arXiv:1906.06431v1 [cs.DS] 14 Jun 2019

A New Family of Tractable Ising Models

Valerii Likhosherstov[1], Yury Maximov[(2,1)] and Michael Chertkov[(3,2,1)]
[1] Skolkovo Institute of Science and Technology, Moscow, Russia
[2] Theoretical Division and Center for Nonlinear Studies,
Los Alamos National Laboratory, Los Alamos, NM, USA
[3] Graduate Program in Applied Mathematics,
University of Arizona, Tucson, AZ, USA

June 18, 2019

Abstract

We present a new family of zero-field Ising models over N binary variables/spins obtained by consecutive "gluing" of planar and $O(1)$-sized components along with subsets of at most three vertices into a tree. The polynomial time algorithm of the dynamic programming type for solving exact inference (partition function computation) and sampling consists of a sequential application of an efficient (for planar) or brute-force (for $O(1)$-sized) inference and sampling to the components as a black box. To illustrate utility of the new family of tractable graphical models, we first build an $O(N^{\frac{3}{2}})$ algorithm for inference and sampling of the K_5-minor-free zero-field Ising models—an extension of the planar zero-field Ising models—which is neither genus- nor treewidth-bounded. Second, we demonstrate empirically an improvement in the approximation quality of the NP-hard problem of the square-grid Ising model (with non-zero field) inference.

1 Introduction

Let $G = (V, E)$ be an undirected graph with a set of vertices $V(G)$ and a set of normal edges $E(G)$ (no loops or multiple edges). We discuss *Ising models* which associate the following probability to each random $N \triangleq |V(G)|$-dimensional binary variable/spin configuration $X \in \{\pm 1\}^N$:

$$\mathbb{P}(X) \triangleq \frac{\mathcal{W}(X)}{Z}, \tag{1}$$

where

$$\mathcal{W}(X) \triangleq \exp\Bigg(\sum_{v\in V(G)} \mu_v x_v + \sum_{e=\{v,w\}\in E(G)} J_e x_v x_w\Bigg) \quad \text{and} \quad Z \triangleq \sum_{X\in\{\pm 1\}^N} \mathcal{W}(X). \tag{2}$$

Here, $\mu = (\mu_v, v \in V(G))$ is a vector of *(magnetic) fields*, $J = (J_e, e \in E(G))$ is a vector of the *(pairwise) spin interactions*, and the normalization constant Z, which is defined as a sum over 2^N spin configurations, is referred to as the *partition function*. Given the model specification $\mathcal{I} = \langle G, \mu, J\rangle$, we address the tasks of finding the exact value of Z (inference) and drawing exact samples with the probability (1).

Related work. It has been known since the seminal contributions of Fisher (Fisher 1966) and Kasteleyn (Kasteleyn 1963) that computation of the partition function in the zero-field ($\mu = 0$) Ising model over a planar graph and sampling from the respective probability distribution are both tractable, that is, these are tasks of complexity polynomial in N. As shown by Barahona (Barahona 1982), even when G is planar or when $\mu = 0$ (*zero field*), the positive results are hard to generalize—both addition of the non-zero (magnetic) field and the extension beyond planar graphs make the computation of the partition function NP-hard. These results are also consistent with the statement from Jerrum and Sinclair (Jerrum and Sinclair 1993) that computation of the partition function of the zero-field Ising model is a #P-complete problem, even in the ferromagnetic case when all components of J are positive. Therefore, describing $\langle G, \mu, J\rangle$ families for which computations of the partition function and sampling are tractable remains an open question.

The simplest tractable (i.e., inference and sampling are polynomial in N) example is one when G is a tree, and the corresponding inference algorithm, known as *dynamic programming* and/or *belief propagation*, has a long history in physics (Bethe 1935; Peierls 1936), optimal control (Bellman 1952), information theory (Gallager 1963), and artificial intelligence (Pearl 1982). Extension to the case when G is a tree of $(t+1)$-sized cliques "glued" together, or more formally when G is of a *treewidth* t, is known as the *junction tree algorithm* (Verner Jensen, Olesen, and Andersen 1990), which has complexity of counting and sampling that grow exponentially with t.

Another insight originates from the foundational statistical physics literature of the last century related to a zero-field version of (1) when G is planar. Onsager (Onsager 1944) found a closed-form solution of (1) in the case of a homogeneous Ising model over an infinite two-dimensional square grid. Kac and Ward (Kac and Ward 1952) reduced the inference of (1) over a finite square lattice to computing a determinant. Kasteleyn (Kasteleyn 1963) generalized this result to an arbitrary planar graph. Kasteleyn's approach consists of expanding each vertex of G into a gadget and reducing the Ising model inference to the problem of counting perfect matchings over the expanded graph. Kasteleyn's construction was simplified by Fisher in (Fisher 1966). The tightest running time estimate for Kasteleyn's method gives $O(N^{\frac{3}{2}})$. Kasteleyn conjectured, which was later proven in (Gallucio and Loebl 1999), that the approach extends to the case of the zero-field Ising model over graphs embedded in a surface of *genus* g with a multiplicative $O(4^g)$ penalty. A slightly different reduction to perfect matching counting (Barahona 1982; Bieche, Uhry, Maynard, and Rammal 1980; Schraudolph and Kamenetsky 2009) also allows one to implement $O(N^{\frac{3}{2}})$ sampling of planar zero-field Ising models using Wilson's algorithm (Likhosherstov, Maximov, and Chertkov 2019; Wilson 1997). A K_{33} (Figure 1(a)) minor-free extension of planar zero-field inference and sampling was constructed in (Likhosherstov, Maximov, and Chertkov 2019).

An upper-bound approximation to a general class of inference problems can be built by utilizing the family of tractable spanning Ising submodels—either trees (Wainwright, Jaakkola, and Willsky 2005) or planar topologies (Globerson and Jaakkola 2007).

Contribution. In this manuscript, we first describe a new family of zero-field Ising models on graphs that are more general than planar. Given a tree decomposition of such graphs into planar and "small" ($O(1)$-sized) components "glued" together along sets of at most three vertices, inference and sampling over the new family of models is of polynomial time. We further show that all the K_5-minor-free graphs are included in this family and, moreover, their aforementioned tree decomposition can be constructed with $O(N)$ efforts. (See Figure 1(a) for an illustration.) This allows us to prove an $O(N^{\frac{3}{2}})$ upper bound on run time complexity of inference and sampling of the K_5-free zero-field Ising models. Planar graphs are included in the set of K_5-free graphs, which are neither genus- nor treewidth-bounded in general.

Second, we show how the newly introduced tractable family of zero-field Ising models allows extension of the approach of (Globerson and Jaakkola 2007) to an upper-bound log-partition function of arbitrary Ising models. Instead of using planar spanning subgraphs as in (Globerson and Jaakkola 2007), we utilize more general (nonplanar) basic tractable elements. Using the methodology of (Globerson and Jaakkola 2007), we illustrate the approach through experiments with a nonzero-field Ising model on a square grid for which inference is NP-hard (Barahona 1982).

Relation to other algorithms. The result presented in this manuscript is similar to the approach used to count perfect matchings in K_5-free graphs (Curticapean 2014; Straub, Thierauf, and Wagner 2014). However, we do not use a transition to perfect matching counting as it is typically done in studies of zero-field Ising models over planar graphs (Fisher 1966; Kasteleyn 1963; Thomas and Middleton 2009). Presumably, a direct transition to perfect matching counting can be done via a construction of an expanded graph in the fashion of (Fisher 1966; Kasteleyn 1963). However, this results in a size increase and, what's more important, there is no direct correspondence between spin configurations and perfect matchings, so sampling is not supported.

Our approach can also be viewed as extending results reported in (Likhosherstov, Maximov, and Chertkov 2019) on the inference and sampling in the K_{33}-free zero-field Ising models. In (Likhosherstov, Maximov, and Chertkov 2019), K_{33}-free graphs are decomposed into planar and K_5 components along pairs of vertices, and the whole construction relies on the underlying planar perfect matching model. In this manuscript, we reformulate the K_{33}-free construction of (Likhosherstov, Maximov, and Chertkov 2019) directly in terms of the Ising model bypassing mapping to perfect matchings. Moreover, an extension of the decomposition to gluing over triplets of vertices generalizes the construction, in particular, yielding novel results for efficient inference and learning for the zero-field Ising models over K_5-free graphs.

Structure. Section 2 formally introduces the concept of the so-called c-nice decomposition of graphs and formulates and proves tractability of c-nice decomposable zero-field Ising models. Section 2.1 introduces basic notations used later in the manuscript. Section 2.2 describes a useful technical instrument, called conditioning, which is then used in Section 2.3 and Section 2.4 to describe algorithms for efficient inference and learning, respectively, of the zero-field Ising models over graphs, which allows for a c-nice decomposition, where c is a positive integer. Section 3 describes an application of the algorithm to the example of the K_5-free zero-field Ising models. Section 4 presents an empirical application of the newly introduced family of tractable models to an upper-bounding log-partition function of a broader family of intractable graphical models (planar nonzero-field Ising models). Section 5 is reserved for conclusions.

2 Algorithm

We commence by introducing the concept of c-nice decomposition of a graph and stating the main result on the tractability of the new family of Ising models in subsection 2.1. We introduce a helpful "conditioning" machinery in subsection 2.2 and then describe the efficient inference (subsection 2.3) and sampling (subsection 2.4) algorithms which constructively prove the statement.

2.1 Decomposition tree and the main result

Throughout the text, we use common graph-theoretic notations and definitions (Diestel 2006) and also restate the most important concepts briefly. We mainly follow (Curticapean 2014; Reed and Li 2008) in the definition of the decomposition tree and its properties sufficient for our goals. Again, we point out that we only consider graphs without loops or multiple edges.

The graph is *planar* when it can be drawn on a plane without edge intersections. Graph G' is a *subgraph* of G whenever $V(G') \subseteq V(G)$ and $E(G') \subseteq E(G)$. For two subgraphs G' and G'' of G, let $G' \cup G'' = (V(G') \cup V(G''), E(G') \cup E(G''))$ (graph *union*).

Consider a tree decomposition $\mathcal{T} = \langle T, \mathcal{G} \rangle$ of a graph G into a set of subgraphs $\mathcal{G} \triangleq \{G_t\}$ of G, where t are *nodes* of a tree T, that is, $t \in V(T)$. One of the nodes of the tree, $r \in V(T)$, is selected as the root. For each node $t \in V(T)$, its *parent* is the first node on the unique path from t to r. $G_{\leq t}$ denotes the graph union of $G_{t'}$ for all the nodes t' in $V(T)$ that are t or its descendants. $G_{\not\leq t}$ denotes the graph union of $G_{t'}$ for all the nodes t' in $V(T)$ that are neither t nor descendants of t. For two neighboring nodes of the tree, $t, p \in V(T)$ and $\{t, p\} \in E(T)$, the set of overlapping vertices of G_t and G_p, $K \triangleq V(G_t) \cap V(G_p)$, is called an *attachment set* of t or p. If p is a parent of t, then K is a *navel* of t. We assume that the navel of the root is empty.

$\mathcal{T}$ is a *c-nice decomposition* of G if the following requirements are satisfied:

1. $\forall t \in V(T)$ with a navel K, it holds that $K = V(G_{\leq t}) \cap V(G_{\not\leq t})$.
2. Every attachment set K is of size 0, 1, 2, or 3.
3. $\forall t \in V(T)$, either $|V(G_t)| \leq c$ or G_t is planar.
4. If $t \in V(T)$ is such that $|V(G_t)| > c$, addition of all edges of type $e = \{v, w\}$, where v, w belong to the same attachment set of t (if e is not yet in $E(G_t)$) does not destroy planarity of G_t.

Stating it informally, the c-nice decomposition of G is a tree decomposition of G into planar and "small" (of size at most c) subgraphs G_t, "glued" via subsets of at most three vertices of G. Figure 1(a) shows an example of a c-nice decomposition with $c = 8$. There are various similar ways to define a graph decomposition in literature, and the one presented above is customized to include only properties significant for our consecutive analysis. The remainder of this section is devoted to a constructive proof of the following statement.

Theorem 1. *Let $\mathcal{I} = \langle G, 0, J \rangle$ be any zero-field Ising model where there exists a c-nice decomposition $\mathcal{T}$ of G, where c is an absolute constant. Then, there is an algorithm which, given $\mathcal{I}, \mathcal{T}$ as an input, does two things: (1) finds Z and (2) samples a configuration from $\mathcal{I}$ in time $O(\sum_{t \in V(T)} |V(G_t)|^{\frac{3}{2}})$.*

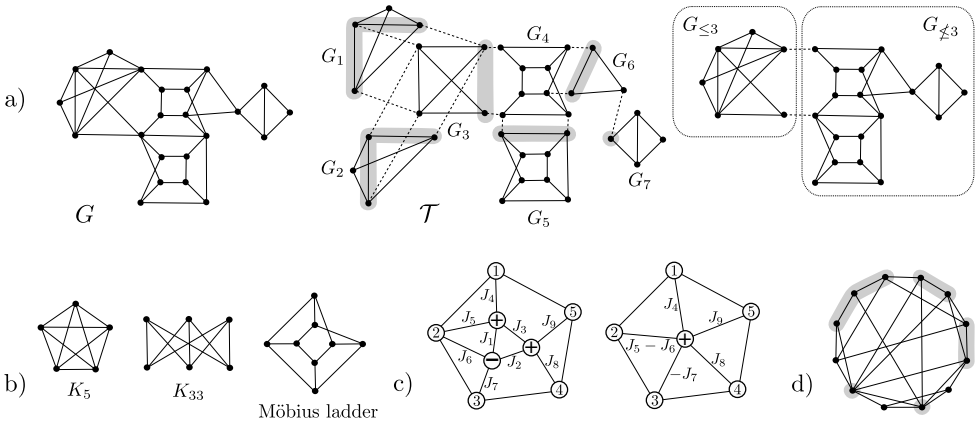


Figure 1: a) An exemplary graph G and its 8-nice decomposition $\mathcal{T}$, where $t \in \{1, \cdots, 7\}$ labels nodes of the decomposition tree T and node 4 is chosen as the root ($r = 4$). Identical vertices of G in its subgraphs G_t are shown connected by dashed lines. Navels of size 1, 2, and 3 are highlighted. Component G_5 is nonplanar, and G_4 becomes nonplanar when all attachment edges are added (according to the fourth item of the definition of the c-nice decomposition). $G_{\leq 3}$ and $G_{\not\leq 3}$ are shown with dotted lines. Note that the decomposition is non-unique for the graph. For instance, edges that belong to the attachment set can go to either of the two subgraphs containing this set or even repeat in both. b) Minors K_5 and K_{33} are forbidden in the planar graphs. Möbius ladder and its subgraphs are the only nonplanar graphs allowed in the 8-nice decomposition of a K_5-free graph. c) The left panel is an example of conditioning on three vertices/spins in the center of a graph. The right panel shows a modified graph where the three vertices (from the left panel) are reduced to one vertex, then leading to a modification of the pairwise interactions within the associated zero-field Ising model over the reduced graph. d) Example of a graph that contains K_5 as a minor: by contracting the highlighted groups of vertices and deleting the remaining vertices, one arrives at the K_5 graph.

2.2 Inference and sampling conditioned on 1, 2, or 3 vertices/spins

Before presenting the algorithm that proves Theorem 1 constructively, let us introduce auxiliary machinery of "conditioning", which describes the partition function of a zero-field Ising model over a planar graph conditioned on 1, 2, or 3 spins. Consider a zero-field Ising model $\mathcal{I} = \langle G, 0, J\rangle$ defined over a planar graph G. Recall the following result, rigorously proven in (Likhosherstov, Maximov, and Chertkov 2019, Corollary under Theorem 1), which we intend to use in the aforementioned tree decomposition as a black box.

Theorem 2. *Given $\mathcal{I} = \langle G, 0, J\rangle$, where G is planar, Z can be found in time $O(N^{\frac{3}{2}})$. Drawing a sample from $\mathcal{I}$ is a task of $O(N^{\frac{3}{2}})$ complexity.*

Let us now introduce the notion of *conditioning*. Consider a spin configuration $X \in \{\pm 1\}^N$, a subset $V' = \{v^{(1)}, \ldots, v^{(\omega)}\} \subseteq V(G)$, and define a *condition* $S = \{x_{v^{(1)}} = s^{(1)}, \ldots, x_{v^{(\omega)}} = s^{(\omega)}\}$ *on* V', where $s^{(1)}, \ldots, s^{(\omega)} = \pm 1$ are fixed values. Conditional versions of the probability distribution (1–2) and the *conditional* partition function become

$$\mathbb{P}(X|S) \triangleq \frac{\mathcal{W}(X) \times \mathbb{1}(X|S)}{Z_{|S}}, \quad \mathbb{1}(X|S) \triangleq \begin{cases} 1, & x_{v^{(1)}} = s^{(1)}, \ldots, x_{v^{(\omega)}} = s^{(\omega)} \\ 0, & \text{otherwise} \end{cases}, \tag{3}$$

$$\text{where } Z_{|S} \triangleq \sum_{X \in \{\pm 1\}^N} \mathcal{W}(X) \times \mathbb{1}(X|S). \tag{4}$$

Notice that when $\omega = 0$, $S = \{\}$ and (3–4) is reduced to (1–2). The subset of $V(G)$ is *connected* whenever the subgraph, induced by this subset is connected. Theorem 2 can be extended as follows (a formal proof can be found in the supplementary materials).

Lemma 1. *Given $\mathcal{I} = \langle G, 0, J\rangle$ where G is planar and a condition S on a connected subset $V' \subseteq V(G)$, $|V'| \leq 3$, computing the conditional partition function $Z_{|S}$ and sampling from $\mathbb{P}(X|S)$ are tasks of $O(N^{\frac{3}{2}})$ complexity.*

We omit here the tedious proof of the Lemma, also mentioning that the conditioning algorithm proving the Lemma takes the subset of connected vertices and "collapses" them into a single vertex. The graph remains planar and the task is reduced to conditioning on one vertex, which is an elementary operation given in Theorem 2. (See Figure 1(c) for an illustration.)

2.3 Inference algorithm

This subsection constructively proves the inference part of Theorem 1. For each $t \in V(T)$, let $\mathcal{I}_{\leq t} \triangleq \langle G_{\leq t}, 0, \{J_e \,|\, e \in E(G_{\leq t}) \subseteq E(G)\}\rangle$ denote a zero-field Ising *submodel induced* by $G_{\leq t}$. Denote the partition function and subvector of X related to $\mathcal{I}_{\leq t}$ as $Z^{\leq t}$ and $X_{\leq t} \triangleq \{x_v | v \in V(G_{\leq t})\}$, respectively.

Further, let K be t's navel and let $S = \{\forall v \in K : x_v = s^{(v)}\}$ denote some condition on K. Recall that $|K| \leq 3$. For each t, the algorithm computes conditional partition functions $Z^{\leq t}_{|S}$ for all choices of condition spin values $\{s^{(v)} = \pm 1\}$. Each t is processed only when its children have already been processed, so the algorithm starts at the leaf and ends at the root. If $r \in G(T)$ is a root, its navel is empty and $G_{\leq r} = G$, hence $Z = Z^{\leq r}_{|\{\}}$ is computed after r's processing.

Suppose all children of t, $c_1, ..., c_m \in V(T)$ with navels $K_1, ..., K_m \subseteq V(G_t)$ have already been processed, and now t itself is considered. Denote a spin configuration on G_t as $Y_t \triangleq \{y_v = \pm 1 \,|\, v \in V(G_t)\}$. $\mathcal{I}_{\leq c_1}, ..., \mathcal{I}_{\leq c_m}$ are $\mathcal{I}_{\leq t}$'s submodels induced by $G_{\leq c_1}, ..., G_{\leq c_m}$, which can only intersect at their navels in G_t. Based on this, one states the following dynamic programming relation:

$$Z^{\leq t}_{|S} = \sum_{Y_t \in \{\pm 1\}^{|V(G_t)|}} \mathbb{1}(Y_t|S) \exp\left(\sum_{e=\{v,w\} \in E(G_t)} J_e y_v y_w\right) \cdot \prod_{i=1}^{m} Z^{\leq c_i}_{|S_i[Y_t]}. \tag{5}$$

Here, $S_i[Y_t]$ denotes a condition $\{\forall v \in K_i : x_v = y_v\}$ on K_i. The goal is to efficiently perform summation in (5). Let $I^{(0)}, I^{(1)}, I^{(2)}, I^{(3)}$ be a partition of $\{1, ..., m\}$ by navel sizes. Figure 2(a,b) illustrates inference in t.

1. **Navels of size 0, 1.** Notice that if $i \in I^{(0)}$, then $Z^{\leq c_i}_{|\{\}} = Z^{\leq c_i}$ is constant, which was computed before. The same is true for $i \in I^{(1)}$ and $Z^{\leq c_i}_{|S^{(i)}[Y_t]} = \frac{1}{2} Z^{\leq c_i}$.

2. **Navels of size 2.** Let $i \in I^{(2)}$ denote $K_i = \{u^i, q^i\}$ and simplify notation $Z^{\leq c_i}_{y_1,y_2} \triangleq Z^{\leq c_i}_{x_{u^i}=y_1, x_{q^i}=y_2}$ for convenience. Notice that $Z^{\leq c_i}_{|S_i[Y_t]}$ is strictly positive, and due to the zero-field nature of $\mathcal{I}_{\leq c_i}$, one finds $Z^{\leq c_i}_{|+1,+1} = Z^{\leq c_i}_{|-1,-1}$ and $Z^{\leq c_i}_{|+1,-1} = Z^{\leq c_i}_{|-1,+1}$. Then, one arrives at $\log Z^{\leq c_i}_{|S_i[Y_t]} = A_i + B_i y_{u^i} y_{q^i}$, where $A_i \triangleq \log Z^{\leq c_i}_{|+1,+1} + \log Z^{\leq c_i}_{|+1,-1}$ and $B_i \triangleq \log Z^{\leq c_i}_{|+1,+1} - \log Z^{\leq c_i}_{|+1,-1}$.

3. **Navels of size 3.** Let $i \in I^{(3)}$, and as above, denote $K_i = \{u^i, q^i, h^i\}$ and $Z^{\leq c_i}_{y_1,y_2,y_3} \triangleq Z^{\leq c_i}_{x_{u^i}=y_1, x_{q^i}=y_2, x_{h^i}=y_3}$.

Due to the zero-field nature of $\mathcal{I}_{\leq c_i}$, it holds that $Z^{\leq c_i}_{|+1,y_2,y_3} = Z^{\leq c_i}_{|-1,y_2,y_3}$. Observe that there are such A_i, B_i, C_i, D_i that $\log Z^{\leq c_i}_{|y_1,y_2,y_3} = A_i + B_i y_1 y_2 + C_i y_1 y_3 + D_i y_2 y_3$ for all $y_1, y_2, y_3 = \pm 1$, which is guaranteed since the following system of equations has a solution:

$$\begin{bmatrix} \log Z^{\leq c_i}_{|+1,+1,+1} \\ \log Z^{\leq c_i}_{|+1,+1,-1} \\ \log Z^{\leq c_i}_{|+1,-1,+1} \\ \log Z^{\leq c_i}_{|+1,-1,-1} \end{bmatrix} = \begin{bmatrix} +1 & +1 & +1 & +1 \\ +1 & +1 & -1 & -1 \\ +1 & -1 & +1 & -1 \\ +1 & -1 & -1 & +1 \end{bmatrix} \times \begin{bmatrix} A_i \\ B_i \\ C_i \\ D_i \end{bmatrix}. \tag{6}$$

Considering three cases, one rewrites Eq. (5) as

$$Z^{\leq t}_{|S} = M \cdot \sum_{Y_t} \mathbb{1}(Y_t|S) \exp\Bigg(\sum_{e=\{v,w\} \in E(G_t)} J_e y_v y_w + \sum_{i \in I^{(2)} \cup I^{(3)}} B_i y_{u^i} y_{q^i} + \sum_{i \in I^{(3)}} (C_i y_{u^i} y_{h^i} + D_i y_{q^i} y_{h^i})\Bigg), \tag{7}$$

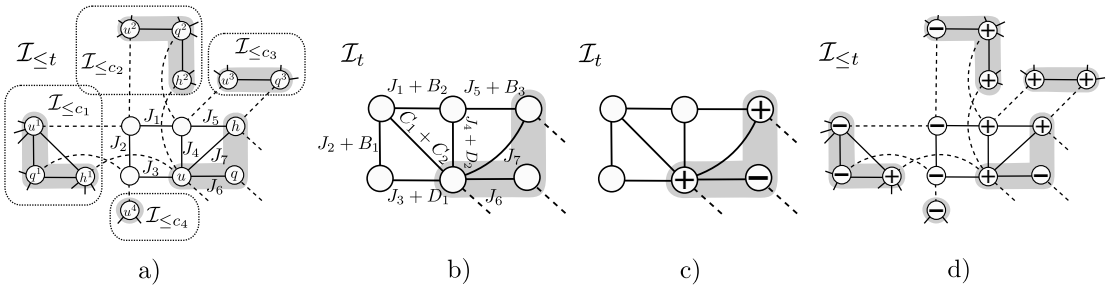


Figure 2: a) Example of inference at node t with children c_1, c_2, c_3, c_4. Navels $K_1 = \{u^1, q^1, h^1\}, K_2 = \{u^2, q^2, h^2\}, K_3 = \{u^2, q^2\}, K_4 = \{u^4\}$, and $K = \{u, q, h\}$ are highlighted. Fragments of $\mathcal{I}_{\leq c_i}$ are shown with dotted lines. Here, $I^{(0)} = \varnothing, I^{(1)} = \{4\}, I^{(2)} = \{3\}$, and $I^{(3)} = \{1, 2\}$, indicating that one child is glued over one node, one child is glued over two nodes, and two children are glued over three nodes. b) "Aggregated" Ising model $\mathcal{I}_t$ and its pairwise interactions are shown. Both c) and d) illustrate sampling over $\mathcal{I}_t$. One sample spins in $\mathcal{I}_t$ conditioned on $S^{(t)}$ and then repeats the procedure at the child nodes.

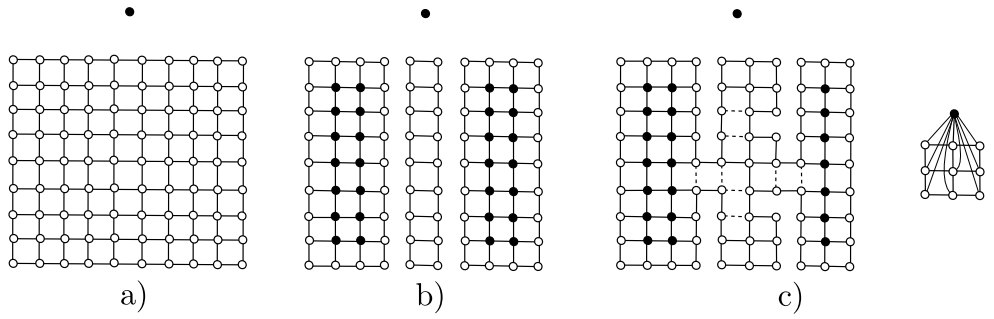


Figure 3: Construction of graphs used for approximate inference on a rectangular lattice. For better visualization, vertices connected to an apex are colored white. a) G' graph. b) One of planar $G^{(r)}$ graphs used in (Globerson and Jaakkola 2007). Such "separator" pattern is repeated for each column and row, resulting in $2(H-1)$ graphs in $\{G^{(r)}\}$. In addition, (Globerson and Jaakkola 2007) adds an *independent variables* graph where only apex edges are drawn. c) A modified "separator" pattern we propose. Again, the pattern is repeated horizontally and vertically resulting in $2(H-2)$ graphs + independent variables graph. This pattern covers more magnetic fields and connects separated parts. Dashed edges indicate the structure of 10-nice decomposition used for inference. (Nonplanar node of size 10 is illustrated on the right.)

where $M \triangleq 2^{-|I^{(1)}|} \cdot \left(\prod_{i \in I^{(0)} \cup I^{(1)}} Z^{\leq c_i}\right) \cdot \exp(\sum_{i \in I^{(2)} \cup I^{(3)}} A_i)$. The sum in Eq. (7) is simply a conditional partition function of a zero-field Ising model $\mathcal{I}_t$ defined over a graph G_t with pairwise interactions of $\mathcal{I}$ adjusted by the addition of B_i, C_i, and D_i summands at the appropriate navel edges (if a corresponding edge is not present in G_t, it has to be added). If $|V(G_t)| \leq c$, then (7) is computed a maximum of four times (depending on navel size) by brute force ($O(1)$ time). Otherwise, if K is a disconnected set in G_t, we add zero-interaction edges inside it to make it connected. Possible addition of edges inside $K, K_1, \ldots, K_m$ doesn't destroy planarity according to the fourth item in the definition of the c-nice decomposition above. Finally, we compute (7) using Lemma 1 in time $O(|V(G_t)|^{\frac{3}{2}})$.

The inference part of Theorem 1 follows directly from the procedure just described.

2.4 Sampling algorithm

Next, we address the sampling part of Theorem 1. We extend the algorithm from section 2.3 so that it supports efficient sampling from $\mathcal{I}$. Assume that the inference pass through T (from leaves to root) has been done so that $\mathcal{I}_t$ for all $t \in V(T)$ are computed. Denote $X_t \triangleq \{x_v \,|\, v \in V(G_t)\}$. The sampling algorithm runs backwards, first drawing spin values X_r at the root r of T from the

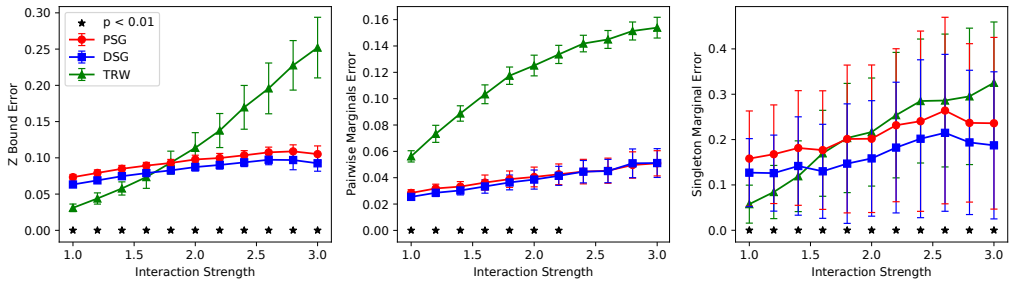


Figure 4: Comparison of tree-reweighted approximation (TRW), planar spanning graph (PSG), and decomposition-based spanning graph (DSG) approaches. The first plot is for normalized log-partition error, the second is for error in pairwise marginals, and the third is for error in singleton central marginal. Standard errors over 100 trials are shown as error bars. An asterisk "*" indicates the statistically significant improvement of DSG over PSG, with a p-value smaller than 0.01 according to the Wilcoxon test with the Bonferroni correction (Wilcoxon 1945).

marginal distribution $\mathbb{P}(X_r)$, and then processing each node t of T after its parent p is processed. Processing consists of drawing spins X_t from $\mathbb{P}(X_t \mid X_p) = \mathbb{P}(X_t \mid X^{(t)} \triangleq \{x_v \mid v \in K\})$, where K is a navel of t. This marginal-conditional scheme generates the correct sample X of spins over G.

Let $\mathbb{P}_{\leq t}(X_{\leq t})$ define a spin distribution of $\mathcal{I}_{\leq t}$. Because the Ising model is an example of Markov Random Field, it holds that $\mathbb{P}_{\leq t}(X_{\leq t} \mid X^{(t)}) = \mathbb{P}(X_{\leq t} \mid X^{(t)})$. We further derive

$$\begin{aligned}
\mathbb{P}(X_t \mid X^{(t)}) = \mathbb{P}_{\leq t}(X_t \mid X^{(t)}) &= \frac{1}{Z^{\leq t}} \sum_{X_{\leq t} \setminus X_t} \exp\left(\sum_{e=\{v,w\} \in E(G_{\leq t})} J_e x_v x_w \right) \\
&= \frac{1}{Z^{\leq t}} \cdot \exp\left(\sum_{e=\{v,w\} \in E(G_t)} J_e x_v x_w \right) \cdot \prod_{i=1}^{m} Z^{\leq c_i}_{|S_i[X_t]} \\
&\propto \exp\left(\sum_{e=\{v,w\} \in E(G_t)} J_e x_v x_w + \sum_{i \in I^{(2)} \cup I^{(3)}} B_i x_{u^i} x_{q^i} + \sum_{i \in I^{(3)}} (C_i x_{u^i} x_{h^i} + D_i x_{q^i} x_{h^i}) \right). \qquad (8)
\end{aligned}$$

In other words, sampling from $\mathbb{P}(X_t \mid X^{(t)})$ is reduced to sampling from $\mathcal{I}_t$ conditional on spins $X^{(t)}$ in the navel K. It is done via brute force if $|V(G_t)| \leq c$; otherwise, Lemma 1 allows one to draw X_t in $O(|V(G_t)|^{\frac{3}{2}})$, since $|K| \leq 3$. Sampling efforts cost as much as inference, which concludes the proof of Theorem 1. Figure 2(c,d) illustrates sampling in t.

3 Application: K_5-free zero-field Ising models

Contraction is an operation of removing two adjacent vertices v and u (and all edges incident to them) from the graph and adding a new vertex w adjacent to all neighbors of v and u. For two graphs G and H, H is G's *minor*, if it is isomorphic to a graph obtained from G's subgraph by a series of contractions. G is *H-free*, if H is not G's minor. We especially focus on the case when $H = K_5$ (Figure 1(d)). Planar graphs are a special type of K_5-free graphs, according to Wagner's theorem (Diestel 2006, Chapter 4.4). Moreover, some nonplanar graphs are K_5-free, for example, K_{33} (Figure 1(b)). K_5-free graphs are neither genus-bounded (a disconnected set of g K_{33} graphs is K_5-free and has a genus of g (Battle, Harary, and Kodama 1962)) and is not treewidth-bounded (planar square grid of size $t \times t$ is K_5-free and has a treewidth of t (Bodlaender 1998)).

Theorem 3. *Let G be a K_5-free graph of size N with no loops or multiple edges. Then, the 8-nice decomposition $\mathcal{T}$ of G exists and can be computed in time $O(N)$.*

Proof (Sketch). An equivalent decomposition is constructed in (Reed and Li 2008) in time $O(N)$. We put a formal proof in the Supplementary materials. □

Remark. *The $O(N)$ construction time of $\mathcal{T}$ guarantees that $\sum_{t \in V(T)} |V(G_t)| = O(N)$. All nonplanar components in $\mathcal{T}$ are isomorphic to the Möbius ladder (Figure 1(b)) or its subgraph.*

The graph in Figure 1(a) is actually K_5-free. Theorems 1 and 3 allow us to conclude:

Theorem 4. *Given $\mathcal{I} = \langle G, 0, J\rangle$ with K_5-free G of size N, finding Z and sampling from $\mathcal{I}$ take $O(N^{\frac{3}{2}})$ total time.*

Proof. Finding 8-nice $\mathcal{T}$ for G takes $O(N)$ time (Theorem 3). Provided with $\mathcal{T}$, the complexity is

$$O\left(\sum_{t\in V(T)} |V(G_t)|^{\frac{3}{2}}\right) = O\left((\sum_{t\in V(T)} |V(G_t)|)^{\frac{3}{2}}\right) = O(N^{\frac{3}{2}}),$$

where we apply convexity of $f(z) = z^{\frac{3}{2}}$ for $z \geq 0$ and the Remark after Theorem 3. □

4 Application: approximate inference of square-grid Ising model

In this section, we consider $\mathcal{I} = \langle G, \mu, J\rangle$ such that G is a square-grid graph of size $H \times H$. Finding $Z(G, \mu, J)$ for arbitrary μ, J is an NP-hard problem (Barahona 1982) in such a setting. Construct G' by adding an *apex* vertex connected to all G's vertices by edge (Figure 3(a)). Now it can easily be seen that $Z(G, \mu, J) = \frac{1}{2}Z(G', 0, J' = (J_\mu \cup J))$, where $J_\mu = \mu$ are interactions assigned for apex edges.

Let $\{G^{(r)}\}$ be a family of spanning graphs ($V(G^{(r)}) = V(G')$, $E(G^{(r)}) \subseteq E(G')$) and $J^{(r)}$ be interaction values on $G^{(r)}$. Also, denote $\hat{J}^{(r)} = J^{(r)} \cup \{0, e \in E(G') \setminus E(G^{(r)})\}$. Assuming that $\log Z(G^{(r)}, 0, J^{(r)})$ are tractable, the convexity of $\log Z(G', 0, J')$ allows one to write the following upper bound:

$$\log Z(G', 0, J') \leq \min_{\substack{\rho(r)\geq 0, \sum_r \rho(r)=1 \\ \{J^{(r)}\}, \sum_r \rho(r)\hat{J}^{(r)}=J'}} \sum_r \rho(r) \log Z(G^{(r)}, 0, J^{(r)}). \tag{9}$$

After graph set $\{G^{(r)}\}$ has been fixed, one can numerically optimize the right-hand side of (9), as shown in (Globerson and Jaakkola 2007) for planar $G^{(r)}$. The extension of the basic planar case is straightforward and can be found in the Supplementary materials for convenience. We also put the description of marginal probabilities approximation suggested in (Globerson and Jaakkola 2007; Wainwright, Jaakkola, and Willsky 2005).

The choice for a planar spanning graph (PSG) family $\{G^{(r)}\}$ of (Globerson and Jaakkola 2007) is illustrated in Figure 3(b). A tractable decomposition-based extension of the planar case presented in this manuscript suggests a more advanced construction—decomposition-based spanning graphs (DSG) (Figure 3(c)). We compare performance of both PSG and DSG approaches as well as the performance of tree-reweighted approximation (TRW) (Wainwright, Jaakkola, and Willsky 2005) in the following setting of *Varying Interaction*: $\mu \sim \mathcal{U}(-0.5, 0.5)$, $J \sim \mathcal{U}(-\alpha, \alpha)$, where $\alpha \in \{1, 1.2, 1.4, \ldots, 3\}$. We opt optimize for grid size $H = 15$ (225 vertices, 420 edges) and compare upper bounds and marginal probability approximations (superscript *alg*) with exact values obtained using a junction tree algorithm (Verner Jensen, Olesen, and Andersen 1990) (superscript *true*). We compute three types of error:

1. normalized log-partition error $\frac{1}{H^2}(\log Z^{alg} - \log Z^{true})$,
2. error in pairwise marginals $\frac{1}{|E(G)|}\sum_{e=\{v,w\}\in E(G)} |\mathbb{P}^{alg}(x_v x_w = 1) - \mathbb{P}^{true}(x_v x_w = 1)|$, and
3. error in singleton central marginal $|\mathbb{P}^{alg}(x_v = 1) - \mathbb{P}^{true}(x_v = 1)|$ where v is a vertex of G with coordinates $(8, 8)$.

We average results over 100 trials (see Fig. 4).[1][2] We use the same quasi-Newton algorithm (Bertsekas 1999) and parameters when optimizing (9) for PSG and DSG, but for most settings, DSG outperforms PSG and TRW. Cases with smaller TRW error can be explained by the fact that TRW implicitly optimizes (9) over the family of *all* spanning trees which can be exponentially big in size, while for PSG and DSG we only use $O(H)$ spanning graphs.

Because PSG and DSG approaches come close to each other, we additionally test for each value of α on each plot, whether the difference $err_{PSG} - err_{DSG}$ is bigger than zero. We apply a one-sided Wilcoxon's test (Wilcoxon 1945) together with the Bonferroni correction because we test 33 times (Jean Dunn 1961). In most settings, the improvement is statistically significant (Figure 4).

[1] Hardware used: 24-core Intel® Xeon® Gold 6136 CPU @ 3.00 GHz

[2] Implementation of the algorithms is available at `https://github.com/ValeryTyumen/planar_ising`

5 Conclusion

In this manuscript, we introduce a new family of zero-field Ising models composed of planar components and graphs of $O(1)$ size. For these models, we describe a polynomial algorithm for exact inference and sampling provided that the decomposition tree is also in the input. A theoretical application is $O(N^{\frac{3}{2}})$ inference and sampling algorithm for K_5-free zero-field Ising models—a superset of planar zero-field models that is neither treewidth- nor genus-bounded. A practical application is an improvement of an approximate inference scheme for arbitrary topologies based on planar spanning graphs (Globerson and Jaakkola 2007) but using tractable spanning decomposition-based graphs instead of planar. We leave the algorithm as it is but substitute planar graphs with a family of spanning decomposition-based graphs that are tractable. This alone gives a tighter upper bound on the true partition function and a more precise approximation of marginal probabilities.

References

Barahona, F. (1982). On the computational complexity of Ising spin glass models. *Journal of Physics A: Mathematical and General 15*(10), 3241.

Battle, J., F. Harary, and Y. Kodama (1962, 11). Additivity of the genus of a graph. *Bull. Amer. Math. Soc. 68*(6), 565–568.

Bellman, R. (1952). On the theory of dynamic programming. *Proceedings of the National Academy of Sciences 38*(8), 716–719.

Bertsekas, D. (1999). *Nonlinear Programming*. Athena Scientific.

Bethe, H. (1935). Statistical theory of superlattices. *Proceedings of Royal Society of London A 150*, 552.

Bieche, L., J. P. Uhry, R. Maynard, and R. Rammal (1980). On the ground states of the frustration model of a spin glass by a matching method of graph theory. *Journal of Physics A: Mathematical and General 13*(8), 2553.

Bodlaender, H. L. (1998). A partial k-arboretum of graphs with bounded treewidth. *Theoretical Computer Science 209*(1), 1 – 45.

Curticapean, R. (2014). Counting perfect matchings in graphs that exclude a single-crossing minor. *arXiv preprint arXiv:1406.4056*.

Diestel, R. (2006). *Graph Theory*. Electronic library of mathematics. Springer.

Fisher, M. E. (1966). On the dimer solution of planar Ising models. *Journal of Mathematical Physics 7*(10), 1776–1781.

Gallager, R. (1963). *Low density parity check codes*. MIT Press, Cambridge, MA.

Gallucio, A. and M. Loebl (1999). On the theory of Pfaffian orientations. I: Perfect matchings and permanents. *The Electronic Journal of Combinatorics [electronic only] 6*(1), Research paper R6, 18 p.

Globerson, A. and T. S. Jaakkola (2007). Approximate inference using planar graph decomposition. In *Advances in Neural Information Processing Systems*, pp. 473–480.

Jean Dunn, O. (1961, 03). Multiple comparisons among means. *Journal of The American Statistical Association - J AMER STATIST ASSN 56*, 52–64.

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

Kac, M. and J. C. Ward (1952, Dec). A combinatorial solution of the two-dimensional Ising model. *Phys. Rev. 88*, 1332–1337.

Kasteleyn, P. W. (1963). Dimer statistics and phase transitions. *Journal of Mathematical Physics 4*(2), 287–293.

Likhosherstov, V., Y. Maximov, and M. Chertkov (2019, 09–15 Jun). Inference and sampling of k_{33}-free ising models. In K. Chaudhuri and R. Salakhutdinov (Eds.), *Proceedings of the 36th International Conference on Machine Learning*, Volume 97 of *Proceedings of Machine Learning Research*, Long Beach, California, USA, pp. 3963–3972. PMLR.

Onsager, L. (1944, Feb). Crystal statistics. I: A two-dimensional model with an order-disorder transition. *Phys. Rev. 65*, 117–149.

Pearl, J. (1982). Reverend bayes on inference engines: A distributed hierarchical approach. In *Proceedings of the Second AAAI Conference on Artificial Intelligence*, AAAI'82, pp. 133–136. AAAI Press.

Peierls, H. (1936). Ising's model of ferromagnetism. *Proceedings of Cambridge Philosophical Society 32*, 477–481.

Reed, B. and Z. Li (2008). Optimization and recognition for K5-minor free graphs in linear time. In E. S. Laber, C. Bornstein, L. T. Nogueira, and L. Faria (Eds.), *LATIN 2008: Theoretical Informatics*, Berlin, Heidelberg, pp. 206–215. Springer Berlin Heidelberg.

Schraudolph, N. N. and D. Kamenetsky (2009). Efficient exact inference in planar Ising models. In D. Koller, D. Schuurmans, Y. Bengio, and L. Bottou (Eds.), *Advances in Neural Information Processing Systems 21*, pp. 1417–1424. Curran Associates, Inc.

Straub, S., T. Thierauf, and F. Wagner (2014, June). Counting the number of perfect matchings in K5-free graphs. In *2014 IEEE 29th Conference on Computational Complexity (CCC)*, pp. 66–77.

Tarjan, R. (1971, Oct). Depth-first search and linear graph algorithms. In *12th Annual Symposium on Switching and Automata Theory (SWAT 1971)*, pp. 114–121.

Thomas, C. K. and A. A. Middleton (2009, Oct). Exact algorithm for sampling the two-dimensional Ising spin glass. *Phys. Rev. E 80*, 046708.

Thomason, A. (2001, March). The extremal function for complete minors. *J. Comb. Theory Ser. B 81*(2), 318–338.

Verner Jensen, F., K. Olesen, and S. Andersen (1990, 08). An algebra of Bayesian belief universes for knowledge based systems. *Networks 20*, 637 – 659.

Wainwright, M. J., T. S. Jaakkola, and A. S. Willsky (2005). A new class of upper bounds on the log partition function. *IEEE Transactions on Information Theory 51*(7), 2313–2335.

Wilcoxon, F. (1945). Individual comparisons by ranking methods. *Biometrics bulletin 1*(6), 80–83.

Wilson, D. B. (1997). Determinant algorithms for random planar structures. In *Proceedings of the Eighth Annual ACM-SIAM Symposium on Discrete Algorithms*, SODA '97, Philadelphia, PA, USA, pp. 258–267. Society for Industrial and Applied Mathematics.

Zhu, C., R. H. Byrd, P. Lu, and J. Nocedal (1997, December). Algorithm 778: L-BFGS-B: Fortran subroutines for large-scale bound-constrained optimization. *ACM Trans. Math. Softw. 23*(4), 550–560.

Appendices

Proof for Lemma 1

Lemma 1. *Given $\mathcal{I} = \langle G, 0, J\rangle$ where G is planar and a condition S on a connected subset $V' \subseteq V(G)$, $|V'| \leq 3$, computing conditional partition function $Z_{|S}$, and sampling from $\mathbb{P}(X|S)$ are tasks of $O(N^{\frac{3}{2}})$ complexity.*

Proof. We consider cases depending on ω and consequently reduce each case to a simpler one. For convenience in cases where applies we denote $u \triangleq v^{(1)}, h \triangleq v^{(2)}, q \triangleq v^{(3)}$:

1. **Conditioning on $\omega = 0$ spins.** See Theorem 2.

2. **Conditioning on $\omega = 1$ spin.** Since configurations X and $-X$ have the same probability in $\mathcal{I}$, one deduces that $Z_{|\,x_u = s^{(1)}} = \frac{1}{2}Z$.

   One also deduces that sampling X from $\mathbb{P}(X \,|\, x_u = s^{(1)})$ is reduced to 1) drawing $\overline{X} = \{\overline{x}_v = \pm 1\}$ from $\mathbb{P}(X)$ and then 2) returning $X = (s^{(1)}\overline{x}_u) \cdot \overline{X}$ as a result.

3. **Conditioning on $\omega = 2$ spins.** There is an edge $e^0 = \{u, h\} \in E(G)$. The following expansion holds:

$$\begin{aligned} Z_{|\,x_u = s^{(1)}, x_h = s^{(2)}} &= \sum_{X,\, x_u = s^{(1)},\, x_h = s^{(2)}} \exp\Big(\sum_{e=\{v,w\} \in E(G)} J_e x_v x_w \Big) \\ &= \exp(J_{e^0} s^{(1)} s^{(2)}) \cdot \sum_{X,\, x_u = s^{(1)},\, x_h = s^{(2)}} \exp\Big(\sum_{\substack{e=\{v,w\} \in E(G) \\ e \neq e^0}} J_e x_v x_w \Big) \\ &= \exp(J_{e^0} s^{(1)} s^{(2)}) \cdot \sum_{X,\, x_u = s^{(1)},\, x_h = s^{(2)}} \exp\Big(\sum_{\substack{e=\{v,w\} \in E(G) \\ e \cap e^0 = \varnothing}} J_e x_v x_w \\ &+ \sum_{\substack{e=\{u,v\} \in E(G) \\ v \neq h}} (J_e s^{(1)}) x_v \cdot 1 + \sum_{\substack{e=\{h,v\} \in E(G) \\ v \neq u}} (J_e s^{(2)}) x_v \cdot 1 \Big) \end{aligned} \tag{10}$$

   Obtain graph G' from G by contracting u, h into z. G' is still planar and has $N - 1$ vertices. Preserve pairwise interactions of edges which were not deleted after contraction. For each edge $e = \{u, v\}$, $v \neq h$ set $J_{\{z,v\}} = J_e s^{(1)}$, for each edge $e = \{h, v\}$, $v \neq u$ set $J_{\{z,v\}} = J_e s^{(2)}$. Collapse double edges in G' which were possibly created by transforming into single edges. A pairwise interaction of the result edge is set to the sum of collapsed interactions.

   Define a zero-field Ising model $\mathcal{I}'$ on the resulted graph G' with its pairwise interactions, inducing a distribution $\mathbb{P}'(X' = \{x'_v = \pm 1 | v \in V(G')\})$. Let Z' denote $\mathcal{I}'$'s partition function. A closer look at (10) reveals that

$$Z_{|\,x_u = s^{(1)}, x_h = s^{(2)}} = \exp(J_{e^0} s^{(1)} s^{(2)}) \cdot Z'_{|\,x'_z = 1} \tag{11}$$

   where $Z'_{|\,z'_y = 1}$ is a partition function conditioned on a single spin and can be found efficiently as shown above.

   Since the equality of sums (11) holds summand-wise, for a given $X'' = \{x''_v = \pm 1 \,|\, v \in V(G) \setminus \{u, h\}\}$ the probabilities $\mathbb{P}(X'' \cup \{x_u = s^{(1)}, x_h = s^{(2)}\} \,|\, x_u = s^{(1)}, x_h = s^{(2)})$ and $\mathbb{P}'(X'' \cup \{x'_z = 1\} \,|\, x'_z = 1)$ are the same. Hence, sampling from $\mathbb{P}(X \,|\, x_u = s^{(1)}, x_h = s^{(2)})$ is reduced to conditional sampling from planar zero-field Ising model $\mathbb{P}'(X' \,|\, x'_z = 1)$ of size $N - 1$.

4. **Conditioning on $w = 3$ spins.** Without loss of generality assume that u, h are connected by an edge e^0 in G. A derivation similar to (10) and (11) reveals that (preserving the notation of Case 2)

$$Z_{|\,x_u = s^{(1)}, x_h = s^{(2)}, x_q = s^{(3)}} = \exp(J_{e^0} s^{(1)} s^{(2)}) \cdot Z'_{|\,x'_z = 1, x'_q = s^{(3)}} \tag{12}$$

   which reduces inference conditional on 3 vertices to a simpler case of 2 vertices. Again, sampling from $\mathbb{P}(X \,|\, x_u = s^{(1)}, x_t = s^{(2)}, x_q = s^{(3)})$ is reduced to a more basic sampling from $\mathbb{P}'(X' \,|\, x'_z = 1, x'_q = s^{(3)})$.

□

In principle, Lemma 1 can be extended to arbitrarily large ω leaving a certain freedom for the Ising model conditioning framework. However, in this manuscript we focus on a given special case which is enough for our goals.

Proof for Theorem 3

Prior to the proof, we introduce a series of definitions used in (Reed and Li 2008). It is assumed that a graph $G = (V, E)$ (no loops and multiple edges) is given.

For any $X \subseteq V(G)$ let $G - X$ denote a graph $(V(G) \setminus X, \{e = \{v, w\} \in E(G) \,|\, v, w \notin X\})$. $X \subseteq V(G)$ is a (i, j)-*cut* whenever $|X| = i$ and $G - X$ has at least j connected components.

The graph is *biconnected* whenever it has no $(1, 2)$-cut. A *biconnected component* of the graph is a maximal biconnected subgraph. Clearly, a pair of biconnected components can intersect in at most one vertex and a graph of components' intersections is a tree when G is connected (*a tree of biconnected components*). The graph is 3-*connected* whenever it has no $(2, 2)$-cut.

A 2-*block tree* of a biconnected graph G, written $\langle T', \mathcal{G}' \rangle$, is a tree T' with a set $\mathcal{G}' = \{G'_t\}_{t \in V(T')}$ with the following properties:

- G'_t is a graph (possibly with multiple edges) for each $t \in V(T')$.
- If G is 3-connected then T' has a single node r which is colored 1 and $G'_r = G$.
- If G is not 3-connected then there exists a color 2 node $t \in V(T')$ such that
  1. G'_t is a graph with two vertices u and v and no edges for some $(2, 2)$-cut $\{u, v\}$ in G.
  2. Let $T'_1, \ldots, T'_k$ be the connected components (subtrees) of $T' - t$. Then $G - \{u, v\}$ has k connected components $U_1, \ldots, U_k$ and there is a labelling of these components such that T'_i is a 2-block tree of $G'_i = (V(U_i) \cup \{u, v\}, E(U_i) \cup \{\{u, v\}\})$.
  3. For each i, there exists exactly one color 1 node $t_i \in V(T'_i)$ such that $\{u, v\} \subseteq V(G'_{t_i})$.
  4. For each i, $\{t, t_i\} \in E(T)$.

A $(3, 3)$-*block tree* of a 3-connected graph G, written $\langle T'', \mathcal{G}'' \rangle$, is a tree T'' with a set $\mathcal{G}'' = \{G''_t\}_{t \in V(T'')}$ with the following properties:

- G''_t is a graph (possibly with multiple edges) for each $t \in V(T'')$.
- If G has no $(3, 3)$-cut then T has a single node r which is colored 1 and $G_r = G$.
- If G has a $(3, 3)$-cut then there exists a color 2 node $t \in V(T'')$ such that
  1. G''_t is a graph with vertices u, v and w and no edges for some $(3, 3)$-cut $\{u, v, w\}$ in G.
  2. Let $T''_1, \ldots, T''_k$ be the connected components (subtrees) of $T'' - t$. Then $G - \{u, v, w\}$ has k connected components $U_1, \ldots, U_k$ and there is a labelling of these components such that T_i is a $(3, 3)$-block tree of $G''_i = (V(U_i) \cup \{u, v, w\}, E(U_i) \cup \{\{u, v\}, \{v, w\}, \{u, w\}\})$.
  3. For each i, there exists exactly one color 1 node $t_i \in V(T''_i)$, such that $\{u, v, w\} \subseteq V(G''_{t_i})$.
  4. For each i, $\{t, t_i\} \in E(T'')$.

Theorem 3. *Let G be K_5-free graph of size N with no loops or multiple edges. Then the 8-nice decomposition $\mathcal{T}$ of G exists and can be computed in time $O(N)$.*

Proof. Since G is K_5-free and has no loops or multiple edges, it holds that $|E(G)| = O(N)$ (Thomason 2001). In time $O(N)$ we can find a forest of G's biconnected components (Tarjan 1971). If we find an 8-nice decomposition for each biconnected component, join them into a single 8-nice decomposition by using attachment sets of size 1 for decompositions inside G's connected component and attachment sets of size 0 for decompositions in different connected components. Hence, further we assume that G is biconnected.

The $O(N)$ algorithm of (Reed and Li 2008) finds a 2-block tree $\langle T', \mathcal{G}' \rangle$ for G and then for each color 1 node $G'_t \in \mathcal{G}'$ it finds $(3, 3)$-block tree $\langle T'', \mathcal{G}'' \rangle$ where all components are either planar or Möbius ladders. To get an 8-nice decomposition from each $(3, 3)$-block tree, 1) for each color 2 node

contract an edge between it and one of its neighbours in T'' and 2) remove all edges which were only created during $\langle T'', \mathcal{G}'' \rangle$ construction (2nd item of $(3,3)$-block tree definition).

Now we have to draw additional edges in the forest F of obtained 8-nice decompositions so that to get a single 8-nice decomposition $\mathcal{T}$ of G. Notice that for each pair of adjacent nodes $G'_t, G'_s \in \mathcal{G}'$ where G'_t is color 1 node and $G'_s = (\{u, v\}, \varnothing)$ is a color 2 node, u, v are in $V(G'_t)$ and $\{u, v\} \in E(G'_t)$. Hence, there is at least one component G''_r of 8-nice decomposition of G'_t where both u and v are present. For each pair of s and t draw an edge between s and r in F. Then 1) for each color 2 node in F (such as s) contract an edge between it and one of its neighbors (such as r) and 2) remove all edges which were created during $\langle T', \mathcal{G}' \rangle$ construction (2nd item of 2-block tree definition). This results is a correct c-nice decomposition for biconnected G. □

Upper Bound Minimization and Marginal Computation in Approximation Scheme

Denote:

$$h(J') \triangleq \min_{\rho(r)\geq 0, \sum_r \rho(r)=1} g(J', \rho), \qquad g(J', \rho) \triangleq \min_{\{J^{(r)}\}, \sum_r \rho(r)\hat{J}^{(r)}=J'} \sum_r \rho(r) \log Z(G^{(r)}, 0, J^{(r)})$$

where $h(J')$ is a tight upper bound for $\log Z(G', 0, J')$.

Given a fixed ρ, we compute $g(J', \rho)$ using L-BFGS-B optimization (Zhu, Byrd, Lu, and Nocedal 1997) by back-propagating through $Z(G^{(r)}, 0, J^{(r)})$ and projecting gradients on the constraint linear manifold. On the upper level we also apply L-BFGS-B algorithm to compute $h(J')$, which is possible since (Wainwright, Jaakkola, and Willsky 2005; Globerson and Jaakkola 2007)

$$\frac{\partial}{\partial \rho(r)} g(J', \rho) = \log Z(G^{(r)}, 0, J^{(r)}_{min}) - (M^{(r)})^\top J^{(r)}_{min}, \;\; M^{(r)} \triangleq \frac{\partial}{\partial J^{(r)}_{min}} \log Z(G^{(r)}, 0, J^{(r)}_{min})$$

where $\{J^{(r)}_{min}\}$ is argmin inside $g(J', \rho)$'s definition and $M^{(r)} = \{M^{(r)}_e \,|\, e \in E(G^{(r)})\}$ is a vector of *pairwise marginal expectations*. We reparameterize $\rho(r)$ into $\frac{w(r)}{\sum_{r'} w(r')}$ where $w(r) > 0$.

For $e = \{v, w\} \in E(G)$ we approximate pairwise marginal probabilities as (Wainwright, Jaakkola, and Willsky 2005; Globerson and Jaakkola 2007)

$$\mathbb{P}^{alg}(x_v x_w = 1) = \frac{1}{2} \cdot [\sum_r \rho(r) M^{(r)}_e] + \frac{1}{2}$$

Let e_A be an edge between central vertex v and apex in G'. We approximate singleton marginal probability at vertex v as

$$\mathbb{P}^{alg}(x_v = 1) = \frac{1}{2} \cdot [\sum_r \rho(r) M^{(r)}_{e_A}] + \frac{1}{2}$$

Future Work

We see the following straightforward extensions of the algorithm presented in the manuscript:

1. The work (Curticapean 2014) extends the polynomial scheme of (Straub, Thierauf, and Wagner 2014) for perfect matching counting in a case when G is H-free, where H is a *single-crossing* graph, i.e. a minor of an arbitrary graph that can be drawn on a plane with no more than one edge crossing. We claim without proofs that the same applies for a setting considered in this manuscript.

2. In (Straub, Thierauf, and Wagner 2014) authors also present a parallel version of their perfect matching counting scheme in K_5-free graphs and show that the problem is in TC^1 parallel complexity class. We claim without proofs that the same applies for inference of K_5-free zero-field Ising models.