Approximating Pairwise Correlations in the Ising Model

Leslie Ann Goldberg* Mark Jerrum†

12 April 2019

Abstract

In the Ising model, we consider the problem of estimating the covariance of the spins at two specified vertices. In the ferromagnetic case, it is easy to obtain an additive approximation to this covariance by repeatedly sampling from the relevant Gibbs distribution. However, we desire a multiplicative approximation, and it is not clear how to achieve this by sampling, given that the covariance can be exponentially small. Our main contribution is a fully polynomial time randomised approximation scheme (FPRAS) for the covariance in the ferromagnetic case. We also show that that the restriction to the ferromagnetic case is essential — there is no FPRAS for multiplicatively estimating the covariance of an antiferromagnetic Ising model unless RP = #P. In fact, we show that even determining the sign of the covariance is #P-hard in the antiferromagnetic case.

1 Introduction

Let $G = (V, E)$ be a graph and let $\beta : E \to \mathbb{Q}$ be an edge weighting of G. A *configuration* of the Ising model is an assignment $\sigma : V \to \{-1, +1\}$ of *spins* from $\{-1, +1\}$ to the vertices of G. The weight of a configuration is

$$\mathrm{wt}^{\mathrm{Ising}}_{G,\beta}(\sigma) = \prod_{\substack{e=\{u,v\}\in E:\\ \sigma(u)=\sigma(v)}} \beta(e).$$

The Ising partition function is $Z^{\mathrm{Ising}}_{G,\beta} = \sum_{\sigma:V\to\{-1,+1\}} \mathrm{wt}^{\mathrm{Ising}}_{G,\beta}(\sigma)$. It is the normalising factor that makes the weights of configurations into a probability distribution, $\pi^{\mathrm{Ising}}_{G,\beta}(\cdot)$, which is called the *Gibbs distribution* of the Ising model. Thus, the probability of observing configuration σ is $\pi^{\mathrm{Ising}}_{G,\beta}(\sigma) = \mathrm{wt}^{\mathrm{Ising}}_{G,\beta}(\sigma)/Z^{\mathrm{Ising}}_{G,\beta}$.

We say that an edge weighting is *ferromagnetic* if $\beta(e) > 1$ for all $e \in E$. The corresponding Ising model is also said to be ferromagnetic in this case. We say that an edge weighting and the corresponding Ising model are *antiferromagnetic* if $0 < \beta(e) < 1$ for all $e \in E$.

Given specified vertices s, t, we are interested in computing $\mathbb{E}_{\pi^{\mathrm{Ising}}_{G,\beta}}[\sigma(s)\sigma(t)]$. Since

$$\mathbb{E}_{\pi^{\mathrm{Ising}}_{G,\beta}}[\sigma(s)] = \mathbb{E}_{\pi^{\mathrm{Ising}}_{G,\beta}}[\sigma(t)] = 0,$$

*University of Oxford, UK. The research leading to these results has received funding from the European Research Council under the European Union's Seventh Framework Programme (FP7/2007–2013) ERC grant agreement no. 334828. The paper reflects only the authors' views and not the views of the ERC or the European Commission. The European Union is not liable for any use that may be made of the information contained therein.

†Queen Mary, University of London, UK. Supported by EPSRC grant EP/N004221/1.

this quantity is equal to the covariance of the spins at s and t.

Interestingly, none of the existing work on computational aspects of the ferromagnetic Ising model provides an efficient algorithm for estimating this covariance. Jerrum and Sinclair [7] presented a polynomial-time algorithm for approximating the partition function $Z^{\text{Ising}}_{G,\beta}$ within specified relative error, and Randall and Wilson [10] observed that this algorithm could be used to produce samples from the Gibbs distribution. Therefore, by repeated sampling we can easily get an additive approximation to the covariance. Specifically, the covariance may be estimated to additive error ε using $O(\varepsilon^{-2})$ samples.

Our main contribution (Theorem 2) is a polynomial-time algorithm to approximate the covariance within small multiplicative error. This is much more challenging than obtaining an additive approximation since the covariance may be exponentially small in n, as will typically be the case when the system is in the uniqueness regime. The computational problem that we study is the following.

Name. FerroIsingCov.

Instance. A graph $G = (V, E)$ with specified vertices s and t. An edge weighting $\beta : E \to \mathbb{Q}_{>1}$ of G.

Output. $\mathbb{E}_{\pi^{\text{Ising}}_{G,\beta}}[\sigma(s)\sigma(t)]$.

The reason that we restrict the range of the edge weighting β to the rationals (rather than allowing real-valued weights) is to avoid the issue of how to represent real numbers in the input. Each weight $\beta(e)$ satisfies $\beta(e) > 1$. For concreteness, we assume that it is represented in the input by two positive integers $P(e)$ and $Q(e)$ (specified in unary[1] in the input) such that and $\beta(e) = 1+P(e)/Q(e)$. Our main result is that there is a polynomial-time approximation algorithm for FerroIsingCov. In order to state the result precisely, we need to recall a definition from computational complexity. We view a problem, such as FerroIsingCov, as a function $f : \Sigma^* \to \mathbb{Q}$ from problem instances to rational numbers.

Definition 1. A *randomised approximation scheme* for $f : \Sigma^* \to \mathbb{Q}$ is a randomised algorithm that takes as input an instance $x \in \Sigma^*$ (e.g., an encoding of a labelled graph) and an error tolerance $\varepsilon > 0$, and outputs a number $z \in \mathbb{Q}$ (a random variable on the "coin tosses" made by the algorithm) such that, for every instance x,

$$\Pr\left[e^{-\varepsilon} \leq \frac{z}{f(x)} \leq e^{\varepsilon}\right] \geq \frac{3}{4},$$

where, by convention, $0/0 = 1$. The randomised approximation scheme is said to be a *fully polynomial randomised approximation scheme*, or *FPRAS*, if it runs in time bounded by a polynomial in $|x|$ and ε^{-1}. (See Mitzenmacher and Upfal [9, Definition 10.2].)

The slight modification of the more familiar definition is to ensure that functions f taking negative values are dealt with correctly.

Theorem 2. *There is an FPRAS for* FerroIsingCov*.*

The restriction to the ferromagnetic case in Theorem 2 is crucial. Consider the unrestricted version of the problem.

[1] The assumption that $P(e)$ and $Q(e)$ are specified in unary is a technical simplification, but is not essential: see Remark 20.

Name. IsingCov.

Instance. A graph $G = (V, E)$ with specified vertices s and t. An edge weighting $\beta : E \to \mathbb{Q}_{>0}$ of G.

Output. $\mathbb{E}_{\pi^{\text{Ising}}_{G,\beta}}[\sigma(s)\sigma(t)]$.

We show the following.

Theorem 3. *There is no FPRAS for* IsingCov *unless* RP = #P.

Theorem 3 holds even in the restricted setting where, for some fixed $b \in (0, 1)$, the edge weighting β is the constant function which assigns every edge weight $\beta(e) = b$. Theorem 23 in Section 5 shows that even showing whether $\mathbb{E}_{\pi^{\text{Ising}}_{G,\beta}}[\sigma(s)\sigma(t)]$ is at least 0 or at most 0 is #P-hard, in this restricted setting. Theorem 3 is an immediate consequence of Theorem 23.

In Section 4 we prove Theorem 2 by providing an FPRAS for FerroIsingCov. Our FPRAS is based on Markov-chain simulation. Like the known MCMC algorithms for approximating the partition function of the Ising model, it is explained in terms of a related model called the even subgraphs model. Our Markov chain is a modification of a process known as the worm process.

2 The even subgraphs model and the worm process

An instance of the even subgraphs model is a graph $G = (V, E)$ with an edge weighting $\lambda : E \to \mathbb{Q}_{>0}$. A configuration of the model is a subset $A \subseteq E$ such that every vertex in the subgraph (V, A) has even degree.

Definition 4. We use the notation $\lambda(A)$ to denote the product $\lambda(A) = \prod_{e \in A} \lambda(e)$ of edge-weights of the edges in A.

It is convenient to generalise the even subgraphs model to allow a small set $S \subseteq V$ of "exceptional vertices" of odd degree. The configuration space of the (extended) even subgraphs model is given by

$$\Omega_S = \{A \subseteq E : \ \deg(v) \text{ is odd in } (V, A) \text{ iff } v \in S\},$$

and the corresponding partition function is given by

$$Z_S(G; \lambda) = \sum_{A \in \Omega_S} \lambda(A).$$

Despite appearances, there is a close connection between the Ising model and the even subgraphs model. Suppose that, for every $e \in E$, $\lambda(e) = (\beta(e) - 1)/(\beta(e) + 1)$. Van der Waerden [14] showed that there is an easily-computable scaling factor C such that $Z^{\text{Ising}}_{G,\beta} = C\, Z_{\emptyset}(G; \lambda)$. Note that a ferromagnetic Ising model corresponds to an even-subgraphs model in which $0 < \lambda(e) < 1$ for all $e \in E$. We do not use precisely van der Waerden's identity, but we do use a closely related one which is captured by the following lemma, which can be found, e.g., in [1, Lemma 2.1].

Lemma 5. *Let $G = (V, E)$ be a graph with edge weighting β. Let λ be the edge weighting of G defined by $\lambda(e) = (\beta(e) - 1)/(\beta(e) + 1)$. Then, for any set $S \subseteq V$,*

$$\mathbb{E}_{\pi^{\mathrm{Ising}}_{G,\beta}}\left[\prod_{v\in S}\sigma(v)\right] = \frac{Z_S(G;\lambda)}{Z_\emptyset(G;\lambda)}. \tag{1}$$

Proof. Observe that

$$\mathrm{wt}^{\mathrm{Ising}}_{G,\beta}(\sigma) = \prod_{e=\{u,v\}\in E}\frac{\beta(e)+1}{2}\left[1 + \frac{\beta(e)-1}{\beta(e)+1}\sigma(u)\sigma(v)\right].$$

since the factor corresponding to $e = \{u, v\}$ contributes $\beta(e)$ if $\sigma(u) = \sigma(v)$ and contributes 1 otherwise. Thus, setting λ as in the statement of the lemma,

$$\begin{aligned}
Z^{\mathrm{Ising}}_{G,\beta} = \sum_\sigma \mathrm{wt}^{\mathrm{Ising}}_{G,\beta}(\sigma) &= \prod_{e\in E}\frac{\beta(e)+1}{2}\sum_\sigma \prod_{e=\{u,v\}\in E}\left[1 + \lambda(e)\sigma(u)\sigma(v)\right] \\
&= \prod_{e\in E}\frac{\beta(e)+1}{2}\sum_\sigma\sum_{A\subseteq E}\prod_{e=\{u,v\}\in A}\lambda(e)\sigma(u)\sigma(v) \\
&= 2^n\prod_{e\in E}\frac{\beta(e)+1}{2}\sum_{A\in\Omega_\emptyset}\prod_{e\in A}\lambda(e) \\
&= c\,Z_\emptyset(G;\lambda),
\end{aligned} \tag{2}$$

where $c = 2^n \prod_{e\in E}[(\beta(e) + 1)/2]$, and σ ranges over configurations $V \to \{-1, +1\}$. The third equality is explained as follows. If (V, A) contains an odd degree vertex u, then $\sigma(u)$ appears to an odd power in the term corresponding to A; the term is then annihilated by the summation over σ.

Arguing similarly,

$$\begin{aligned}
\sum_\sigma \mathrm{wt}^{\mathrm{Ising}}_{G,\beta}(\sigma)\prod_{w\in S}\sigma(w) &= \prod_{e\in E}\frac{\beta(e)+1}{2}\sum_\sigma\prod_{e=\{u,v\}\in E}\left[1 + \lambda(e)\sigma(u)\sigma(v)\right]\prod_{w\in S}\sigma(w) \\
&= \prod_{e\in E}\frac{\beta(e)+1}{2}\sum_\sigma\sum_{A\subseteq E}\prod_{e=\{u,v\}\in A}\lambda(e)\sigma(u)\sigma(v)\prod_{w\in S}\sigma(w) \\
&= 2^n\prod_{e\in E}\frac{\beta(e)+1}{2}\sum_{A\in\Omega_S}\prod_{e\in A}\lambda(e) \\
&= c\,Z_S(G;\lambda).
\end{aligned} \tag{3}$$

The identity in the statement of the lemma is obtained by dividing (3) by (2). □

We remark that the interesting case of the lemma is when $|S|$ is even. If $|S|$ is odd, then both sides of identity (1) are zero. Lemma 5 provides a way to approximate the correlation $\mathbb{E}[\sigma(s)\sigma(t)]$ in the Ising model by estimating the ratio of two partition functions in the even-subgraphs model. At first sight it might seem that existing Markov chain Monte Carlo approaches might be up to this task. One such Markov chain is the so-called "worm process". The state space of this chain is defined as follows.

Definition 6. Let $\Omega = \bigcup_{S\subseteq V:|S|\leq 2} \Omega_S = \bigcup_{S\subseteq V:|S|\in\{0,2\}} \Omega_S$.

The "worm process" is a Markov chain on Ω whose stationary distribution assigns probability proportional to $\lambda(A) = \prod_{e\in A} \lambda(e)$ to each configuration $A \in \Omega$. A transition of the worm process simply flips a single edge of the graph from being in the configuration A to being out of A or vice versa. Thus, as transitions occur, the two odd degree vertices move in random paths along the edges of G, occasionally becoming adjacent and disappearing.

The worm process is rapidly mixing, as was shown by Collevecchio, Garoni, Hyndman and Tokarev [1, Theorem 1.3]. In principle, to estimate the ratio appearing in the right-hand side of equation (1) with $S = \{s,t\}$, we could just run the worm process and observe the relative time that the process spends in states in $\Omega_{\{s,t\}}$ compared with the time that it spends in states in $\Omega_\emptyset$, However, if the spins at s and t are only weakly correlated, then the ratio $Z_{\{s,t\}}(G;\lambda)/Z_\emptyset(G;\lambda)$ will be small, and the process will spend a small (possibly exponentially small) proportion of time in $\Omega_{\{s,t\}}$.

Following [8], we modify the worm process by artificially weighting configurations so that each subset in the partition $\{\Omega_S : |S| \leq 2\}$ of Ω has roughly equal weight in the stationary distribution. We will give the details of the modified process in Section 3. First, we need to define the Random Cluster model [6] (which, in the special case we consider, is also equivalent to the Ising model) and use the Random Cluster model to prove a lemma (Lemma 10 below), which will help with the analysis of the weighted worm process.

An instance of the Random cluster model is a graph $G = (V,E)$ with an edge weighting $p : E \to \mathbb{Q} \cap (0,1)$. A configuration of this model is a subset $A \subseteq E$. The weight of configuration A is

$$\mathrm{wt}^{\mathrm{RC}}_{G,p}(A) = \prod_{e\in A} p(e) \prod_{e\in E\setminus A} (1-p(e))\, 2^{\kappa(A)},$$

where $\kappa(A)$ is the number of connected components in the graph (V,A). There is an associated partition function $Z^{\mathrm{RC}}_{G;p} = \sum_{A\subseteq E} \mathrm{wt}^{\mathrm{RC}}_{G,p}(A)$, but we are more concerned with the probability distribution on configurations given by $\pi^{\mathrm{RC}}_{G,p} = \mathrm{wt}^{\mathrm{RC}}_{G,p}(A)/Z^{\mathrm{RC}}_{G;p}$ for all $A \subseteq E$. Following Fortuin and Kasteleyn [4], Edwards and Sokal [3] showed that there is a simple coupling between the distributions $\pi^{\mathrm{Ising}}_{G,\beta}$ and $\pi^{\mathrm{RC}}_{G,p}$ given by the following trial.

Definition 7. (Edwards-Sokal Distribution) Given a graph $G = (V,E)$ with an edge weighting $p : E \to \mathbb{Q} \cap (0,1)$, let $\mathcal{D}_{G,p}$ be the following distribution on pairs (A,σ).

1. Select $A \subseteq E$ according to the distribution $\pi^{\mathrm{RC}}_{G,p}$.
2. Independently and uniformly, for each connected component of (V,A), choose a spin from $\{-1,+1\}$ and assign that spin to all vertices in the connected component. Let $\sigma : V \to \{-1,+1\}$ be the resulting spin configuration.

The following lemma shows that the output of the Edwards-Sokal coupling is a sample from $\pi^{\mathrm{Ising}}_{G,\beta}$.

Lemma 8. *(Edwards and Sokal [3]) Let $G = (V,E)$ be a graph with edge weighting $\beta : E \to \mathbb{Q}_{>1}$. Let p be the edge weighting of G defined by $p(e) = 1 - 1/\beta(e)$. Let (A,σ) be drawn from the Edwards-Sokal distribution $\mathcal{D}_{G,p}$. Then the distribution of σ is $\pi^{\mathrm{Ising}}_{G,\beta}$.*

We say that an event $\mathcal{E} \subseteq 2^E$ is monotonically increasing if, for all $A \subset A' \subseteq E$, we have $A \in \mathcal{E}$ implies $A' \in \mathcal{E}$. In the random cluster model as defined here, monotonically increasing events are positively correlated.

Lemma 9. *Suppose that events $\mathcal{E}_1, \mathcal{E}_2 \subseteq 2^E$ are monotonically increasing. Then*

$$\Pr_{\pi^{\mathrm{RC}}_{G,p}}(\mathcal{E}_1 \wedge \mathcal{E}_2) \geq \Pr_{\pi^{\mathrm{RC}}_{G,p}}(\mathcal{E}_1) \Pr_{\pi^{\mathrm{RC}}_{G,p}}(\mathcal{E}_2).$$

Proof. This inequality is stated as Part (b) of Theorem (3.8) of [6], for the situation where $p(e)$ is the same for all edges e. However the proof is essentially the same when $p(e)$ varies with e. The main step, in order to apply the FKG inequality, is to prove the well-known fact that the distribution $\pi^{\mathrm{RC}}_{G,p}$ satisfies the FKG lattice condition, which says that, for any sets $A_1, A_2 \subseteq E$,

$$\Pr_{\pi^{\mathrm{RC}}_{G,p}}(A_1 \cup A_2) \Pr_{\pi^{\mathrm{RC}}_{G,p}}(A_1 \cap A_2) \geq \Pr_{\pi^{\mathrm{RC}}_{G,p}}(A_1) \Pr_{\pi^{\mathrm{RC}}_{G,p}}(A_2).$$

To see this, recall the definition of $\pi^{\mathrm{RC}}_{G,p}$. The denominators cancel, so the FKG lattice condition is equivalent to

$$\mathrm{wt}^{\mathrm{RC}}_{G,p}(A_1 \cup A_2)\mathrm{wt}^{\mathrm{RC}}_{G,p}(A_1 \cap A_2) \geq \mathrm{wt}^{\mathrm{RC}}_{G,p}(A_1)\mathrm{wt}^{\mathrm{RC}}_{G,p}(A_2).$$

Recalling the definition of $\mathrm{wt}^{\mathrm{RC}}_{G,p}$, note that, for any edge e, the quantities $p(e)$ and $1 - p(e)$ occur the same number of times on the left-hand-side and right-hand-side. Thus, the FKG lattice condition is equivalent to $2^{\kappa(A_1 \cup A_2)} 2^{\kappa(A_1 \cap A_2)} \geq 2^{\kappa(A_1)} 2^{\kappa(A_2)}$. The proof in [6] now applies without any further changes. □

The following lemma will be used in the analysis of the weighted worm process.

Lemma 10. *Let $G = (V, E)$ be a graph with edge weighting $\lambda : E \to \mathbb{Q} \cap (0, 1)$. Suppose $S, S' \subseteq V$ are subsets of V of even cardinality, and assume that it is not the case that $\emptyset \subset S' \subset S$. Then*

$$\frac{Z_\emptyset(G;\lambda)}{Z_S(G;\lambda)} \leq \frac{Z_\emptyset(G;\lambda)}{Z_{S'}(G;\lambda)} \times \frac{Z_\emptyset(G;\lambda)}{Z_{S \oplus S'}(G;\lambda)}.$$

Proof. Fix $G = (V, E)$, λ, S and S' as in the statement of the lemma. Let β be the edge weighting of G defined by $\beta(e) = (1 + \lambda(e))/(1 - \lambda(e))$. Taking reciprocals, the inequality in the statement of the lemma is equivalent by Lemma 5 to

$$\mathbb{E}_{\pi^{\mathrm{Ising}}_{G,\beta}}\left[\prod_{v \in S} \sigma(v)\right] \geq \mathbb{E}_{\pi^{\mathrm{Ising}}_{G,\beta}}\left[\prod_{v \in S'} \sigma(v)\right] \times \mathbb{E}_{\pi^{\mathrm{Ising}}_{G,\beta}}\left[\prod_{v \in S \oplus S'} \sigma(v)\right]. \tag{4}$$

Now let p be the edge weighting defined by $p(e) = 1 - 1/\beta(e)$. Let (A, σ) be drawn from the Edwards-Sokal distribution $\mathcal{D}_{G,p}$. For any subset T of V, let "T is connected" be a shorthand for the event "T is contained within a single connected component of (V, A)". Let Y_T be the random variable $Y_T = \prod_{v \in T} \sigma(v)$. Then

$$\begin{aligned}\mathbb{E}_{\mathcal{D}_{G,p}}[Y_T] = {} & \Pr_{\mathcal{D}_{G,p}}(T \text{ is connected})\, \mathbb{E}_{\mathcal{D}_{G,p}}[Y_T \mid T \text{ is connected}] \\ & + \Pr_{\mathcal{D}_{G,p}}(\neg\, T \text{ is connected})\, \mathbb{E}_{\mathcal{D}_{G,p}}[Y_T \mid \neg\, T \text{ is connected}].\end{aligned}$$

The definition of $\mathcal{D}_{G,p}$ (Definition 7) ensures that, for any set T with even cardinality, $\mathbb{E}_{\mathcal{D}_{G,p}}[Y_T \mid T \text{ is connected}] = 1$ and $\mathbb{E}_{\mathcal{D}_{G,p}}[Y_T \mid \neg\, T \text{ is connected}] = 0$. Hence, $\mathbb{E}_{\mathcal{D}_{G,p}}[Y_T] = \Pr_{\mathcal{D}_{G,p}}(T \text{ is connected})$. Using Lemma 8,

$$\mathbb{E}_{\pi^{\mathrm{Ising}}_{G,\beta}}[Y_T] = \mathbb{E}_{\mathcal{D}_{G,p}}[Y_T] = \Pr_{\mathcal{D}_{G,p}}(T \text{ is connected}) = \Pr_{\pi^{\mathrm{RC}}_{G,p}}(T \text{ is connected}).$$

Plugging this into (4) with $T = S$, $T = S'$ and $T = S \oplus S'$, we find that (4) is equivalent to the following inequality.

$$\Pr_{\pi^{\mathrm{RC}}_{G,p}}(S \text{ is connected}) \geq \Pr_{\pi^{\mathrm{RC}}_{G,p}}(S' \text{ is connected}) \times \Pr_{\pi^{\mathrm{RC}}_{G,p}}(S \oplus S' \text{ is connected}). \tag{5}$$

By considering the possible intersections of S and S', recalling from the statement of the lemma that it is not the case that $\emptyset \subset S' \subset S$, it is easy to see that

$$\Pr_{\pi^{\mathrm{RC}}_{G,p}}(S \text{ is connected}) \geq \Pr_{\pi^{\mathrm{RC}}_{G,p}}(S' \text{ is connected } \wedge \ S \oplus S' \text{ is connected}). \tag{6}$$

Now observe that "S' is connected" and "$S \oplus S'$ is connected" are both monotonically increasing events, and hence (5) follows from (6) by Lemma 9. $\square$

3 The weighted worm process

Consider a graph $G = (V, E)$ with an edge weighting $\lambda : E \to \mathbb{Q} \cap (0, 1)$.

Definition 11. A *subset weighting* of G is a function w that assigns a weight $w_S \in \mathbb{Q}_{>0}$ to each subset S of $V(G)$ with $|S| \in \{0, 2\}$. We refer to the pair (λ, w) as a *weighting* of G. Given a subset $A \subseteq E(G)$, there is a unique $S(A) \subseteq V(G)$ such that $A \in \Omega_{S(A)}$. If $|S(A)| \leq 2$ we define $\Lambda(A) = \lambda(A) w_{S(A)}$. The partition function that we study is

$$\widehat{Z}_S(G; \lambda, w) = \sum_{A \in \Omega_S} \Lambda(A) = \sum_{A \in \Omega_S} \lambda(A) w_S = w_S Z_S(G; \lambda).$$

We also define $\widehat{Z}(G; \lambda, w) = \sum_{S \subseteq V; |S| \leq 2} \widehat{Z}_S(G; \lambda, w)$.

Later, we shall need to extend the above definition to subsets $S \subseteq V(G)$ with $|S| \leq 4$ in the obvious way.

Recall from Definition 6 that $\Omega = \bigcup_{S \subseteq V : |S| \leq 2} \Omega_S$. The weighted worm process is a Markov chain with state space Ω. The transitions of the process are given in Figure 1. It is easy to see from the definition of the transitions that the weighted worm process is ergodic and time-reversible and that the stationary probability of a configuration $A \in \Omega$ is $\pi(A) = \Lambda(A)/\widehat{Z}(G; \lambda, w)$.

Given a subset S of $V(G)$ with $|S| \leq 2$, the probability of Ω_S in the stationary distribution of the weighted worm process is

$$\sum_{A \in \Omega_S} \pi(A) = \frac{w_S Z_S(G; \lambda)}{\widehat{Z}(G; \lambda, w)} = \frac{w_S Z_S(G; \lambda)}{\sum_{S'} w_{S'} Z_{S'}(G; \lambda)},$$

where the sum is over all subsets $S' \subseteq V(G)$ with $|S'| \leq 2$.

Thus, we will be most interested in the weighted worm process when the weighting satisfies $w_S = Z_\emptyset(G; \lambda)/Z_S(G; \lambda)$ so that all subsets S have equal weight. We show in Section 3.2 how to "learn" such a weighting by running the process multiple times. First, we consider the mixing rate of the process itself.

(* One transition from state $A \in \Omega$ *)
Choose the type of transition T uniformly at random from {"self-loop", "move"}
if $T =$ "self-loop" **then**
  the next state is A
else
  Choose an edge $e \in E$ uniformly at random
  if $A \oplus \{e\} \in \Omega$ **then**
    $A' \leftarrow A \oplus \{e\}$
  else
    $A' \leftarrow A$
  end if
  With probability $\min\{\Lambda(A')/\Lambda(A), 1\}$ the next state is A', otherwise A
end if

Figure 1: One transition of the weighted worm process for graph $G = (V, E)$ with weighting (λ, w), starting at state $A \in \Omega$, where $\Lambda(A) = \lambda(A) w_{S(A)}$.

3.1 Rapid mixing of the weighted worm process

In broad outline, the proof of rapid mixing follows existing work [7, 1], but is complicated by the need to deal with the subset weightings.

We use $\mathcal{W}(G)$ to denote the set of weightings (λ, w) where $\lambda : E \to \mathbb{Q} \cap (0, 1)$ is an edge weighting of G and w is a subset weighting of G satisfying

$$\begin{aligned} w_S &= 1, \quad \text{if } |S| = 0, \\ w_S &= 0, \quad \text{if } |S| = 1, \text{ and} \\ \frac{1}{2} \leq \frac{\widehat{Z}_S(G; \lambda, w)}{\widehat{Z}_\emptyset(G; \lambda, w)} &\leq 2, \quad \text{if } |S| = 2. \end{aligned} \tag{7}$$

The purpose of this section is to prove that the weighted worm process is rapidly mixing if $(\lambda, w) \in \mathcal{W}(G)$ (see Lemma 15 below).

In order to prove rapid mixing, given a weighting (λ, w) of G it will be useful to extend the subset weighting w by defining $w_S = Z_\emptyset(G; \lambda)/Z_S(G; \lambda)$ for every S with $|S| = 4$. The extended weighting will be used in the proof, but not in the Markov chain. The following lemma will be used in the proof of rapid mixing.

Lemma 12. *If $(\lambda, w) \in \mathcal{W}(G)$ then, for every subset S of $V(G)$ with $|S| \in \{0, 4\}$ we have $w_S = Z_\emptyset(G; \lambda)/Z_S(G; \lambda)$. For every size-2 subset S of $V(G)$ we have*

$$\frac{Z_\emptyset(G; \lambda)}{2 Z_S(G; \lambda)} \leq w_S \leq \frac{2 Z_\emptyset(G; \lambda)}{Z_S(G; \lambda)}.$$

Proof. The lemma follows trivially from the definition of w_S if $|S| = 0$ or $|S| = 4$, so suppose that $|S| = 2$. From (7) and the definitions of $Z_S(G; \lambda)$ and $\widehat{Z}_S(G; \lambda, w)$ we have

$$\frac{1}{2} \leq \frac{w_S Z_S(G; \lambda)}{Z_\emptyset(G; \lambda)} \leq 2,$$

as required. □

In order to bound the mixing time of the weighted worm process we use the canonical path method or, more precisely, a well-known generalisation of the method that replaces paths by flows. We briefly describe the method, using notation that is slightly more general than that of the weighted worm process. Consider a Markov chain $\mathcal{M}$ on a state space Ω^* with transition matrix P and stationary distribution π^*. A *path* from a state $I \in \Omega^*$ to a state $F \in \Omega^*$ is a sequence $I = T_0, \ldots, T_k = F$ of states, all of which are distinct except possibly I and F, such that, for each $i \in \{0, \ldots, k-1\}$, $P_{T_i,T_{i+1}} > 0$. A *flow* $f_{I,F}$ is a distribution whose support is the set of paths from I to F which is normalised so that $\sum_p f_{I,F}(p) = \pi^*(I)\pi^*(F)$. Typically, when we refer to a flow $f_{I,F}$, we refer to I as the "initial state" and to F as the "final state". The collection of all flows is $\mathcal{F} = \{f_{I,F} : I, F \in \Omega^*\}$. The *congestion* of this collection of flows is

$$\varrho(\mathcal{F}) = \max_{(T,T')} \left\{ \frac{1}{\pi(T)P(T,T')} \sum_{I,F\in\Omega^*} \sum_{p=I,\ldots,T,T',\ldots,F} f_{I,F}(p)\,|p| \right\},$$

where the maximisation is over all transitions (T,T') with $P(T,T') > 0$, the second sum is over all paths p from I to F that use transition (T,T'), and $|p|$ denotes the length of path p.

The mixing time $t_{\mathrm{mix},T}(\delta)$ of $\mathcal{M}$, when starting from state T, is defined to be the minimum time t such that the total variation distance between the t-step distribution $P^t(T,\cdot)$ and the stationary distribution π^* of $\mathcal{M}$ is at most δ. The existence of a collection of flows with small congestion implies rapid mixing. The following lemma is due to Sinclair [12], building on work of Diaconis and Stroock [2]. The explicit statement that we use is taken from [8, Lemma 2.2]

Lemma 13. *Let $\mathcal{M}$ be an ergodic time-reversible Markov chain with state space Ω^* and stationary distribution π^* whose self-loop probabilities satisfy $P(T,T) \geq 1/2$ for all states T. Suppose that $\mathcal{M}$ supports a collection $\mathcal{F}$ of flows. Given any state $T_0 \in \Omega^*$,*

$$t_{\mathrm{mix},T_0}(\delta) \leq \varrho(\mathcal{F}) \left(\ln\left(\frac{1}{\pi^*(T_0)}\right) + \ln\left(\frac{1}{\delta}\right)\right).$$

A standard method for defining a collection of flows is to partition the state space Ω^* into two parts Ω_1^* and Ω_2^*, define canonical paths from every state $I \in \Omega_1^*$ to every state $F \in \Omega_2^*$, and then use an idea similar to Valiant's randomised routing [13] to obtain a collection of flows. Thus, for each pair or initial and final states $(I,F) \in \Omega_1^* \times \Omega_2^*$ we specify a path $\gamma(I,F)$ from I to F. The collection of all such canonical paths is $\Gamma = \{\gamma(I,F) : (I,F) \in \Omega_1^* \times \Omega_2^*\}$. For each possible transition (T,T') of the Markov chain, denote by

$$\mathrm{cp}(T,T') = \big\{(I,F) \in \Omega_1^* \times \Omega_2^* : \gamma(I,F) \text{ uses the transition } (T,T')\big\}$$

the set of canonical paths using (T,T'). The *congestion* of Γ is then given by

$$\varrho(\Gamma) = \max_{(T,T')} \left\{ \frac{1}{\pi^*(T)P(T,T')} \sum_{(I,F)\in\mathrm{cp}(T,T')} \pi^*(I)\pi^*(F)\,|\gamma(I,F)| \right\}. \tag{8}$$

The next step is to use the canonical paths in Γ to induce a collection $\mathcal{F}$ of flows, via randomised routing: If I and F are in Ω_1^* then the flow $f_{I,F}$ is constructed by choosing intermediate states $T \in \Omega_2^*$ and routing flow via the path $\gamma(I,T)$ followed by the reversal of the path $\gamma(F,T)$. Similarly, flow from Ω_2^* to Ω_2^* is routed via a random intermediate state in Ω_1^*. The following lemma shows that if the congestion $\varrho(\Gamma)$ is low then the resulting collection $\mathcal{F}$ also has low congestion. The lemma is a direct translation of Lemma 4.4 of [8] into the more general language of this section. A similar lemma was used earlier by Schweinsberg [11].

Lemma 14. *Given a partition* $\{\Omega_1^*, \Omega_2^*\}$ *of the state space* Ω^* *of a time-reversible Markov chain, and a collection* Γ *of canonical paths from* Ω_1^* *to* Ω_2^* *with congestion* $\varrho(\Gamma)$*, there exists a collection of flows* $\mathcal{F}$ *with congestion*

$$\varrho(\mathcal{F}) \leq \left(2 + 4\left(\frac{\pi^*(\Omega_1^*)}{\pi^*(\Omega_2^*)} + \frac{\pi^*(\Omega_2^*)}{\pi^*(\Omega_1^*)}\right)\right)\varrho(\Gamma).$$

A bound on the mixing time of the Markov chain $\mathcal{M}$ can be derived by constructing low-congestion canonical paths from Ω_1^* to Ω_2^*, using Lemma 14 to derive a collection of flows with low congestion, and then applying Lemma 13. We next apply these methods to bound the mixing time of the weighted worm process.

Lemma 15. *Suppose that* $G = (V, E)$ *is a connected graph with* n *vertices and* m *edges and* $(\lambda, w) \in \mathcal{W}(G)$*. Let* $\lambda_{\min} = \min_{e \in E} \lambda(e)$*. Then the weighted worm process, started in the empty configuration on* G*, has mixing time* $t_{\mathrm{mix},\emptyset}(\delta) = O(\lambda_{\min}^{-2} n^4 m^2)\left(O(m) + \ln\left(\frac{1}{\delta}\right)\right)$*.*

Proof. As we observed earlier, the weighted worm process is a time-reversible Markov chain with state space $\Omega = \bigcup_{S \subseteq V : |S| \leq 2} \Omega_S$. Our goal will be to apply Lemma 14. To this end, let $\overline{\Omega}_\emptyset = \Omega \setminus \Omega_\emptyset = \bigcup_{S \subseteq V : |S| = 2} \Omega_S$. We will define a collection Γ of canonical paths from $\overline{\Omega}_\emptyset$ to $\Omega_\emptyset$. We will bound the congestion $\varrho(\Gamma)$ and use Lemma 14 and Lemma 13 to bound the mixing time.

We start by constructing a canonical path from any configuration $I \in \overline{\Omega}_\emptyset$ to any configuration $F \in \Omega_\emptyset$. Let a and b be the two odd-degree vertices in I. The vertices of the graph $(V, I \oplus F)$ all have even degree, except for a and b. Choose a canonical partition of $I \oplus F$ into a path Π from a to b, and a number of cycles $C_1, C_2, \ldots, C_k$; also choose a distinguished end vertex for Π and a distinguished vertex and orientation for each cycle. To *unwind* a path or cycle, start at the distinguished vertex and travel along the path or around the oriented cycle flipping all edges along the way. The act of *flipping* changes the status of an edge from absent to present or vice versa. The canonical path $\gamma(I, F)$ is obtained by unwinding first the path Π and then the cycles $C_1, C_2, \ldots, C_k$, in order. Note that all of the flips are transitions of the Markov chain corresponding to the weighted worm process.

Fix any transition (T, T') that can be made by the Markov chain, and let

$$\mathrm{cp}(T, T') = \{(I, F) \in \overline{\Omega}_\emptyset \times \Omega_\emptyset \colon \gamma(I, F) \text{ uses the transition } (T, T')\}$$

be the set of canonical paths using this transition. Our goal is to bound the congestion through the transition (T, T').

Denote by $\widehat{\Omega}$ the extended state space $\widehat{\Omega} = \bigcup_{S \subseteq V : |S| \leq 4}$. Consider the function $\eta_{T,T'} : \mathrm{cp}(T, T') \to \widehat{\Omega}$ defined by $\eta_{T,T'}(I, F) = I \oplus F \oplus T$. (Note that the range of $\eta_{T,T'}$ is contained in $\widehat{\Omega}$.) We claim that $\eta_{T,T'}$ is injective. To see this, suppose $(I, F) \in \mathrm{cp}(T, T')$ and let $X = \eta_{T,T'}(I, F)$. Since $I \oplus F = T \oplus X$, we can use the fixed configuration T from the transition and the known value X to recover $I \oplus F$ and hence the path Π and the cycles $C_1, C_2, \ldots, C_k$. The set $T \oplus T'$ contains a single edge e, which tells which of $\Pi, C_1, C_2, \ldots, C_k$ is having its edges flipped by the particular transition (T, T'). Using this information, we can apportion the edges in $\Pi \cup C_1 \cup C_2 \cup \cdots \cup C_k = I \oplus F$ between I and F. (Each edge is either in I or F but not both.) Say that $I \oplus F$ is the disjoint union of I' and F', with $I' \subseteq I$ and $F' \subseteq F$.

We can then recover I and F themselves using the equalities $I = I' \cup (I \cap F) = I' \cup (T \cap X)$[2] and $F = F' \cup (I \cap F) = F' \cup (T \cap X)$. Thus, we have shown that $\eta_{T,T'}$ is injective.

We now proceed to bound the congestion through the transition (T, T'). Note that $\lambda(I)\lambda(F) = \lambda(T)\lambda(X)$, which is exactly what we would need for the analysis of the unweighted case, i.e., when $w_S = 1$ for all S. To analyse the weighted case we need to relate $\Lambda(I)\Lambda(F)$ to $\Lambda(T)\Lambda(X)$. There are three cases, depending on where the transition (T, T') occurs on the canonical path from I to F.

- The transition is the first one of all. Then $T = I$ and $X = F$, and so $\Lambda(I)\Lambda(F) = \Lambda(T)\Lambda(X)$.

- The transition is on the unwinding of the path Π. Then $T \in \Omega_{\{b,c\}}$, where c is a vertex on the path Π, from which it follows that $X \in \Omega_{\{a,c\}}$. We will show

$$\Lambda(I)\Lambda(F) = w_{\{a,b\}}\lambda(I)w_\emptyset\lambda(F) \le 8w_{\{b,c\}}\lambda(T)w_{\{a,c\}}\lambda(X) = 8\Lambda(T)\Lambda(X).$$

  To establish the inequality recall that $\lambda(I)\lambda(F) = \lambda(T)\lambda(X)$ so, cancelling these out, and noting that $w_\emptyset = 1$, it suffices to show $w_{\{a,b\}} \le 8w_{\{b,c\}}w_{\{a,c\}}$. Using Lemma 12, it suffices to show

$$\frac{2Z_\emptyset(G;\lambda)}{Z_{\{a,b\}}(G;\lambda)} \le 8\frac{Z_\emptyset(G;\lambda)}{2Z_{\{b,c\}}(G;\lambda)}\,\frac{Z_\emptyset(G;\lambda)}{2Z_{\{a,c\}}(G;\lambda)},$$

  which follows from Lemma 10 taking $S = \{a, b\}$ and $S' = \{b, c\}$.

- The transition is the first one in the unwinding of a cycle. Then $T \in \Omega_\emptyset$ and and $X \in \Omega_{\{a,b\}}$, and hence

$$\Lambda(I)\Lambda(F) = w_{\{a,b\}}\lambda(I)w_\emptyset\lambda(F) = w_\emptyset\lambda(T)w_{\{a,b\}}\lambda(X) = \Lambda(T)\Lambda(X).$$

- The transition arises during the unwinding of a cycle but is not the first such transition. Then $T \in \Omega_{\{c,d\}}$ for vertices c and d on the cycle, and $X \in \Omega_{\{a,b,c,d\}}$. We will show

$$\Lambda(I)\Lambda(F) = w_{\{a,b\}}\lambda(I)w_\emptyset\lambda(F) \le 8w_{\{c,d\}}\lambda(T)w_{\{a,b,c,d\}}\lambda(X) = 8\Lambda(T)\Lambda(X).$$

  As in the second case, it suffices to show $w_{\{a,b\}} \le 8w_{\{c,d\}}w_{\{a,b,c,d\}}$. Using Lemma 12, it suffices to show

$$\frac{2Z_\emptyset(G;\lambda)}{Z_{\{a,b\}}(G;\lambda)} \le 8\frac{Z_\emptyset(G;\lambda)}{2Z_{\{c,d\}}(G;\lambda)}\,\frac{Z_\emptyset(G;\lambda)}{Z_{\{a,b,c,d\}}(G;\lambda)},$$

  which follows from Lemma 10 (with a factor of 2 to spare) taking $S = \{a, b\}$ and $S' = \{c, d\}$.

Note that in all instances, $\Lambda(I)\Lambda(F) \le 8\Lambda(T)\Lambda(X)$. Given a set $\Psi \subseteq \widehat{\Omega}$, we use $\Lambda(\Psi)$ to denote $\sum_{C\in\Psi}\Lambda(C)$. The probability of a configuration $C \in \Omega$ in the stationary distribution

[2] To see that $I \cap F = T \cap X$ consider some $e \in I \cap F$. Note from the definition of $\gamma(I, F)$ that e is in every configuration along way from I to F. Hence e is in T. By the definition of X, e is also in X. The other direction is similar.

of the weighted worm process is then $\pi(C) = \Lambda(C)/\Lambda(\Omega)$. We can then bound the congestion through transition (T, T') arising from the canonical paths Γ as follows.

$$
\begin{aligned}
\sum_{(I,F)\in \mathrm{cp}(T,T')} \pi(I)\pi(F) &= \frac{1}{\Lambda(\Omega)^2} \sum_{(I,F)\in \mathrm{cp}(T,T')} \Lambda(I)\Lambda(F) \\
&\leq \frac{8}{\Lambda(\Omega)^2} \sum_{(I,F)\in \mathrm{cp}(T,T')} \Lambda(T)\Lambda(\eta_{T,T'}(I,F)) \\
&\leq \frac{8}{\Lambda(\Omega)^2} \sum_{X\in\widehat{\Omega}} \Lambda(T)\Lambda(X) \\
&= 8 \times \frac{\Lambda(\widehat{\Omega})}{\Lambda(\Omega)} \times \frac{\Lambda(T)}{\Lambda(\Omega)} \\
&= O(n^2)\pi(T). \qquad (9)
\end{aligned}
$$

The second inequality uses the fact that $\eta_{T,T'}$ is injective. The final equality follows from the observation that

$$\Lambda(\Omega) = \sum_{S:|S|\in\{0,2\}} \widehat{Z}_S(G;\lambda,w) \quad \text{and} \quad \Lambda(\widehat{\Omega}) = \sum_{S:|S|\in\{0,2,4\}} \widehat{Z}_S(G;\lambda,w).$$

The first sum has $O(n^2)$ terms and the second $O(n^4)$. Thus,

$$\frac{\Lambda(\widehat{\Omega})}{\Lambda(\Omega)} \leq O(n^2) \frac{\max_{S:|S|\in\{0,2,4\}} \widehat{Z}_S(G;\lambda,w)}{\min_{S:|S|\in\{0,2\}} \widehat{Z}_S(G;\lambda,w)} = O(n^2) \frac{\max_{S:|S|\in\{0,2,4\}} w_S Z_S(G;\lambda)}{\min_{S:|S|\in\{0,2\}} w_S Z_S(G;\lambda)}.$$

By Lemma 12, this is at most

$$O(n^2) \frac{2Z_\emptyset(G;\lambda)}{\frac{1}{2}Z_\emptyset(G;\lambda)} = O(n^2),$$

so the final equality holds.

By establishing (9), we have done most of the work required to estimate the congestion $\varrho(\Gamma)$ in (8). Since the paths have length at most m, the only remaining task is to lower bound the transition probability $P(T, T')$. Let $e = \{u, v\} \in E$ be any edge, and let $S \subseteq V$ be any subset of vertices of even cardinality. There is a bijection between Ω_S and $\Omega_{S\oplus\{u,v\}}$ obtained by flipping the edge e. Since this operation changes only a single edge, we see that

$$\lambda(e) Z_S(G;\lambda) \leq Z_{S\oplus\{u,v\}}(G;\lambda) \leq \lambda(e)^{-1} Z_S(G;\lambda).$$

Then, from Lemma 12,

$$\frac{\lambda_{\min}}{4} \leq \frac{Z_S(G;\lambda)}{4Z_{S\oplus\{u,v\}}(G;\lambda)} \leq \frac{w_{S\oplus\{u,v\}}}{w_S} \leq \frac{4Z_S(G;\lambda)}{Z_{S\oplus\{u,v\}}(G;\lambda)} \leq \frac{4}{\lambda_{\min}}.$$

Since T and T' differ by a single edge, this implies

$$\frac{\lambda_{\min}^2}{4} \leq \frac{\Lambda(T')}{\Lambda(T)} \leq \frac{4}{\lambda_{\min}^2}.$$

Going back to the definition of the weighted worm process in Figure 1, it follows that $P(T,T') \geq \frac{1}{2}\frac{1}{m}\min\{\Lambda(T')/\Lambda(T),1\} \geq \lambda_{\min}^2/(8m)$.

Now, starting from (8), and plugging in the bound that path-lengths are at most m and (9) and then this bound, we get

$$\begin{aligned}\varrho(\Gamma) &= \max_{(T,T')}\left\{\frac{1}{\pi(T)P(T,T')}\sum_{(I,F)\in\mathrm{cp}(T,T')}\pi(I)\pi(F)\,|\gamma(I,F)|\right\}\\ &\leq \max_{(T,T')}\left\{\frac{1}{\pi(T)P(T,T')}O(n^2)\pi(T)m\right\} = O(\lambda_{\min}^{-2}n^2m^2)\end{aligned}$$

In order to apply Lemma 14 we must find an upper bound for $\pi(\overline{\Omega}_\emptyset)/\pi(\Omega_\emptyset)$ and $\pi(\Omega_\emptyset)/\pi(\overline{\Omega}_\emptyset)$. Using the upper bound in Lemma 12,

$$\frac{\pi(\overline{\Omega}_\emptyset)}{\pi(\Omega_\emptyset)} = \frac{\sum_{A\in\overline{\Omega}_\emptyset}\Lambda(A)}{\sum_{A\in\Omega_\emptyset}\Lambda(A)} = \frac{\sum_{S:|S|=2}\widehat{Z}_S(G;\lambda,w)}{\widehat{Z}_\emptyset(G;\lambda,w)} = \frac{\sum_{S:|S|=2}w_S Z_S(G;\lambda)}{w_\emptyset Z_\emptyset(G;\lambda)} = O(n^2).$$

Similarly, $\pi(\Omega_\emptyset)/\pi(\overline{\Omega}_\emptyset) = O(1/n^2) = O(n^2)$.

Now applying Lemma 14, there is a collection of flows $\mathcal{F}$ with $\varrho(\mathcal{F}) \leq O(n^2)\varrho(\Gamma) = O(\lambda_{\min}^{-2}n^4m^2)$.

In order to apply Lemma 13 starting from state $T_0 = \emptyset$ we need an upper bound for $\ln(1/\pi(\emptyset))$. For this we use

$$\ln\left(\frac{1}{\pi(\emptyset)}\right) = \ln\left(\frac{\widehat{Z}(G;\lambda,w)}{\Lambda(\emptyset)}\right) = \ln(\widehat{Z}(G;\lambda,w)).$$

By the definition of $\widehat{Z}(G;\lambda,w)$ and (7),

$$\ln(\widehat{Z}(G;\lambda,w)) \leq \ln(n^2\widehat{Z}_\emptyset(G;\lambda,w)) = \ln(n^2 Z_\emptyset(G;\lambda)) \leq \ln(n^2 2^m) = O(m),$$

where the asymptotic bound uses the fact that G is connected.

Finally, by Lemma 13,

$$t_{\mathrm{mix},\emptyset}(\delta) \leq \varrho(\mathcal{F})\left(O(m)+\ln\left(\frac{1}{\delta}\right)\right) = O(\lambda_{\min}^{-2}n^4m^2)\left(O(m)+\ln\left(\frac{1}{\delta}\right)\right).$$

□

The following lemma captures how we will use Lemma 15.

Lemma 16. *There is an algorithm that takes as input an n-vertex connected graph $G=(V,E)$ with a weighting $(\lambda,w)\in\mathcal{W}(G)$ and a set $S\subseteq V$ with $|S|=2$, also an accuracy parameter $\varepsilon\in(0,1)$ and a desired failure probability δ^*. With probability at least $1-\delta^*$, the algorithm produces as estimate $\widehat{R}$ such that*

$$e^{-\varepsilon}\widehat{R} \leq \frac{\widehat{Z}_\emptyset(G;\lambda,w)}{\widehat{Z}_S(G;\lambda,w)} \leq e^{\varepsilon}\widehat{R}.$$

Let $\lambda_{\min} = \min_{e\in E}\lambda(e)$. The running time of the algorithm is at most a polynomial in n, $1/\lambda_{\min}$, $1/\varepsilon$, and $\log(1/\delta^*)$.

Proof. Let $\theta = \varepsilon/8$, $\delta = \varepsilon/(32n^2)$ and $T = \lceil \ln(6/\delta^*) e^{8n^2\delta} 12n^2/\theta^2 \rceil$. Let $\lambda_{\min} = \min_{e\in E} \lambda(e)$. Let t be the upper bound on the mixing time $t_{\mathrm{mix},\emptyset}(\delta)$ of the weighted worm process, from Lemma 15. Given the definition of δ, t is at most a polynomial in n, $1/\lambda_{\min}$, and $\log(1/\varepsilon)$. For $i \in [T]$, the algorithm will run the weighted worm process for t steps, starting from the empty configuration, computing x_i, the indicator for the event that the output is in $\Omega_\emptyset$, Similarly, for $i \in [T]$, the algorithm will run the weighted worm process for t steps, starting from the empty configuration, computing y_i, the indicator for the event that the output is in Ω_S. Let $x = \sum_{i=1}^T x_i$ and $y = \sum_{y=1}^T y_i$. The output is then $\widehat{R} = x/y$. The calculation of errors is standard. Let $p_\emptyset = \Lambda(\emptyset) = \widehat{Z}_\emptyset(G;\lambda,w)/\widehat{Z}(G;\lambda,w)$. Since $(\lambda, w) \in \mathcal{W}(G)$, by the definition of $\mathcal{W}(G)$, we have the loose inequality $1/(4n^2) \le p_\emptyset \le 4/n^2$. By the total variation distance guarantee of Lemma 15, the probability $\hat{p}_\emptyset$ that $x_i = 1$ satisfies $\hat{p}_\emptyset \le p_\emptyset + \delta = (1+\delta/p_\emptyset)p_\emptyset \le e^{\delta/p_\emptyset} p_\emptyset \le e^{4n^2\delta} p_\emptyset$ and $\hat{p}_\emptyset \ge p_\emptyset - \delta = (1-\delta/p_\emptyset)p_\emptyset \ge e^{-2\delta/p_\emptyset} p_\emptyset \ge e^{-8n^2\delta} p_\emptyset$. Then by a Chernoff bound, for any $\theta \in (0,1)$,

$$\Pr(x \ge e^\theta T e^{4n^2\delta} p_\emptyset) \le \Pr(x \ge (1+\theta)T\hat{p}_\emptyset) \le 2\exp(-\theta^2 \hat{p}_\emptyset T/3) \le 2\exp(-\theta^2 T/(e^{8n^2\delta}12n^2)).$$

Similarly,

$$\Pr(x \le e^{-2\theta} T e^{-8n^2\delta} p_\emptyset) \le \Pr(x \le (1-\theta)T\hat{p}_\emptyset) \le \exp(-\theta^2 \hat{p}_\emptyset T/2) \le \exp(-\theta^2 T/(e^{8n^2\delta}8n^2)).$$

Similarly, with $p_S = \Lambda(S) = \widehat{Z}_S(G;\lambda,w)/\widehat{Z}(G;\lambda,w)$, the probability that y fails to satisfy $e^{-2\theta} e^{-8n^2\delta} p_S T \le y \le e^\theta e^{4n^2\delta} p_S T$ is at most $3\exp(-\theta^2 T/(e^{8n^2\delta}12n^2))$. The accuracy guarantee follows from the choice of θ and δ, which ensure that $e^{2\theta} e^{8n^2\delta} = e^{\varepsilon/2}$.

The failure probability guarantee comes from the fact that $6\exp(-\theta^2 T/(e^{8n^2\delta}12n^2)) \le \delta^*$. The worm process is simulated for t steps $O(T)$ times, giving the running time bound in the statement of the lemma. □

3.2 Learning appropriate weights for the worm process

Lemma 15 shows that the weighted worm process is rapidly mixing as long as the weighting (λ, w) is in $\mathcal{W}(G)$. Let $G = (V, E)$ be a connected graph with $|V| = n$ and $|E| = m$. Let $\lambda : E \to \mathbb{Q} \cap (0,1)$ be an edge weighting of G.

In this section we show how to learn a sequence $(\lambda^{[0]}, w^{[0]}), \ldots, (\lambda^{[t]}, w^{[t]})$ of weightings so that each weighting $(\lambda^{[i]}, w^{[i]})$ is in $\mathcal{W}(G)$. The sequence will satisfy

$$\lambda^{[i]}(e) = \max(1/(1+\tfrac{1}{2m})^i, \lambda(e)), \tag{10}$$

so taking $t = \max_{e\in E} \lceil \log(1/\lambda(e))/\log(1+\frac{1}{2m}) \rceil$, we have $\lambda^{[t]} = \lambda$. The results of the section are summarised in Lemma 19.

Although the definition of $\lambda^{[i]}$, from Equation (10), is straightforward, the definition of the subset weighting $w^{[i]}$ is more complicated. In order to conform with the definition (7) of $\mathcal{W}(G)$, we will set $w_\emptyset^{[i]} = 1$ for all $i \in \{0, \ldots, t\}$. Also, for sets S with $|S| = 1$, we set $w_S^{[i]} = 0$. This leaves the definition of $w_S^{[i]}$ where $|S| = 2$. For this, we start by defining the base case, which is $i = 0$. Then, we show how to learn $w^{[i+1]}$ from $w^{[i]}$ by running the weighted worm process. As quantified by Lemma 19, there is a probability that the process does not converge sufficiently quickly to its stationary distribution. Thus, throughout this section we take δ to be the desired failure probability, from Lemma 19. We will give an algorithm which, with probability at least $1 - \delta$, learns the weights. We start by defining the base case. For every size-2 set $S \subseteq V$, we set $w_S^{[0]} = 1$.

Observation 17. The weighting $(\lambda^{[0]}, w^{[0]})$ is in $\mathcal{W}(G)$.

Proof. First note that, for every $e \in E$, $\lambda(e) \le 1$ so $\lambda^{[0]}(e) = 1$.

Consider any $S \subseteq V$ with $|S| = 2$ and note that

$$\frac{\widehat{Z}_S(G; \lambda^{[0]}, w^{[0]})}{\widehat{Z}_\emptyset(G; \lambda^{[0]}, w^{[0]})} = \frac{\sum_{A \in \Omega_S} 1}{\sum_{A \in \Omega_\emptyset} 1}.$$

We will show below that $|\Omega_S| = |\Omega_\emptyset|$. This ensures that $(\lambda^{[0]}, w^{[0]})$ satisfies Equation (7) so it is in $\mathcal{W}(G)$ and the observation follows.

To see that $|\Omega_S| = |\Omega_\emptyset|$, we establish a bijection τ between Ω_S and $\Omega_\emptyset$. Let $S = \{u, v\}$ and let P be the set of edges in any fixed path from u to v in G — such a path exists since G is connected. The bijection is straightforward. Given any $A \in \Omega_S$, let $\tau(A) = A \oplus P$ and note that $\tau(A) \in \Omega_\emptyset$. $\square$

Now consider the weighting $(\lambda^{[i]}, w^{[i]})$. If $i < t$ then, for every size-2 subset S of V, we define $w_S^{[i+1]}$ by running the weighted worm process, as follows. Set $\varepsilon = 1/8$ and set $\delta^* = \delta/(n^2 t)$. Now run the weighted worm process with weighting $(\lambda^{[i]}, w^{[i]})$ to obtain (by Lemma 16) an estimate $\widehat{R}_S^{[i]}$ which, with probability at least $1 - \delta^*$, satisfies

$$e^{-\varepsilon} \widehat{R}_S^{[i]} \le \frac{\widehat{Z}_\emptyset(G; \lambda^{[i]}, w^{[i]})}{\widehat{Z}_S(G; \lambda^{[i]}, w^{[i]})} \le e^{\varepsilon} \widehat{R}_S^{[i]}.$$

In the proof of Lemma 19, we will use Lemma 16 to account for how long this run of the weighted worm process takes. To conclude with the definition of $w_S^{[i+1]}$, let $w_S^{[i+1]} = w_S^{[i]} \widehat{R}_S^{[i]}$.

Lemma 18. *Assuming that the algorithm from Lemma 16 does not fail when it is called to learn $\widehat{R}_S^0, \ldots, \widehat{R}_S^{[i]}$, The weighting $(\lambda^{[i+1]}, w^{[i+1]})$ is in $\mathcal{W}(G)$.*

Proof. Consider any $S \subseteq V$ with $|S| = 2$. Then

$$\begin{aligned}
\frac{\widehat{Z}_S(G; \lambda^{[i+1]}, w^{[i+1]})}{\widehat{Z}_\emptyset(G; \lambda^{[i+1]}, w^{[i+1]})} &= \frac{\sum_{A \in \Omega_S} w_S^{[i+1]} \prod_{e \in A} \lambda^{[i+1]}(e)}{\sum_{A \in \Omega_\emptyset} w_\emptyset^{[i+1]} \prod_{e \in A} \lambda^{[i+1]}(e)} \\
&= \frac{w_S^{[i]} \widehat{R}_S^{[i]} \sum_{A \in \Omega_S} \prod_{e \in A} \lambda^{[i+1]}(e)}{\sum_{A \in \Omega_\emptyset} \prod_{e \in A} \lambda^{[i+1]}(e)}.
\end{aligned}$$

Using the upper bound on $\widehat{R}_S^{[i]}$ and $\frac{\lambda^{[i]}(e)}{(1+\frac{1}{2m})} \le \lambda^{[i+1]}(e) \le \lambda^{[i]}(e)$, this quantity is at most

$$\begin{aligned}
\frac{w_S^{[i]} e^{\varepsilon} \widehat{Z}_\emptyset(G; \lambda^{[i]}, w^{[i]}) \sum_{A \in \Omega_S} \prod_{e \in A} \lambda^{[i]}(e)}{\widehat{Z}_S(G; \lambda^{[i]}, w^{[i]}) \sum_{A \in \Omega_\emptyset} \prod_{e \in A} \frac{\lambda^{[i]}(e)}{1+\frac{1}{2m}}} &\le e^{\varepsilon}(1 + \tfrac{1}{2m})^m \frac{w_S^{[i]} \widehat{Z}_\emptyset(G; \lambda^{[i]}, w^i) \sum_{A \in \Omega_S} \prod_{e \in A} \lambda^{[i]}(e)}{\widehat{Z}_S(G; \lambda^{[i]}, w^i) \sum_{A \in \Omega_\emptyset} \prod_{e \in A} \lambda^{[i]}(e)} \\
&= e^{\varepsilon}(1 + \tfrac{1}{2m})^m \le 2.
\end{aligned}$$

Similarly, the quantity is at least 1/2. $\square$

Collecting what we have done in this section, we get the following lemma.

Lemma 19. *There is a randomised algorithm that takes as input a connected graph $G = (V, E)$ with $|V| = n$ and $|E| = m$ and an edge weighting $\lambda : E \to \mathbb{Q} \cap (0, 1)$. The algorithm also takes a failure probability δ. With probability at least $1-\delta$, it computes a subset weighting w of G so that (λ, w) is in $\mathcal{W}(G)$. Let $\lambda_{\min} = \min_{e\in E} \lambda(e)$. The running time of the algorithm is at most a polynomial in n, $1/\lambda_{\min}$, and $\log(1/\delta)$.*

Proof. Let $t = \max_{e\in E} \lceil \log(1/\lambda(e))/\log(1 + \frac{1}{2m})\rceil$. Note that $t = O(m \log(1/\lambda_{\min}))$. The algorithm constructs the sequence $(\lambda^{[0]}, w^{[0]}), \ldots, (\lambda^{[t]}, w^{[t]})$ of weightings as described in this section where $\lambda^{[t]} = \lambda$.

We just have to collect the failure probabilities and running times. As note earlier, for $i \in \{0, \ldots, t-1\}$, for each size-2 subset S of V, we estimate $\widehat{R}_S^{[i]}$ using the weighted worm process (Lemma 16) with $\varepsilon = 1/8$ and specified failure probability $\delta^* = \delta/(n^2 t)$.

The running time from Lemma 16 is at most a polynomial in n, $1/\lambda_{\min}^{[i]}$, and $\log(1/\delta^*)$. Recall that $\lambda^{[i]}(e) \geq \lambda(e)$. Thus, the overall running time is at most a polynomial in n, $1/\lambda_{\min}$, and $\log(1/\delta)$. By a union bound, the overall failure probability is at most δ. □

4 The Proof of Theorem 2

Theorem 2. There is an FPRAS for FerroIsingCov.

Proof. We start by reviewing what are the inputs and outputs of an FPRAS for FerroIsingCov.

The input consists of an input to FerroIsingCov, an accuracy parameter $\varepsilon \in (0, 1)$, and a failure probability $\delta \in (0, 1)$. An input to FerroIsingCov consists of a graph $G = (V, E)$ with specified vertices s and t and an edge weighting $\beta : E \to \mathbb{Q}_{>1}$ of G. Let $n = |V|$. We need to be more specific about how the edge weighting β is represented. Recall from the introduction that each weight $\beta(e)$ satisfies $\beta(e) > 1$ and is represented in the input by two positive integers $P(e)$ and $Q(e)$ (specified in unary in the input) such that and $\beta(e) = 1 + P(e)/Q(e)$.

With probability at least $1-\delta$, the output $\widehat{C}$ of the FPRAS should satisfy

$$e^{-\varepsilon}\widehat{C} \leq \mathbb{E}_{\pi^{\text{Ising}}_{G,\beta}}[\sigma(s)\sigma(t)] \leq e^{\varepsilon}\widehat{C}. \tag{11}$$

Finally, in order to be an FPRAS, the running time should be at most a polynomial in n, $\sum_{e\in E}(P(e) + Q(e))$, $1/\varepsilon$, and $\log(1/\delta)$.

If s and t are in different connected components of G then $\mathbb{E}_{\pi^{\text{Ising}}_{G,\beta}}[\sigma(s)\sigma(t)]=0$, so we can just output 0 in this case. If s and t are in the same connected component, G', of G then $\mathbb{E}_{\pi^{\text{Ising}}_{G,\beta}}[\sigma(s)\sigma(t)] = \mathbb{E}_{\pi^{\text{Ising}}_{G',\beta}}[\sigma(s)\sigma(t)]$. So assume without loss of generality, for the rest of the proof, that G is connected.

Now let λ be the edge-weighting of G defined by $\lambda(e) = (\beta(e) - 1)/(\beta(e) + 1)$. Let $\lambda_{\min} = \min_{e\in E} \lambda(e)$.

The FPRAS should first run the algorithm of Lemma 19 with input G, λ and $\delta/2$. With probability at least $1-\delta/2$, it computes a subset weighting w of G so that (λ, w) is in $\mathcal{W}(G)$.

Let $S = \{s, t\}$. Suppose that the algorithm of Lemma 19 has succeeded. Recall from Lemma 5 that $\mathbb{E}_{\pi^{\text{Ising}}_{G,\beta}}[\sigma(s)\sigma(t)] = \frac{Z_S(G;\lambda)}{Z_\emptyset(G;\lambda)}$. Also, by plugging in the definitions of $\widehat{Z}_S(G; \lambda, w)$ (from the beginning of Section 3) and $Z_S(G; \lambda)$ (from the beginning of Section 2) we have

$$\mathbb{E}_{\pi^{\text{Ising}}_{G,\beta}}[\sigma(s)\sigma(t)] = \frac{Z_S(G; \lambda)}{Z_\emptyset(G; \lambda)} = \frac{\widehat{Z}_S(G; \lambda, w)}{w_S \widehat{Z}_\emptyset(G; \lambda, w)}.$$

Since we already know w_S, the goal is to compute a quantity $\widehat{Q}$ such that

$$e^{-\varepsilon}\widehat{Q} \le \frac{\widehat{Z}_S(G;\lambda,w)}{\widehat{Z}_\emptyset(G;\lambda,w)} \le e^{\varepsilon}\widehat{Q}.$$

Then we satisfy (11) by taking $\widehat{C} = \widehat{Q}/w_S$.

The estimate $\widehat{Q}$ can be obatined by running the algorithm of Lemma 16 with input G, weighting (λ, w), set $S = \{s,t\}$, accuracy parameter ε, desired failure probability $\delta^* = \delta/2$, letting $\widehat{R}$ be the output of the algorithm, and taking $\widehat{Q} = 1/\widehat{R}$.

The running time of both algorithms is at most a polynomial in n, $\log(1/\delta)$, $1/\varepsilon$ and

$$1/\lambda_{\min} = \max_{e\in E} \frac{\beta(e)+1}{\beta(e)-1} = \max_{e\in E}\left\{1 + \frac{2}{\beta(e)-1}\right\} = \max_{e\in E}\left\{1+\frac{2Q(e)}{P(e)}\right\}.$$

□

Remark 20. It is possible to improve the run-time of the algorithm of Theorem 2 so that the dependence on $1/\lambda_{\min} = \max_{e\in E}\left\{1+\frac{2}{\beta(e)-1}\right\}$ is logarithmic, rather than polynomial. To do this, we pre-process the graph G. If an edge e has a weight $\beta(e)$ that is very close to 1 then it is replaced with a subgraph J. The weights of the edges of J are constants bounded above 1, but the overall effect of J is to simulate the weight $\beta(e)$ with exponential precision. The technical details of the simulation are very similar to what we do in Lemma 22 of Section 5. We omit the details since polynomial-time algorithms (as opposed to strongly polynomial-time algorithms) are sufficient for our purposes.

5 The Antiferromagnetic Case

Consider the antiferromagnetic Ising model on a graph $G = (V, E)$ as defined in Section 1. We will consider the situation where, for some $b \in (0,1)$, $G = (V,E)$ is a graph with an edge weighting β that assigns value $\beta(e) = b$ to every $e \in E$. We will simplify the notation by defining

$$\mathrm{wt}^{\mathrm{Ising}}_{G,b}(\sigma) = b^{|\{e=\{u,v\}\in E:\sigma(u)=\sigma(v)\}|}$$

with the corresponding partition function $Z^{\mathrm{Ising}}_{G,b} = \sum_{\sigma:V\to\{-1,+1\}} \mathrm{wt}^{\mathrm{Ising}}_{G,b}(\sigma)$ and Gibbs distribution $\pi^{\mathrm{Ising}}_{G,b}(\cdot)$. We will be interested in the following computational problem, with parameter $b \in (0,1)$.

Name. $\mathsf{SignIsingCov}_b$.

Instance. A graph G with specified vertices s and t

Output. A correct statement of the form "$\mathbb{E}_{\pi^{\mathrm{Ising}}_{G,b}}[\sigma(s)\sigma(t)] \ge 0$" or "$\mathbb{E}_{\pi^{\mathrm{Ising}}_{G,b}}[\sigma(s)\sigma(t)] \le 0$".

The purpose of this section is to prove Theorem 23 which states that, for any $b \in (0,1)$, $\mathsf{SignIsingCov}_b$ is #P-hard. Note that the #P-hardness does not come from the difficulty of determining whether or not $\mathbb{E}_{\pi^{\mathrm{Ising}}_{G,b}}[\sigma(s)\sigma(t)]$ is zero — an algorithm for $\mathsf{SignIsingCov}_b$ is allowed to give either answer in this case. Theorem 23 implies that it is #P-hard to approximate $\mathbb{E}_{\pi^{\mathrm{Ising}}_{G,b}}[\sigma(s)\sigma(t)]$ within any specified factor, since such an approximation would allow one to determine either $\mathbb{E}_{\pi^{\mathrm{Ising}}_{G,b}}[\sigma(s)\sigma(t)] \ge 0$ or $\mathbb{E}_{\pi^{\mathrm{Ising}}_{G,b}}[\sigma(s)\sigma(t)] \le 0$.

We start with some notation. Given vertices s and t of G, let $\Psi_{s+,t+}$ be the set of assignments $\sigma\colon V \to \{-1,+1\}$ that satisfy $\sigma(s) = +1$ and $\sigma(t) = +1$. Let

$$Z^{\text{Ising}}_{G,b,s+,t+} = \sum_{\sigma \in \Psi_{s+,t+}} \text{wt}(\sigma).$$

Define the sets of assignments $\Psi_{s+,t-}$, $\Psi_{s-,t+}$, $\Psi_{s-,t-}$ and the partition functions $Z^{\text{Ising}}_{G,b,s+,t-}$, $Z^{\text{Ising}}_{G,b,s-,t+}$, and $Z^{\text{Ising}}_{G,b,s-,t-}$ similarly. We use the following notation of implementation.

Definition 21. A graph J is said to *b-implement* a rational number b' if there are vertices s and t of J such that $Z^{\text{Ising}}_{J,b,s+,t+}/Z^{\text{Ising}}_{G,b,s+,t-} = b'$. We call s and t the terminals of J.

We will use the following lemma for implementations.

Lemma 22. *Fix $b \in (0,1)$. There is a polynomial-time algorithm that takes as input*

- *A positive integer n, in unary,*
- *a target-edge weight $b' \in [b^n, b^{-n}]$, and*
- *a rational accuracy parameter $\varepsilon \in (0,1)$, in binary.*

The algorithm produces a graph J with terminals s and t that b-implements a value $\hat{b}$ satisfying $|\hat{b} - b'| \le \varepsilon$. The size of J is at most a polynomial in n and $\log(1/\varepsilon)$, independently of b'.

Proof. If $b' = 1$ then J is the graph with vertices s and t and no edges. So suppose $b' \neq 1$.

Let P_ℓ be an ℓ-edge path with endpoints s and t. Let $f_\ell = Z^{\text{Ising}}_{P_\ell,b,s+,t+}$ and $a_\ell = Z^{\text{Ising}}_{P_\ell,b,s+,t-}$. Then $f_1 = b$, $a_1 = 1$ and we have the system $f_\ell = bf_{\ell-1} + a_{\ell-1}$ and $a_\ell = f_{\ell-1} + ba_{\ell-1}$. Thus, $f_\ell = (1/2)((b+1)^\ell + (b-1)^\ell)$ and $a_\ell = (1/2)((b+1)^\ell - (b-1)^\ell)$. Let $\zeta_\ell = f_\ell/a_\ell$. Then

$$\zeta_\ell = \frac{f_\ell}{a_\ell} = \frac{(b+1)^\ell + (b-1)^\ell}{(b+1)^\ell - (b-1)^\ell} = 1 + \frac{2}{c^\ell - 1},$$

where $c = (b+1)/(b-1) < -1$. Note that for odd ℓ the values of ζ_ℓ are in $(0,1)$ and they increase. Also, for even ℓ the values of ζ_ℓ are greater than 1 and they decrease.

The graph J is constructed by combining copies of P_ℓ (for different values of ℓ), identifying the vertex s in all copies, and identifying the vertex t in all copies. Let

$$L = \left\lceil \frac{\log\left(\frac{2}{b^n\varepsilon} + 1\right)}{\log(c^2)} \right\rceil.$$

We will use paths P_ℓ with $\ell \le 2L+1$. The value of L is defined so that L is at most a polynomial in n and $\log(1/\varepsilon)$, as required in the statement of the lemma, and also

$$c^{2L} \ge 2b^{-n}/\varepsilon + 1. \tag{12}$$

The graph J is constructed as follows.

- If $b' > 1$: Set $B_0 = b' > 1$. For odd $j \in \{2, \dots, 2L+1\}$, let $d_j = 0$. For even $j \in \{2, \dots, 2L+1\}$, let d_j be the largest non-negative integer such that $\zeta_j^{d_j} \le B_{j-2}$ and let $B_j = B_{j-2}/\zeta_j^{d_j} \ge 1$.

- If $b' < 1$: Set $B_1 = b' < 1$. For even $j \in \{2, \ldots, 2L+1\}$, let $d_j = 0$. For odd $j \in \{2, \ldots, 2L+1\}$, let d_j be the largest non-negative integer such that $\zeta_j^{d_j} \geq B_{j-2}$ and let $B_j = B_{j-2}/\zeta_j^{d_j} \leq 1$.

The graph J is constructed by taking d_j copies of P_j for $j \in \{2, \ldots, 2L+1\}$, identifying the vertex s in all copies and identifying the vertex t in all copies. The value that J b-implements is

$$\hat{b} = \frac{Z^{\text{Ising}}_{J,b,s+,t+}}{Z^{\text{Ising}}_{J,b,s+,t-}} = \prod_{j=2}^{2L+1} \zeta_j^{d_j}$$

We next show that $|\hat{b} - b'| \leq \varepsilon$.

- If $b' > 1$: The construction guarantees $\frac{b'}{\zeta_{2L}} \leq \hat{b} \leq b'$. But (12) implies $\zeta_{2L} \leq 1 + \frac{\varepsilon}{b^{-n}} \leq 1 + \frac{\varepsilon}{b'} \leq \frac{1}{1-\frac{\varepsilon}{b'}}$, so $b' - \varepsilon \leq b'/\zeta_{2L}$.

- If $b' < 1$: The construction guarantees $b' \leq \hat{b} \leq \frac{b'}{\zeta_{2L+1}}$. But (12) implies $|c|^{2L+1} \geq c^{2L} \geq 2b^{-n}/\varepsilon + 1$ so $\zeta_{2L+1} \geq 1 - \frac{2}{|c|^{2L+1}+1} \geq 1 - \frac{\varepsilon}{b^{-n}+\varepsilon} \geq 1 - \frac{\varepsilon}{b'+\varepsilon} = \frac{b'}{b'+\varepsilon}$, so $b'/\zeta_{2L+1} \leq b' + \varepsilon$.

To finish the bound on the size of J, we will show that d_2 and d_3 are $O(n)$ and that, for every $j \in \{4, \ldots, 2L+1\}$, $d_j = O(1)$. First, $d_2 \leq \log_{\zeta_2}(b')$ where $\zeta_2 = (b^2+1)/(2b)$. Also, $d_3 \leq \log_{1/\zeta_3}(1/b')$ where $1/\zeta_3 = (1+3b^2)/(b(3+b^2))$.

Finally, let $d = \lceil c^4/(c^2-1) \rceil$. Note that we could replace "c" with "$|c|$" in the definition of d without changing the definition, so, plugging the definition in, we find, for $j \geq 4$, that $|c|^j - 1 \leq d(|c|^{j-2} - 1)$. This implies

$$\left(1 + \frac{2}{|c|^j - 1}\right)^d \geq 1 + \frac{2d}{|c|^j - 1} \geq 1 + \frac{2}{|c|^{j-2} - 1}. \tag{13}$$

If $j \geq 4$ is even then (13) implies $\zeta_j^d \geq \zeta_{j-2}$, so $d_j \leq d$. If $j \geq 4$ is odd then (13) gives

$$\left(\frac{|c|^j - 1}{|c|^j + 1}\right)^d \leq \frac{|c|^{j-2} - 1}{|c|^{j-2} + 1},$$

and all numerators and denominators are negative, so multiplying them by -1 we get

$$\left(\frac{c^j + 1}{c^j - 1}\right)^d \leq \frac{c^{j-2} + 1}{c^{j-2} - 1},$$

so $\zeta_j^d \leq \zeta_{j-2}$ and $d_j \leq d$. □

Theorem 23. *Let $b \in (0,1)$ be a rational number. Then* $\textsf{SignIsingCov}_b$ *is #P-hard.*

Proof. Fix $b \in (0,1)$. We will show how to use an oracle for $\textsf{SignIsingCov}_b$ to give a polynomial time algorithm for exactly computing $Z^{\text{Ising}}_{G,b}$, a problem that is known to be #P-hard (see [7, Theorem 14] for #P-hardness of a multi-variate version and [15, Corollary 2] for a result that implies #P-hardness of the version considered here).

Let $G = (V, E)$ be a graph with n vertices and $m > 0$ edges. We will show how to compute $Z^{\text{Ising}}_{G,b}$ using the given oracle for $\textsf{SignIsingCov}_b$.

As we will see, the information provided by the oracle for $\mathsf{SignIsingCov}_b$ can naturally be used to provide a multiplicative approximation to $Z^{\text{Ising}}_{G,b}$. Since we need additive approximations in order to compute $Z^{\text{Ising}}_{G,b}$ precisely, we have to switch back and forth between additive and multiplicative approximations. To this end, let $b = p/q$ for integers p and q. Choose $m' = O(m)$ such that $b^{m'} \leq (1/q)^m$.

Note that $Z^{\text{Ising}}_{G,b} = \sum_{j=0}^{m} b^j c_j$, where c_j is the number of configurations $\sigma\colon V \to \{-1,+1\}$ which induce j edges with like spins in G. This implies that $b^m 2^n \leq Z^{\text{Ising}}_{G,b} \leq 2^n$. Now let $\delta = b^{m'} 2^{-(n+3)}$. Suppose that $\widehat{Z}$ satisfies

$$e^{-\delta} Z^{\text{Ising}}_{G,b} \leq \widehat{Z} \leq e^{\delta} Z^{\text{Ising}}_{G,b} \tag{14}$$

so that $(1-\delta) Z^{\text{Ising}}_{G,b} \leq \widehat{Z} \leq (1+2\delta) Z^{\text{Ising}}_{G,b}$. We conclude that

$$|\widehat{Z} - Z^{\text{Ising}}_{G,b}| \leq 2\delta Z^{\text{Ising}}_{G,b} \leq 2\delta 2^n \leq b^{m'}/4$$

so from $\widehat{Z}$ we learn $Z^{\text{Ising}}_{G,b}$ precisely. To see this, note that any interval of length $b^{m'}/2$ contains the value of at most one polynomial of the form $\sum_{j=0}^{m} b^j c_j$ with integer coefficients. Consider two such polynomials $Z_1 = \sum_{j=0}^{m} b^j c_j$ and $Z_2 = \sum_{j=0}^{m} b^j c'_j$, both with integer coefficients. Set $a_j = c_j - c'_j$. Then

$$Z_1 - Z_2 = \sum_{j=0}^{m} \frac{a_j p^j}{q^j} \leq \frac{\sum_{j=0}^{m} a_j p^j q^{m-j}}{q^m},$$

but the numerator is an integer, so if $Z_1 \neq Z_2$ then $|Z_1 - Z_2| \geq 1/q^m \geq b^{m'}$, so Z_1 and Z_2 cannot both be in an interval of length $b^{m'}/2$.

Thus, from now on, our goal will be to show how to use the given oracle for $\mathsf{SignIsingCov}_b$ to obtain $\widehat{Z}$ satisfying (14). This will complete the proof of Theorem 23 and, a fortiori, Theorem 3.

Let the edges of G be $e_1, \dots, e_m$ and, for $j \in [m]$, let $G_j = (V, \{e_1, \dots, e_j\})$. Denote the endpoints of e_j by s_j and t_j. Using the notation from the beginning of the section, let $\nu_j = Z^{\text{Ising}}_{G_{j-1},b,s_j+,t_j-} / Z^{\text{Ising}}_{G_{j-1},b,s_j+,t_j+}$ and let $\alpha_j = (b+\nu_j)/(1+\nu_j)$. Observe that $Z^{\text{Ising}}_{G_j,b} = 2(bZ^{\text{Ising}}_{G_{j-1},b,s_j+,t_j+} + Z^{\text{Ising}}_{G_{j-1},b,s_j+,t_j-})$ and $Z^{\text{Ising}}_{G_{j-1},b} = 2(Z^{\text{Ising}}_{G_{j-1},b,s_j+,t_j+} + Z^{\text{Ising}}_{G_{j-1},b,s_j+,t_j-})$, and hence $Z^{\text{Ising}}_{G_j,b} = \alpha_j Z^{\text{Ising}}_{G_{j-1},b}$. Therefore,

$$Z^{\text{Ising}}_{G,b} = Z^{\text{Ising}}_{G_m,b} = \left(\prod_{j=1}^{m} \alpha_j\right) Z^{\text{Ising}}_{G_0,b} = 2^n \prod_{j=1}^{m} \alpha_j,$$

so to finish it suffices to show, for $j \in [m]$, that we can use an oracle for $\mathsf{SignIsingCov}_b$ to approximate α_j with multiplicative error $\exp(\pm\delta/m)$.

Suppose that we could produce $\hat{\nu}_j$ satisfying $|\hat{\nu}_j - \nu_j| \leq b\delta/(5m)$. Then, setting $\hat{\alpha}_j = (b+\hat{\nu}_j)/(1+\hat{\nu}_j)$, we have $\alpha_j \exp(-\delta/m) \leq \hat{\alpha}_j \leq \alpha_j \exp(\delta/m)$, as required.

So, to finish it suffices to show, for $j \in [m]$, that we can use an oracle for $\mathsf{SignIsingCov}_b$ to approximate ν_j with additive error at most $\delta' = b\delta/(5m)$. Our basic approach is the binary-search method that the authors used in [5] to show that it is #P-hard to compute the sign of the Tutte polynomial.

The invariant that we will maintain is that ν_j lies in an interval $[\nu_{\min}, \nu_{\max}]$. We will repeatedly use the oracle to reduce the length of the interval by a constant factor, until the length is at most δ' (in which case we can take $\hat{\nu_j}$ to be any point in the interval). To initialise the search interval, we take $[\nu_{\min}, \nu_{\max}] = [b^n, b^{-n}]$. It is clear that ν_j lies in this interval, since flipping the spin at t_j affects at most $n-1$ incident edges, and therefore changes the weight of a configuration by a factor that is at least b^n and at most b^{-n}.

Our basic approach is as follows. Suppose that we can construct a graph J with terminals s and t to b-implement a point $\hat{b}$ in the middle third of the interval. Let β be the edge-labelling of G_j that assigns value $\hat{b}$ to edge e_j and b to all other edges. Let $G_j(J)$ be the graph formed from G_j by replacing the edge e_j with the graph J (identifying the terminal s of J with the vertex s_j of G_j and identifying the terminal t of J with the vertex t_j of G_j). Then

$$\mathbb{E}_{\pi^{\text{Ising}}_{G_j(J),b}}[\sigma(s_j)\sigma(t_j)] = \mathbb{E}_{\pi^{\text{Ising}}_{G_j,\beta}}[\sigma(s_j)\sigma(t_j)] = \frac{\hat{b} Z^{\text{Ising}}_{G_{j-1},b,s_j+,t_j+} - Z^{\text{Ising}}_{G_{j-1},b,s_j+,t_j-}}{\hat{b} Z^{\text{Ising}}_{G_{j-1},b,s_j+,t_j+} + Z^{\text{Ising}}_{G_{j-1},b,s_j+,t_j-}}.$$

Using an oracle for $\mathsf{SignIsingCov}_b$ we can determine either that this quantity is at least 0 (in which case $\hat{b} \geq \nu_j$ and we can recurse on $[\nu_{\min}, \hat{b}]$) or that it is at most 0 (in which $\hat{b} \leq \nu_j$ and we can recurse on $[\hat{b}, \nu_{\max}]$). Either way, the length of the new interval is at most $2/3$ of the length that it was, so after $\mathrm{O}(\log(1/b)(m'+n))$ iterations, the length of the interval will have shrunk to length at most δ', as required.

Finding the required J is straightforward — this can be done by taking $b' = (\nu_{\min} + \nu_{\max})/2$ to be the centre of the interval and using Lemma 22 with inputs n, b' and $\varepsilon = \delta'/6$. The size of J is at most a polynomial in n and $\log(1/\varepsilon)$, which is polynomial in n. $\square$

References

[1] Andrea Collevecchio, Timothy M. Garoni, Timothy Hyndman, and Daniel Tokarev. The worm process for the Ising model is rapidly mixing. *J. Stat. Phys.*, 164(5):1082–1102, 2016.

[2] Persi Diaconis and Daniel Stroock. Geometric bounds for eigenvalues of Markov chains. *Ann. Appl. Probab.*, 1(1):36–61, 1991.

[3] Robert G. Edwards and Alan D. Sokal. Generalization of the Fortuin-Kasteleyn-Swendsen-Wang representation and monte carlo algorithm. *Phys. Rev. D*, 38:2009–2012, Sep 1988.

[4] C. M. Fortuin and P. W. Kasteleyn. On the random-cluster model. I. Introduction and relation to other models. *Physica*, 57:536–564, 1972.

[5] Leslie Ann Goldberg and Mark Jerrum. The complexity of computing the sign of the Tutte polynomial. *SIAM J. Comput.*, 43(6):1921–1952, 2014.

[6] Geoffrey Grimmett. *The random-cluster model*, volume 333 of *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*. Springer-Verlag, Berlin, 2006.

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

[8] Mark Jerrum, Alistair Sinclair, and Eric Vigoda. A polynomial-time approximation algorithm for the permanent of a matrix with nonnegative entries. *J. ACM*, 51(4):671–697, 2004.

[9] Michael Mitzenmacher and Eli Upfal. *Probability and computing.* Cambridge University Press, Cambridge, 2005. Randomized algorithms and probabilistic analysis.

[10] Dana Randall and David Wilson. Sampling spin configurations of an Ising system. In *Proceedings of the Tenth Annual ACM-SIAM Symposium on Discrete Algorithms*, SODA '99, pages 959–960, Philadelphia, PA, USA, 1999. Society for Industrial and Applied Mathematics.

[11] Jason Schweinsberg. An $O(n^2)$ bound for the relaxation time of a Markov chain on cladograms. *Random Structures Algorithms*, 20(1):59–70, 2002.

[12] Alistair Sinclair. Improved bounds for mixing rates of Markov chains and multicommodity flow. *Combin. Probab. Comput.*, 1(4):351–370, 1992.

[13] L. G. Valiant. A scheme for fast parallel communication. *SIAM J. Comput.*, 11(2):350–361, 1982.

[14] Bartel L. van der Waerden. Die lange Reichweite der regelmäßigen Atomanordnung in Mischkristallen. *Zeitschrift für Physik*, 118(7):473–488, 1941.

[15] Dirk Vertigan and Dominic J. A. Welsh. The compunational complexity of the Tutte plane: the bipartite case. *Combinatorics, Probability & Computing*, 1:181–187, 1992.