SAMPLING FROM THE FERROMAGNETIC ISING MODEL WITH EXTERNAL FIELDS

WEIMING FENG, HENG GUO, JIAHENG WANG

Abstract. We study the sampling problem for ferromagnetic Ising models with consistent external fields. We show that the edge-flipping dynamics for a corresponding weighted random cluster model is rapidly mixing. Consequences include rapid mixing of the Swendsen-Wang dynamics and perfect samplers for both models.

1. Introduction

The Ising model is a classical statistical physics model for ferromagnetism that had far-reaching impact in many areas. In computer science / combinatorics terms, the model defines a weighted distribution over cuts of a graph. To be more precise, let $G = (V, E)$ be a simple undirected graph. For each edge $e \in E$, we have the local interaction strength $\beta_e \in \mathbb{R}_{>0}$, and for each vertex $v \in V$, we have the external magnetic field (namely vertex weight) $\lambda_v \in \mathbb{R}_{>0}$. An Ising model is specified by the tuple $(G; \boldsymbol{\beta}, \boldsymbol{\lambda})$, where $\boldsymbol{\beta} = (\beta_e)_{e\in E}$ and $\boldsymbol{\lambda} = (\lambda_v)_{v\in V}$. We assign spins $\{0, 1\}$ to the vertices V. For each spin configuration $\sigma \in \{0, 1\}^V$, the *weight* of σ is defined by

$$\mathrm{wt}_{\mathrm{Ising}}(\sigma) := \prod_{e=(u,v)\in E} \beta_e^{\mathbb{I}[\sigma(u)=\sigma(v)]} \prod_{u\in V} \lambda_u^{\sigma(u)},$$

where $\mathbb{I}[\sigma(u) = \sigma(v)]$ is the indicator variable of the event $\sigma(u) = \sigma(v)$. The *Gibbs distribution* π_{Ising} is defined by

$$\forall \sigma \in \{0, 1\}^V, \quad \pi_{\mathrm{Ising}}(\sigma) = \frac{\mathrm{wt}_{\mathrm{Ising}}(\sigma)}{Z_{\mathrm{Ising}}}, \tag{1}$$

where

$$Z_{\mathrm{Ising}} = Z_{\mathrm{Ising}}(G; \boldsymbol{\beta}, \boldsymbol{\lambda}) := \sum_{\tau\in\{0,1\}^V} \mathrm{wt}_{\mathrm{Ising}}(\tau)$$

is the *partition function*. In this paper we focus on the *ferromagnetic* case, where $\beta_e > 1$ for all $e \in E$, with *consistent* fields, where $\lambda_v \in (0, 1]$ for all $v \in V$. Note that by flipping the spins, the last assumption is equivalent to assuming $\lambda_v \in [1, \infty)$ for all $v \in V$.

There is extensive computational interest in simulating the Ising model and in evaluating various quantities related to it. A major contribution in the rigorous algorithmic study of the model is the Jerrum-Sinclair algorithm [JS93], which is the first *fully polynomial-time randomised approximation scheme* (FPRAS) for the partition function Z_{Ising} of the ferromagnetic Ising model with consistent fields on any graph. The main ingredient of their algorithm is to show that a natural Markov chain mixes in polynomial-time to sample from the so-called "subgraph-world" model, which has the same partition function up to some easy to compute factors.

Usually, using self-reducibility, approximately evaluating the partition function is computationally inter-reducible to approximate sampling [JVV86]. However, in the case of the Ising model, the original algorithm by Jerrum and Sinclair does not directly yield a sampling algorithm for spin configurations. This is because inconsistent fields may be created during the self-reduction, making the algorithm no longer applicable. To circumvent this issue, Randall and Wilson [RW99] showed that when there is no external field, an efficient approximate sampler for spin configurations exists by doing self-reductions

(Weiming Feng, Heng Guo, Jiaheng Wang) School of Informatics, University of Edinburgh, Informatics Forum, Edinburgh, EH8 9AB, United Kingdom. E-mail: wfeng@ed.ac.uk, hguo@inf.ed.ac.uk, jiaheng.wang@ed.ac.uk

in the so-called random cluster model. This is a model introduced by Fortuin and Kasteleyn [FK72] and also has the same partition function as the previous two models up to some easy to compute factors.[1]

On the other hand, a different Markov chain introduced by Swendsen and Wang [SW87] has shown great performance on sampling Ising configurations in practice. This dynamics is best understood via the Edwards-Sokal distribution [ES88], which is a joint distribution on both edges and vertices. The marginal distribution on vertices is the Ising model, and the marginal distribution on edges is the random cluster model. Sokal and later Peres[2] conjectured that the Swendsen-Wang (SW) dynamics mixes in polynomial-time for ferromagnetic Ising models, and this was resolved in affirmative by Guo and Jerrum [GJ18]. They showed that the edge-flipping dynamics for the random cluster model mixes in polynomial-time, and this dynamics is known to be no faster than the SW dynamics [Ull14]. Another consequence of [GJ18] is that there is a perfect sampler for the ferromagnetic Ising model and the corresponding random cluster model, improving upon the approximate sampler of [RW99]. This is done via monotone coupling from the past (CFTP) [PW96] as the random cluster model is monotone.

One restriction of [GJ18] is that their result only applies to the ferromagnetic Ising model without external fields. The original random cluster formulation of [FK72] does not incorporate external fields, although it is not hard to do so by generalising to a weighted random cluster formulation. Indeed, Park, Jang, Galanis, Shin, Štefankovič, and Vigoda [PJG+17] generalised the SW dynamics $P_{\mathrm{SW}}^{\mathrm{Ising}}$ (see Section 2.2.2 for detailed description) in the presence of external fields. They also showed efficiency of this algorithm in certain parameter regimes and on random graphs. This left open the question if the generalised SW dynamics is efficient in general.

To state our main results, let us first define the *mixing time* of Markov chains, which measures the convergence rate and efficiency of Markov chain based algorithms. Let P be a Markov chain whose stationary distribution is π over the state space Ω. The *mixing time* of P is defined by

$$\forall 0 < \varepsilon < 1, \quad T_{\mathrm{mix}}(P, \varepsilon) = \max_{X_0 \in \Omega} \min \left\{ t \mid d_{\mathrm{TV}}\left(P^t(X_0, \cdot), \pi\right) \le \varepsilon \right\},$$

where $d_{\mathrm{TV}}\left(P^t(X_0,\cdot),\pi\right) = \frac{1}{2}\sum_{\sigma\in\Omega}|P^t(X_0,\sigma)-\pi(\sigma)|$ is the *total variation distance* between the distribution of X_t and the stationary distribution π.

In this paper, we show that the edge-flipping dynamics for the weighted random cluster model mixes in polynomial-time. This implies that the generalised SW dynamics has a polynomial running time for any ferromagnetic Ising model with consistent fields on any graph, answering the question above. This implication is done via adapting [Ull14] to the case with fields.

Theorem 1.1. *For any ferromagnetic Ising model on $G = (V, E)$ with local interaction strength $\beta_e > 1$ for all $e \in E$ and consistent external fields $0 < \lambda_v \le 1$ for all $v \in V$, the mixing time of the (generalised) Swendsen-Wang dynamics satisfies*

$$T_{\mathrm{mix}}(P_{\mathrm{SW}}^{\mathrm{Ising}}, \varepsilon) = O\left(N^4 m^3 \log \frac{\beta_{\max}}{\varepsilon \lambda_{\min}}\right),$$

where $N = \min\{n, \frac{1}{1-\lambda_{\max}}\}$, $n = |V|$, $m = |E|$, $\beta_{\max} = \max_{e\in E}\beta_e$ and $\lambda_{\min} = \min_{v\in V}\lambda_v$.

Also, by applying the monotone CFTP [PW96], we obtain a perfect sampler for such models, which extends the previous result from without fields to with consistent fields.

Theorem 1.2. *There exists a perfect sampler such that given any ferromagnetic Ising model on $G = (V, E)$ with local interaction strength $\beta_e > 1$ for all $e \in E$ and consistent external fields $0 < \lambda_v \le 1$ for all $v \in V$, it returns a perfect sample from π_{Ising} in expected time*

$$O\left(N^4 m^4 \log n \log\left(\beta_{\max} + \frac{\beta_{\min}}{\beta_{\min} - 1}\right)\right),$$

where $N = \min\{n, \frac{1}{1-\lambda_{\max}}\}$, $n = |V|$, $m = |E|$, $\beta_{\max} = \max_{e\in E}\beta_e$ and $\beta_{\min} = \min_{e\in E}\beta_e$.

---

[1] The random cluster model has a parameter $q > 0$. The Ising model corresponds to the case of $q = 2$.

[2] Peres further conjectured that the sharp mixing time bound is $O(|V|^{1/4})$.

Both Theorem 1.1 and Theorem 1.2 indicate that our algorithms are less efficient when there are vertices with no field (namely $\lambda_v = 1$) present. On the other hand, if all vertices have a field bounded away from 1, the running times are much faster.

Corollary 1.3. *Let $0 < \delta < 1$ be a constant. For any ferromagnetic Ising model on graph $G = (V, E)$ with local interaction strength $\beta_e > 1$ for all $e \in E$ and consistent external fields $0 < \lambda_v \le 1 - \delta$ for all $v \in V$,*

- *the mixing time of the Swendsen-Wang dynamics satisfies $T_{\text{mix}}(P_{\text{SW}}^{\text{Ising}}, \varepsilon) = O(m^3 \log \frac{\beta_{\max}}{\varepsilon \lambda_{\min}})$;*
- *the expected running time of the perfect sampler in Theorem 1.2 is $O(m^4 \log n \log(\beta_{\max} + \frac{\beta_{\min}}{\beta_{\min}-1}))$.*

To analyse the mixing time for the edge-flipping dynamics for the weighted random cluster model, we use a lifting argument showing that its spectral gap has the same order as the spectral gap of edge-flipping dynamics for the subgraph-world model. A polynomial bound of the latter has been obtained by Jerrum and Sinclair [JS93]. This lifting argument has been done for the case without field in [GJ18], where the main tool is the coupling between random cluster models and even subgraphs by Grimmett and Janson [GJ07b]. The even subgraphs model is a special case of the subgraph-world model where all vertices must have an even degree in the subgraph. Our main technical contribution is to extend this coupling in the presence of fields. In [GJ07b], it is crucial that the number of even subgraphs has an explicit formula. However, when there are fields, the corresponding subgraph-world allows odd-degree vertices, but with a penalty related to the field. There appears to be no known formula to count the contribution of such subgraphs, and we provide a calculation using Valiant's holographic transformations [Val08] to show a generalised coupling (Lemma 3.3).

The factor $N = \min\{n, \frac{1}{1-\lambda_{\max}}\}$ in Theorem 1.1 and Theorem 1.2 comes from the case where some vertices have fields very close to 1. As λ_v goes to 1, the weight when the degree of v is odd in the subgraph world goes to 0. When these weights are close to 0, the state space of subgraph worlds is very rigid and makes the Markov chain move slowly (or even stop the Markov chain from moving altogether). In order to bypass this issue, we consider a perturbed subgraph world model, where all weights smaller than $1/n$ are replaced by $1/n$. We show that these perturbations only change the marginal probabilities and transition probabilities by constant multiplicative factors (Lemma 3.7 and Lemma 3.8). Thus, the mixing time bound lifted from the perturbed distribution is valid for the original chain up to a constant factor. Once again, to show these marginal probability bounds, we find holographic transformations to be a useful tool.

A natural question is if we can relax the assumptions on the parameters in Theorem 1.1 and Theorem 1.2. For anti-ferromagnetic Ising models, the sampling problem (either approximate or perfect) has no polynomial-time algorithm unless **NP** = **RP**. Even restricted to the ferromagnetic case, Goldberg and Jerrum [GJ07a] showed that the problem becomes #BIS-equivalent when inconsitent fields are allowed, where #BIS stands for counting bipartite independent sets. Its approximation complexity is a major open problem and is usually conjectured to have no polynomial-time algorithm. Thus, it is unlikely to extend the range of parameters in Theorem 1.1 and Theorem 1.2.

Another question is to improve the mixing time in Theorem 1.1 and Theorem 1.2. Essentially, the bottleneck of our results comes from the overhead in the canonical path [JS93] or multicommodity flow method [Sin92]. Unfortunately, there does not seem to be any progress in improving the mixing time bound of these methods in the last three decades. It is an interesting question if recent rapid progress on obtaining optimal mixing times for various models [CLV21, AJK+21, CFYZ22, CE22] may sharpen these classical analysis methods.

The rest of the paper is organised as follows. In Section 2, we introduce the models and a few preliminary results. In Section 3, we bound the mixing time of the edge-flipping dynamics for the weighted random cluster model. In Section 4, we relate the SW dynamics with the edge-flipping dynamics. Finally, in Section 5, we give the perfect sampler via CFTP.

2. Preliminaries

2.1. Models and equivalence result.

2.1.1. *Weighted random cluster model.* The standard random cluster model (at $q = 2$) is equivalent to the ferromagnetic Ising model without external field. To handle Ising models with fields, we need to introduce weights to the random cluster model. Given a graph $G = (V, E)$, the parameters of this model are $\boldsymbol{p} = (p_e)_{e\in E}$ and $\boldsymbol{\lambda} = (\lambda_v)_{v\in V}$, where $0 < p_e < 1$ and $\lambda_v > 0$. The weight of any subset of edges $S \subseteq E$ is given by

$$\mathrm{wt}_{\mathrm{wrc}}(S) := \prod_{e\in S} p_e \prod_{f\in E\setminus S} (1-p_f) \prod_{C\in\kappa(V,S)} \left(1 + \prod_{u\in C} \lambda_u\right), \tag{2}$$

where $\kappa(V, S)$ is the set of all connected components of the graph (V, S), where each $C \in \kappa(V, S)$ is a subset of vertices that forms a connected subgraph. The probability that S is drawn is

$$\pi_{\mathrm{wrc}}(S) = \frac{\mathrm{wt}_{\mathrm{wrc}}(S)}{Z_{\mathrm{wrc}}} \tag{3}$$

where

$$Z_{\mathrm{wrc}} = Z_{\mathrm{wrc}}(G; \boldsymbol{p}, \boldsymbol{\lambda}) := \sum_{S\subseteq E} \mathrm{wt}_{\mathrm{wrc}}(S)$$

is the partition function of the weighted random cluster model. Notice that the weight of each cluster depends on the vertices inside it, which makes this weighted model different from the standard random cluster models at different q's. In the special case $p_e = p \in (0, 1)$ for all $e \in E$ and $\lambda_v = 1$ for all $v \in V$, the weighted random cluster model degenerates to the standard random cluster model.

2.1.2. *Subgraph-world model.* Fix a graph $G = (V, E)$. For any subset of edges $S \subseteq E$, denote by $\mathrm{odd}(S)$ the set of vertices with odd degree in S. The subgraph-world model [JS93] with parameters $\boldsymbol{p} = (p_e)_{e\in E}$ and $\boldsymbol{\eta} = (\eta_v)_{v\in V}$ is defined by following: each subset of edges S has weight

$$\mathrm{wt}_{\mathrm{sg}}(S) := \prod_{e\in S} p_e \prod_{f\in E\setminus S} (1-p_f) \prod_{v\in\mathrm{odd}(S)} \eta_v. \tag{4}$$

The probability that S is drawn is

$$\pi_{\mathrm{sg}}(S) = \frac{\mathrm{wt}_{\mathrm{sg}}(S)}{Z_{\mathrm{sg}}} \tag{5}$$

where

$$Z_{\mathrm{sg}} = Z_{\mathrm{sg}}(G; \boldsymbol{p}, \boldsymbol{\eta}) := \sum_{S\subseteq E} \mathrm{wt}_{\mathrm{sg}}(S)$$

is the partition function of the subgraph-world model. In the special case where $p_e = p \in (0, 1)$ for all $e \in E$ and $\eta_v = 0$ for all $v \in V$, the weight of any subgraph S does not vanish if and only if S is an even subgraph, i.e., $\mathrm{odd}(S) = \varnothing$. This yields the even subgraph model, or the so-called "high-temperature expansion" in the context of statistical mechanics.

2.1.3. *Equivalence of the three models.* We have the following equivalence result among the ferromagnetic Ising model with external fields, the subgraph-world model and the weighted random cluster model. The proof of the equivalence result is given in Appendix A for completeness.

Proposition 2.1. *Given any graph $G = (V, E)$, any $\boldsymbol{\beta} = (\beta_e)_{e\in E}$ and $\boldsymbol{\lambda} = (\lambda_v)_{v\in V}$ satisfying $\beta_e > 1$ for all $e \in E$ and $0 < \lambda_v \le 1$ for all $v \in V$, it holds that*

$$\left(\prod_{e\in E} \beta_e\right) \cdot Z_{\mathrm{wrc}}(G; 2\boldsymbol{p}, \boldsymbol{\lambda}) = Z_{\mathrm{Ising}}(G; \boldsymbol{\beta}, \boldsymbol{\lambda}) = \left(\prod_{v\in V}(1+\lambda_v)\right)\left(\prod_{e\in E} \beta_e\right) Z_{\mathrm{sg}}(G; \boldsymbol{p}, \boldsymbol{\eta}), \tag{6}$$

where $\boldsymbol{p} = (p_e)_{e\in E}$ satisfying $p_e = \frac{1}{2}\left(1 - \frac{1}{\beta_e}\right)$ and $\boldsymbol{\eta} = (\eta_v)_{v\in V}$ satisfying $\eta_v = \frac{1-\lambda_v}{1+\lambda_v}$.

Remark 2.2. For the ferromagnetic Ising model $(G; \boldsymbol{\beta}, \boldsymbol{\lambda}) = (G; \beta, 1)$, where $\beta_e = \beta > 1$ for all $e \in E$ and $\lambda_v = 1$ for all $v \in V$, its relationship with the even subgraph model and the random cluster model is well known [vdW41, FK72, Gri06]. Formally,

$$\beta^{|E|} Z_{\mathrm{wrc}}(G; 2p, 1) = Z_{\mathrm{Ising}}(G; \beta, 1) = 2^{|V|}\beta^{|E|} Z_{\mathrm{sg}}(G; p, 0) \text{ where } p = \frac{1}{2}\left(1 - \frac{1}{\beta}\right),$$

which is a special case of Proposition 2.1.

2.2. **Markov chains.** Let Ω be a state space. Let $(X_t)_{t=1}^{\infty}$ be a Markov chain over Ω. Let $P : \Omega \times \Omega \to \mathbb{R}_{\geq 0}$ denote the transition matrix, i.e. for any $x, y \in \Omega$, $P(x, y) = \mathbf{Pr}[X_t = y \mid X_{t-1} = x]$. We often use transition matrix P to refer the corresponding Markov chain. A distribution π over Ω is a *stationary distribution* of P if $\pi = \pi P$. The Markov chain P is *irreducible* if for any $x, y \in \Omega$, there exists t such that $P^t(x, y) > 0$. The Markov chain P is *aperiodic* if for any $x \in \Omega$, $\gcd\{t \mid P^t(x, x) > 0\} = 1$. If the Markov chain P is both irreducible and aperiodic, then it has a unique stationary distribution. The Markov chain P is *reversible* with respect to distribution π if the following *detailed balance equation* holds

$$\forall x, y \in \Omega, \quad \pi(x)P(x, y) = \pi(y)P(y, x),$$

which implies π is a stationary distribution of P.

Let P be a Markov chain that is reversible with respect to π. The *mixing time* of P is defined by

$$\forall 0 < \varepsilon < 1, \quad T_{\mathrm{mix}}(P, \varepsilon) = \max_{X_0 \in \Omega} \min \left\{ t \mid d_{\mathrm{TV}}\left(P^t(X_0, \cdot), \pi\right) \leq \varepsilon \right\},$$

where $d_{\mathrm{TV}}\left(P^t(X_0, \cdot), \pi\right) = \frac{1}{2} \sum_{\sigma \in \Omega} |P^t(X_0, \sigma) - \pi(\sigma)|$ is the *total variation distance* between the distribution of X_t and the stationary distribution π. The mixing of Markov chain is related to the eigenvalues of the transition matrix. For a reversible Markov chain, its transition matrix P is a self-adjoint operator in the inner product space $\langle \cdot, \cdot \rangle_\pi$:

$$\forall f, g : \Omega \to \mathbb{R}, \quad \langle f, Pg \rangle_\pi = \langle Pf, g \rangle_\pi,$$

where $\langle f, g \rangle_\pi = \sum_{x \in \Omega} \pi(x) f(x) g(x)$. By standard linear algebra results, the matrix P has $|\Omega|$ real eigenvalues $1 = \lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_{|\Omega|}$. The spectral gap of P is defined by

$$\mathfrak{Gap} = 1 - \lambda_2.$$

Let $\mathrm{E}_\pi[f] = \sum_{\sigma \in \Omega} \pi(\sigma) f(\sigma)$ be the *expectation* of f with respect to π, $\mathrm{Var}_\pi[f] = \langle f, f \rangle_\pi - \mathrm{E}_\pi^2[f]$ be the *variance* of f with respect to π, and $\mathcal{E}_P(f, f) = \langle f, (I - P) f \rangle_\pi$ be the *Dirichlet form* of P. The spectral gap has an alternative variational characterisation

$$\mathfrak{Gap} = \mathfrak{Gap}(P) = \inf \left\{ \frac{\mathcal{E}_P(f, f)}{\mathrm{Var}_\pi(f)} \mid f : \Omega \to \mathbb{R}, \mathrm{Var}_\pi(f) \neq 0 \right\}. \tag{7}$$

Without loss of generality, we may assume that P only has non-negative eigenvalues, as otherwise we can consider its lazy version $\frac{1}{2}(I + P)$, where I is the identity matrix. For a positive semidefinite transition matrix P, the following relation between mixing time and spectral gap is well-known [LP17]

$$T_{\mathrm{mix}}(P, \varepsilon) \leq \frac{1}{\mathfrak{Gap}} \left(\log \frac{1}{\pi_{\min}} + \log \frac{1}{\varepsilon} \right), \tag{8}$$

where $\pi_{\min} = \min_{\sigma \in \Omega} \pi(\sigma)$ is the minimum probability in π.

2.2.1. *Metropolis algorithm.* Let μ be a joint distribution over the state space Ω. Let $\Psi : \Omega \times \Omega \to \mathbb{R}_{\geq 0}$ be a symmetric (i.e. $\Psi(x, y) = \Psi(y, x)$) stochastic matrix, where each row $\Psi(x, \cdot)$ is the *proposal distribution* for state x. Suppose the current state is $X_t = x \in \Omega$. For the transition step $X_t \to X_{t+1}$ of the Metropolis algorithm, a candidate state y is sampled from distribution $\Psi(x, \cdot)$. Then

- with probability $a(x, y) = \min\{1, \frac{\mu(y)}{\mu(x)}\}$, accept the candidate state and let $X_{t+1} = y$;
- with probability $1 - a(x, y)$, reject the candidate state and let $X_{t+1} = X_t = x$.

It is straightforward to verify that the Metropolis algorithm for μ satisfies the detailed balance equation with respect to μ. We say a Markov chain is the *lazy Metropolis algorithm* for μ if for each transition step $X_t \to X_{t+1}$,

- with probability $1/2$, the chain is lazy, i.e. $X_{t+1} = X_t$;
- with probability $1/2$, the chain evolves as the Metropolis algorithm for μ.

Edge flipping dynamics for weighted random cluster models. We call the lazy Metropolis algorithm for weighted random cluster models the *edge flipping dynamics* for weighted random cluster models. Here the proposal distribution is to propose a change of a uniformly at random chosen edge. Let $G = (V, E)$ be a graph with $m = |E|$ edges. Consider the weighted random cluster model with parameters $\boldsymbol{p} = (p_e)_{e\in E}$ and $\boldsymbol{\lambda} = (\lambda_v)_{v\in V}$, where $0 < p_e < 1$ and $\lambda_v > 0$. Let π_{wrc} denote the distribution over $\Omega = 2^E$ of the weighted random cluster model. The transition matrix P_{wrc} of the edge flipping dynamics is given by

$$\forall x, y \in \Omega, \quad P_{\mathrm{wrc}}(x, y) = \begin{cases} \frac{1}{2m} \min\left\{1, \frac{\pi_{\mathrm{wrc}}(y)}{\pi_{\mathrm{wrc}}(x)}\right\}, & \text{if } |x \oplus y| = 1; \\ 1 - \frac{1}{2m} \sum_{e\in E} \min\left\{1, \frac{\pi_{\mathrm{wrc}}(x\oplus\{e\})}{\pi_{\mathrm{wrc}}(x)}\right\} & \text{if } x = y; \\ 0 & \text{otherwise.} \end{cases} \tag{9}$$

Here, we use $A \oplus B = (A \setminus B) \cup (B \setminus A)$ to denote the symmetric difference between sets A and B.

Edge flipping dynamics for subgraph-world models. We call the lazy Metropolis algorithm for subgraph-world models the *edge flipping dynamics* for subgraph-world models. Again the proposal distribution is to propose a change of a uniformly at random chosen edge. Consider the subgraph-world model on $G = (V, E)$ with parameters $\boldsymbol{p} = (p_e)_{e\in E}$ and $\boldsymbol{\eta} = (\eta_v)_{v\in V}$. Let π_{sg} denote the distribution over $\Omega = 2^E$ induced by the subgraph-world model. The transition matrix P_{sg} of the edge flipping dynamics is defined by

$$\forall x, y \in \Omega, \quad P_{\mathrm{sg}}(x, y) = \begin{cases} \frac{1}{2m} \min\left\{1, \frac{\pi_{\mathrm{sg}}(y)}{\pi_{\mathrm{sg}}(x)}\right\}, & \text{if } |x \oplus y| = 1; \\ 1 - \frac{1}{2m} \sum_{e\in E} \min\left\{1, \frac{\pi_{\mathrm{sg}}(x\oplus\{e\})}{\pi_{\mathrm{sg}}(x)}\right\} & \text{if } x = y; \\ 0 & \text{otherwise.} \end{cases}$$

2.2.2. *Swendsen-Wang dynamics.* Let $G = (V, E)$ be a graph. Consider the ferromagnetic Ising model on G with parameters $\boldsymbol{\beta} = (\beta_e)_{e\in E}$ and $\boldsymbol{\lambda} = (\lambda_v)_{v\in V}$, where $\beta_e > 1$ for all $e \in E$, and the weighted random cluster model on G with parameters $\boldsymbol{p} = (p_e)_{e\in E}$ and $\boldsymbol{\lambda} = (\lambda_v)_{v\in V}$, where $p_e = 1 - \frac{1}{\beta_e}$ for all $e \in E$. Recall π_{Ising} from (1) and π_{WRC} from (3).

Define the following two transformations between Ising and weighted random cluster models.

- $P_{\mathcal{I}\to\mathcal{R}} : \{0,1\}^V \to 2^E$: Given any Ising configuration $\sigma \in \{0,1\}^V$, $P_{\mathcal{I}\to\mathcal{R}}$ transforms σ into a weighted random cluster model configuration $S \subseteq E$. For each edge $e = \{u, v\} \in E$ with $\sigma(u) = \sigma(v)$, add e independently into S with probability $p_e = 1 - \frac{1}{\beta_e}$. Formally,

$$\forall \sigma \in \{0,1\}^V, S \subseteq E, \quad P_{\mathcal{I}\to\mathcal{R}}(\sigma, S) = \mathbb{I}[S \subseteq M(\sigma)] \cdot \prod_{e\in S} \left(1 - \frac{1}{\beta_e}\right) \cdot \prod_{f\in M(\sigma)\setminus S} \frac{1}{\beta_f}, \tag{10}$$

  where $M(\sigma) = \{e = \{u, v\} \in E \mid \sigma_u = \sigma_v\}$ is the set of monochromatic edges with respect to σ.
- $P_{\mathcal{R}\to\mathcal{I}} : \{0,1\}^E \to \{0,1\}^V$: Given any weighted random cluster model configuration $S \subseteq E$, $P_{\mathcal{R}\to\mathcal{I}}$ transforms S to an Ising configuration $\sigma \in \{0,1\}^V$. For each connected component $C \subseteq V$ in graph $G' = (V, S)$, sample $x_C \in \{0,1\}$ independently according to the following distribution

$$x_C = \begin{cases} 1 & \text{with probability } \frac{\prod_{v\in C}\lambda_v}{1+\prod_{v\in C}\lambda_v}; \\ 0 & \text{with probability } \frac{1}{1+\prod_{v\in C}\lambda_v}, \end{cases}$$

  and then let $\sigma(v) = x_C$ for all vertices $v \in C$. Formally,

$$\forall \sigma \in \{0,1\}^V, S \subseteq E, \quad P_{\mathcal{R}\to\mathcal{I}}(S, \sigma) = \mathbb{I}[S \subseteq M(\sigma)] \cdot \prod_{C\in\kappa(V,S)} \frac{\prod_{v\in C}\lambda_v^{\sigma(v)}}{1 + \prod_{v\in C}\lambda_v}, \tag{11}$$

  where $\kappa(V, S)$ is the set of connected components in graph $G' = (V, S)$.

The Swendsen-Wang dynamics for Ising models is defined by

$$P_{\mathrm{SW}}^{\mathrm{Ising}} := P_{\mathcal{I}\to\mathcal{R}} P_{\mathcal{R}\to\mathcal{I}}, \tag{12}$$

and the Swendsen-Wang dynamics for weighted random cluster models is defined by

$$P_{\mathrm{SW}}^{\mathrm{wrc}} := P_{\mathcal{R}\to\mathcal{I}} P_{\mathcal{I}\to\mathcal{R}}. \tag{13}$$

The following proposition about Swendsen-Wang dynamics is well-known. However, here we consider more general Ising models with external fields and weighted random cluster models. For completeness, we prove of the following proposition in Appendix C.

Proposition 2.3. *The following results hold for $P_{\mathcal{I}\to\mathcal{R}}$, $P_{\mathcal{R}\to\mathcal{I}}$, $P_{\mathrm{SW}}^{\mathrm{Ising}}$ and $P_{\mathrm{SW}}^{\mathrm{wrc}}$:*

- *for any $\sigma \in \{0,1\}^V$, $S \subseteq 2^E$, $\pi_{\mathrm{Ising}}(\sigma) P_{\mathcal{I}\to\mathcal{R}}(\sigma, S) = \pi_{\mathrm{wrc}}(S) P_{\mathcal{R}\to\mathcal{I}}(\sigma)$.*
- *$\pi_{\mathrm{Ising}} P_{\mathcal{I}\to\mathcal{R}} = \pi_{\mathrm{wrc}}$ and $\pi_{\mathrm{wrc}} P_{\mathcal{R}\to\mathcal{I}} = \pi_{\mathrm{Ising}}$;*
- *$P_{\mathrm{SW}}^{\mathrm{Ising}}$ is reversible with respect to π_{Ising}; $P_{\mathrm{SW}}^{\mathrm{wrc}}$ is reversible with respect to π_{wrc}.*
- *$P_{\mathrm{SW}}^{\mathrm{Ising}}$ and $P_{\mathrm{SW}}^{\mathrm{wrc}}$ are both positive semidefinite.*

The proposition above shows that both $P_{\mathrm{SW}}^{\mathrm{Ising}}$ and $P_{\mathrm{SW}}^{\mathrm{wrc}}$ have non-negative real eigenvalues. By definitions in (12) and (13), the following equation is straightforward to verify

$$\mathfrak{Gap}\left(P_{\mathrm{SW}}^{\mathrm{Ising}}\right) = \mathfrak{Gap}\left(P_{\mathrm{SW}}^{\mathrm{wrc}}\right). \tag{14}$$

2.3. **Canonical path and multicommodity flow.** A useful tool to analyse the spectral gap of a reversible Markov chains is the canonical path introduced by Jerrum and Sinclair [JS89] and its later multicommodity flow generalisation. Let P be a reversible Markov chain over the state space Ω with stationary distribution π. Let $\gamma_{xy} = (z_0 = x, z_1, z_2, \dots, z_\ell = y)$ be a path moving in the state space using transitions of P, i.e. for any $1 \le i < \ell$, $P(z_i, z_{i+1}) > 0$. Denote by $x \overset{\gamma}{\rightsquigarrow} y$ that γ moves from x to y. Let Γ be a flow, i.e., a collection of paths, with each path $\gamma \in \Gamma$ assigned a weight $w(\gamma)$ satisfying the following condition

$$\forall x, y \in \Omega, \quad \pi(x)\pi(y) = \sum_{\gamma \in \Gamma : x \overset{\gamma}{\rightsquigarrow} y} w(\gamma). \tag{15}$$

The congestion of the flow Γ is defined by

$$\varrho(\Gamma) := \max_{(z,z')\in\Omega^2, P(z,z')>0} \frac{L}{\pi(z)P(z,z')} \sum_{\gamma\in\Gamma:(z,z')\in\gamma} w(\gamma) \tag{16}$$

where L is the maximum length of path in Γ. The original canonical path formulation is a special case where there is exactly one path γ_{xy} from x to y whose weight is $\pi(x)\pi(y)$.

Sinclair [Sin92] showed that the congestion of any flow Γ for a Markov chain P is a upper bound of the inverse of its spectral gap, namely, $\frac{1}{\mathfrak{Gap}(P)} \le \varrho(\Gamma)$. Furthermore, if P is positive semidefinite, by (8),

$$T_{\mathrm{mix}}(P, \varepsilon) \le \varrho(\Gamma)\left(\log \frac{1}{\pi_{\min}} + \log \frac{1}{\varepsilon}\right).$$

2.4. **Holographic transformation.** We will need holographic transformations to show couplings between the subgraph-world model and the weighted random cluster model. Let $f : \{0,1\}^d \to \mathbb{C}$ be a function. We may represent it by a vector (either row or column vector) $\left(f_0, \cdots, f_x, \cdots, f_{2^d-1}\right)$ where f_x is the value of f on $x \in \{0,1\}^d$ by regarding x as a binary representation. In the symmetric case where f is invariant under permutations of indices, we use a succinct "signature" $[f_0, \cdots, f_w, \cdots, f_d]$ to express f, where f_w is the value of f on inputs of Hamming weight w, i.e. all $x \in \{0,1\}^d$ satisfying $|x| = w$.

Given a bipartite graph $H = (V, E)$ with partition $V = V_1 \uplus V_2$. Let $\mathcal{F} = (f_v)_{v\in V_1}$ and $\mathcal{G} = (g_v)_{v\in V_2}$ be two sets of functions such that the arity of the function is the degree of the corresponding vertex. The *Holant* (an edge weighted partition function) is defined by

$$\mathrm{Holant}(H; \mathcal{F} \mid \mathcal{G}) := \sum_{\sigma: E\to\{0,1\}} \prod_{v\in V_1} f_v\left(\sigma\mid_{E(v)}\right) \prod_{u\in V_2} g_u\left(\sigma\mid_{E(u)}\right),$$

where $\sigma\mid_{E(v)}$ stands for the restriction of the assignment σ to the incident edges of v.

Let $\boldsymbol{M}$ be a 2×2 matrix and f be a function of arity d. If f is represented by a column (resp. row) vector, we write $\boldsymbol{M}f = \boldsymbol{M}^{\otimes d} f$ (resp. $f\boldsymbol{M} = f\boldsymbol{M}^{\otimes d}$) as the transformed signature. Given $\mathrm{Holant}(H; \mathcal{F} \mid \mathcal{G})$ and an invertible matrix $\boldsymbol{T} \in \mathbb{C}^{2\times 2}$, we view signatures in $\mathcal{F}$ as row vectors and define $\mathcal{F}\boldsymbol{T} = \{f'_v \mid v \in V_1 \wedge f'_v = f_v \boldsymbol{T}\}$; and view signatures in $\mathcal{G}$ as column vectors and define $\boldsymbol{T}^{-1}\mathcal{G} = \{g'_v \mid v \in V_2 \wedge g'_v = \boldsymbol{T}^{-1} g_v\}$. Valiant's celebrated Holant Theorem [Val08] states

Theorem 2.4. $\mathrm{Holant}(H; \mathcal{F} \mid \mathcal{G}) = \mathrm{Holant}(H; \mathcal{F}\boldsymbol{T} \mid \boldsymbol{T}^{-1}\mathcal{G})$.

3. Mixing of edge flipping dynamics

Let P_{EF}^{wrc} denote the edge flipping dynamics for weighted random cluster model. Note that P_{EF}^{wrc} is lazy, implying that it is positive semi-definite. We prove the following result in this section.

Lemma 3.1. *For any weighted random cluster model on graph $G = (V, E)$ with parameters $2\boldsymbol{p} = (2p_e)_{e\in E}$ and $\boldsymbol{\lambda} = (\lambda_v)_{v\in V}$, if $0 < p_e < \frac{1}{2}$ for all $e \in E$ and $0 < \lambda_v \le 1$ for all $v \in V$. Let $b = \max_{e\in E} \max\{\frac{1}{2p_e}, \frac{1}{1-2p_e}\}$. Let $N = \min\{n, \frac{1}{1-\lambda_{\max}}\}$, where $\lambda_{\max} = \max_{v\in V} \lambda_v$. Then the spectral gap for P_{EF}^{wrc} is $\Omega(\frac{1}{N^4 m^2})$ and the mixing time satisfies $T_{\mathrm{mix}}(P_{EF}^{\mathrm{wrc}}, \varepsilon) = O\left(N^4 m^3 \log \frac{b}{\varepsilon}\right)$.*

Lemma 3.1 is proved by two steps. We first prove a weaker result in Lemma 3.2 that works for the case $0 < \lambda_v < 1$ for all $v \in V$. In Section 3.3, we boost the weaker result via a comparison argument.

Lemma 3.2. *For any weighted random cluster model on graph $G = (V, E)$ with parameters $2\boldsymbol{p} = (2p_e)_{e\in E}$ and $\boldsymbol{\lambda} = (\lambda_v)_{v\in V}$, if $0 < p_e < \frac{1}{2}$ for all $e \in E$ and $0 < \lambda_v < 1$ for all $v \in V$, there exists a multicommodity flow Γ for P_{EF}^{wrc} such that $\varrho(\Gamma) \le \frac{10m^2}{a^4}$, where $m = |E|$ and $a = \min_{v\in V} \frac{1-\lambda_v}{1+\lambda_v}$. As a consequence, the spectral gap of P_{EF}^{wrc} is at least $\frac{a^4}{10m^2}$.*

The multicommodity flow in Lemma 3.2 is obtained by lifting canonical paths for subgraph-world models. In Section 3.1, we give a new coupling result between subgraph-world models and weighted random cluster models. In Section 3.2, we use this coupling to lift canonical paths and prove Lemma 3.2.

3.1. Coupling via holographic transformations.

Under the unweighted setting, Grimmett and Janson [GJ07b, Thm 3.5] discovered a coupling between random even subgraphs and random cluster configurations. The following lemma is a generalisation to the weighted case via holographic transformations.

Lemma 3.3. *Let $G = (V, E)$ be a graph, $\boldsymbol{p} = (p_e)_{e\in E}$ and $\boldsymbol{\eta} = (\eta_v)_{v\in V}$, where $0 \le p_e < 1/2$ for all $e \in E$ and $\eta_v \ge 0$ for all $v \in V$. Let $\mathcal{S} \subseteq E$ be a random sample from the subgraph-world model $(G; \boldsymbol{p}, \boldsymbol{\eta})$. Let $\mathcal{R}$ be $\mathcal{S}$ with each remaining edge $e \in E \setminus \mathcal{S}$ added into $\mathcal{R}$ independently with probability $p_e/(1-p_e)$. Then the random subgraph $\mathcal{R}$ satisfies the distribution of the weighted random cluster model with parameter $(2\boldsymbol{p}, \boldsymbol{\lambda})$ where $\eta_v = \frac{1-\lambda_v}{1+\lambda_v}$ for all $v \in V$.*

We need the following lemma to prove Lemma 3.3.

Lemma 3.4. *Let $G = (V, E)$ be a graph. Let $\boldsymbol{\lambda} = (\lambda_v)_{v\in V}$ where $0 \le \lambda_v < 1$ for all $v \in V$. For each $v \in V$, let $\eta_v = \frac{1-\lambda_v}{1+\lambda_v}$. It holds that*

$$\prod_{C\in\kappa(V,E)} \left(1 + \prod_{u\in C} \lambda_u\right) = \left(\prod_{v\in V} (1+\lambda_v)\right) \left(\frac{1}{2}\right)^{|E|} \sum_{E'\subseteq E} \prod_{u\in odd(E')} \eta_u, \tag{17}$$

where $\kappa(V, E)$ is the set of connected components in graph $G = (V, E)$.

Proof. Define a bipartite graph H with left part $V_1 = V$ corresponding to vertices in G and right part $V_2 = E$ corresponding to edges in G. Two vertices $v \in V_1$ and $e \in V_2$ are adjacent in H if v is incident to e in G. Let d_v denote the degree of v in G. Consider the following set of signatures

$$\begin{aligned}
\mathcal{F}^{(1)} &= \left\{ f_v^{(1)} = [1,0]^{\otimes d_v} + \lambda_v [0,1]^{\otimes d_v} \mid v \in V \right\}, \\
\mathcal{F}^{(2)} &= \left\{ f_v^{(2)} = \frac{1}{1+\lambda_v} \left([1,1]^{\otimes d_v} + \lambda_v [1,-1]^{\otimes d_v} \right) \mid v \in V \right\}, \\
\mathcal{G} &= \{ g_e = [1,0,1] \mid e \in E \}.
\end{aligned}$$

We remark that $f_v^{(2)} = [1, \eta_v, 1, \eta_v, \dots]$. Let $\boldsymbol{T} = \left(\begin{smallmatrix} 1 & 1 \\ 1 & -1 \end{smallmatrix}\right)$. Observe that $f_v^{(1)}\boldsymbol{T} = (1+\lambda_v) f_v^{(2)}$ and $\boldsymbol{T}^{-1} g_e = \frac{1}{2} g_e$. By Theorem 2.4, it holds that

$$\text{Holant}\left(H; \mathcal{F}^{(1)} \mid \mathcal{G}\right) = \left(\prod_{v \in V} (1+\lambda_v) \right) \left(\frac{1}{2}\right)^{|E|} \text{Holant}\left(H; \mathcal{F}^{(2)} \mid \mathcal{G}\right). \tag{18}$$

This equation is indeed (17) in disguise. The equivalence between the left-hand sides of (18) and (17) is a simple observation that the signature $[1,0,1]$ on the edge forces the spins of vertices in each connected component C to be the same. Each component contributes a weight $1 + \prod_{u \in C} \lambda_u$. The equivalence between the right-hand sides of (18) and (17) follows from how $\mathcal{F}^{(2)}$ and $\mathcal{G}$ are defined. This proves the lemma. □

Proof of Lemma 3.3. For each subgraph $R \subseteq E$ of $G = (V, E)$,

$$\begin{aligned}
\mathbf{Pr}[\mathcal{R} = R] &= \frac{1}{Z_{\text{sg}}(G; \boldsymbol{p}, \boldsymbol{\eta})} \sum_{S \subseteq R} \prod_{u \in \text{odd}(S)} \eta_u \prod_{e \in S} p_e \prod_{f \in E \setminus S} (1 - p_f) \prod_{g \in R \setminus S} \frac{p_g}{1-p_g} \prod_{h \in E \setminus R} \frac{1 - 2p_h}{1 - p_h} \\
&= \frac{1}{Z_{\text{sg}}(G; \boldsymbol{p}, \boldsymbol{\eta})} \sum_{S \subseteq R} \prod_{u \in \text{odd}(S)} \eta_u \prod_{e \in R} p_e \prod_{f \in E \setminus R} (1 - 2p_f) \\
&= \frac{1}{Z_{\text{sg}}(G; \boldsymbol{p}, \boldsymbol{\eta})} 2^{-|R|} \prod_{e \in R} (2p_e) \prod_{f \in E \setminus R} (1 - 2p_f) \sum_{S \subseteq R} \prod_{u \in \text{odd}(S)} \eta_u \\
\text{(By (17) on } (V,R)) \quad &= \frac{1}{Z_{\text{sg}}(G; \boldsymbol{p}, \boldsymbol{\eta})} \prod_{e \in R} (2p_e) \prod_{f \in E \setminus R} (1 - 2p_f) \prod_{v \in V} \frac{1}{1+\lambda_v} \prod_{C \in \kappa(V,R)} \left(1 + \prod_{u \in C} \lambda_u \right) \\
\text{(By (6))} \quad &= Z_{\text{wrc}}(G; 2\boldsymbol{p}, \boldsymbol{\lambda}) \prod_{e \in R} (2p_e) \prod_{f \in E \setminus R} (1 - 2p_f) \prod_{C \in \kappa(V,R)} \left(1 + \prod_{u \in C} \lambda_u \right). \\
&= \pi_{\text{wrc}}(R).
\end{aligned}$$

□

3.2. **Lifting from subgraph-world.** Let $P_{\text{EF}}^{\text{sg}}$ denote the edge flipping dynamics for subgraph-world model. Given a set of canonical paths Γ_{sg} of $P_{\text{EF}}^{\text{sg}}$, we construct a flow Γ_{wrc} for $P_{\text{EF}}^{\text{wrc}}$, using the coupling Lemma 3.3. This is the adaption of the same idea as Guo and Jerrum [GJ18] while we are working on different dynamics. The construction contains two stages, the lifting stage from Γ_{sg} to Γ'_{wrc}, and the rerandomising stage from Γ'_{wrc} to Γ_{wrc}. Prior to that, we have to construct a set of canonical paths for the subgraph-world model for our setting, where the fugacity on odd-degree vertices may not necessarily be uniform. This is a straightforward generalisation of the original argument in [JS93]. Its proof is given in Appendix B for completeness.

Lemma 3.5. *For any subgraph-world model on graph $G = (V, E)$ with parameters $\boldsymbol{p} = (p_e)_{e \in E}$ and $\boldsymbol{\eta} = (\eta_v)_{v \in V}$, if $0 < \eta_v < 1$ for all $v \in V$, then there exists a set of canonical paths $\gamma_{\text{sg}} = \{\gamma_{X,Y} : X, Y \subseteq E\}$ for the $P_{\text{EF}}^{\text{sg}}$ such that (1) $w_{\text{sg}}(\gamma_{X,Y}) = \pi_{\text{sg}}(X)\pi_{\text{sg}}(Y)$; (2) $|\gamma_{X,Y}| \leq m$; (3) for any transition (Z, Z') with $|Z \oplus Z'| = 1$, it holds that*

$$\sum_{\gamma : (Z, Z') \in \gamma} w_{\text{sg}}(\gamma) \leq \eta_{\min}^{-4} \pi_{\text{sg}}(Z) \text{ and } \sum_{\gamma : (Z, Z') \in \gamma} w_{\text{sg}}(\gamma) \leq \eta_{\min}^{-4} \pi_{\text{sg}}(Z'),$$

where $\eta_{\min} := \min_v \eta_v$.

Next we use Lemma 3.5 to prove Lemma 3.2.

3.2.1. *Construction of* Γ_{wrc}. We now describe the construction of multicommodity flow for $P_{\mathrm{EF}}^{\mathrm{wrc}}$. For the lifting stage, fix any canonical path $\gamma = (S_0, \cdots, S_\ell)$ in Γ_{sg}. We lift γ to a flow with some weight function, which can virtually be regarded as a random path $\mathcal{Z} = (\mathcal{Z}_0, \cdots, \mathcal{Z}_\ell)$.

We first construct the initial state $\mathcal{Z}_0$ by adding each remaining edge $e \notin S_0$ with probability $p'_e := p_e/(1-p_e)$[3]. This is equivalent to drawing a superset $\mathcal{Z}_0 \supseteq S_0$ according to the distribution

$$\mathbf{Pr}_\gamma[\mathcal{Z}_0 = Z] = \delta(S_0, Z) \qquad \text{where } \delta(S, Z) = \prod_{e \in Z \setminus S} p'_e \prod_{f \in E \setminus Z} (1 - p'_f).$$

Here $\mathbf{Pr}_\gamma$ emphasis that the construction is with respect to a fixed canonical path γ. We may drop the subscript γ when it is clear from context. Using the same notation, Lemma 3.3 implies

$$\pi_{\mathrm{wrc}}(Z) = \sum_{S \subseteq Z} \pi_{\mathrm{sg}}(S)\delta(S, Z). \tag{19}$$

For any $1 \le k \le \ell$, one possible way to construct $\mathcal{Z}_k$ based on S_k is to lift from S_k such that $\mathbf{Pr}[\mathcal{Z}_k = Z] = \delta(S_k, Z)$. However, $\mathcal{Z}_{k-1}$ and $\mathcal{Z}_k$ may differ by a lot of edges and hence it is not a transition in the edge-flipping dynamics. A potential fix is to unwind the edges in $\mathcal{Z}_{k-1} \oplus \mathcal{Z}_k$ one by one, at the cost of at most m extra intermediate steps. To bypass this extra m, we mimic the transition from S_{k-1} to S_k while ensuring

$$\mathbf{Pr}[\mathcal{Z}_k = Z] = \delta(S_k, Z) \quad \forall Z \supseteq S_k. \tag{20}$$

We do this inductively. Assume (20) is true for $k-1$. In the Markov chain $P_{\mathrm{EF}}^{\mathrm{sg}}$ for the subgraph-world model, the transition from S_{k-1} to S_k is either one of the following cases:

- $S_k = S_{k-1} + e$ where $e \notin S_{k-1}$. In this case, we simply construct $\mathcal{Z}_k = \mathcal{Z}_{k-1} + e$. Be cautious that $\mathcal{Z}_{k-1}$ may already contain the edge e. Verify (20) by a direct calculation using induction hypothesis that

$$\begin{aligned}\mathbf{Pr}[\mathcal{Z}_k = Z] &= \mathbf{Pr}[\mathcal{Z}_{k-1} = Z] + \mathbf{Pr}[\mathcal{Z}_{k-1} = Z - e] \\ &= \delta(S_{k-1}, Z) + \delta(S_{k-1}, Z - e) = \delta(S_k, Z)p'_e + \delta(S_k, Z)(1 - p'_e) = \delta(S_k, Z).\end{aligned}$$

- $S_k = S_{k-1} - e$ where $e \in S_{k-1}$. In this case, with probability $1 - p'_e$ we remove e from $\mathcal{Z}_{k-1}$ to obtain $\mathcal{Z}_k$ (be cautious again that $\mathcal{Z}_{k-1} \supseteq S_{k-1}$ must contain e), and with probability p'_e, just keep $\mathcal{Z}_k$ as $\mathcal{Z}_{k-1}$. Verify (20) as follows:
  - $e \in Z$ where $Z \supseteq S_k$. This mean that $\mathcal{Z}_k$ is obtained in the latter way, which gives

$$\mathbf{Pr}[\mathcal{Z}_k = Z] = \mathbf{Pr}[\mathcal{Z}_{k-1} = Z]p'_e = \delta(S_{k-1}, Z)p'_e = \delta(S_k, Z).$$

  - $e \notin Z$. This means e is removed, and therefore

$$\mathbf{Pr}[\mathcal{Z}_k = Z] = \mathbf{Pr}[\mathcal{Z}_{k-1} = Z + e](1 - p'_e) = \delta(S_{k-1}, Z + e)(1 - p'_e) = \delta(S_k, Z).$$

In this way, we construct a random path each of whose states is a random variable from a fixed path γ for subgraph-world model. For a fixed path $\xi = \{Z_0, \cdots Z_\ell\}$ in Γ'_{wrc}, note that it may arise from multiple different canonical paths γ of Γ_{sg}.

The problem of Γ'_{wrc} is that it is difficult to find a weight function w_{wrc} to satisfy (15) due to the dependency between $\mathcal{Z}_0$ and $\mathcal{Z}_\ell$ (see below), and hence not a valid flow for the random cluster dynamics, despite the marginal distribution of $\mathcal{Z}_\ell$ being correct. In the second stage, we remove such dependency by rerandomising every edge not in S_ℓ whilst keeping the correct marginal distribution. Take $\{e_1, \cdots, e_k\} = E \setminus S_\ell$ to be the set of edges not present in S_ℓ. In each step from $\mathcal{Z}_\ell$ onwards, given $\mathcal{Z}_{\ell+i-1}$ where $1 \le i \le k$, construct $\mathcal{Z}_{\ell+i}$ adhering to the following rule.

- If $e_i \in \mathcal{Z}_{\ell+i-1}$, then with probability $1 - p'_{e_i}$, exclude e_i to obtain $\mathcal{Z}_{\ell+i}$, and with probability p'_{e_i}, keep $\mathcal{Z}_{\ell+i}$ to be the same as $\mathcal{Z}_{\ell+i-1}$.
- If $e_i \notin \mathcal{Z}_{\ell+i-1}$, then with probability $1 - p'_{e_i}$, keep $\mathcal{Z}_{\ell+i}$ to be the same as $\mathcal{Z}_{\ell+i-1}$, and with probability p'_{e_i}, include e_i to obtain $\mathcal{Z}_{\ell+i}$.

[3]The probability p'_e is valid because we only consider $0 < p_e < \frac{1}{2}$.

We first verify that the rerandomisation of each edge keeps the marginal distribution, i.e.,

$$\mathbf{Pr}[\mathcal{Z}_{\ell+i} = Z] = \delta(S_\ell, Z) \tag{21}$$

for all $1 \leq i \leq k$. This follows from a simple induction that

- if $e_i \in Z$, then

$$\begin{aligned}\mathbf{Pr}[\mathcal{Z}_{\ell+i} = Z] &= \mathbf{Pr}[\mathcal{Z}_{\ell+i-1} = Z]p'_{e_i} + \mathbf{Pr}[\mathcal{Z}_{\ell+i-1} = Z - e_i]p'_{e_i} \\ &= \delta(S_\ell, Z)p'_{e_i} + \delta(S_\ell, Z - e_i)p'_{e_i} = \delta(S_\ell, Z)p'_{e_i} + \delta(S_\ell, Z)(1 - p'_{e_i}) = \delta(S_\ell, Z),\end{aligned}$$

  and
- if $e_i \notin Z$, then

$$\begin{aligned}\mathbf{Pr}[\mathcal{Z}_{\ell+i} = Z] &= \mathbf{Pr}[\mathcal{Z}_{\ell+i-1} = Z](1 - p'_{e_i}) + \mathbf{Pr}[\mathcal{Z}_{\ell+i-1} = Z + e_i](1 - p'_{e_i}) \\ &= \delta(S_\ell, Z)(1 - p'_{e_i}) + \delta(S_\ell, Z + e_i)(1 - p'_{e_i}) = \delta(S_\ell, Z)(1 - p'_{e_i}) + \delta(S_\ell, Z)p'_{e_i} = \delta(S_\ell, Z),\end{aligned}$$

with the base case that trivially holds. To see why the rerandomisation removes the correlation conditioned on γ, for each edge e_i not in S_ℓ, the event that it is present in $\mathcal{Z}_{\ell+i}$ onwards is independent of $\mathcal{Z}_{\ell+i-1}$ because

$$\mathbf{Pr}_\gamma[e_i \in \mathcal{Z}_{\ell+i} \mid e_i \in \mathcal{Z}_{\ell+i-1} = Z] = p'_{e_i}$$

for any Z.

To construct Γ_{wrc}, for any path $\zeta = \{Z_0, \cdots Z_{\ell+k}\}$ in Γ_{wrc}, assign its weight to be

$$w_{\mathrm{wrc}}(\zeta) = \sum_\gamma w_{\mathrm{sg}}(\gamma)\,\mathbf{Pr}_\gamma[Z = \zeta],$$

where w_{sg} is the weight function in Lemma 3.5. We argue that Γ_{wrc} is a valid flow by checking (15) that

$$\begin{aligned}
\sum_{\zeta:X \overset{\zeta}{\rightsquigarrow} Y} w_{\mathrm{wrc}}(\zeta) &= \sum_{\zeta:X \overset{\zeta}{\rightsquigarrow} Y} \sum_{\gamma\in\Gamma_{\mathrm{sg}}} w_{\mathrm{sg}}(\gamma)\,\mathbf{Pr}_\gamma[\mathcal{Z} = \zeta] \\
&= \sum_{S\subseteq X, S'\subseteq Y} \sum_{\gamma:S \overset{\gamma}{\rightsquigarrow} S'} w_{\mathrm{sg}}(\gamma)\,\mathbf{Pr}_\gamma[\mathcal{Z}_0 = X, \mathcal{Z}_{\ell+k} = Y] \quad \text{where } k = |E \setminus S'| \\
\text{(By independence)} \quad &= \sum_{S\subseteq X, S'\subseteq Y} \sum_{\gamma:S \overset{\gamma}{\rightsquigarrow} S'} w_{\mathrm{sg}}(\gamma)\,\mathbf{Pr}_\gamma[\mathcal{Z}_0 = X]\,\mathbf{Pr}_\gamma[\mathcal{Z}_{\ell+k} = Y] \\
\text{(By (20) and (21))} \quad &= \sum_{S\subseteq X, S'\subseteq Y} \sum_{\gamma:S \overset{\gamma}{\rightsquigarrow} S'} w_{\mathrm{sg}}(\gamma)\delta(S, X)\delta(S', Y) \\
\text{(By validity of } \Gamma_{\mathrm{sg}}) \quad &= \sum_{S\subseteq X, S'\subseteq Y} \pi_{\mathrm{sg}}(S)\pi_{\mathrm{sg}}(S')\delta(S, X)\delta(S', Y) \\
&= \left(\sum_{S\subseteq X} \pi_{\mathrm{sg}}(S)\delta(S, X)\right)\left(\sum_{S'\subseteq Y} \pi_{\mathrm{sg}}(S')\delta(S', Y)\right) \\
\text{(By (19))} \quad &= \pi_{\mathrm{wrc}}(X)\pi_{\mathrm{wrc}}(Y).
\end{aligned}$$

Notice that in the construction above, there are transitions that do not change anything in the flow. These transitions can be removed without changing the congestion, but we keep them for ease of analysis for congestion later.

3.2.2. *Traffic through a single transition.* For each pair of (Z, Z') in the state space of weighted random cluster models, there are three different cases. In any of them, the traffic from Z to Z' in the flow Γ_{wrc} emerges from two sources, either the lifting stage or the rerandomisation. The detailed computation is shown as below. Let $i(\gamma, S)$ be the index of S in γ in the lifting stage, and $k(S, e)$ be the index of e in $E \setminus S$ in the rerandomisation stage. It is easy to see both indices are well defined.

In the case $Z' = Z + e$ and $e \notin Z$, we have

$$
\begin{aligned}
\sum_{\substack{\xi \in \Gamma_{\mathrm{wrc}}:\\ (Z,Z')\in\xi}} w_{\mathrm{wrc}}(\xi) = \sum_{S\subseteq Z}\Bigg(& \sum_{\substack{\gamma\in\Gamma_{\mathrm{sg}}:(S,S')\in\gamma\\ S'=S+e}} w_{\mathrm{sg}}(\gamma)\,\mathbf{Pr}_\gamma[\mathcal{Z}_{i(\gamma,S)} = Z, \mathcal{Z}_{i(\gamma,S)+1} = Z'] \\
& + \sum_{\substack{\gamma\in\Gamma_{\mathrm{sg}}:S_0\overset{\gamma}{\rightsquigarrow}S\\ \text{for some } S_0}} w_{\mathrm{sg}}(\gamma)\,\mathbf{Pr}_\gamma[\mathcal{Z}_{|\gamma|+k(S,e)-1} = Z, \mathcal{Z}_{|\gamma|+k(S,e)} = Z']\Bigg) \\
= \sum_{S\subseteq Z}\Bigg(& \sum_{\substack{\gamma\in\Gamma_{\mathrm{sg}}:(S,S')\in\gamma\\ S'=S+e}} w_{\mathrm{sg}}(\gamma)\,\mathbf{Pr}_\gamma[\mathcal{Z}_{i(\gamma,S)} = Z] + \sum_{\substack{\gamma\in\Gamma_{\mathrm{sg}}:S_0\overset{\gamma}{\rightsquigarrow}S\\ \text{for some } S_0}} w_{\mathrm{sg}}(\gamma)\,\mathbf{Pr}_\gamma[\mathcal{Z}_{|\gamma|+k(S,e)-1} = Z]p'_e\Bigg) \\
= \sum_{S\subseteq Z}\delta(S,Z)\Bigg(& \sum_{\substack{\gamma\in\Gamma_{\mathrm{sg}}:(S,S')\in\gamma\\ S'=S+e}} w_{\mathrm{sg}}(\gamma) + \sum_{\substack{\gamma\in\Gamma_{\mathrm{sg}}:S_0\overset{\gamma}{\rightsquigarrow}S\\ \text{for some } S_0}} w_{\mathrm{sg}}(\gamma)p'_e\Bigg) \\
\le \sum_{S\subseteq Z}\delta(S,Z) & \left(\pi_{\mathrm{sg}}(S+e)\eta_{\min}^{-4} + \pi_{\mathrm{sg}}(S)p'_e\right) \quad \text{(by Lemma 3.5)} \\
= (1-p'_e)&\eta_{\min}^{-4}\pi_{\mathrm{wrc}}(Z+e) + p'_e\pi_{\mathrm{wrc}}(Z). \quad \text{(by (19))}
\end{aligned}
$$

In the case $Z' = Z - e$ and $e \in Z$, we have

$$
\begin{aligned}
\sum_{\xi \in \Gamma_{\mathrm{wrc}}:(Z,Z')\in\xi} w_{\mathrm{wrc}}(\xi) = & \sum_{\substack{S\subseteq Z\\ e\in S}} \sum_{\substack{\gamma\in\Gamma_{\mathrm{sg}}:(S,S')\in\gamma\\ S'=S-e}} w_{\mathrm{sg}}(\gamma)\,\mathbf{Pr}_\gamma[\mathcal{Z}_{i(\gamma,S)} = Z, \mathcal{Z}_{i(\gamma,S)+1} = Z'] \\
& + \sum_{\substack{S\subseteq Z\\ e\notin S}} \sum_{\substack{\gamma\in\Gamma_{\mathrm{sg}}:S_0\overset{\gamma}{\rightsquigarrow}S\\ \text{for some } S_0}} w_{\mathrm{sg}}(\gamma)\,\mathbf{Pr}_\gamma[\mathcal{Z}_{|\gamma|+k(S,e)-1} = Z, \mathcal{Z}_{|\gamma|+k(S,e)} = Z'] \\
= & \sum_{\substack{S\subseteq Z\\ e\in S}} \sum_{\substack{\gamma\in\Gamma_{\mathrm{sg}}:(S,S')\in\gamma\\ S'=S-e}} w_{\mathrm{sg}}(\gamma)\,\mathbf{Pr}_\gamma[\mathcal{Z}_{i(\gamma,S)} = Z](1-p'_e) \\
& + \sum_{\substack{S\subseteq Z\\ e\notin S}} \sum_{\substack{\gamma\in\Gamma_{\mathrm{sg}}:S_0\overset{\gamma}{\rightsquigarrow}S\\ \text{for some } S_0}} w_{\mathrm{sg}}(\gamma)\,\mathbf{Pr}_\gamma[\mathcal{Z}_{|\gamma|+k(S,e)-1} = Z](1-p'_e) \\
= & \sum_{\substack{S\subseteq Z\\ e\in S}} \delta(S,Z) \sum_{\substack{\gamma\in\Gamma_{\mathrm{sg}}:(S,S')\in\gamma\\ S'=S-e}} w_{\mathrm{sg}}(\gamma)(1-p'_e) + \sum_{\substack{S\subseteq Z\\ e\notin S}} \delta(S,Z) \sum_{\substack{\gamma\in\Gamma_{\mathrm{sg}}:S_0\overset{\gamma}{\rightsquigarrow}S\\ \text{for some } S_0}} w_{\mathrm{sg}}(\gamma)(1-p'_e) \\
\le & \sum_{\substack{S\subseteq Z\\ e\in S}} \delta(S,Z)\pi_{\mathrm{sg}}(S)\eta_{\min}^{-4}(1-p'_e) + \sum_{\substack{S\subseteq Z\\ e\notin S}} \delta(S,Z)\pi_{\mathrm{sg}}(S)(1-p'_e) \quad \text{(by Lemma 3.5)} \\
= & \eta_{\min}^{-4}(1-p'_e)\pi_{\mathrm{wrc}}(Z). \quad \text{(by (19))}
\end{aligned}
$$

In the case $Z' = Z$, we have

$$
\begin{aligned}
\sum_{\xi \in \Gamma_{\mathrm{wrc}}:(Z,Z)\in\xi} w_{\mathrm{wrc}}(\xi) &\le \sum_{S\subseteq Z}\sum_{\gamma\in\Gamma_{\mathrm{sg}}:S\in\gamma} w_{\mathrm{sg}}(\gamma)\,\mathbf{Pr}_{\gamma}[\mathcal{Z}_{i(\gamma,S)} = Z, \mathcal{Z}_{i(\gamma,S)+1} = Z] \\
&\quad + \sum_{S\subseteq Z}\sum_{\substack{\gamma\in\Gamma_{\mathrm{sg}}:S_0\overset{\gamma}{\rightsquigarrow}S\\ \text{for some } S_0}} w_{\mathrm{sg}}(\gamma)\sum_{i=1}^{|E\setminus S|}\mathbf{Pr}_{\gamma}[\mathcal{Z}_{|\gamma|+i-1} = Z, \mathcal{Z}_{|\gamma|+i} = Z] \\
&\le \sum_{S\subseteq Z}\sum_{\gamma\in\Gamma_{\mathrm{sg}}:S\in\gamma} w_{\mathrm{sg}}(\gamma)\,\mathbf{Pr}_{\gamma}[\mathcal{Z}_{i(\gamma,S)} = Z] + \sum_{S\subseteq Z}\sum_{\substack{\gamma\in\Gamma_{\mathrm{sg}}:S_0\overset{\gamma}{\rightsquigarrow}S\\ \text{for some } S_0}} w_{\mathrm{sg}}(\gamma)|E\setminus S|\delta(S,Z) \\
&= \sum_{S\subseteq Z}\delta(S,Z)\left(\sum_{e\in Z}\sum_{\substack{\gamma\in\Gamma_{\mathrm{sg}}:(S,S')\in\gamma\\ S'=S\oplus\{e\}}} w_{\mathrm{sg}}(\gamma) + \sum_{\substack{\gamma\in\Gamma_{\mathrm{sg}}:S_0\overset{\gamma}{\rightsquigarrow}S\\ \text{for some } S_0}} w_{\mathrm{sg}}(\gamma)|E\setminus S|\right) \\
\text{(By Lemma 3.5)} \qquad &\le m\sum_{S\subseteq Z}\delta(S,Z)\left(\pi_{\mathrm{sg}}(S)\eta_{\min}^{-4} + \pi_{\mathrm{sg}}(S)\right) \\
\text{(By (19))} \qquad &= m(\eta_{\min}^{-4}+1)\pi_{\mathrm{wrc}}(Z).
\end{aligned}
$$

3.2.3. *Bound the total congestion.* We are now able to bound the congestion of Γ_{wrc}. There are also three cases. Before continuing, we need to study the transition probability in the edge flipping dynamics for weighted random cluster models (9). Recall that the parameter for weighted random cluster is $2\boldsymbol{p}$. For $Z \subseteq E$ and $e \notin Z$, to provide a bound on $\frac{\pi_{\mathrm{wrc}}(Z+e)}{\pi_{\mathrm{wrc}}(Z)}$, note that it can be explicitly written by

$$
\frac{\pi_{\mathrm{wrc}}(Z+e)}{\pi_{\mathrm{wrc}}(Z)} = \begin{cases} \frac{2p_e}{1-2p_e}\frac{1+\prod_{u\in C_1\cup C_2}\lambda_u}{(1+\prod_{u\in C_1}\lambda_u)(1+\prod_{u\in C_2}\lambda_u)}, & \text{if adding } e \text{ connects two components } C_1 \text{ and } C_2; \\ \frac{2p_e}{1-2p_e}, & \text{otherwise.} \end{cases}
$$

In either case, it holds that

$$
\frac{2p_e}{1-2p_e} \le \frac{\pi_{\mathrm{wrc}}(Z+e)}{\pi_{\mathrm{wrc}}(Z)} \le 2\cdot\frac{2p_e}{1-2p_e} \tag{22}
$$

by using the inequality $1 \le (1+xy)/((1+x)(1+y)) \le 2$ whenever $0 \le x, y \le 1$.

- For transition (Z, Z') such that $Z' = Z + e$ and $e \notin Z$,

$$
\frac{L}{\pi_{\mathrm{wrc}}(Z)P_{\mathrm{wrc}}(Z,Z')}\sum_{\xi\in\Gamma_{\mathrm{wrc}}:(Z,Z')\in\xi} w(\xi) \le \frac{m(1-p_e')\eta_{\min}^{-4}\pi_{\mathrm{wrc}}(Z+e)}{\pi_{\mathrm{wrc}}(Z)P_{\mathrm{wrc}}(Z,Z+e)} + \frac{m}{P_{\mathrm{wrc}}(Z,Z+e)}p_e'
$$

  To bound the first term, note that if $\frac{\pi_{\mathrm{wrc}}(Z+e)}{\pi_{\mathrm{wrc}}(Z)} > 1$, then

$$
\begin{aligned}
\frac{m(1-p_e')\eta_{\min}^{-4}\pi_{\mathrm{wrc}}(Z+e)}{\pi_{\mathrm{wrc}}(Z)P_{\mathrm{wrc}}(Z,Z+e)} &= \frac{2m^2}{\min\left\{1, \frac{\pi_{\mathrm{wrc}}(Z+e)}{\pi_{\mathrm{wrc}}(Z)}\right\}}(1-p_e')\eta_{\min}^{-4}\frac{\pi_{\mathrm{wrc}}(Z+e)}{\pi_{\mathrm{wrc}}(Z)} \\
\text{(By (22))} \qquad &\le 4m^2\frac{1-2p_e}{1-p_e}\eta_{\min}^{-4}\frac{2p_e}{1-2p_e} \\
\text{(By } p_e < 1/2\text{)} \qquad &\le 8m^2\eta_{\min}^{-4}.
\end{aligned}
$$

  The above inequality trivially holds if $\frac{\pi_{\mathrm{wrc}}(Z+e)}{\pi_{\mathrm{wrc}}(Z)} \le 1$ too.

  Assuming $\frac{\pi_{\mathrm{wrc}}(Z+e)}{\pi_{\mathrm{wrc}}(Z)} < 1$ again, the second term is

$$
\frac{m}{P_{\mathrm{wrc}}(Z,Z')}p_e' = \frac{2m^2}{\frac{\pi_{\mathrm{wrc}}(Z+e)}{\pi_{\mathrm{wrc}}(Z)}}\frac{p_e}{1-p_e} \overset{(22)}{\le} \frac{1-2p_e}{1-p_e}m^2 \le m^2.
$$

If $\frac{\pi_{\mathrm{wrc}}(Z+e)}{\pi_{\mathrm{wrc}}(Z)} \geq 1$, the second term is trivially bounded by $2m^2$. In all, we have

$$\frac{L}{\pi_{\mathrm{wrc}}(Z)P_{\mathrm{wrc}}(Z,Z')} \sum_{\xi \in \Gamma_{\mathrm{wrc}}:(Z,Z') \in \xi} w(\xi) \leq 10m^2 \eta_{\min}^{-4}.$$

- For transition (Z,Z') such that $Z' = Z - e$ and $e \in Z$,

$$\frac{L}{\pi_{\mathrm{wrc}}(Z)P_{\mathrm{wrc}}(Z,Z')} \sum_{\xi \in \Gamma_{\mathrm{wrc}}:(Z,Z') \in \xi} w(\xi) \leq \frac{2m^2(1-p'_e)\eta_{\min}^{-4}}{\min\left\{1, \frac{\pi_{\mathrm{wrc}}(Z-e)}{\pi_{\mathrm{wrc}}(Z)}\right\}}.$$

  If $\frac{\pi_{\mathrm{wrc}}(Z-e)}{\pi_{\mathrm{wrc}}(Z)} < 1$, then we have

$$\frac{2m^2(1-p'_e)\eta_{\min}^{-4}}{\min\left\{1, \frac{\pi_{\mathrm{wrc}}(Z-e)}{\pi_{\mathrm{wrc}}(Z)}\right\}} = 2m^2(1-p'_e)\eta_{\min}^{-4} \frac{\pi_{\mathrm{wrc}}(Z)}{\pi_{\mathrm{wrc}}(Z-e)} \overset{(22)}{\leq} 4m^2\eta_{\min}^{-4} \frac{1-2p_e}{1-p_e} \frac{2p_e}{1-2p_e} \leq 8m^2\eta_{\min}^{-4},$$

  where the last step is due to $p_e < 1/2$. The inequality trivially holds if $\frac{\pi_{\mathrm{wrc}}(Z-e)}{\pi_{\mathrm{wrc}}(Z)} \geq 1$.
- For transition (Z,Z), first note that the laziness of the chain ensures $P_{\mathrm{wrc}}(Z,Z) \geq 1/2$. Therefore,

$$\frac{L}{\pi_{\mathrm{wrc}}(Z)P_{\mathrm{wrc}}(Z,Z)} \sum_{\xi \in \Gamma_{\mathrm{wrc}}:(Z,Z') \in \xi} w(\xi) \leq 2m^2(\eta_{\min}^{-4} + 1) \leq 4m^2\eta_{\min}^{-4}$$

Lemma 3.2 follows from combining all three cases.

3.3. **Faster mixing via perturbed chains.** One issue with Lemma 3.2 is the dependency of mixing time on $a = \min_{v \in V} \frac{1-\lambda_v}{1+\lambda_v}$. Especially, when some λ_v is very close to 1, say $1 - e^{-n}$, or even simply 1, the relaxation time is no longer a polynomial in n. In this subsection, we eliminate such extreme cases at the cost of some extra factors of n. The main lemma of this subsection is stated as follows.

Lemma 3.6. *For any weighted random cluster model on graph $G = (V, E)$ with parameters $2\boldsymbol{p} = (2p_e)_{e \in E}$ and $\boldsymbol{\lambda} = (\lambda_v)_{v \in V}$, if $0 < p_e < \frac{1}{2}$ for all $e \in E$ and $0 < \lambda_v \leq 1$ for all $v \in V$. Then the spectral gap for P_{EF}^{wrc} is at least $\frac{1}{4410n^4m^2}$.*

The core of the refined analysis is to cap λ_v from the above, or equivalently for the subgraph-world model, to cap η_v from the below. Given a subgraph-world model $(G; \boldsymbol{p}, \boldsymbol{\eta})$, we define the "perturbed" model $(G; \boldsymbol{p}, \widehat{\boldsymbol{\eta}})$ by

$$\widehat{\eta}_v = \begin{cases} \frac{1}{n}, & \text{if } 0 \leq \eta_v \leq \frac{1}{n} \\ \eta_v, & \text{otherwise.} \end{cases} \tag{23}$$

Call the induced distribution $\widehat{\pi_{\mathrm{sg}}}$. Take a random subgraph $\mathcal{S}$ according to $\widehat{\pi_{\mathrm{sg}}}$, and add each remaining edge $e \in E \setminus \mathcal{S}$ with probability $p_e/(1-p_e)$ to obtain $\mathcal{R}$. By Lemma 3.3, the resulting distribution is $\pi_{\mathrm{wrc}}(G; 2\boldsymbol{p}, \widehat{\boldsymbol{\lambda}}) =: \widehat{\pi_{\mathrm{wrc}}}$, where $\widehat{\lambda}_v = \frac{1-\widehat{\eta}_v}{1+\widehat{\eta}_v}$. Applying Lemma 3.2 to the perturbed random-cluster model $(G; 2\boldsymbol{p}, \widehat{\boldsymbol{\lambda}})$ yields a $\frac{1}{10m^2n^4}$ lower bound on spectral gap of the corresponding edge flipping dynamics. Based on this, the main effort of this subsection is to bound the spectral gap of the original model $(G; 2\boldsymbol{p}, \boldsymbol{\lambda})$ via the bounds for $(G; 2\boldsymbol{p}, \widehat{\boldsymbol{\lambda}})$.

We start with comparing the two distributions.

Lemma 3.7. *For any $R \subseteq E$,*

$$\frac{1}{9} \leq \frac{\widehat{\pi_{wrc}}(R)}{\pi_{wrc}(R)} < \mathrm{e}.$$

Proof. Let $n = |V|$. If $n = 1$, the only possible R is $\varnothing$ and the lemma holds. We assume $n \geq 2$ in the rest. To prove the first inequality,

$$\frac{\widehat{\pi_{\mathrm{wrc}}}(R)}{\pi_{\mathrm{wrc}}(R)} = \frac{Z_{\mathrm{wrc}}}{\widehat{Z_{\mathrm{wrc}}}} \cdot \frac{\widehat{\mathrm{wt}_{\mathrm{wrc}}}(R)}{\mathrm{wt}_{\mathrm{wrc}}(R)} = \frac{Z_{\mathrm{wrc}}}{\widehat{Z_{\mathrm{wrc}}}} \cdot \prod_{C \in \kappa(V,\mathcal{S})} \frac{1 + \prod_{u \in C} \widehat{\lambda}_u}{1 + \prod_{u \in C} \lambda_u}.$$

Note that $\frac{Z_{\text{wrc}}}{\widehat{Z_{\text{wrc}}}} \geq 1$ because $\widehat{\lambda}_u \leq \lambda_u$, which implies that the weight of each configuration of the weighted random cluster model decreases after replacing $\boldsymbol{\lambda}$ with $\widehat{\boldsymbol{\lambda}}$. The second term can be handled by

$$\prod_{C\in\kappa(V,S)} \frac{1+\prod_{u\in C}\widehat{\lambda}_u}{1+\prod_{u\in C}\lambda_u} \geq \prod_{C\in\kappa(V,S)} \frac{\prod_{u\in C}\widehat{\lambda}_u}{\prod_{u\in C}\lambda_u} \geq \left(\frac{n-1}{n+1}\right)^n \geq \frac{1}{9}$$

as $n \geq 2$.

For the second inequality, the definition of π_{wrc}, together with the relation between Z_{wrc} and Z_{sg} in Equation (6), gives

$$\frac{\widehat{\pi_{\text{wrc}}}(R)}{\pi_{\text{wrc}}(R)} = \frac{Z_{\text{sg}}(G;\boldsymbol{p},\boldsymbol{\eta})}{Z_{\text{sg}}(G;\boldsymbol{p},\widehat{\boldsymbol{\eta}})} \cdot \frac{\prod_{v\in V}\frac{1}{1+\widehat{\lambda}_v}}{\prod_{v\in V}\frac{1}{1+\lambda_v}} \cdot \frac{\prod_{C\in\kappa(V,R)}\left(1+\prod_{u\in C}\widehat{\lambda}_u\right)}{\prod_{C\in\kappa(V,R)}\left(1+\prod_{u\in C}\lambda_u\right)}.$$

There are three terms. For the first one, note that $\widehat{\eta}_v > \eta_v$ for all v, indicating that the weight of each configuration of the subgraph-world model is increased after replacing $\boldsymbol{\eta}$ with $\widehat{\boldsymbol{\eta}}$. As such, it is less or equal than 1. The third term is also less or equal than 1 due to $\widehat{\lambda}_v < \lambda_v$. The second term can be bounded by

$$\frac{\prod_{v\in V}\frac{1}{1+\widehat{\lambda}_v}}{\prod_{v\in V}\frac{1}{1+\lambda_v}} = \frac{\prod_{v\in V}(1+\widehat{\eta}_v)}{\prod_{v\in V}(1+\eta_v)} \leq \left(1+\frac{1}{n}\right)^n < \mathrm{e}$$

which concludes this lemma. □

We also have a bound on the ratio of the transition probability between the original and perturbed model in the Metropolis dynamics (9).

Lemma 3.8. *Let P_{wrc} and $\widehat{P_{wrc}}$ be the transition of the edge flipping dynamics on the weighted random cluster models $(G;2\boldsymbol{p},\boldsymbol{\lambda})$ and $(G;2\boldsymbol{p},\widehat{\boldsymbol{\lambda}})$ respectively. Then it holds that*

$$\frac{1}{2} \leq \frac{\widehat{P_{wrc}}(Z,Z')}{P_{wrc}(Z,Z')} \leq 2 \qquad \text{for all } |Z\oplus Z'| \leq 1.$$

Proof. The case $Z = Z'$ immediately follows from the chain being lazy with probability at least $1/2$. Therefore, assume $Z' = Z + e$ where $e \notin Z$. The case $Z' = Z - e$ where $e \in Z$ follows by a similar argument. We then have

$$\frac{\widehat{P_{\text{wrc}}}(Z,Z')}{P_{\text{wrc}}(Z,Z')} = \frac{\min\left\{1, \frac{\widehat{\pi_{\text{wrc}}}(Z')}{\widehat{\pi_{\text{wrc}}}(Z)}\right\}}{\min\left\{1, \frac{\pi_{\text{wrc}}(Z')}{\pi_{\text{wrc}}(Z)}\right\}} = \frac{\min\{1, Ax\}}{\min\{1, Bx\}},$$

where

$$x := \frac{2p_e}{1-2p_e}, \qquad Ax := \frac{\widehat{\pi_{\text{wrc}}}(Z')}{\widehat{\pi_{\text{wrc}}}(Z)} \qquad \text{and} \qquad Bx := \frac{\pi_{\text{wrc}}(Z')}{\pi_{\text{wrc}}(Z)}.$$

By (22), it holds that $1 \leq A, B \leq 2$. We claim that $\frac{1}{2} \leq \frac{\min\{1,Ax\}}{\min\{1,Bx\}} \leq 2$, which follows from elementary algebra, finishing the proof. □

Now we are ready to prove Lemma 3.6.

Proof of Lemma 3.6. Fix a test function f. Denote by $\mathcal{E}(f,f)$, $\widehat{\mathcal{E}}(f,f)$ the Dirichlet form of P_{wrc} and $\widehat{P_{\text{wrc}}}$ respectively. Denote by $\mathrm{Var}[f]$ and $\widehat{\mathrm{Var}}[f]$ the variance of f with respect to π_{wrc} and $\widehat{\pi_{\text{wrc}}}$ respectively.

Then by Lemma 3.7 and Lemma 3.8,

$$\frac{\mathcal{E}(f,f)}{\mathrm{Var}[f]}=\frac{\sum\limits_{\substack{X,Y\subseteq E\\|X\oplus Y|\le 1}}\pi_{\mathrm{wrc}}(X)P_{\mathrm{wrc}}(X,Y)\,(f(X)-f(Y))^2}{\sum\limits_{\substack{X,Y\subseteq E\\|X\oplus Y|\le 1}}\pi_{\mathrm{wrc}}(X)\pi_{\mathrm{wrc}}(Y)\,(f(X)-f(Y))^2}$$

$$\ge\frac{\frac{1}{2\mathrm{e}}\sum\limits_{\substack{X,Y\subseteq E\\|X\oplus Y|\le 1}}\widehat{\pi_{\mathrm{wrc}}}(X)\widehat{P_{\mathrm{wrc}}}(X,Y)\,(f(X)-f(Y))^2}{81\sum\limits_{\substack{X,Y\subseteq E\\|X\oplus Y|\le 1}}\widehat{\pi_{\mathrm{wrc}}}(X)\widehat{\pi_{\mathrm{wrc}}}(Y)\,(f(X)-f(Y))^2}>\frac{1}{441}\frac{\widehat{\mathcal{E}}(f,f)}{\widehat{\mathrm{Var}}[f]}.$$

Therefore, $\mathfrak{Gap}(P_{\mathrm{wrc}})\ge\frac{1}{441}\mathfrak{Gap}(\widehat{P_{\mathrm{wrc}}})\ge\frac{1}{4410n^4m^2}$.

□

3.4. **Proof of Lemma 3.1.** Note that $a=\min_{v\in V}\frac{1-\lambda_v}{1+\lambda_v}=\frac{1-\lambda_{\max}}{1+\lambda_{\max}}\ge\frac{1-\lambda_{\max}}{2}$. By Lemma 3.2, the spectral gap of P^{wrc}_{EF} satisfies $\mathfrak{Gap}(P^{\mathrm{wrc}}_{EF})=\Omega(\frac{(1-\lambda_{\max})^4}{m^2})$. On the other hand, by Lemma 3.6, $\mathfrak{Gap}(P^{\mathrm{wrc}}_{EF})=\Omega(\frac{1}{n^4m^2})$. This proves the spectral gap result in Lemma 3.1. To prove the mixing time, the partition function Z_{wrc} of weighted random cluster model satisfies $Z_{\mathrm{wrc}}\le 2^m2^n$. For any $R\subseteq E$,

$$\mathrm{wt}_{\mathrm{wrc}}(R)\ge\min\{2p_e,1-2p_e\}^m=\frac{1}{b^m}.$$

Since P^{wrc}_{EF} is positive semi-definite, the mixing of P^{wrc}_{EF} satisfies

$$T_{\mathrm{mix}}(P^{\mathrm{wrc}}_{EF},\varepsilon)=\frac{1}{\mathfrak{Gap}(P^{\mathrm{wrc}}_{EF})}\left(\log(b^m2^{m+n})+\log\frac{1}{\varepsilon}\right)=O\left(N^4m^3\log\frac{b}{\varepsilon}\right),$$

where the last equation holds because $m\ge n-1$ (we may assume the graph is connected).

4. Comparing Swendsen-Wang and edge-flipping dynamics

Having analysed the edge-flipping dynamics, now we turn to relating it with the Swendsen-Wang dynamics. From this point on, we no longer need the subgraph-world model. We first reiterate the settings for clarity. Let $G=(V,E)$ be a graph. We consider the Ising model on G with parameters $\boldsymbol{\lambda}=(\lambda_v)_{v\in V}$ and $\boldsymbol{\beta}=(\beta_e)_{e\in E}$, where $0<\lambda_v\le 1$ for all $v\in V$ and $\beta_e>1$ for all $e\in E$, as well as the weighted random cluster model on G with parameters $\boldsymbol{p}=(p_e)_{e\in E}$ and $\boldsymbol{\lambda}=(\lambda_v)_{v\in V}$, where $p_e=1-\frac{1}{\beta_e}$ for all $e\in E$. Let π_{Ising} over $\Omega_{\mathcal{I}}=\{0,1\}^V$ denote the Gibbs distribution of the Ising model, and π_{wrc} over $\Omega_{\mathcal{R}}=2^E$ denote the distribution of the weighted random cluster model.

Let $P^{\mathrm{wrc}}_{\mathrm{SW}}$ denote the transition matrix of the Swendsen-Wang dynamics for weighted random cluster models as defined in Section 2.2.2, and P_{EF} denote the transition matrix of the edge flipping dynamics for weighted random cluster models as defined in Section 2.2.1. The main result of this section is the following lemma.

Lemma 4.1. *If $0<\lambda_v\le 1$ for all $v\in V$, then it holds that*

$$\mathfrak{Gap}(P^{\mathrm{wrc}}_{\mathrm{SW}})\ge\frac{\mathfrak{Gap}\,(P_{\mathrm{EF}})}{3}.$$

Let $P^{\mathrm{Ising}}_{\mathrm{SW}}$ denote the transition matrix of Swendsen-Wang dynamics for the Ising model. By the relation in (14), we have the following corollary.

Corollary 4.2. *If $0<\lambda_v\le 1$ for all $v\in V$, then it holds that*

$$\mathfrak{Gap}(P^{\mathrm{Ising}}_{\mathrm{SW}})\ge\frac{\mathfrak{Gap}\,(P_{\mathrm{EF}})}{3}.$$

Now, we are ready to prove Theorem 1.1.

Proof of Theorem 1.1. Combining Lemma 3.1 and Corollary 4.2, we have

$$\mathfrak{Gap}(P_{\mathrm{SW}}^{\mathrm{Ising}}) = \Omega\left(\frac{1}{N^4 m^2}\right).$$

Let Z_{Ising} denote the partition function of the Ising model. Since $\lambda_v \leq 1$ for all $v \in V$, we have $Z_{\mathrm{Ising}} \leq 2^n \beta_{\max}^m$. For any configuration $\sigma \in \{0,1\}^V$, it holds that $\mathrm{wt}_{\mathrm{Ising}}(\sigma) = \lambda_{\min}^n$. Without loss of generality, we assume the graph is connected, and thus $m \geq n-1$. Hence

$$T_{\mathrm{mix}}(\varepsilon) = \frac{1}{\mathfrak{Gap}(P_{\mathrm{SW}}^{\mathrm{Ising}})}\left(\log\left(\frac{2^n\beta_{\max}^m}{\lambda_{\min}^n}\right) + \log\frac{1}{\varepsilon}\right) = O\left(N^4 m^3 \log\frac{\beta_{\max}}{\lambda_{\min}\varepsilon}\right). \qquad \square$$

4.1. **Compare edge flipping dynamics with Swendsen-Wang dynamics.** We now prove Lemma 4.1. We introduce a Markov chain called the (lazy) single-bond dynamics for weighted random cluster models. For any $S \subseteq E$, any two vertices $u, v \in V$, we use $u \overset{S}{\leftrightarrow} v$ to denote the event that u and v belong to the same connected component in the subgraph (V, S), and we use $u \overset{S}{\not\leftrightarrow} v$ to denote the event that u and v belong to different connected components in the subgraph (V, S). Suppose the current state is $S \in \Omega_{\mathcal{R}}$, the single-bond dynamics does the following update $S \to S'$. With probability $\frac{1}{2}$, let $S' = S$; otherwise

- pick an edge $e = \{u, v\} \in E$ uniformly at random;
- If $u \overset{S}{\leftrightarrow} v$, then
  (1) $S' = S \cup \{e\}$ with probability p_e;
  (2) $S' = S \setminus \{e\}$ with probability $1 - p_e$.

  If $u \overset{S}{\not\leftrightarrow} v$, suppose u and v are in connected components C_u and C_v of subgraph (V, S) respectively, then
  (1) $S' = S \cup \{e\}$ with probability $\frac{p_e(1+\prod_{w \in C_u \cup C_v}\lambda_w)}{(1+\prod_{w\in C_u}\lambda_w)(1+\prod_{w' \in C_v}\lambda_{w'})}$;
  (2) $S' = S \setminus \{e\} = S$ with probability $1 - \frac{p_e(1+\prod_{w \in C_u \cup C_v}\lambda_w)}{(1+\prod_{w\in C_u}\lambda_w)(1+\prod_{w' \in C_v}\lambda_{w'})}$.

  Let P_{SB} denote the transition matrix of (lazy) single-bond dynamics.

Lemma 4.3 ([Ull14])**.** *The transition matrix P_{SB} is reversible with respect to π_{wrc}. It holds that*

$$\mathfrak{Gap}(P_{\mathrm{SB}}) \leq \mathfrak{Gap}(P_{\mathrm{SW}}).$$

The above result is proved in [Ull14] in the case where $p_e = p \in (0,1)$ for all $e \in E$ and $\lambda_v = 1$ for all $v \in V$.[4] The model we consider allows that each e has different $p_e \in (0,1)$ and each v has different $\lambda_v \in (0,1]$. Lemma 4.3 can be proved by the same method in [Ull14].

We have the following relation between P_{SB} and P_{EF}. The proof is given in Section 4.2.

Lemma 4.4. *If $0 < \lambda_v \leq 1$ for all $v \in V$, then for any $S, S' \in \Omega_{\mathcal{R}}$ with $S \neq S'$, it holds that*

$$P_{\mathrm{SB}}(S, S') \geq \frac{P_{\mathrm{EF}}(S, S')}{3}.$$

Combining Lemma 4.4 with (7), it holds that

$$\mathfrak{Gap}\left(P_{\mathrm{SB}}\right) \geq \frac{\mathfrak{Gap}\left(P_{\mathrm{EF}}\right)}{3}.$$

Lemma 4.1 is a consequence of above inequality and Lemma 4.3.

[4] In [Ull14], Ullrich proved this for general random cluster models with an arbitrary $q \geq 1$, but when $q \neq 2$ that model cannot be easily translated to the notation we use.

4.2. **Compare edge flipping dynamics with single-bond dynamics.** We prove Lemma 4.4. Fix $S \subseteq E$ and $e = \{u, v\} \in E$. We consider the following two cases $u \overset{S}{\leftrightarrow} v$ or $u \overset{S}{\not\leftrightarrow} v$. We remark that in each case, we only need to consider $S' \neq S$ satisfying $S' = S \cup \{e\}$ or $S' = S \setminus \{e\}$.

Suppose $u \overset{S}{\leftrightarrow} v$. We have two sub-cases for S: $e \in S$ or $e \notin S$.

- Case $e \in S$. Then we only need to consider $S' = S \setminus \{e\}$. It holds that

$$P_{\mathrm{SB}}(S, S') = \frac{1 - p_e}{2}. \tag{24}$$

In the edge flipping dynamics, if $u \overset{S'}{\leftrightarrow} v$, then

$$P_{\mathrm{EF}}(S, S') = \frac{1}{2} \min\left\{1, \frac{1 - p_e}{p_e}\right\}; \tag{25}$$

if $u \overset{S'}{\not\leftrightarrow} v$, suppose u and v belong to connected components C_u and C_v in graph (V, S') respectively. Let $x = \prod_{w \in C_u} \lambda_w$ and $y = \prod_{w \in C_v} \lambda_v$. Note that $0 < x, y \leq 1$. We have

$$P_{\mathrm{EF}}(S, S') = \frac{1}{2} \min\left\{1, \frac{(1 - p_e)(1 + x)(1 + y)}{p_e(1 + xy)}\right\} \leq \frac{1}{2} \min\left\{1, \frac{2(1 - p_e)}{p_e}\right\}. \tag{26}$$

Combining (24), (25) and (26), we have if $0 < p_e \leq \frac{2}{3}$, it holds that $P_{\mathrm{EF}}(S, S') \leq \frac{1}{2} = \frac{P_{\mathrm{SB}}(S,S')}{1 - p_e} \leq 3P_{\mathrm{SB}}(S, S')$; if $\frac{2}{3} < p_e \leq 1$, it holds that $P_{\mathrm{EF}}(S, S') \leq \frac{1-p_e}{p_e} \leq \frac{2}{p_e} P_{\mathrm{SB}}(S, S') \leq 3P_{\mathrm{SB}}(S, S')$. Hence, we have

$$P_{\mathrm{EF}}(S, S') \leq 3P_{\mathrm{SB}}(S, S').$$

- Case $e \notin S$. Then we only need to consider $S' = S \cup \{e\}$. It holds that

$$P_{\mathrm{SB}}(S, S') = \frac{p_e}{2} \quad \text{and} \quad P_{\mathrm{EF}}(S, S') = \frac{1}{2} \min\left\{1, \frac{p_e}{1 - p_e}\right\}.$$

If $0 < p_e \leq \frac{1}{2}$, it holds that $P_{\mathrm{EF}}(S, S') = \frac{p_e}{2(1-p_e)} = \frac{P_{\mathrm{SB}}(S,S')}{1-p_e} \leq 2P_{\mathrm{SB}}(S, S')$. If $\frac{1}{2} < p_e < 1$, it holds that $P_{\mathrm{EF}}(S, S') = \frac{1}{2} = \frac{P_{\mathrm{SB}}(S,S')}{p_e} \leq 2P_{\mathrm{SB}}(S, S')$. Hence, we have

$$P_{\mathrm{EF}}(S, S') \leq 2P_{\mathrm{SB}}(S, S').$$

Now suppose $u \overset{S}{\not\leftrightarrow} v$. It must hold that $e \notin S$. We only need to consider $S' = S \cup \{e\}$. Suppose u and v belong to connected components C_u and C_v in graph (V, S) respectively. Let $x = \prod_{w \in C_u} \lambda_w$ and $y = \prod_{w \in C_v} \lambda_v$. We have

$$P_{\mathrm{SB}}(S, S') = \frac{p_e(1 + xy)}{2(1 + x)(1 + y)} \quad \text{and} \quad P_{\mathrm{EF}}(S, S') = \frac{1}{2} \min\left\{1, \frac{p_e(1 + xy)}{(1 - p_e)(1 + x)(1 + y)}\right\}.$$

Note that $0 < x, y \leq 1$, we have

$$\frac{1}{2} \leq \frac{(1 + xy)}{(1 + x)(1 + y)} \leq 1,$$

which implies

$$P_{\mathrm{EF}}(S, S') \leq \begin{cases} \frac{1}{2} & \text{if } \frac{2}{3} \leq p_e < 1 \\ \frac{p_e(1+xy)}{2(1-p_e)(1+x)(1+y)} & \text{if } 0 < p_e < \frac{2}{3} \end{cases}$$

If $\frac{2}{3} \leq p_e < 1$, then $P_{\mathrm{EF}}(S, S') \leq \frac{1}{2} = \frac{P_{\mathrm{SB}}(S,S')(1+x)(1+y)}{p_e(1+xy)} \leq \frac{2P_{\mathrm{SB}}}{p_e} \leq 3P_{\mathrm{SB}}$; if $0 < p_e < \frac{2}{3}$, then it holds that $P_{\mathrm{EF}}(S, S') \leq \frac{P_{\mathrm{SB}}(S,S')}{1-p_e} \leq 3P_{\mathrm{SB}}(S, S')$.

Combining all cases together, we have

$$P_{\mathrm{EF}}(S, S') \leq 3P_{\mathrm{SB}}(S, S').$$

5. Perfect sampler via CFTP

In this section, we give a perfect sampler for the ferromagnetic Ising model. We first give a perfect sampler for the weighted random cluster model, then turn it into a perfect sampler for the Ising model.

Theorem 5.1. *There exists a perfect sampling algorithm such that given any weighted random cluster model on graph $G = (V, E)$ with parameters $\boldsymbol{p} = (p_e)_{e\in E}$ and $\boldsymbol{\lambda} = (\lambda_v)_{v\in V}$, if $0 < p_e < 1$ for all $e \in E$ and $0 < \lambda_v \leq 1$ for all $v \in V$, the algorithm returns a perfect sample from weighted random cluster models in expected time $O(N^4 m^4 \log n \log b)$, where $b = \max_{e\in E} \max\{\frac{1}{p_e}, \frac{1}{1-p_e}\}$, $N = \min\{n, \frac{1}{1-\lambda_{\max}}\}$ and $\lambda_{\max} = \max_{v\in V} \lambda_v$.*

5.1. Perfect ferromagnetic Ising sampler.

We now prove Theorem 1.2. We give the perfect ferromagnetic Ising sampler assuming the algorithm in Theorem 5.1. Let $G = (V, E)$ be a graph. Let $\boldsymbol{\beta} = (\beta_e)_{e\in E}$ and $\boldsymbol{\lambda} = (\lambda_v)_{v\in V}$ be parameters for the Ising model, where $\beta_e > 1$ for all $e \in E$ and $0 < \lambda_v < 1$ for all $v \in V$. Let $p_e = 1 - \frac{1}{\beta_e}$ for all $e \in E$. We first use algorithm in Theorem 5.1 to draw a perfect random sample $\mathcal{S} \subseteq E$ from the weighted random cluster model with parameters $\boldsymbol{p}$ and $\boldsymbol{\lambda}$. Then we using the Markov chain $\mathcal{P}_{\mathcal{R}\to\mathcal{I}}$ in (11) to transform $\mathcal{S}$ into a random Ising configuration $\sigma \in \{0, 1\}^V$. By Proposition 2.3, since $\mathcal{S} \sim \pi_{\mathrm{wrc}}$, σ is a perfect sample from the Ising model. The running time of the transformation step is $O(n + m)$. Note that

$$b = \max_{e\in E} \max\left\{\frac{1}{p_e}, \frac{1}{1-p_e}\right\} = \max_{e\in E} \max\left\{\frac{\beta_e}{\beta_e - 1}, \beta_e\right\} \leq \beta_{\max} + \frac{\beta_{\min}}{\beta_{\min} - 1}.$$

By Theorem 5.1, the total running time is $O(N^4 m^4 \log n \log(\beta_{\max} + \frac{\beta_{\min}}{\beta_{\min}-1}))$.

5.2. CFTP for weighted random cluster models.

We give a perfect sampler for weighted random cluster models based on the coupling form the past (CFTP), where the CFTP is applied to the edge flipping dynamics. Here is an equivalent definition of the edge flipping dynamics defined in (9). There is a one-to-one correspondence between vectors in $\{0, 1\}^E$ and subsets in 2^E (i.e. for any $X \in \{0, 1\}^E$, let $S_X = \{e \in E \mid X_e = 1\}$). We assume that the Markov chain is defined over the state space $\{0, 1\}^E$. The edge flipping dynamics starts from an arbitrary subset of edges $X_0 \in \{0, 1\}^E$. For the t-th transition step, sample a bit $\ell_t \in \{0, 1\}$ uniformly at random, if $\ell_t = 0$, let $X_t = X_{t-1}$; otherwise, the chain does as follows:

- pick an edge $e_t \in E$ uniformly at random;
- sample a bit $b_t \in \{0, 1\}$ uniformly at random, construct the candidate state $X' \in \{0, 1\}^E$ by setting $X'(e) = X_{t-1}(e)$ for all $e \in E \setminus \{e_t\}$ and $X'(e_t) = b_t$.
- sample a real number $r_t \in [0, 1]$ uniformly at random; if $r_t < a_t$, accept the candidate and let $X_t = X'$; if $r_t \geq a_t$, reject the candidate and let $X_t = X_{t-1}$, where

$$a_t := \min\left\{1, \frac{\pi_{\mathrm{wrc}}(X')}{\pi_{\mathrm{wrc}}(X_{t-1})}\right\}.$$

The edge flipping dynamics for weighted random cluster models admits a *grand monotone coupling*. Let $\Omega = \{0, 1\}^E$. Let $P : \Omega \times \Omega \to \mathbb{R}_{\geq 0}$ denote the transition matrix of the edge flipping dynamics. We use the function $\varphi(\cdot, \cdot)$ to represent each transition step of edge flipping dynamics. For any t, given the current configuration $X_{t-1} \in \Omega$, the next configuration can be generated by $X_t = \varphi(X_{t-1}, U_t)$, where U_t is the randomness used in the t-th transition step. Specifically,

$$U_t \sim \mathcal{D} \text{ and } U_t = (\ell_t, e_t, b_t, r_t) \in \Omega_R := \{0, 1\} \times E \times \{0, 1\} \times [0, 1],$$

where $\mathcal{D}$ is a distribution such that ℓ_t and b_t are uniform random bits, e_t is a uniform random edge in E, r_t is a uniform random real number, and they are mutually independent. The function φ uses the transition rule defined above to map X_{t-1} to a random state $X_t = \varphi(X_{t-1}, U_t)$, where the randomness of X_t is determined by the randomness of $U_t \sim \mathcal{D}$. The function $\varphi(\cdot, \cdot)$ is called a *grand coupling* of flipping dynamics because

$$\forall \sigma, \tau \in \Omega, \quad \mathbf{Pr}_{U\sim D}\left[\varphi(\sigma, U) = \tau\right] = P(\sigma, \tau).$$

Define a partial ordering $\preceq$ among all vectors in $\{0,1\}^E$: for any $X, Y \in \{0,1\}^E$,

$$X \preceq Y \quad \text{if } X(e) \leq Y(e) \text{ for all } e \in E.$$

Let $X^{\min} = \mathbf{0}$ be the constant vector with value 0 and $X^{\max} = \mathbf{1}$ be the constant vector with value 1. It is straightforward to verify that $X^{\min} \preceq X \preceq X^{\max}$ for all $X \in \{0,1\}^E$. The next lemma shows that the grand coupling φ is monotone with respect to the partial ordering $\preceq$.

Lemma 5.2. *Suppose $0 \leq p_e < 1$ for all $e \in E$ and $0 < \lambda_v \leq 1$ for all $v \in V$. The grand coupling φ of the flipping dynamics for weighted random cluster models is monotone, i.e. for any $\sigma, \tau \in \Omega$ with $\sigma \preceq \tau$, any $U \in \Omega_R$, it holds that $\varphi(\sigma, U) \preceq \varphi(\tau, U)$.*

The proof of Lemma 5.2 is deferred to Section 5.3. With the monotone grand coupling φ, we apply CFTP to the edge flipping dynamics for weighted random cluster models in Algorithm 1.

Algorithm 1: CFTP of the flipping dynamics for weighted random cluster models

---

Input: a weighted random cluster model on graph $G = (V, E)$ with parameters $\boldsymbol{\lambda} = (\lambda_v)_{v \in V}$ and $\boldsymbol{p} = (p_e)_{e \in E}$, where $0 < p_e < 1$ for all $e \in E$ and $0 < \lambda_v \leq 1$ for all $v \in V$.

Output: a perfect sample $X \sim \pi_{wrc}$, where π_{wrc} is the distribution over $\{0,1\}^E$ defined by the input weighted random cluster model.

1 generate $U_t \sim \mathcal{D}$ independently for all integers $t \in (-\infty, -1]$;
2 $T = 1$;
3 **repeat**
4     $X^{\min} = \mathbf{0}$ and $X^{\max} = \mathbf{1}$;
5   **for** $t = -T$ *to* -1 **do**
6         $X^{\min} \leftarrow \varphi(X^{\min}, U_t)$;
7         $X^{\max} \leftarrow \varphi(X^{\max}, U_t)$;
        `//` φ `is the monotone grand coupling in Lemma 5.2`
8     $T \leftarrow 2T$
9 **until** $X^{\min} = X^{\max}$;
10 **return** $X^{\min}$;

Remark 5.3. In Algorithm 1, infinitely many U_t are generated in Line 1. To implement the algorithm, we can first generate U_{-1}, and then generate $(U_t)_{-2T \leq t \leq -T}$ when updating $T \leftarrow 2T$.

Let $T_{\mathcal{D}}$ be the time cost for generating a random sample from $\mathcal{D}$. Let T_φ be the time cost for computing the value of the function φ. Let $T_{\text{mix}}(\cdot)$ denote the mixing time of the edge flipping dynamics for weighted random cluster models. By the standard result of the CFTP for monotone systems [PW96] (also see [LP17, Chapter 25]), we have the following proposition about Algorithm 1.

Proposition 5.4 ([PW96]). *Suppose the input weighted random cluster model satisfies $0 \leq p_e < 1$ for all $e \in E$ and $0 < \lambda_v \leq 1$ for all $v \in V$. Algorithm 1 returns a perfect sample for the stationary distribution of edge flipping dynamics for weighted random cluster models, i.e. the distribution π_{wrc}. The expected running time of Algorithm 1 is $O((T_{\mathcal{D}} + T_\varphi) T_{\text{mix}}(\frac{1}{4e}) \log n)$.*

Now, we are ready to prove Theorem 5.1.

Proof of Theorem 5.1. By definitions of $\mathcal{D}$ and φ, it is straightforward to verify that $T_{\mathcal{D}} = O(1)$ and $T_\varphi = O(n + m)$. By Lemma 3.1, the mixing of edge flipping dynamics is $O(N^4 m^3 \log b)$, where $b = \max_{e \in E} \max\{\frac{1}{p_e}, \frac{1}{1-p_e}\}$, $N = \min\{n, \frac{1}{1-\lambda_{\max}}\}$ and $\lambda_{\max} = \max_{v \in V} \lambda_v$. Hence, the expected running time of Algorithm 1 can be bounded by $O(N^4 m^4 \log n \log b)$. $\square$

5.3. **Proof of monotonicity.** We prove Lemma 5.2. Fix $\sigma, \tau \in \{0,1\}^E$ such that $\sigma \preceq \tau$. Fix $U = (\ell, e, b, r) \in \Omega_R$. Without loss of generality, we assume $\ell = 1$. Otherwise, we have $\varphi(\sigma, U) = \sigma \preceq \tau = \varphi(\tau, U)$. Define two candidate states σ' and τ' by

$$\forall h \in E, \quad \sigma'(h) = \begin{cases} \sigma(h) & \text{if } h \neq e \\ b & \text{if } h = e \end{cases} \quad \text{and} \quad \tau'(h) = \begin{cases} \tau(h) & \text{if } h \neq e \\ b & \text{if } h = e \end{cases}.$$

We prove the lemma by considering the two cases $b = 0$ and $b = 1$.

Suppose $b = 1$. To prove $\varphi(\sigma, U) \preceq \varphi(\tau, U)$, we only need to show that if σ' is accepted, then τ' is also accepted. It suffices to show that

$$\min\left\{1, \frac{\pi_{\mathrm{wrc}}(\sigma')}{\pi_{\mathrm{wrc}}(\sigma)}\right\} \leq \min\left\{1, \frac{\pi_{\mathrm{wrc}}(\tau')}{\pi_{\mathrm{wrc}}(\tau)}\right\}. \tag{27}$$

Since $\sigma \preceq \tau$, here are two cases (1) $\tau(e) = 1$; (2) $\tau(e) = \sigma(e) = 0$. The first case $\tau(e) = 1$ is trivial, because $\tau = \tau'$ and the RHS of (27) is 1, thus (27) holds. We consider the second case $\tau(e) = \sigma(e) = 0$. Let $S_\sigma = \{h \in E \mid \sigma(h) = 1\}$ and $S_\tau = \{h \in E \mid \tau(h) = 1\}$. Define two subgraphs $G_\sigma = (V, S_\sigma)$ and $G_\tau = (V, S_\tau)$. Suppose $e = \{w_1, w_2\}$. Note that $S_\sigma \subseteq S_\tau$. Here are three sub-cases.

- w_1 and w_2 belong to the same connected component in both G_σ and G_τ. It holds that $\frac{\pi_{\mathrm{wrc}}(\sigma')}{\pi_{\mathrm{wrc}}(\sigma)} = \frac{\pi_{\mathrm{wrc}}(\tau')}{\pi_{\mathrm{wrc}}(\tau)} = \frac{p_e}{1-p_e}$, which implies (27).
- w_1 and w_2 belong to the different connected components C_1, C_2 in G_σ, but they belong to the same connected component in G_τ. In this case, we have
$$\frac{\pi_{\mathrm{wrc}}(\sigma')}{\pi_{\mathrm{wrc}}(\sigma)} = \frac{p_e}{1-p_e} \cdot \frac{1 + \prod_{v \in C_1 \cup C_2} \lambda_v}{(1 + \prod_{v \in C_1} \lambda_v)(1 + \prod_{u \in C_2} \lambda_u)} \leq \frac{p_e}{1-p_e} = \frac{\pi_{\mathrm{wrc}}(\tau')}{\pi_{\mathrm{wrc}}(\tau)},$$
which implies (27).
- w_1 and w_2 belong to the different connected components C_1, C_2 in G_σ, and they belong to different connected components C_1', C_2' in G_τ, where $C_1 \subseteq C_1'$ and $C_2 \subseteq C_2'$. In this case, we have
$$\frac{\pi_{\mathrm{wrc}}(\sigma')}{\pi_{\mathrm{wrc}}(\sigma)} = \frac{p_e}{1-p_e} \cdot \frac{1 + \prod_{v \in C_1 \cup C_2} \lambda_v}{(1 + \prod_{v \in C_1} \lambda_v)(1 + \prod_{u \in C_2} \lambda_u)}$$
$$\frac{\pi_{\mathrm{wrc}}(\tau')}{\pi_{\mathrm{wrc}}(\tau)} = \frac{p_e}{1-p_e} \cdot \frac{1 + \prod_{v \in C_1' \cup C_2'} \lambda_v}{(1 + \prod_{v \in C_1'} \lambda_v)(1 + \prod_{u \in C_2'} \lambda_u)}$$
It suffices to show that $\frac{\pi_{\mathrm{wrc}}(\sigma')}{\pi_{\mathrm{wrc}}(\sigma)} \leq \frac{\pi_{\mathrm{wrc}}(\tau')}{\pi_{\mathrm{wrc}}(\tau)}$. Since $0 < \lambda_v \leq 1$ for all $v \in V$, it suffices to show that
$$\forall 0 < x' \leq x \leq 1, 0 < y' \leq y \leq 1, \quad \frac{1 + xy}{(1+x)(1+y)} \leq \frac{1 + x'y'}{(1+x')(1+y')}, \tag{28}$$
which is equivalent to $(x - x')(1 - yy') + (y - y')(1 - xx') \geq 0$. This proves (27).

Suppose $b = 0$. To prove $\varphi(\sigma, U) \preceq \varphi(\tau, U)$, we only need to show that if τ' is accepted, then σ' is also accepted. It suffices to show that

$$\min\left\{1, \frac{\pi_{\mathrm{wrc}}(\tau')}{\pi_{\mathrm{wrc}}(\tau)}\right\} \leq \min\left\{1, \frac{\pi_{\mathrm{wrc}}(\sigma')}{\pi_{\mathrm{wrc}}(\sigma)}\right\}. \tag{29}$$

Since $\sigma \preceq \tau$, here are two cases (1) $\sigma(e) = 0$; (2) $\sigma(e) = \tau(e) = 1$. The first case $\sigma(e) = 0$ is trivial, because $\sigma = \sigma'$ and the RHS of (29) is 1, thus (29) holds. We consider the second case $\tau(e) = \sigma(e) = 0$. Note that σ' and τ' are obtained by removing e from σ and τ respectively. Let $S_{\sigma'} = \{h \in E \mid \sigma'(h) = 1\}$ and $S_{\tau'} = \{h \in E \mid \tau'(h) = 1\}$. Define two subgraphs $G_{\sigma'} = (V, S_{\sigma'})$ and $G_{\tau'} = (V, S_{\tau'})$. Suppose $e = \{w_1, w_2\}$. Note that $S_{\sigma'} \subseteq S_{\tau'}$. Here are three sub-cases.

- w_1 and w_2 belong to the same connected component in both $G_{\sigma'}$ and $G_{\tau'}$. It holds that $\frac{\pi_{\mathrm{wrc}}(\tau')}{\pi_{\mathrm{wrc}}(\tau)} = \frac{\pi_{\mathrm{wrc}}(\sigma')}{\pi_{\mathrm{wrc}}(\sigma)} = \frac{1-p_e}{p_e}$, which implies (29).

- w_1 and w_2 belong to the different connected components C_1, C_2 in $G_{\sigma'}$, but they belong to the same connected component in $G_{\tau'}$. In this case, we have

$$\frac{\pi_{\mathrm{wrc}}(\tau')}{\pi_{\mathrm{wrc}}(\tau)} = \frac{1-p_e}{p_e} \leq \frac{1-p_e}{p_e} \cdot \frac{(1+\prod_{v\in C_1}\lambda_v)(1+\prod_{u\in C_2}\lambda_u)}{1+\prod_{v\in C_1\cup C_2}\lambda_v} = \frac{\pi_{\mathrm{wrc}}(\sigma')}{\pi_{\mathrm{wrc}},(\sigma)}$$

which implies (29).
- w_1 and w_2 belong to the different connected components C_1, C_2 in $G_{\sigma'}$, and they belong to different connected components C_1', C_2' in $G_{\tau'}$, where $C_1 \subseteq C_1'$ and $C_2 \subseteq C_2'$. In this case, we have

$$\frac{\pi_{\mathrm{wrc}}(\tau')}{\pi_{\mathrm{wrc}}(\tau)} = \frac{1-p_e}{p_e} \cdot \frac{(1+\prod_{v\in C_1'}\lambda_v)(1+\prod_{u\in C_2'}\lambda_u)}{1+\prod_{v\in C_1'\cup C_2'}\lambda_v}$$
$$\frac{\pi_{\mathrm{wrc}}(\sigma')}{\pi_{\mathrm{wrc}}(\sigma)} = \frac{1-p_e}{p_e} \cdot \frac{(1+\prod_{v\in C_1}\lambda_v)(1+\prod_{u\in C_2}\lambda_u)}{1+\prod_{v\in C_1\cup C_2}\lambda_v}$$

It suffices to show that $\frac{\pi_{\mathrm{wrc}}(\tau')}{\pi_{\mathrm{wrc}}(\tau)} \leq \frac{\pi_{\mathrm{wrc}}(\sigma)}{\pi_{\mathrm{wrc}}(\sigma')}$, which follows from (28).

Acknowledgement

We thank Mary Cryan for bringing this question to our attention and for some preliminary discussion. This project has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement No. 947778).

References

[AJK+21] Nima Anari, Vishesh Jain, Frederic Koehler, Huy Tuan Pham, and Thuy-Duong Vuong. Entropic independence II: optimal sampling and concentration via restricted modified log-Sobolev inequalities. *arXiv*, abs/2111.03247, 2021.

[CE22] Yuansi Chen and Ronen Eldan. Localization schemes: A framework for proving mixing bounds for Markov chains. *arXiv*, abs/2203.04163, 2022.

[CFYZ22] Xiaoyu Chen, Weiming Feng, Yitong Yin, and Xinyuan Zhang. Optimal mixing for two-state anti-ferromagnetic spin systems. *arXiv*, abs/2203.07771, 2022.

[CLV21] Zongchen Chen, Kuikui Liu, and Eric Vigoda. Optimal mixing of Glauber dynamics: entropy factorization via high-dimensional expansion. In *STOC*, pages 1537–1550. ACM, 2021.

[ES88] Robert G. Edwards and Alan D. Sokal. Generalization of the Fortuin-Kasteleyn-Swendsen-Wang representation and Monte Carlo algorithm. *Phys. Rev. D (3)*, 38(6):2009–2012, 1988.

[FK72] C. M. Fortuin and P. W. Kasteleyn. On the random-cluster model. I. Introduction and relation to other models. *Physica*, 57:536–564, 1972.

[GJ07a] Leslie Ann Goldberg and Mark Jerrum. The complexity of ferromagnetic Ising with local fields. *Comb. Probab. Comput.*, 16(1):43–61, 2007.

[GJ07b] Geoffrey R. Grimmett and Svante Janson. Random even graphs. *Electronic Journal of Combinatorics*, 16, 10 2007.

[GJ18] Heng Guo and Mark Jerrum. Random cluster dynamics for the Ising model is rapidly mixing. *Ann. Appl. Probab.*, 28(2):1292–1313, 2018.

[Gri06] Geoffrey R. Grimmett. *The random-cluster model*, volume 333. Springer Science & Business Media, 2006.

[JS89] Mark Jerrum and Alistair Sinclair. Approximating the permanent. *SIAM J. Comput.*, 18(6):1149–1178, 1989.

[JS93] Mark Jerrum and Alistair Sinclair. Polynomial-time approximation algorithms for the Ising model. *SIAM J. Comput.*, 22(5):1087–1116, 1993.

[JVV86] Mark Jerrum, Leslie G. Valiant, and Vijay V. Vazirani. Random generation of combinatorial structures from a uniform distribution. *Theoretical Computer Science*, 43:169–188, 1986.

[LP17] David A. Levin and Yuval Peres. *Markov chains and mixing times*. American Mathematical Soc., 2017.

[PJG+17] Sejun Park, Yunhun Jang, Andreas Galanis, Jinwoo Shin, Daniel Štefankovič, and Eric Vigoda. Rapid mixing Swendsen-Wang sampler for stochastic partitioned attractive models. In *AISTATS*, volume 54 of *Proceedings of Machine Learning Research*, pages 440–449. PMLR, 2017.

[PW96] James G. Propp and David B. Wilson. Exact sampling with coupled Markov chains and applications to statistical mechanics. *Random Struct. Algorithms*, 9(1-2):223–252, 1996.

[RW99] Dana Randall and David B. Wilson. Sampling spin configurations of an Ising system. In *SODA*, pages 959–960. ACM/SIAM, 1999.

[Sin92] Alistair Sinclair. Improved bounds for mixing rates of Markov chains and multicommodity flow. *Comb. Probab. Comput.*, 1:351–370, 1992.

[SW87] Robert Swendsen and Jian-Sheng Wang. Nonuniversal critical dynamics in Monte Carlo simulations. *Phys. Rev. Lett.*, 58:86–88, 1987.

[Ull14] Mario Ullrich. Swendsen-Wang is faster than single-bond dynamics. *SIAM J. Discret. Math.*, 28(1):37–48, 2014.

[Val08] Leslie G. Valiant. Holographic algorithms. *SIAM J. Comput.*, 37(5):1565–1594, 2008.

[vdW41] Bartel L van der Waerden. Die lange reichweite der regelmässigen atomanordnung in mischkristallen. *Zeitschrift für Physik*, 118(7):473–488, 1941.

Appendix A. Proof of the equivalence result

A.1. **Equivalence between Ising and weighted random cluster models.** Fix a graph $G = (V, E)$. We first show the first equation in (6). Observe that we can decompose the Ising model interaction matrix as

$$f_e^{\mathrm{Ising}} = \begin{pmatrix} \beta_e & 1 \\ 1 & \beta_e \end{pmatrix} = \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix} + \begin{pmatrix} \beta_e - 1 & 0 \\ 0 & \beta_e - 1 \end{pmatrix} =: f_e^{(0)} + f_e^{(1)}.$$

By definition, $f_e^{(1)}$ forces the two endpoints to take the same spin, while $f_e^{(0)}$ poses no requirements. In this way, we can perform an extra enumeration over all the assignments over the edges $\tau : E \to \{0,1\}$, the decompose the effect of f_e^{Ising} into $f_e^{(0)}$ and $f_e^{(1)}$. The partition function of Ising model then becomes

$$\begin{aligned}
\sum_{\sigma \in \{0,1\}^V} \mathrm{wt}_{\mathrm{Ising}}(\sigma) &= \sum_{\sigma \in \{0,1\}^V} \prod_{e=(u,v)\in E} f_e^{\mathrm{Ising}}(\sigma(u), \sigma(v)) \prod_{u \in V} \lambda_u^{\sigma(u)} \\
&= \sum_{\sigma \in \{0,1\}^V} \prod_{e=(u,v)\in E} \left(\sum_{\tau(e)\in\{0,1\}} f_e^{(\tau(e))}(\sigma(u), \sigma(v)) \right) \prod_{u \in V} \lambda_u^{\sigma(u)} \\
(*) \qquad &= \sum_{\tau \in \{0,1\}^E} \sum_{\sigma \in \{0,1\}^V} \prod_{e=(u,v)\in E} f_e^{(\tau(e))}(\sigma(u), \sigma(v)) \prod_{u \in V} \lambda_u^{\sigma(u)}.
\end{aligned}$$

Fix τ. Consider the subgraph $G' = (V, S)$ where S is the set of edges assigned to 1 under τ. Each connected component $C \subseteq V$ of G' must take the same spin in σ, otherwise the contribution to the sum is 0. Let $E_C \subseteq S$ denote all the edges in component C. The total weight of the component C is $\prod_{e \in E_C}(\beta_e - 1)\,(1 + \prod_{u \in C} \lambda_u)$. Combining all components yields

$$\sum_{\sigma \in \{0,1\}^V} \prod_{e=(u,v)\in E} f_e^{(\tau(e))}(\sigma(u), \sigma(v)) \prod_{u \in V} \lambda_u^{\sigma(u)} = \prod_{e \in S}(\beta_e - 1) \prod_{C \in \kappa(V,S)} \left(1 + \prod_{u \in C} \lambda_u\right).$$

And hence

$$\begin{aligned}
(*) &= \sum_{S \subseteq E} \prod_{e \in S} (\beta_e - 1) \prod_{C \in \kappa(V,S)} \left(1 + \prod_{u \in C} \lambda_u\right) \\
&= \left(\prod_{e \in E} \beta_e\right) \cdot \sum_{S \subseteq E} \prod_{e \in S} \left(1 - \frac{1}{\beta_e}\right) \prod_{f \in E \setminus S} \frac{1}{\beta_f} \prod_{C \in \kappa(V,S)} \left(1 + \prod_{u \in C} \lambda_u\right) = Z_{\mathrm{wrc}}(G; 2\boldsymbol{p}, \boldsymbol{\lambda})
\end{aligned}$$

by taking $2p_e = 1 - 1/\beta_e$.

A.2. **Equivalence between Ising and subgraph-world.** To apply Theorem 2.4, we express the Ising model $(G=(V,E);\boldsymbol{\beta},\boldsymbol{\lambda})$ as a Holant problem. Given an Ising model on graph $G=(V,E)$. We define a bipartite graph H with left part $V_1=V$ corresponding to vertices in G and right part $V_2=E$ corresponding to edges in G. Two vertices $v\in V_1$ and $e\in V_2$ are adjacent in graph H if v is incident to e in graph G. By definition, each edge $e=(u,v)$ in G is decomposed into two half-edges (v,e) and (u,e) in graph H.

For any vertex $v\in V_1$, we force the assignment to its incident half-edges to be equal, and furthermore, if they are all ones, then we multiply the weight by λ_v. This yields the signature $[1,0,\cdots,0,\lambda_v]=[1,0]^{\otimes d_v}+\lambda_v[0,1]^{\otimes d_v}$ on each vertex v, where d_v is the degree of v in G. For any edge e in G, its signature is $[\beta_e,1,\beta_e]$ to model the ferromagnetic Ising interaction. Define

$$\mathcal{F}_{\text{Ising}}=\left\{[1,0]^{\otimes d_v}+\lambda_v\,[0,1]^{\otimes d_v}\mid v\in V\right\}\text{ and }\mathcal{G}_{\text{Ising}}=\{[\beta_e,1,\beta_e]\mid e\in E\}\,.$$

It is straightforward to verify

$$\text{Holant}(H;\mathcal{F}_{\text{Ising}}\mid\mathcal{G}_{\text{Ising}})=Z_{\text{Ising}}(G;\boldsymbol{\beta},\boldsymbol{\lambda}).\tag{30}$$

For subgraph-world models, we define a Holant problem on the same bipartite graph H. The signature on each vertex v is defined by $[1,\eta_v,1,\eta_v,\cdots]$, and on each edge $e\in E$, it is defined by $[1-p_e,0,p_e]$. Define

$$\mathcal{F}_{\text{sg}}=\{[1,\eta_v,1,\eta_v,\cdots]\mid v\in V\}\text{ and }\mathcal{G}_{\text{sg}}=\{[1-p_e,0,p_e]\mid e\in E\}\,.$$

It is straightforward to verify

$$\text{Holant}(H;\mathcal{F}_{\text{sg}}\mid\mathcal{G}_{\text{sg}})=Z_{\text{sg}}(G;\boldsymbol{p},\boldsymbol{\eta}).\tag{31}$$

Take $T=\left(\begin{smallmatrix}1&1\\1&-1\end{smallmatrix}\right)$. Let $p_e=\frac{1}{2}\left(1-\frac{1}{\beta_e}\right)$. It holds that

$$\left(\boldsymbol{T}^{-1}\right)^{\otimes 2}(\beta_e,1,1,\beta_e)^{\top}=\left(\frac{\beta_e+1}{2},0,0,\frac{\beta_e-1}{2}\right)^{\top}=\beta_e\left[\frac{\beta_e+1}{2\beta_e},0,\frac{\beta_e-1}{2\beta_e}\right]=\beta_e[1-p_e,0,p_e].$$

Let $\eta_v=\frac{1-\lambda_v}{1+\lambda_v}$. We have

$$\left((1,0)^{\otimes d_v}+\lambda_v\,(0,1)^{\otimes d_v}\right)\boldsymbol{T}^{\otimes d_v}=(1,1)^{\otimes d_v}+\lambda_v\,(1,-1)^{\otimes d_v}=(1+\lambda_v)\,[1,\eta_v,1,\eta_v,\cdots]\,.$$

Combining Theorem 2.4, (30) and (31) with the above, it holds that

$$Z_{\text{Ising}}(G;\boldsymbol{\beta},\boldsymbol{\lambda})=\left(\prod_{v\in V}(1+\lambda_v)\right)\left(\prod_{e\in E}\beta_e\right)Z_{\text{sg}}(G;\boldsymbol{p},\boldsymbol{\eta}).$$

Appendix B. Canonical path for the subgraph-world model

In this subsection we prove Lemma 3.5. Suppose there is a fixed ordering $\prec$ for all paths and cycles of the graph G. For any pair of subgraphs X,Y, construct the canonical path $\gamma_{X,Y}$ as follows. Consider the symmetric difference $X\oplus Y$. Let $2k$ be the number of its odd-degree vertices. Then, $X\oplus Y$ can be decomposed into an edge-disjoint union of exactly k paths $P_1,\cdots,P_k$ and cycles $C_1,\cdots,C_{k'}$. We pick the unique one such that $P_1,\cdots,P_k,C_1,\cdots,C_{k'}$ is the first one in lexicographic order yielded by $\prec$. Then the path $\gamma_{X,Y}$ is defined by processing $P_1,\cdots,P_k,C_1,\cdots,C_{k'}$ one by one, and for each of them, we unwind all the edges (if the current edge is present in X, then delete it, or otherwise, add it). To ensure the path is unique, when unwinding P_i for some i, we start at the open vertex with smaller index, and move along the path; when unwinding C_i, we start at the vertex of smallest index, and move in the direction that the neighbour of starting vertex has a smaller index than the other one.

Consider the transition from Z to Z' where $|Z\oplus Z'|=1$, and let $\Lambda(Z,Z')$ be all the pairs (X,Y) such that the path $\gamma_{X,Y}$ uses the transition (Z,Z'). Define the mapping $\varphi_{Z,Z'}(X,Y):=X\oplus Y\oplus Z$. We claim that $\varphi_{Z,Z'}$ is an injective mapping. This is because we can first recover $X\oplus Y$ by knowing Z and $U:=\varphi_{Z,Z'}(X,Y)$, and hence the unique edge-disjoint decomposition $P_1,\cdots,P_k,C_1,\cdots,C_{k'}$. By checking $Z\oplus Z'$, we know the stage of unwinding we are currently on, and therefore we can continue the construction from Z', and when finished, Y is obtained. Similarly, the starting state X can be obtained by reversing the process.

Finally, we compute $\sum_{\gamma:(Z,Z')\in\gamma} w(\gamma)$. Recall that $U \oplus Z = X \oplus Y$. We then consider what may contribute to the ratio $\frac{\pi_{\text{sg}}(X)\pi_{\text{sg}}(Y)}{\pi_{\text{sg}}(U)\pi_{\text{sg}}(Z)}$, or equivalently, $\frac{\text{wt}_{\text{sg}}(X)\text{wt}_{\text{sg}}(Y)}{\text{wt}_{\text{sg}}(U)\text{wt}_{\text{sg}}(Z)}$. Recall from (4) that there may be two kinds of factors in this ratio, the one from the edges p_e and the one from the vertices η_v. Note that the construction of $\gamma_{X,Y}$ ensures that $Z \subseteq X \cup Y$ and $Z \supseteq X \cap Y$, implying that $U \cup Z = X \cup Y$, and hence $U \cap Z = X \cap Y$. This means all the p_e terms in the numerator and denominator of $\frac{\text{wt}_{\text{sg}}(X)\text{wt}_{\text{sg}}(Y)}{\text{wt}_{\text{sg}}(U)\text{wt}_{\text{sg}}(Z)}$ cancel, and the remaining terms can only come from the odd vertices that cause discrepancy. It is not hard to see that such discrepancy only arises from the current cycle or path that is being processed, and more specifically, the vertex incident to the two edges wound before and after Z, which contributes twice, and the starting vertex of the current cycle, which contributes twice as well. Therefore,

$$\frac{\pi_{\text{sg}}(X)\pi_{\text{sg}}(Y)}{\pi_{\text{sg}}(U)\pi_{\text{sg}}(Z)} \leq \eta_{\min}^{-4}. \tag{32}$$

Be cautious that $0 < \eta_v < 1$ for all v.

Using (32), it can be obtained that

$$\begin{aligned}
&\text{(By definition)} && \sum_{\gamma:(Z,Z')\in\gamma} w(\gamma) = \sum_{X,Y:(Z,Z')\in\gamma_{X,Y}} \pi_{\text{sg}}(X)\pi_{\text{sg}}(Y) \\
&\text{(By (32))} && \leq \eta_{\min}^{-4} \sum_{X,Y:(Z,Z')\in\gamma_{X,Y}} \pi_{\text{sg}}(Z)\pi_{\text{sg}}(\varphi_{Z,Z'}(X,Y)) \\
&(\varphi_{Z,Z'} \text{ is injective}) && \leq \eta_{\min}^{-4}\pi_{\text{sg}}(Z).
\end{aligned}$$

We construct another mapping $\varphi'_{Z,Z'}(X,Y) := X \oplus Y \oplus Z'$, and the same argument shows that

$$\sum_{\gamma:(Z,Z')\in\gamma} w(\gamma) \leq \eta_{\min}^{-4}\pi_{\text{sg}}(Z').$$

Appendix C. Proof of the self-adjointness

To prove Proposition 2.3, we need the following proposition.

Proposition C.1. *For any functions $f : \{0,1\}^V \to \mathbb{R}$ and $g : 2^E \to \mathbb{R}$, it holds that*

$$\langle f, P_{\mathcal{I}\to\mathcal{R}}g\rangle_{\pi_{\text{Ising}}} = \langle P_{\mathcal{R}\to\mathcal{I}}f, g\rangle_{\pi_{\text{wrc}}}. \tag{33}$$

Proof. Let $D_{\text{Ising}} = \text{diag}(\pi_{\text{Ising}})$ and $D_{\text{wrc}} = \text{diag}(\pi_{\text{wrc}})$ denote the diagonal matrices induced from vectors π_{Ising} and π_{wrc} respectively. We have

$$\langle f, P_{\mathcal{I}\to\mathcal{R}}g\rangle_{\pi_{\text{Ising}}} = f^T D_{\text{Ising}} P_{\mathcal{I}\to\mathcal{R}} g \quad \text{and} \quad \langle P_{\mathcal{R}\to\mathcal{I}}f, g\rangle_{\pi_{\text{wrc}}} = f^T P_{\mathcal{R}\to\mathcal{I}}^T D_{\text{wrc}} g.$$

For any $\sigma \in \{0,1\}^V$ and $S \subseteq E$, we show that

$$\left(D_{\text{Ising}}P_{\mathcal{I}\to\mathcal{R}}\right)(\sigma, S) = \left(P_{\mathcal{R}\to\mathcal{I}}^T D_{\text{wrc}}\right)(\sigma, S)$$

Recall $M(\sigma) = \{\{u,v\} \in E \mid \sigma_u = \sigma_v\}$. It holds that

$$\begin{aligned}
\left(D_{\text{Ising}}P_{\mathcal{I}\to\mathcal{R}}\right)(\sigma, S) &= \mathbb{I}[S \subseteq M(\sigma)] \cdot \pi_{\text{Ising}}(\sigma) \cdot \prod_{e\in S}\left(1 - \frac{1}{\beta_e}\right) \prod_{f\in M(\sigma\setminus S)} \frac{1}{\beta_f} \\
&= \mathbb{I}[S \subseteq M(\sigma)] \cdot \frac{1}{Z_{\text{Ising}}} \cdot \prod_{v\in V} \lambda_v^{\sigma(v)} \prod_{h\in M(\sigma)} \beta_h \prod_{e\in S}\left(1 - \frac{1}{\beta_e}\right) \prod_{f\in M(\sigma\setminus S)} \frac{1}{\beta_f} \\
&= \mathbb{I}[S \subseteq M(\sigma)] \cdot \frac{1}{Z_{\text{Ising}}} \cdot \prod_{v\in V} \lambda_v^{\sigma(v)} \prod_{e\in S}(\beta_e - 1).
\end{aligned} \tag{34}$$

Recall $\kappa(V,S)$ is the set of all connected components of graph (V,S). It holds that

$$\begin{aligned}\left(P_{\mathcal{R}\to\mathcal{I}}^T D_{\mathrm{wrc}}\right)(\sigma,S) &= \mathbb{I}[S\subseteq M(\sigma)]\cdot\pi_{\mathrm{wrc}}(S)\cdot\prod_{C\in\kappa(V,S)}\frac{\prod_{v\in C}\lambda_v^{\sigma(v)}}{1+\prod_{v\in C}\lambda_v}\\
&=\mathbb{I}[S\subseteq M(\sigma)]\cdot\frac{1}{Z_{\mathrm{wrc}}}\cdot\prod_{e\in S}\left(1-\frac{1}{\beta_e}\right)\prod_{f\in E\setminus S}\frac{1}{\beta_f}\prod_{C\in\kappa(V,S)}\left(1+\prod_{u\in C}\lambda_u\right)\cdot\prod_{C\in\kappa(V,S)}\frac{\prod_{v\in C}\lambda_v^{\sigma(v)}}{1+\prod_{v\in C}\lambda_v}\\
&=\mathbb{I}[S\subseteq M(\sigma)]\cdot\frac{1}{Z_{\mathrm{wrc}}}\cdot\prod_{e\in S}\left(1-\frac{1}{\beta_e}\right)\prod_{f\in E\setminus S}\frac{1}{\beta_f}\prod_{v\in V}\lambda_v^{\sigma(v)}\\
&=\mathbb{I}[S\subseteq M(\sigma)]\cdot\frac{1}{Z_{\mathrm{wrc}}}\cdot\prod_{h\in E}\frac{1}{\beta_h}\prod_{v\in V}\lambda_v^{\sigma(v)}\prod_{e\in S}(\beta_e-1)\end{aligned}\tag{35}$$

By Proposition 2.1, we know that

$$\left(\prod_{e\in E}\beta_e\right)Z_{\mathrm{wrc}}=Z_{\mathrm{Ising}}.$$

Combining above equation with (34) and (35) proves $\left(D_{\mathrm{Ising}}P_{\mathcal{I}\to\mathcal{R}}\right)(\sigma,S)=\left(P_{\mathcal{R}\to\mathcal{I}}^T D_{\mathrm{wrc}}\right)(\sigma,S)$. □

Proof of Proposition 2.3. Fix $\sigma\in\{0,1\}^V$ and $S\subseteq E$. Define $f_\sigma:\{0,1\}^V\to\mathbb{R}$ by $f_\sigma(\sigma)=1$ and $f_\sigma(\tau)=0$ for all $\tau\in\{0,1\}^V$. Define $g_S:2^E\to\mathbb{R}$ by $g_S(S)=1$ and $g_S(S')=0$ for all $S'\neq S$. By Proposition C.1,

$$\pi_{\mathrm{Ising}}(\sigma)P_{\mathcal{I}\to\mathcal{R}}(\sigma,S)=\langle f_\sigma,P_{\mathcal{I}\to\mathcal{R}}g_S\rangle_{\pi_{\mathrm{Ising}}}=\langle P_{\mathcal{R}\to\mathcal{I}}f_\sigma,g_S\rangle_{\pi_{\mathrm{wrc}}}=\pi_{\mathrm{wrc}}(S)P_{\mathcal{R}\to\mathcal{I}}(S,\sigma),$$

which implies the first part.

For the second part, consider equation (33). We take f as all 1 vector. Fix $S\subseteq E$. Let $g(S)=1$ and $g(S')=0$ for all $S'\neq S$. We have

$$\sum_{x\in\{0,1\}^V}\pi_{\mathrm{Ising}}(x)P_{\mathcal{I}\to\mathcal{R}}(x,S)=\langle f,P_{\mathcal{I}\to\mathcal{R}}g\rangle_{\pi_{\mathrm{Ising}}}=\langle P_{\mathcal{R}\to\mathcal{I}}f,g\rangle_{\pi_{\mathrm{wrc}}}=\pi_{\mathrm{wrc}}(S),$$

which implies $\pi_{\mathrm{Ising}}P_{\mathcal{I}\to\mathcal{R}}=\pi_{\mathrm{wrc}}$. The equation $\pi_{\mathrm{wrc}}P_{\mathcal{R}\to\mathcal{I}}=\pi_{\mathrm{Ising}}$ can be proved in a similar way.

Next, consider two functions $h_1,h_2:\{0,1\}^V\to\mathbb{R}$. We have

$$\begin{aligned}&\langle h_1,P_{\mathrm{SW}}^{\mathrm{Ising}}h_2\rangle_{\pi_{\mathrm{Ising}}}=\langle h_1,P_{\mathcal{I}\to\mathcal{R}}P_{\mathcal{R}\to\mathcal{I}}h_2\rangle_{\pi_{\mathrm{Ising}}}\overset{(*)}{=}\langle P_{\mathcal{R}\to\mathcal{I}}h_1,P_{\mathcal{R}\to\mathcal{I}}h_2\rangle_{\pi_{\mathrm{wrc}}};\\
&\langle P_{\mathrm{SW}}^{\mathrm{Ising}}h_1,h_2\rangle_{\pi_{\mathrm{Ising}}}=\langle P_{\mathcal{I}\to\mathcal{R}}P_{\mathcal{R}\to\mathcal{I}}h_1,h_2\rangle_{\pi_{\mathrm{Ising}}}=\langle h_2,P_{\mathcal{I}\to\mathcal{R}}P_{\mathcal{R}\to\mathcal{I}}h_1\rangle_{\pi_{\mathrm{Ising}}}\overset{(\star)}{=}\langle P_{\mathcal{R}\to\mathcal{I}}h_2,P_{\mathcal{R}\to\mathcal{I}}h_1\rangle_{\pi_{\mathrm{wrc}}},\end{aligned}$$

where equations $(*)$ and $(\star)$ holds due to (33). Combining above two equations together, we have $\langle h_1,P_{\mathrm{SW}}^{\mathrm{Ising}}h_2\rangle_{\pi_{\mathrm{Ising}}}=\langle P_{\mathrm{SW}}^{\mathrm{Ising}}h_1,h_2\rangle_{\pi_{\mathrm{Ising}}}$, which implies $P_{\mathrm{SW}}^{\mathrm{Ising}}$ is reversible with respect to π_{Ising}. The reversibility of $P_{\mathrm{SW}}^{\mathrm{wrc}}$ can be proved in a similar way.

Finally, let $h_1=h_2=h$. We have

$$\langle h,P_{\mathrm{SW}}^{\mathrm{Ising}}h\rangle_{\pi_{\mathrm{Ising}}}=\langle P_{\mathcal{R}\to\mathcal{I}}h,P_{\mathcal{R}\to\mathcal{I}}h\rangle_{\pi_{\mathrm{wrc}}}=\sum_{S\subseteq E}\pi_{\mathrm{wrc}}(S)\left(\sum_{\sigma\in\{0,1\}^V}P_{\mathcal{R}\to\mathcal{I}}(S,\sigma)h(\sigma)\right)^2\geq 0,$$

which implies $P_{\mathrm{SW}}^{\mathrm{Ising}}$ is positive semidefinite. The result for $P_{\mathrm{SW}}^{\mathrm{wrc}}$ can be proved in a similar way. □