only terms in the expansion that do not vanish are the ones supported on a subgraph of G where all nodes have even degree. These are closed paths (not necessarily connected). Each of these closed paths contribute a factor 2^n to the partition function. We can therefore rewrite the partition function in the following form, called *high temperature expansion*

$$\mathcal{Z} = 2^n \left(\prod_{(ij) \in E(G)} a_{ij}\right) \left[1 + \sum_{g \in \mathcal{C}} \prod_{(ij) \in E(g)} b_{ij}\right], \tag{37}$$

where $\mathcal{C}$ is the set of closed paths, possibly disconnected.

3.2   Proof of Theorem 2.1

We now introduce the set $\mathcal{C}_l^c$ of *connected* closed paths of length l. We denote by $\mathcal{W}_l^c$ the sum of all contributions to the partition function coming from connected closed paths of length $l \geq 1$, i.e.

$$\mathcal{W}_l^c = \sum_{g \in \mathcal{C}_l^c} \prod_{(ij) \in E(g)} b_{ij}. \tag{38}$$

We have the inequality

$$\mathcal{Z} \leq 2^n \left(\prod_{(ij) \in E(G)} a_{ij}\right) \left[1 + \sum_{l \geq 1} \sum_{n_c \geq 1} \frac{1}{n_c!} \sum_{l_1 + l_2 + \cdots + l_{n_c} = l} \mathcal{W}_{l_1}^c \mathcal{W}_{l_2}^c \cdots \mathcal{W}_{l_{n_c}}^c\right] \tag{39}$$

$$\leq 2^n \left(\prod_{(ij) \in E(G)} a_{ij}\right) \left[1 + \sum_{n_c \geq 1} \frac{1}{n_c!} \left(\sum_{l \geq 1} \mathcal{W}_l^c\right)^{n_c}\right]. \tag{40}$$

Indeed, the product $\mathcal{W}_{l_1}^c \mathcal{W}_{l_2}^c \cdots \mathcal{W}_{l_{n_c}}^c$ contains all the contributions coming from disconnected closed paths which connected components are of size $l_1, \cdots, l_{n_c}$. The factor $n_c!$ accounts for all the permutations of the factors in this product. This is only an inequality because we are counting graphs in which some edges appear more than once. However, the factors $\mathcal{W}_l^c$ are still hard to compute, and we will look for an upper bound. A simple upper bound can be derived by considering a weighted version of the adjacency matrix of the graph. Define $A \in \mathbb{R}^{n \times n}$ by its entries

$$A_{ij} = \begin{cases} b_{ij} & \text{if } (ij) \in E(G) \\ 0 & \text{otherwise} \end{cases}, \tag{41}$$

then it holds that

$$\mathcal{W}_l^c \leq \frac{\text{Tr}A^l}{l}. \tag{42}$$

Indeed, the right hand side contains all contributions coming from closed walks that start and end at the same point. The factor $1/l$ accounts for the choice of the starting point. However, as hinted at in section 2.1, this upper bound is loose, because the adjacency matrix allows backtracking, and therefore going back and forth on the same edge, whereas such contributions do not appear in the partition sum. To improve this bound, we use a matrix that forbids backtracking: the operator B of equation (6). Since connected closed paths are closed non-backtracking walks (although the converse is not true), it holds that

$$\mathcal{W}_l^c \leq \frac{\text{Tr}B^l}{2l}. \tag{43}$$

where the additional factor $1/2$ accounts for the degeneracy due to the orientation of the non-backtracking closed walk. Under the assumption that $\rho(B) < 1$, the following bound holds:

$$\sum_{l \geq 1} \mathcal{W}_l^c \leq \sum_{l \geq 1} \frac{\text{Tr}B^l}{2l} = -\frac{1}{2}\text{Tr}\log(I - B) = -\frac{1}{2}\log\det(I - B), \tag{44}$$

where $I \in \mathbb{R}^{2m \times 2m}$ is the identity matrix. This, along with the Ihara-Bass formula (12) completes the proof of Theorem 2.1.

3.3 Proof of Theorem 2.2

The correlation functions are given, for any $x, y \in [n]$, by

$$\mathbb{E}(s_x s_y) = \frac{\sum\limits_{s \in \{\pm 1\}^n} s_x s_y \prod\limits_{(ij) \in E(G)} (1 + b_{ij} s_i s_j)}{\sum\limits_{s \in \{\pm 1\}^n} \prod\limits_{(ij) \in E(G)} (1 + b_{ij} s_i s_j)} = \frac{\sum\limits_{g \in \mathcal{P}_{xy} + \mathcal{C}} \prod\limits_{(ij) \in E(g)} b_{ij}}{1 + \sum\limits_{g' \in \mathcal{C}} \prod\limits_{(ij) \in E(g')} b_{ij}} \tag{45}$$

where:

- $\mathcal{P}_{xy}$ is the set of paths from x to y.

- $\mathcal{C}$ is, as in the previous section, the set of closed paths.

- $\mathcal{P}_{xy} + \mathcal{C}$ is the set of diagrams made of a path from x to y and any number of disconnected closed paths such that each edge is selected at most once. The closed paths therefore do not intersect the path from x to y.

We assume without loss of generality that $x \neq y$. From the previous definitions, we have that

$$\mathbb{E}(s_x s_y) \leq \frac{\left(\sum\limits_{g \in \mathcal{P}_{xy}} \prod\limits_{(ij) \in E(g)} b_{ij}\right)\left(1 + \sum\limits_{g' \in \mathcal{C}} \prod\limits_{(ij) \in E(g')} b_{ij}\right)}{1 + \sum\limits_{g' \in \mathcal{C}} \prod\limits_{(ij) \in E(g')} b_{ij}} = \sum\limits_{g \in \mathcal{P}_{xy}} \prod\limits_{(ij) \in E(g)} b_{ij}. \tag{46}$$

Indeed, developing the product at the numerator gives a sum of positive contributions including the ones in $\mathcal{P}_{xy} + \mathcal{C}$ and also (positive) spurious contributions coming from diagrams where the closed

12

loops intersect the path from x to y. We now introduce the set $\mathcal{N}^l_{(x \to x'),(y' \to y)}$ of non-backtracking walks of length l, starting on the directed edge $(x \to x')$ and terminating on the edge $(y' \to y)$. For a non-backtracking walk w, we denote by $\mathcal{E}(w)$ the list of edges crossed by w, where each edge appears with a multiplicity equal to the number of times w crosses this edge. Since any path in $\mathcal{P}_{xy}$ is a non-backtracking walk (though the reverse is again, in the presence of loops, not true), it holds that

$$\mathbb{E}(s_x s_y) \le \sum_{l \ge 1} \sum_{\substack{x' \in \partial x \\ y' \in \partial y}} \sum_{w \in \mathcal{N}^l_{(x \to x'),(y' \to y)}} \prod_{(ij) \in \mathcal{E}(w)} b_{ij} \tag{47}$$

where ∂x denotes the set of neighbors of x in the graph G. In order to write this last expression in terms of the non-backtracking operator, we introduce the vector $u_{x \to x'} \in \mathbb{R}^{2m}$ with entries all equal to 0 except for the $(x \to x')$ entry, which is equal to 1. Similarly, we introduce the vector $v_{y' \to y} \in \mathbb{R}^{2m}$ which only non-zero entry is equal to $b_{yy'}$, in position $(y \to y')$. Then we have

$$\sum_{g \in \mathcal{N}^l_{(x \to x'),(y' \to y)}} \prod_{(ij) \in E(g)} b_{ij} = (v_{y' \to y})^\mathsf{T} \, \mathrm{B}^{l-1} \, u_{x \to x'} \, , \tag{48}$$

so that

$$\mathbb{E}(s_x s_y) \le \left(\sum_{y' \in \partial y} v_{y' \to y} \right)^\mathsf{T} \sum_{l \ge 1} \mathrm{B}^{l-1} \left(\sum_{x' \in \partial x} u_{x \to x'} \right) \tag{49}$$

$$\le \left(\sum_{y' \in \partial y} v_{y' \to y} \right)^\mathsf{T} (\mathrm{I} - \mathrm{B})^{-1} \left(\sum_{x' \in \partial x} u_{x \to x'} \right) \tag{50}$$

where we have used the assumption $\rho(\mathrm{B}) < 1$, so that the series of powers of B is summable. To write this last equation in a more compact way, recall the definition of the *susceptibility matrix* $\chi \in \mathbb{R}^{n \times n}$, with elements $\chi_{xy} = \mathbb{E}(s_x s_y)$. Then it holds element-wise that

$$\chi \le \mathrm{P}(\mathrm{I} - \mathrm{B})^{-1}\mathrm{Q}^\mathsf{T} + \mathrm{I}_{n \times n} \tag{51}$$

where $\mathrm{P}, \mathrm{Q} \in \mathbb{R}^{n \times 2m}$ are defined in equation (14). Note that the addition of an identity matrix ensures that the inequality also holds on the diagonal of χ. This completes the proof of Theorem 2.2.

3.4   Proof of Corollary 2.3

The corollary will follow from the following lemma, which is of independent interest because it provides another connection between the non-backtracking matrix and the Bethe Hessian, besides the Ihara-Bass formula (12).

Lemma 3.1. *Let* B *be the non-backtracking matrix matrix defined in (6), and* H *the Bethe Hessian defined in (7). Let* P, Q *be the matrices defined in (14). Assume that* $\mathrm{I} - \mathrm{B}$ *is invertible. Then* H *is also invertible, and*

$$\mathrm{P}(\mathrm{I} - \mathrm{B})^{-1}\mathrm{Q}^\mathsf{T} + \mathrm{I}_{n \times n} = \mathrm{H}^{-1} \, . \tag{52}$$

Proof: The fact that H is invertible if $\mathrm{I} - \mathrm{B}$ is invertible follows from the Ihara-Bass formula (12). Take $x \in \mathbb{R}^n$, and call $y = (\mathrm{I} - \mathrm{B})^{-1}\mathrm{Q}^\mathsf{T}x$. We wish to show that $\mathrm{H}\mathrm{P}y + \mathrm{H}x = x$. Denoting by $(Py)_i$

13

for $i = 1 \cdots n$ the i-th component of the vector Py, we have

$$(Py)_i = \sum_{(k \to l)} b_{kl} \delta_{il} y_{k \to l} \tag{53}$$

$$= \sum_{k \in \partial i} b_{ki} y_{k \to i} \,. \tag{54}$$

y verifies the following equation

$$(\mathrm{I} - \mathrm{B})y = \mathrm{Q}^\intercal x \tag{55}$$

so that for all directed edges $(i \to j)$

$$y_{i \to j} - \sum_{k \in \partial i \backslash j} b_{ik} y_{k \to i} = x_i \tag{56}$$

which can be rewritten as

$$y_{i \to j} - (\mathrm{P}y)_i + b_{ij} y_{j \to i} = x_i \,, \tag{57}$$

$$y_{j \to i} - (\mathrm{P}y)_j + b_{ij} y_{i \to j} = x_j \,, \tag{58}$$

where the second equation is for the link $(j \to i)$. Together, these two questions form a closed set of equations allowing to compute $y_{i \to j}$ and $y_{j \to i}$. More precisely, we get

$$y_{i \to j} = \frac{1}{1 - b_{ij}^2} \Big(x_i + (\mathrm{P}y)_i - b_{ij} \left(x_j + (\mathrm{P}y)_j \right) \Big) \,. \tag{59}$$

We can now insert this expression in eq. 54 and get

$$(\mathrm{P}y)_i = \sum_{k \in \partial i} b_{ik} y_{k \to i} = \sum_{k \in \partial i} \frac{b_{ik}}{1 - b_{ik}^2} \Big(x_k + (\mathrm{P}y)_k - b_{ik} \left(x_i + (\mathrm{P}y)_i \right) \Big) \tag{60}$$

which is equivalent to

$$\left(1 + \sum_{k \in \partial i} \frac{b_{ik}^2}{1 - b_{ik}^2} \right) (\mathrm{P}y)_i - \sum_{k \in \partial i} \frac{b_{ik}}{1 - b_{ik}^2} (\mathrm{P}y)_k = \sum_{k \in \partial i} \frac{b_{ik}}{1 - b_{ik}^2} x_k - \left(\sum_{k \in \partial i} \frac{b_{ik}^2}{1 - b_{ik}^2} \right) x_i \tag{61}$$

which in matrix form reads

$$\mathrm{H}\mathrm{P}y = -\mathrm{H}x + x \tag{62}$$

which completes the proof.

Acknowledgements

References

[1] Michael E Fisher. Critical temperatures of anisotropic ising lattices. ii. general upper bounds. *Physical Review*, 162(2):480, 1967.

[2] Marc Mezard and Thierry Mora. Constraint satisfaction problems and neural networks: A statistical physics perspective. *Journal of Physiology-Paris*, 103(1):107–113, 2009.

[3] M. J. Wainwright and M. I. Jordan. Graphical models, exponential families, and variational inference. *Foundations and Trends in Machine Learning*, 1, 2008.

[4] Daphne Koller and Nir Friedman. *Probabilistic graphical models: principles and techniques*. MIT press, 2009.

[5] DM Greig, BT Porteous, and Allan H Seheult. Exact maximum a posteriori estimation for binary images. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 271–279, 1989.

[6] Marcelo Blatt, Shai Wiseman, and Eytan Domany. Superparamagnetic clustering of data. *Physical review letters*, 76(18):3251, 1996.

[7] John P Barton, Eleonora De Leonardis, Alice Coucke, and Simona Cocco. Ace: adaptive cluster expansion for maximum entropy graphical model inference. *bioRxiv*, page 044677, 2016.

[8] Faruck Morcos, Andrea Pagnani, Bryan Lunt, Arianna Bertolino, Debora S Marks, Chris Sander, Riccardo Zecchina, José N Onuchic, Terence Hwa, and Martin Weigt. Direct-coupling analysis of residue coevolution captures native contacts across many protein families. *Proceedings of the National Academy of Sciences*, 108(49):E1293–E1301, 2011.

[9] David H Ackley, Geoffrey E Hinton, and Terrence J Sejnowski. A learning algorithm for boltzmann machines. *Cognitive science*, 9(1):147–169, 1985.

[10] Mark Kac and John C Ward. A combinatorial solution of the two-dimensional ising model. *Physical Review*, 88(6):1332, 1952.

[11] Pieter W Kasteleyn. Dimer statistics and phase transitions. *Journal of Mathematical Physics*, 4(2):287–293, 1963.

[12] Michael E Fisher. On the dimer solution of planar ising models. *Journal of Mathematical Physics*, 7(10):1776–1781, 1966.

[13] Nicol N Schraudolph and Dmitry Kamenetsky. Efficient exact inference in planar ising models. In *Advances in Neural Information Processing Systems*, pages 1417–1424, 2009.

[14] Jason K Johnson, Diane Oyen, Michael Chertkov, and Praneeth Netrapalli. Learning planar ising models. *arXiv preprint arXiv:1502.00916*, 2015.

[15] Vladimir Kolmogorov and Ramin Zabin. What energy functions can be minimized via graph cuts? *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 26(2):147–159, 2004.

[16] Yuri Boykov, Olga Veksler, and Ramin Zabih. Fast approximate energy minimization via graph cuts. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 23(11):1222–1239, 2001.

[17] Yuri Boykov, Olga Veksler, and Ramin Zabih. Markov random fields with efficient approximations. In *Computer vision and pattern recognition, 1998. Proceedings. 1998 IEEE computer society conference on*, pages 648–655. IEEE, 1998.

[18] Nicholas Ruozzi. The bethe partition function of log-supermodular graphical models. In *Advances in Neural Information Processing Systems*, pages 117–125, 2012.

[19] Alan S Willsky, Erik B Sudderth, and Martin J Wainwright. Loop series and bethe variational bounds in attractive graphical models. In *Advances in neural information processing systems*, pages 1425–1432, 2008.

[20] Michael Chertkov and Vladimir Y Chernyak. Loop series for discrete statistical models on graphs. *Journal of Statistical Mechanics: Theory and Experiment*, 2006(06):P06009, 2006.

[21] Robert H Swendsen and Jian-Sheng Wang. Nonuniversal critical dynamics in monte carlo simulations. *Physical review letters*, 58(2):86, 1987.

[22] Ulli Wolff. Collective monte carlo updating for spin systems. *Physical Review Letters*, 62(4):361, 1989.

[23] Massimo Campostrini, Andrea Pelissetto, Paolo Rossi, and Ettore Vicari. 25th-order high-temperature expansion results for three-dimensional ising-like systems on the simple-cubic lattice. *Physical Review E*, 65(6):066127, 2002.

[24] Simona Cocco and Rémi Monasson. Adaptive cluster expansion for inferring boltzmann machines with noisy data. *Physical review letters*, 106(9):090601, 2011.

[25] Ki-ichiro Hashimoto. Zeta functions of finite graphs and representations of p-adic groups. *Automorphic forms and geometry of arithmetic varieties.*, pages 211–280, 1989.

[26] F. Krzakala, C. Moore, E. Mossel, J. Neeman, A. Sly, L. Zdeborová, and P. Zhang. Spectral redemption in clustering sparse networks. *Proc. Natl. Acad. Sci. U.S.A.*, 110(52):20935–20940, 2013.

[27] Pan Zhang. Non-backtracking operator for ising model and its application in attractor neural networks. *arXiv:1409.3264*, 2014.

[28] Alaa Saade, Marc Lelarge, Florent Krzakala, and Lenka Zdeborová. Clustering from sparse pairwise measurements. *arXiv preprint arXiv:1601.06683*, 2016.

[29] Alaa Saade, Florent Krzakala, and Lenka Zdeborová. Spectral clustering of graphs with the bethe hessian. In *Advances in Neural Information Processing Systems*, pages 406–414, 2014.

[30] Alaa Saade, Florent Krzakala, and Lenka Zdeborová. Matrix completion from fewer entries: Spectral detectability and rank estimation. In *Advances in Neural Information Processing Systems*, pages 1261–1269, 2015.

[31] Yusuke Watanabe and Kenji Fukumizu. Graph zeta function in the bethe free energy and loopy belief propagation. In *NIPS*, pages 2017–2025, 2009.

[32] Michael E Fisher and David S Gaunt. Ising model and self-avoiding walks on hypercubical lattices and "high-density" expansions. *Physical Review*, 133(1A):A224, 1964.

[33] Amir Dembo, Andrea Montanari, et al. Ising models on locally tree-like graphs. *The Annals of Applied Probability*, 20(2):565–592, 2010.

[34] Sander Dommers, Cristian Giardinà, and Remco van der Hofstad. Ising critical exponents on random trees and graphs. *Communications in Mathematical Physics*, 328(1):355–395, 2014.

[35] Marc Mezard and Andrea Montanari. *Information, physics, and computation.* Oxford University Press, 2009.

[36] Jonathan S Yedidia, William T Freeman, and Yair Weiss. Bethe free energy, kikuchi approximations, and belief propagation algorithms. *Advances in neural information processing systems*, 13, 2001.

[37] Thierry Mora. *Géométrie et inférence dans l'optimisation et en théorie de l'information.* PhD thesis, Université Paris Sud-Paris XI, 2007.

[38] F. Ricci-Tersenghi. The bethe approximation for solving the inverse ising problem: a comparison with other inference methods. *J. Stat. Mech.: Th. and Exp.*, page P08015, 2012.

[39] Tobias Preis, Peter Virnau, Wolfgang Paul, and Johannes J Schneider. Gpu accelerated monte carlo simulation of the 2d and 3d ising model. *Journal of Computational Physics*, 228(12):4468–4477, 2009.