Optimal Mixing via Tensorization for Random Independent Sets on Arbitrary Trees

Charilaos Efthymiou ✉
Computer Science, University of Warwick, Coventry, UK

Thomas P. Hayes ✉
Dept. of Computer Science and Engineering, University at Buffalo, NY, USA

Daniel Štefankovič ✉
Department of Computer Science, University of Rochester, NY, USA

Eric Vigoda ✉
Department of Computer Science, University of California, Santa Barbara, CA, USA

Abstract

We study the mixing time of the single-site update Markov chain, known as the Glauber dynamics, for generating a random independent set of a tree. Our focus is obtaining optimal convergence results for arbitrary trees. We consider the more general problem of sampling from the Gibbs distribution in the hard-core model where independent sets are weighted by a parameter $\lambda > 0$; the special case $\lambda = 1$ corresponds to the uniform distribution over all independent sets. Previous work of Martinelli, Sinclair and Weitz (2004) obtained optimal mixing time bounds for the complete Δ-regular tree for all λ. However, Restrepo et al. (2014) showed that for sufficiently large λ there are bounded-degree trees where optimal mixing does not hold. Recent work of Eppstein and Frishberg (2022) proved a polynomial mixing time bound for the Glauber dynamics for arbitrary trees, and more generally for graphs of bounded tree-width.

We establish an optimal bound on the relaxation time (i.e., inverse spectral gap) of $O(n)$ for the Glauber dynamics for unweighted independent sets on arbitrary trees. Moreover, for $\lambda \leq .44$ we prove an optimal mixing time bound of $O(n \log n)$. We stress that our results hold for arbitrary trees and there is no dependence on the maximum degree Δ. Interestingly, our results extend (far) beyond the uniqueness threshold which is on the order $\lambda = O(1/\Delta)$. Our proof approach is inspired by recent work on spectral independence. In fact, we prove that spectral independence holds with a constant independent of the maximum degree for any tree, but this does not imply mixing for general trees as the optimal mixing results of Chen, Liu, and Vigoda (2021) only apply for bounded degree graphs. We instead utilize the combinatorial nature of independent sets to directly prove approximate tensorization of variance/entropy via a non-trivial inductive proof.

2012 ACM Subject Classification Theory of computation → Randomness, geometry and discrete structures; Mathematics of computing → Discrete mathematics; Theory of computation → Design and analysis of algorithms

Keywords and phrases MCMC, Mixing Time, Independent Sets, Hard-Core Model, Approximate Counting Algorithms, Sampling Algorithms

Digital Object Identifier 10.4230/LIPIcs.APPROX/RANDOM.2023.33

Category RANDOM

Related Version *Full Version*: `https://arxiv.org/pdf/2307.07727.pdf`

Funding *Charilaos Efthymiou*: EPSRC New Investigator Award (grant no. EP/V050842/1) and Centre of Discrete Mathematics and Applications (DIMAP), The University of Warwick.
Daniel Štefankovič: Supported by NSF grant CCF-1563757.
Eric Vigoda: Supported by NSF grant CCF-2147094.

Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques (APPROX/RANDOM 2023).
Editors: Nicole Megow and Adam D. Smith; Article No. 33; pp. 33:1–33:16
Leibniz International Proceedings in Informatics
LIPICS Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

1 Introduction

This paper studies the mixing time of the Glauber dynamics for the hard-core model assuming that the underlying graph is an *arbitrary* tree. In the hard-core model, we are given a graph $G = (V, E)$ and an activity $\lambda > 0$. The model is defined on the collection of all independent sets of G (regardless of size), which we denote as Ω.

Each independent set $\sigma \in \Omega$ is assigned a weight $w(\sigma) = \lambda^{|\sigma|}$ where $|\sigma|$ is the number of vertices contained in the independent set σ. The Gibbs distribution μ is defined on Ω: for $\sigma \in \Omega$, let $\mu(\sigma) = w(\sigma)/Z$ where $Z = \sum_{\tau\in\Omega} w(\tau)$ is known as the partition function. When $\lambda = 1$ then every independent set has weight one and hence the Gibbs distribution μ is the uniform distribution over (unweighted) independent sets.

Our goal is to sample from μ (or estimate Z) in time polynomial in $n = |V|$. Our focus is on trees. These sampling and counting problems are computationally easy on trees using dynamic programming algorithms. Nevertheless, our interest is to understand the convergence properties of a simple Markov Chain Monte Carlo (MCMC) algorithm known as the Glauber dynamics for sampling from the Gibbs distribution.

The Glauber dynamics (also known as the Gibbs sampler) is the simple single-site update Markov chain for sampling from the Gibbs distribution of a graphical model. For the hard-core model with activity λ, the transitions $X_t \to X_{t+1}$ of the Glauber dynamics are defined as follows: first, choose a random vertex v. Then, with probability $\frac{\lambda}{1+\lambda}$ set $X' = X_t \cup \{v\}$ and with the complementary probability set $X' = X_t \setminus \{v\}$. If X' is an independent set, then set $X_{t+1} = X'$ and otherwise set $X_{t+1} = X_t$.

We consider two standard notions of convergence to stationarity. The *relaxation time* is the inverse spectral gap, i.e., $(1 - \lambda^*)^{-1}$ where $\lambda^* = \max\{\lambda_2, |\lambda_N|\}$ and $1 = \lambda_1 > \lambda_2 \geq \cdots \geq \lambda_N > -1$ are the eigenvalues of the transition matrix P for the Glauber dynamics. The relaxation time is a key quantity in the running time for approximate counting algorithms (see, e.g., [29]). The *mixing time* is the number of steps, from the worst initial state, to reach within total variation distance $\leq 1/2e$ of the stationary distribution, which in our case is the Gibbs distribution μ.

We say that $O(n)$ is the optimal relaxation time and that $O(n \log n)$ is the optimal mixing time (see Hayes and Sinclair [18] for a matching lower bound for any constant degree graph). Here, n denotes the size of the underlying graph. More generally, we say the Glauber dynamics is rapidly mixing when the mixing time is poly(n).

We establish bounds on the mixing time of the Glauber dynamics by means of *approximate tensorization inequalities* for the variance and the entropy of the hard-core model. Interestingly, our analysis utilizes nothing further than the inductive nature of the tree, e.g., we do not make any assumptions about spatial mixing properties of the Gibbs distribution. As a consequence, the bounds we obtain have no dependence on the maximum degree of the graph.

To be more specific we derive the following two group of results: We establish approximate tensorization of variance of the hard-core model on the tree for all $\lambda < 1.1$. This implies *optimal* $O(n)$ relaxation time for the Glauber dynamics. Notably this also includes the uniform distribution over independent sets, i.e., $\lambda = 1$. Furthermore, we establish approximate tensorization of entropy for the hard-core model on any tree for all $\lambda < 0.44$. In turn, this implies *optimal* mixing time $O(n \log n)$ for the Glauber dynamics.

We can now state our main results.

▶ **Theorem 1.** *For any n-vertex tree, for any $\lambda < 1.1$ the Glauber dynamics for sampling λ-weighted independent sets in the hard-core model has an optimal relaxation time of $O(n)$.*

Moreover, when $\lambda \leq 0.44$ then we can prove an optimal bound on the mixing time. (This extends to $\lambda \leq 1.05$ under an intriguing conjecture that we can numerically verify.)

▶ **Theorem 2.** *For any n-vertex tree, for any $\lambda \leq 0.44$ the Glauber dynamics for sampling λ-weighted independent sets in the hard-core model has an optimal mixing time of $O(n \log n)$.*

We believe the optimal mixing results of Theorems 1 and 2 are related to the reconstruction threshold, which we describe now. Consider the complete Δ-regular tree of height h; this is the rooted tree where all nodes at distance $\ell < h$ from the root have $\Delta - 1$ children and all nodes at distance h from the root are leaves. We are interested in how the configuration at the leaves affects the configuration at the root.

Consider fixing an assignment/configuration σ to the leaves (i.e., specifying which leaves are fixed to occupied and which are unoccupied), we refer to this fixed assignment to the leaves as a boundary condition σ. Let μ_σ denote the Gibbs distribution conditional on this fixed boundary condition σ, and let p_σ denote the marginal probability that the root is occupied in μ_σ.

The uniqueness threshold $\lambda_c(\Delta)$ measures the affect of the *worst-case* boundary condition on the root. For all $\lambda < \lambda_c(\Delta)$, all $\sigma \neq \sigma'$, in the limit $h \to \infty$, we have $p_\sigma = p'_\sigma$; this is known as the (tree) uniqueness region. In contrast, for $\lambda > \lambda_c(\Delta)$ there are pairs $\sigma \neq \sigma'$ (namely, all even occupied vs. odd occupied) for which the limits are different; this is the non-uniqueness region. The uniqueness threshold is at $\lambda_c(\Delta) = (\Delta - 1)^{\Delta-1}/(\Delta - 2)^{\Delta} = O(1/\Delta)$.

In contrast, the reconstruction threshold $\lambda_r(\Delta)$ measures the affect on the root of a random/typical boundary condition. In particular, we fix an assignment c at the root and then generate the Gibbs distribution via an appropriately defined broadcasting process. Finally, we fix the boundary configuration σ and ask whether, in the conditional Gibbs distribution μ_σ, the root has a bias towards the initial assignment c. The non-reconstruction region $\lambda < \lambda_r(\Delta)$ corresponds to when we cannot infer the root's initial value, in expectation over the choice of the boundary configuration σ and in the limit $h \to \infty$, see Mossel [24] for a more complete introduction to reconstruction.

The reconstruction threshold is not known exactly but close bounds were established by Bhatnagar, Sly, and Tetali [3] and Brightwell and Winkler [5] who showed that: $C_1 \log^2 \Delta / \log \log \Delta \leq \lambda_r(\Delta) \leq C_2 \log^2 \Delta$ for sufficiently large Δ, and hence $\lambda_r(\Delta)$ is "increasing asymptotically" with Δ whereas the uniqueness threshold is a decreasing function of Δ. Martin [21] showed that $\lambda_r(\Delta) > e - 1$ for all Δ. As a consequence, we conjecture that Theorems 1 and 2 holds for all trees for all $\lambda < e - 1$, which is close to the bound we obtain. A slowdown in the reconstruction region is known: as described below, Restrepo et al. [26] showed that there are trees for which there is a polynomial slow down for $\lambda > C$ for a constant $C > 0$; an explicit constant C is not stated in [26] but using the Kesten-Stigum bound one can show $C \approx 28$.

For general graphs the appropriate threshold is the uniqueness threshold, which is $\lambda_c(\Delta) = O(1/\Delta)$. In particular, for bipartite random Δ-regular graphs the Glauber dynamics has optimal mixing in the uniqueness region [13], and is exponentially slow in the non-uniqueness region [25, 17]. Moreover, for general graphs there is a computational phase transition at the uniqueness threshold: optimal mixing on all graphs of maximum degree Δ in the uniqueness region [13, 9, 10], and NP-hardness to approximately count/sample in the non-uniqueness region [27, 17, 28].

There are a variety of mixing results for the special case on trees, which is the focus of this paper. In terms of establishing optimal mixing time bounds for the Glauber dynamics, previous results only applied to complete, Δ-regular trees. Seminal work of Martinelli, Sinclair, and Weitz [22, 23] proved optimal mixing on complete Δ-regular trees for all λ. The intuitive reason this holds for all λ is that the complete tree corresponds to one of the two extremal phases (all even boundary or all odd boundary) and hence it does not exhibit the phase co-existence which causes mixing. As mentioned earlier, Restrepo et al. [26] shows that there is a there is a fixed assignment τ for the leaves of the complete, Δ-regular tree so that the mixing time slows down in the reconstruction region; intuitively, this boundary condition τ corresponds to the assignment obtained by the broadcasting process.

For more general trees the following results were known. A classical result of Berger et al. [2] proves polynomial mixing time for trees with constant maximum degree [2]. A very recent result of Eppstein and Frishberg [16] proved polynomial mixing time $n^{C(\lambda)}$ of the Glauber dynamics for graphs with bounded tree-width which includes arbitrary trees, however the polynomial exponent is roughly $C(\lambda) = 11 + 6\log(\lambda)$ for trees. For other combinatorial models, rapid mixing for the Glauber dynamics on trees with bounded maximum degree was established for k-colorings in [20] and edge-colorings in [15].

Spectral independence is a powerful notion in the analysis of the convergence rate of Markov Chain Monte Carlo (MCMC) algorithms. For independent sets on an n-vertex graph $G = (V, E)$, spectral independence considers the $n \times n$ pairwise influence matrix I_G where $I_G(v, w) = \mathrm{Prob}_{\sigma\sim\mu}(v \in \sigma \mid w \in \sigma) - \mathrm{Prob}_{\sigma\sim\mu}(v \in \sigma \mid w \notin \sigma)$; this matrix is closely related to the vertex covariance matrix. We say that spectral independence holds if the maximum eigenvalue of $I_{G'}$ for all vertex-induced subgraphs G' of G are bounded by a constant. Spectral independence was introduced by Anari, Liu, and Oveis Gharan [1]. Chen, Liu, and Vigoda [13] proved that spectral independence, together with a simple condition known as marginal boundedness which is a lower bound on the marginal probability of a valid vertex-spin pair, implies optimal mixing time of the Glauber dynamics for *constant-degree graphs*. This has led to a flurry of optimal mixing results, e.g., [14, 4, 19, 12, 11].

The limitation of the above spectral independence results is that they only hold for graphs with constant maximum degree Δ. There are several noteworthy results that achieve a stronger form of spectral independence which establishes optimal mixing for unbounded degree graphs [9, 10]; however all of these results for general graphs only achieve rapid mixing in the tree uniqueness region which corresponds to $\lambda = O(1/\Delta)$ whereas we are aiming for $\lambda = \Theta(1)$.

The inductive approach we use to establish approximate tensorization inequalities can also be utilized to establish spectral independence. In fact, we show that spectral independence holds for any tree when $\lambda < 1.3$, including the case where $\lambda = 1$, see Appendix A of the full version of this paper. Applying the results of Anari, Liu, and Oveis Gharan [1] we obtain a poly(n) bound on the mixing time, but with a large constant in the exponent of n. For constant degree trees we obtain the following optimal mixing result by applying the results of Chen, Liu, and Vigoda [13] (see also [4, 9, 10]).

▶ **Theorem 3.** *For all constant Δ, all $\lambda \le 1.3$, for any tree T with maximum degree Δ, the Glauber dynamics for sampling λ-weighted independent sets has an optimal mixing time of $O(n \log n)$.*

In the next section we recall the key functional definitions and basic properties of variance/entropy that will be useful later in the proofs. In Section 3 we prove approximate tensorization of variance which establishes Theorem 1. Then in Section 4 we prove Theorem 2. We establish spectral independence and prove Theorem 3 in Appendix A of the full version of this paper.

2 Preliminaries

2.1 Standard Definitions

Let P be the transition matrix of a Markov chain $\{X_t\}$ with a finite state space Ω and equilibrium distribution μ. For $t \geq 0$ and $\sigma \in \Omega$, let $P^t(\sigma, \cdot)$ denote the distribution of X_t when the initial state of the chain satisfies $X_0 = \sigma$. The *mixing time* of the Markov chain $\{X_t\}_{t \geq 0}$ is defined by

$$T_{\text{mix}} = \max_{\sigma \in \Omega} \min \left\{ t > 0 \mid \| P^t(\sigma, \cdot) - \mu \|_{\text{TV}} \leq \frac{1}{2\text{e}} \right\} . \tag{1}$$

The transition matrix P with stationary distribution μ is called *time reversible* if it satisfies the so-called *detailed balance relation*, i.e., for any $x, y \in \Omega$ we have $\mu(x)P(x, y) = P(y, x)\mu(y)$. For P that is time reversible the set of eigenvalues are real numbers and we denote them as $1 = \lambda_1 \geq \lambda_2 \geq \ldots \lambda_{|\Omega|} \geq -1$. Let $\lambda^* = \max\{|\lambda_2|, |\lambda_{|\Omega|}|\}$, then we define the *relaxation time* T_{relax} by

$$T_{\text{relax}}(P) = \frac{1}{1 - \lambda^*} . \tag{2}$$

The quantity $1 - \lambda^*$ is also known as the *spectral gap* of P. We use T_{relax} to bound T_{mix} by using the following inequality

$$T_{\text{mix}}(P) \leq T_{\text{relax}}(P) \cdot \log \left(\frac{2e}{\min_{x \in \Omega} \mu(x)} \right) . \tag{3}$$

2.2 Gibbs Distributions and Functional Analytic Definitions

For a graph $G = (V, E)$ and $\lambda > 0$, let $\mu = \mu_{G,\lambda}$ be the hard-core model on this graph with activity λ, while let $\Omega \subseteq 2^V$ be the support of μ. For any $\Lambda \subseteq V$ and any $\tau \subseteq \Lambda$, we let $\mu^{\Lambda,\tau}$ be the distribution μ conditional on that from Λ we choose exactly the vertices in τ. When there is no danger of confusion, we omit Λ. We let $\Omega^\tau \subseteq \Omega$ be the support of $\mu^{\Lambda,\tau}$, while we call τ *feasible* if Ω^τ is nonempty.

For any subset $S \subseteq V$, let μ_S denote the marginal of μ at S, while let $\Omega_S \subseteq 2^S$ denote the support of μ_S. That is, for any $\sigma \subseteq S$, we have that

$$\mu_S(\sigma) = \sum_{\eta \in 2^V} \mathbf{1}\{\eta \cap S = \sigma\} \mu(\eta) . \tag{4}$$

In a natural way, we define the conditional marginal. That is, for $\Lambda \subseteq V \setminus S$ and $\tau \subseteq \Lambda$, we let $\mu_S^{\Lambda,\tau}$ denote the marginal at S conditional on the configuration at Λ being τ. Similarly to what we had before, when there is no danger of confusion, we omit Λ. We let Ω_S^τ denote the support of μ_S^τ.

For any function $f : \Omega \to \mathbb{R}_{\geq 0}$, we let $\mu(f)$ is the expected value of f with respect to μ, i.e.,

$$\mu(f) = \sum_{\sigma \in \Omega} \mu(\sigma) f(\sigma) .$$

Analogously, we define variance of f with respect to μ by

$$\text{Var}(f) = \mu(f^2) - (\mu(f))^2 . \tag{5}$$

We are also using the following equation for $\mathrm{Var}(f)$,

$$\mathrm{Var}(f) = \tfrac{1}{2}\textstyle\sum_{\sigma,\tau\in\Omega}\mu(\sigma)\mu(\tau)\left(f(\sigma)-f(\tau)\right)^2 \quad . \tag{6}$$

For any $S \subseteq V$, for any $\tau \in \Omega_{V\setminus S}$, we define the function $f_\tau : \Omega_S^\tau \to \mathbb{R}_{\geq 0}$ such that $f_\tau(\sigma) = f(\tau \cup \sigma)$ for all $\sigma \in \Omega_S^\tau$. Let $\mathrm{Var}_S^\tau(f_\tau)$ denote the variance of f_τ with respect to the conditional distribution μ_S^τ:

$$\mathrm{Var}_S^\tau(f_\tau) = \mu_S^\tau(f_\tau^2) - \left(\mu_S^\tau(f_\tau)\right)^2 \tag{7}$$

$$= \frac{1}{2}\sum_{\sigma,\eta\in\Omega}\frac{\mathbf{1}\{\sigma\setminus S=\tau,\ \eta\setminus S=\tau\}\mu(\sigma)\mu(\eta)}{\left(\sum_{\hat{\sigma}\in\Omega}\mathbf{1}\{\hat{\sigma}\setminus S=\tau\}\mu(\hat{\sigma})\right)^2}\left(f(\sigma)-f(\eta)\right)^2 \quad . \tag{8}$$

Furthermore, we let

$$\mu(\mathrm{Var}_S(f)) = \sum_{\tau\in\Omega_{V\setminus S}}\mu_{V\setminus S}(\tau)\cdot\mathrm{Var}_S^\tau(f_\tau) \quad , \tag{9}$$

i.e., $\mu(\mathrm{Var}_S(f))$ is the average of $\mathrm{Var}_S^\tau(f_\tau)$ with respect to the τ being distributed as in $\mu_{V\setminus S}(\cdot)$. For the sake of brevity, when $S=\{v\}$, i.e., the set S is a singleton, we use $\mu(\mathrm{Var}_v(f))$.

Similarly to $\mu(f)$ and $\mathrm{Var}(f)$ we define the entropy with respect to μ by

$$\mathrm{Ent}(f) = \mu\left(f\log\tfrac{f}{\mu(f)}\right) \quad , \tag{10}$$

where we use the convention that $0\log 0 = 0$. Analogously to $\mu(\mathrm{Var}_S(f))$, we let

$$\mu(\mathrm{Ent}_S(f)) = \sum_{\tau\in\Omega_{V\setminus S}}\mu_{V\setminus S}(\tau)\mathrm{Ent}_S^\tau(f_\tau) \quad . \tag{11}$$

That is, $\mu(\mathrm{Ent}_S(f))$ is the average of the entropy $\mathrm{Ent}_S^\tau(f_\tau)$ with respect to the measure $\mu_{V\setminus S}(\cdot)$.

When X is a Bernoulli random variable, i.e.,

$$X = \begin{cases} A & \text{with probability } p \\ B & \text{with probability } 1-p, \end{cases}$$

one formulation for the variance that will be convenient for us is

$$\mathrm{Var}(X) = p(1-p)(A-B)^2. \tag{12}$$

2.3 Approximate Tensorization of Variance/Entropy

To bound the convergence rate of the Glauber dynamics we consider the approximate tensorization of variance/entropy as introduced in [7].

We begin with the definition of approximate tensorization of variance.

▶ **Definition 4** (Variance Tensorization). *A distribution μ with support $\Omega \subseteq \{\pm 1\}^V$ satisfies the approximate tensorisation of Variance with constant $C > 0$, denoted using the predicate $VT(C)$, if for all $f : \Omega \to \mathbb{R}_{\geq 0}$ we have that*

$$\mathrm{Var}(f) \leq C \cdot \sum_{v\in V}\mu\left(\mathrm{Var}_v(f)\right) \quad .$$

For a vertex x, recall that $\mathrm{Var}_x[f] = \sum_\sigma \mu_{V\setminus\{x\}}(\sigma)\mathrm{Var}_x^\sigma[f_\sigma]$. Variance tensorization $VT(C)$ yields the following bound on the relaxation time of the Glauber dynamics [7, 6]:

$$T_{\text{relax}} \le Cn. \tag{13}$$

We continue with the analog for entropy, which is the key step in our proofs establishing optimal mixing bounds of the Glauber dynamics.

▶ **Definition 5** (Entropy Tensorization). *A distribution μ with support $\Omega \subseteq \{\pm 1\}^V$ satisfies the approximate tensorisation of Entropy with constant $C > 0$, denoted using the predicate $ET(C)$, if for all $f : \Omega \to \mathbb{R}_{\ge 0}$ we have that*

$$\mathrm{Ent}(f) \le C \cdot \sum_{v\in V} \mu\left(\mathrm{Ent}_v(f)\right) \quad .$$

For a vertex x, recall that $\mathrm{Ent}_x[f] = \sum_\sigma \mu_{V\setminus\{x\}}(\sigma)\mathrm{Ent}_x^\sigma[f_\sigma]$. Entropy tensorization $ET(C)$ immediately yields the following mixing time bound for the Glauber dynamics [7, 6]:

$$T_{\text{mix}} \le Cn\left(\log(\log(1/\mu^*)) + \log 2 + 2\right) \quad . \tag{14}$$

2.4 Decomposition of Variance/Entropy

We use the following basic decomposition properties for entropy and variance. The following lemma follows from a decomposition of relative entropy, see [8, Lemma 3.1] (see also [4, Lemma 2.3]).

▶ **Lemma 6.** *For any $S \subset V$, for any $f \ge 0$:*

$$\mathrm{Var}(f) = \mu[\mathrm{Var}_S(f)] + \mathrm{Var}(\mu_S(f)) \quad , \tag{15}$$

$$\mathrm{Ent}(f) = \mu[\mathrm{Ent}_S(f)] + \mathrm{Ent}(\mu_S(f)) \quad . \tag{16}$$

We utilize the following decomposition of a product measure, see [6, Eqn (4.7)].

▶ **Lemma 7.** *Consider $U, W \subset V$ where $\mathrm{dist}(U, W) \ge 2$. Then for all $f \ge 0$ we have:*

$$\mu[\mathrm{Var}_U(\mu_W f)] \le \mu[\mathrm{Var}_U(f)] \quad , \tag{17}$$

$$\mu[\mathrm{Ent}_U(\mu_W f)] \le \mu[\mathrm{Ent}_U(f)] \quad . \tag{18}$$

Proof. We apply [6, Eqn (4.7)], which reaches the same conclusion under the assumptions that $U \cap W = \emptyset$ and $\mu_U\mu_W = \mu_W\mu_U$. In the current context, the reason these conditional expectation operators commute here is that, because $\mathrm{dist}(U, W) \ge 2$, if we let S be an independent set sampled according to distribution μ, then the random variables $S \cap U$ and $S \cap W$ are conditionally independent given $S \setminus (U \cup W)$. ◀

3 Variance Factorization

In this section we prove Theorem 1, establishing an optimal bound on the relaxation time for the Glauber dynamics on any tree for $\lambda < 1.1$. We will prove this by establishing variance tensorization, see Definition 4.

Let $T' = (V', E')$ be a tree, let $\{\lambda'_w\}_{w \in V'}$ be a collection of fugacities and let μ' be the corresponding hard-core measure. We will establish the following variance tensorization inequality: for all $f' : 2^{V'} \to \mathbb{R}$

$$\mathrm{Var}(f') \leq \sum_{x \in V'} F(\lambda'_x)\mu'(\mathrm{Var}_x(f')), \tag{19}$$

where $F : \mathbb{R}_{\geq 0} \to \mathbb{R}_{\geq 0}$ is a function to be determined later (in Lemma 8). We refer to $\mathrm{Var}(f')$ as the "global" variance and we refer to $\mu'(\mathrm{Var}_x(f'))$ as the "local" variance (of f' at x).

We will establish (19) using induction. Let v be a vertex of degree 1 in T' and let u be the unique neighbor of v. Let $T = (V, E)$ be the tree which is the induced subgraph of G' on $V = V' \setminus \{v\}$. Let $\{\lambda_w\}_{w \in V}$ be a collection of fugacities where $\lambda_w = \lambda'_w$ for $w \neq u$ and $\lambda_u = \lambda'_u/(1 + \lambda'_v)$. Let μ be the hard-core measure on T with fugacities $\{\lambda_w\}_{w \in V}$.

Note that for $S \subseteq V$ we have

$$\mu(S) = \mu'(S) + \mu'(S \cup \{v\}) = \mu'_V(S). \tag{20}$$

Fix a function $f' : 2^{V'} \to \mathbb{R}$. Let $f : 2^V \to \mathbb{R}$ be defined by

$$f(S) = \frac{\mu'(S)f'(S) + \mu'(S \cup \{v\})f'(S \cup \{v\})}{\mu'(S) + \mu'(S \cup \{v\})} = \mathbb{E}_{Z \sim \mu'}[f'(Z) \mid Z \cap V = S] = \mu'_v(f')(S). \tag{21}$$

Note that f' is defined on independent sets of the tree T' and f is the natural projection of f' to the tree T. Since $f = \mu'_v(f')$, then by Lemma 6 we have that:

$$\mathrm{Var}(f') = \mu'(\mathrm{Var}_v(f')) + \mathrm{Var}(f). \tag{22}$$

When we condition on the configuration at u then μ' becomes a product measure on $V \setminus \{u\}$ and $\{v\}$. Hence, from Equation (17) then for any $x \notin \{u, v\}$ (by setting $U = \{x\}$ and $W = \{v\}$) we have:

$$\mu'(\mathrm{Var}_x(f)) \leq \mu'(\mathrm{Var}_x(f')).$$

Since by (20) we have $\mu(\mathrm{Var}_x(f)) = \mu'(\mathrm{Var}_x(f))$, the above implies that

$$\mu(\mathrm{Var}_x(f)) \leq \mu'(\mathrm{Var}_x(f')). \tag{23}$$

The following lemma is the main technical ingredient. It bounds the local variance at u for the smaller graph T in terms of the local variance at u and v in the original graph T'.

▶ **Lemma 8.** *For $F(x) = 1000(1 + x)^2 \exp(1.3x)$ and any $\lambda_v, \lambda_u \in (0, 1.1]$ we have:*

$$F(\lambda_u)\mu(\mathrm{Var}_u(f)) \leq (F(\lambda'_v) - 1)\mu'(\mathrm{Var}_v(f')) + F(\lambda'_u)\mu'(\mathrm{Var}_u(f')). \tag{24}$$

We now utilize the above lemma to prove the main theorem for the relaxation time. We then go back to prove Lemma 8.

Proof of Theorem 1. Note Equation (24) is equivalent to:

$$\mu'(\mathrm{Var}_v(f')) + F(\lambda_u)\mu(\mathrm{Var}_u(f)) \leq F(\lambda'_v)\mu'(\mathrm{Var}_v(f')) + F(\lambda'_u)\mu'(\mathrm{Var}_u(f')). \tag{25}$$

We can prove variance tensorization by induction as follows:

$$\begin{aligned}
\mathrm{Var}(f') &= \mu'(\mathrm{Var}_v(f')) + \mathrm{Var}(f) \\
&\le \mu'(\mathrm{Var}_v(f')) + \sum_{x\in V} F(\lambda_x)\mu(\mathrm{Var}_x(f)) \\
&\le \mu'(\mathrm{Var}_v(f')) + F(\lambda_u)\mu(\mathrm{Var}_u(f)) + \sum_{x\in V\setminus\{u\}} F(\lambda'_x)\mu'(\mathrm{Var}_x(f')) \\
&\le F(\lambda'_v)\mu'(\mathrm{Var}_v(f')) + F(\lambda'_u)\mu'(\mathrm{Var}_u(f')) + \sum_{x\in V\setminus\{u\}} F(\lambda'_x)\mu'(\mathrm{Var}_x(f')) \\
&= \sum_{x\in V'} F(\lambda'_x)\mu'(\mathrm{Var}_x(f')),
\end{aligned}$$

where the first line follows by by Equation (22), the second line by induction, the third line by Equation (23), and the fourth line by Equation (25). ◀

Our task now is to prove Lemma 8. The following technical inequality will be useful.

▶ **Lemma 9.** *Let $p \in [0,1]$. Suppose $s_1, s_2 > 0$ satisfy $s_1 s_2 \ge 1$. Then for all $A, B, C \in \mathbb{R}$ we have*

$$(C - pA - (1-p)B)^2 \le (1+s_1)(C-A)^2 + (1-p)^2(1+s_2)(B-A)^2. \tag{26}$$

The proof of Lemma 9 is in Appendix B of the full version of this paper. We can now prove the main lemma.

Proof of Lemma 8. Our goal is to prove Equation (24), let us recall its statement:

$$F(\lambda_u)\mu(\mathrm{Var}_u(f)) \le (F(\lambda'_v) - 1)\mu'(\mathrm{Var}_v(f')) + F(\lambda'_u)\mu'(\mathrm{Var}_u(f')). \tag{24}$$

We will consider each of these local variances $\mu(\mathrm{Var}_u(f))$, $\mu'(\mathrm{Var}_v(f'))$, and $\mu'(\mathrm{Var}_u(f'))$. We will express each of them as a sum over independent sets S of V'. We can then establish Equation (24) in a pointwise manner by considering the corresponding inequality for each independent set S.

Let us begin by looking at the general definition of the expected local variance $\mu'(\mathrm{Var}_x(f'))$ for any $x \in V'$. Applying the definition in Equation (9) and then simplifying we obtain the following (a reader familar with the notation can apply Equation (12) to skip directly to the last line):

$$\begin{aligned}
&\mu'(\mathrm{Var}_x(f')) \\
&= \sum_{S\subseteq V'\setminus\{x\}} \mu'_{V'\setminus\{x\}}(S)\cdot \mathrm{Var}_x^S(f_S) \\
&= \sum_{S\subseteq V'\setminus\{x\}} \left(\sum_{T\subseteq\{x\}} \mu'(S\cup T)\right)\left(\frac{1}{2}\sum_{T,U\subseteq\{x\}, T\neq U} {\mu'}_x^S(T){\mu'}_x^S(U)(f'(S\cup T) - f'(S\cup U))^2\right) \\
&= \sum_{S\subseteq V'\setminus\{x\}} \left(\sum_{T\subseteq\{x\}} \mu'(S\cup T)\right)\left({\mu'}_x^S(x){\mu'}_x^S(\emptyset)(f'(S) - f'(S\cup\{x\}))^2\right) \\
&= \sum_{S\subseteq V'\setminus\{x\}} \Big(\mu'(S) + \mu'(S\cup\{x\})\Big)\frac{\mu'(S)\mu'(S\cup\{x\})}{(\mu'(S)+\mu'(S\cup\{x\}))^2}\Big(f'(S) - f'(S\cup\{x\})\Big)^2.
\end{aligned} \tag{27}$$

Notice in Equation (27) that the only $S \subset V' \setminus \{x\}$ which contribute are those where x is unblocked (i.e., no neighbor of x is included in the independent set S) because we need that S and $S \cup \{x\}$ are both independent sets and hence have positive measure in μ'.

Let us now consider each of the local variances appearing in Equation (24) (expressed using carefully chosen summations that will allow us to prove (24) term-by-term in terms of S).

Let $Q_1 := \mu(\mathrm{Var}_u(f))$ denote the expected local variance of f at u. We will use (27); note that only S where u is unblocked (that is, when no neighbor of u is occupied) contribute to the local variance. Such an S where u is unblocked and $u \notin S$ has the same contribution as $S \cup \{u\}$ times $1/\lambda_u$ (since $\mu(S \cup \{u\}) = \lambda_u \mu(S)$). Hence the expected local variance of f at u is given by

$$Q_1 := \mu(\mathrm{Var}_u(f)) = \sum_{S \subseteq V; u \in S} \mu(S) \left(1 + \frac{1}{\lambda_u}\right) \frac{\lambda_u}{1+\lambda_u} \frac{1}{1+\lambda_u} \left(f(S \setminus \{u\}) - f(S)\right)^2 .$$

We have $f(S) = f'(S)$ (since $u \in S$) and $f(S \setminus \{u\}) = \frac{1}{1+\lambda'_v} f'(S \setminus \{u\}) - \frac{\lambda'_v}{1+\lambda'_v} f'(S \setminus \{u\} \cup \{v\})$. Plugging these in and simplifying we obtain the following:

$$Q_1 = \frac{1+\lambda'_v}{1+\lambda'_u+\lambda'_v} \sum_{S \subseteq V; u \in S} \mu(S) \left(f'(S) - \frac{1}{1+\lambda'_v} f'(S-u) - \frac{\lambda'_v}{1+\lambda'_v} f'(S-u+v)\right)^2 . \tag{28}$$

We now consider $Q_2 := \mu'(\mathrm{Var}_u(f'))$. To compute the expected local variance of f' at u we need to generate Z from μ' but only Z where u is unblocked contribute to the local variance. We can generate Z by first generating S from μ and if $u \notin S$ adding v with probability $\lambda'_v/(1+\lambda'_v)$. The S where $u \notin S$ contribute only if u is unblocked; contributing the same amount as $S \cup \{u\}$ multiplied by $1/\lambda'_u$ (since $\mu'(S \cup \{u\}) = \lambda'_u \mu'(S)$) and by $1/(1+\lambda'_v)$ (since they only contribute if we do not add v). Hence, we have the following:

$$\begin{aligned} Q_2 := \mu(\mathrm{Var}_u(f')) &= \sum_{S \subseteq V; u \in S} \mu(S) \left(1 + \frac{1}{\lambda'_u} \frac{1}{1+\lambda'_v}\right) \frac{\lambda'_u}{1+\lambda'_u} \frac{1}{1+\lambda'_u} \left(f'(S-u) - f'(S)\right)^2 \\ &= \left(\lambda'_u + \frac{1}{1+\lambda'_v}\right) \frac{1}{(1+\lambda'_u)^2} \sum_{S \subseteq V; u \in S} \mu(S) \left(f'(S) - f'(S-u)\right)^2 . \end{aligned} \tag{29}$$

Finally, we consider $\mu'(\mathrm{Var}_v(f'))$, the expected local variance of f' at v. We will establish a lower bound which we will denote by Q_3 (note, Q_1 and Q_2 were identities but here we will have an inequality).

To compute $\mu'(\mathrm{Var}_v(f'))$, the expected local variance of f' at v, we need to generate an independent set Z from μ'. Only those Z where v is unblocked (that is where u is missing) contribute to the local variance. We can generate Z by generating S from μ (whether we add or do not add v does not change the contribution to the local variance). As in Equation (27), we obtain the following:

$$\begin{aligned} \mu'(\mathrm{Var}_v(f')) &= \sum_{S \subseteq V; u \notin S} \mu(S) \frac{1}{1+\lambda'_v} \frac{\lambda'_v}{1+\lambda'_v} \left(f'(S \cup \{v\}) - f'(S)\right)^2 \\ &\geq \sum_{S \subseteq V; u \in S} \mu(S) \frac{1}{\lambda'_u} \frac{1}{1+\lambda'_v} \frac{\lambda'_v}{1+\lambda'_v} \left(f'(S \cup \{v\} \setminus \{u\}) - f'(S \setminus \{u\})\right)^2 , \end{aligned}$$

where in the summation in the second line we effectively sum over a subset of sets not containing u (the ones that can be obtained by removing u from a set containing u).

Let Q_3 denote the lower bound we obtained above:

$$Q_3 := \frac{1}{\lambda'_u}\frac{1}{1+\lambda'_v}\frac{\lambda'_v}{1+\lambda'_v}\sum_{S\subseteq V;u\in S}\mu(S)\left(f'(S\cup\{v\}\setminus\{u\})-f'(S\setminus\{u\})\right)^2 \geq \mu'(\mathrm{Var}_v(f')). \tag{30}$$

Plugging in (28), (29), (30) we obtain that Equation (24) follows from the following inequality:

$$F(\lambda_u)Q_1 \leq (F(\lambda'_v)-1)Q_3 + F(\lambda'_u)Q_2. \tag{31}$$

We will establish (31) term-by-term, that is, for each S in the sums of (28), (29), (30). Fix $S\subseteq V$ such that $u\in S$ and let $A=f'(S-u)$, $B=f'(S-u+v)$, and $C=f'(S)$. We need to show

$$\begin{aligned}&\frac{1+\lambda'_v}{1+\lambda'_u+\lambda'_v}\left(C-\frac{1}{1+\lambda'_v}A-\frac{\lambda'_v}{1+\lambda'_v}B\right)^2 F\left(\frac{\lambda'_u}{1+\lambda'_v}\right)\\ &\quad\leq \frac{1+\lambda'_v+\lambda'_u}{1+\lambda'_v}\frac{1}{(1+\lambda'_u)^2}\left(C-A\right)^2F(\lambda'_u)+\frac{1}{\lambda'_u(1+\lambda'_v)^2}(B-A)^2\left(F(\lambda'_v)-1\right).\end{aligned} \tag{32}$$

Let $p=1/(1+\lambda'_v)$. We will match (26) to (32), by first dividing both sides of (32) by $\frac{1+\lambda'_v}{1+\lambda'_u+\lambda'_v}F\left(\frac{\lambda'_u}{1+\lambda'_v}\right)$ and then choosing

$$1+s_1=\left(\frac{1+\lambda'_u+\lambda'_v}{(1+\lambda'_v)(1+\lambda'_u)}\right)^2\cdot\frac{F(\lambda'_u)}{F\left(\frac{\lambda'_u}{1+\lambda'_v}\right)} \quad\text{and}\quad 1+s_2=\frac{1+\lambda'_u+\lambda'_v}{\lambda'_u(1+\lambda'_v)\lambda'^{\,2}_v}\cdot\frac{F(\lambda'_v)-1}{F\left(\frac{\lambda'_u}{1+\lambda'_v}\right)}.$$

Note that with this choice of s_1 and s_2 equations (26) and (32) are equivalent, and hence to prove (32) it is enough to show $s_1s_2\geq 1$.

▷ Claim 10. $s_1s_2\geq 1$.

We defer the proof of this technical inequality to Appendix B of the full version of this paper. This completes the proof of the lemma. ◀

4 Entropy Factorization

Here we will prove Theorem 2 establishing $O(n\log n)$ mixing time for $\lambda\leq .44$. We will accomplish this task by proving the following approximate tensorization inequality:

$$\mathrm{Ent}(f')\leq\sum_{x\in V'}F(\lambda'_x)\mu'(\mathrm{Ent}_x(f')), \tag{33}$$

where $F:\mathbb{R}_{\geq 0}\to\mathbb{R}_{\geq 0}$ is a function to be determined later. By Equation (14) this implies a mixing time bound of $O(n\log n)$ and hence Theorem 2 follows from Equation (33).

We use the same notation as Section 3. Let T' be a tree and μ' be the hard-core measure on T' with fugacities $\{\lambda'_w\}_{w\in V'}$. Let v be a vertex of degree 1 in T' and u is the unique neighbor of v. Let v be a vertex of degree 1 in T' and let u be the unique neighbor of v. Let $T=(V,E)$ be the tree which is the induced subgraph of G' on $V=V'\setminus\{v\}$. Let $\{\lambda_w\}_{w\in V}$ be a collection of fugacities where $\lambda_w=\lambda'_w$ for $w\neq u$ and $\lambda_u=\lambda'_u/(1+\lambda'_v)$. Let μ be the hard-core measure on T with fugacities $\{\lambda_w\}_{w\in V}$.

Fix a function $f':2^{V'}\to\mathbb{R}_{\geq 0}$. Let $f:2^V\to\mathbb{R}_{\geq 0}$ be defined by $f(S)=\mathbb{E}_{Z\sim\mu'}[f'(Z)\mid Z\cap V=S]=\mu'_v(f')(S)$, which is the same definition as in Equation (21).

Our main technical lemma is the following.

▶ **Lemma 11.** *For* $F(x) = 1000(1+x)\exp(x)$ *and* $\lambda_u, \lambda'_v, \lambda'_u \in [0, 0.44]$ *we have the following.*

$$F(\lambda_u)\mu(\mathrm{Ent}_u(f)) \le (F(\lambda'_v) - 1)\mu'(\mathrm{Ent}_v(f')) + F(\lambda'_u)\mu'(\mathrm{Ent}_u(f')). \tag{34}$$

Using the above lemma we will establish (33) using induction as done in Section 3 for variance.

Proof of Theorem 2.

$$\begin{aligned}
\mathrm{Ent}(f') &= \mu'(\mathrm{Ent}_v(f')) + \mathrm{Ent}(f) && \text{by Equation (16)}\\
&\le \mu'(\mathrm{Ent}_v(f')) + \sum_{x\in V} F(\lambda_x)\mu(\mathrm{Ent}_x(f)) && \text{by induction}\\
&\le \mu'(\mathrm{Ent}_v(f')) + F(\lambda_u)\mu(\mathrm{Ent}_u(f)) + \sum_{x\in V\setminus\{u\}} F(\lambda'_x)\mu'(\mathrm{Ent}_x(f')) && \text{by Equation (18)}\\
&= \sum_{x\in V'} F(\lambda'_x)\mu'(\mathrm{Var}_x(f')) && \text{by Equation (34).}
\end{aligned}$$

Hence, Theorem 2 follows from Lemma 11. ◀

Let

$$G(p, A, B) = pA\ln A + (1-p)B\ln B - (pA + (1-p)B)\ln(pA + (1-p)B).$$

Let $Q_1 := \mu(\mathrm{Ent}_u(f))$. Then we have the following analog of Equation (28):

$$Q_1 = \left(1 + \frac{1+\lambda'_v}{\lambda'_u}\right) \sum_{S\subseteq V; u\in S} \mu(S) G\left(\frac{1+\lambda'_v}{1+\lambda'_v+\lambda'_u}, \frac{f'(S\setminus\{u\})}{1+\lambda'_v} + \frac{\lambda'_v f'(S\cup\{v\}\setminus\{u\})}{1+\lambda'_v}, f'(S)\right). \tag{35}$$

Let $Q_2 := \mu'(\mathrm{Ent}_u(f'))$ denote the expected local entropy of f' at u. Then we have the analog of Equation (29):

$$Q_2 = \left(1 + \frac{1}{\lambda'_u(1+\lambda'_v)}\right) \sum_{S\subseteq V; u\in S} \mu(S) G\left(\frac{1}{1+\lambda'_u}, f'(S-u), f(S)\right). \tag{36}$$

Finally, as in Equation (30), we use Q_3 for a lower bound on the expected local entropy of f' at v. We prove $\mu'(\mathrm{Ent}_v(f')) \ge Q_3$ where:

$$Q_3 = \frac{1}{\lambda'_u} \sum_{S\subseteq V; u\in S} \mu(S) G\left(\frac{1}{1+\lambda'_v}, f'(S-u), f(S-u+v)\right). \tag{37}$$

Plugging in (35), (36), (37) we obtain that Equation (34) follows from the following inequality

$$F(\lambda_u)Q_1 \le F(\lambda'_u)Q_2 + (F(\lambda'_v) - 1)Q_3. \tag{38}$$

We will establish (34) term-by-term, that is, for each S in the sums of (35), (36), (37). Fix $S \subseteq V$ such that $u \in S$ and let $A = f'(S-u)$, $B = f'(S-u+v)$, and $C = f'(S)$. We need to show

$$\begin{aligned}
&\left(1 + \frac{1+\lambda'_v}{\lambda'_u}\right) G\left(\frac{1+\lambda'_v}{1+\lambda'_v+\lambda'_u}, \frac{1}{1+\lambda'_v}A + \frac{\lambda'_v}{1+\lambda'_v}B, C\right) F\left(\frac{\lambda'_u}{1+\lambda'_v}\right) \le\\
&\left(1 + \frac{1}{\lambda'_u(1+\lambda'_v)}\right) G\left(\frac{1}{1+\lambda'_u}, A, C\right) F(\lambda'_u) + \frac{1}{\lambda'_u} G\left(\frac{1}{1+\lambda'_v}, A, B\right)(F(\lambda'_v) - 1).
\end{aligned} \tag{39}$$

We will show that the following lemma implies an optimal mixing time bound of $O(n \log n)$ for all $\lambda \leq .44$.

▶ **Lemma 12.** *Let $b, p \in (0,1)$. For any $A, B, C \geq 0$ we have*

$$G(b, A, C) + (1-b)G(p, A, B) - (b+p-bp)G\left(\frac{b}{b+p-bp}, pA + (1-p)B, C\right) \geq 0. \quad (40)$$

The proof of Lemma 12 appears in Appendix D of the full version of this paper.

Before finishing the proof of Lemma 11 using Lemma 12, we have the following conjecture which is a generalization of Lemma 12. This conjecture implies (39) for all $\lambda \leq 1.05$ and hence an optimal mixing time bound of $O(n \log n)$ for all $\lambda \leq 1.05$. Lemma 12 corresponds to Conjecture 13 "around" $W = 1$ (note that for $W = 1$ both sides of the below inequality are zero). We can numerical verify the conjecture.

▶ **Conjecture 13.** *Let $b, p \in (0,1)$. For any $W \in (1-b, 1/(1-p))$ and any $A, B, C \geq 0$ we have*

$$\left|\ln W\right|(b+p-bp)G\left(\frac{b}{b+p-bp}, pA + (1-p)B, C\right) \leq$$
$$p\left|\ln \frac{pW}{pW - W + 1}\right|G(b, A, C) + b\left|\ln \frac{W+b-1}{bW}\right|G(p, A, B).$$

In Appendix C of the full version of this paper we prove that Conjecture 13 implies a strengthening of Lemma 11 with the interval $[0, 1.05]$ and hence $O(n \log n)$ mixing time.

4.1 Proof of Lemma 11

Here we prove that Lemma 12 implies Equation (39), and hence Lemma 11. Recall, $F(x) = 1000(1+x)\exp(x)$.

Let

$$p = \frac{1}{1+\lambda'_v} \quad \text{and} \quad b = \frac{1}{1+\lambda'_u} \quad \text{and} \quad \alpha = \frac{1}{p(1-b)} F\left(\frac{p(1-b)}{b}\right).$$

Note that

$$1 + \frac{1+\lambda'_v}{\lambda'_u} = \frac{b+p-bp}{p(1-b)}, \qquad \frac{1+\lambda'_v}{1+\lambda'_v+\lambda'_u} = \frac{b}{b+p-bp}, \qquad \frac{\lambda'_u}{1+\lambda'_v} = \frac{p(1-b)}{b}.$$

We aim to prove (39) using the following sequence of inequalities

$$\left(1 + \frac{1+\lambda'_v}{\lambda'_u}\right) G\left(\frac{1+\lambda'_v}{1+\lambda'_v+\lambda'_u}, \frac{1}{1+\lambda'_v}A + \frac{\lambda'_v}{1+\lambda'_v}B, C\right) F\left(\frac{\lambda'_u}{1+\lambda'_v}\right)$$
$$= \alpha(b+p-bp)G\left(\frac{b}{b+p-bp}, pA + (1-p)B, C\right) \quad (41)$$
$$\leq \alpha p\left(\frac{1}{p}\right)G(b, A, C) + \alpha b\left(\frac{1-b}{b}\right)G(p, A, B) \quad (42)$$
$$\leq \left(1 + \frac{1}{\lambda'_u(1+\lambda'_v)}\right) G\left(\frac{1}{1+\lambda'_u}, A, C\right) F(\lambda'_u) + \frac{1}{\lambda'_u} G\left(\frac{1}{1+\lambda'_v}, A, B\right)(F(\lambda'_v) - 1), \quad (43)$$

Here, Eq. (41) follows from the definitions.

Inequality (42) follows from Lemma 12 (we avoided simplifying (42) in order to make the match to Lemma 12 easier).

Inequality (43) follows from the following two inequalities:

$$\alpha p\Big(\frac{1}{p}\Big) \le \frac{b+p-bp}{p(1-b)} F\Big(\frac{1-b}{b}\Big), \text{ and} \tag{44}$$

$$\alpha b\Big(\frac{1-b}{b}\Big) \le \frac{b}{1-b}\Big(F\Big(\frac{1-p}{p}\Big)-1\Big). \tag{45}$$

Replacing α by its definition, equations (44) and (45) become

$$F\left(\frac{p(1-b)}{b}\right) \le (b+p-bp)F\Big(\frac{1-b}{b}\Big), \quad \text{and} \tag{46}$$

$$F\left(\frac{p(1-b)}{b}\right)(1-b) \le bp\left(F\Big(\frac{1-p}{p}\Big)-1\right). \tag{47}$$

Let x and y be such that $b = 1/(1+x)$ and $p = 1/(1+y)$. Equations (46) and (47) simplify to the following

$$F\left(\frac{x}{1+y}\right) \le \frac{1+x+y}{(1+x)(1+y)}F(x) \quad \text{and} \quad F\left(\frac{x}{1+y}\right)x \le \frac{F(y)-1}{1+y}.$$

Note that $F(y) \ge 1000$ and hence it is enough to satisfy the following inequalities.

$$F\left(\frac{x}{1+y}\right) \le \frac{1+x+y}{(1+x)(1+y)}F(x) \quad \text{and} \quad F\left(\frac{x}{1+y}\right)x \le \frac{999}{1000}\frac{F(y)}{1+y}.$$

Recalling the definition of F, the constraints further simplify to:

$$\exp\left(\frac{x}{1+y}\right) \le \exp(x) \quad \text{and} \quad \exp\left(\frac{x}{1+y}\right)\frac{(1+x+y)x}{1+y} \le \frac{999}{1000}\exp(y). \tag{48}$$

Note that the first constraint follows from the fact that exp() is increasing and $x, y > 0$. The second constraint is addressed in the following lemma.

▶ **Lemma 14.** *For $x, y \in [0, u]$, where $u = 0.44$, we have*

$$\exp\left(\frac{x}{1+y}\right)\frac{(1+x+y)x}{1+y} \le \frac{999}{1000}\exp(y). \tag{49}$$

Proof. Let

$$Q = \frac{999}{1000}\exp(y) - \exp\left(\frac{x}{1+y}\right)\frac{(1+x+y)x}{1+y}.$$

We have

$$\frac{\partial}{\partial y}Q = \frac{999}{1000}\exp(y) + \frac{x^2(x+2y+2)\exp\left(\frac{x}{1+y}\right)}{(1+y)^3} > 0,$$

that is, Q is increasing in y and hence we only need to prove (49) for $y = 0$. We need to show

$$\frac{999}{1000} \ge \exp(x)(1+x)x. \tag{50}$$

Note that RHS of (50) is increasing in x and hence we need to check (50) for $x = u$. For $x = u = 0.44$ we have that (50) is satisfied (checked using interval arithmetic). ◀

▶ Remark 15. In order for (48) to hold for $x, y \in [0, u]$ we need it to hold for $x = y(y+1) \le u$. Equation (48) then simplifies to

$$(1+y)^2 y \le 1.$$

For (48) to hold (for all $x, y \in [0, u]$) we need $y \le 0.47$ which in turn implies $u \le 0.7$. This means that if we want to prove rapid mixing for unweighted independent sets ($\lambda = 1$) we have to go beyond Lemma 12, see Conjecture 13.

References

1 Nima Anari, Kuikui Liu, and Shayan Oveis Gharan. Spectral independence in high-dimensional expanders and applications to the hardcore model. In *Proceedings of the 61st Annual IEEE Symposium on Foundations of Computer Science (FOCS)*, pages 1319–1330, 2020.

2 Noam Berger, Claire Kenyon, Elchanan Mossel, and Yuval Peres. Glauber dynamics on trees and hyperbolic graphs. *Probability Theory and Related Fields*, 131(3):311–340, 2005.

3 Nayantara Bhatnagar, Allan Sly, and Prasad Tetali. Decay of correlations for the hardcore model on the d-regular random graph. *Electron. J. Probab.*, 21:1–42, 2016.

4 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 *Proceedings of the 33rd Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pages 3670–3692, 2022.

5 Graham Brightwell and Peter Winkler. A second threshold for the hard-core model on a Bethe lattice. *Random structures and algorithms*, 24:303–314, 2004.

6 Pietro Caputo. Lecture notes on Entropy and Markov Chains. *Preprint, available from: http://www.mat.uniroma3.it/users/caputo/entropy.pdf*, 2023.

7 Pietro Caputo, Georg Menz, and Prasad Tetali. Approximate tensorization of entropy at high temperature. *Annales de la Faculté des sciences de Toulouse: Mathématiques*, 24(4):691–716, 2015.

8 Pietro Caputo and Daniel Parisi. Block factorization of the relative entropy via spatial mixing. *Communications in Mathematical Physics*, 388:793–818, 2021.

9 Xiaoyu Chen, Weiming Feng, Yitong Yin, and Xinyuan Zhang. Optimal mixing for two-state anti-ferromagnetic spin systems. In *Proceedings of the 63rd Annual IEEE Symposium on Foundations of Computer Science (FOCS)*, 2022.

10 Yuansi Chen and Ronen Eldan. Localization schemes: A framework for proving mixing bounds for Markov chains. In *Proceedings of the 63rd Annual IEEE Symposium on Foundations of Computer Science (FOCS)*, 2022.

11 Zongchen Chen and Yuzhou Gu. Fast sampling of b-matchings and b-edge covers. *arXiv preprint arXiv:2304.14289*, 2023.

12 Zongchen Chen, Kuikui Liu, Nitya Mani, and Ankur Moitra. Strong spatial mixing for colorings on trees and its algorithmic applications. *arXiv preprint arXiv:2304.01954*, 2023.

13 Zongchen Chen, Kuikui Liu, and Eric Vigoda. Optimal mixing of Glauber dynamics: Entropy factorization via high-dimensional expansion. In *Proceedings of the 53rd Annual ACM Symposium on Theory of Computing (STOC)*, pages 1537–1550, 2021.

14 Zongchen Chen, Kuikui Liu, and Eric Vigoda. Spectral independence via stability and applications to holant-type problems. In *Proceedings of the 63rd Annual IEEE Symposium on Foundations of Computer Science (FOCS)*, 2022.

15 Michelle Delcourt, Marc Heinrich, and Guillem Perarnau. The Glauber dynamics for edge-colorings of trees. *Random Structures & Algorithms*, 57(4):1050–1076, 2020.

16 David Eppstein and Daniel Frishberg. Rapid mixing of the hardcore Glauber dynamics and other Markov chains in bounded-treewidth graphs. *arXiv preprint arXiv:2111.03898*, 2022.

17 Andreas Galanis, Daniel Štefankovič, and Eric Vigoda. Inapproximability of the partition function for the antiferromagnetic Ising and hard-core models. *Combinatorics, Probability and Computing*, 25(4):500–559, 2016.

18 Thomas P. Hayes and Alistair Sinclair. A general lower bound for mixing of single-site dynamics on graphs. *The Annals of Applied Probability*, 17(3):931–952, 2007.

19 Kuikui Liu. From coupling to spectral independence and blackbox comparison with the down-up walk. In *APPROX-RANDOM*, pages 32:1–32:21, 2021.

20 Brendan Lucier, Michael Molloy, and Yuval Peres. The Glauber dynamics for colourings of bounded degree trees. In *RANDOM*, pages 631–645, 2009.

21 James B. Martin. Reconstruction thresholds on regular trees. In *Discrete Random Walks*, DMTCS Proceedings, pages 191–204, 2003.

22 Fabio Martinelli, Alistair Sinclair, and Dror Weitz. The Ising model on trees: boundary conditions and mixing time. In *44th Annual IEEE Symposium on Foundations of Computer Science (FOCS)*, pages 628–639, 2003.

23 Fabio Martinelli, Alistair Sinclair, and Dror Weitz. Fast mixing for independent sets, colorings and other models on trees. In *Proceedings of the 15th Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pages 456–465, 2004.

24 Elchanan Mossel. Survey: Information flow on trees. In *Graphs, Morphisms and Statistical Physics*, DIMACS series in discrete mathematics and theoretical computer science, pages 155–170, 2004.

25 Elchanan Mossel, Dror Weitz, and Nicholas C. Wormald. On the hardness of sampling independent sets beyond the tree threshold. *Probability Theory and Related Fields*, 143:401–439, 2007.

26 Ricardo Restrepo, Daniel Štefankovič, Juan C. Vera, Eric Vigoda, and Linji Yang. Phase transition for Glauber dynamics for independent sets on regular trees. *SIAM Journal on Discrete Mathematics*, 28, 2014.

27 Allan 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.

28 Allan Sly and Nike Sun. The computational hardness of counting in two-spin models on d-regular graphs. *The Annals of Probability*, 42(6):2383–2416, 2014.

29 Daniel Štefankovič, Santosh Vempala, and Eric Vigoda. Adaptive simulated annealing: A near-optimal connection between sampling and counting. *Journal of the ACM*, 56(3):1–36, 2009.