SWENDSEN-WANG DYNAMICS FOR THE FERROMAGNETIC ISING MODEL WITH EXTERNAL FIELDS

WEIMING FENG, HENG GUO, JIAHENG WANG

Abstract. We study the sampling problem for the ferromagnetic Ising model with consistent external fields, and in particular, Swendsen-Wang dynamics on this model. We introduce a new grand model unifying two closely related models: the subgraph world and the random cluster model. Through this new viewpoint, we show:

(1) polynomial mixing time bounds for Swendsen-Wang dynamics and (edge-flipping) Glauber dynamics of the random cluster model, generalising the bounds and simplifying the proofs for the no-field case by Guo and Jerrum (2018);
(2) near linear mixing time for the two dynamics above if the maximum degree is bounded and all fields are (consistent and) bounded away from 1.

Keywords. Ising model, random cluster model, Markov chain, mixing time, Swendsen-Wang dynamics, holographic transformation

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.

(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

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 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 2.3.2 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 start stating our 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

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

where $d_{\mathrm{TV}}(\mu, \pi) = \frac{1}{2} \sum_{\sigma \in \Omega} |\mu(\sigma) - \pi(\sigma)|$ is the *total variation distance* between two distributions.

First, we show that the edge-flipping dynamics for the weighted random cluster model mixes in polynomial-time. By adapting [Ull14] to the case with fields, 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.

Theorem 1.1. *Let $1 < \beta_{\min} \leq \beta_{\max}$ be constants. For any ferromagnetic Ising model on graph $G = (V, E)$ with parameters $(\beta_e)_{e \in E}$ and $(\lambda_v)_{v \in V}$, where $\beta_{\min} \leq \beta_e \leq \beta_{\max}$ and $0 < \lambda_v \leq 1$, the mixing time of Swendsen-Wang dynamics is $O(N^4 m^2 \left(m + \log \frac{1}{\varepsilon}\right))$, where $N = \min\left\{n, \frac{1}{1-\lambda_{\max}}\right\}$, $\lambda_{\max} = \max_{v \in V} \lambda_v$, $n = |V|$ and $m = |E|$.*

Note that if $\beta_e = 1$ for some $e \in E$, it is equivalent to remove such an edge. Also if $\lambda_v = 0$ for some $v \in V$, it is equivalent to pin v to 0 and then absorb v into its neighbours external fields. Thus, any ferromagnetic Ising model with consistent external fields can be transformed into one satisfying the condition of Theorem 1.1. The big-O notation hides a constant factor depending only on $\beta_{\min}$ and $\beta_{\max}$. See (33) for the details of the hidden constant.

The main technical innovation of ours is to introduce a grand model, which incorporates both the so-called subgraph world [JS93] and the random cluster model. The subgraph world assigns weights

[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 $O(|V|^{1/4})$.

to subsets of edges, where each vertex of an odd degree in the induced graph suffers a penalty corresponding to its external field (or the lack thereof). Detailed definitions of the basic models are given in Section 2.1.

The main inspiration of our grand model is the coupling given by Grimmett and Janson [GJ07b] between the two models above without external fields. Our model assigns 3 states to each edge: 0, 1, 2. A sample of our model can be generated as follows: first, we sample a subset of edges from the subgraph world model, and assign 1 to them; then, we assign 0 or 2 to each remaining independently with a carefully chosen probability. Detailed definitions are in Section 3.1. The marginal distribution of edges assigned 1 clearly follow the subgraph world distribution, and we show that the non-zero edges follow the weighted random cluster model (Lemma 3.3). This last step is done using Valiant's holographic transformations [Val08]. It is also a generalisation of [GJ07b] in the presence of external fields.

We give a polynomial upper bound of the mixing time of the Glauber dynamics for the grand model in Section 4 via the method of canonical paths [JS89]. Our construction of the canonical path is a variation of the original paths by Jerrum and Sinclair [JS93]. The projection of this dynamics to the non-zero edges is exactly the Glauber dynamics for the weighted random cluster model. We show that this project does not slow down the dynamics (Section 6), and therefore mixing time bounds for the weighted random cluster model is a direct consequence. This implies Theorem 1.1. When there is no field, our argument recovers the result of Guo and Jerrum [GJ18]. However, our argument is both simpler and more general.

Another important feature of the grand model is that it gives a Gibbs distribution, in the sense that variables are independent if we condition on a subset of edges which disconnect the graph. This is a feature absent in the random cluster models. Recently, there is a lot of progress in analysing the mixing time of dynamics for Gibbs distributions, especially using the notion of spectral independence [ALO20]. Since the domain in our case is not Boolean, we use a generalisation of [FGYZ21] (see also [CGŠV21] for a different generalisation). An important development along this line is that in bounded degree graphs, spectral independence implies near-linear mixing time of dynamics for the Gibbs distribution [CLV21a]. To be more precise, they showed a constant decay rate for the relative entropy in this setting.

Back to the Ising model, when the maximum degree is bounded and all external fields are bounded away from 1, Chen, Liu, and Vigoda [CLV21b] established spectral independence for the subgraph world model. Using our grand model, this implies spectral independence for the random cluster model as well. However, since the random cluster model does not have conditional independence, the method of [CLV21a] does not apply. Instead, we show spectral independence for the grand model in this setting. Thus, via the method of [CLV21a] and exploiting the fact that the grand model is indeed a Gibbs distribution, we obtain a constant decay rate for the relative entropy for the (edge-flipping) Glauber dynamics for the weighted random cluster model. (We apply the result of projecting chains in Section 6 here again.)

However, this is still not quite enough to obtain desired mixing time bounds for the Swendsen-Wang dynamics. The reason is that the aforementioned comparison techniques of [Ull14] is an analysis of the eigenvalues of transition matrices, and thus it works only for spectral gaps but not for relative entropies. For this last step, we introduce a new comparison argument for the decay rate of relative entropies between the (edge-flipping) Glauber dynamics and the Swendsen-Wang dynamics in Section 7.

To be more precise, we perform a careful analysis between the Glauber dynamics and the so-called "single-bond" dynamics introduced in [Ull14]. Our analysis utilises ideas from high-dimensional random walks [ALOV19, CGM21]. For both the Glauber dynamics and the single-bond dynamics, we decompose them into two sub-steps: the down-walk and the up-walk. Using our grand model, we bound the decay rate of relative entropy for the down-walk of Glauber dynamics. By a new comparison argument, we show that the relative entropy also decays for the down-walk of "single-bond" dynamics with a slightly weaker rate. Finally, we compare the down-walk of "single-bond" dynamics to the Swendsen-Wang dynamics via a simple application of the data processing inequality. Our analysis not only works for the decay of relative entropy, but also gives a very simple proof (see Remark 7.5) to the main result in [Ull14].

Theorem 1.2. *Let $1 < \beta_{\min} \le \beta_{\max}, \Delta \ge 3$ and $0 < \delta < 1$ be constants. For any ferromagnetic Ising model on graph $G = (V, E)$ with parameters $(\beta_e)_{e\in E}$ and $(\lambda_v)_{v\in V}$, where $\beta_{\min} \le \beta_e \le \beta_{\max}$, $0 < \lambda_v \le 1-\delta$ and the maximum degree of G is at most Δ, the mixing time of Swendsen-Wang dynamics is $O(n \log \frac{n}{\varepsilon})$, where $n = |V|$.*

By the same reasoning below Theorem 1.1, we do not lose generality by assuming $\beta_{\min} > 1$ and $\lambda_v > 0$ in Theorem 1.2. The big-O notation hides a constant factor depending only on $\beta_{\min}, \beta_{\max}, \delta$ and Δ. See (34) for the details of the hidden constant.

Comparing to Theorem 1.1, Theorem 1.2 has a faster mixing time bound but comes with two further assumptions: constant degree bound and no trivial field. It would be very interesting to relax either restriction. Essentially, the bottleneck in Theorem 1.1 comes from the overhead in the canonical path [JS93] or multicommodity flow method [Sin92] arguments. Unfortunately, there does not seem to be any progress in improving the mixing time bound of these methods in the last three decades. Instead, Theorem 1.2 relies on recent progress of analysing spin systems via high-dimensional random walks [CLV21a, CLV21b]. This method has very recently been generalised to bypass the bounded degree restriction [AJK+21, CFYZ22, CE22] in various models. It is an interesting question whether this is also possible in the setting of Theorem 1.2. To bypass the no trivial field restriction, we would need a new spectral independence bound, for which there seems to be less progress. In particular, it seems hard to explain the $\Theta(n^{1/4})$ mixing time on the complete graph without fields [LNNP14] with spectral independence.

Previously, most studies on Swendsen-Wang dynamics focus on the case without fields (with the exception of [PJG+17] discussed earlier), and are usually for the more general Potts model instead of just the Ising model. Very sharp mixing time bounds have been obtained recently, either for special cases of graphs such as $\mathbb{Z}^d$ [BCP+21], or in the tree uniqueness region for general graphs [BCC+22]. Our Theorem 1.2 does not have these restrictions, but it only works with the presence of non-trivial external fields for the Ising model. In the settings of Theorem 1.2, we conjecture that the sharp mixing time bound is $O(\log n)$. The current argument reduces the analysis of SW dynamics to that of the single-bond dynamics, as the latter is "no-faster" in a technical sense. However, since the diameter of the single-bond dynamics is $\Omega(n)$, it cannot mix in $o(n)$ time [LP17, §7.1.2], making this line of argument difficult to approach the conjectured sharp bound for SW dynamics.

Lastly, by applying the monotone CFTP [PW96], we obtain perfect sampling versions of the (edge-flipping) Glauber dynamics in Section 8 for the weighted random cluster models. Using that, we achieve perfectly sampling for the ferromagnetic Ising model with consistent external fields.

Theorem 1.3. *Let $1 < \beta_{\min} \le \beta_{\max}$ be two constants. There is a perfect sampling algorithm such that given any ferromagnetic Ising model on graph $G = (V, E)$ with parameters $(\beta_e)_{e\in E}$ and $(\lambda_v)_{v\in V}$, where $\beta_{\min} \le \beta_e \le \beta_{\max}$ and $0 < \lambda_v < 1$, the algorithm returns a perfect sample in expected time $O(N^4 m^4 \log n)$, where $N = \min\left\{n, \frac{1}{1-\lambda_{\max}}\right\}$ and $\lambda_{\max} = \max_{v\in V} \lambda_v$.*

Furthermore, if G has bounded maximum degree $\Delta = O(1)$ and there exists a constant $\delta > 0$ such that $\lambda_v \le 1 - \delta$ for all $v \in V$, the algorithm runs in expected time $O(n^2 \log^2 n)$.

We remark that the overhead in monotone CFTP is $O(\log |V|)$ and there is an extra factor $m = |E|$ to implement each step of CFTP. The hidden constants can be found in (44).

A natural question is if we can relax the assumptions on the parameters in Theorem 1.1, 1.2, and 1.3. For anti-ferromagnetic Ising models, the sampling problem (either approximate or perfect) has no polynomial-time algorithm unless **NP** = **RP** [JS93]. Even restricted to the ferromagnetic case, Goldberg and Jerrum [GJ07a] showed that the problem becomes #BIS-equivalent when inconsistent 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 in Theorem 1.1, 1.2, and 1.3.

Subsequent work. After our paper was posted on arXiv, Chen and Zhang [CZ23] gave a sampling algorithm of the ferromagnetic Ising model on *any graph* with running time $\tilde{O}(m)$, where m is the number of edges, providing all the external fields are bounded away from 1 and all the edge interactions are consistent and bounded away from 1. This is a setting similar to our Theorem 1.2 without the

bounded degree assumption. Their work relies heavily on our coupling result, Lemma 3.3. Furthermore, their algorithm is based on the field dynamics introduced in [CFYZ21], and does not imply mixing time bounds for either the Glauber dynamics or the Swendsen-Wang dynamics considered in the current paper.

2. Preliminaries

2.1. **The models and their equivalences.** Here we define the weighted random cluster model, and the subgraph-world model. Then we give some equivalence results between them and the ferromagnetic Ising model.

2.1.1. *Weighted random cluster model.* The standard random cluster model (at $q = 2$) is equivalent to the ferromagnetic Ising model without external fields. 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. The (general) standard random cluster model allows a uniform weight q for each connected component, and in the special case of $\lambda_v = 1$ for all $v \in V$, the weighted random cluster model degenerates to the standard random cluster model at $q = 2$. On the other hand, in our model the weight of each cluster depends on the vertices inside it, which makes it different from the standard random cluster models.

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. *Equivalences of the three models.* We have the following equivalence result among the ferromagnetic Ising model with external fields, the weighted random cluster model, and the subgraph-world 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*

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

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

In addition, there are also probabilistic equivalence relations among the models, which will be the topic in Section 3.

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 (see e.g. [vdW41, FK72, Gri06]). Formally,

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

which is a special case of Proposition 2.1.

2.2. **f-divergences.** A widely-used quantity for measuring the difference between two distributions is the f-divergence. Let $f : \mathbb{R}_{\ge 0} \to \mathbb{R}$ be a *convex* function satisfying $f(1) = 0$. Let μ be a distribution with (finite) support $\Omega = \Omega(\mu)$. Let ν be a distributions with support $\Omega(\nu) \subseteq \Omega$. The f-divergence between ν and μ is defined by

$$D_f(\nu \parallel \mu) := \mathbf{E}_{X\sim\mu}\left[f\left(\frac{\nu(X)}{\mu(X)}\right)\right].$$

In this paper, we consider three important f-divergences: the total variation distance, the χ^2-divergence, and the Kullback-Leibler divergence (KL divergence).

Let $f(x) = \frac{1}{2}|x - 1|$. The *total variation distance* between ν and μ is defined by

$$d_{\text{TV}}(\nu, \mu) := \frac{1}{2}\sum_{x\in\Omega}|\nu(x) - \mu(x)|.$$

We say the random variable $(X, Y) \in \Omega\times\Omega$ is a *coupling* between ν and μ if the marginal distributions satisfy $X \sim \nu$ and $Y \sim \mu$. The *coupling inequality* states that for any coupling (X, Y),

$$\mathbf{Pr}[X \ne Y] \ge d_{\text{TV}}(\nu, \mu), \tag{7}$$

and there exists an *optimal coupling* between ν and μ such that equality holds.

Let $f(x) = x^2 - 1$. The *χ^2 divergence* between ν and μ is defined by

$$D_{\chi^2}(\nu \parallel \mu) := \sum_{x\in\Omega}\frac{\nu^2(x)}{\mu(x)} - 1.$$

A similar notion is the relative variance of a function $g : \Omega \to \mathbb{R}_{\ge 0}$ over μ:

$$\text{Var}_\mu(g) = \mathbf{E}_\mu[g^2] - \mathbf{E}^2_\mu[g] = \sum_{x\in\Omega}\mu(x)g^2(x) - \left(\sum_{x\in\Omega}\mu(x)g(x)\right)^2.$$

Clearly, if $g(x) = \frac{\nu(x)}{\mu(x)}$, then $\text{Var}_\mu(g) = D_{\chi^2}(\nu \parallel \mu)$. The following relation is well-known

$$d_{\text{TV}}(\nu, \mu) \le \sqrt{D_{\chi^2}(\nu \parallel \mu)}. \tag{8}$$

Let $f(x) = x \log x$. The *Kullback-Leibler divergence* (KL divergence) is defined by

$$D_{\mathrm{KL}}(\nu \parallel \mu) := \sum_{x \in \Omega} \nu(x) \log\left(\frac{\nu(x)}{\mu(x)}\right).$$

A similar notion is the relative entropy of a function $g : \Omega \to \mathbb{R}_{\geq 0}$ over μ:

$$\mathrm{Ent}_\mu(g) := \mathbf{E}_\mu[g \log g] - \mathbf{E}_\mu[g] \log \mathbf{E}_\mu[g] = \sum_{x \in \Omega} \mu(x) g(x) \log g(x) - \left(\sum_{x \in \Omega} \mu(x) g(x)\right) \log\left(\sum_{x \in \Omega} \mu(x) g(x)\right),$$

where the convention is that $0 \log 0 = 0$. Clearly, if $g(x) = \frac{\nu(x)}{\mu(x)}$, $\mathrm{Ent}_\mu(g) = D_{\mathrm{KL}}(\nu \parallel \mu)$. The following Pinsker's inequality is well known

$$d_{\mathrm{TV}}(\nu, \mu) \leq \sqrt{\frac{D_{\mathrm{KL}}(\nu \parallel \mu)}{2}}. \tag{9}$$

For any stochastic matrix P that transforms any $x \in \Omega$ to a random $y \in \Omega'$ (Ω' is not necessarily the same as Ω), the following data-processing inequality is well-known: for any f-divergence,

$$D_f(\nu P \parallel \mu P) \leq D_f(\nu \parallel \mu).$$

2.3. **Markov chains and down-up walks.** Let Ω be a finite state space. Let $(X_t)_{t \geq 0}$ be a Markov chain over Ω and P denote the transition matrix. We say P is

- *irreducible* if for any $x, y \in \Omega$, there exists $t > 0$ such that $P^t(x, y) > 0$;
- *aperiodic* if $\gcd\{t \mid P^t(x, x) > 0\} = 1$ for all $x \in \Omega$;
- *reversible* with respect to μ if the following detailed balance equation holds

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

We say the distribution μ is a stationary distribution of P if $\mu P = \mu$. If P is reversible with respect to μ, then μ is a stationary distribution of P. If P is both irreducible and aperiodic, then P has a unique stationary distribution. The *mixing time* of P is defined by

$$\forall \varepsilon > 0, \quad t_{\mathrm{mix}}(P, \varepsilon) := \max_{x \in \Omega} \min\{t \mid d_{\mathrm{TV}}\left(P^t(x, \cdot), \mu\right) \leq \varepsilon\}.$$

In this paper, we consider two Markov chains: Glauber dynamics and Swendsen-Wang dynamics. It will be convenient for us to view Glauber dynamics as a so-called "down-up" walk, which we will define next.

Let Ω_0 and Ω_1 denote two finite state spaces. Let μ_0 and μ_1 denote two distributions over Ω_0 and Ω_1 respectively. For $f, g : \Omega_i \to \mathbb{R}$, define $\langle f, g \rangle_{\mu_i} = \sum_{x \in \Omega_i} \mu_i(x) f(x) g(x)$. Let $P^\uparrow : \Omega_0 \times \Omega_1 \to \mathbb{R}_{\geq 0}$ and $P^\downarrow : \Omega_1 \times \Omega_0 \to \mathbb{R}_{\geq 0}$ denote two transition matrices. We say $P^\uparrow$ and $P^\downarrow$ are a pair of adjoint operator if

$$\forall f : \Omega_0 \to \mathbb{R},\ g : \Omega_1 \to \mathbb{R}, \quad \langle f, P^\uparrow g \rangle_{\mu_0} = \langle P^\downarrow f, g \rangle_{\mu_1}.$$

The following equation holds for adjoint $P^\uparrow$ and $P^\downarrow$:

$$\forall x_0 \in \Omega_0, x_1 \in \Omega_1, \quad \mu_0(x_0) P^\uparrow(x_0, x_1) = \mu_1(x_1) P^\downarrow(x_1, x_0).$$

Moreover, for any distribution ν over Ω_1 and $f = \frac{\nu}{\mu_1}$, it holds that

$$D_{\mathrm{KL}}\left(\nu P^\downarrow \parallel \mu_0\right) = \mathrm{Ent}_{\mu_1}\left(P^\uparrow f\right) \text{ and } D_{\chi^2}\left(\nu P^\downarrow \parallel \mu_0\right) = \mathrm{Var}_{\mu_1}\left(P^\uparrow f\right).$$

It is straightforward to verify $P^\vee = P^\downarrow P^\uparrow$ and $P^\wedge = P^\uparrow P^\downarrow$ are self-adjoint, i.e. $\langle f, P^\vee g \rangle_{\mu_1} = \langle P^\vee f, g \rangle_{\mu_1}$ and $\langle f, P^\wedge g \rangle_{\mu_0} = \langle P^\wedge f, g \rangle_{\mu_0}$. Hence, $P^\vee$ and $P^\wedge$ are reversible with respect to μ_1 and μ_0 respectively.

2.3.1. *Glauber dynamics.* Given a distribution μ with support Q^V, let $\Omega_1 = Q^V$ and $\Omega_0 = \{\sigma \in Q^{V\setminus\{v\}} \mid v \in V\}$. and the current state $X \in \Omega$, the transition $X \to X'$ of Glauber dynamics can be interpreted as the following two steps

- down walk $P^{\downarrow}_{\text{Glauber}}$: pick $v \in V$ uniformly at random and transform $X \in \Omega_1$ to $X_{V\setminus v} \in \Omega_0$;
- up walk $P^{\uparrow}_{\text{Glauber}}$: sample $c \sim \mu_v^{X_{V\setminus\{v\}}}$ and transform $X_{V\setminus v} \in \Omega_0$ to $X' \in \Omega_1$ such that $X'_v = c$ and $X'_{V\setminus\{v\}} = X_{V\setminus\{v\}}$.

Let $\mu_0 := \mu P^{\downarrow}_{\text{Glauber}}$ be a distribution over Ω_0. Then $P^{\downarrow}_{\text{Glauber}}$ and $P^{\uparrow}_{\text{Glauber}}$ is a pair of adjoint operators with respect to distributions $\mu_1 = \mu$ and μ_0. Thus, Glauber dynamics is a down-up walk and is reversible with respect to μ.

2.3.2. *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

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

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

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

The following adjoint result 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 provide a proof of the following proposition in Appendix B.

Proposition 2.3. *For any functions $f : \{0,1\}^V \to \mathbb{R}$ and $g : 2^E \to \mathbb{R}$, it holds that*

$$\langle f, P_{\mathcal{I}\to\mathcal{R}} g\rangle_{\pi_{\text{Ising}}} = \langle P_{\mathcal{R}\to\mathcal{I}} f, g\rangle_{\pi_{\text{wrc}}}. \tag{14}$$

By Proposition 2.3, it holds that $\pi_{\text{Ising}} P_{\mathcal{I}\to\mathcal{R}} = \pi_{\text{wrc}}$ and $\pi_{\text{wrc}} P_{\mathcal{R}\to\mathcal{I}} = \pi_{\text{Ising}}$. Both $P^{\text{Ising}}_{\text{SW}}$ and $P^{\text{wrc}}_{\text{SW}}$ are down-up walks, and their stationary distributions are π_{Ising} and π_{wrc} respectively.

Finally, the mixing times of $P^{\text{Ising}}_{\text{SW}}$ and $P^{\text{wrc}}_{\text{SW}}$ have the following relationships:

$$T_{\text{mix}}\left(P^{\text{Ising}}_{\text{SW}}, \varepsilon\right) \le T_{\text{mix}}\left(P^{\text{wrc}}_{\text{SW}}, \varepsilon\right) + 1 \quad \text{and} \quad T_{\text{mix}}\left(P^{\text{wrc}}_{\text{SW}}, \varepsilon\right) \le T_{\text{mix}}\left(P^{\text{Ising}}_{\text{SW}}, \varepsilon\right) + 1. \tag{15}$$

We prove the first one, the second one holds similarly. Let $T = T_{\mathrm{mix}}\left(P_{\mathrm{SW}}^{\mathrm{wrc}}, \varepsilon\right)$. For any distribution ν over $\{0,1\}^V$, we have

$$\begin{aligned} d_{\mathrm{TV}}\left(\nu(P_{\mathrm{SW}}^{\mathrm{Ising}})^{T+1}, \pi_{\mathrm{Ising}}\right) &= d_{\mathrm{TV}}\left((\nu P_{\mathcal{I}\to\mathcal{R}})(P_{\mathrm{SW}}^{\mathrm{wrc}})^T P_{\mathcal{R}\to\mathcal{I}}, \pi_{\mathrm{wrc}} P_{\mathcal{R}\to\mathcal{I}}\right) \\ \text{(by data processing inequality)} \quad &\le d_{\mathrm{TV}}\left((\nu P_{\mathcal{I}\to\mathcal{R}})(P_{\mathrm{SW}}^{\mathrm{wrc}})^T, \pi_{\mathrm{wrc}}\right) \le \varepsilon. \end{aligned}$$

2.4. **Canonical paths and variance decay.** Let P denote a random walk over Ω that is reversible with respect to μ. It is well-known that P has real eigenvalues $1 = \lambda_1 \ge \lambda_2 \ge \ldots \ge \lambda_{|\Omega|}$. The *spectral gap* is defined by

$$\mathfrak{Gap}(P) = 1 - \lambda_2.$$

Define the Dirichlet form of P by for any functions $f, g : \Omega \to \mathbb{R}$,

$$\mathcal{E}_P(f,g) = \langle f, (I-P)g\rangle_\mu = \frac{1}{2}\sum_{x,y\in\Omega} \mu(x)P(x,y)(f(x)-f(y))(g(x)-g(y)).$$

We can also characterise the spectral gap $\mathfrak{Gap}(P)$ in a variational form:

$$\mathfrak{Gap}(P) = \inf\left\{\frac{\mathcal{E}_P(f,f)}{\mathrm{Var}_\mu(f)} \mid f:\Omega\to\mathbb{R} \wedge \mathrm{Var}_\mu(f) \neq 0\right\}. \tag{16}$$

A useful tool to analyse the spectral gap of a reversible Markov chain is the canonical path introduced by Jerrum and Sinclair [JS89]. Let P be a reversible Markov chain over the state space Ω with stationary distribution π. Let $\gamma_{XY} = (Z_0 = X, Z_1, Z_2, \ldots, Z_\ell = Y)$ be a path of length ℓ moving in the state space using transitions of P, i.e. for any $i \in [\ell]$, $P(Z_{i-1}, Z_i) > 0$. For each pair of $X, Y \in \Omega$, its path γ_{XY} is assigned a weight $w(\gamma_{XY}) = \mu(X)\mu(Y)$. Let Γ be the collection of all canonical paths. The congestion of Γ is defined by

$$\varrho(\Gamma) := \max_{(Z,Z')\in\Omega^2, P(Z,Z')>0} \frac{L}{\mu(Z)P(Z,Z')} \sum_{\gamma\in\Gamma:(Z,Z')\in\gamma} w(\gamma) \tag{17}$$

where L is the maximum length of path in Γ. Sinclair [Sin92] showed that the congestion of any collection of paths Γ for a Markov chain P is an upper bound of the inverse of its spectral gap, namely,

$$\frac{1}{\mathfrak{Gap}(P)} \le \varrho(\Gamma).$$

Consider the down-up walk $P^\vee = P^\downarrow P^\uparrow$ over Ω_1, where $P^\downarrow : \Omega_1 \times \Omega_0 \to \mathbb{R}_{\ge 0}$ and $P^\uparrow : \Omega_0 \times \Omega_1 \to \mathbb{R}_{\ge 0}$ are a pair of adjoint operators with respect to distribution μ_0 over Ω_0 and μ_1 over Ω_1. For simplicity, we denote Ω_1 by Ω, and we denote μ_1 by μ. The following result holds for $P^\vee$.

Proposition 2.4. *Let $P^\vee = P^\downarrow P^\uparrow$ be a down-up walk over Ω that is reversible with respect to μ. For any $0 < \delta < 1$, the spectral gap $\mathfrak{Gap}(P^\vee) \ge \delta$ if and only if for any distribution ν over Ω,*

$$D_{\chi^2}\left(\nu P^\downarrow \parallel \mu P^\downarrow\right) \le (1-\delta)\, D_{\chi^2}(\nu \parallel \mu). \tag{18}$$

Proof. Let $f = \frac{\nu}{\mu}$. It holds that

$$\mathcal{E}_{P^\vee}(f,f) = \langle f, f\rangle_\mu - \langle f, P^\vee f\rangle_\mu = \langle f, f\rangle_\mu - \langle P^\uparrow f, P^\uparrow f\rangle_{\mu_0} = \mathrm{Var}_\mu(f) - \mathrm{Var}_{\mu_0}\left(P^\uparrow f\right).$$

Then the lemma follows from $D_{\chi^2}\left(\nu P^\downarrow \parallel \mu P^\downarrow\right) = \mathrm{Var}_{\mu_0}\left(P^\uparrow f\right)$, $D_{\chi^2}(\nu \parallel \mu) = \mathrm{Var}_\mu(f)$, and (16). □

2.5. **Spectral independence and entropy decay.** Let Q be a finite set. Let μ be a distribution with support Q^V. Fix a partial pinning $\tau \in Q^\Lambda$ for some $\Lambda \subseteq V$. Define the *absolute influence matrix* Ψ_μ^τ by

$$\forall u, v \in V \setminus \Lambda \text{ with } u \neq v, \quad \Psi_\mu^\tau(u,v) := \max_{i,j \in Q} d_{\mathrm{TV}}\left(\mu_v^{\tau \wedge (u \leftarrow i)}, \mu_v^{\tau \wedge (u \leftarrow j)}\right)$$
$$\forall v \in V \setminus \Lambda, \quad \Psi_\mu^\tau(v,v) := 0.$$

where $d_{\mathrm{TV}}(\cdot,\cdot)$ denotes the total variation distance and $\mu_v^{\tau \wedge (u \leftarrow i)}$ denotes the marginal distribution on v conditional on that variables in Λ take the value τ and u takes the value i. We say that the distribution μ is ℓ_∞-*spectrally independent* with parameter ζ if

$$\forall \Lambda \subset V, \sigma \in Q^\Lambda, \quad \left\|\Psi_\mu^\sigma\right\|_\infty = \max_{u \notin \Lambda} \sum_{v \notin \Lambda} \Psi_\mu^\sigma(u,v) \leq \zeta.$$

Call μ *b-marginally bounded* if

$$\min_{\Lambda \subseteq V, v \notin \Lambda} \min_{\sigma \in Q^\Lambda, c \in Q} \mu_v^\sigma(c) \geq b.$$

In this paper, we are particularly interested in *Gibbs distributions*. We will consider a slightly more general than usual version defined over hypergraphs. Let $H = (V, \mathcal{E})$ be a hypergraph. Given weight functions $(\phi_v)_{v \in V}$ and $(\phi_e)_{e \in \mathcal{E}}$, where $\phi_v : Q \to \mathbb{R}_{>0}$ and $\phi_e : Q^e \to \mathbb{R}_{>0}$, define the Gibbs distribution μ over Q^V by

$$\forall \sigma \in Q^V, \quad \mu(\sigma) \propto \prod_{v \in V} \phi_v(\sigma_v) \prod_{e \in \mathcal{E}} \phi_e(\sigma_e).$$

Let $G_\mu = (V, E)$ be a graph such that $\{u, v\} \in E$ if $u \in e'$ and $v \in e'$ for some $e' \in \mathcal{E}$. For any disjoint $A, B, C \subseteq V$, if the removal of C disconnects A and B in G_μ, it holds that variables in A and B are independent in μ conditional on any assignment on C. Define *maximum degree* D_μ of the Gibbs distribution μ as the maximum degree of the graph G_μ.

The spectral independence is related to the mixing time of Glauber dynamics. The following result is proved in [CLV21a, BCC+22] (see also [CLV21b, Theorem 13])

Theorem 2.5 ([CLV21a, BCC+22]). *Let $\zeta, b, D > 0$. For any Gibbs distribution μ over Q^V, where $|V| = n$, if μ is ℓ_∞-*spectrally independent *with parameter ζ, b-marginally bounded and has the maximum degree at most D, then the down walk of the Glauber dynamics satisfies that*

$$\forall \textit{distribution } \nu \textit{ over } Q^V, \quad D_{\mathrm{KL}}\left(\nu P^\downarrow_{\text{Glauber}} \parallel \mu P^\downarrow_{\text{Glauber}}\right) \leq \left(1 - \frac{1}{Cn}\right) D_{\mathrm{KL}}\left(\nu \parallel \mu\right),$$

where $C = \left(\frac{D}{b}\right)^{1+2\left\lceil \frac{\zeta}{b} \right\rceil} > 1$ is a constant depending only on ζ, b and D.

In [CLV21a, BCC+22], they mainly establish the so-called "approximate tensorization of entropy" property for μ. However this is equivalent to the contraction of relative entropy by $P^\downarrow_{\text{Glauber}}$ [CLV21a].

2.6. **Holographic transformation.** We will need holographic transformations [Val08] 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) $(f_0, \cdots, f_x, \cdots, f_{2^d-1})$ 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 (bipartite) *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 $\sigma\,|_{E(v)}$ stands for the restriction of the assignment σ to the incident edges of v.[3]

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 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 2.6. Holant$(H; \mathcal{F} \mid \mathcal{G})$ = Holant$(H; \mathcal{F}\boldsymbol{T} \mid \boldsymbol{T}^{-1}\mathcal{G})$ *for any invertible* $\boldsymbol{T} \in \mathbb{C}^{2\times 2}$.

3. The grand model and a generalised Grimmett–Janson coupling

We introduce a grand model, inspired by [GJ07b], that unifies the subgraph and random cluster models introduced in Section 2.1. We also generalise the coupling of Grimmett and Janson [GJ07b] for ferromagnetic Ising models with external fields. It is possible to also include vertex configurations in this grand model à la Edwards and Sokal [ES88], so that the Ising model is also unified under this framework. However it does not appear to have much benefit and we choose not to do so.

3.1. **The grand model.** Let $G = (V, E)$ be a simple undirected graph. The *grand model*, specified by parameters $\boldsymbol{p} = (p_e)_{e\in E}$ and $\boldsymbol{\eta} = (\eta_v)_{v\in V}$ where $0 \le p_e \le 1/2$ and $0 \le \eta_v \le 1$, defines a distribution π_{gm} over all configurations on the edges of three states $X : E \to \{0, 1, 2\}$. Given an assignment X in the grand model, denote by $X^{-1}(q)$ the set of edges that are assigned q under X where $q = 0, 1, 2$. The weight of each configuration is given by

$$\mathrm{wt}_{\mathrm{gm}}(X) = \prod_{e\in X^{-1}(\{1,2\})} p_e \prod_{f\in X^{-1}(0)} (1 - 2p_f) \prod_{v\in O(X)} \eta_v, \tag{19}$$

where $O(X)$ is the set of vertices of odd degree in the subgraph $(V, X^{-1}(1))$. The probability of each configuration X is

$$\pi_{\mathrm{gm}}(X) = \frac{\mathrm{wt}_{\mathrm{gm}}(X)}{Z_{\mathrm{gm}}} \tag{20}$$

where

$$Z_{\mathrm{gm}} = Z_{\mathrm{gm}}(G; \boldsymbol{p}, \boldsymbol{\eta}) := \sum_{X\in\Omega_{\mathrm{gm}}(G)} \mathrm{wt}_{\mathrm{gm}}(X)$$

is the partition function of the grand model.

Equivalently, a random sample from the grand model can be generated by the following procedure.

- **Step-I**: Sample $S \sim \pi_{\mathrm{sg}}$, where π_{sg} is the distribution specified by the subgraph-world model with parameters $(\boldsymbol{p}, \boldsymbol{\eta})$; for each $e \in E$, let $X(e) = 1$ if $e \in S$ and let $X(e) = *$ if $e \notin S$.
- **Step-II**: Independently for each $e \in E$ with $X_e = *$, set $X(e) = 2$ with probability $\frac{p_e}{1-p_e}$, and $X(e) = 0$ otherwise.

It is straightforward to verify that the outcome distribution is exactly the grand model distribution.

Recall the definition of a Gibbs distribution and its maximum degree in Section 2.5. The grand model is indeed a Gibbs distribution in the sense of Theorem 2.5. Each edge of G corresponds to a variable, and each vertex $v \in V$ corresponds to a weight function. In other words, this is a Holant-type problem [CLX11]. Theorem 2.5 applies to Holant-type problems, as explained in [CLV21b, Section 2.2]. The underlying graph of π_{gm} (as defined in Section 2.5) is the line graph of G, whose maximum degree is at most $2\Delta - 1$. Thus we have the following observation.

Observation 3.1. *The distribution* π_{gm} *is a Gibbs distribution with maximum degree* $D \le 2\Delta - 1$*, where* Δ *is the maximum degree of the graph* $G = (V, E)$.

The next lemma gives the relation among the grand model, the subgraph-world model and the random cluster model.

Lemma 3.2. *Let* $X \sim \pi_{\mathrm{gm}}$ *be a random sample from the grand model with parameter* $\boldsymbol{p} = (p_e)_{e\in E}$ *and* $\boldsymbol{\eta} = (\eta_v)_{v\in V}$*, where* $0 \le p_e \le 1/2$ *and* $0 \le \eta_v \le 1$*. It holds that*

[3] Holant problems can also be defined for not necessarily bipartite graphs, but we do not need those here.

- $\mathcal{S} = \{e \in E \mid X(e) = 1\}$ *follows the distribution specified by the subgraph-world model with parameters* $(\boldsymbol{p}, \boldsymbol{\eta})$*;*
- $\mathcal{R} = \{e \in E \mid X(e) = 1 \vee X(e) = 2\}$ *follows the distribution specified by the random cluster model with parameters* $(2\boldsymbol{p}, \boldsymbol{\lambda})$*, where* $\lambda_v = \frac{1-\eta_v}{1+\eta_v}$ *for all* $v \in V$*.*

Namely, $X(e) = 1$ means e is present in the subgraph-world model (**Step-I**), and $X(e) = 2$ means e is absent in the subgraph-world model, but gets added into the random cluster model in **Step-II**. $X(e) = 0$ means e is absent in both models.

The first part of Lemma 3.2 holds trivially. The second part is proved by a generalised Grimmett–Janson coupling [GJ07b]. The proof of the second part is given in Section 3.2.

3.2. **Coupling via holographic transformation.** Under the unweighted setting, Grimmett and Janson [GJ07b, Theorem 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* $0 \le p_e \le 1/2$ *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 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 remark that the second part of Lemma 3.2 is a straightforward consequence of Lemma 3.3. 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*

$$\prod_{C\in\kappa(V,E)} \left(1 + \prod_{u\in C} \lambda_u\right) = \left(\prod_{v\in V}(1+\lambda_v)\right)\left(\frac{1}{2}\right)^{|E|} \sum_{E'\subset E} \prod_{u\in odd(E')} \eta_u, \tag{21}$$

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 $f_v^{(2)} = [1, \eta_v, 1, \eta_v, \ldots]$. 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 2.6, it holds that

$$\operatorname{Holant}\left(H; \mathcal{F}^{(1)} \mid \mathcal{G}\right) = \left(\prod_{v\in V}(1+\lambda_v)\right)\left(\frac{1}{2}\right)^{|E|} \operatorname{Holant}\left(H; \mathcal{F}^{(2)} \mid \mathcal{G}\right). \tag{22}$$

This equation is indeed (21) in disguise. The equivalence between the left-hand sides of (22) and (21) 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 (22) and (21) 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 $G = (V, E)$,

$$\begin{aligned}
\mathbf{Pr}[\mathcal{R} = R] &= \frac{1}{Z_{\mathrm{sg}}(G; \boldsymbol{p}, \boldsymbol{\eta})} \sum_{S \subseteq R} \prod_{u \in \mathrm{odd}(S)} \eta_u \prod_{e \in S} p_e \prod_{f \in E \setminus S} (1 - p_f) \prod_{g \in R \setminus S} \frac{p_g}{1 - p_g} \prod_{h \in E \setminus R} \frac{1 - 2p_h}{1 - p_h} \\
&= \frac{1}{Z_{\mathrm{sg}}(G; \boldsymbol{p}, \boldsymbol{\eta})} \sum_{S \subseteq R} \prod_{u \in \mathrm{odd}(S)} \eta_u \prod_{e \in R} p_e \prod_{f \in E \setminus R} (1 - 2p_f) \\
&= \frac{1}{Z_{\mathrm{sg}}(G; \boldsymbol{p}, \boldsymbol{\eta})} 2^{-|R|} \prod_{e \in R} (2p_e) \prod_{f \in E \setminus R} (1 - 2p_f) \sum_{S \subseteq R} \prod_{u \in \mathrm{odd}(S)} \eta_u \\
\text{(By (21) on } (V, R)) \quad &= \frac{1}{Z_{\mathrm{sg}}(G; \boldsymbol{p}, \boldsymbol{\eta})} \prod_{e \in R} (2p_e) \prod_{f \in E \setminus R} (1 - 2p_f) \prod_{v \in V} \frac{1}{1 + \lambda_v} \prod_{C \in \kappa(V,R)} \left(1 + \prod_{u \in C} \lambda_u\right) \\
\text{(By (6))} \quad &= \frac{1}{Z_{\mathrm{wrc}}(G; 2\boldsymbol{p}, \boldsymbol{\lambda})} \prod_{e \in R} (2p_e) \prod_{f \in E \setminus R} (1 - 2p_f) \prod_{C \in \kappa(V,R)} \left(1 + \prod_{u \in C} \lambda_u\right). \\
&= \pi_{\mathrm{wrc}}(R).
\end{aligned}$$

□

4. Variance decay of Glauber dynamics on the grand model

Let $G = (V, E)$ be a graph. Let $\boldsymbol{p} = (p_e)_{e \in E}$ and $\boldsymbol{\eta} = (\eta_v)_{v \in V}$, where $0 < p_e < 1/2$ and $0 < \eta_v < 1$. Let π_{gm} denote the distribution specified by the grand model with parameters $\boldsymbol{p}$ and $\boldsymbol{\eta}$. Let $\Omega(\pi_{\mathrm{gm}})$ denote the support of π_{gm}. We use $P_{\mathrm{GlauberGM}}$ to denote Glauber dynamics on π_{gm} as defined in Section 2.3.1.

Lemma 4.1. *The Glauber dynamics $P_{\mathrm{GlauberGM}}$ satisfies that for any distribution ν with support $\Omega(\nu) \subseteq \Omega(\pi_{\mathrm{gm}})$,*

$$D_{\chi^2}\left(\nu P^{\downarrow}_{\mathrm{GlauberGM}} \parallel \pi_{\mathrm{gm}} P^{\downarrow}_{\mathrm{GlauberGM}}\right) \leq \left(1 - \frac{\eta^4_{\min} \min\{p_{\min}, 1 - 2p_{\max}\}}{m^2}\right) D_{\chi^2}\left(\nu \parallel \pi_{\mathrm{gm}}\right),$$

where $\eta_{\min} = \min_{v \in V} \eta_v$ and $m = |E|$.

By Proposition 2.4, we only need to bound the spectral gap of the Glauber dynamics. The rest of this section endeavours to show

$$\mathfrak{Gap}\left(P_{\mathrm{GlauberGM}}\right) \geq \frac{\eta^4_{\min}}{m^2} \min\{p_{\min}, 1 - 2p_{\max}\}. \tag{23}$$

This will be proved using the canonical path method adapted from [JS93].

4.1. Construction of the canonical path.

Below is the main lemma of this subsection.

Lemma 4.2. *For any grand model on a 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 $\Gamma = \{\gamma_{XY} : X, Y \in \Omega\}$ for the Glauber dynamics P_{gm} such that*

(1) $w_{\mathrm{gm}}(X, Y) = \pi_{\mathrm{gm}}(X)\pi_{\mathrm{gm}}(Y)$;
(2) $|\gamma_{XY}| \leq m$;
(3) for any transition (Z, Z') with $|\{e : Z(e) \neq Z'(e)\}| = 1$, where the only edge e of discrepancy is assigned 1 in either Z or Z', it holds that

$$\sum_{\gamma \in \Gamma : (Z, Z') \in \gamma} w_{\mathrm{gm}}(\gamma) \leq \eta^{-4}_{\min} \min\left\{\pi_{\mathrm{gm}}(Z), \pi_{\mathrm{gm}}(Z')\right\} \tag{24}$$

where $\eta_{\min} := \min_v \eta_v$;
(4) for any transition (Z, Z') with $|\{e : Z(e) \neq Z'(e)\}| = 1$, where the only edge e of discrepancy is assigned 1 in neither Z nor Z', it holds that

$$\sum_{\gamma \in \Gamma : (Z, Z') \in \gamma} w_{\mathrm{gm}}(\gamma) \leq \min\left\{\pi_{\mathrm{gm}}(Z), \pi_{\mathrm{gm}}(Z')\right\}. \tag{25}$$

Proof. We begin the proof with the construction of the paths. Suppose all vertices and edges are indexed by distinct integers, and there is a fixed ordering $\prec$ for all paths and cycles of the graph G. For any pair of assignments X, Y in the grand model, the canonical path γ_{XY} contains two stages, moving from X to W and W to Y respectively.

Stage 1. (1-*edge mending*.) Midst this stage we mend the edges assigned 1 in either X or Y but not the other. Denote the set of such edges $D := X^{-1}(1) \oplus Y^{-1}(1)$. The resulting configuration W has the property that (1) for any edge $e \in D$, it holds that $W(e) = Y(e)$, and (2) for any other edge $e \notin D$, it holds that $W(e) = X(e)$.

Let $2k$ be the number of the odd-degree vertices in D. Then, D 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 the lexicographic order induced by $\prec$.

To move from X to W, we process each of the paths and cycles one by one. For each of them, we first choose the vertex and edge to start with. When winding (handling) a path, the starting vertex is one of the two open vertices of the path that has a smaller index; when winding a cycle, the starting vertex is the one with the smallest index, and the next vertex (which together with the starting one defines a starting edge) is one of the two neighbours of the starting vertex of the cycle that has a smaller index than the other one. After deciding the starting vertex and edge, we just move along the path/cycle. For each of the edge, we set the assignment to it as that in Y. Obviously this gives W satisfying the properties aforementioned because every edge in D is mended while the rest are left untouched.

Stage 2. (0, 2-*edge mending*.) None of the conflicting edges between W and Y can be assigned 1 in either of them. In this stage, we simply change all remaining disagreeing edges from the value in W to the value in Y one by one according to the order of their indices.

We then show that the set of canonical paths Γ constructed above fulfills Lemma 4.2. Assign weight $w_{\mathrm{gm}}(\gamma) = \pi_{\mathrm{gm}}(X)\pi_{\mathrm{gm}}(Y)$ to the path γ_{XY}. The length (number of transitions) of each path γ_{XY} is at most m, because each edge is mended at most once.

We first prove (24). Let (Z, Z') be a transition with $|\{e : Z(e) \neq Z'(e)\}| = 1$, where the only edge e of discrepancy is assigned 1 in either Z or Z'. Note that (Z, Z') will only be used by any path in its first stage described above. Define a mapping $\varphi_{Z,Z'} : \Omega \times \Omega \to \Omega$ over any pair of configurations X, Y whose corresponding path γ_{XY} uses the transition (Z, Z') by

$$\varphi_{Z,Z'}(X, Y) = U \quad \text{where} \quad U(e) = X(e) + Y(e) - Z(e), \forall e \in E(G). \tag{26}$$

We claim that $\varphi_{Z,Z'}$ is an injection. Given U and Z, we can recover $X(e) + Y(e)$ for any edge e. First we can find D, the set of conflicting 1-edge in Stage 1, as it is simply $\{e : X(e) + Y(e) = 1 \text{ or } 3\}$. This gives rise to the unique edge-disjoint decomposition $P_1, \cdots, P_k, C_1, \cdots, C_{k'}$. By looking at Z and Z', we know the edge that is currently being wound, and, together with the edge-disjoint decomposition, the stage of the whole winding process. Therefore, we can continue the winding from Z' with these information, and when finished, W (defined in the process Stage 1) is obtained. To further recover Y, note that e gets mended in Stage 2 if any only if $U(e) + Z(e) = 2$ and $Z(e) \neq 1$. This follows from the fact that $Z(e)$ (in the first stage) is in line with $X(e)$ so long as $Z(e) \neq 1$. Therefore, we can decide all such edges and mend the assignment to obtain Y. To get X, we just reverse the operations backwards from Z.

Given this injection, we compute $\sum_{\gamma \in \Gamma : (Z,Z') \in \gamma} w_{\mathrm{gm}}(\gamma)$. The goal here is to bound the following ratio

$$\frac{\pi_{\mathrm{gm}}(X)\pi_{\mathrm{gm}}(Y)}{\pi_{\mathrm{gm}}(U)\pi_{\mathrm{gm}}(Z)}, \quad \text{or equivalently,} \quad \frac{\mathrm{wt}_{\mathrm{gm}}(X)\mathrm{wt}_{\mathrm{gm}}(Y)}{\mathrm{wt}_{\mathrm{gm}}(U)\mathrm{wt}_{\mathrm{gm}}(Z)}. \tag{27}$$

Recall that this ratio may contain two kinds of factors, emerging from both the vertices and edges. For the factor from edges, the construction of U ensures that (1) if $X(e) + Y(e) = U(e) + Z(e) \in \{0, 1, 3, 4\}$, or $X(e) + Y(e) = 2$ and $X(e) \neq 1$, then it must holds that either $X(e) = U(e)$ and $Y(e) = Z(e)$, or $X(e) = Z(e)$ and $Y(e) = U(e)$; (2) if $X(e) = Y(e) = 1$, then e never gets mended throughout the canonical path, and hence $Z(e) = U(e) = 1$. In either case, all the terms rising from the edges in the numerator and denominator cancel. The terms rising from the vertices come from those in $O(X), O(Y), O(U), O(Z)$. It is not hard to see that the ones that do not get cancelled only arise 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,

$$\frac{\pi_{\mathrm{gm}}(X)\pi_{\mathrm{gm}}(Y)}{\pi_{\mathrm{gm}}(U)\pi_{\mathrm{gm}}(Z)} \le \eta_{\min}^{-4}, \tag{28}$$

as $0 < \eta_v < 1$ for all v.

Then, (24) follows from (28) that

$$\begin{aligned}
\text{(By definition)} \quad & \sum_{\gamma\in\Gamma:(Z,Z')\in\gamma} w_{\mathrm{gm}}(\gamma) = \sum_{X,Y:(Z,Z')\in\gamma_{XY}} \pi_{\mathrm{gm}}(X)\pi_{\mathrm{gm}}(Y) \\
\text{(By (28))} \quad & \le \eta_{\min}^{-4} \sum_{X,Y:(Z,Z')\in\gamma_{XY}} \pi_{\mathrm{gm}}(Z)\pi_{\mathrm{gm}}(\varphi_{Z,Z'}(X,Y)) \\
(\varphi_{Z,Z'} \text{ is injective}) \quad & \le \eta_{\min}^{-4}\pi_{\mathrm{gm}}(Z).
\end{aligned}$$

We construct the other mapping $\varphi'_{Z,Z'}(X,Y)$ by taking $\varphi_{Z,Z'}(X,Y)(e) = X(e) + Y(e) - Z'(e)$. The same argument shows that $\sum_{\gamma\in\Gamma:(Z,Z')\in\gamma} w_{\mathrm{gm}}(\gamma) \le \eta_{\min}^{-4}\pi_{\mathrm{gm}}(Z')$.

To prove (25), we look at the transition step (Z,Z') with $|\{e : Z(e) \neq Z'(e)\}| = 1$ where the only edge e of discrepancy is assigned 1 in neither Z nor Z'. We use the same mapping $\varphi_{Z,Z'}(X,Y)$ as above, and claim it is still injective in this case. Recall that e gets mended in Stage 2 if and only if $U(e) + Z(e) = 2$ and $Z(e) \neq 1$, and we can again determine the edges to be mended in Stage 2. Moreover, by looking at the difference of Z and Z', we know the index of the edge being mended, and therefore we can continue this process manually according to the instruction of Stage 2, knowing which edges to mend, to obtain Y. To get X, we first go backwards from Z to the beginning of Stage 2 to obtain W, and revert the whole Stage 1 using the same argument aforementioned.

To show (25), note that the edge factors in the ratio of (27) again cancel, and because no edge with assignment 1 is involved, the vertex factors cancel as well. Hence the ratio is exactly 1, and (25) follows according to the same calculation. □

4.2. **Total congestion and rapid mixing.** We next bound the total congestion for Γ_{gm}. For each transition (Z,Z') such that $|\{e : Z(e) \neq Z'(e)\}| = 1$, where the only edge of discrepancy is assigned 1 in either Z or Z', we have

$$\frac{L}{\pi_{\mathrm{gm}}(Z)P_{\mathrm{gm}}(Z,Z')} \sum_{\substack{\gamma\in\Gamma:\\ (Z,Z')\in\gamma}} w_{\mathrm{gm}}(\gamma) \le \frac{m\eta_{\min}^{-4}\min\{\pi_{\mathrm{gm}}(Z),\pi_{\mathrm{gm}}(Z')\}}{\pi_{\mathrm{gm}}(Z)P_{\mathrm{gm}}(Z,Z')} =: (\spadesuit)$$

by Lemma 4.2. To continue the calculation, there are several cases $(Z(e),Z'(e)) = (0,1),(2,1),(1,0),(1,2)$. Below we only prove the case $(Z(e),Z'(e)) = (0,1)$. The rest cases can be argued the same way and yield the same bound. Let $e = (u,v)$. There are some more subcases, depending on if u or v is in $O(Z)$.

- $u,v \notin O(Z)$. In this case, setting the edge to 1 leads to extra factors from both vertices in Z'. Cancelling all the edges and vertices not involved, we obtain

$$(\spadesuit) = \frac{m^2\eta_{\min}^{-4}\min\{1-2p_e, p_e\eta_u\eta_v\}}{(1-2p_e)\frac{p_e\eta_u\eta_v}{(1-2p_e)+(p_e\eta_u\eta_v)+p_e}} \le \frac{m^2\eta_{\min}^{-4}\min\{1-2p_e, p_e\eta_u\eta_v\}}{(1-2p_e)(p_e\eta_u\eta_v)} \le \frac{m^2\eta_{\min}^{-4}}{1-2p_e}$$

  where we use the fact that $\eta_u,\eta_v \le 1$.
- $u,v \in O(Z)$. In this case, setting the edge to 1 removes the factors from both vertices in Z'. Cancelling all the edges and vertices not involved, we obtain

$$(\spadesuit) = \frac{m^2\eta_{\min}^{-4}\min\{(1-2p_e)\eta_u\eta_v, p_e\}}{(1-2p_e)\eta_u\eta_v\frac{p_e}{(1-2p_e)\eta_u\eta_v+p_e+p_e\eta_u\eta_v}} \le \frac{m^2\eta_{\min}^{-4}\min\{(1-2p_e)\eta_u\eta_v, p_e\}}{(1-2p_e)\eta_u\eta_v p_e} \le \frac{m^2\eta_{\min}^{-4}}{p_e}$$

  where we use the fact that $\eta_u,\eta_v \le 1$ again.

- WLOG suppose $u \in O(Z), v \notin O(Z)$. In this case, setting the edge to 1 causes the vertex factor to switch. Cancelling all the edges and vertices not involved, we obtain

$$(\spadesuit) = \frac{m^2\eta_{\min}^{-4}\min\{(1-2p_e)\eta_u, p_e\eta_v\}}{(1-2p_e)\eta_u \frac{p_e\eta_v}{(1-2p_e)\eta_u+(p_e\eta_v)+(p_e\eta_u)}}.$$

If $\eta_u < \eta_v$, then above becomes

$$\frac{m^2\eta_{\min}^{-4}\min\{(1-2p_e)\frac{\eta_u}{\eta_v}, p_e\}}{(1-2p_e)\frac{\eta_u}{\eta_v}\frac{p_e}{(1-2p_e)\frac{\eta_u}{\eta_v}+p_e+p_e\frac{\eta_u}{\eta_v}}} \le \frac{m^2\eta_{\min}^{-4}}{p_e}.$$

Otherwise, it can be written as

$$\frac{m^2\eta_{\min}^{-4}\min\{(1-2p_e), p_e\frac{\eta_v}{\eta_u}\}}{(1-2p_e)\frac{p_e\frac{\eta_v}{\eta_u}}{(1-2p_e)+p_e\frac{\eta_v}{\eta_u}+p_e}} \le \frac{m^2\eta_{\min}^{-4}}{1-2p_e}.$$

For each transition (Z, Z') such that $|\{e : Z(e) \neq Z'(e)\}| = 1$, where the only edge of discrepancy is assigned 1 in none of Z or Z', the calculation is similar as above but simpler. WLOG assume $Z(e) = 0$ and $Z'(e) = 2$.

(Lemma 4.2)
$$\frac{L}{\pi_{\mathrm{gm}}(Z)P_{\mathrm{gm}}(Z,Z')}\sum_{\substack{\gamma\in\Gamma:\\(Z,Z')\in\gamma}} w_{\mathrm{gm}}(\gamma) \le \frac{m\min\{\pi_{\mathrm{gm}}(Z),\pi_{\mathrm{gm}}(Z')\}}{\pi_{\mathrm{gm}}(Z)P_{\mathrm{gm}}(Z,Z')}$$

(Worst case of η terms)
$$\le \frac{m^2\min\{1-2p_e, p_e\}}{(1-2p_e)\frac{p_e}{1-2p_e+p_e+p_e\frac{1}{\eta_u\eta_v}}} \le \min\left\{\frac{1}{p_e}, \frac{1}{1-2p_e}\right\} m^2\eta_{\min}^{-2}.$$

There is no canonical path using the self loop (Z, Z), so the congestion is zero. In all, the congestion is bounded by $m^2\eta_{\min}^{-4}\max\left\{\frac{1}{p_{\min}}, \frac{1}{1-2p_{\max}}\right\}$, from which (23) follows.

5. Entropy decay of Glauber dynamics on the grand model

In Section 4, we analysed the variance decay of Glauber dynamics on the grand model. We now continue to analyse its relative entropy decay. Let $G = (V, E)$ be a graph, and $\boldsymbol{p} = (p_e)_{e\in E}$ and $\boldsymbol{\eta} = (\eta_v)_{v\in V}$ be the parameters, where $0 < p_e < 1/2$ for any $e \in E$ and $\eta_v > 0$ for any $v \in V$. Let π_{gm} denote the distribution specified by the grand model with parameters $\boldsymbol{p}$ and $\boldsymbol{\eta}$. Let $\Omega(\pi_{\mathrm{gm}})$ denote the support of π_{gm}. We use $P_{\mathrm{GlauberGM}}$ to denote Glauber dynamics on π_{gm}.

Lemma 5.1. *If $0 < \eta_v < 1$ for all $v \in V$, then for any distribution ν with support $\Omega(\nu) \subseteq \Omega(\pi_{\mathrm{gm}})$, Glauber dynamics $P_{\mathrm{GlauberGM}}$ satisfies*

$$D_{\mathrm{KL}}\left(\nu P^{\downarrow}_{\mathrm{GlauberGM}} \,\|\, \pi_{\mathrm{gm}}P^{\downarrow}_{\mathrm{GlauberGM}}\right) \le \left(1-\frac{1}{Cn}\right) D_{\mathrm{KL}}\left(\nu \,\|\, \pi_{\mathrm{gm}}\right),$$

where $C = C(\Delta, \eta_{\min}, p_{\min}, p_{\max})$, $\eta_{\min} = \min_{v\in V}\eta_v$, $p_{\min} = \min_{e\in E} p_e$, $p_{\max} = \max_{e\in E} p_e$, Δ is the maximum degree of G and $n = |V|$.

Remark 5.2. For interested readers, the constant C in the lemma above can be taken as

$$C = \Delta\left(\frac{2\Delta}{\eta_{\min}^2\min\{1-2p_{\max}, p_{\min}\}}\right)^{2+\frac{16\Delta^2}{\eta_{\min}^4\min\{1-2p_{\max},p_{\min}\}}}.$$

Lemma 5.1 is proved by Theorem 2.5. To apply Theorem 2.5, we need to verify (1) π_{gm} is a Gibbs distribution with maximum degree $D = 2\Delta - 1$; (2) π_{gm} is ℓ_∞-spectrally independent; (3) π_{gm} is marginally bounded. The rest of this section is dedicated to the proof of Lemma 5.1.

Lemma 5.3. *π_{gm} is ℓ_∞-spectrally independent with parameter $\zeta = O(\Delta^2/\eta_{\min}^2)$.*

We need the following result in [CLV21b] to prove Lemma 5.3. We view the subgraph world as a distribution over $\{0,1\}^E$, where each $Y \in \{0,1\}^E$ corresponds to $S = \{e \in E \mid Y_e = 1\}$.

Lemma 5.4 ([CLV21b]). *Let $G = (V, E)$ be a graph with the maximum degree $\Delta \geq 3$. Let $\boldsymbol{p} = (p_e)_{e\in E}$ and $\boldsymbol{\eta} = (\eta_v)_{v\in V}$, where $0 \leq p_e < 1/2$ and $0 < \eta_v \leq 1$. The distribution π_{sg} specified by the subgraph-world model with parameters $(\boldsymbol{p}, \boldsymbol{\eta})$ is ℓ_∞-spectrally independent with parameter $\zeta = O(\Delta^2/\eta_{\min}^2)$.*

Remark 5.5. In [CLV21b], the authors only formalise the proof for the uniform case (i.e., all η_v's take the same value) while stating that the argument works for non-uniform case without a proof. This in fact holds true by going through the proof and taking the worst region of stability. The final spectral independence parameter is

$$\zeta = 8\left(\frac{\left(\frac{1+\eta_{\min}}{1-\eta_{\min}}\right)^{1/\Delta}+1}{\left(\frac{1+\eta_{\min}}{1-\eta_{\min}}\right)^{1/\Delta}-1}\right)^2 \sim 8\Delta^2/\eta_{\min}^2.$$

Note that the λ in their paper is actually $p/(1-p)$ in our formulation of the subgraph-world model (under the uniform edge parameter setting). Also note that we are only considering the region $0 < p < 1/2$, so the λ in their paper is bounded from above by 1.

Proof of Lemma 5.3. Fix a pinning $\sigma \in \{0,1,2\}^\Lambda$ for some $\Lambda \subseteq E$. According to the definition of the grand model, to draw $X \sim \pi_{\mathrm{gm}}$, we first sample $Y \sim \pi_{\mathrm{sg}}$ (where $Y \in \{0,1\}^E$ as we view π_{sg} as a distribution over $\{0,1\}^E$), then flip independent coins for each $e \in E$ with $Y_e = 0$. Define the pinning $\tau \in \{0,1\}^\Lambda$ by $\tau_e = 1$ if $\sigma_e = 1$ and $\tau_e = 0$ if $\sigma_e = 0$ or $\sigma_e = 2$. Consider the influence

$$\Psi^\sigma_{\pi_{\mathrm{gm}}}(e,f) = \max\left\{d_{\mathrm{TV}}\left(\pi^{\sigma\wedge e\leftarrow 0}_{\mathrm{gm},f}, \pi^{\sigma\wedge e\leftarrow 1}_{\mathrm{gm},f}\right), d_{\mathrm{TV}}\left(\pi^{\sigma\wedge e\leftarrow 0}_{\mathrm{gm},f}, \pi^{\sigma\wedge e\leftarrow 2}_{\mathrm{gm},f}\right), d_{\mathrm{TV}}\left(\pi^{\sigma\wedge e\leftarrow 1}_{\mathrm{gm},f}, \pi^{\sigma\wedge e\leftarrow 2}_{\mathrm{gm},f}\right)\right\},$$

where $e, f \in E \setminus \Lambda$ and $e \neq f$. Since each coin flipping is independent with the random sample from π_{gm}, we can couple two distributions $\pi^{\sigma\wedge e\leftarrow 0}_{\mathrm{gm},f}$ and $\pi^{\sigma\wedge e\leftarrow 1}_{\mathrm{gm},f}$ as follows:

- sample Y_f, Y'_f from the optimal coupling between $\pi^{\tau\wedge e\leftarrow 0}_{\mathrm{sg},f}$ and $\pi^{\tau\wedge e\leftarrow 1}_{\mathrm{sg},f}$;
- flip a coin C independently with probability of HEADS being $\frac{p_f}{1-p_f}$;
- if $Y_f = 1$, let $X_f = 1$; otherwise, if the outcome of C is HEADS, let $X_f = 2$, if the outcome of C is not HEADS, let $X_f = 0$;
- if $Y'_f = 1$, let $X'_f = 1$; otherwise, if the outcome of C is HEADS, let $X'_f = 2$, if the outcome of C is not HEADS, let $X'_f = 0$;

It is straightforward to verify that (X_f, X'_f) is sampled from a coupling between $\pi^{\sigma\wedge e\leftarrow 0}_{\mathrm{gm},f}$ and $\pi^{\sigma\wedge e\leftarrow 1}_{\mathrm{gm},f}$. By the coupling inequality (7) and as Y_f and Y'_f are optimally coupled, we have

$$d_{\mathrm{TV}}\left(\pi^{\sigma\wedge e\leftarrow 0}_{\mathrm{gm},f}, \pi^{\sigma\wedge e\leftarrow 1}_{\mathrm{gm},f}\right) \leq \mathbf{Pr}\left[X_f \neq X'_f\right] = \mathbf{Pr}\left[Y_f \neq Y'_f\right] = d_{\mathrm{TV}}\left(\pi^{\tau\wedge e\leftarrow 0}_{\mathrm{sg},f}, \pi^{\tau\wedge e\leftarrow 1}_{\mathrm{sg},f}\right).$$

Similarly, we have

$$d_{\mathrm{TV}}\left(\pi^{\sigma\wedge e\leftarrow 0}_{\mathrm{gm},f}, \pi^{\sigma\wedge e\leftarrow 2}_{\mathrm{gm},f}\right) = 0 \quad \text{and} \quad d_{\mathrm{TV}}\left(\pi^{\sigma\wedge e\leftarrow 1}_{\mathrm{gm},f}, \pi^{\sigma\wedge e\leftarrow 2}_{\mathrm{gm},f}\right) \leq d_{\mathrm{TV}}\left(\pi^{\tau\wedge e\leftarrow 0}_{\mathrm{sg},f}, \pi^{\tau\wedge e\leftarrow 1}_{\mathrm{sg},f}\right).$$

Hence, by Lemma 5.4,

$$\left\|\Psi^\sigma_{\mathrm{gm}}\right\|_\infty \leq \left\|\Psi^\tau_{\mathrm{sg}}\right\|_\infty \leq \zeta. \qquad \square$$

Lemma 5.6. *π_{gm} is b-marginally bounded, where $b = \eta_{\min}^2 \min\{1-2p_{\max}, p_{\min}\}$.*

Proof. Consider the marginal distribution of an edge $e = (u,v)$. Let $e_1, \ldots, e_k$ be the edges adjacent to either u or v (but not both). Suppose we have an arbitrary pinning X on $\Lambda \subset E$ and $e \notin \Lambda$. Let Y be an arbitrary pinning on $\Lambda \cup \{e_1, \ldots, e_k\}$ that is consistent with X. The true marginal of e under X is a linear combination of marginals conditioned on all possibilities of Y (namely, we first sample Y and then sample e conditioned on Y). Thus, to establish a lower bound, it suffices to establish a lower bound under any Y. Given Y, the marginal of e depends only on p_e and whether u or v is in $O(Y)$. These cases are verified as follows.

- $u, v \notin O(Y^{e\to 0})$, where $Y^{e\to 0}$ is the configuration of Y with e further pinned to 0. In this case the marginal is at least
$$\frac{\min\{1-2p_e, p_e\eta_u\eta_v, p_e\}}{1-2p_e+p_e\eta_u\eta_v+p_e} \geq \min\{1-2p_e, p_e\eta_u\eta_v\}.$$
Note that the denominator is no greater than 1 because $\eta_u, \eta_v \leq 1$.
- $u, v \in O(Y^{e\to 0})$. Then the marginal is at least
$$\frac{\min\{1-2p_e\eta_u\eta_v, p_e, p_e\eta_u\eta_v\}}{(1-2p_e)\eta_u\eta_v+p_e+p_e\eta_u\eta_v} \geq \min\{(1-2p_e)\eta_u\eta_v, p_e\}.$$
- In the remaining cases, assume w.l.o.g. $u \in O(Y^{e\to 0})$ while $v \notin O(Y^{e\to 0})$. Then the marginal is at least

$$\frac{\min\{(1-2p_e)\eta_u, p_e\eta_v, p_e\eta_u\}}{(1-2p_e)\eta_u+p_e\eta_v+p_e\eta_u} = \begin{cases} \frac{\min\{(1-2p_e)\frac{\eta_u}{\eta_v}, p_e\frac{\eta_u}{\eta_v}\}}{(1-2p_e)\frac{\eta_u}{\eta_v}+p_e+p_e\frac{\eta_u}{\eta_v}} \geq \min\{(1-2p_e)\frac{\eta_u}{\eta_v}, p_e\frac{\eta_u}{\eta_v}\}, & \text{if } \eta_u < \eta_v; \\ \frac{\min\{(1-2p_e), p_e\frac{\eta_v}{\eta_u}\}}{(1-2p_e)+p_e\frac{\eta_v}{\eta_u}+p_e} \geq \min\{(1-2p_e), p_e\frac{\eta_v}{\eta_u}\}, & \text{otherwise.} \end{cases}$$

In all cases, the value
$$b = \eta_{\min}^2 \min\{1-2p_{\max}, p_{\min}\}$$
suffices as a marginal lower bound. □

Proof of Lemma 5.1. Combine Theorem 2.5, Observation 3.1, Lemma 5.3, Lemma 5.6 and $m \leq n\Delta$. □

6. Rapid mixing of Glauber dynamics on the random cluster model

Let $\boldsymbol{p} = (p_e)_{e\in E}$ and $\boldsymbol{\eta} = (\eta_v)_{v\in V}$, where $0 < p_e < 1/2$ and $0 < \eta_v < 1$. Let π_{wrc} denote the distribution specified by the random cluster model with parameters $2\boldsymbol{p}$ and $\boldsymbol{\lambda}$, where $\lambda_v = \frac{1-\eta_v}{1+\eta_v}$. Let $\Omega(\pi_{\text{wrc}})$ denote the support of π_{wrc}. We use $P_{\text{GlauberRC}}$ to denote Glauber dynamics on π_{wrc}.

Lemma 6.1. *Let π_{wrc} be the distribution specified by weighted random cluster model with parameters $(2\boldsymbol{p}, \boldsymbol{\lambda})$. The Glauber dynamics $P_{\text{GlauberRC}}$ satisfies that for any distribution ν with support $\Omega(\nu) \subseteq \Omega(\pi_{\text{wrc}})$,*

- $D_{\chi^2}\left(\nu P^{\downarrow}_{\text{GlauberRC}} \parallel \pi_{\text{wrc}} P^{\downarrow}_{\text{GlauberRC}}\right) \leq \left(1-\frac{\alpha}{m^2}\right) D_{\chi^2}(\nu \parallel \pi_{\text{wrc}})$,
- $D_{\text{KL}}\left(\nu P^{\downarrow}_{\text{GlauberRC}} \parallel \pi_{\text{wrc}} P^{\downarrow}_{\text{GlauberRC}}\right) \leq \left(1-\frac{1}{Cn}\right) D_{\text{KL}}(\nu \parallel \pi_{\text{wrc}})$,

where
$$\alpha = \left(\frac{1-\lambda_{\max}}{1+\lambda_{\max}}\right)^4 \min\{p_{\min}, 1-2p_{\max}\},$$
$$C = \Delta\left(\frac{8\Delta}{(1-\lambda_{\max})^2 \min\{1-2p_{\max}, p_{\min}\}}\right)^{2+\frac{256\Delta^2}{(1-\lambda_{\max})^4 \min\{1-2p_{\max}, p_{\min}\}}},$$
$\lambda_{\max} = \max_{v\in V}\lambda_v$, $\lambda_{\min} = \min_{v\in V}\lambda_v$, $p_{\max} = \max_{e\in E} p_e$, $p_{\min} = \min_{e\in E} p_e$, Δ is the maximum degree of G, $n = |V|$ and $m = |E|$.

Lemma 6.1 projects the decay results (Lemma 4.1 and Lemma 5.1) from the grand model to the random cluster model. Lemma 6.1 is proved by a comparison lemma in Section 6.1 that works for general projections and f-divergences.

Lemma 6.1 provides an entropy decay rate and a χ^2-divergence decay rate. When $\lambda_{\max}$ is bounded away from 1, the entropy decay rate is better. On the other hand, the χ^2-divergence decay rate has a better dependency on $1-\lambda_{\max}$. In particular, when $\lambda_{\max} = 1$, namely when some vertices do not have external fields, neither statement provides any decay. In such cases, we can perturb λ by a factor of $1/n$. This incurs a cost of a polynomial factor in n for α and an exponentially large factor for C. Thus, we need to apply the χ^2-divergence decay rate in Lemma 6.1 after perturbation in the $\lambda_{\max} = 1$ case. Specifically, in Section 6.2 we showed the following.

Lemma 6.2. *Let π_{wrc} be the distribution specified by the weighted random cluster model with parameters $(2\boldsymbol{p}, \boldsymbol{\lambda})$. The Glauber dynamics $P_{\mathrm{GlauberRC}}$ satisfies that for any distribution ν with support $\Omega(\nu) \subseteq \Omega(\pi_{\mathrm{wrc}})$,*

$$D_{\chi^2}\left(\nu P^{\downarrow}_{\mathrm{GlauberRC}} \parallel \pi_{\mathrm{wrc}} P^{\downarrow}_{\mathrm{GlauberRC}}\right) \leq \left(1 - \frac{\min\{p_{\min}, 1-2p_{\max}\}}{10^4 n^4 m^2}\right) D_{\chi^2}\left(\nu \parallel \pi_{\mathrm{wrc}}\right).$$

We remark that both Lemma 6.1 and Lemma 6.2 consider the random cluster model specified by parameters $(2\boldsymbol{p}, \boldsymbol{\lambda})$. Combining Lemma 6.1 and Lemma 6.2, we have the following mixing result for the Glauber dynamics on random cluster model.

Theorem 6.3. *Let $G = (V, E)$ be a n-vertex and m-edge graph with maximum degree Δ. Let $\boldsymbol{p} = (p_e)_{e\in E}$ and $\boldsymbol{\lambda} = (\lambda_v)_{v\in V}$, where $0 < p_e < 1$ and $0 < \lambda_v \leq 1$. Let π_{wrc} be the distribution specified by the random cluster model with parameters $(\boldsymbol{p}, \boldsymbol{\lambda})$. The mixing of Glauber dynamics $P_{\mathrm{GlauberRC}}$ on π_{wrc} satisfies*

$$T_{\mathrm{mix}}(P_{\mathrm{GlauberRC}}, \varepsilon) \leq C_1(p_{\min}, p_{\max}) \cdot \min\left\{n^4, \left(\frac{1}{1-\lambda_{\max}}\right)^4\right\} \cdot m^2 \cdot \left(\log\frac{1}{\varepsilon} + m\right),$$

where $C_1(p_{\min}, p_{\max}) = O\left(\frac{1}{\min\{p_{\min}, 1-p_{\max}\}} \log \frac{1}{\min\{p_{\min}, 1-p_{\max}\}}\right)$.

Furthermore, if there exists $\delta > 0$ such that $\lambda_v \leq 1 - \delta$ for all $v \in V$, then the mixing time satisfies

$$T_{\mathrm{mix}}(P_{\mathrm{GlauberRC}}, \varepsilon) \leq C_2(\Delta, \delta, p_{\min}, p_{\max}) \cdot n\left(\log n + \log\frac{1}{\varepsilon}\right),$$

where $C_2(\Delta, \delta, p_{\min}, p_{\max}) = \left(\frac{\Delta}{\delta^2 \min\{p_{\min}, 1-p_{\max}\}}\right)^{O\left(\frac{\Delta^2}{\delta^4 \min\{p_{\min}, 1-p_{\max}\}}\right)}$.

Proof. Let $\pi_{\mathrm{wrc,min}} = \min_{S\subseteq E} \pi_{\mathrm{wrc}}(S)$ denote the minimum probability in π_{wrc}. It is straightforward to verify that $\pi_{\mathrm{wrc,min}} \geq \min\{p_{\min}, 1-p_{\max}\}^m / 2^{m+n}$. By the data processing inequality,

$$D_f\left(\nu P_{\mathrm{GlauberRC}} \parallel \pi_{\mathrm{wrc}}\right) = D_f\left(\nu P_{\mathrm{GlauberRC}} \parallel \pi_{\mathrm{wrc}} P_{\mathrm{GlauberRC}}\right) \leq D_f\left(\nu P^{\downarrow}_{\mathrm{GlauberRC}} \parallel \pi_{\mathrm{wrc}} P^{\downarrow}_{\mathrm{GlauberRC}}\right).$$

By Lemma 6.1 and Lemma 6.2, we know that after each transition step of Glauber dynamics, the χ^2-divergence and KL-divergence between the current distribution of the stationary distribution decays by factors specified earlier. The χ^2-divergence between the initial distribution and the stationary distribution is at most $\frac{1}{\pi_{\mathrm{wrc,min}}}$, and the KL-divergence is at most $\log\frac{1}{\pi_{\mathrm{wrc,min}}}$. By Lemma 6.1, Lemma 6.2, and (8),

$$\begin{aligned} T_{\mathrm{mix}}(P_{\mathrm{GlauberRC}}, \varepsilon) &\leq \frac{10^4}{\min\{p_{\min}/2, 1-p_{\max}\}} \cdot \min\left\{n^4, \left(\frac{1+\lambda_{\max}}{1-\lambda_{\max}}\right)^4\right\} \cdot m^2 \left(\log\frac{1}{\varepsilon^2 \pi_{\mathrm{wrc,min}}}\right) \\ &\leq C_1(p_{\min}, p_{\max}) \cdot \min\left\{n^4, \left(\frac{1}{1-\lambda_{\max}}\right)^4\right\} \cdot m^2 \cdot \left(\log\frac{1}{\varepsilon} + m\right). \end{aligned}$$

Note that $1 < 1 + \lambda_{\max} \leq 2$.

By Lemma 6.1, (9) and $m \leq \Delta n$, if for all $\lambda_v \leq 1 - \delta$, then we have $1 - \lambda_{\max} \geq \delta$ and

$$\begin{aligned} T_{\mathrm{mix}}(P_{\mathrm{GlauberRC}}, \varepsilon) &\leq \Delta\left(\frac{8\Delta}{\delta^2 \min\{1-p_{\max}, p_{\min}/2\}}\right)^{2+\frac{256\Delta^2}{\delta^4 \min\{1-p_{\max}, p_{\min}/2\}}} \cdot n\left(\log\log\frac{1}{\pi_{\mathrm{wrc,min}}} + \log\frac{1}{2\varepsilon^2}\right) \\ &\leq C_2(\Delta, \delta, p_{\min}, p_{\max}) \cdot n\left(\log n + \log\frac{1}{\varepsilon}\right). \end{aligned}$$

□

6.1. **Comparing the decay rates of down walks.** Here we consider a general projection from a larger state space to a smaller one. Let Q and R be two finite sets, and let $\Omega \subseteq Q^V$ be the state space. Consider a mapping $g : Q \to R$. (Note that here we can restrict R to the range of g without changing the rest of the argument. In other words, after the mapping the effective domain is never larger than Q, although we do not need to require $|Q| \geq |R|$ a priori.) Given any $\sigma \in \Omega$, we map σ to $\tau = (\tau_v)_{v\in V}$,

where $\tau_v = g(\sigma_v)$. We abuse the notation and denote $\tau = g(\sigma)$. Let $\Omega' = \{g(\sigma) \mid \sigma \in \Omega\} \subseteq R^V$. Define the projection matrix $P : \Omega \times \Omega' \to \{0, 1\}$:

$$\forall \sigma \in \Omega, \tau \in \Omega', \quad P(\sigma, \tau) = \mathbb{I}[\tau = g(\sigma)].$$

We remark that P is a stochastic matrix.

Let π be a distribution with support Ω. Define the distribution $\mu = \pi P$ with support Ω'. Let $P^{\downarrow}_{\text{Glauber},\pi} : \Omega \times \Omega_{\text{down}} \to \mathbb{R}_{\geq 0}$ denote the down walk of Glauber dynamics on π, where $\Omega_{\text{down}} = \{\sigma_{V\setminus\{v\}} \mid v \in V \wedge \sigma \in \Omega\}$. Given any configuration $\sigma \in \Omega$, $P^{\downarrow}_{\text{Glauber},\pi}$ picks a variable $v \in V$ uniformly at random, and then transforms σ to $\sigma_{V\setminus\{v\}}$ by dropping the value of v. Similarly, let $P^{\downarrow}_{\text{Glauber},\mu}$ denote the down walk of Glauber dynamics on the distribution $\mu = \pi P$.

Lemma 6.4. *Let $0 < \delta < 1$. Let $f : \mathbb{R}_{\geq 0} \to \mathbb{R}$ be a convex function with $f(1) = 0$. If $P^{\downarrow}_{\text{Glauber},\pi}$ satisfies that for any distribution ν with support Ω,*

$$D_f\left(\nu P^{\downarrow}_{\text{Glauber},\pi} \parallel \pi P^{\downarrow}_{\text{Glauber},\pi}\right) \leq (1-\delta) D_f\left(\nu \parallel \pi\right),$$

then $P^{\downarrow}_{\text{Glauber},\mu}$ satisfies that for any distribution φ with support Ω',

$$D_f\left(\varphi P^{\downarrow}_{\text{Glauber},\mu} \parallel \mu P^{\downarrow}_{\text{Glauber},\mu}\right) \leq (1-\delta) D_f\left(\varphi \parallel \mu\right).$$

Proof. Given any $\rho \in \Omega_{\text{down}}$, we can map ρ to $\eta = g(\rho)$, where $\eta_u = g(\rho_u)$ for any variable u. Let $\Omega'_{\text{down}} = \{g(\rho) \mid \rho \in \Omega_{\text{down}}\}$. Define the projection matrix $P' : \Omega_{\text{down}} \times \Omega'_{\text{down}} \to \{0, 1\}$:

$$\forall \rho \in \Omega_{\text{down}}, \eta \in \Omega'_{\text{down}}, \quad P'(\rho, \eta) = \mathbb{I}[\eta = g(\rho)].$$

We remark that P' is a stochastic matrix. Since both P and P' project the value of each variable independently, the following equation is straightforward to verify

$$P^{\downarrow}_{\text{Glauber},\pi} \cdot P' = P \cdot P^{\downarrow}_{\text{Glauber},\mu}. \tag{29}$$

For any configuration $\tau \in \Omega'$, define the distribution π^τ over Ω by

$$\forall \sigma \in \Omega, \quad \pi^\tau(\sigma) = \frac{\mathbb{I}[g(\sigma) = \tau]\pi(\sigma)}{\mu(\tau)}.$$

For any $\sigma \in \Omega$, let $\tau = g(\sigma)$, it holds that $\pi(\sigma) = \mu(\tau)\pi^\tau(\sigma)$. Fix a distribution φ with support Ω'. Define the distribution ν by

$$\forall \sigma \in \Omega, \quad \nu(\sigma) = \varphi(\tau)\pi^\tau(\sigma), \quad \text{where } \tau = g(\sigma). \tag{30}$$

We have

$$D_f\left(\nu \parallel \pi\right) = \mathbf{E}_{\sigma\sim\pi}\left[f\left(\frac{\nu(\sigma)}{\pi(\sigma)}\right)\right] = \mathbf{E}_{\tau\sim\mu}\,\mathbf{E}_{\sigma\sim\pi^\tau}\left[f\left(\frac{\varphi(\tau)\pi^\tau(\sigma)}{\mu(\tau)\pi^\tau(\sigma)}\right)\right] = \mathbf{E}_{\tau\sim\mu}\left[f\left(\frac{\varphi(\tau)}{\mu(\tau)}\right)\right] = D_f\left(\varphi \parallel \mu\right). \tag{31}$$

By the definition in (30), we have for all $\tau \in \Omega'$,

$$(\nu P)(\tau) = \sum_{\sigma : g(\sigma) = \tau} \nu(\sigma) = \varphi(\tau) \sum_{\sigma : g(\sigma) = \tau} \pi^\tau(\sigma) = \varphi(\tau),$$

which implies $\varphi = \nu P$. Recall that $\mu = \pi P$. We have

$$\begin{aligned}
D_f\left(\varphi P^{\downarrow}_{\text{Glauber},\mu} \parallel \mu P^{\downarrow}_{\text{Glauber},\mu}\right) &= D_f\left(\nu P P^{\downarrow}_{\text{Glauber},\mu} \parallel \pi P P^{\downarrow}_{\text{Glauber},\mu}\right) \\
\text{(by (29))} \quad &= D_f\left(\nu P^{\downarrow}_{\text{Glauber},\pi} P' \parallel \pi P^{\downarrow}_{\text{Glauber},\pi} P'\right) \\
\text{(by data processing inequality)} \quad &\leq D_f\left(\nu P^{\downarrow}_{\text{Glauber},\pi} \parallel \pi P^{\downarrow}_{\text{Glauber},\pi}\right) \\
\text{(by assumption)} \quad &\leq (1-\delta) D_f\left(\nu \parallel \pi\right) \\
\text{(by (31))} \quad &= (1-\delta) D_f\left(\varphi \parallel \mu\right).
\end{aligned}$$

□

We are now ready to prove Lemma 6.1.

Proof of Lemma 6.1. Let $\Omega = \{0,1,2\}^E$ denote the support of π_{gm}. Define the map g by $g(0) = 0$, $g(1) = 1$ and $g(2) = 1$. By Lemma 3.2, it holds that $\pi_{\mathrm{wrc}} = \pi_{\mathrm{gm}} P$. Lemma 6.1 follows from Lemma 4.1, Lemma 5.1 and Lemma 6.4. □

6.2. **Faster mixing via perturbed chains.** Given a subgraph-world model $(G; \boldsymbol{p}, \boldsymbol{\eta})$, we define the "perturbed" model $(G; \boldsymbol{p}, \widehat{\boldsymbol{\eta}})$ by

$$\widehat{\eta}_v = \begin{cases} \frac{1}{n}, & \text{if } 0 \le \eta_v \le \frac{1}{n} \\ \eta_v, & \text{otherwise.} \end{cases} \tag{32}$$

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 $\widehat{\lambda}_v = \frac{1-\widehat{\eta}_v}{1+\widehat{\eta}_v}$. Let $\widehat{P_{\mathrm{wrc}}}$ denote the Glauber dynamics on $\widehat{\pi_{\mathrm{wrc}}}$. Let $\widehat{P_{\mathrm{wrc}}}^{\downarrow}$ denote the down-walk of $\widehat{P_{\mathrm{wrc}}}$. Applying the first item of Lemma 6.1 to the perturbed random-cluster model $(G; 2\boldsymbol{p}, \widehat{\boldsymbol{\lambda}})$ yields that for any distribution ν,

$$D_{\chi^2}\left(\nu \widehat{P_{\mathrm{wrc}}}^{\downarrow} \,\|\, \widehat{\pi_{\mathrm{wrc}}} \widehat{P_{\mathrm{wrc}}}^{\downarrow}\right) \le \left(1 - \frac{\min\{p_{\min}, 1-2p_{\max}\}}{m^2 n^4}\right) D_{\chi^2}\left(\nu \,\|\, \widehat{\pi_{\mathrm{wrc}}}\right)$$

By Proposition 2.4, we know that

$$\mathfrak{Gap}(\widehat{P_{\mathrm{wrc}}}) \ge \frac{\min\{p_{\min}, 1-2p_{\max}\}}{m^2 n^4}.$$

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 6.5. *For any $R \subseteq E$,*

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

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

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

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

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

as $n \ge 2$.

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

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

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

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

which concludes this lemma. □

We also have a bound on the ratio of the transition probability between the original and perturbed model in the Glauber dynamics.

Lemma 6.6. *Let P_{wrc} and $\widehat{P_{\mathrm{wrc}}}$ be the transition matrices of Glauber dynamics on the random cluster models $(G; 2\boldsymbol{p}, \boldsymbol{\lambda})$ and $(G; 2\boldsymbol{p}, \widehat{\boldsymbol{\lambda}})$ respectively. Then it holds that*

$$\frac{1}{9\mathrm{e}} \le \frac{\widehat{P_{\mathrm{wrc}}}(Z, Z')}{P_{\mathrm{wrc}}(Z, Z')} \le 9\mathrm{e} \qquad \text{for all } |Z \oplus Z'| = 1.$$

Proof. 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

$$\frac{1}{9\mathrm{e}} \le \frac{\widehat{P_{\mathrm{wrc}}}(Z, Z')}{P_{\mathrm{wrc}}(Z, Z')} = \frac{\widehat{\pi_{\mathrm{wrc}}}(Z')(\pi_{\mathrm{wrc}}(Z) + \pi_{\mathrm{wrc}}(Z'))}{(\widehat{\pi_{\mathrm{wrc}}}(Z) + \widehat{\pi_{\mathrm{wrc}}}(Z'))\pi_{\mathrm{wrc}}(Z')} \le 9\mathrm{e}. \qquad \square$$

Now we are ready to prove Lemma 6.2.

Proof of Lemma 6.2. Fix a test function f. Denote by $\mathcal{E}(f, f), \widehat{\mathcal{E}}(f, f)$ the Dirichlet form of P_{wrc} and $\widehat{P_{\mathrm{wrc}}}$ respectively. Denote by $\mathrm{Var}[f]$ and $\widehat{\mathrm{Var}}[f]$ the variance of f with respect to π_{wrc} and $\widehat{\pi_{\mathrm{wrc}}}$ respectively. Then by Lemma 6.5 and Lemma 6.6,

$$\begin{aligned}
\frac{\mathcal{E}(f, f)}{\mathrm{Var}[f]} &= \frac{\sum_{\substack{X, Y \subseteq E \\ |X \oplus Y| = 1}} \pi_{\mathrm{wrc}}(X) P_{\mathrm{wrc}}(X, Y) \left(f(X) - f(Y)\right)^2}{\sum_{\substack{X, Y \subseteq E \\ |X \oplus Y| = 1}} \pi_{\mathrm{wrc}}(X) \pi_{\mathrm{wrc}}(Y) \left(f(X) - f(Y)\right)^2} \\
&\ge \frac{\frac{1}{9\mathrm{e}^2} \sum_{\substack{X, Y \subseteq E \\ |X \oplus Y| = 1}} \widehat{\pi_{\mathrm{wrc}}}(X) \widehat{P_{\mathrm{wrc}}}(X, Y) \left(f(X) - f(Y)\right)^2}{81 \sum_{\substack{X, Y \subseteq E \\ |X \oplus Y| = 1}} \widehat{\pi_{\mathrm{wrc}}}(X) \widehat{\pi_{\mathrm{wrc}}}(Y) \left(f(X) - f(Y)\right)^2} > \frac{1}{10^4} \frac{\widehat{\mathcal{E}}(f, f)}{\widehat{\mathrm{Var}}[f]}.
\end{aligned}$$

Therefore, $\mathfrak{Gap}(P_{\mathrm{wrc}}) \ge \frac{1}{10^4} \mathfrak{Gap}(\widehat{P_{\mathrm{wrc}}}) \ge \frac{\min\{p_{\min}, 1 - 2p_{\max}\}}{10^4 n^4 m^2}$. Lemma 6.2 follows from Proposition 2.4. $\square$

7. Rapid mixing of Swendsen-Wang 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 grand 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 $\Omega_{\mathcal{R}} = \{0, 1\}^E$ denote the distribution of the weighted random cluster model. We remark that we view π_{wrc} as a distribution over $\{0, 1\}^E$ instead of 2^E.

Let $P_{\mathrm{SW}}^{\mathrm{wrc}} = P_{\mathcal{R} \to \mathcal{I}} P_{\mathcal{I} \to \mathcal{R}}$ denote the transition matrix of the Swendsen-Wang dynamics for weighted random cluster models as defined in Section 2.3.2, and $P_{\mathrm{GlauberRC}}$ denote the transition matrix of the Glauber dynamics for weighted random cluster models. In this section, we compare the Swendsen-Wang dynamics with the Glauber dynamics. Ullrich [Ull14] showed the following result about the variance decay (spectral gap) of the Swendsen-Wang dynamics.

Lemma 7.1 ([Ull14, Remark 2 and Theorem 5])**.** *Suppose $0 < \lambda_v \le 1$ for all $v \in V$. It holds that*

$$\mathfrak{Gap}(P_{\mathrm{SW}}^{\mathrm{wrc}}) \ge \frac{\mathfrak{Gap}\left(P_{\mathrm{GlauberRC}}\right)}{2}.$$

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 $\lambda_v \in (0,1]$. However, there is no substantial change required to generalise it to our setting. Alternatively, we provide a somewhat simpler proof of Lemma 7.1 in Remark 7.5.

Lemma 7.1 only compares the decay rate of the variance. The main technical result in this section is the following comparison lemma on the decay rate of the relative entropy.

Lemma 7.2. *Suppose $0 < \lambda_v \le 1$ for all $v \in V$. Let $0 < \delta < 1$. For any distribution ν over $\Omega_{\mathcal{R}}$, if*

$$D_{\mathrm{KL}}\left(\nu P^{\downarrow}_{\text{GlauberRC}} \parallel \pi_{\mathrm{wrc}} P^{\downarrow}_{\text{GlauberRC}}\right) \le (1-\delta) D_{\mathrm{KL}}\left(\nu \parallel \pi_{\mathrm{wrc}}\right),$$

then it holds that

$$D_{\mathrm{KL}}\left(\nu P^{\mathrm{wrc}}_{\mathrm{SW}} \parallel \pi_{\mathrm{wrc}} P^{\mathrm{wrc}}_{\mathrm{SW}}\right) \le \left(1 - \frac{\delta}{4}\right) D_{\mathrm{KL}}\left(\nu \parallel \pi_{\mathrm{wrc}}\right).$$

We are now ready to prove the main results in Theorem 1.1 and Theorem 1.2.

Proofs of Theorem 1.1 and Theorem 1.2. Let $\pi_{\mathrm{wrc,min}} = \min_{S \subseteq E} \pi_{\mathrm{wrc}}(S)$ denote the minimum probability in π_{wrc}. It is straightforward to verify that $\pi_{\mathrm{wrc,min}} \ge \min\{p_{\min}, 1-p_{\max}\}^m / 2^{m+n}$. By the data processing inequality, Proposition 2.3 and Proposition 2.4, we have

$$\begin{aligned} D_{\chi^2}\left(\nu P^{\mathrm{wrc}}_{\mathrm{SW}} \parallel \pi_{\mathrm{wrc}}\right) &= D_{\chi^2}\left(\nu P^{\mathrm{wrc}}_{\mathrm{SW}} \parallel \pi_{\mathrm{wrc}} P^{\mathrm{wrc}}_{\mathrm{SW}}\right) \le D_{\chi^2}\left(\nu P_{\mathcal{R}\to\mathcal{I}} \parallel \pi_{\mathrm{wrc}} P_{\mathcal{R}\to\mathcal{I}}\right) \\ &\le \left(1 - \mathfrak{Gap}(P^{\mathrm{wrc}}_{\mathrm{SW}})\right) D_{\chi^2}\left(\nu \parallel \pi_{\mathrm{wrc}}\right). \end{aligned}$$

A lower bound of $\mathfrak{Gap}(P^{\mathrm{wrc}}_{\mathrm{SW}})$ can be obtained by Proposition 2.4, Lemma 6.2 and Lemma 7.1. Let C_1 be the constant in Theorem 6.3. By a similar calculation as that in the proof of Theorem 6.3, we have

$$T_{\mathrm{mix}}(P^{\mathrm{wrc}}_{\mathrm{SW}}, \varepsilon) \le 2C_1(p_{\min}, p_{\max}) \cdot \min\left\{n^4, \left(\frac{1}{1-\lambda_{\max}}\right)^4\right\} \cdot m^2 \cdot \left(\log\frac{1}{\varepsilon} + m\right).$$

By (15), the mixing time of Swendsen-Wang dynamics on Ising model satisfies

$$T_{\mathrm{mix}}(P^{\mathrm{Ising}}_{\mathrm{SW}}, \varepsilon) \le C_1'(\beta_{\min}, \beta_{\max}) \cdot \min\left\{n^4, \left(\frac{1}{1-\lambda_{\max}}\right)^4\right\} \cdot m^2 \cdot \left(\log\frac{1}{\varepsilon} + m\right),$$

where $p_{\min} = 1 - \frac{1}{\beta_{\min}}$, $p_{\max} = 1 - \frac{1}{\beta_{\max}}$, and thus

$$\begin{aligned} C_1'(\beta_{\min}, \beta_{\max}) &= O\left(\frac{1}{\min\{p_{\min}, 1-p_{\max}\}} \log \frac{1}{\min\{p_{\min}, 1-p_{\max}\}}\right) \\ &= O\left(\left(\frac{\beta_{\min}}{1-\beta_{\min}} + \beta_{\max}\right) \log\left(\frac{\beta_{\min}}{1-\beta_{\min}} + \beta_{\max}\right)\right). \end{aligned} \tag{33}$$

This proves Theorem 1.1.

For the decay of the relative entropy, the initial KL-divergence is at most $\log \frac{1}{\pi_{\mathrm{wrc,min}}}$. Let C_2 be the constant in Theorem 6.3. By Lemma 7.2, Lemma 6.1, and (9), we can use a similar calculation as that in the proof of Theorem 6.3 to obtain

$$T_{\mathrm{mix}}(P^{\mathrm{wrc}}_{\mathrm{SW}}, \varepsilon) \le 4C_2(\Delta, \delta, p_{\min}, p_{\max}) \cdot n\left(\log n + \log\frac{1}{\varepsilon}\right).$$

By (15), the mixing time of Swendsen-Wang dynamics on Ising model satisfies

$$T_{\mathrm{mix}}(P^{\mathrm{Ising}}_{\mathrm{SW}}, \varepsilon) \le C_2'(\Delta, \delta, \beta_{\min}, \beta_{\max}) \cdot n\left(\log n + \log\frac{1}{\varepsilon}\right)$$

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

where

$$C_2'(\Delta,\delta,\beta_{\min},\beta_{\max}) = \left(\frac{\Delta}{\delta^2\min\{p_{\min},1-p_{\max}\}}\right)^{O\left(\frac{\Delta^2}{\delta^4\min\{p_{\min},1-p_{\max}\}}\right)}$$
$$= \left(\frac{\Delta}{\delta^2}\left(\frac{\beta_{\min}}{1-\beta_{\min}}+\beta_{\max}\right)\right)^{O\left(\frac{\Delta^2}{\delta^4}\left(\frac{\beta_{\min}}{1-\beta_{\min}}+\beta_{\max}\right)\right)}. \tag{34}$$

This proves Theorem 1.2. □

The rest of this section is dedicated to the proof of Lemma 7.2.

7.1. **FKES distribution and single-bond dynamics.** To compare the Swendsen-Wang dynamics to the Glauber dynamics, we first introduce the FKES (Fortuin-Kasteleyn-Edwards-Sokal) distribution [FK72, ES88] π_{FKES} over $\Omega_{\mathcal{I}} \times \Omega_{\mathcal{R}}$, which couples the Ising distribution π_{Ising} and the random cluster distribution π_{wrc}:

$$\forall \sigma \in \Omega_{\mathcal{I}}, \tau \in \Omega_{\mathcal{R}}, \quad \pi_{\text{FKES}}(\sigma\tau) := \pi_{\text{Ising}}(\sigma)P_{\mathcal{I}\to\mathcal{R}}(\sigma,\tau) \overset{(\star)}{=} \pi_{\text{wrc}}(\tau)P_{\mathcal{R}\to\mathcal{I}}(\tau,\sigma), \tag{35}$$

where $\Omega_{\mathcal{I}} = \{0,1\}^V$, $\Omega_{\mathcal{R}} = \{0,1\}^E$, $P_{\mathcal{I}\to\mathcal{R}}$ and $P_{\mathcal{R}\to\mathcal{I}}$ are defined in (10) and (11) respectively. The equation $(\star)$ holds due to Proposition 2.3. We use $\Omega_{\text{FKES}} \subseteq \Omega_{\mathcal{I}} \times \Omega_{\mathcal{R}}$ to denote the support of the distribution π_{FKES}. The above equation shows that

- the marginal distribution projected from π_{FKES} to $\Omega_{\mathcal{I}}$ is π_{Ising};
- the marginal distribution projected from π_{FKES} to $\Omega_{\mathcal{R}}$ is π_{wrc};
- conditional on $\sigma \in \Omega_{\mathcal{I}}$, the marginal distribution projected from π_{FKES} to $\Omega_{\mathcal{R}}$ is $\mathbf{P}_{\mathcal{I}\to\mathcal{R}}(\sigma,\cdot)$;
- conditional on $\tau \in \Omega_{\mathcal{R}}$, the marginal distribution projected from π_{FKES} to $\Omega_{\mathcal{I}}$ is $\mathbf{P}_{\mathcal{R}\to\mathcal{I}}(\tau,\cdot)$.

Define the following stochastic matrix from the weighted random cluster model to the FKES model

$$\forall \tau_1 \in \Omega_{\mathcal{R}}, \sigma\tau_2 \in \Omega_{\text{FKES}}, \quad P_{\mathcal{R}\to\text{FKES}}(\tau_1,\sigma\tau_2) = P_{\mathcal{R}\to\mathcal{I}}(\tau_1,\sigma)\cdot\mathbb{I}[\tau_1=\tau_2],$$

The operator $P_{\mathcal{R}\to\text{FKES}}$ maps from $L_2(\pi_{\text{FKES}})$ to $L_2(\pi_{\text{wrc}})$, where $L_2(\pi)$ is the vector space with the inner product $\langle\cdot,\cdot\rangle_\pi$. The adjoint operator $P_{\text{FKES}\to\mathcal{R}}$ is defined by

$$\forall \sigma\tau_1 \in \Omega_{\text{FKES}}, \tau_2 \in \Omega_{\mathcal{R}}, \quad P_{\text{FKES}\to\mathcal{R}}(\sigma\tau_1,\tau_2) = \mathbb{I}[\tau_1=\tau_2].$$

For any $f \in L_2(\pi_{\text{FKES}})$ and $g \in L_2(\pi_{\text{wrc}})$, it holds that $\langle P_{\mathcal{R}\to\text{FKES}}f, g\rangle_{\pi_{\text{wrc}}} = \langle f, P_{\text{FKES}\to\mathcal{R}}g\rangle_{\pi_{\text{FKES}}}$.

Next, we define the edge down-walk on the joint distribution. Fix an edge $e \in E$. Given $\sigma\tau \in \Omega_{\text{FKES}}$, let $P_e^{\downarrow}$ denote the edge down-walk that drops the value on edge e. Formally, $P_e^{\downarrow}$ is defined on any $\sigma\tau \in \Omega_{\text{FKES}}$ and any $\sigma'\tau' \in \Omega^e_{\text{FKES}}$,

$$P_e^{\downarrow}(\sigma\tau,\sigma'\tau') = \mathbb{I}[\sigma=\sigma' \wedge \tau' = \tau_{E-e}],$$

where we use $E-e$ to denote $E\setminus\{e\}$. Let $\pi^e_{\text{FKES}} = \pi_{\text{FKES}}P_e^{\downarrow}$. Let Ω^e_{FKES} denote the support of $\pi_{\text{FKES}}P_e^{\downarrow}$. Suppose $e=\{u,v\}$. We then define the edge up-walk $P_e^{\uparrow}$, for all $\sigma'\tau' \in \Omega^e_{\text{FKES}}$ and $\sigma\tau \in \Omega_{\text{FKES}}$,

$$P_e^{\uparrow}(\sigma'\tau',\sigma\tau) = \mathbb{I}[\sigma=\sigma' \wedge \tau_{E-e}=\tau'] \times \begin{cases} p_e & \text{if } \tau_e = 1 \text{ and } \sigma(u)=\sigma(v);\\ 1-p_e & \text{if } \tau_e = 0 \text{ and } \sigma(u)=\sigma(v);\\ 0 & \text{if } \tau_e = 1 \text{ and } \sigma(u)\neq\sigma(v);\\ 1 & \text{if } \tau_e = 0 \text{ and } \sigma(u)\neq\sigma(v).\end{cases}$$

For any $f \in L_2(\pi_{\text{FKES}})$ and $g \in L_2(\pi^e_{\text{FKES}})$, it holds that $\langle P_e^{\uparrow}f, g\rangle_{\pi^e_{\text{FKES}}} = \langle f, P_e^{\downarrow}g\rangle_{\pi_{\text{FKES}}}$.

Since in each transition step of $P_e^{\uparrow}$, $\sigma'\tau'_{E-e} = \sigma\tau_{E-e}$ and the distribution of τ'_e depends only on σ_u and σ_v, the following observation is straightforward to verify.

Observation 7.3. *For any $e, f \in E$, it holds that*

- $(P_e^{\downarrow}P_e^{\uparrow})(P_f^{\downarrow}P_f^{\uparrow}) = (P_f^{\downarrow}P_f^{\uparrow})(P_e^{\downarrow}P_e^{\uparrow})$.
- $P_e^{\downarrow}P_e^{\uparrow} = (P_e^{\downarrow}P_e^{\uparrow})^2$.

The single bond dynamics $P_{\mathrm{SB}} : \Omega_{\mathcal{R}} \times \Omega_{\mathcal{R}} \to \mathbb{R}_{\geq 0}$ is defined as follows

$$P_{\mathrm{SB}} = P_{\mathcal{R}\to\mathrm{FKES}} \left(\frac{1}{m} \sum_{e\in E} P_e^{\downarrow} P_e^{\uparrow} \right) P_{\mathrm{FKES}\to\mathcal{R}}.$$

Intuitively, given any $\tau \in \Omega_{\mathcal{R}}$, P_{SB} first transforms τ into a joint configuration $\sigma\tau \in \Omega_{\mathrm{FKES}}$; samples an edge $e \in E$ uniformly at random; updates τ_e conditional on σ; drops σ and keeps the random cluster configuration τ. Similarly, we can decompose the single bond dynamics as $P_{\mathrm{SB}} = P_{\mathrm{SB}}^{\downarrow} P_{\mathrm{SB}}^{\uparrow}$:

$$P_{\mathrm{SB}}^{\downarrow} = P_{\mathcal{R}\to\mathrm{FKES}} \left(\frac{1}{m} \sum_{e\in E} P_e^{\downarrow} \right) \text{ and } P_{\mathrm{SB}}^{\uparrow} = P_E^{\uparrow} P_{\mathrm{FKES}\to\mathcal{R}}, \tag{36}$$

where for convenience, we treat $(\frac{1}{m}\sum_{e\in E} P_e^{\downarrow})$ as a matrix defined on $\Omega_{\mathrm{FKES}} \times (\cup_{e\in E}\Omega_{\mathrm{FKES}}^e)$ and $P_E^{\uparrow} : (\cup_{e\in E}\Omega_{\mathrm{FKES}}^e) \times \Omega_{\mathrm{FKES}} \to \mathbb{R}_{\geq 0}$ is defined by $P_E^{\uparrow}(x,y) = P_e^{\uparrow}(x,y)$ where $x \in \Omega_{\mathrm{FKES}}^e$ for some $e \in E$ and $y \in \Omega_{\mathrm{FKES}}$. Note that once x is given, e is uniquely determined, and $P_E^{\uparrow}$ agrees with $P_e^{\uparrow}$. It is straightforward to check $(\frac{1}{m}\sum_{e\in E} P_e^{\downarrow})$ and $P_E^{\uparrow}$ is a pair of adjoint operators.

Lemma 7.4. *Suppose $0 < \lambda_v \leq 1$ for all $v \in V$. Let $0 < \delta < 1$. For any distribution ν over $\Omega_{\mathcal{R}}$, if*

$$D_{\mathrm{KL}}\left(\nu P_{\mathrm{GlauberRC}}^{\downarrow} \parallel \pi_{\mathrm{wrc}} P_{\mathrm{GlauberRC}}^{\downarrow}\right) \leq (1-\delta) D_{\mathrm{KL}}\left(\nu \parallel \pi_{\mathrm{wrc}}\right),$$

then it holds that

$$D_{\mathrm{KL}}\left(\nu P_{\mathrm{SB}}^{\downarrow} \parallel \pi_{\mathrm{wrc}} P_{\mathrm{SB}}^{\downarrow}\right) \leq \left(1 - \frac{\delta}{4}\right) D_{\mathrm{KL}}\left(\nu \parallel \pi_{\mathrm{wrc}}\right). \tag{37}$$

The proof of Lemma 7.4 is deferred to Section 7.2. We are now ready to prove Lemma 7.2.

Proof of Lemma 7.2. By Observation 7.3, the Swendsen-Wang dynamics P_{SW}^{wrc} can be written as

$$\begin{aligned} P_{\mathrm{SW}}^{\mathrm{wrc}} &= P_{\mathcal{R}\to\mathrm{FKES}} \left(\prod_{e\in E} P_e^{\downarrow} P_e^{\uparrow} \right) P_{\mathrm{FKES}\to\mathcal{R}} = P_{\mathcal{R}\to\mathrm{FKES}} \left(\frac{1}{m} \sum_{e\in E} P_e^{\downarrow} P_e^{\uparrow} \right) \left(\prod_{e\in E} P_e^{\downarrow} P_e^{\uparrow} \right) P_{\mathrm{FKES}\to\mathcal{R}} \\ &= P_{\mathcal{R}\to\mathrm{FKES}} \left(\frac{1}{m} \sum_{e\in E} P_e^{\downarrow} \right) P_E^{\uparrow} \left(\prod_{e\in E} P_e^{\downarrow} P_e^{\uparrow} \right) P_{\mathrm{FKES}\to\mathcal{R}} = P_{\mathrm{SB}}^{\downarrow} P_E^{\uparrow} \left(\prod_{e\in E} P_e^{\downarrow} P_e^{\uparrow} \right) P_{\mathrm{FKES}\to\mathcal{R}}. \end{aligned}$$

Hence, by the data processing inequality, we have

$$D_{\mathrm{KL}}\left(\nu P_{\mathrm{SW}}^{\mathrm{wrc}} \parallel \pi_{\mathrm{wrc}} P_{\mathrm{SW}}^{\mathrm{wrc}}\right) \leq D_{\mathrm{KL}}\left(\nu P_{\mathrm{SB}}^{\downarrow} \parallel \pi_{\mathrm{wrc}} P_{\mathrm{SB}}^{\downarrow}\right) \leq \left(1 - \frac{\delta}{4}\right) D_{\mathrm{KL}}\left(\nu \parallel \pi_{\mathrm{wrc}}\right),$$

where the last inequality holds due to Lemma 7.4. □

Remark 7.5 (a simple proof of the main result in [Ull14] and Lemma 7.1)**.** If we replace KL-divergence in the above proof with χ^2-divergence, the same proof shows that for any distribution ν,

$$D_{\chi^2}\left(\nu P_{\mathrm{SW}}^{\mathrm{wrc}} \parallel \pi_{\mathrm{wrc}} P_{\mathrm{SW}}^{\mathrm{wrc}}\right) \leq D_{\chi^2}\left(\nu P_{\mathrm{SB}}^{\downarrow} \parallel \pi_{\mathrm{wrc}} P_{\mathrm{SB}}^{\downarrow}\right).$$

By Proposition 2.4, we have the following result

$$\mathfrak{Gap}((P_{\mathrm{SW}}^{\mathrm{wrc}})^2) \geq \mathfrak{Gap}(P_{\mathrm{SB}}) \quad \Longrightarrow \quad \mathfrak{Gap}(P_{\mathrm{SW}}^{\mathrm{wrc}}) \geq \frac{\mathfrak{Gap}(P_{\mathrm{SB}})}{2},$$

which recovers the main result in [Ull14] up to the factor 2.

The above analysis loses a factor of 2 because we compare $P_{\mathrm{SW}}^{\mathrm{wrc}}$ with $P_{\mathrm{SB}}^{\downarrow}$. Note that $P_{\mathrm{SW}}^{\mathrm{wrc}}$ can be decomposed as $P_{\mathcal{R}\to\mathcal{I}} \cdot P_{\mathcal{I}\to\mathcal{R}}$. This factor 2 can be saved by comparing the intermediate step $P_{\mathcal{R}\to\mathcal{I}}$ with $P_{\mathrm{SB}}^{\downarrow}$. Define the intermediate state space $\Omega_{\mathcal{R}}^* = \{0, 1, *\}^E$, where for any $\tau \in \Omega_{\mathcal{R}}^*$ and $e \in E$, $\tau_e = *$ means that e is not assigned with any value, in other words, the value on e is dropped. We can view $P_e^{\downarrow}$ as a random walk on $\Omega_{\mathcal{I}} \times \Omega_{\mathcal{R}}^*$ such that given any $\sigma\tau \in \Omega_{\mathcal{I}} \times \Omega_{\mathcal{R}}^*$, $P_e^{\downarrow}$ drops the value τ_e (i.e.

sets $\tau_e = *$) and keeps $\sigma\tau_{E\setminus\{e\}}$ unchanged. It is straightforward to verify that $P_{\mathcal{R}\to\mathcal{I}}$ is equivalent to $P_{\mathcal{R}\to\mathrm{FKES}}\prod_{e\in E}P_e^{\downarrow}$. Note that

$$P_{\mathcal{R}\to\mathrm{FKES}}\prod_{e\in E}P_e^{\downarrow} = P_{\mathcal{R}\to\mathrm{FKES}}\left(\frac{1}{m}\sum_{e\in E}P_e^{\downarrow}\right)\prod_{e\in E}P_e^{\downarrow},$$

as updating an edge twice is the same as updating it once. Recall (36) that $P_{\mathcal{R}\to\mathrm{FKES}}(\frac{1}{m}\sum_{e\in E}P_e^{\downarrow})$ is equivalent to $P_{\mathrm{SB}}^{\downarrow}$. By the data processing inequality, we have the following stronger result

$$D_{\chi^2}\left(\nu P_{\mathcal{R}\to\mathcal{I}} \parallel \pi_{\mathrm{wrc}}P_{\mathcal{R}\to\mathcal{I}}\right) \le D_{\chi^2}\left(\nu P_{\mathrm{SB}}^{\downarrow} \parallel \pi_{\mathrm{wrc}}P_{\mathrm{SB}}^{\downarrow}\right),$$

which gives a better bound $\mathfrak{Gap}(P_{\mathrm{SW}}^{\mathrm{wrc}}) \ge \mathfrak{Gap}(P_{\mathrm{SB}})$, matching [Ull14].

For Lemma 7.1, we still need to compare $\mathfrak{Gap}(P_{\mathrm{SB}})$ with $\mathfrak{Gap}(P_{\mathrm{GlauberRC}})$. We claim that $\mathfrak{Gap}(P_{\mathrm{SB}}) \ge \mathfrak{Gap}(P_{\mathrm{GlauberRC}})/2$. By a simple comparison argument through the Dirichlet form (see for example [LP17, Section 13.3]), it suffices to show

$$\frac{P_{\mathrm{GlauberRC}}(A,B)}{P_{\mathrm{SB}}(A,B)} \le 2$$

for all $A, B \subseteq E$ such that $|A \oplus B| = 1$. Let e be the edge where A and B differ. By writing down the transition probability explicitly, the above ratio is 1 if A and B give the same connected components, and

$$\frac{1}{1-\left(1-\frac{1+XY}{(1+X)(1+Y)}\right)p_e}$$

otherwise, where $X = \prod_{v\in C_1}\lambda_v$, $Y = \prod_{w\in C_2}\lambda_w$, and C_1, C_2 are the two components created by disconnecting e. Using the inequality that $1/2 \le (1+XY)/((1+X)(1+Y)) \le 1$ for all $0 \le X, Y \le 1$, the above ratio is bounded by 2.

7.2. **Comparing Glauber dynamics to single-bond dynamics.** We first introduce some notations. Let μ be a distribution with support $\Omega \subseteq Q^V$.

For any $S \subseteq V$, we use μ_S to denote the marginal distribution on S induced by μ. Let $\Omega(\mu_S)$ denote the support of μ_S. Given any $x_S \in \Omega(\mu_S)$, we use μ^{x_S} to denote the distribution over Ω obtained from μ conditional on x_S. Formally, for any $y \in \Omega$, $\mu^{x_S}(y) = \mathbb{I}[y_S = x_S]\mu(y)/\mu_S(x_S)$, where y_S is the restriction of y on S. For any $\Lambda \subseteq V$, we use $\mu_\Lambda^{x_S}$ to denote the marginal distribution on Λ induced by μ^{x_S}. We need the following chain rule of the KL-divergence. Such a result is very well-known. See for example [CP21, Lemma 3.1].

Lemma 7.6. *For any distribution ν be a distribution over Ω, any $S \subseteq V$, it holds that*

$$D_{\mathrm{KL}}(\nu \parallel \mu) = D_{\mathrm{KL}}(\nu_S \parallel \mu_S) + \mathbf{E}_{x_S\sim\nu_S} D_{\mathrm{KL}}(\nu^{x_S} \parallel \mu^{x_S}) = D_{\mathrm{KL}}(\nu_S \parallel \mu_S) + \mu[\mathrm{Ent}_{V-S}(f)],$$

where $V - S = V \setminus S$ and $f : \Omega \to \mathbb{R}_{\ge 0}$ is defined by $f(x) = \nu(x)/\mu(x)$ and

$$\mu[\mathrm{Ent}_{V-S}(f)] = \sum_{x_S\in\Omega(\mu_S)} \mu_S(x_S)\mathrm{Ent}_{\mu^{x_S}}(f).$$

Proof. The first equation $D_{\mathrm{KL}}(\nu \parallel \mu) = D_{\mathrm{KL}}(\nu_S \parallel \mu_S) + \mathbf{E}_{x_S\sim\nu_S} D_{\mathrm{KL}}(\nu^{x_S} \parallel \mu^{x_S})$ follows directly from the standard chain rule of KL-divergence. To prove the second equation, for any $x_S \in \Omega(\nu_S)$, define

$$\forall y \in \Omega, \quad g^{x_S}(y) := \begin{cases} \frac{\nu^{x_S}(y)}{\mu^{x_S}(y)} = \frac{\nu(y)\mu_S(x_S)}{\mu(y)\nu_S(x_S)} = \frac{\mu_S(x_S)}{\nu_S(x_S)}f(y) & \text{if } y_S = x_S; \\ 0 & \text{otherwise.} \end{cases}$$

Since $\Omega(\nu_S) \subseteq \Omega(\mu_S)$, we have

$$\mathbf{E}_{x_S\sim\nu_S} D_{\mathrm{KL}}(\nu^{x_S} \parallel \mu^{x_S}) = \sum_{x_S\in\Omega(\nu_S)} \nu(x_S)\mathrm{Ent}_{\mu^{x_S}}(g^{x_S}) = \sum_{x_S\in\Omega(\nu_S)} \nu(x_S)\mathrm{Ent}_{\mu^{x_S}}\left(\frac{\mu_S(x_S)}{\nu_S(x_S)}f\right)$$

$$(\text{as } \mathrm{Ent}_{\mu^{x_S}}(cf) = c\,\mathrm{Ent}_{\mu^{x_S}}(f)) \qquad = \sum_{x_S\in\Omega(\nu_S)} \mu(x_S)\mathrm{Ent}_{\mu^{x_S}}(f).$$

Note that for all $\sigma \in \Omega$ such that $\sigma_S \in \Omega(\mu_S) \setminus \Omega(\nu_S)$, it holds that $f(\sigma) = \frac{\nu(\sigma)}{\mu(\sigma)} = 0$, implying that $\mathrm{Ent}_{\mu^{\sigma_S}}(f) = 0$. We have

$$\begin{aligned}\mathbf{E}_{x_S \sim \nu_S} D_{\mathrm{KL}}\left(\nu^{x_S} \parallel \mu^{x_S}\right) &= \sum_{x_S \in \Omega(\nu_S)} \mu(x_S)\mathrm{Ent}_{\mu^{x_S}}(f) + \sum_{x_S \in \Omega(\mu_S)\setminus\Omega(\nu_S)} \mu(x_S)\mathrm{Ent}_{\mu^{x_S}}(f) \\ &= \sum_{x_S \in \Omega(\mu_S)} \mu(x_S)\mathrm{Ent}_{\mu^{x_S}}(f) = \mu[\mathrm{Ent}_{V-S}(f)]. \end{aligned}$$

□

Now we are ready to prove Lemma 7.4.

Proof of Lemma 7.4. For any $e \in E$, let $E - e = E \setminus \{e\}$, using Lemma 7.6, it holds that

$$D_{\mathrm{KL}}\left(\nu \parallel \pi_{\mathrm{wrc}}\right) = D_{\mathrm{KL}}\left(\nu_{E-e} \parallel \pi_{\mathrm{wrc},E-e}\right) + \pi_{\mathrm{wrc}}[\mathrm{Ent}_e(f)], \quad \text{where } f(\tau) = \frac{\nu(\tau)}{\pi_{\mathrm{wrc}}(\tau)}.$$

Averaging over all $e \in E$, we get

$$\begin{aligned}\frac{1}{m}\sum_{e\in E} \pi_{\mathrm{wrc}}[\mathrm{Ent}_e(f)] &= \frac{1}{m}\sum_{e\in E} D_{\mathrm{KL}}\left(\nu \parallel \pi_{\mathrm{wrc}}\right) - \frac{1}{m}\sum_{e\in E} D_{\mathrm{KL}}\left(\nu_{E-e} \parallel \pi_{\mathrm{wrc},E-e}\right) \\ &= D_{\mathrm{KL}}\left(\nu \parallel \pi_{\mathrm{wrc}}\right) - D_{\mathrm{KL}}\left(\nu P^{\downarrow}_{\mathrm{GlauberRC}} \parallel \pi_{\mathrm{wrc}} P^{\downarrow}_{\mathrm{GlauberRC}}\right).\end{aligned}$$

By the assumption of Lemma 7.4, we have

$$\frac{1}{m}\sum_{e\in E} \pi_{\mathrm{wrc}}[\mathrm{Ent}_e(f)] \geq \delta D_{\mathrm{KL}}\left(\nu \parallel \pi_{\mathrm{wrc}}\right). \tag{38}$$

Next, by (36), we have

$$\begin{aligned}D_{\mathrm{KL}}\left(\nu P^{\downarrow}_{\mathrm{SB}} \parallel \pi_{\mathrm{wrc}} P^{\downarrow}_{\mathrm{SB}}\right) &= D_{\mathrm{KL}}\left(\nu P_{\mathcal{R}\to\mathrm{FKES}}\left(\frac{1}{m}\sum_{e\in E} P^{\downarrow}_e\right) \parallel \pi_{\mathrm{wrc}} P_{\mathcal{R}\to\mathrm{FKES}}\left(\frac{1}{m}\sum_{e\in E} P^{\downarrow}_e\right)\right) \\ &= D_{\mathrm{KL}}\left(\nu_{\mathrm{joint}}\left(\frac{1}{m}\sum_{e\in E} P^{\downarrow}_e\right) \parallel \pi_{\mathrm{FKES}}\left(\frac{1}{m}\sum_{e\in E} P^{\downarrow}_e\right)\right),\end{aligned}$$

where $\nu_{\mathrm{joint}} = \nu P_{\mathcal{R}\to\mathrm{FKES}}$ so that for any $\sigma\tau \in \Omega_{\mathrm{FKES}}$, $\nu_{\mathrm{joint}}(\sigma\tau) = \nu(\tau)\pi^{\tau}_{\mathrm{FKES},V}(\sigma)$. Hence, we have

$$D_{\mathrm{KL}}\left(\nu_{\mathrm{joint}} \parallel \pi_{\mathrm{FKES}}\right) = \sum_{\sigma\tau\in\Omega_{\mathrm{FKES}}} \nu_{\mathrm{joint}}(\sigma\tau)\log\frac{\nu(\tau)\pi^{\tau}_{\mathrm{FKES},V}(\sigma)}{\pi_{\mathrm{wrc}}(\tau)\pi^{\tau}_{\mathrm{FKES},V}(\sigma)} = D_{\mathrm{KL}}\left(\nu \parallel \pi_{\mathrm{wrc}}\right).$$

With these two equations, our goal, (37), is equivalent to

$$D_{\mathrm{KL}}\left(\nu_{\mathrm{joint}} \parallel \pi_{\mathrm{FKES}}\right) - D_{\mathrm{KL}}\left(\nu_{\mathrm{joint}}\left(\frac{1}{m}\sum_{e\in E} P^{\downarrow}_e\right) \parallel \pi_{\mathrm{FKES}}\left(\frac{1}{m}\sum_{e\in E} P^{\downarrow}_e\right)\right) \geq \frac{\delta}{4} D_{\mathrm{KL}}\left(\nu_{\mathrm{joint}} \parallel \pi_{\mathrm{FKES}}\right). \tag{39}$$

Using Lemma 7.6, for any $e \in E$, let $V + E - e$ be $V \cup E \setminus \{e\}$, it holds that

$$D_{\mathrm{KL}}\left(\nu_{\mathrm{joint}} \parallel \pi_{\mathrm{FKES}}\right) = D_{\mathrm{KL}}\left(\nu_{\mathrm{joint},V+E-e} \parallel \pi_{\mathrm{FKES},V+E-e}\right) + \pi_{\mathrm{FKES}}\left[\mathrm{Ent}_e\left(\overline{f}\right)\right],$$

where

$$\overline{f}(\sigma\tau) = \frac{\nu_{\mathrm{joint}}(\sigma\tau)}{\pi_{\mathrm{FKES}}(\sigma\tau)} = \frac{\nu(\tau)\pi^{\tau}_{\mathrm{FKES},V}(\sigma)}{\pi_{\mathrm{wrc}}(\tau)\pi^{\tau}_{\mathrm{FKES},V}(\sigma)} = \frac{\nu(\tau)}{\pi_{\mathrm{wrc}}(\tau)} = f(\tau).$$

Hence, (39) is equivalent to

$$\frac{1}{m}\sum_{e\in E} \pi_{\mathrm{FKES}}\left[\mathrm{Ent}_e\left(\overline{f}\right)\right] \geq \frac{\delta}{4} D_{\mathrm{KL}}\left(\nu_{\mathrm{joint}} \parallel \pi_{\mathrm{FKES}}\right) = \frac{\delta}{4} D_{\mathrm{KL}}\left(\nu \parallel \pi_{\mathrm{wrc}}\right).$$

Given (38), to prove the above inequality, it suffices to show that for any $e \in E$,

$$4 \cdot \pi_{\mathrm{FKES}}\left[\mathrm{Ent}_e\left(\overline{f}\right)\right] \geq \pi_{\mathrm{wrc}}[\mathrm{Ent}_e(f)]. \tag{40}$$

We now prove (40). We use σ to denote the vertex configuration in $\{0,1\}^V$ and τ to denote the edge configuration $\tau \in \{0,1\}^E$. Suppose $e = \{u, v\}$. We use τ_{-e} to denote a configuration in $\{0,1\}^{E-e}$. To ease the notation, we use $\pi_{\mathrm{FKES}}(\sigma\tau_{-e})$ to denote $\pi_{\mathrm{FKES},E-e}(\sigma\tau_{-e})$. For any $\tau_e \in \{0,1\}$, we use $\tau_{-e}\tau_e$ to denote a full configuration τ in $\{0,1\}^E$. We have

$$\begin{aligned}\pi_{\mathrm{FKES}}\left[\mathrm{Ent}_e\left(\overline{f}\right)\right] &= \sum_{\sigma\tau_{-e}} \pi_{\mathrm{FKES}}(\sigma\tau_{-e})\mathrm{Ent}_{\pi_{\mathrm{FKES}}^{\sigma\tau_{-e}}}\left(\overline{f}\right)\\ = \sum_{\sigma\tau_{-e}} \pi_{\mathrm{FKES}}(\sigma\tau_{-e}) &\sum_{\tau_e\in\{0,1\}} \pi_{\mathrm{FKES},e}^{\sigma\tau_{-e}}(\tau_e)\overline{f}(\sigma\tau_{-e}\tau_e)\log\frac{\overline{f}(\sigma\tau_{-e}\tau_e)}{\sum_{\tau_e\in\{0,1\}}\pi_{\mathrm{FKES},e}^{\sigma\tau_{-e}}(\tau_e)\overline{f}(\sigma\tau_{-e}\tau_e)}.\end{aligned}$$

If $\sigma_u \neq \sigma_v$, then $\pi_{\mathrm{FKES},e}^{\sigma\tau_{-e}}(0) = 1$, and in this case

$$\sum_{\tau_e\in\{0,1\}} \pi_{\mathrm{FKES},e}^{\sigma\tau_{-e}}(\tau_e)\overline{f}(\sigma\tau_{-e}\tau_e)\log\frac{\overline{f}(\sigma\tau_{-e}\tau_e)}{\sum_{\tau_e\in\{0,1\}}\pi_{\mathrm{FKES},e}^{\sigma\tau_{-e}}(\tau_e)\overline{f}(\sigma\tau_{-e}\tau_e)} = 0.$$

Thus we only need to consider the case where the two endpoints of e get the same spin. Note that this always happens if $\tau_{-e} \in C_e$, where $C_e \subseteq \{0,1\}^{E-e}$ is the set of τ_{-e} such that u and v are connected by edges assigned 1 in τ_{-e}. Again, to ease the notation, let $\pi_{\mathrm{wrc}}(\tau_{-e})$ be $\pi_{\mathrm{wrc},E-e}(\tau_{-e})$. Hence, we have

$$\begin{aligned}\pi_{\mathrm{FKES}}\left[\mathrm{Ent}_e\left(\overline{f}\right)\right] &= \sum_{\tau_{-e}\in C_e} \pi_{\mathrm{wrc}}(\tau_{-e})h(p_e,\tau_{-e}) + \sum_{\tau_{-e}\notin C_e} \pi_{\mathrm{wrc}}(\tau_{-e})\,\mathbf{Pr}_{\sigma\sim\pi_{\mathrm{FKES},V}^{\tau_{-e}}}[\sigma_u=\sigma_v]h(p_e,\tau_{-e})\\ &\geq \sum_{\tau_{-e}\in C_e} \pi_{\mathrm{wrc}}(\tau_{-e})h(p_e,\tau_{-e}) + \frac{1}{2}\sum_{\tau_{-e}\notin C_e} \pi_{\mathrm{wrc}}(\tau_{-e})h(p_e,\tau_{-e}),\end{aligned} \tag{41}$$

where

$$\begin{aligned}h(p_e,\tau_{-e}) \coloneqq\ & p_e f(\tau_{-e}1)\log f(\tau_{-e}1) + (1-p_e)f(\tau_{-e}0)\log f(\tau_{-e}0)\\ &- (p_e f(\tau_{-e}1) + (1-p_e)f(\tau_{-e}0))\log(p_e f(\tau_{-e}1) + (1-p_e)f(\tau_{-e}0)).\end{aligned}$$

(Recall that $\tau_{-e}\tau_e$ is a full configuration on E, where $\tau_e = 0$ or 1.) To see (41), since all external fields are consistent, $\mathbf{Pr}_{\sigma\sim\pi_{\mathrm{FKES},V}^{\tau_{-e}}}[\sigma_u = \sigma_v] \geq 1/2$. This is because we can further condition on τ_e: if $\tau_e = 1$, then $\sigma_u = \sigma_v$ with probability 1, and if $\tau_e = 0$, then σ_u and σ_v are independent and biased towards the same direction, in which case they are equal with probability at least $1/2$. The final probability is a linear combination of the two cases.

Similarly, we can expand the right hand side of (40),

$$\begin{aligned}\pi_{\mathrm{wrc}}[\mathrm{Ent}_e(f)] &= \sum_{\tau_{-e}} \pi_{\mathrm{wrc}}(\tau_{-e})\mathrm{Ent}_{\pi_{\mathrm{wrc}}^{\tau_{-e}}}(f)\\ &= \sum_{\tau_{-e}} \pi_{\mathrm{wrc}}(\tau_{-e}) \sum_{\tau_e\in\{0,1\}} \pi_{\mathrm{wrc},e}^{\tau_{-e}}(\tau_e)f(\tau_{-e}\tau_e)\log\frac{f(\tau_{-e}\tau_e)}{\sum_{\tau_e\in\{0,1\}}\pi_{\mathrm{wrc},e}^{\tau_{-e}}(\tau_e)f(\tau_{-e}\tau_e)}\\ &= \sum_{\tau_{-e}\in C_e} \pi_{\mathrm{wrc}}(\tau_{-e})h(p_e,\tau_{-e}) + \sum_{\tau_{-e}\notin C_e} \pi_{\mathrm{wrc}}(\tau_{-e})h\left(\frac{p_e}{1-\alpha(\tau_{-e})(p_e-1)},\tau_{-e}\right).\end{aligned} \tag{42}$$

In the last step above we use $\frac{p_e}{1-\alpha(\tau_{-e})(p_e-1)} = \pi_{\mathrm{wrc},e}^{\tau_{-e}}(1)$ where $\alpha(\tau_{-e})$ is a factor depending on τ_{-e}, derived as follows. Suppose $e = \{u, v\}$. Consider the random cluster configuration with e set not to be taken, and adding e causes the two connected components C_1 and C_2 to be merged as one, where u is in C_1 and v is in C_2. Let $X = X(\tau_{-e}) = \prod_{w\in C_1}\lambda_w$ and $Y = Y(\tau_{-e}) = \prod_{w\in C_2}\lambda_w$. We have

$$\pi_{\mathrm{wrc},e}^{\tau_{-e}}(1) = \frac{p_e(1+XY)}{p_e(1+XY)+(1-p_e)(1+X)(1+Y)} = \frac{p_e}{1-\frac{X+Y}{1+XY}(p_e-1)},$$

which means we can take $\alpha(\tau_{-e}) = \frac{X+Y}{1+XY}$. Moreover, we have $0 \leq \alpha(\tau_{-e}) \leq 1$ since $0 < X \leq 1$ and $0 < Y \leq 1$.

To finish the proof, define the following functions

$$t(x,p,\alpha) := \frac{g(x,p)}{g(x,p/(1-\alpha(p-1)))} \qquad \text{where} \qquad g(x,p) := px\log x - (px+1-p)\log(px+1-p)$$

for $0 \le p \le 1$ and $0 \le \alpha \le 1$. Define $t(0,p,\alpha) := \lim_{x\downarrow 0} t(x,p,\alpha)$ and $t(1,p,\alpha) := \lim_{x\to 1} t(x,p,\alpha)$. It is not hard to verify that $t(x,p,\alpha)$ is continuous with respect to x over $[0,\infty)$ for any fixed p and α, and $t\left(\frac{f(\tau_{-e},1)}{f(\tau_{-e},0)}, p_e, \alpha(\tau_{-e})\right) = \frac{h(p_e,\tau_{-e})}{h\left(\frac{p_e}{1-\alpha(\tau_{-e})(p_e-1)},\tau_{-e}\right)}$. This function admits the following monotonicity property, whose proof is postponed till Appendix C.

Lemma 7.7. *For any $0 \le p \le 1$ and $0 \le \alpha \le 1$, $t(x,p,\alpha)$ is monotone decreasing in x over $x \ge 0$.*

Given this, $t(x,p,\alpha)$ has a lower bound

$$t(x,p,\alpha) \ge \lim_{x\to\infty} t(x,p,\alpha) = \frac{(1-\alpha(p-1))\log p}{\log p - \log(1-\alpha(p-1))} =: C(p,\alpha).$$

We remark that the constant $C = C(p,\alpha)$ satisfies

$$0.5 \le C(p,\alpha) \le 2. \tag{43}$$

The proof is given in Appendix C, too. Using this fact, we conclude (40) by comparing (41) with (42).

This finishes the proof of Lemma 7.4. □

8. Perfect sampling via coupling from the past

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

Theorem 8.1. *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 $0 < \lambda_v \le 1$ for all $v \in V$, the algorithm returns a perfect sample from weighted random cluster models in expected time $C_1(p_{\min}, p_{\max})N^4m^4 \log n$, where $N = \min\left\{n, \frac{1}{1-\lambda_{\max}}\right\}$, $\lambda_{\max} = \max_{v\in V}\lambda_v$, $C_1(p_{\min}, p_{\max}) = O\left(\frac{1}{\min\{p_{\min},1-p_{\max}\}}\log\frac{1}{\min\{p_{\min},1-p_{\max}\}}\right)$, $p_{\max} = \max_{e\in E} p_e$ and $p_{\min} = \min_{e\in E} p_e$.*

Furthermore, if there exists $\delta > 0$ such that $\lambda_v \le 1-\delta$ for all $v \in V$, then the algorithm runs in time $C_2(\Delta,\delta,p_{\min},p_{\max})n^2\log^2 n$, where $C_2(\Delta,\delta,p_{\min},p_{\max}) = \left(\frac{\Delta}{\delta^2\min\{p_{\min},1-p_{\max}\}}\right)^{O\left(\frac{\Delta^2}{\delta^4\min\{p_{\min},1-p_{\max}\}}\right)}$.

Note that if $p_e = 0$, we can simply remove e, and if $p_e = 1$, we can contract e. Similarly if $\lambda_v = 0$, we may pin v to 0 and absorb it into its neighbours external fields. Thus for any weighted random cluster model, we can modify it so that it satisfies the condition of Theorem 8.1.

8.1. Perfect ferromagnetic Ising sampler.

We now prove Theorem 1.3. We give the perfect ferromagnetic Ising sampler assuming the algorithm in Theorem 8.1. 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 8.1 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 $\mathcal{S} \sim \pi_{\text{wrc}}$, σ is a perfect sample from the Ising model. The running time of the transformation step is $O(n+m)$. Note that

$$p_{\min} = 1 - \frac{1}{\beta_{\min}}, \quad 1 - p_{\max} = \frac{1}{\beta_{\max}}.$$

By Theorem 8.1, the total running time is $C_1 N^4 m^4 \log n$ and $C_2 n^2 \log^2 n$ for all $\lambda_v \le 1-\delta$, where

$$
\begin{aligned}
C_1 = C_1(\beta_{\min}, \beta_{\max}) &= O\left(\left(\beta_{\max} + \frac{\beta_{\min}}{\beta_{\min}-1}\right)\log\left(\beta_{\max} + \frac{\beta_{\min}}{\beta_{\min}-1}\right)\right), \\
C_2 = C_2(\Delta, \delta, \beta_{\min}, \beta_{\max}) &= \left(\frac{\Delta}{\delta^2}\left(\beta_{\max} + \frac{\beta_{\min}}{\beta_{\min}-1}\right)\right)^{O\left(\frac{\Delta^2}{\delta^4}\left(\beta_{\max}+\frac{\beta_{\min}}{\beta_{\min}-1}\right)\right)}.
\end{aligned} \tag{44}
$$

8.2. **CFTP for weighted random cluster models.** We give a perfect sampler for weighted random cluster models based on coupling form the past (CFTP) applied to the Glauber dynamics. Here is an equivalent definition of the Glauber dynamics. 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 Glauber dynamics starts from an arbitrary subset of edges $X_0 \in \{0,1\}^E$. For the t-th transition step, the chain does the following:

- pick an edge $e_t \in E$ uniformly at random;
- sample a real number $r_t \in [0,1]$ uniformly at random; if $r_t < a_t$, let $X_t = X_{t-1}^{e\leftarrow 1}$; if $r_t \ge a_t$, let $X_t = X_{t-1}^{e\leftarrow 0}$, where $X_{t-1}^{e\leftarrow c}$ satisfies $X_{t-1}^{e\leftarrow c}(E\setminus\{e\}) = X_{t-1}(E\setminus\{e\})$ and $X_{t-1}^{e\leftarrow c}(e) = c$, and

$$
a_t = a(X_{t-1}, e) := \frac{\pi_{\mathrm{wrc}}(X_{t-1}^{e\leftarrow 1})}{\pi_{\mathrm{wrc}}(X_{t-1}^{e\leftarrow 0}) + \pi_{\mathrm{wrc}}(X_{t-1}^{e\leftarrow 1})}. \tag{45}
$$

The Glauber dynamics for weighted random cluster models admits a *grand monotone coupling*. Let $\Omega = \{0,1\}^E$. Let $P : \Omega\times\Omega \to \mathbb{R}_{\ge 0}$ denote the transition matrix of the Glauber 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,

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

where $\mathcal{D}$ is a distribution such that e_t is a uniform random edge in E, r_t is a uniform random real number in $[0,1]$, and they are 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 $X, Y \in \{0,1\}^E$,

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

Let $X^{\min} = \mathbf{0}$ be the constant 0 vector and $X^{\max} = \mathbf{1}$ be the constant 1 vector, so 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 8.2. *Suppose $0 \le p_e < 1$ for all $e \in E$ and $0 < \lambda_v \le 1$ for all $v \in V$. The grand coupling φ of the Glauber 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 8.2 is deferred to Section 8.3. With the monotone grand coupling φ, we apply CFTP to the Glauber dynamics for weighted random cluster models in Algorithm 1.

Remark 8.3. 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 $(U_t)_{-2T\le t<-T}$ 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 $T_{\mathrm{mix}}(\cdot)$ 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.

Algorithm 1: CFTP of the Glauber 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 $0 < \lambda_v \le 1$ for all $v \in V$.

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

generate $U_t \sim \mathcal{D}$ independently for all integers $t \in (-\infty, -1]$;
$T = 1$;
repeat
  $X^{\min} = \mathbf{0}$ and $X^{\max} = \mathbf{1}$;
  for $t = -T$ *to* -1 **do**
    $X^{\min} \leftarrow \varphi(X^{\min}, U_t)$;
    $X^{\max} \leftarrow \varphi(X^{\max}, U_t)$;
    `// φ is the monotone grand coupling in Lemma 8.2`
  $T \leftarrow 2T$
until $X^{\min} = X^{\max}$;
return $X^{\min}$;

Proposition 8.4 ([PW96]). *Suppose the input weighted random cluster model satisfies $0 \le p_e < 1$ for all $e \in E$ and $0 < \lambda_v \le 1$ 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 $O((T_{\mathcal{D}} + T_{\varphi})T_{\mathrm{mix}}(\frac{1}{4e}) \log n)$.*

Now, we are ready to prove Theorem 8.1.

Proof of Theorem 8.1. By definitions of $\mathcal{D}$ and φ, it is straightforward to verify that $T_{\mathcal{D}} = O(1)$ and $T_{\varphi} = O(n + m)$. The mixing time can be obtained from Theorem 6.3. □

8.3. **Proof of monotonicity.** Here we prove Lemma 8.2. Fix $\sigma, \tau \in \{0,1\}^E$ such that $\sigma \preceq \tau$. Fix $U = (e, r) \in \Omega_R$. Let $e = \{u, v\}$. Let σ_{-e} and τ_{-e} denote $\sigma(E \setminus \{e\})$ and $\tau(E \setminus \{e\})$ respectively, and G_σ and G_τ be the graphs with vertices V and edges in σ_{-e} and τ_{-e} respectively. Note that G_σ is a subgraph of G_τ. We prove the lemma by considering three cases (1) u, v are connected in both G_σ and G_τ (2) u, v are neither connected in neither G_σ nor G_τ (3) u, v are connected in G_τ but not in G_σ.

First suppose u, v are connected in both G_σ and G_τ. In this case $a(\sigma, e) = a(\tau, e) = p_e$, where $a(\cdot, \cdot)$ is defined in (45). The lemma holds trivially.

Next assume u, v are neither connected in neither G_σ nor G_τ. Suppose u, v belong to connected components C_1, C_2 (or C_1', C_2') in G_σ (or G_τ) respectively. Define

$$x_1^\sigma := \prod_{w\in C_1} \lambda_w, \quad x_2^\sigma := \prod_{w\in C_2} \lambda_w, \quad x_1^\tau := \prod_{w\in C_1'} \lambda_w, \quad x_2^\tau := \prod_{w\in C_2'} \lambda_w.$$

We have

$$a(\sigma, e) = \frac{p_e(1 + x_1^\sigma x_2^\sigma)}{p_e(1 + x_1^\sigma x_2^\sigma) + (1 - p_e)(1 + x_1^\sigma)(1 + x_2^\sigma)},$$
$$a(\tau, e) = \frac{p_e(1 + x_1^\tau x_2^\tau)}{p_e(1 + x_1^\tau x_2^\tau) + (1 - p_e)(1 + x_1^\tau)(1 + x_2^\tau)}.$$

Since $\lambda_w \le 1$ for all $w \in V$, $x_1^\sigma \ge x_1^\tau$ and $x_2^\sigma \ge x_2^\tau$, which implies

$$\frac{(1 + x_1^\sigma)(1 + x_2^\sigma)}{(1 + x_1^\sigma x_2^\sigma)} \ge \frac{(1 + x_1^\tau)(1 + x_2^\tau)}{(1 + x_1^\tau x_2^\tau)}.$$

Hence $a(\sigma, e) \le a(\tau, e)$, which implies the lemma.

Lastly suppose u, v are connected in G_τ but not in G_σ. Suppose u, v belong to connected components C_1, C_2 in G_σ. Define x_1^σ and x_2^σ in the same way.

$$a(\sigma, e) = \frac{p_e(1 + x_1^\sigma x_2^\sigma)}{p_e(1 + x_1^\sigma x_2^\sigma) + (1 - p_e)(1 + x_1^\sigma)(1 + x_2^\sigma)}, \quad a(\tau, e) = p_e.$$

Since $(1 + x_1^\sigma)(1 + x_2^\sigma) \geq 1 + x_1^\sigma x_2^\sigma$, $a(\sigma, e) \leq a(\tau, e)$, which implies the lemma.

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.

[ALO20] Nima Anari, Kuikui Liu, and Shayan Oveis Gharan. Spectral independence in high-dimensional expanders and applications to the hardcore model. In *FOCS*, pages 1319–1330. IEEE, 2020.

[ALOV19] Nima Anari, Kuikui Liu, Shayan Oveis Gharan, and Cynthia Vinzant. Log-concave polynomials II: high-dimensional walks and an FPRAS for counting bases of a matroid. In Moses Charikar and Edith Cohen, editors, *STOC*, pages 1–12. ACM, 2019.

[BCC+22] Antonio Blanca, Pietro Caputo, Zongchen Chen, Daniel Parisi, Daniel Štefankovič, and Eric Vigoda. On mixing of Markov chains: Coupling, spectral independence, and entropy factorization. In *SODA*, pages 3670–3692. SIAM, 2022.

[BCP+21] Antonio Blanca, Pietro Caputo, Daniel Parisi, Alistair Sinclair, and Eric Vigoda. Entropy decay in the Swendsen-Wang dynamics on $\mathbb{Z}^d$. In *STOC*, pages 1551–1564. ACM, 2021.

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

[CFYZ21] Xiaoyu Chen, Weiming Feng, Yitong Yin, and Xinyuan Zhang. Rapid mixing of glauber dynamics via spectral independence for all degrees. In *62nd IEEE Annual Symposium on Foundations of Computer Science, FOCS 2021, Denver, CO, USA, February 7-10, 2022*, pages 137–148. IEEE, 2021.

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

[CGM21] Mary Cryan, Heng Guo, and Giorgos Mousa. Modified log-Sobolev inequalities for strongly log-concave distributions. *Ann. Probab.*, 49(1):506–525, 2021.

[CGŠV21] Zongchen Chen, Andreas Galanis, Daniel Štefankovič, and Eric Vigoda. Rapid mixing for colorings via spectral independence. In *SODA*, pages 1548–1557. SIAM, 2021.

[CLV21a] 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.

[CLV21b] Zongchen Chen, Kuikui Liu, and Eric Vigoda. Spectral independence via stability and applications to Holant-type problems. In *FOCS*, pages 149–160. IEEE, 2021.

[CLX11] Jin-Yi Cai, Pinyan Lu, and Mingji Xia. Computational complexity of Holant problems. *SIAM J. Comput.*, 40(4):1101–1132, 2011.

[CP21] Pietro Caputo and Daniel Parisi. Block factorization of the relative entropy via spatial mixing. *Comm. Math. Phys.*, 388(2):793–818, 2021.

[CZ23] Xiaoyu Chen and Xinyuan Zhang. A near-linear time sampler for the ising model. In *SODA*. SIAM, 2023. to appear.

[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.

[FGYZ21] Weiming Feng, Heng Guo, Yitong Yin, and Chihao Zhang. Rapid mixing from spectral independence beyond the Boolean domain. In *SODA*, pages 1558–1577. SIAM, 2021.

[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.

[LNNP14] Yun Long, Asaf Nachmias, Weiyang Ning, and Yuval Peres. A power law of order 1/4 for critical mean field Swendsen-Wang dynamics. *Mem. Amer. Math. Soc.*, 232(1092):vi+84, 2014.

[LP17] David A. Levin and Yuval Peres. *Markov chains and mixing times*. American Mathematical Society, Providence, RI, 2017. Second edition of [MR2466937], With contributions by Elizabeth L. Wilmer, With a chapter on "Coupling from the past" by James G. Propp and David B. Wilson.

[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

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

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

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

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

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

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

A.2. **Equivalence between Ising and subgraph-world.** To apply Theorem 2.6, 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 further more, 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

$$\text{Holant}(H; \mathcal{F}_{\text{Ising}} \mid \mathcal{G}_{\text{Ising}}) = Z_{\text{Ising}}(G; \boldsymbol{\beta}, \boldsymbol{\lambda}). \tag{46}$$

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

$$\text{Holant}(H; \mathcal{F}_{\text{sg}} \mid \mathcal{G}_{\text{sg}}) = Z_{\text{sg}}(G; \boldsymbol{p}, \boldsymbol{\eta}). \tag{47}$$

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 2.6, (46) and (47) 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. Proof of the adjointness

Proof of Proposition 2.3. 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

$$\begin{aligned}
\left(D_{\text{Ising}} P_{\mathcal{I} \to \mathcal{R}}\right)(\sigma, S) &= \mathbb{I}[S \subseteq M(\sigma)] \cdot \pi_{\text{Ising}}(\sigma) \cdot \prod_{e \in S} \left(1 - \frac{1}{\beta_e}\right) \prod_{f \in M(\sigma) \setminus S} \frac{1}{\beta_f} \\
&= \mathbb{I}[S \subseteq M(\sigma)] \cdot \frac{1}{Z_{\text{Ising}}} \cdot \prod_{v \in V} \lambda_v^{\sigma(v)} \prod_{h \in M(\sigma)} \beta_h \prod_{e \in S} \left(1 - \frac{1}{\beta_e}\right) \prod_{f \in M(\sigma) \setminus S} \frac{1}{\beta_f} \\
&= \mathbb{I}[S \subseteq M(\sigma)] \cdot \frac{1}{Z_{\text{Ising}}} \cdot \prod_{v \in V} \lambda_v^{\sigma(v)} \prod_{e \in S} (\beta_e - 1). \qquad (48)
\end{aligned}$$

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

$$\begin{aligned}
\left(P_{\mathcal{R} \to \mathcal{I}}^T D_{\text{wrc}}\right)(\sigma, S) &= \mathbb{I}[S \subseteq M(\sigma)] \cdot \pi_{\text{wrc}}(S) \cdot \prod_{C \in \kappa(V,S)} \frac{\prod_{v \in C} \lambda_v^{\sigma(v)}}{1 + \prod_{v \in C} \lambda_v} \\
&= \mathbb{I}[S \subseteq M(\sigma)] \cdot \frac{1}{Z_{\text{wrc}}} \cdot \prod_{e \in S} \left(1 - \frac{1}{\beta_e}\right) \prod_{f \in E \setminus S} \frac{1}{\beta_f} \prod_{C \in \kappa(V,S)} \left(1 + \prod_{u \in C} \lambda_u\right) \cdot \prod_{C \in \kappa(V,S)} \frac{\prod_{v \in C} \lambda_v^{\sigma(v)}}{1 + \prod_{v \in C} \lambda_v} \\
&= \mathbb{I}[S \subseteq M(\sigma)] \cdot \frac{1}{Z_{\text{wrc}}} \cdot \prod_{e \in S} \left(1 - \frac{1}{\beta_e}\right) \prod_{f \in E \setminus S} \frac{1}{\beta_f} \prod_{v \in V} \lambda_v^{\sigma(v)} \\
&= \mathbb{I}[S \subseteq M(\sigma)] \cdot \frac{1}{Z_{\text{wrc}}} \cdot \prod_{h \in E} \frac{1}{\beta_h} \prod_{v \in V} \lambda_v^{\sigma(v)} \prod_{e \in S} (\beta_e - 1) \qquad (49)
\end{aligned}$$

By Proposition 2.1, we know that

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

Combining above equation with (48) and (49) proves $\left(D_{\text{Ising}}P_{\mathcal{I}\to\mathcal{R}}\right)(\sigma,S) = \left(P^T_{\mathcal{R}\to\mathcal{I}}D_{\text{wrc}}\right)(\sigma,S)$. □

Appendix C. Proof of analytic lemmata

This section of appendix proves Lemma 7.7 and (43).

Proof of Lemma 7.7. The goal is to show $\partial t(x,p,\alpha)/\partial x < 0$ for all $x \in (0,1)\cup(1,+\infty)$. The lemma then follows by combining this with continuity.

A straightforward calculation shows that

$$\frac{\partial t(x,p,\alpha)}{\partial x} = \frac{-(1-\alpha(1-p))(1-p)p}{\left(xp\log x - ((1+\alpha)(1-p)+px)\log\left(1+\frac{p(x-1)}{1+\alpha(1-p)}\right)\right)^2} s(x,p,\alpha)$$

where

$$s(x,p,\alpha) := (1+\alpha)(\log x)\log\left(1+\frac{p(x-1)}{1+\alpha(1-p)}\right) - \left(\log x + \alpha\log\left(1+\frac{p(x-1)}{1+\alpha(1-p)}\right)\right)\log(1+p(x-1)).$$

This means $\operatorname{sgn}(\partial t(x,p,\alpha)/\partial x) = -\operatorname{sgn}(s(x,p,\alpha))$, and hence we only need to show $s(x,p,\alpha) > 0$ whenever $x \in (0,1)\cup(1,+\infty)$.

From now on in this section, we use the notation $A \gtrless_x B$ to represent that $A > B$ when $x > 1$, and $A < B$ when $0 < x < 1$. In other words, when $x > 1$, $\gtrless_x$ should be read as $>$, and vice versa.

We first claim the following inequalities:

$$(1+\alpha)\log x - \alpha\log(1+p(x-1)) \gtrless_x 0; \tag{50}$$

$$\log\left(1+\frac{p(x-1)}{1+\alpha(1-p)}\right) \gtrless_x 0; \tag{51}$$

$$\log(1+(x-1)p) \gtrless_x 0. \tag{52}$$

We focus on $s(x,p,\alpha)$ and postpone the proof of these simple inequalities till the end. By collecting terms of $\log\left(1+\frac{p(x-1)}{1+\alpha(1-p)}\right)$, one can find that $s(x,p,\alpha) > 0$ if and only if

$$((1+\alpha)\log x - \alpha\log(1+p(x-1)))\log\left(1+\frac{p(x-1)}{1+\alpha(1-p)}\right) > (\log x)\log(1+p(x-1)).$$

By using (50), it is equivalent to show that

$$\log\left(1+\frac{p(x-1)}{1+\alpha(1-p)}\right) \gtrless_x \frac{(\log x)\log(1+p(x-1))}{(1+\alpha)\log x - \alpha\log(1+p(x-1))},$$

or equivalently, using (50)(51)(52), to show that

$$\frac{1}{\log(1+(x-1)p)} \gtrless_x \frac{\alpha}{1+\alpha}\cdot\frac{1}{\log x} + \frac{1}{1+\alpha}\cdot\frac{1}{\log\left(1+\frac{p(x-1)}{1+\alpha(1-p)}\right)}. \tag{53}$$

Note that the following function

$$u_{x,p}(y) := \frac{1}{\log\left(1+\frac{p(x-1)}{y}\right)}$$

reveals the essence of (53) in the way that (53) is equivalent to

$$u_{x,p}(1) \gtrless_x \frac{\alpha}{1+\alpha}\cdot u_{x,p}(p) + \frac{1}{1+\alpha}\cdot u_{x,p}(1-\alpha(p-1)), \tag{54}$$

and note that

$$1 = \frac{\alpha}{1+\alpha} \cdot p + \frac{1}{1+\alpha} \cdot (1 - \alpha(p-1)).$$

This means (54) follows if for fixed $x > 1$ (resp., $0 < x < 1$) and p, $u_{x,p}(y)$ is a concave (resp., convex) function over $y \in (p, 2) \supseteq (p, 1-\alpha(p-1))$, which would conclude the proof. We verify this as follows.

A straightforward calculation shows that

$$\frac{\mathrm{d}^2}{\mathrm{d}y^2} u_{x,p}(y) = \frac{(x-1)p}{y(y+(x-1)p)^2 \log^3\left(1+\frac{p(x-1)}{y}\right)} \left(2 \cdot \frac{p(x-1)}{y} - \left(2 + \frac{p(x-1)}{y}\right) \log\left(1 + \frac{p(x-1)}{y}\right)\right).$$

It is not hard to verify that

$$\log\left(1 + \frac{p(x-1)}{y}\right) \gtrless_x 0, \tag{55}$$

which we prove later. With a bit more endeavour, we can also show that

$$-\left(2 \cdot \frac{p(x-1)}{y} - \left(2 + \frac{p(x-1)}{y}\right) \log\left(1 + \frac{p(x-1)}{y}\right)\right) \gtrless_x 0, \tag{56}$$

whose proof is postponed as well. Concavity/Convexity is then established by combining the expression for the second-order derivative, (55) and (56). □

Proof of (50), (51), (52), (55), and (56). For (50), because $\log x \gtrless_x 0$, we only need to show

$$\frac{x}{1+(x-1)p} \gtrless_x 1.$$

Note that $1 + (x-1)p$ is positive. The above is hence equivalent to

$$(x-1)(1-p) \gtrless_x 0,$$

which is obvious.

All of (51), (52) and (55), after simple calculation, are equivalent to $(x-1)p \gtrless_x 0$, which is obvious, too.

Finally, we show (56). Let $z := p(x-1)/y$. LHS is then $r(z) := (2+z)\log(1+z) - 2z$. It is not hard to show that $r(z)$ is monotone in z over $z \in (-1, +\infty)$, by observing that $r'(z) = \frac{1}{1+z} - 1 - \log\frac{1}{1+z}$, which is non-negative as $\log x \le x - 1$ for $x > 0$. Moreover, $r(0) = 0$. Therefore, when $x > 1$, we have $z > 0$, and (56) holds. When $0 < x < 1$, we have $-1 < (x-1) \le z < 0$, and (56) holds too. □

Proof of (43). For convenient reference, the expression of interest is

$$C(p, \alpha) := \frac{(1-\alpha(p-1))\log p}{\log p - \log(1-\alpha(p-1))}.$$

Taking derivative with respect to α, we get

$$\frac{\partial}{\partial \alpha} C(p,\alpha) = \frac{(1-p)\log(p)\left(1 + \log\left(\frac{p}{1+\alpha(1-p)}\right)\right)}{\left(\log\left(\frac{p}{1+\alpha(1-p)}\right)\right)^2}.$$

A simple calculation shows that

- if $p \le 1/e$, then $C(p,\alpha)$ is increasing with α, and hence lies between $C(p,0) = 1$ and $C(p,1) = \frac{(2-p)\log p}{\log p - \log(2-p)}$;
- if $1/e < p < 2/(1+e)$, then $C(p,\alpha)$ is decreasing within $\alpha \in (0, (ep-1)/(1-p))$ and increasing within $\alpha \in ((ep-1)/(1-p), 1)$, and hence lies between $C(p, (ep-1)/(1-p)) = -ep\log p \ge 2e\log((1+e)/2)/(1+e) > 0.90$ and $\max\{C(p,0), C(p,1)\}$; and
- if $p \ge 2/(1+e)$, then $C(p,\alpha)$ is decreasing, and hence lies between $C(p,1) = \frac{(2-p)\log p}{\log p - \log(2-p)}$ and $C(p,0) = 1$.

From the case-by-case analysis, it suffices to show that $0.5 \le \frac{(2-p)\log p}{\log p - \log(2-p)} \le 2$, which is a simple analytic exercise. □