*Combinatorics, Probability and Computing* (2021), **30**, pp. 905–921
doi:10.1017/S0963548321000080

CAMBRIDGE
UNIVERSITY PRESS

ARTICLE

# Polynomial-time approximation algorithms for the antiferromagnetic Ising model on line graphs

Martin Dyer[1], Marc Heinrich[1], Mark Jerrum[2,*] and Haiko Müller[1]

[1]School of Computing, University of Leeds, Leeds LS2 9JT, UK
[2]School of Mathematical Sciences, Queen Mary, University of London, Mile End Road, London E1 4NS, UK
*Corressponding author. Email: m.jerrum@qmul.ac.uk

(Received 17 June 2020; first published online 12 April 2021)

**Abstract**

We present a polynomial-time Markov chain Monte Carlo algorithm for estimating the partition function of the antiferromagnetic Ising model on any line graph. The analysis of the algorithm exploits the 'winding' technology devised by McQuillan [CoRR abs/1301.2880 (2013)] and developed by Huang, Lu and Zhang [Proc. 27th Symp. on Disc. Algorithms (SODA16), 514–527]. We show that exact computation of the partition function is #P-hard, even for line graphs, indicating that an approximation algorithm is the best that can be expected. We also show that Glauber dynamics for the Ising model is rapidly mixing on line graphs, an example being the kagome lattice.

**2020 MSC Codes:** Primary 68Q25; secondary 60J10, 68Q17, 68W20, 68W25, 82B20

## 1. Introduction

In statistical mechanics, the Ising model was proposed by Lenz in 1920 as a mathematical model of ferromagnetism. The model is defined in terms of simple local interactions between sites representing atoms. Despite the simplicity of its description, the model exhibits complex behaviours, which have been hard to unravel. Indeed several years were to pass before the most fundamental property of the Ising model was established, namely that small changes in the strength of the local interactions could lead to dramatic changes in macroscopic behaviour. The model has remained an active object of study ever since, with applications to statistical inference and image analysis helping to sustain interest [24]. An accessible account of the Ising model can be found in Friedli and Velenik's book [8, Section 3].

For $\beta, \nu \in \mathbb{R}$, the partition function of the Ising model on a (simple, undirected) graph $\Gamma$ is defined by

$$Z_{\beta,\nu}(\Gamma) = \sum_{\sigma : V(\Gamma) \to \{0,1\}} w(\sigma), \tag{1}$$

†All authors are supported by EPSRC grants EP/S016562/1 and EP/S016694/1, 'Sampling in hereditary classes'.

© The Author(s), 2021. Published by Cambridge University Press.

https://doi.org/10.1017/S0963548321000080 Published online by Cambridge University Press

where

$$w(\sigma) = \prod_{\{i,j\}\in E(\Gamma)} \exp\left(\beta\,|\sigma(i)-\sigma(j)|\right) \prod_{k\in V(\Gamma)} \exp\left(\nu\sigma(k)\right).$$

The partition function is a sum over *spin configurations* $\sigma : V(\Gamma) \to \{0,1\}$, which assign 0/1 'spins' to the vertices of $\Gamma$. The parameter $\beta$ is an 'interaction energy' that controls the strength of the interaction between spins at adjacent vertices. The parameter $\nu$ indicates the strength of the 'external field' that biases the spins at individual vertices. The partition function is the normalising factor in a probability distribution $\mathcal{D}_{\Gamma,\beta,\nu}$, the *Gibbs distribution*, that assigns weight $\mathcal{D}_{\Gamma,\beta,\nu}(\sigma) = w(\sigma)/Z_{\beta,\nu}(\Gamma)$ to spin configuration $\sigma$. More generally, the uniform interaction energy $\beta$ may be replaced by individual interaction energies $\beta_{i,j}$ for each edge $\{i,j\} \in E(\Gamma)$, and the uniform field strength $\nu$ by individual field strengths $\nu_k$ for each vertex $k \in V(\Gamma)$. For our main result, it is crucial that the interaction energies are uniform over edges, but uniformity of the field is inessential. For simplicity, we carry through the calculation for the uniform case, and return to varying fields towards the end, in Section 5.

In this article, we analyse the computational complexity of evaluating the partition function (1) when $\Gamma$ is a line graph and $\beta > 0$. We also address the related problem of sampling configurations according to the Gibbs distribution $\mathcal{D}_{\Gamma,\beta,\nu}$. The class of line graphs, which is well studied in graph theory, will be defined presently. When $\beta > 0$ the spins $\sigma(i)$ and $\sigma(j)$ associated with adjacent vertices $i$ and $j$ tend to differ; that is, configurations $\sigma : V(\Gamma) \to \{0,1\}$ with $\sigma(i) \neq \sigma(j)$ make a greater contribution to the sum (1). The qualifier *antiferromagnetic* is applied to this situation, while *ferromagnetic* describes the case $\beta < 0$.[a] When $\beta = 0$ the spins are probabilistically independent and the model is trivial.

We now briefly describe the context for our investigation. Exact evaluation of $Z_{\beta,\nu}(\Gamma)$ is generally #P-hard, and hence almost certainly computationally intractable [14, Theorem 15]. The main exception is when the graph $\Gamma$ is restricted to be planar and the external field is absent ($\nu = 0$), a situation which is covered by a classical algorithm due to Fisher [7] and Kasteleyn [16]. (Planarity of $\Gamma$ can be relaxed somewhat, to the condition that $\Gamma$ admits a Pfaffian orientation.)

Turning to approximate computation, Jerrum and Sinclair [14] present an efficient algorithm to estimate (in the sense of Fully Polynomial Randomised Approximation Scheme or FPRAS) the partition function $Z_{\beta,\nu}(\Gamma)$ for general graphs $\Gamma$ in the ferromagnetic case, even when $\nu \neq 0$. This algorithm extends to the situation of varying fields, provided that the sign of $\nu_k$ is consistent, but breaks down if the sign varies. In general, the partition function is hard to approximate in the antiferromagnetic case. The reason, intuitively, is that the ground states of the model, i.e. the configurations of greatest weight, correspond to maximum cardinality cuts in $\Gamma$. Thus, an efficient approximation algorithm for the partition function would most likely yield a polynomial-time algorithm for finding a maximum cut in a graph, which is an NP-hard problem [9]. So, to find tractable classes of instances in the antiferromagnetic situation, we need to focus on graph classes in which a maximum cardinality cut can be found in polynomial time. One obvious example is the class of bipartite graphs. Graphs in this class are easily dealt with: just invert the sign of $\beta$ and flip the interpretation of the spins on one side of the bipartition to yield an equivalent ferromagnetic Ising model. For this we have an FPRAS, as mentioned earlier.

Here, we focus on another graph class for which maximum cardinality cut is polynomial-time solvable, namely the class of line graphs. The *line graph* $L(\Gamma)$ associated with an underlying graph $\Gamma = (V(\Gamma), E(\Gamma))$ is the graph with vertex set $V(L(\Gamma)) = E(\Gamma)$ and edge set

$$E(L(\Gamma)) = \{\{e,f\} : e,f \in E(\Gamma) \text{ and } e \cap f \neq \emptyset\}.$$

[a] This notation is slightly non-standard, in that our interaction energy $\beta$ has the opposite sign to usual, but this convention becomes convenient later on.

https://doi.org/10.1017/S0963548321000080 Published online by Cambridge University Press

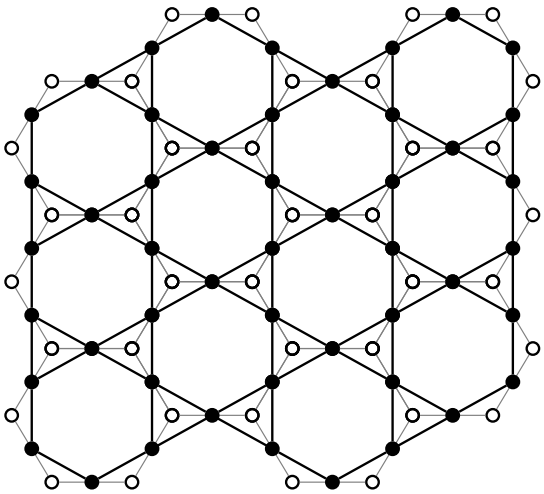

**Figure 1.** A portion of the hexagonal lattice (open vertices/thin grey edges) and its line graph, the kagome lattice (filled vertices/thick black edges).

Thus, there is an edge between two vertices in the line graph $L(\Gamma)$ whenever the corresponding edges in $\Gamma$ share an endpoint. See Figure 1 for an example graph and its associated line graph. That a maximum cardinality cut can found in polynomial time in line graphs was shown by Arbib [1]; see also Guruswami [11]. Our main contribution (Theorem 1.1) is an efficient approximation algorithm for estimating the partition function $Z_{\beta,\nu}(\Gamma)$ in the antiferromagnetic case, when $\Gamma$ is a line graph. The algorithm is efficient in the sense of Fully Polynomial Randomised Approximation Scheme (FPRAS). An FPRAS is a randomised algorithm that produces an output that is within ratio $1 \pm \epsilon$ of the desired solution, with high probability, and within time that is bounded by a polynomial in the instance size $n$ and $1/\epsilon$. See [20, Definition 11.2] for details. Along the way, we also find an efficient algorithm (Theorem 1.2) for sampling configurations in the same setting. In contrast to the case of bipartite graphs, the reason for tractability of these two problems, counting and sampling, is non-trivial. Our precise results are the following.

**Theorem 1.1.** *Suppose $\beta > 0$ (the antiferromagnetic case) and $\nu \in \mathbb{R}$. There is an FPRAS that, given a line graph $L(\Gamma)$, estimates the partition function $Z_{\beta,\nu}(L(\Gamma))$ of the Ising model* (1).

**Theorem 1.2.** *Suppose $\beta > 0$ (the antiferromagnetic case) and $\nu \in \mathbb{R}$. There is an algorithm that, given a line graph $L(\Gamma)$ and tolerance $\delta > 0$, produces a spin configuration $\sigma$ from a distribution that is within variation distance $\delta$ of the Gibbs distribution $\mathcal{D}_{L(\Gamma),\beta,\nu}$. The running time of the algorithm is polynomial in $\log \delta^{-1}$ and the instance size.*

One consequence of our results (see Section 5 for details) is that Glauber dynamics for the antiferromagnetic Ising model on a line graph is rapidly mixing. Glauber dynamics moves around the space of configurations by making single spin updates according to a simple probabilistic rule. Its unique stationary distribution is the Gibbs distribution defined earlier. The phrase 'rapidly mixing' indicates that the dynamics converges close to the Gibbs distribution in a number of updates that is polynomial in the size of the graph. In a recent paper, Chen, Liu and Vigoda [6] show that Glauber dynamics mixes rapidly up to a natural threshold for $\beta$, depending on the maximum degree of the graph, beyond which it was already known that the mixing time is exponential. Here we show that, by suitably restricting the underlying graph, we can have rapid mixing for all $\beta$.

Rapid mixing at all $\beta$ is perhaps a surprising phenomenon in the context of the Ising model, and has an interesting relationship to phase transitions in infinite systems. As a concrete example, take the kagome lattice, which is the line graph of the familiar hexagonal lattice. Figure 1 depicts

https://doi.org/10.1017/S0963548321000080 Published online by Cambridge University Press

a portion of the kagome lattice superimposed on the underlying hexagonal lattice. Consider an $L \times L$ patch of the kagome lattice with top and bottom, and left and right side identified (toroidal or periodic boundary conditions). This finite graph is a line graph, so Glauber dynamics mixes in polynomial time on this graph. Heuristically, this ought to imply that the antiferromagnetic Ising model on the infinite kagome lattice does not exhibit a phase transition. In other words, as the interaction strength $\beta$ is continuously varied, no dramatic change occurs in the typical macroscopic configurations. Indeed, Syôzi [22] has shown that the antiferromagnetic Ising model on the kagome lattice does not exhibit a phase transition at any $\beta > 0$ (which is taken to mean that a certain observable called the specific heat is continuous in $\beta$). This is another example of the emerging connection between algorithmic tractability of finite systems and absence of a phase transition in the corresponding infinite system. We leave the connection to phase transitions as a heuristic observation as there are currently no rigorous results in this direction. Indeed, Goldberg, Martin and Paterson [10] show polynomial-time mixing of Glauber dynamics on three colourings of finite portions of the square lattice, and yet the infinite lattice has 'frozen' colourings from which no transition is possible.

We note that the kagome lattice is not the only line graph that has been studied in the context of statistical physics. Another example, this time three-dimensional, is the pyrochlore lattice, which is the line graph of a degree-4 graph: see Jurčišinová and Jurčišin [15].

As the kagome lattice exhibits a phase transition in the ferromagnetic situation, we cannot expect Glauber dynamics to be rapidly mixing for all $\beta < 0$, even on line graphs. However, our approach does show rapid mixing for $\beta_c \leq \beta \leq 0$, for some $\beta_c < 0$ depending on the maximum degree $\Delta$ of $\Gamma$. This is so at least for small $\Delta$ by direct computation, and probably for all $\Delta$, though we have not shown this. See Section 5.

It is natural to question whether *approximation* algorithms are necessary in the context of line graphs. Lemmas 6.1 and 6.3 in Section 6 can be interpreted as asserting that we must settle for approximate results, assuming RP $\neq$ NP: computing the partition function *exactly* is hard even in the case of zero external field (Lemma 6.1), or in the case of a planar graph (Lemma 6.3). As we have noted, imposing both conditions simultaneously leads to tractability. The Ising model on the kagome lattice is exactly solvable [22] with zero external field, but in the presence of an external field, it seems we have to settle for the approximate solution promised by Theorem 1.1.

Finally, we can speculate on the existence of other graph classes on which the antiferromagnetic Ising model can be solved approximately. The maximum cut problem is known to be NP-complete on many graph classes. For example, Bodlaender and Jansen [3] show that maximum cut is NP-hard on several graph classes including chordal graphs and tripartite graphs. For such classes, it is unlikely that the antiferromagnetic Ising model is tractable, even in the approximate sense. Indeed, whenever maximum cut cannot be solved to within a constant factor in polynomial time (under some complexity theoretic assumption) this heuristic claim becomes rigorous. The general argument is given, in the context of maximum independent set, by Luby and Vigoda [18].

Very recently, Bencs, Csikvári and Regts [2], using a completely different approach to the one described here, gave a *deterministic* algorithm for approximating the partition function of an antiferromagnetic Ising model on a line graph.

## 2. Overview of the approach

We use an off-the-shelf analytical technique, known as 'winding', devised by McQuillan [19] and developed and systematised by Huang, Lu and Zhang [12]. As far as the proofs in this article are concerned, we can treat this technique for the most part as a black box. Nevertheless, we do have to peer a little way inside the black box. We'll cover enough of the method for our current needs and leave the interested reader to consult the original sources for details.

The first step is to change our viewpoint away from vertex configurations towards edge configurations. Recall that the spin configurations of the Ising model we are working with live on the

https://doi.org/10.1017/S0963548321000080 Published online by Cambridge University Press

*vertices* of a line graph $L(\Gamma)$. These configurations can equally be thought of as spin configurations living on the *edges* of $\Gamma$. Such a spin model is called a 'holant' in the computer science literature [4]. In the holant view, the partition function of the Ising model has the following form:

$$Z_{\beta,\nu}(L(\Gamma)) = \sum_{\sigma:E(\Gamma)\to\{0,1\}} \prod_{k\in V(\Gamma)} f_k(\sigma|_{E(k)}), \tag{2}$$

where $\sigma|_{E(k)}$ denotes the restriction of $\sigma$ to the edges incident at vertex $k$. The functions $\{f_k\}$ are specific to the Ising model and depend on the vertex degrees; they will be derived in the next section.

The basic approach is Markov chain Monte Carlo (MCMC). As usual, we concentrate on solving the sampling problem (Theorem 1.2) first, as it is a short step from there to a solution of the estimation problem (Theorem 1.1). Our first thought would be to construct a Markov chain whose state space is the set of configurations $\sigma:E(\Gamma)\to\{0,1\}$. However, the winding approach, as developed in [12], requires us to expand the state space somewhat. Think of each edge $\{i,j\}$ of $\Gamma$ as being subdivided into two *half-edges* one incident at $i$ and the other at $j$. Denote by $\mathcal{E}=\mathcal{E}(\Gamma)$ the set of all half-edges of $\Gamma$. Now consider an enlarged set of configurations $\sigma:\mathcal{E}\to\{0,1\}$. We say that a configuration $\sigma$ is *consistent* if, for every edge, the two half-edges that compose it are assigned the same spin. Let $\Omega_0$ denote the set of all consistent configurations, and define

$$\mathcal{H}_0(\Gamma) = \mathcal{H}_0(\Gamma;\{f_k\}) = \sum_{\sigma\in\Omega_0} \prod_{k\in V(\Gamma)} f_k\left(\sigma|_{\mathcal{E}(k)}\right),$$

where $\sigma|_{\mathcal{E}(k)}$ denotes the restriction of $\sigma$ to the half-edges incident at vertex $k$. Clearly $\mathcal{H}_0(\Gamma) = Z_{\beta,\nu}(L(\Gamma))$. We say that a configuration $\sigma$ is *nearly consistent* if it is consistent except for precisely two edges. Denote by $\Omega_2$ be the set of all nearly consistent configurations, and define

$$\mathcal{H}_2(\Gamma) = \mathcal{H}_2(\Gamma;\{f_k\}) = \sum_{\sigma\in\Omega_2} \prod_{k\in V(\Gamma)} f_k\left(\sigma|_{\mathcal{E}(k)}\right).$$

Nearly consistent configurations are needed in general to connect the state space of the Markov chain from [12] that we are about to define, but could be avoided in this specific application: see Section 5.2.

We now consider a Markov chain on state space $\Omega=\Omega_0\cup\Omega_2$ in which a transition is available between every pair of configurations that differ on exactly two half-edges. It is a simple matter to choose transition probabilities so that, in the stationary distribution, configuration $\sigma$ occurs with probability proportional to $\hat{w}(\sigma)$, where

$$\hat{w}(\sigma) = \prod_{k\in V(\Gamma)} f_k\left(\sigma|_{\mathcal{E}(k)}\right).$$

Specifically, Huang et al. use transition probabilities $P:\Omega^2\to[0,1]$ based on the Metropolis filter. Denote by $d(\sigma,\sigma')$ the number of half-edges on which $\sigma$ and $\sigma'$ differ, and let $m=|E(\Gamma)|$. Then

$$P(\sigma,\sigma') = \begin{cases} \frac{1}{4m^2}\min\left\{\frac{\hat{w}(\sigma')}{\hat{w}(\sigma)},1\right\}, & \text{if } d(\sigma,\sigma')=2; \\ 0, & \text{if } d(\sigma,\sigma')\notin\{0,2\}, \end{cases} \tag{3}$$

and the loop probabilities $P(\sigma,\sigma)$ are chosen to satisfy $\sum_{\sigma'\in\Omega}P(\sigma,\sigma')=1$. It may be verified that $P(\sigma,\sigma)>\frac{1}{2}$; this 'laziness' property is convenient in the analysis of mixing time.[b] It is an immediate consequence of the Metropolis transition rule that the stationary distribution assigns probability proportional to $\hat{w}(\sigma)$ to each state $\sigma\in\Omega$.

[b]Note that the 'proposal probability' $1/4m^2$ here is smaller than the one appearing in [12], which is itself too large to guarantee $P(\sigma,\sigma)\geq 0$.

https://doi.org/10.1017/S0963548321000080 Published online by Cambridge University Press

Note that, in particular, the stationary distribution, restricted to $\Omega_0$, is the Gibbs distribution $\mathcal{D}_{\Gamma,\beta,\nu}$ (under the obvious bijection between $\Omega_0$ and assignments $E(\Gamma) \to \{0,1\}$). So we will have a polynomial-time approximate sampler for $\mathcal{D}_{\Gamma,\beta,\nu}$ if (a) the Markov chain is rapidly mixing, i.e. comes close to stationarity in polynomially many steps and (b) the ratio $\mathcal{H}_2(\Gamma)/\mathcal{H}_0(\Gamma)$ is polynomially bounded. Condition (b) ensures that, when we stop the Markov chain after sufficiently many steps, we have a good chance of being at a state in $\Omega_0$. (Intriguingly, the ratio $\mathcal{H}_2(\Gamma)/\mathcal{H}_0(\Gamma)$ also plays a critical role in bounding the mixing time.)

Condition (b) can be swiftly dispensed with.

**Lemma 2.1.** $\mathcal{H}_2(\Gamma)/\mathcal{H}_0(\Gamma) = O(m^2)$, where $m = |E(\Gamma)|$. The implicit constant in the O-notation depends on $\beta$, $\nu$ and the maximum degree $\Delta$ of $\Gamma$.

*Proof.* Given a configuration $\sigma_2 \in \Omega_2$, we can reach a configuration $\sigma_0 \in \Omega_0$ by flipping the spin on two half-edges. This operation changes the weight by a constant factor, i.e. $\hat{w}(\sigma_0) \geq C\hat{w}(\sigma_2)$ where $C$ depends on $\beta$, $\nu$ and $\Delta$. Going backwards, there are $\binom{2m}{2} \leq 2m^2$ half-edges on which we could choose to flip the spin. □

To prove rapid mixing of the Markov chain we use the 'canonical paths' method or, more accurately, its generalisation to 'multi-commodity flow' [21]. The key parameter of interest is the so-called congestion of the flow, as the mixing time scales linearly with congestion. Generally speaking, the construction of appropriate canonical paths and the analysis of their congestion requires expertise and insight. However, for the holant setting with 0/1 spins, this forbidding step can to a large extent be reduced to linear algebra by the winding technique. In this approach, the selection of canonical paths follows automatically from the condition of 'windability' of the functions $f_k$. The concept of windability of functions will be discussed a little later. Then, the conductance of this set of canonical paths can be bounded in terms of the ratio appearing in Lemma 2.1. Thus, in light of that lemma, the canonical paths method and associated concept of conductance can also be treated as a black box.

In principle, windability of any given function can be tested, at least for reasonable arities, with the aid of a computer algebra system. The main obstacle we face in the current application is the need to demonstrate that the particular functions arising from the antiferromagnetic Ising model are windable for all arities (i.e. for all vertex degrees). As will become apparent, overcoming this obstacle presents a challenge. There are few natural infinite families of functions that are known to be uniformly windable, possibly just the examples given by McQuillan [19], namely parity (Even and Odd) and not-all-equal (NAE). So it is significant that the Ising model has this property.

## 3. The Ising model and holants

In this section, we formalise the intuition that the partition function $Z_{\beta,\nu}(L(\Gamma))$ of the Ising model on a line graph $L(\Gamma)$ can be expressed as a holant on the underlying graph $\Gamma$.

Denote the non-negative real numbers by $\mathbb{R}^+$ and the positive integers by $\mathbb{N}_{>0}$. A symmetric function $F{:}\{0,1\}^d \to \mathbb{R}^+$ of arity $d$ has an associated *signature vector* $[f_0, \ldots f_d]$ if $F(x) = f_i$ when $x$ has exactly $i$ coordinates set to 1. If $F$ is also *self-complementary*, that is, $F(x) = F(\bar{x})$ for all $x$, then we have $f_i = f_{d-i}$. In that case, the vector $h = [f_0, \ldots, f_{\lfloor d/2 \rfloor}]$ completely characterises $F$, and will be called the *half-vector* of $F$. Henceforward, all functions $F$ will be symmetric.

For $\beta, \mu \in \mathbb{R}$ and $d \in \mathbb{N}_{>0}$, denote by $F_{\beta,\mu,d}{:}\{0,1\}^d \to \mathbb{R}^+$ the function with signature vector $[e^{\beta i(d-i)+\mu i}{:}0 \leq i \leq d]$. We will write $F_{\beta,d}$ for $F_{\beta,0,d}$. Note that $F_{\beta,d}$ is self-complementary, but $F_{\beta,\mu,d}$ is not, if $\mu \neq 0$. In the derivation below, we show that $Z_{\beta,\nu}(L(\Gamma))$ can be expressed in the form (2), in which, for each vertex $k$, we set $f_k = F_{\beta,\mu,d(k)}$, where $\mu = \nu/2$ and $d(k)$ is the degree of $k$. In what follows, $E(\Gamma)^{(2)}$ denotes the set of unordered pairs of edges of $\Gamma$.

https://doi.org/10.1017/S0963548321000080 Published online by Cambridge University Press

$$\begin{aligned}
Z_{\beta,\nu}(L(\Gamma)) &= \sum_{\sigma:V(L(\Gamma))\to\{0,1\}} \prod_{\{i,j\}\in E(L(\Gamma))} \exp\left(\beta\,|\sigma(i)-\sigma(j)|\right) \prod_{k\in V(L(\Gamma))} \exp\left(\nu\sigma(k)\right) \\
&= \sum_{\sigma:E(\Gamma)\to\{0,1\}} \prod_{\substack{\{e,f\}\in E(\Gamma)^{(2)} \\ e\cap f\neq\emptyset}} \exp\left(\beta\,|\sigma(e)-\sigma(f)|\right) \prod_{e\in E(\Gamma)} \exp\left(\nu\sigma(e)\right) \\
&= \sum_{\sigma:E(\Gamma)\to\{0,1\}} \prod_{k\in V(\Gamma)} \prod_{\substack{\{e,f\}\in E(\Gamma)^{(2)} \\ e,f\ni k}} \exp\left(\beta\,|\sigma(e)-\sigma(f)|\right) \prod_{e\in E(\Gamma)} \exp\left(\nu\sigma(e)\right) \\
&= \sum_{\sigma:E(\Gamma)\to\{0,1\}} \prod_{k\in V(\Gamma)} F_{\beta,d(k)}(\sigma(e_{k,1}),\sigma(e_{k,2}),\ldots,\sigma(e_{k,d(k)})) \prod_{e\in E(\Gamma)} \exp\left(\nu\sigma(e)\right) \\
&= \sum_{\sigma:E(\Gamma)\to\{0,1\}} \prod_{k\in V(\Gamma)} F_{\beta,\mu,d(k)}(\sigma(e_{k,1}),\sigma(e_{k,2}),\ldots,\sigma(e_{k,d(k)})),
\end{aligned}$$

where $e_{k,1}, e_{k,2}, \ldots, e_{k,d(k)}$ is an enumeration of edges incident at vertex $k$, and $\mu = \nu/2$.

Comparing with (2), it can be seen that we have successfully expressed the partition function $Z_{\beta,\nu}(L(\Gamma))$ as a holant on $\Gamma$.

## 4. Windability

This section will be a proof of the following.

**Theorem 4.1.** *The function $F_{\beta,\mu,d}$ is windable for all $\beta \in \mathbb{R}^+$, $\mu \in \mathbb{R}$ and $d \in \mathbb{N}$.*

The definition of windable can be found in [19] or [12]. However, a knowledge of the definition is not essential for an understanding of this paper, as we shall be using the characterisation of windable that was given by Huang et al., and that is captured in Theorem 4.2. In order to understand that theorem we need some preliminary definitions.

Suppose $m \in \mathbb{N}_{>0}$ and let $n = \lfloor m/2 \rfloor$. Recall from [12] the definition of the matrix $A_m \in \mathbb{R}^{(n+1)\times(n+1)}$, whose coefficients $a_{i,j}^{(m)}$ $(0 \le, i, j \le n)$ are, for $m = 2n$ even,

$$a_{i,j}^{(m)} = \begin{cases} \binom{i}{j}\binom{2n-i}{j} j!(i-j-1)!!(2n-i-j-1)!! & \text{if } i \equiv j \bmod 2, \\ 0 & \text{otherwise;} \end{cases} \tag{4}$$

or, for $m = 2n+1$ odd,

$$a_{i,j}^{(m)} = \begin{cases} \binom{i}{j}\binom{2n+1-i}{j} j!(i-j-1)!!(2n+1-i-j)!! & \text{if } i \equiv j \bmod 2, \\ \binom{i}{j}\binom{2n+1-i}{j} j!(i-j)!!(2n-i-j)!! & \text{otherwise,} \end{cases} \tag{5}$$

where $x!!$ is the double factorial $x!! := x \cdot (x-2) \cdot (x-4) \ldots$ for $x \in \mathbb{N}$, and we use the convention $(-1)!! = 1$.[c]

The matrix entries $a_{i,j}^{(m)}$ have a combinatorial meaning: the number of ways of pairing $m$ objects of two types, $i$ of the first type and $m-i$ of the second, such that exactly $j$ pairs have both types. There will be one singleton if $m = 2n+1$. It follows that $\sum_{j=0}^{i} a_{i,j}^{(m)}$ is simply the total number of

[c]The second formula has a typo in [12], and is missing '!!'.

https://doi.org/10.1017/S0963548321000080 Published online by Cambridge University Press

ways of pairing $m$ objects, so

$$\sum_{j=0}^{i} a_{i,j}^{(m)} = \begin{cases} (2n-1)!! & \text{if } m = 2n; \\ (2n+1)!! & \text{if } m = 2n+1. \end{cases} \tag{6}$$

The expressions for $a_{i,j}^{(m)}$ may be rewritten in a more convenient form, which separates, as far as possible, factors depending only on $i$ from those depending only on $j$. Starting with definition (4) and assuming $i$ and $j$ have the same parity,

$$\begin{aligned} a_{i,j}^{(m)} &= (2n)!\,\frac{i!(2n-i)!}{(2n)!}\,\frac{1}{j!}\,\frac{(i-j-1)!!}{(i-j)!}\,\frac{(2n-i-j-1)!!}{(2n-i-j)!} \\ &= (2n)!\binom{2n}{i}^{-1}\frac{1}{j!(n-j)!}\,\frac{(n-j)!}{2^{(i-j)/2}[(i-j)/2]!\,2^{n-(i+j)/2}[n-(i+j)/2]!} \\ &= \frac{(2n)!}{2^n n!}\binom{2n}{i}^{-1} 2^j \binom{n}{j}\binom{n-j}{(i-j)/2}. \end{aligned}$$

Thus we reach the following equivalent form of (4), valid for $m = 2n$ even:

$$a_{i,j}^{(m)} = \begin{cases} (2n-1)!!\binom{2n}{i}^{-1} 2^j \binom{n}{j}\binom{n-j}{(i-j)/2} & \text{if } i \equiv j \bmod 2\,, \\ 0 & \text{otherwise}\,. \end{cases} \tag{7}$$

Starting with (5) and performing very similar manipulations we obtain the following, valid for $m = 2n+1$ odd:

$$a_{i,j}^{(m)} = (2n+1)!!\binom{2n+1}{i}^{-1} 2^j \binom{n}{j}\binom{n-j}{\lfloor (i-j)/2 \rfloor}. \tag{8}$$

Clearly $a_{i,j}^{(m)} = 0$ for $i < j$, so $A_m$ is lower triangular. (We use the usual convention that $\binom{b}{a} = 0$ if $a < 0$ or $b < a$.)

A *pinning* of a function $F{:}\{0,1\}^d \to \mathbb{R}^+$ is a function $G{:}\{0,1\}^m \to \mathbb{R}^+$ of smaller or equal arity that is obtained from $F$ by setting certain variables to 0 or 1. If $F$ is symmetric, then the it does not matter which variables are set to 0 (or 1), only their number. Thus, a pinning of $F$ may be written as

$$G(x_1,\ldots,x_m) = F\left(x_1,\ldots,x_m,\underbrace{0,\ldots,0}_{d-b},\underbrace{1,\ldots,1}_{a}\right),$$

where $0 \le a \le b \le d$ and $m = b - a$. Note that, if the signature vector of $F$ is $[f_0,\ldots,f_d]$, then that of $G$ is $[f_a,\ldots,f_b]$. Now, to prove that $F_{\beta,\mu,d}$ is windable, we use the following result from [12].

**Theorem 4.2.** [12], Theorem 7. *A symmetric function $F{:}\{0,1\}^d \mapsto \mathbb{R}^+$ is windable if and only if, for every $1 \le m \le d$ and every pinning $G$ of $F$ with arity $m$, the self-complementary function $H(x) \stackrel{def}{=} G(x)G(\bar{x})$ with half-vector $h = [h_0, h_1, \ldots, h_n]$ (where $n = \lfloor m/2 \rfloor$) satisfies the following condition: The linear equations $A_m x = h$ have a non-negative solution $x \in (\mathbb{R}^+)^{n+1}$.*

To apply this, we first determine $H(x) = G(x)G(\bar{x})$ for all pinnings $G$ of $F_{\beta,\mu,d}$. To help guide the reader through the remainder of the section, the naming conventions in Theorem 4.2 will be maintained throughout. So, $d$ will be the arity of some function $F$, integer $m$ with $1 \le m \le d$ the

https://doi.org/10.1017/S0963548321000080 Published online by Cambridge University Press

arity of some self-complementary function $H$ derived from a pinning of $F$, and $n = \lfloor m/2 \rfloor$. Thus the indices of the half-vector $h$ representing $H$ run from 0 to $n$.

**Lemma 4.1.** Let $d \in \mathbb{N}_{>0}$ and $\beta, \mu \in \mathbb{R}$, and let $G$ be a pinning of $F_{\beta,\mu,d}$. Define $H$ as in Theorem 4.2. Then, there exists $m \leq d$ and a constant $K \geq 0$ such that $H = K \cdot F_{2\beta,m}$.

*Proof.* Let $[f_0, \ldots, f_d]$ be the signature vector of $F_{\beta,\mu,d}$, where $f_i = e^{\beta i(d-i)+\mu i}$. Let $G$ be a pinning of $F$, with $a$ variables pinned to 1, and $d-b$ to 0. Its signature vector is of the form $[f_i{:}a \leq i \leq b]$. Let $[z_i{:}0 \leq i \leq m]$ be the signature vector of $H$, where $m = b - a \leq d$. Then we have:

$$z_i = e^{\beta(a+i)(d-a-i)+\mu(a+i)} \cdot e^{\beta(b-i)(d-b+i)+\mu(b-i)} = e^{Z_i},$$

where the exponent $Z_i$ is

$$\begin{aligned} Z_i &= \beta((a+i)(d-a-i) + (b-i)(d-b+i)) + \mu(a+b) \\ &= \beta(a(d-a) + b(d-b)) + \mu(a+b) + 2\beta i(b-a-i). \end{aligned}$$

If $m = b - a$ and $K = e^{\beta(a(d-a)+b(d-b))+\mu(a+b)}$, this implies that we have:

$$z_i = Ke^{2\beta i(m-i)} = KF_{2\beta,m}.$$

Thus $H$ is a positive multiple of the function $F_{2\beta,m}$. □

From Theorem 4.2 and Lemma 4.1, it follows that

**Corollary 4.2.** *$F_{\beta,\mu,d}$ is windable if and only if for all $1 \leq m \leq d$ the following holds: the system $A_m x = h$ has a solution with $x \geq \mathbf{0}$, where $h$ is the half-vector representing the function $F_{2\beta,m}$.*

We can put the equations $A_m x = h$ in a more convenient form. Let

$$b_{i,j}^{(m)} = \begin{cases} \binom{n-j}{\lfloor (i-j)/2 \rfloor} & \text{if } d = 2n+1 \text{ or } i \equiv j \bmod 2 \\ 0 & \text{otherwise.} \end{cases}$$

It follows from (7,8) that the equations $A_m x = h$ can be rewritten in the following form:

$$\sum_{j=0}^{n} b_{i,j}^{(m)} y_j = \begin{cases} \dfrac{1}{(2n-1)!!}\dbinom{m}{i} h_i & \text{if } m = 2n, \\ \dfrac{1}{(2n+1)!!}\dbinom{m}{i} h_i & \text{if } m = 2n+1, \end{cases} \tag{9}$$

where $y_j = 2^j \binom{n}{j} x_j$ for all $j \in [0, n]$. Let $B_m$ be the $n \times n$ matrix $(b_{i,j}^{(m)}{:}i, j \in [0, n])$, and let $b_j$ be its $j$th column, so $B_m = [b_0\ b_1\ \cdots\ b_n]$. Then we can express the equations $A_m x = h$ in a form which suppresses irrelevant scalars.

The (convex) *cone* $\mathcal{C}(S) \subseteq \mathbb{R}^{n+1}$ generated by the set $S \subseteq \mathbb{R}^{n+1}$ is the closure of $S$ under addition and multiplication by non-negative scalars. That is, for all $x \in \mathcal{C}(S)$, $\kappa x \in \mathcal{C}(S)$ for any $\kappa \geq 0$ and, for all $x, y \in \mathcal{C}(S)$, $x + y \in \mathcal{C}(S)$. Clearly we always have $\mathbf{0} \in \mathcal{C}(S)$. Also, establishing $x \in \mathcal{C}(S)$ is clearly equivalent to establishing $\kappa x \in \mathcal{C}(S)$ for any $\kappa > 0$. Then we may restate Theorem 4.2 as

**Theorem 4.3.** *Let $F{:}\{0,1\}^d \mapsto \mathbb{R}^+$ be a symmetric function and, let $G$ be a pinning of $F$ with arity $m$. Let $H(x) \stackrel{def}{=} G(x)G(\bar{x})$ with half-vector $h = [h_0, h_1, \ldots, h_n]$, where $n \stackrel{def}{=} \lfloor m/2 \rfloor$. Let $C_n \subset \mathbb{R}^{n+1}$ be the cone generated by the columns of $B_m$, and let $z(h)$ be the vector such that $z(h)_i = \binom{m}{i} h_i$ $(i \in [0, n])$. Then $F$ is windable if and only if $z(h) \in C_n$ for every such $G$.*

https://doi.org/10.1017/S0963548321000080 Published online by Cambridge University Press

Lemma 4.1 implies that we need only apply this criterion to functions $H$ of the form $F_{\beta,d}$.

From (9), we have $z(\mathbf{1}) \in C_n$, where $\mathbf{1}$ is the all-1's vector. (To see this, substitute $x = \mathbf{1}$ in (6) to give $A\mathbf{1} = h$, where $h = (2n \pm 1)!!\,\mathbf{1}$. Now transform variables from $x$ to $y$ and compare with (9).) Observe that $b_{i,j}^{(m)} = b_{i-k,j-k}^{(m-2k)}$, for all $k \in [0, n]$. (Informally, deleting the first row and column of the matrix $B_m$ yields the matrix $B_{m-2}$.) Then it follows that there exists a vector of the form

$$\hat{y} = \Big(\underbrace{0, 0, \ldots, 0}_{k \text{ copies}}, y_0, y_1, \ldots, y_{n-k}\Big)$$

such that

$$\sum_{j=0}^{n} b_{i,j}^{(m)} \hat{y}_j = \sum_{j=k}^{n} b_{i,j}^{(m)} y_{j-k} = \sum_{j=0}^{n-k} b_{i-k,j}^{(m-2k)} y_j = \begin{cases} \dfrac{1}{(2n-2k-1)!!}\dbinom{m-2k}{i-k} & \text{if } m = 2n, \\ \dfrac{1}{(2n-2k+1)!!}\dbinom{m-2k}{i-k} & \text{if } m = 2n+1. \end{cases} \tag{10}$$

Specifically, we can choose $(y_0, y_1, \ldots, y_{n-k})$ to be the vector that witnesses, via (9), the membership of $z(\mathbf{1}_{n-k+1})$ in the cone $C_{n-k}$. Thus, for all $k \in [0, n]$, the vector

$$v_k = \left[\binom{m-2k}{i-k} : i \in [0, n]\right]$$

is in the cone $C_n$. Let $D_n$ be the cone generated by the set $\{v_0, v_1, \ldots, v_n\}$. Then $D_n \subseteq C_n$. Thus, to show $z(h) \in C_n$, it suffices to show that $z(h) \in D_n$.

**Lemma 4.6.** $z(h) \in D_n$, where $h$ is the half-vector representing $F_{\beta,m}$

*Proof.* Expanding the exponential function in the definition of $h_i$,

$$z(h)_i = \binom{m}{i} h_i = \sum_{j=0}^{\infty} \frac{\beta^j}{j!} \big(i(m-i)\big)^j \binom{m}{i} \qquad (i \in [0, n]).$$

Thus, the claim $z(h) \in D_n$ will follow from showing that the vector with entries $(i(m-i))^j \binom{m}{i}$ $(i \in [0, n])$ is in $D_n$ for all $j \in \mathbb{N}$.

Now, since $v_0 = z(\mathbf{1}) \in D_n$, the inclusion $z(h) \in D_n$ will follow, by induction on $j$, from the following

**Claim**: If $u \in D_n$, then the the vector $u^*$ with entries $u_i^* = i(m-i)u_i$ is in $D_n$.

*Proof.* It suffices to prove this for the generators of $D_n$, namely $v_k = [v_{0,k}, v_{1,k}, \ldots, v_{n,k}]$. Then, if $v_{-1} \stackrel{\text{def}}{=} \mathbf{0}$,

$$\begin{aligned} i(m-i)v_{i,k} &= [(i-k)(m-k-i) + k(m-k)]\binom{m-2k}{i-k} \\ &= \frac{(m-2k)!}{(i-k-1)!(m-k-i-1)!} + k(m-k)\binom{m-2k}{i-k} \\ &= (m-2k)(m-2k-1)\frac{(m-2k-2)!}{(i-k-1)!(m-k-i-1)!} + k(m-k)\binom{m-2k}{i-k} \\ &= (m-2k)(m-2k-1)v_{i,k+1} + k(m-k)v_{i,k} \end{aligned}$$

https://doi.org/10.1017/S0963548321000080 Published online by Cambridge University Press

So $v_k^* = (m-2k)(m-2k-1)v_{k+1} + k(m-k)v_k \in \mathcal{C}(v_{k+1}, v_k) \subseteq D_n$. Note that the scalars appearing in the above linear combination of $v_{k+1}$ and $v_k$ are non-negative: if $m = 2n$ is even and $k = n$ then factor $m - 2k - 1$ is negative, but in this case $m - 2k = 0$. □

Theorem 4.1 follows by combining Lemma 4.1, Theorem 4.3, Lemma 4.6 and the fact that $D_n \subseteq C_n$.

## 5. Positive results for approximate computation

It is now a short step to the two results claimed in the Introduction. Note that in this section $n = |V(\Gamma)|$ and $m = |E(\Gamma)|$ will always denote the number of vertices and edges of an instance graph $\Gamma$.

### 5.1 Proofs of the main results

*Proof of Theorem 1.2.* This result follows directly from [12, Lemma 29] using Theorem 4.1 and Lemma 2.1. Note that the quantity $\mu_\Lambda(\Omega_0)$ appearing in that lemma is just $\mathcal{H}_0(\Gamma)/(\mathcal{H}_0(\Gamma) + \mathcal{H}_2(\Gamma))$ in our notation, and hence $\mu_\Lambda(\Omega_0)^{-1} = O(m^2)$. □

*Proof of Theorem 1.1.* The reduction from estimating the partition function to sampling configurations is standard. Choose a suitably spaced sequence of values $0 = \beta_0 < \beta_1 < \cdots < \beta_r = \beta$. The ratios $Z_{\beta_i,\nu}(\Gamma)/Z_{\beta_{i-1},\nu}(\Gamma)$, for $1 \le i \le r$, may be estimated from polynomially many samples from the distribution $\mathcal{D}_{\Gamma,\beta_{i-1},\nu}$ using the Markov chain described in Section 2, which mixes in polynomial time by Theorem 1.2. The partition function $Z_{0,\nu}(\Gamma)$ factorises, and is hence easy to compute. Thus, $Z_{\beta,\nu}(\Gamma)$ itself may be estimated from the telescoping product

$$Z_{\beta,\nu}(\Gamma) = \frac{Z_{\beta_r,\nu}(\Gamma)}{Z_{\beta_{r-1},\nu}(\Gamma)} \times \frac{Z_{\beta_{r-1},\nu}(\Gamma)}{Z_{\beta_{r-2},\nu}(\Gamma)} \times \cdots \times \frac{Z_{\beta_1,\nu}(\Gamma)}{Z_{\beta_0,\nu}(\Gamma)} \times Z_{0,\nu}(\Gamma).$$

Details can be found in [23], together with refinements.

Alternatively, one could show self-reducibility of the partition function and apply [12, Thm 3]. Details of this approach can be found in [13, Section 3.2]. □

### 5.2 Glauber dynamics

Theorem 1.2 was established by considering a Markov chain based on updates to half-edges of $\Gamma$ (see Section 2). A more natural approach would be to work with whole edges. So, the configurations would simply be the natural ones, namely $\sigma : E(\Gamma) \to \{0, 1\}$, and a transition would update a single (whole) edge. We might call this the (single-site) Glauber dynamics. As usual, it is easy to arrange transition probabilities so that the stationary distribution of the Markov chain is the Gibbs distribution.

A convenient way to define Glauber dynamics for our purposes is the following. Start with the half-edge Markov chain on $\Omega = \Omega_0 \cup \Omega_2$ as defined in (3), and censor all transitions from a state in $\Omega_0$ to a state in $\Omega_2$. That is to say, instead of making a transition from $\tau \in \Omega_0$ to $\tau' \in \Omega_2$, the Markov chain remains at $\tau$. Since the number of available transitions out of any state is now $m$ instead of $\binom{2m}{2}$, we take the opportunity to increase the proposal probability from $1/4m^2$ to $1/2m$. This simple modification speeds up the Markov chain and reduces the mixing time by a factor $2m$. Assuming the initial state is in $\Omega_0$, the censored Markov chain on state space $\Omega_0$ is ergodic with stationary distribution $\mathcal{D}_{\Gamma,\beta,\nu}$. Note that every transition of this derived Markov chain has the following property: the two half-edges whose spins are flipped belong to the same whole edge. Since $\Omega_0$ is isomorphic to $2^{E(\Gamma)}$ it would have been more natural to define the Glauber dynamics directly on whole-edge configurations, but the above view simplifies the analysis that follows.

https://doi.org/10.1017/S0963548321000080 Published online by Cambridge University Press

To show that Glauber dynamics is rapidly mixing, we will compare it with the already analysed half-edge Markov chain. To follow the argument in detail, some acquaintance with the canonical paths method is required, and in particular the use made of it by Huang et al. [12]. However, it should be possible to obtain an intuitive grasp of the argument, without that prerequisite. Denote the transition probabilities of the half-edge Markov chain by $P{:}\Omega^2 \to [0,1]$ and the stationary distribution by $\pi{:}\Omega \to [0,1]$. Imagine the Markov chain as a transition graph with states as vertices and transitions as edges. Each edge (transition) $(\tau, \tau')$ is assigned a capacity $Q(\tau,\tau') = \pi(\tau)P(\tau,\tau')$. Since the Markov chain is time-reversible, $Q(\tau,\tau') = Q(\tau',\tau)$ and the capacity is well defined. For each pair of states $\sigma, \sigma'$, we specify a collection of (canonical) paths from $\sigma$ to $\sigma'$ through which we route a flow of $\pi(\sigma)\pi(\sigma')$. Suppose we can do this in such a way that the total flow through any transition $\tau \to \tau'$ is small, specifically, less than or equal to $\rho\, Q(\tau,\tau')$ for some uniform bound $\varrho$. Then the relaxation time of the Markov chain (a quantity strongly related to the mixing time) is known to be $O(\rho)$. The quantity $\rho$ is the *congestion* of the chosen canonical paths. For more details see, e.g. [13, Section 5.2].

The canonical paths exploited first by McQuillan and later by Huang et al., are derived from collection of (graph theoretic) paths and cycles in the instance graph $\Gamma$. Given such a collection, the corresponding canonical path in the transition graph is obtained by 'unwinding' these paths and cycles in sequence. Unwinding consists in tracing along the path or around the cycle, at each vertex flipping the two half-edges incident at that vertex. (See [12, Section B] for details.) Suppose that $e_0, e_1, e_2, e_3, \ldots, e_{2\ell-2}, e_{2\ell-1}$ is a path or cycle in $\Gamma$ of length $\ell$, composed of half-edges. For each $0 \le i \le \ell - 1$, the half-edges $e_{2i}, e_{2i+1}$ form a single edge, and for each $1 \le i \le \ell - 1$, the half-edges $e_{2i-1}, e_{2i}$ are incident at a common vertex. The half-edges $e_0$ and $e_{2\ell-1}$ may or may not be incident at a common vertex. In Huang et al's analysis, the half-edges are flipped in the sequence $(e_{2\ell-1}, e_0), (e_1, e_2), \ldots, (e_{2\ell-3}, e_{2\ell-2})$. If we pair the half-edges in the other natural way, i.e. $(e_0, e_1), (e_2, e_3), \ldots, (e_{2\ell-2}, e_{2\ell-1})$, then we obtain a canonical path that flips whole edges at each step. In this way, a canonical path on half-edges induces a canonical path on (whole) edges. As we shall see, these canonical paths witness rapid mixing of the Glauber dynamics. It is important in this context to note that the signatures we are working with are all permissive, in the sense that each function $f_v$ is supported on the whole of $\{0,1\}^{E(v)}$. In other situations, the derived paths might be invalid.

In more detail, suppose

$$\sigma = \sigma_0 \to \sigma_1 \to \cdots \to \sigma_i \to \sigma_{i+1} \to \cdots \to \sigma_\ell = \sigma',$$

is the canonical path in $\Omega_0 \cup \Omega_2$ from $\sigma \in \Omega_0$ to $\sigma' \in \Omega_0$ that results from the sequence $(e_{2\ell-1}, e_0), (e_1, e_2), \ldots, (e_{2\ell-3}, e_{2\ell-2})$ of half-edge flips. Then let

$$\sigma = \tilde{\sigma}_0 \to \tilde{\sigma}_1 \to \cdots \to \tilde{\sigma}_i \to \tilde{\sigma}_{i+1} \to \cdots \to \tilde{\sigma}_\ell = \sigma',$$

be the derived path in $\Omega_0$ resulting from the sequence $(e_0, e_1), (e_2, e_3), \ldots, (e_{2\ell-2}, e_{2\ell-1})$. Thus, for $0 \le i \le \ell - 1$,

$$\tilde{\sigma}_{i+1}(e) = \begin{cases} 1 - \tilde{\sigma}_i(e), & \text{if } e \in \{e_{2i}, e_{2i+1}\}; \\ \tilde{\sigma}_i(e), & \text{otherwise.} \end{cases}$$

The derived path stays close to the original path at all times; specifically, for $1 \le i \le \ell - 1$,

$$\tilde{\sigma}_i(e) = \begin{cases} 1 - \sigma_i(e), & \text{if } e \in \{e_{2i-1}, e_{2\ell-1}\}; \\ \sigma_i(e), & \text{otherwise.} \end{cases} \tag{11}$$

Consider a transition $\tilde{\sigma}_i \to \tilde{\sigma}_{i+1}$ in the derived path. By comparing $\tilde{\sigma}_i$ and $\tilde{\sigma}_{i+1}$ we may deduce the unordered pair of edges $\{e_{2i}, e_{2i+1}\}$, and with one extra bit of information we may identify

https://doi.org/10.1017/S0963548321000080 Published online by Cambridge University Press

the edges $e_{2i}$ and $e_{2i+1}$ individually. If we now specify $e_{2i-1}$ and $e_{2\ell-1}$, we may recover the corresponding transition $\sigma_i \to \sigma_{i+1}$ in the original path, by putting (11) into reverse. (The first and last transitions are special, but the given data is still sufficient to recover the transition $\sigma_i \to \sigma_{i+1}$.) Since $e_{2i-1}$ shares a vertex with edge $e_{2i}$, we have at most $n$ choices for $e_{2i-1}$, where $n = |V(\Gamma)|$. The upshot of this is that, given the transition $\tilde{\sigma}_i \to \tilde{\sigma}_{i+1}$, there are at most $2nm$ possibilities for the corresponding transition $\sigma_i \to \sigma_{i+1}$. Suppose $\tilde{\tau} \to \tilde{\tau}'$ is any transition with $\tilde{\tau}, \tilde{\tau}' \in \Omega_0$. Define the *halo* around $\tilde{\tau} \to \tilde{\tau}'$ to be the set of $2nm$ transitions $\tau \to \tau'$ that correspond to $\tilde{\tau} \to \tilde{\tau}'$ via the above procedure. Note that the flow in any derived path through $\tilde{\tau} \to \tilde{\tau}'$ must originally have passed through some transition in this (small) halo.

Now observe that the capacity of any transition $\tau \to \tau'$ in the halo of transition $\tilde{\tau} \to \tilde{\tau}'$ is not too much bigger that the capacity of $\tilde{\tau} \to \tilde{\tau}'$. Since $\tau$ differs from $\tilde{\tau}$ at at most two half-edges, the weights $\hat{w}(\tau)$ and $\hat{w}(\tilde{\tau})$ are within a constant factor of each other; the same applies to $\hat{w}(\tau')$ and $\hat{w}(\tilde{\tau}')$. It follows that $Q(\tau, \tau') \le CQ(\tilde{\tau}, \tilde{\tau}')$, where $C = C(\beta, \nu, \Delta)$. Suppose $\rho$ is the (known) congestion of the half-edge Markov chain. Then the flow in transition $\tau \to \tau'$ is bounded by $\rho\, Q(\tau, \tau') \le C\rho\, Q(\tilde{\tau}, \tilde{\tau}')$. So if we reroute the flow between pairs of states in $\Omega_0$ through the derived paths, the new flow in transition $\tilde{\tau} \to \tilde{\tau}'$ is bounded by $2nm \times C\rho\, Q(\tilde{\tau}, \tilde{\tau}')$. So rerouting flow through paths lying entirely within $\Omega_0$ increases congestion by at most a factor $2Cnm$.

Since the derived paths lie entirely within $\Omega_0$, they can be used to bound the congestion of the censored Markov chain (Glauber dynamics). Denote the transition probabilities of this full-edge Markov chain by $\widetilde{P}:\Omega_0^2 \to [0, 1]$ and the stationary distribution by $\tilde{\pi}:\Omega_0 \to [0, 1]$. Also define the capacities of transitions in the censored Markov chain by $\widetilde{Q}(\tau, \tau') = \tilde{\pi}(\tau)\widetilde{P}(\tau, \tau')$. Let $R = \pi(\Omega_0)^{-1}$, and note that $\tilde{\pi}(\tau) = R\pi(\tau)$ and

$$\widetilde{Q}(\tau, \tau') = \tilde{\pi}(\tau)\widetilde{P}(\tau, \tau') = R\pi(\tau) \times 2mP(\tau, \tau') = 2RmQ(\tau, \tau'),$$

for all $\tau, \tau' \in \Omega_0$. Thus, the flows are increased by a factor $R^2$ but the capacities are increased by a factor $2Rm$. Thus, the congestion $\tilde{\rho}$ of the censored Markov chain is increased by a factor $R/2m$. Putting it all together, $\tilde{\rho} = 2Cnm\rho R^2/2mR = CnR\rho$. Since $R = O(m^2)$, we see that the relaxation time (and hence the mixing time) of Glauber dynamics is at most $O(nm^2)$ longer than that of the half-edge Markov chain. The comparison argument could be tightened by delving further into the details of the canonical paths analysis in [12].

### *5.3 Ferromagnetic interactions*

Direct calculation from Theorem 4.2 shows that $[1, e^{2\beta}, e^{2\beta}, 1]$ is windable when $e^{4\beta} \ge \frac{1}{2}$. Thus, Glauber dynamics is rapidly mixing on line graphs of cubic graphs (such as fragments of the kagome lattice) when $\beta > -0.173286$. The kagome lattice exhibits a phase transition when $e^{4\beta} = 1/(3 + 2\sqrt{3})$ [22], so we cannot expect Glauber dynamics to be rapidly mixing when $\beta < -0.466567$.

Rapid mixing can also be deduced from Theorem 4.2 for graphs of maximum degree $\Delta$, for small values of $\Delta > 3$ and all $\beta$ in some range $\beta_\Delta \le \beta < 0$. It seems likely that a general argument could be found covering all $\Delta$, though we have not done this. However, $\beta_\Delta$ will necessarily converge to 0 as $\Delta \to \infty$. For $\Delta$ even, consider the star $\Gamma = K_{1,\Delta}$, the line graph of which is the complete graph $K_\Delta$. Let $\Omega$ denote the set of all configurations $E(\Gamma) \to \{0, 1\}$ and $\Omega_{1/2}$ denote the subset of balanced configurations with equal numbers of 0s and 1s. Then

$$\sum_{\sigma \in \Omega} \hat{w}(\sigma) \ge 2 \qquad \text{and} \qquad \sum_{\sigma \in \Omega_{1/2}} \hat{w}(\sigma) \le 2^\Delta \exp(\beta\Delta^2/4),$$

where the first inequality comes from considering just the all 0s and all 1s configurations. When $\beta < -4/\Delta$, the second sum is exponentially smaller than the first. Since $\Omega_{1/2}$ forms a cut in the state space, a standard conductance bound [21, Theorem 2] then shows that mixing time of Glauber dynamics is exponential in $\Delta$.

https://doi.org/10.1017/S0963548321000080 Published online by Cambridge University Press

### 5.4 Local fields

Finally, varying fields, with an individual strength $\nu_k$ for each vertex $k$ of $L(\Gamma)$ can be handled. One way is to repeat the calculations in Sections 3 and 4 with $\nu_k$ and $\mu_k$ in place of $\mu$ and $\nu$. The notational complexity increases, but there are no essential changes. However, the same end can be achieved with less effort as follows. Subdivide each edge of $\Gamma$ by a single vertex with signature vector $[1, 0, e^{\nu_k}]$. This signature is windable by [12, Lemma 21], and correctly implements the field acting on a single spin. The Markov chain on the extended configurations is rapidly mixing, as before.

## 6. Negative results for exact computation

The problem Holant($[x_0, x_1, x_2, x_3]$) is the following. Given a cubic graph $\Gamma$, evaluate the holant (2), where $f_k$ is the function with signature vector $[x_0, x_1, x_2, x_3]$, for all $k \in V(\Gamma)$. The problem Pl-Holant($[x_0, x_1, x_2, x_3]$) is the same with the instance $\Gamma$ restricted to be planar. According to the correspondence derived in Section 3, the partition function of the antiferromagnetic Ising model on a line graph of a cubic graph can be expressed as an instance of Holant($[1, a, a, 1]$) with $a > 1$.[d] Lemma 6.1 thus implies that computing *exactly* the partition function of the antiferromagnetic Ising model on a graph $\Gamma$ is #P-hard, even in the absence of an external field, and even when $\Gamma$ is the line graph of a cubic graph. The following results exploit a tiny fraction of a vast theory of the computational complexity of holant problems. To see the wider scope, refer to Cai and Chen's monograph [4].

**Lemma 6.1.** Holant($[1, a, a, 1]$) is #P-hard, for all $a \in \mathbb{Q} \cap (1, \infty)$.

The route to proving this lemma is via an intermediate computational problem. Suppose $\Gamma'$ is a bipartite graph with vertices of degree two on one side of the bipartition and of degree three on the other. The problem Holant($[y_0, y_1, y_2] \mid [1, 0, 0, 1]$) is the following. Given a graph $\Gamma'$ of the specified form, evaluate the holant (2), where vertices of degree two have signature vector $[y_0, y_1, y_2]$ and those of degree three have signature vector $[1, 0, 0, 1]$ (i.e. ternary equality). Following Cai, Huang and Lu, [5, Section 3] we derive (possibly complex) numbers $y_0, y_1, y_2$ such that Holant($[x_0, x_1, x_2, x_3]$) is polynomial-time interreducible with Holant($[y_0, y_1, y_2] \mid [1, 0, 0, 1]$). We then complete the proof of Lemma 6.1 by invoking the following dichotomy result.

**Lemma 6.2.** [5, Theorem 2]. The problem Holant($[y_0, y_1, y_2] \mid [1, 0, 0, 1]$) is #P-hard for all $y_0, y_1, y_2 \in \mathbb{C}$, except in the following cases, for which the problem is in P: (1) $y_1^2 = y_0 y_2$; (2) $y_0^{12} = y_1^{12}$ and $y_0 y_2 = -y_1^2$ ($y_1 \neq 0$); (3) $y_1 = 0$; and (4) $y_0 = y_2 = 0$.

*Proof of Lemma 6.1.* In the notation of [5, Section 3], we have $x_0 = x_3 = 1$ and $x_1 = x_2 = a$. The first step is a change of variables. Write

$$\begin{aligned} x_0 &= \alpha_1^3 + \beta_1^3 \\ x_1 &= \alpha_1^2 \alpha_2 + \beta_1^2 \beta_2 \\ x_2 &= \alpha_1 \alpha_2^2 + \beta_1 \beta_2^2 \\ x_3 &= \alpha_2^3 + \beta_2^3, \end{aligned}$$

[d] To avoid issues with the representation of real numbers, and so obtain a clean computational problem, we introduce $a = e^{2\beta}$ and assume $a \in \mathbb{Q}$.

https://doi.org/10.1017/S0963548321000080 Published online by Cambridge University Press

where $\alpha_1, \alpha_2, \beta_1, \beta_2$ are to be determined. Setting $\alpha_1 = \beta_2 = e^{i\theta}$ and $\alpha_2 = \beta_1 = e^{-i\theta}$ we have

$$\begin{aligned} x_0 &= 2\cos 3\theta \\ x_1 &= 2\cos\theta \\ x_2 &= 2\cos\theta \\ x_3 &= 2\cos 3\theta. \end{aligned}$$

Then, by setting $0 < \theta < \frac{\pi}{6}$ appropriately, the signature vector $[x_0, x_1, x_2, x_3]$ can be made to match $[1, a, a, 1]$ up to a scaling factor. Following Cai et al., introduce the matrix

$$T = \begin{bmatrix} \alpha_1 & \beta_1 \\ \alpha_2 & \beta_2 \end{bmatrix} = \begin{bmatrix} e^{i\theta} & e^{-i\theta} \\ e^{-i\theta} & e^{i\theta} \end{bmatrix},$$

and (in the compact notation used in work on holants) define $[y_0, y_1, y_2] = [1, 0, 1]T^{\otimes 2} = [2\cos 2\theta, 2, 2\cos 2\theta]$. This last identity can be expanded into a less compact but more generally accessible matrix equation:

$$\begin{bmatrix} y_0 & y_1 \\ y_1 & y_2 \end{bmatrix} = T^{\mathsf{T}} I_2 T = \begin{bmatrix} e^{i\theta} & e^{-i\theta} \\ e^{-i\theta} & e^{i\theta} \end{bmatrix} \begin{bmatrix} e^{i\theta} & e^{-i\theta} \\ e^{-i\theta} & e^{i\theta} \end{bmatrix} = \begin{bmatrix} 2\cos 2\theta & 2 \\ 2 & 2\cos 2\theta \end{bmatrix}.$$

(See the exercise preceding Theorem 1.6 of [4].) So we have $y_0 = y_2 = 2\cos\theta$ and $y_1 = 2$. Since none of the four conditions involving $y_0, y_1, y_2$ in the statement of Lemma 6.2 hold, it follows that Holant$([1, a, a, 1])$ is #P-hard. □

We can even restrict the graph $\Gamma$ to be planar, at the expense of introducing a non-zero field.

**Lemma 6.3.** Pl-Holant$([1, ab, ab^2, b^3])$ is #P-hard, for all $a \in \mathbb{Q} \cap (1, \infty)$ and $b \in \mathbb{Q}^+ \setminus \{0, 1\}$.

The proof follows a similar line to the previous one, using the following dichotomy result for planar holants.

**Lemma 6.4.** Kowalczyk and Cai [17, Theorem 4.4]. The problem Pl-Holant$([y_0, y_1, y_2] \mid [1, 0, 0, 1])$ is #P-hard for all $y_0, y_1, y_2 \in \mathbb{C}$, except in the four cases listed in the statement of Lemma 6.2, together with a fifth, namely (5) $y_0^3 = y_2^3$. In these five cases, the problem is in P.

Note that in the statement of this result in [17, Theorem 4.4], Hol$(a, b)$ is shorthand for Holant$([a, 1, b] \mid [1, 0, 0, 1])$.

*Proof of Lemma 6.3.* Define $\alpha_1 = e^{-\lambda + i\theta}$, $\beta_1 = e^{-\lambda - i\theta}$, $\alpha_2 = e^{\lambda - i\theta}$ and $\beta_2 = e^{\lambda + i\theta}$; then define

$$\begin{aligned} x_0 &= \alpha_1^3 + \beta_1^3 = 2e^{-3\lambda}\cos 3\theta \\ x_1 &= \alpha_1^2\alpha_2 + \beta_1^2\beta_2 = 2e^{-\lambda}\cos\theta \\ x_2 &= \alpha_1\alpha_2^2 + \beta_1\beta_2^2 = 2e^{\lambda}\cos\theta \\ x_3 &= \alpha_2^3 + \beta_2^3 = 2e^{3\lambda}\cos 3\theta. \end{aligned}$$

By setting $0 < \theta < \frac{\pi}{6}$ and $\lambda \in \mathbb{R}$ appropriately, the signature vector $[x_0, x_1, x_2, x_3]$ can be made to match $[1, ab, ab^2, b^3]$ up to a scaling factor. Again define

$$T = \begin{bmatrix} \alpha_1 & \beta_1 \\ \alpha_2 & \beta_2 \end{bmatrix} = \begin{bmatrix} e^{-\lambda + i\theta} & e^{-\lambda - i\theta} \\ e^{\lambda - i\theta} & e^{\lambda + i\theta} \end{bmatrix},$$

https://doi.org/10.1017/S0963548321000080 Published online by Cambridge University Press

and

$$[y_0, y_1, y_2] = [1, 0, 1]T^{\otimes 2} = [e^{-2\lambda+2i\theta} + e^{2\lambda-2i\theta},\ e^{2\lambda} + e^{-2\lambda},\ e^{2\lambda+2i\theta} + e^{-2\lambda-2i\theta}].$$

To establish #P-hardness for planar graphs, we just need to verify that none of the five conditions in Lemma 6.4 hold. To this end, observe that

$$y_0 y_2 = 2\cosh 4\lambda + 2\cos 4\theta \quad \text{and} \quad y_1^2 = 2\cosh 4\lambda + 2.$$

Noting that $\cosh 4\lambda \geq 1$ and $0 < \theta < \frac{\pi}{6}$, we find that $0 < y_0 y_2 < y_1^2$. Also, $y_2$ is the complex conjugate of $y_0$ and so $|y_0| = |y_2| < |y_1|$. These facts rule out the four polynomial-time cases (1)–(4) listed in the statement of Lemma 6.4. The fifth condition is ruled out by $\arg y_0 = -\arg y_2 \neq 0$ and $-\frac{\pi}{3} < \arg y_0, \arg y_2 < \frac{\pi}{3}$. It follows that Pl-Holant($[1, ab, ab^2, b^3]$) is #P-hard. □

Note that the final step of the proof crucially uses the fact that $\lambda \neq 0$, which is equivalent to $b \neq 1$. Indeed, the partition function can be evaluated in polynomial time on planar graphs in the zero field case, using the algorithm of Fisher and Kasteleyn mentioned earlier. Lemma 6.3 can be interpreted as asserting that exactly computing the partition function of the Ising model with a constant non-zero external field is hard even when restricted to line graphs of cubic planar graphs. An example of such a graph is the kagome lattice with toroidal boundary conditions, as it is the line graph of the hexagonal lattice with the same boundary conditions.

## Acknowledgements

MJ thanks Heng Guo, Alan Sokal and Eric Vigoda for helpful remarks.

## References

[1] Arbib, C. (1988) A polynomial characterization of some graph partitioning problems. *Inform. Process. Lett.* **26**(5) 223–230.

[2] Bencs, F., Csikvári, P. and Regts, G. (2020) Some applications of Wagner's weighted subgraph counting polynomial. arXiv:2012.00806.

[3] Bodlaender, H. L. and Jansen, K. (2000) On the complexity of the maximum cut problem. *Nordic J. Comput.* **7**(1) 14–31.

[4] Cai, J.-Y. and Chen, X. (2017) *Complexity Dichotomies for Counting Problems. Vol. 1: Boolean Domain*. Cambridge University Press, Cambridge.

[5] Cai, J.-Y., Huang, S. and Lu, P. (2012) From Holant to #CSP and back: dichotomy for Holant$^c$ problems. *Algorithmica* **64**(3) 511–533.

[6] Chen, Z., Liu, K. and Vigoda, E. (2020) Rapid mixing of Glauber dynamics up to uniqueness via contraction. *CoRR*, abs/2004.09083.

[7] Fisher, M. E. (1966) On the dimer solution of planar Ising models. *J. Math. Phys.* **7**(10) 1776–1781.

[8] Friedli, S. and Velenik, Y. (2018) *Statistical Mechanics of Lattice Systems*. Cambridge University Press, Cambridge. A concrete mathematical introduction.

[9] Garey, M. R., Johnson, D. S. and Stockmeyer, L. (1976) Some simplified NP-complete graph problems. *Theoret. Comput. Sci.* **1**(3) 237–267.

[10] Goldberg, L. A., Martin, R. and Paterson, M. (2004) Random sampling of 3-colorings in $\mathbb{Z}^2$. *Random Struct. Algorithms* **24**(3) 279–302.

[11] Guruswami, V. (1999) Maximum cut on line and total graphs. *Discrete Appl. Math.* **92**(2-3) 217–221.

[12] Huang, L., Lu, P. and Zhang, C. (2016) Canonical paths for MCMC: from art to science. In *Proceedings of the Twenty-Seventh Annual ACM-SIAM Symposium on Discrete Algorithms, SODA 2016, Arlington, VA, USA, January 10–12, 2016*, (R. Krauthgamer, ed), SIAM, pp. 514–527.

[13] Jerrum, M. (2003) *Counting, Sampling and Integrating: Algorithms and Complexity*. Lectures in Mathematics ETH Zürich. Birkhäuser Verlag, Basel, 2003.

[14] Jerrum, M. and Sinclair, A. (1993) Polynomial-time approximation algorithms for the Ising model. *SIAM J. Comput.* **22**(5) 1087–1116.

[15] Jurčišinová, E. and Jurčišin, M. (2020) Ground states, residual entropies, and specific heat capacity properties of frustrated Ising system on pyrochlore lattice in effective field theory cluster approximations. *Phys. A* **554** 124671.

[16] Kasteleyn, P. W. (1963) Dimer statistics and phase transitions. *J. Mathematical Phys.* **4** 287–293.

https://doi.org/10.1017/S0963548321000080 Published online by Cambridge University Press

[17] Kowalczyk, M. and Cai, J.-Y. (2016) Holant problems for 3-regular graphs with complex edge functions. *Theory Comput. Syst.* **59**(1) 133–158.
[18] Luby, M. and Vigoda, E. (1999) Fast convergence of the Glauber dynamics for sampling independent sets. *Random Struct. Algorithms* **15**(3–4) 229–241. *Statistical Physics Methods in Discrete Probability, Combinatorics, and Theoretical Computer Science*. Princeton, NJ, 1997.
[19] McQuillan, C. (2013) Approximating Holant problems by winding. *CoRR*, abs/1301.2880.
[20] Motwani, R. and Raghavan, P. (1995) *Randomized algorithms*. Cambridge University Press, Cambridge.
[21] Sinclair, A. (1992) Improved bounds for mixing rates of Markov chains and multicommodity flow. *Combin. Probab. Comput.* **1**(4) 351–370.
[22] Syôzi, I. (1951) Statistics of Kagomé lattice. *Progr. Theoret. Phys.* **6** 306–308.
[23] Štefankovič, D., Vempala, S. and Vigoda, E. (2009) Adaptive simulated annealing: a near-optimal connection between sampling and counting. *J. ACM* **56**(3) Art. 18, 36.
[24] Wainwright, M. J. and Jordan, M. I. (2008) Graphical models, exponential families, and variational inference. *Found. Trends Mach. Learn.* **1**(1-2) 1–305.

**Cite this article:** Dyer M, Heinrich M, Jerrum M and Müller H (2021). Polynomial-time approximation algorithms for the antiferromagnetic Ising model on line graphs. *Combinatorics, Probability and Computing* **30**, 905–921. https://doi.org/10.1017/S0963548321000080

https://doi.org/10.1017/S0963548321000080 Published online by Cambridge University Press