Swendsen-Wang Dynamics for General Graphs in the Tree Uniqueness Region

Antonio Blanca
School of Computer Science, Georgia Institute of Technology, Atlanta GA 30332, USA
ablanca3@gatech.edu

Zongchen Chen
School of Mathematics, Georgia Institute of Technology, Atlanta GA 30332, USA
chenzongchen@gatech.edu

Eric Vigoda
School of Computer Science, Georgia Institute of Technology, Atlanta GA 30332, USA
vigoda@gatech.edu

Abstract

The Swendsen-Wang dynamics is a popular algorithm for sampling from the Gibbs distribution for the ferromagnetic Ising model on a graph $G = (V, E)$. The dynamics is a "global" Markov chain which is conjectured to converge to equilibrium in $O(|V|^{1/4})$ steps for any graph G at any (inverse) temperature β. It was recently proved by Guo and Jerrum (2017) that the Swendsen-Wang dynamics has polynomial mixing time on any graph at all temperatures, yet there are few results providing $o(|V|)$ upper bounds on its convergence time.

We prove fast convergence of the Swendsen-Wang dynamics on general graphs in the tree uniqueness region of the ferromagnetic Ising model. In particular, when $\beta < \beta_c(d)$ where $\beta_c(d)$ denotes the uniqueness/non-uniqueness threshold on infinite d-regular trees, we prove that the *relaxation time* (i.e., the inverse spectral gap) of the Swendsen-Wang dynamics is $\Theta(1)$ on any graph of maximum degree $d \geq 3$. Our proof utilizes a version of the Swendsen-Wang dynamics which only updates isolated vertices. We establish that this variant of the Swendsen-Wang dynamics has mixing time $O(\log |V|)$ and relaxation time $\Theta(1)$ on any graph of maximum degree d for all $\beta < \beta_c(d)$. We believe that this Markov chain may be of independent interest, as it is a *monotone* Swendsen-Wang type chain. As part of our proofs, we provide modest extensions of the technology of Mossel and Sly (2013) for analyzing mixing times and of the censoring result of Peres and Winkler (2013). Both of these results are for the Glauber dynamics, and we extend them here to general monotone Markov chains. This class of dynamics includes for example the *heat-bath block dynamics*, for which we obtain new tight mixing time bounds.

2012 ACM Subject Classification Theory of computation → Random walks and Markov chains, Theory of computation → Design and analysis of algorithms

Keywords and phrases Swendsen-Wang dynamics, mixing time, relaxation time, spatial mixing, censoring

Digital Object Identifier 10.4230/LIPIcs.APPROX-RANDOM.2018.32

Related Version A full version of the paper is available at `https://arxiv.org/pdf/1806.04602.pdf`.

Funding Research supported in part by NSF grants CCF-1617306 and CCF-1563838.

Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques (APPROX/RANDOM 2018).
Editors: Eric Blais, Klaus Jansen, José D. P. Rolim, and David Steurer; Article No. 32; pp. 32:1–32:18
Leibniz International Proceedings in Informatics
LIPICS Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

1 Introduction

For spin systems, sampling from the associated Gibbs distribution is a key computational task with a variety of applications, notably including inference/learning [19] and approximate counting [28, 48]. In the study of spin systems, a model of prominent interest is the Ising model. This is a classical model in statistical physics, which was introduced in the 1920's to study the ferromagnet and its physical phase transition [25, 31]. More recently, the Ising model has found numerous applications in theoretical computer science, computer vision, social network analysis, game theory, biology, discrete probability and many other fields [7, 17, 12, 13, 37].

An instance of the (ferromagnetic) Ising model is given by an undirected graph $G = (V, E)$ on $n = |V|$ vertices and an (inverse) temperature $\beta > 0$. A configuration $\sigma \in \{+, -\}^V$ assigns a spin value ($+$ or $-$) to each vertex $v \in V$. The probability of a configuration σ is proportional to

$$w(\sigma) = \exp\Big(\beta \sum_{\{v,w\}\in E} \sigma(v)\sigma(w)\Big), \tag{1}$$

where $\sigma(v)$ is the spin of v. The associated Gibbs distribution $\mu = \mu_{G,\beta}$ is given by $\mu(\sigma) = w(\sigma)/Z$, where the normalizing factor Z is known as the *partition function*. Since $\beta > 0$ the system is ferromagnetic as neighboring vertices prefer to align their spins.

For general graphs Jerrum and Sinclair [26] presented an FPRAS for the partition function (which yields an efficient sampler); however, its running time is a large polynomial in n. Hence, there is significant interest in obtaining tight bounds on the convergence rate of Markov chains for the Ising model, namely, Markov chains on the space of Ising configurations $\{+, -\}^V$ that converge to Gibbs distribution μ. A standard notion for measuring the speed of convergence to stationarity is the *mixing time*, which is defined as the number of steps until the Markov chain is close to its stationary distribution in total variation distance, starting from the worst possible initial configuration.

A simple, popular Markov chain for sampling from the Gibbs distribution is the Glauber dynamics, commonly referred to as the Gibbs sampler in some communities. This dynamics works by updating a randomly chosen vertex in each step in a reversible fashion. Significant progress has been made in understanding the mixing properties of the Glauber dynamics and its connections to the spatial mixing (i.e., decay of correlation) properties of the underlying spin system. In general, in the high-temperature region (small β) correlations typically decay exponentially fast, and one expects the Glauber dynamics to converge quickly to stationarity. For example, for the special case of the integer lattice $\mathbb{Z}^2$, in the high-temperature region it is well known that the Glauber dynamics has mixing time $\Theta(n \log n)$ [35, 6, 10]. For general graphs, Mossel and Sly [38] proved that the Glauber dynamics mixes in $O(n \log n)$ steps on any graph of maximum degree d in the tree uniqueness region. Tree uniqueness is defined as follows: let T_h denote a (finite) complete tree of height h (by complete we mean all internal vertices have degree d). Fix the leaves to be all $+$ spins, consider the resulting conditional Gibbs distribution on the internal vertices, and let p_h^+ denote the probability the root is assigned spin $+$ in this conditional distribution; similarly, let p_h^- denote the corresponding marginal probability with the leaves fixed to spin $-$. When $\beta < \beta_c(d)$, where $\beta_c(d)$ is such that

$$(d - 1) \tanh \beta_c(d) = 1, \tag{2}$$

then $p_\infty^+ = p_\infty^-$ and we say *tree uniqueness* holds since there is a unique Gibbs measure on the infinite d-regular tree [41]. In the same setting, building upon the approach of Weitz [52] for the hard-core model, Li, Lu and Yin [33] provide an FPTAS for the partition function, but the running time is a large polynomial in n.

In practice, it is appealing to utilize non-local (or global) chains which possibly update $\Omega(n)$ vertices in a step; these chains are more popular due to their presumed speed-up and for their ability to be naturally parallelized [30].

A notable example for the ferromagnetic Ising model is the Swendsen-Wang (SW) dynamics [49] which utilizes the random-cluster representation to derive an elegant Markov chain in which every vertex can change its spin in every step. The SW dynamics works in the following manner. From the current spin configuration $\sigma_t \in \{+,-\}^V$:

1. Consider the set of agreeing edges $E(\sigma_t) = \{(v,w) \in E : \sigma_t(v) = \sigma_t(w)\}$;
2. Independently for each edge $e \in E(\sigma_t)$, "percolate" by deleting e with probability $\exp(-2\beta)$ and keeping e with probability $1 - \exp(-2\beta)$; this yields $F_t \subseteq E(\sigma_t)$;
3. For each connected component C in the subgraph (V, F_t), choose a spin s_C uniformly at random from $\{+,-\}$, and then assign spin s_C to all vertices in C, yielding $\sigma_{t+1} \in \{+,-\}^V$.

The proof that the stationary distribution of the SW dynamics is the Gibbs distribution is non-trivial; see [11] for an elegant proof. The SW dynamics is also well-defined for the *ferromagnetic Potts model*, a natural generalization of the Ising model that allows vertices to be assigned q different spins.

The SW dynamics for the Ising model is quite appealing as it is conjectured to mix quickly at all temperatures.

Its behavior for the Potts model (which corresponds to $q > 2$ spins) is more subtle, as there are multiple examples of classes of graphs where the SW dynamics is torpidly mixing; i.e., mixing time is exponential in the number of vertices of the graph; see, e.g., [20, 15, 3, 18, 4, 5].

Despite the popularity [51, 42, 43] and rich mathematical structure [21] of the SW dynamics there are few results with tight bounds on its speed of convergence to equilibrium. In fact, there are few results proving the SW dynamics is faster than the Glauber dynamics (or the edge dynamics analog in the random-cluster representation). Most results derive as a consequence of analyses of these local dynamics. Recently, Guo and Jerrum [22] established that the mixing time of the SW dynamics on *any* graph and at any temperature is $O(|V|^{10})$.

This bound, however, is far from the conjectured universal upper bound of $O(|V|^{1/4})$ [39], and once again their result derives from a bound on a local chain (the edge dynamics in the random-cluster representation).

In the special case of the *mean-field* Ising model, which corresponds to the underlying graph G being the complete graph on n vertices, Long, Nachmias, Ning and Peres [34] provided a tight analysis of the mixing time of the SW dynamics. They prove that the mixing time of the mean-field SW dynamics is $\Theta(|V|^{1/4})$; this is expected to be the worst case and thus yields the aforementioned conjecture [39].

Another relevant case for which the speed of convergence is known is the two-dimensional integer lattice $\mathbb{Z}^2$ (more precisely, finite subsections of it). Blanca, Caputo, Sinclair and Vigoda [1] recently established that the *relaxation time* of the SW dynamics is $\Theta(1)$ in the high-temperature region. The relaxation time measures the speed of convergence to μ when the initial configuration is reasonably close to this distribution (a so-called "warm start") [27, 29]. More formally, the relaxation time is equal to the inverse spectral gap of the transition matrix of the chain and is another well-studied notion of rate of convergence [32]. This result [1] applied a well-established proof approach [35, 10] which utilizes that $\mathbb{Z}^2$ is an amenable graph. Our goal in this paper is to establish results for general graphs of bounded degree.

Our inspiration is the result of Mossel and Sly [38] who proved $O(n \log n)$ mixing time of the Glauber dynamics for every graph of maximum degree d. When $\beta < \beta_c(d)$, in addition to uniqueness on the infinite d-regular tree, the ferromagnetic Ising model is also known to exhibit several key spatial mixing properties. For instance, Mossel and Sly [38] showed that when $\beta < \beta_c(d)$ a rather strong form of spatial mixing holds on graphs of maximum degree d; see Definition 9 and Lemma 10 in Section 3. Using this, together with the censoring result of Peres and Winkler [40] for the Glauber dynamics, they establish optimal bounds for the mixing and relaxation times of the Glauber dynamics. At a high-level, the censoring result [40] says that extra updates by the Markov chain do not slow it down, and hence one can ignore transitions outside a local region of interest in the analysis of mixing times.

A Markov chain is *monotone* if it preserves the natural partial order on states; see Section 2 for a detailed definition. We generalize the proof approach of Mossel and Sly to apply to general (non-local) monotone Markov chains. This allows us to analyze a monotone variant of the SW dynamics, and a direct comparison of these two chains yields a new bound for the relaxation time of the SW dynamics.

▶ **Theorem 1.** *Let G be an arbitrary n-vertex graph of maximum degree d. If $\beta < \beta_c(d)$, then the relaxation time of the Swendsen-Wang dynamics is $\Theta(1)$.*

This tight bound for the relaxation time is a substantial improvement over the best previously known $O(n)$ bound which follows from Ullrich's comparison theorem [50] combined with Mossel and Sly's result [38] for the Glauber dynamics. We note that in Theorem 1, d is assumed to be a constant independent of n and thus the result holds for arbitrary graphs of *bounded* degree. We also mention that while spatial mixing properties are known to imply optimal mixing of local dynamics, only recently the effects of these properties on the rate of convergence of non-local dynamics have started to be investigated [1]. In general, spatial mixing properties have proved to have a number of powerful algorithmic applications in the design of efficient approximation algorithms for the partition function using the associated self-avoiding walk trees (see, e.g., [52, 45, 33, 16, 44, 46, 47]).

There are three key components in our proof approach. First, we generalize the recursive/inductive argument of Mossel and Sly [38] from the Glauber dynamics to general (non-local) monotone dynamics. Since this approach relies crucially on the censoring result of Peres and Winkler [40] which only applies to the Glauber dynamics, we also need to establish a modest extension of the censoring result. For this, we use the framework of Fill and Kahn [14]. Finally, we require a monotone Markov chain that can be analyzed with these new tools and which is naturally comparable to the SW dynamics. To this end we utilize the *Isolated-vertex dynamics* which was previously used in [1].

The Isolated-vertex dynamics operates in the same manner as the SW dynamics, except in step 3 only components of size 1 choose a new random spin (other components keep the same spin as in σ_t). We prove that the Isolated-vertex dynamics is *monotone*. Combining these new tools we obtain the following result.

▶ **Theorem 2.** *Let G be an arbitrary n-vertex graph of maximum degree d. If $\beta < \beta_c(d)$, then the mixing time of the Isolated-vertex dynamics is $O(\log n)$, and its relaxation time is $\Theta(1)$.*

Our result for censoring may be of independent interest, as it applies to a fairly general class of non-local monotone Markov chains. Indeed, combined with our generalization of Mossel and Sly's results [38], it gives a general method for analyzing monotone Markov chains.

As the first application of this technology, we are able to establish tight bounds for the mixing and relaxation times of the *block dynamics*. Let $\{B_1, ..., B_r\}$ be a collection of sets (or blocks) such that $B_i \subseteq V$ and $V = \cup_i B_i$. The *heat-bath block dynamics* with blocks $\{B_1, ..., B_r\}$ is a Markov chain that in each step picks a block B_i uniformly at random and updates the configuration in B_i with a new configuration distributed according to the conditional measure in B_i given the configuration in $V \setminus B_i$.

▶ **Theorem 3.** *Let G be an arbitrary n-vertex graph of maximum degree d and let $\{B_1, \ldots, B_r\}$ be an arbitrary collection of blocks such that $V = \cup_{i=1}^r B_i$. If $\beta < \beta_c(d)$, then the mixing time of the block dynamics with blocks $\{B_1, \ldots, B_r\}$ is $O(r \log n)$, and its relaxation time is $O(r)$.*

We observe that there are no restrictions on the geometry of the blocks B_i in the theorem other than $V = \cup_i B_i$. These optimal bounds were only known before for certain specific collections of blocks.

As a second application of our technology, we consider another monotone variant of the SW dynamics, which we call the *Monotone SW dynamics*. This chain proceeds exactly like the SW dynamics, except that in step 3 each connected component C is assigned a new random spin only with probability $1/2^{|C|-1}$ and is not updated otherwise. We derive the following bounds.

▶ **Theorem 4.** *Let G be an arbitrary n-vertex graph of maximum degree d. If $\beta < \beta_c(d)$, then the mixing time of the Monotone SW dynamics is $O(\log n)$, and its relaxation time is $\Theta(1)$.*

The remainder of the paper is structured as follows. Section 2 contains some basic definitions and facts used throughout the paper. In Section 3 we study the Isolated-vertex dynamics and establish Theorem 2. Theorem 1 for the SW dynamics will follow as an easy corollary of these results. In Section 3 we also state our generalization of Mossel and Sly's approach [38] for non-local dynamics (Theorem 11) and our censoring result (Theorem 7). The proofs of these theorems are included in Appendix A and B, respectively. Finally, the proofs of Theorems 3 and 4 are provided in the full version [2].

2 Background

In this section we provide a number of standard definitions that we will refer to in our proofs. For more details see the book [32].

Ferromagnetic Ising model. Given a graph $G = (V, E)$ and a real number $\beta > 0$, the ferromagnetic Ising model on G consists of the probability distribution over $\Omega_G = \{+, -\}^V$ given by

$$\mu_{G,\beta}(\sigma) = \frac{1}{Z(G, \beta)} \exp\left[\beta \sum\nolimits_{\{u,v\}\in E} \sigma(u)\sigma(v)\right], \tag{3}$$

where $\sigma \in \Omega_G$ and

$$Z(G, \beta) = \sum\nolimits_{\sigma\in\Omega_G} \exp\left[\beta \sum\nolimits_{\{u,v\}\in E} \sigma(u)\sigma(v)\right]$$

is called the *partition function*.

Mixing and relaxation times. Let P be the transition matrix of an ergodic (i.e., irreducible and aperiodic) Markov chain over Ω_G with stationary distribution $\mu = \mu_{G,\beta}$. Let $P^t(X_0, \cdot)$ denote the distribution of the chain after t steps starting from $X_0 \in \Omega_G$, and let

$$T_{\rm mix}(P, \varepsilon) = \max_{X_0\in\Omega} \min\left\{t \geq 0 : \|P^t(X_0, \cdot) - \mu(\cdot)\|_{\rm TV} \leq \varepsilon\right\}.$$

The *mixing time* of P is defined as $T_{\rm mix}(P) = T_{\rm mix}(P, 1/4)$.

If P is reversible with respect to (w.r.t.) μ, the spectrum of P is real. Let $1 = \lambda_1 > \lambda_2 \geq \ldots \geq \lambda_{|\Omega|} \geq -1$ denote its eigenvalues. The *absolute spectral gap* of P is defined by $\lambda(P) = 1 - \lambda^*$, where $\lambda^* = \max\{|\lambda_2|, |\lambda_{|\Omega|}|\}$. $T_{\rm rel}(P) = \lambda(P)^{-1}$ is called the *relaxation time* of P, and is another well-studied notion of rate of convergence to μ [27, 29].

Couplings and grand couplings. A *(one step) coupling* of a Markov chain $\mathcal{M}$ over Ω_G specifies, for every pair of states $(X_t, Y_t) \in \Omega_G \times \Omega_G$, a probability distribution over (X_{t+1}, Y_{t+1}) such that the processes $\{X_t\}$ and $\{Y_t\}$, viewed in isolation, are faithful copies of $\mathcal{M}$, and if $X_t = Y_t$ then $X_{t+1} = Y_{t+1}$. Let $\{X_t^\sigma\}_{t\geq 0}$ denote an instance of $\mathcal{M}$ started from $\sigma \in \Omega_G$. A *grand coupling* of $\mathcal{M}$ is a simultaneous coupling of $\{X_t^\sigma\}_{t\geq 0}$ for all $\sigma \in \Omega_G$.

Monotonicity. For two configurations $\sigma, \tau \in \Omega_G$, we say $\sigma \geq \tau$ if $\sigma(v) \geq \tau(v)$ for all $v \in V$ (assuming "$+$" $>$ "$-$"). This induces a partial order on Ω_G. The ferromagnetic Ising model is *monotone* w.r.t. this partial order, since for every $B \subseteq V$ and every pair of configurations τ_1, τ_2 on B such that $\tau_1 \geq \tau_2$ we have $\mu(\cdot \mid \tau_1) \succeq \mu(\cdot \mid \tau_2)$, where $\succeq$ denotes stochastic domination. (For two distributions ν_1, ν_2 on Ω_G, we say that ν_1 stochastically dominates ν_2 if for any increasing function $f \in \mathbb{R}^{|\Omega_G|}$ we have $\sum_{\sigma\in\Omega_G} \nu_1(\sigma) f(\sigma) \geq \sum_{\sigma\in\Omega_G} \nu_2(\sigma) f(\sigma)$, where a vector or function $f \in \mathbb{R}^{|\Omega_G|}$ is increasing if $f(\sigma) \geq f(\tau)$ for all $\sigma \geq \tau$.)

Suppose $\mathcal{M}$ is an ergodic Markov chain over Ω_G with stationary distribution μ and transition matrix P. A coupling of two instances $\{X_t\}$, $\{Y_t\}$ of $\mathcal{M}$ is a *monotone coupling* if $X_{t+1} \geq Y_{t+1}$ whenever $X_t \geq Y_t$. We say that $\mathcal{M}$ is a *monotone Markov chain* and P is a *monotone transition matrix* if $\mathcal{M}$ has a monotone grand coupling.

Comparison inequalities. The *Dirichlet form* of a Markov chain with transition matrix P reversible w.r.t. μ is defined for any $f, g \in \mathbb{R}^{|\Omega_G|}$ as

$$\mathcal{E}_P(f, g) = \langle f, (I - P)g\rangle_\mu = \frac{1}{2} \sum_{\sigma,\tau\in\Omega_G} \mu(\sigma) P(\sigma, \tau)(f(\sigma) - f(\tau))(g(\sigma) - g(\tau)),$$

where $\langle f, g\rangle_\mu = \sum_{\sigma\in\Omega_G} \mu(\sigma) f(\sigma) g(\sigma)$ for all $f, g \in \mathbb{R}^{|\Omega_G|}$.

If P and Q are the transition matrices of two monotone Markov chains reversible w.r.t. μ, we say that $P \leq Q$ if $\langle Pf, g\rangle_\mu \leq \langle Qf, g\rangle_\mu$ for every increasing and positive $f, g \in \mathbb{R}^{|\Omega_G|}$. Note that $P \leq Q$ is equivalent to $\mathcal{E}_P(f, g) \geq \mathcal{E}_Q(f, g)$ for every increasing and positive $f, g \in \mathbb{R}^{|\Omega_G|}$.

3 Isolated-vertex dynamics

In this section we consider a variant of the SW dynamics known as the *Isolated-vertex dynamics* which was first introduced in [1]. We shall use this dynamics to introduce a general framework for analyzing monotone Markov chains for the Ising model and to derive our bounds for the SW dynamics. Specifically, we will prove Theorems 1 and 2 from the introduction.

Throughout the section, let $G = (V, E)$ be an arbitrary n-vertex graph of maximum degree d, $\mu = \mu_{G,\beta}$ and $\Omega = \Omega_G$. Given an Ising model configuration $\sigma_t \in \Omega$, one step of the Isolated-vertex dynamics is given by:

1. Consider the set of agreeing edges $E(\sigma_t) = \{(v, w) \in E : \sigma_t(v) = \sigma_t(w)\}$;
2. Independently for each edge $e \in E(\sigma_t)$, delete e with probability $\exp(-2\beta)$ and keep e with probability $1 - \exp(-2\beta)$; this yields $F_t \subseteq E(\sigma_t)$;

3. For each *isolated* vertex v in the subgraph (V, F_t) (i.e., those vertices with no incident edges in F_t), choose a spin uniformly at random from $\{+,-\}$ and assign it to v to obtain σ_{t+1}; all other (non-isolated) vertices keep the same spin as in σ_t.

We use $\mathcal{IV}$ to denote the transition matrix of this chain. The reversibility of $\mathcal{IV}$ with respect to μ was established in [1]. Observe also that in step 3, only *isolated vertices* are updated with new random spins, whereas in the SW dynamics all connected components are assigned new random spins. It is thus intuitive that the SW dynamics converges faster to stationarity than the Isolated-vertex dynamics. This intuition was partially captured in [1], where it was proved that

$$T_{\text{rel}}(\mathcal{SW}) \leq T_{\text{rel}}(\mathcal{IV}). \tag{4}$$

The Isolated-vertex dynamics exhibits various properties that vastly simplify its analysis. These properties allow us to deduce, for example, strong bounds for both its relaxation and mixing times. Specifically, we show (see Theorem 2) that when $\beta < \beta_c(d)$, $T_{\text{mix}}(\mathcal{IV}) = O(\log n)$ and $T_{\text{rel}}(\mathcal{IV}) = \Theta(1)$; see also (2) for the definition of $\beta_c(d)$. Theorem 1 from the introduction then follows from (4).

A comparison inequality like (4) but for mixing times is not known, so Theorem 2 does not yield a $O(\log n)$ bound for the mixing time of the SW dynamics as one might hope. Direct comparison inequalities for mixing times are rare, since almost all known techniques involve the comparison of Dirichlet forms, and there are inherent penalties in using such inequalities to derive mixing times bounds.

The first key property of the Isolated-vertex dynamics is that, unlike the SW dynamics, this Markov chain is *monotone*. Monotonicity is known to play a key role in relating spatial mixing (i.e., decay of correlation) properties to fast convergence of the Glauber dynamics. For instance, for spin systems in lattice graphs, sophisticated functional analytic techniques are required to establish the equivalence between a spatial mixing property known as *strong spatial mixing* and optimal mixing of the Glauber dynamics [35, 36, 37]. For monotone spin systems such as the Ising model a simpler combinatorial argument yields the same sharp result [10]. This combinatorial argument is in fact more robust, since it can be used to analyze a larger class of Markov chains, including for example the systematic scan dynamics [1].

▶ **Lemma 5.** *For all graphs G and all $\beta > 0$, the Isolated-vertex dynamics for the Ising model is monotone.*

The proof of Lemma 5 is given in Section 3.1. The second key property of the Isolated-vertex dynamics concerns whether moves (or partial moves) of the dynamics could be *censored* from the evolution of the chain without possibly speeding up its convergence. Censoring of Markov chains is a well-studied notion [40, 14, 24] that has found important applications [38, 8, 9].

We say that a stochastic $|\Omega| \times |\Omega|$ matrix Q acts on a set $A \subseteq V$ if for all $\sigma, \sigma' \in \Omega$:

$$Q(\sigma, \sigma') \neq 0 \text{ iff } \sigma(V \setminus A) = \sigma'(V \setminus A).$$

Also recall that $P \leq P_A$ if $\langle Pf, g \rangle_\mu \leq \langle P_A f, g \rangle_\mu$ for any pair of increasing positive functions $f, g \in \mathbb{R}^{|\Omega|}$.

▶ **Definition 6.** Let G be an arbitrary graph and let $\beta > 0$. Consider an ergodic and monotone Markov chain for the Ising model on G, reversible w.r.t. $\mu = \mu_{G,\beta}$ with transition matrix P. Let $\{P_A\}_{A \subseteq V}$ be a collection of monotone stochastic matrices reversible w.r.t. μ with the property that P_A acts on A for every $A \subseteq V$. We say that $\{P_A\}_{A \subseteq V}$ is a *censoring* for P if $P \leq P_A$ for all $A \subseteq V$.

As an example, consider the *heat-bath Glauber dynamics* for the Ising model on the graph $G = (V, E)$. Recall that in this Markov chain a vertex $v \in V$ is chosen uniformly at random (u.a.r.) and a new spin is sampled for v from the conditional distribution at v given the configuration on $V \setminus v$. For every $A \subseteq V$, we may take P_A to be the $|\Omega| \times |\Omega|$ transition matrix of the censored heat-bath Glauber dynamics that ignores all moves outside of A. That is, if the randomly chosen vertex $v \in V$ is not in A, then the move is ignored; otherwise the chain proceeds as the standard heat-bath Glauber dynamics.

It is easy to check that P_A is monotone and reversible w.r.t. μ. Moreover, it was established in [40, 14] that $P \leq P_A$ for every $A \subseteq V$, and thus the collection $\{P_A\}_{A\subseteq V}$ is a censoring for the heat-bath Glauber dynamics. This particular censoring has been used to analyze the speed of convergence of the Glauber dynamics in various settings (see [38, 40, 8, 9]), since it can be proved that censored variants of the Glauber dynamics—where moves of P are replaced by moves of P_A—converge more slowly to the stationary distribution [40, 14]. Consequently, it suffices to analyze the speed of convergence of the censored chain, and this could be much simpler for suitably chosen censoring schemes.

Using the machinery from [40, 14], we can show that given a censoring (as defined in Definition 6), the strategy just mentioned for Glauber dynamics can be used for general monotone Markov chains.

▶ **Theorem 7.** *Let G be an arbitrary graph and let $\beta > 0$. Let $\{X_t\}$ be an ergodic monotone Markov chain for the Ising model on G, reversible w.r.t. $\mu = \mu_{G,\beta}$ with transition matrix P. Let $\{P_A\}_{A\subseteq V}$ be a censoring for P and let $\{\hat{X}_t\}$ be a censored version of $\{X_t\}$ that sequentially applies $P_{A_1}, P_{A_2}, P_{A_3} \dots$ where $A_i \subseteq V$. If X_0, Y_0 are both sampled from a distribution ν over Ω such that ν/μ is increasing, then the following hold:*
1. *$X_t \preceq \hat{X}_t$ for all $t \geq 0$;*
2. *Let $\hat{P}^t = P_{A_1} \dots P_{A_t}$. Then, for all $t \geq 0$*

$$\|P^t(X_0,\cdot) - \mu(\cdot)\|_{\text{TV}} \leq \|\hat{P}^t(X_0,\cdot) - \mu(\cdot)\|_{\text{TV}}.$$

If ν/μ is decreasing, then $X_t \succeq \hat{X}_t$ for all $t \geq 0$.

The proof of this theorem is provided in Appendix B.

We define next a specific censoring for the Isolated-vertex dynamics. For $A \subseteq V$, let $\mathcal{IV}_A$ be the transition matrix for the Markov chain that given an Ising model configuration σ_t generates σ_{t+1} as follows:

1. Consider the set of agreeing edges $E(\sigma_t) = \{(v, w) \in E : \sigma_t(v) = \sigma_t(w)\}$;
2. Independently for each edge $e \in E(\sigma_t)$, delete e with probability $\exp(-2\beta)$ and keep e with probability $1 - \exp(-2\beta)$; this yields $F_t \subseteq E(\sigma_t)$;
3. For each isolated vertex v of the subgraph (V, F_t) in the subset A, choose a spin uniformly at random from $\{+, -\}$ and assign it to v to obtain σ_{t+1}; all other vertices keep the same spin as in σ_t.

▶ **Lemma 8.** *The collection of matrices $\{\mathcal{IV}_A\}_{A\subseteq V}$ is a censoring for the Isolated-vertex dynamics.*

The proof of Lemma 8 is provided in Secion 3.2. To establish Theorem 2 we show that a strong form of spatial mixing, which is known to hold for all $\beta < \beta_c(d)$ [38], implies the desired mixing and relaxation times bounds for the Isolated-vertex dynamics. We define this notion of spatial mixing next.

For $v \in V$ and $R \in \mathbb{N}$, let $B(v, R) = \{u \in V : \text{dist}(u, v) \leq R\}$ denote the ball of radius R around v, where $\text{dist}(\cdot, \cdot)$ denotes graph distance. Also, let $S(v, R) = B(v, R + 1) \backslash B(v, R)$

be the external boundary of $B(v,R)$. For any $A \subseteq V$, let $\Omega_A = \{+,-\}^A$ be the set of all configurations on A; hence $\Omega = \Omega_G = \Omega_V$. For $v \in V$, $u \in S(v,R)$ and $\tau \in \Omega_{S(v,R)}$, let τ_u^+ (resp., τ_u^-) be the configuration obtained from τ by changing the spin of u to $+$ (resp., to $-$) and define

$$a_u = \sup_{\tau \in \Omega_{S(v,R)}} \left| \mu\left(v = + \mid S(v,R) = \tau_u^+\right) - \mu\left(v = + \mid S(v,R) = \tau_u^-\right) \right|, \tag{5}$$

where "$v = +$" represents the event that the spin of v is $+$ and "$S(v,R) = \tau_u^+$" (resp., "$S(v,R) = \tau_u^-$") stands for the event that $S(v,R)$ has configuration τ_u^+ (resp., τ_u^-).

▶ **Definition 9.** We say that *Aggregate Strong Spatial Mixing (ASSM)* holds for $R \in \mathbb{N}$, if for all $v \in V$

$$\sum_{u \in S(v,R)} a_u \le \frac{1}{4}.$$

▶ **Lemma 10** (Lemma 3, [38])**.** *For all graphs G of maximum degree d and all $\beta < \beta_c(d)$, there exists an integer $R = R(\beta, d) \in \mathbb{N}$ such that ASSM holds for R.*

Theorem 2 is then a direct corollary of the following more general theorem. The proof of this general theorem, which is provided in Appendix A, follows closely the approach in [38] for the case of the Glauber dynamics, but key additional considerations are required to establish such result for general (non-local) monotone Markov chains. The main new innovation in our proof is the use of the more general Theorem 7, instead of the standard censoring result in [40].

▶ **Theorem 11.** *Let $\beta > 0$ and G be an arbitrary n-vertex graph of maximum degree d where d is a constant independent of n. Consider an ergodic monotone Markov chain for the Ising model on G, reversible w.r.t. $\mu = \mu_{G,\beta}$ with transition matrix P. Suppose $\{P_A\}_{A \subseteq V}$ is a censoring for P. If ASSM holds for a constant $R > 0$, and for any $v \in V$ and any starting configuration $\sigma \in \Omega$*

$$T_{\rm mix}(P_{B(v,R)}) \le T, \tag{6}$$

then $T_{\rm mix}(P) = O(T \log n)$ and $T_{\rm rel}(P) = O(T)$.

We note that $T_{\rm mix}(P_{B(v,R)})$ denotes the mixing time from the worst possible starting configuration, both in $B(v,R)$ and in $V \setminus B(v,R)$. (Since $P_{B(v,R)}$ only acts in $B(v,R)$, the configuration in $V \setminus B(v,R)$ remains fixed throughout the evolution of the chain and determines its stationary distribution.)

We now use Theorem 11 to establish Theorem 2. Theorem 11 is also used to establish Theorems 3 and 4 from the introduction, concerning the mixing time of the block dynamics and a monotone variant of the SW dynamics; see the full version of this paper [2].

Proof of Theorem 2. By Lemma 5 the Isolated-vertex dynamics is monotone, and by Lemma 8 the collection $\{\mathcal{IV}_A\}_{A \subseteq V}$ is a censoring for $\mathcal{IV}$. Moreover, Lemma 10 implies that there exists a constant R such that ASSM holds. Thus, to apply Theorem 11 all that is needed is a bound for $T_{\rm mix}(\mathcal{IV}_{B(v,R)})$ for all $v \in V$. For this, we can use a crude coupling argument. Since $|B(v,R)| \le d^R$, the probability that every vertex in $B(v,R)$ becomes isolated is at least

$$\mathrm{e}^{-2\beta d |B(v,R)|} \ge \mathrm{e}^{-2\beta d^{R+1}}.$$

Starting from two arbitrary configurations in $B(v,R)$, if all vertices become isolated in both configurations, then we can couple them with probability 1. Hence, we can couple two arbitrary configurations in one step with probability at least $\exp(-2\beta d^{R+1})$. Thus, $T_{\mathrm{mix}}(\mathcal{IV}_{B(v,R)}) = \exp(O(\beta d^{R+1})) = O(1)$, and the result then follows from Theorem 11. ◀

Proof of Theorem 1. Follows from Theorem 2 and the fact that $T_{\mathrm{rel}}(\mathcal{SW}) \le T_{\mathrm{rel}}(\mathcal{IV})$, which was established in Lemma 4.1 from [1]. ◀

3.1 Monotonicity of the Isolated-vertex dynamics

In this section, we show that the Isolated-vertex dynamics is monotone by constructing a monotone grand coupling; see Section 2 for the definition of a grand coupling. In particular, we prove Lemma 5.

Proof of Lemma 5. Let $\{X_t^\sigma\}_{t\ge 0}$ be an instance of the Isolated-vertex dynamics starting from $\sigma \in \Omega$; i.e., $X_0^\sigma = \sigma$. We construct a grand coupling for the Isolated-vertex dynamics as follows. At time t:

1. For every edge $e \in E$, pick a number $r_t(e)$ uniformly at random from $[0,1]$;
2. For every vertex $v \in V$, choose a uniform random spin $s_t(v)$ from $\{+,-\}$;
3. For every $\sigma \in \Omega$:
   - **(i)** Obtain $F_t^\sigma \subseteq E$ by including the edge $e = \{u,v\}$ in F_t^σ iff $X_t^\sigma(u) = X_t^\sigma(v)$ and $r_t(e) \le 1 - \mathrm{e}^{-2\beta}$;
   - **(ii)** For every $v \in V$, set $X_{t+1}^\sigma(v) = s_t(v)$ if v is an *isolated vertex* in the subgraph (V, F_t^σ); otherwise, set $X_{t+1}^\sigma(v) = X_t^\sigma(v)$.

This is clearly a valid grand coupling for the Isolated-vertex dynamics. We show next that it is also monotone.

Suppose $X_t^\sigma \ge X_t^\tau$. We need to show that $X_{t+1}^\sigma \ge X_{t+1}^\tau$ after one step of the grand coupling. Let $v \in V$. If v is not isolated in either (V, F_t^σ) or (V, F_t^τ), then the spin of v remains unchanged in both X_{t+1}^σ and X_{t+1}^τ, and $X_{t+1}^\sigma(v) = X_t^\sigma(v) \ge X_t^\tau(v) = X_{t+1}^\tau(v)$. On the other hand, if v is isolated in both (V, F_t^σ) and (V, F_t^σ), then the spin of v is set to $s_t(v)$ in both instances of the chain; hence, $X_{t+1}^\sigma(v) = s_t(v) = X_{t+1}^\tau(v)$.

Suppose next that v is isolated in (V, F_t^σ) but not in (V, F_t^τ). Then, $X_{t+1}^\sigma(v) = s_t(v)$ and $X_{t+1}^\tau(v) = X_t^\tau(v)$. The only possibility that would violate $X_{t+1}^\sigma(v) \ge X_{t+1}^\tau(v)$ is that $X_{t+1}^\sigma(v) = -, X_t^\sigma(v) = +$ and $X_{t+1}^\tau(v) = X_t^\tau(v) = +$. If this is the case, then $X_t^\sigma(v) = X_t^\tau(v) = +$. Moreover, since $X_t^\sigma \ge X_t^\tau$, all neighbors of v assigned "+" in X_t^τ are also "+" in X_t^σ; thus if v is isolated in (V, F_t^σ) then v is also isolated in (V, F_t^τ). This leads to a contradiction, and so $X_{t+1}^\sigma(v) \ge X_{t+1}^\tau(v)$. The case in which v is isolated in (V, F_t^τ) but not in (V, F_t^σ) follows from an analogous argument. ◀

We can use the same grand coupling to show that $\mathcal{IV}_A$ is also monotone for all $A \subseteq V$. The only required modification in the construction is that if $v \in V \setminus A$, then the spin of v is not updated in either copy. This gives the following corollary.

▶ **Corollary 12.** *$\mathcal{IV}_A$ is monotone for all $A \subseteq V$.*

3.2 Censoring for the Isolated-vertex dynamics

In this section we show that the collection $\{\mathcal{IV}_A\}_{A\subseteq V}$ is a censoring for $\mathcal{IV}$. Specifically, we prove Lemma 8.

Proof of Lemma 8. For all $A \subseteq V$, we need to establish that $\mathcal{IV}_A$ is reversible w.r.t. $\mu = \mu_{G,\beta}$, monotone and that $\mathcal{IV} \leq \mathcal{IV}_A$. Monotonicity follows from Corollary 12. To establish the other two facts we use an alternative representation of the matrices $\mathcal{IV}$ and $\mathcal{IV}_A$ that was already used in [1] and is inspired by the methods in [50].

Let $\Omega_{\mathrm{J}} = 2^E \times \Omega$ be the *joint* configuration space, where configurations consist of a spin assignment to the vertices together with a subset of the edges of G. The joint Edwards-Sokal measure ν on Ω_{J} is given by

$$\nu(F,\sigma) = \frac{1}{Z_{\mathrm{J}}} p^{|F|}(1-p)^{|E \setminus F|}\, \mathbb{1}(F \subseteq E(\sigma)), \tag{7}$$

where $p = 1 - \mathrm{e}^{-2\beta}$, $F \subseteq E$, $\sigma \in \Omega$, $E(\sigma) = \{\{u,v\} \in E : \sigma(u) = \sigma(v)\}$, and Z_{J} is the partition function [11].

Let T be the $|\Omega| \times |\Omega_{\mathrm{J}}|$ matrix given by:

$$T(\sigma,(F,\tau)) = \mathbb{1}(\sigma = \tau)\, \mathbb{1}(F \subseteq E(\sigma)) p^{|F|}(1-p)^{|E(\sigma)\setminus F|}, \tag{8}$$

where $\sigma \in \Omega$ and $(F,\tau) \in \Omega_{\mathrm{J}}$. The matrix T corresponds to adding each edge $\{u,v\} \in E$ with $\sigma(u) = \sigma(v)$ independently with probability p, as in step 1 of the Isolated-vertex dynamics. Let $L_2(\nu)$ and $L_2(\mu)$ denote the Hilbert spaces $(\mathbb{R}^{|\Omega_{\mathrm{J}}|}, \langle \cdot,\cdot \rangle_\nu)$ and $(\mathbb{R}^{|\Omega|}, \langle \cdot,\cdot \rangle_\mu)$ respectively. The matrix T defines an operator from $L_2(\nu)$ to $L_2(\mu)$ via vector-matrix multiplication. Specifically, for any $f \in \mathbb{R}^{|\Omega_{\mathrm{J}}|}$ and $\sigma \in \Omega$

$$Tf(\sigma) = \sum_{(F,\tau) \in \Omega_{\mathrm{J}}} T(\sigma,(F,\tau)) f(F,\tau).$$

It is easy to check that the adjoint operator $T^* : L_2(\mu) \to L_2(\nu)$ of T is given by the $|\Omega_{\mathrm{J}}| \times |\Omega|$ matrix

$$T^*((F,\tau),\sigma) = \mathbb{1}(\tau = \sigma), \tag{9}$$

with $(F,\tau) \in \Omega_{\mathrm{J}}$ and $\sigma \in \Omega$. Finally, for $A \subseteq V$, $F_1, F_2 \subseteq E$ and $\sigma, \tau \in \Omega$ let

$$\begin{aligned} Q_A((F_1,\sigma),(F_2,\tau)) = {} & \mathbb{1}(F_1 = F_2)\, \mathbb{1}(F_1 \subseteq E(\sigma) \cap E(\tau)) \\ & \mathbb{1}(\sigma(\mathcal{I}_A^c(F_1)) = \tau(\mathcal{I}_A^c(F_1))) \cdot 2^{-|\mathcal{I}_A(F_1)|} \end{aligned}$$

where $\mathcal{I}_A(F_1)$ is the set of isolated vertices of (V, F_1) in A and $\mathcal{I}_A^c(F_1) = V \setminus \mathcal{I}_A(F_1)$, and similarly for F_2. For ease of notation we set $Q = Q_V$. It follows straightforwardly from the definition of these matrices that $\mathcal{IV} = TQT^*$ and $\mathcal{IV}_A = TQ_AT^*$ for all $A \subseteq V$. It is also easy to verify that $Q = Q^2 = Q^*$, $Q_A = Q_A^2 = Q_A^*$ and that $Q = Q_AQQ_A$; see [1].

The reversibility of $\mathcal{IV}_A$ w.r.t. μ follows from the fact that $\mathcal{IV}_A^* = (TQ_AT^*)^* = TQ_AT^* = \mathcal{IV}_A$. This implies that $\mathcal{IV}_A$ is self-adjoint and thus reversible w.r.t. μ [32].

To establish that $\mathcal{IV} \leq \mathcal{IV}_A$, it is sufficient to show that for every pair of increasing and positive functions $f_1, f_2 : \mathbb{R}^{|\Omega|} \to \mathbb{R}$ on Ω, we have

$$\langle f_1, \mathcal{IV} f_2 \rangle_\mu \leq \langle f_1, \mathcal{IV}_A f_2 \rangle_\mu. \tag{10}$$

Now,

$$\langle f_1, \mathcal{IV}_A f_2 \rangle_\mu = \langle f_1, TQ_AT^* f_2 \rangle_\mu = \langle f_1, TQ_A^2T^* f_2 \rangle_\mu = \langle Q_AT^* f_1, Q_AT^* f_2 \rangle_\nu = \langle \hat{f}_1, \hat{f}_2 \rangle_\nu,$$

where $\hat{f}_1 = Q_AT^* f_1$ and $\hat{f}_2 = Q_AT^* f_2$. Similarly,

$$\langle f_1, \mathcal{IV} f_2 \rangle_\mu = \langle f_1, TQ_AQ^2Q_AT^* f_2 \rangle_\mu = \langle QQ_AT^* f_1, QQ_AT^* f_2 \rangle_\nu = \langle Q\hat{f}_1, Q\hat{f}_2 \rangle_\nu.$$

Thus, it is sufficient for us to show that $\langle Q\hat{f}_1, Q\hat{f}_2 \rangle_\nu \leq \langle \hat{f}_1, \hat{f}_2 \rangle_\nu$.

Consider the partial order on Ω_{J} where $(F,\sigma) \geq (F',\sigma')$ iff $F = F'$ and $\sigma \geq \sigma'$.

▶ **Claim 13.** *Suppose $f : \mathbb{R}^{|\Omega|} \to \mathbb{R}$ is an increasing positive function. Then, $\hat{f} : \mathbb{R}^{|\Omega_{\mathrm{J}}|} \to \mathbb{R}$ where $\hat{f} = Q_A T^* f$ is also increasing and positive.*

Given $\omega \in \Omega_{\mathrm{J}}$, let $\rho_\omega(\cdot) = Q(\omega, \cdot)$; i.e., ρ_ω is the distribution over Ω_{J} after applying Q from ω. We have

$$Q\hat{f}_1(\omega) = \sum_{\omega' \in \Omega_{\mathrm{J}}} Q\big(\omega, \omega'\big)\hat{f}_1(\omega') = \mathbb{E}_{\rho_\omega}[\hat{f}_1].$$

Similarly, we get $Q\hat{f}_2(\omega) = \mathbb{E}_{\rho_\omega}[\hat{f}_2]$.

For a distribution π on a partially ordered set S, we say π is positively correlated if for any increasing functions $f, g \in \mathbb{R}^{|S|}$ we have $\mathbb{E}_\pi[fg] \geq \mathbb{E}_\pi[f]\,\mathbb{E}_\pi[g]$. Since ρ_ω is a product distribution over the isolated vertices in ω, ρ_ω is positively correlated for any $\omega \in \Omega_{\mathrm{J}}$ by Harris inequality (see, e.g., Lemma 22.14 in [32]). By Claim 13, $\hat{f}_1$ and $\hat{f}_2$ are increasing. We then deduce that for any $\omega \in \Omega_{\mathrm{J}}$:

$$Q\hat{f}_1(\omega)\,Q\hat{f}_2(\omega) = \mathbb{E}_{\rho_\omega}[\hat{f}_1]\,\mathbb{E}_{\rho_\omega}[\hat{f}_2] \leq \mathbb{E}_{\rho_\omega}[\hat{f}_1\,\hat{f}_2].$$

Putting all these facts together, we get

$$\begin{aligned}\langle Q\hat{f}_1, Q\hat{f}_2\rangle_\nu &= \sum_{\omega \in \Omega_{\mathrm{J}}} Q\hat{f}_1(\omega)\,Q\hat{f}_2(\omega)\nu(\omega) \leq \sum_{\omega \in \Omega_{\mathrm{J}}} \mathbb{E}_{\rho_\omega}[\hat{f}_1\,\hat{f}_2]\nu(\omega) \\ &= \sum_{\omega, \omega' \in \Omega_{\mathrm{J}}} \hat{f}_1(\omega')\,\hat{f}_2(\omega')\rho_\omega(\omega')\nu(\omega) = \sum_{\omega, \omega' \in \Omega_{\mathrm{J}}} \hat{f}_1(\omega')\,\hat{f}_2(\omega')\rho_{\omega'}(\omega)\nu(\omega') \\ &= \langle \hat{f}_1, \hat{f}_2\rangle_\nu,\end{aligned}$$

where the second to last equality follows from the reversibility of Q w.r.t. ν; namely,

$$\rho_\omega(\omega')\nu(\omega) = Q(\omega, \omega')\nu(\omega) = Q(\omega', \omega)\nu(\omega') = \rho_{\omega'}(\omega)\nu(\omega').$$

This implies that (10) holds for every pair of increasing positive functions, and the theorem follows. ◀

We conclude this section with the proof of Claim 13.

Proof of Claim 13. From the definition of T^* we get $T^*f(F, \sigma) = f(\sigma)$ for any $(F, \sigma) \in \Omega_{\mathrm{J}}$. Let $(F, \sigma), (F, \tau) \in \Omega_{\mathrm{J}}$ be such that $\sigma \geq \tau$. Then,

$$\hat{f}(F, \sigma) = Q_A T^* f(F, \sigma) = \sum_{(F', \sigma') \in \Omega_{\mathrm{J}}} Q_A\big((F, \sigma), (F', \sigma')\big) f(\sigma').$$

Recall that $Q_A\big((F, \sigma), (F', \sigma')\big) > 0$ iff $F = F'$ and σ, σ' differ only in $\mathcal{I}_A(F)$, the set of isolated vertices in A. If this is the case, then

$$Q_A\big((F, \sigma), (F, \sigma')\big) = \frac{1}{2^{|\mathcal{I}_A(F)|}}.$$

For $\xi \in \Omega_{\mathcal{I}_A(F)}$, let σ_ξ denote the configuration obtained from σ by changing the spins of vertices in $\mathcal{I}_A(F)$ to ξ; τ_ξ is defined similarly. (Recall that $\Omega_{\mathcal{I}_A(F)}$ denotes the set of Ising configurations on the set $\mathcal{I}_A(F)$.) Then, $\sigma_\xi \geq \tau_\xi$ for any $\xi \in \Omega_{\mathcal{I}_A(F)}$ and

$$\hat{f}(F, \sigma) = \frac{1}{2^{|\mathcal{I}_A(F)|}} \sum_{\xi \in \Omega_{\mathcal{I}_A(F)}} f(\sigma_\xi) \geq \frac{1}{2^{|\mathcal{I}_A(F)|}} \sum_{\xi \in \Omega_{\mathcal{I}_A(F)}} f(\tau_\xi) = \hat{f}(F, \tau).$$

This shows that $\hat{f}$ is increasing. ◀

References

1 A. Blanca, P. Caputo, A. Sinclair, and E. Vigoda. Spatial Mixing and Non-local Markov chains. In *Proceedings of the 29th Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pages 1965–1980, 2018.

2 A. Blanca, Z. Chen, and E. Vigoda. Swendsen-Wang Dynamics for General Graphs in the Tree Uniqueness Region. *ArXiv preprint arXiv:1806.04602*, 2018.

3 A. Blanca and A. Sinclair. Dynamics for the mean-field random-cluster model. In *Proceedings of the 19th International Workshop on Randomization and Computation (RANDOM)*, pages 528–543, 2015.

4 C. Borgs, J. Chayes, and P. Tetali. Tight bounds for mixing of the Swendsen-Wang algorithm at the Potts transition point. *Probability Theory and Related Fields*, 152(3-4):509–557, 2012.

5 C. Borgs, A. M. Frieze, J. H. Kim, P. Tetali, E. Vigoda, and V. Vu. Torpid mixing of some Monte Carlo Markov chain algorithms in statistical physics. In *Proceedings of the 40th Annual IEEE Symposium on Foundations of Computer Science (FOCS)*, pages 218–229, 1999.

6 F. Cesi. Quasi–factorization of the entropy and logarithmic Sobolev inequalities for Gibbs random fields. *Probability Theory and Related Fields*, 120(4):569–584, 2001.

7 C. Daskalakis, E. Mossel, and S. Roch. Evolutionary trees and the Ising model on the Bethe lattice: a proof of Steel's conjecture. *Probability Theory and Related Fields*, 149(1-2):149–189, 2011.

8 J. Ding, E. Lubetzky, and Y. Peres. Mixing time of critical Ising model on trees is polynomial in the height. *Communications in Mathematical Physics*, 295(1):161–207, 2010.

9 J. Ding and Y. Peres. Mixing time for the Ising model: a uniform lower bound for all graphs. *Annales de l'Institut Henri Poincaré, Probabilités et Statistiques*, 47(4):1020–1028, 2011.

10 M. Dyer, A. Sinclair, E. Vigoda, and D. Weitz. Mixing in time and space for lattice spin systems: A combinatorial view. *Random Structures & Algorithms*, 24(4):461–479, 2004.

11 R. G. Edwards and A. D. Sokal. Generalization of the Fortuin-Kasteleyn-Swendsen-Wang representation and Monte Carlo algorithm. *Physical Review D*, 38(6):2009–2012, 1988.

12 G. Ellison. Learning, local interaction, and coordination. *Econometrica: Journal of the Econometric Society*, pages 1047–1071, 1993.

13 J. Felsenstein. *Inferring Phylogenies*, volume 2. Sinauer Associates Sunderland, MA, 2004.

14 J. A. Fill and J. Kahn. Comparison inequalities and fastest-mixing Markov chains. *The Annals of Applied Probability*, 23(5):1778–1816, 2013.

15 A. Galanis, D. Štefankovič, and E. Vigoda. Swendsen-Wang algorithm on the mean-field Potts model. In *Proceedings of the 19th International Workshop on Randomization and Computation (RANDOM)*, pages 815–828, 2015.

16 D. Gamarnik and D. Katz. Correlation decay and deterministic FPTAS for counting list-colorings of a graph. In *Proceedings of the 18th Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pages 1245–1254, 2007.

17 S. Geman and C. Graffigne. Markov random field image models and their applications to computer vision. In *Proceedings of the International Congress of Mathematicians*, pages 1496–1517, 1986.

18 R. Gheissari, E. Lubetzky, and Y. Peres. Exponentially slow mixing in the mean-field Swendsen-Wang dynamics. In *Proceedings of the 29th Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pages 1981–1988, 2018.

19 W. R. Gilks, S. Richardson, and D. J. Spiegelhalter. *Markov chain Monte Carlo in practice*. London: Chapman and Hall, 1996.

20 V. K. Gore and M. R. Jerrum. The Swendsen-Wang process does not always mix rapidly. *Journal of Statistical Physics*, 97(1-2):67–86, 1999.

21 G. R. Grimmett. *The Random-Cluster Model*, volume 333. Springer-Verlag, 2009.

22 H. Guo and M. Jerrum. Random cluster dynamics for the Ising model is rapidly mixing. In *Proceedings of the 28th Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pages 1818–1827, 2017.

23 T. P. Hayes and A. Sinclair. A general lower bound for mixing of single-site dynamics on graphs. In *Proceedings of the 46th Annual IEEE Symposium on Foundations of Computer Science (FOCS)*, pages 511–520, 2005.

24 A. E. Holroyd. Some Circumstances Where Extra Updates Can Delay Mixing. *Journal of Statistical Physics*, 145(6):1649–1652, 2011.

25 E. Ising. Beitrag zur theorie des ferromagnetismus. *Zeitschrift für Physik*, 31(1):253–258, 1925.

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

27 M. Jerrum, A. Sinclair, and E. Vigoda. A polynomial-time approximation algorithm for the permanent of a matrix with nonnegative entries. *Journal of the ACM*, 51(4):671–697, 2004.

28 M. R. Jerrum, L. G. Valiant, and V. V. Vazirani. Random generation of combinatorial structures from a uniform distribution. *Theoretical Computer Science*, 43(2-3):169–188, 1986.

29 R. Kannan, L. Lovász, and M. Simonovits. Random walks and an $O^*(n^5)$ volume algorithm for convex bodies. *Random Structures & Algorithms*, 11(1):1–50, 1997.

30 D. P. Landau and K. A. Binder. *A guide to Monte Carlo simulations in statistical physics*. Cambridge University Press, 2014.

31 W. Lenz. Beiträge zum verstandnis der magnetischen eigenschaften in festen korpern. *Physikalische Zeitschrift*, 21:613–615, 1920.

32 D. A. Levin and Y. Peres. *Markov Chains and Mixing Times, 2nd edition*, volume 107. American Mathematical Society, 2017.

33 L. Li, P. Lu, and Y. Yin. Correlation decay up to uniqueness in spin systems. In *Proceedings of the 24th Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pages 67–84, 2013.

34 Y. Long, A. Nachmias, W. Ning, and Y. Peres. A power law of order 1/4 for critical mean-field Swendsen-Wang dynamics. *Memoirs of the American Mathematical Society*, 232(1092):84 pages, 2014.

35 F. Martinelli and E. Olivieri. Approach to equilibrium of Glauber dynamics in the one phase region. I. The attractive case. *Communications in Mathematical Physics*, 161(3):447–486, 1994.

36 F. Martinelli and E. Olivieri. Approach to equilibrium of Glauber dynamics in the one phase region. II. The general case. *Communications in Mathematical Physics*, 161(3):458–514, 1994.

37 A. Montanari and A. Saberi. The spread of innovations in social networks. *Proceedings of the National Academy of Sciences*, 107(47):20196–20201, 2010.

38 E. Mossel and A. Sly. Exact thresholds for Ising–Gibbs samplers on general graphs. *The Annals of Probability*, 41(1):294–328, 2013.

39 Y. Peres. Personal communication, 2016.

40 Y. Peres and P. Winkler. Can Extra Updates Delay Mixing? *Communications in Mathematical Physics*, 323(3):1007–1016, 2013.

41 C. J. Preston. Gibbs States on Countable Sets. *Cambridge Tracts in Mathematics*, 1974.

42 J. Salas and A. D. Sokal. Dynamic critical behavior of a Swendsen-Wang-type algorithm for the Ashkin-Teller model. *Journal of Statistical Physics*, 85(3-4):297–361, 1996.

43 J. Salas and A. D. Sokal. Dynamic critical behavior of the Swendsen-Wang algorithm: The two-dimensional three-state potts model revisited. *Journal of Statistical Physics*, 87(1-2):1–36, 1997.

44 A. Sinclair, P. Srivastava, D. Štefankovič, and Y. Yin. Spatial mixing and the connective constant: Optimal bounds. *Probability Theory and Related Fields*, 168(1-2):153–197, 2017.

45 A. Sinclair, P. Srivastava, and M. Thurley. Approximation algorithms for two-state anti-ferromagnetic spin systems on bounded degree graphs. *Journal of Statistical Physics*, 155(4):666–686, 2014.

46 A. Sly. Computational transition at the uniqueness threshold. In *Proceedings of the 51st Annual IEEE Symposium on Foundations of Computer Science (FOCS)*, pages 287–296, 2010.

47 A. Sly and N. Sun. The computational hardness of counting in two-spin models on d-regular graphs. In *Proceedings of the 53rd Annual IEEE Symposium on Foundations of Computer Science (FOCS)*, pages 361–369, 2012.

48 D. Štefankovič, S. Vempala, and E. Vigoda. Adaptive Simulated Annealing: A Near-optimal Connection between Sampling and Counting. *Journal of the ACM*, 56(3):1–36, 2009.

49 R. H. Swendsen and J. S. Wang. Nonuniversal critical dynamics in Monte Carlo simulations. *Physical Review Letters*, 58(2):86–88, 1987.

50 M. Ullrich. Rapid mixing of Swendsen-Wang and single-bond dynamics in two dimensions. *Dissertationes Mathematicae*, 2014.

51 J. S. Wang. Critical dynamics of the Swendsen-Wang algorithm in the three-dimensional Ising model. *Physica A*, 164(2):240–244, 1990.

52 D. Weitz. Counting independent sets up to the tree threshold. In *Proceedings of the 38th Annual ACM Symposium on Theory of Computing (STOC)*, pages 140–149, 2006.

A Proof of Theorem 11

In [38], Mossel and Sly show that ASSM (see Definition 9) implies optimal $O(n \log n)$ mixing of the Glauber dynamics on any n-vertex graph of bounded degree [23]. Our proof of Theorem 11 follows the approach in [38]. The key new novelty is the use of Theorem 7.

Proof of Theorem 11. Let $\{X_t^+\}$, $\{X_t^-\}$ be two instances of the chain such that X_0^+ is the "all plus" configuration and X_0^- is the "all minus" one. Since the chain is monotone there exists a monotone grand coupling of $\{X_t^+\}$ and $\{X_t^-\}$ such that $X_t^+ \geq X_t^-$ for all $t \geq 0$. The existence of a monotone grand coupling implies that the extremal "all plus" and "all minus" are the worst possible starting configurations, and thus

$$T_{\text{mix}}(P, \varepsilon) \leq T_{\text{coup}}(\varepsilon)$$

where $T_{\text{coup}}(\varepsilon)$ is the minimum t such that $\Pr[X_t^+ \neq X_t^-] \leq \varepsilon$, assuming $\{X_t^+\}$ and $\{X_t^-\}$ are coupled using the monotone coupling. Hence, it is sufficient to find t such that for all $v \in V$

$$\Pr[X_t^+(v) \neq X_t^-(v)] \leq \frac{\varepsilon}{n},$$

since the result would follow from a union bound over the vertices.

Choose $R \in \mathbb{N}$ such that ASSM holds; see Lemma 10. Let $s \in \mathbb{N}$ be arbitrary and fixed. For each $v \in V$, we define two instances $\{Y_t^+\}$ and $\{Y_t^-\}$ of the censored chain that until time s evolves as the chain P and after time s it evolves according to $P_{B(v,R)}$. By assumption

$P_{B(v,R)}$ is also monotone, so the evolutions of $\{Y_t^+\}$ and $\{Y_t^-\}$ can be coupled as follows: up to time s, $\{Y_t^+\}$ and $\{Y_t^-\}$ are coupled by setting $Y_t^+ = X_t^+$ and $Y_t^- = X_t^-$ for all $0 \le t \le s$; for $t > s$ the monotone coupling for $P_{B(v,R)}$ is used. Then, we have $X_t^+ \ge X_t^-$ and $Y_t^+ \ge Y_t^-$ for all $t \ge 0$.

Since $P \le P_{B(v,R)}$ by assumption, and the distribution ν^+ (resp., ν^-) of X_0^+ (resp., X_0^-) is such that ν^+/μ (resp., ν^-/μ) is trivially increasing (resp., decreasing), Theorem 7 implies $Y_t^+ \succeq X_t^+$ and $X_t^- \succeq Y_t^-$ for all $t \ge 0$. Hence,

$$Y_t^+ \succeq X_t^+ \succeq X_t^- \succeq Y_t^-.$$

Thus,

$$\begin{aligned}\Pr[X_t^+(v) \neq X_t^-(v)] &= \Pr[X_t^+(v) = +] - \Pr[X_t^-(v) = +] \\ &\le \Pr[Y_t^+(v) = +] - \Pr[Y_t^-(v) = +] \\ &= \Pr[Y_t^+(v) \neq Y_t^-(v)],\end{aligned}$$

where the first and third equations follow from the monotonicity of $\{X_t^+\}$, $\{X_t^-\}$, $\{Y_t^+\}$ and $\{Y_t^-\}$ and the inequality from the fact that $Y_t^+ \succeq X_t^+$ and $Y_t^- \preceq X_t^-$.

Recall our earlier definitions of $B(v, R)$ as the ball of radius R and $S(v, R)$ as the external boundary of $B(v, R)$; i.e., $B(v, R) = \{u \in V : \mathrm{dist}(u, v) \le R\}$ and let $S(v, R) = B(v, R + 1)\backslash B(v, R)$. For ease of notation let $A = B(v, R + 1) = B(v, R) \cup S(v, R)$ and for $\sigma^+, \sigma^- \in \Omega_A$ let $\mathcal{F}_s(\sigma^+, \sigma^-)$ be the event $\{X_s^+(A) = \sigma^+, X_s^-(A) = \sigma^-\}$. Then, for $t > s$ we have

$$\begin{aligned}\Pr[Y_t^+(v) \neq Y_t^-(v) \mid \mathcal{F}_s(\sigma^+, \sigma^-)] \le &\left|\Pr[Y_t^+(v) = + \mid \mathcal{F}_s(\sigma^+, \sigma^-)] - \mu(v = + \mid \tau^+)\right| \\ &+ \left|\mu(v = + \mid \tau^+) - \mu(v = + \mid \tau^-)\right| \\ &+ \left|\Pr[Y_t^-(v) = + \mid \mathcal{F}_s(\sigma^+, \sigma^-)] - \mu(v = + \mid \tau^-)\right|,\end{aligned} \tag{11}$$

where $\mu = \mu_{G,\beta}$, $\tau^+ = \sigma^+(S(v, R))$ and $\tau^- = \sigma^-(S(v, R))$.

Observe that $\mu(\cdot \mid \tau^+)$ and $\mu(\cdot \mid \tau^-)$ are the stationary measures of $\{Y_t^+\}$ and $\{Y_t^-\}$ respectively, and recall that by assumption

$$\max_{\sigma \in \Omega} T_{\mathrm{mix}}(P_{B(v,R)}, \sigma) \le T.$$

Hence, for $t = s + T \log_4\lceil 8|A|\rceil$, we have

$$\left|\Pr[Y_t^+(v) = + \mid \mathcal{F}_s(\sigma^+, \sigma^-)] - \mu(v = + \mid \tau^+)\right| \le \frac{1}{8|A|}, \tag{12}$$

and similarly

$$\left|\Pr[Y_t^-(v) = + \mid \mathcal{F}_s(\sigma^+, \sigma^-)] - \mu(v = + \mid \tau^-)\right| \le \frac{1}{8|A|}. \tag{13}$$

We bound next $|\mu(v = + \mid \tau^+) - \mu(v = + \mid \tau^-)|$. For $u \in S(v, R)$, let a_u be defined as in (5) and let $S(v, R) = \{u_1, u_2, \ldots, u_l\}$ with $l = |S(v, R)|$. Let $\tau_0, \tau_1, \ldots, \tau_l$ be a sequence of configurations on $S(v, R)$ such that $\tau_j(u_k) = \tau^+(u_k)$ for $j < k \le l$ and $\tau_j(u_k) = \tau^-(u_k)$ for

$1 \le k \le j$. That is, $\tau_0 = \tau^+$, $\tau_l = \tau^-$ and τ_j is obtained from τ_{j-1} by changing the spin of u_j from $\tau^+(u_j)$ to $\tau^-(u_j)$. The triangle inequality then implies that

$$\begin{aligned}\left|\mu(v=+\mid\tau^+)-\mu(v=+\mid\tau^-)\right| &\le \sum_{j=1}^{l}\left|\mu(v=+\mid\tau_{j-1})-\mu(v=+\mid\tau_j)\right| \\ &\le \sum_{j=1}^{l}\mathbb{1}\{\tau^+(u_j)\neq\tau^-(u_j)\}\cdot a_{u_j} \\ &= \sum_{u\in S(v,R)}\mathbb{1}\{\sigma^+(u)\neq\sigma^-(u)\}\cdot a_u. \qquad (14)\end{aligned}$$

Hence, plugging (12), (13) and (14) into (11), we get

$$\Pr[Y_t^+(v)\neq Y_t^-(v)\mid\mathcal{F}_s(\sigma^+,\sigma^-)] \le \frac{1}{4|A|} + \sum_{u\in S(v,R)}\mathbb{1}\{\sigma^+(u)\neq\sigma^-(u)\}\cdot a_u.$$

Now, if $X_s^+(A) = X_s^-(A)$, then $Y_t^+(A) = Y_t^-(A)$ for all $t \ge s$. Therefore,

$$\begin{aligned}&\Pr[Y_t^+(v)\neq Y_t^-(v)] = \sum_{\sigma^+\neq\sigma^-\in\Omega_A}\Pr[Y_t^+(v)\neq Y_t^-(v)\mid\mathcal{F}_s(\sigma^+,\sigma^-)]\Pr[\mathcal{F}_s(\sigma^+,\sigma^-)] \\ &\le \frac{\Pr[X_s^+(A)\neq X_s^-(A)]}{4|A|} + \sum_{\sigma^+\neq\sigma^-\in\Omega_A}\sum_{u\in S(v,R)}\mathbb{1}\{\sigma^+(u)\neq\sigma^-(u)\}\cdot a_u\cdot\Pr[\mathcal{F}_s(\sigma^+,\sigma^-)] \\ &= \frac{\Pr[X_s^+(A)\neq X_s^-(A)]}{4|A|} + \sum_{u\in S(v,R)}\Pr[X_s^+(u)\neq X_s^-(u)]\cdot a_u.\end{aligned}$$

By union bound,

$$\frac{\Pr[X_s^+(A)\neq X_s^-(A)]}{4|A|} \le \frac{1}{4|A|}\sum_{u\in A}\Pr[X_s^+(u)\neq X_s^-(u)] \le \frac{1}{4}\max_{u\in V}\Pr[X_s^+(u)\neq X_s^-(u)].$$

Moreover, the ASSM property (see Lemma 10) implies that

$$\begin{aligned}\sum_{u\in S(v,R)}\Pr[X_s^+(u)\neq X_s^-(u)]\cdot a_u &\le \max_{u\in V}\Pr[X_s^+(u)\neq X_s^-(u)]\sum_{u\in S(v,R)}a_u \\ &\le \frac{1}{4}\max_{u\in V}\Pr[X_s^+(u)\neq X_s^-(u)].\end{aligned}$$

Thus, we conclude that for every $v\in V$

$$\Pr[X_t^+(v)\neq X_t^-(v)] \le \Pr[Y_t^+(v)\neq Y_t^-(v)] \le \frac{1}{2}\max_{u\in V}\Pr[X_s^+(u)\neq X_s^-(u)]$$

for $t = s + T\log_4\lceil 8|A|\rceil$. Taking the maximum over v

$$\max_{v\in V}\Pr[X_t^+(v)\neq X_t^-(v)] \le \frac{1}{2}\max_{v\in V}\Pr[X_s^+(v)\neq X_s^-(v)].$$

Iteratively, we get that for $\hat{T} = T\log_4\lceil 8|A|\rceil\log_2\lceil\frac{n}{\varepsilon}\rceil$

$$\max_{v\in V}\Pr[X_{\hat{T}}^+(v)\neq X_{\hat{T}}^-(v)] \le \frac{\varepsilon}{n}.$$

This implies that $T_{\rm mix}(P,\varepsilon) \le T\log_4\lceil 8|A|\rceil\log_2\lceil\frac{n}{\varepsilon}\rceil$, so taking $\varepsilon = 1/4$ it follows that $T_{\rm mix}(P) = O(T\log n)$ as desired. Moreover, since for $\varepsilon > 0$

$$(T_{\rm rel}(P)-1)\log(2\varepsilon)^{-1} \le T_{\rm mix}(P,\varepsilon),$$

taking $\varepsilon = n^{-1}$ yields that $T_{\rm rel}(P) = O(T)$; see Theorem 12.5 in [32]. ◀

B Proof of Theorem 7

Proof of Theorem 7. By assumption, X_t has distribution νP^t while $\hat{X}_t$ has distribution $\nu\hat{P}^t$ where $\hat{P}^t = P_{A_1}\dots P_{A_t}$. Since $\{P_A\}_{A\subseteq V}$ is a censoring for P, we have $P \le P_A$ for all $A \subseteq V$. We show first that this implies $P^t \le \hat{P}^t$.

Recall that P_{A_i} may be viewed as an operator from $L_2(\mu)$ to $L_2(\mu)$. The reversibility of P_{A_i} w.r.t. μ implies that P_{A_i} is self-adjoint; i.e., $P^*_{A_i} = P_{A_i}$. Also, since P is monotone, $P^k f$ is increasing for any integer $k > 0$ and any increasing function f; see Proposition 22.7 in [32]. Combining these facts, we have that for any pair of increasing positive functions $f, g : \mathbb{R}^{|\Omega|} \to \mathbb{R}$

$$\langle f, P^t g\rangle_\mu = \langle f, P(P^{t-1}g)\rangle_\mu \le \langle f, P_{A_1}(P^{t-1}g)\rangle_\mu = \langle P_{A_1}f, P^{t-1}g\rangle_\mu.$$

Note also that P_{A_1} is monotone, so $P_{A_1}f$ is increasing. Iterating this argument, we obtain

$$\langle f, P^t g\rangle_\mu \le \langle P_{A_1}f, P^{t-1}g\rangle_\mu \le \dots \le \langle P_{A_t}\dots P_{A_1}f, g\rangle_\mu = \langle f, \hat{P}^t g\rangle_\mu.$$

This shows that $P^t \le \hat{P}^t$.

To prove $X_t \preceq \hat{X}_t$, we need to show that for any increasing function g

$$\sum_{\sigma\in\Omega} \nu P^t(\sigma)g(\sigma) \le \sum_{\sigma\in\Omega} \nu\hat{P}^t(\sigma)g(\sigma). \tag{15}$$

Let $h : \mathbb{R}^{|\Omega|} \to \mathbb{R}$ be the function given by $h(\tau) = \nu(\tau)/\mu(\tau)$ for $\tau \in \Omega$. Then we have

$$\begin{aligned}\sum_{\sigma\in\Omega} \nu P^t(\sigma)g(\sigma) &= \sum_{\sigma\in\Omega}\Big(\sum_{\tau\in\Omega}\nu(\tau)P^t(\tau,\sigma)\Big)g(\sigma) = \sum_{\sigma,\tau\in\Omega}\nu(\tau)P^t(\tau,\sigma)g(\sigma)\\ &= \sum_{\sigma,\tau\in\Omega}\mu(\tau)P^t(\tau,\sigma)g(\sigma)h(\tau) = \langle h, P^t g\rangle_\mu.\end{aligned}$$

Similarly,

$$\sum_{\sigma\in\Omega} \nu\hat{P}^t(\sigma)g(\sigma) = \langle h, \hat{P}^t g\rangle_\mu.$$

The function h is increasing by assumption, and thus (15) follows immediately from the fact that $P^t \le \hat{P}^t$. This establishes part 1 of the theorem. Part 2 of the theorem follows from part 1 and Lemma 2.4 in [40]. ◀