Combinatorics, Probability and Computing (2024), 1–17
doi:10.1017/S0963548324000348

ARTICLE

Optimal mixing via tensorization for random independent sets on arbitrary trees

Charilaos Efthymiou[1], Thomas P. Hayes[2], Daniel Štefankovič[3] and Eric Vigoda[4]

[1]Department of Computer Science, University of Warwick, Coventry, UK, [2]Department of Computer Science and Engineering, University at Buffalo, New York, USA, [3]Department of Computer Science, University of Rochester, Rochester, USA, and [4]Department of Computer Science, University of California, Santa Barbara, USA
Corresponding author: Charilaos Efthymiou; Email: charilaos.efthymiou@warwick.ac.uk

(Received 9 April 2024; revised 15 September 2024; accepted 21 September 2024)

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, Stefankovic, Vera, Vigoda, and Yang (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. 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 via a non-trivial inductive proof.

Keywords: Markov Chain Monte Carlo; mixing time; independent sets; hard-core model; approximate counting algorithms; sampling algorithms

2020 MSC Codes: Primary: 60J10

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

Charilaos Efthymiou: Research supported in part by EPSRC NIA, grant EP/V050842/1, and Centre of Discrete Mathematics and Applications (DIMAP). Daniel Stefankovic: Research supported in part by NSF grant CCF-1563757. Eric Vigoda: Research supported in part by NSF grant CCF-2147094.

© The Author(s), 2024. Published by Cambridge University Press.

https://doi.org/10.1017/S0963548324000348 Published online by Cambridge University Press

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., Štefankovič, Vempala, and Vigoda [33]). 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 [22] 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 of the hard-core model. Interestingly, our analysis utilises 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., $\lambda = 1$.

We can now state our main results.

Theorem 1.1. *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 $O(n)$.*

We believe the optimal mixing results of Theorem 1.1 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

https://doi.org/10.1017/S0963548324000348 Published online by Cambridge University Press

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 [28] 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]; they showed that for constants $C_1, C_2 > 0$ and sufficiently large Δ: $C_1 \log^2 \Delta / \log\log \Delta \le \lambda_r(\Delta) \le C_2 \log^2 \Delta$, and hence $\lambda_r(\Delta)$ is "increasing asymptotically" with Δ whereas the uniqueness threshold is a decreasing function of Δ. Martin [25] showed that $\lambda_r(\Delta) > e - 1$ for all Δ. As a consequence, we conjecture that Theorem 1.1 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: Restrepo, Štefankovič, Vera, Vigoda, and Yang [30] 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 [30] but using the Kesten–Stigum bound one can show $C \approx 28$ (by considering binary trees).

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 by Chen, Liu and Vigoda [16], and is exponentially slow in the non-uniqueness region by Mossel, Weitz, and Wormald [29] (see also [21]). 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 [11,12,16], and NP-hardness to approximately count/sample in the non-uniqueness region by Sly [31] (see also, [21,32]).

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 [26,27] 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, [30] shows that 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, Kenyon, Mossel and Peres [2] proves polynomial mixing time for trees with constant maximum degree [2]. A very recent result of Eppstein and Frishberg [19] 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 $C(\lambda) = O(1 + |\log(\lambda)|)$ for trees; see more recent work of Chen [10] for further improvements. For other combinatorial models, rapid mixing for the Glauber dynamics on trees with bounded maximum degree was established for k-colorings in [24] and edge-colorings in [18].

Spectral independence is a powerful notion in the analysis of the convergence rate of MCMC algorithms. For independent sets on an n-vertex graph $G = (V, E)$, spectral independence considers the $n \times n$ pairwise influence matrix $\mathcal{I}_G$ where $\mathcal{I}_G(v, w) = \mathrm{Prob}_{\sigma\sim\mu}(v \in \sigma \mid w \in \sigma) - \mathrm{Prob}_{\sigma\sim\mu}(v \in \sigma \mid w \notin \sigma)$; this matrix is closely related to the vertex covariance matrix. We say that spectral independence holds if the maximum eigenvalue of $\mathcal{I}_{G'}$ for all vertex-induced subgraphs G' of G are bounded by a constant. Spectral independence was introduced by Anari, Liu,

https://doi.org/10.1017/S0963548324000348 Published online by Cambridge University Press

and Oveis Gharan [1]. Chen, Liu, and Vigoda [16] 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., [4,14,15,17,23].

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 [11,12]; 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 utilised 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 Section 4. 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 [16] (see also [4,11,12]).

Theorem 1.2. *For all constant Δ, all $\lambda \leq 1.3$, for any tree T with maximum degree Δ, the Glauber dynamics for sampling λ-weighted independent sets has an optimal mixing time of $O(n \log n)$.*

Interestingly, combining the spectral independence results from the proof Theorem 1.2 with results of Chen, Feng, Yin, and Zhang [[13], Theorem 1.9], we are able to strengthen Theorem 1.1 by allowing larger fugacities, i.e., $\lambda \leq 1.3$.

Corollary 1.3. *For any n-vertex tree, for any $\lambda \leq 1.3$ the Glauber dynamics for sampling λ-weighted independent sets in the hard-core model has an optimal relaxation time of $O(n)$.*

In the next section we recall the key functional definitions and basic properties of variance that will be useful later in the proofs. In Section 3 we prove approximate tensorization of variance which establishes Theorem 1.1. We establish spectral independence and prove Theorem 1.2 in Section 4.

Remark 1.4. An earlier version of this paper [20] claimed to prove $O(n \log n)$ mixing time for $\lambda \leq .44$ for any tree (without any constraint on the maximum degree). There was a mistake in that proof. In particular, inequality (54) is false. Moreover, Zongchen Chen pointed out a simple test function f which shows that entropy tensorization and modified log-Sobolev inequality do not hold for the star graph with constants independent of the degree.

A recent paper of Chen, Yang, Yin, and Zhang [9] improves Corollary 1.3 to obtain optimal relaxation time on trees for $\lambda < e^2$.

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

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

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

https://doi.org/10.1017/S0963548324000348 Published online by Cambridge University Press

$1 = \lambda_1 \geq \lambda_2 \geq \ldots \lambda_{|\Omega|} \geq -1$. 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, i.e., Ω are the collection of independent sets of G.

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

$$\text{Var}(f) = \mu(f^2) - \left(\mu(f)\right)^2.$$

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

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

For any subset $S \subseteq V$, let Ω_S denote the set of independent sets on S. Then, let μ_S denote the marginal of μ on S; that is, for any $\sigma \in \Omega_S$, we have that

$$\mu_S(\sigma) = \sum_{\eta\in\Omega} \mathbf{1}\{\eta \cap S = \sigma\}\mu(\eta).$$

For any $S \subset V$, any $\tau \in \Omega_{V\setminus S}$, we let μ_S^τ be the distribution μ conditional on the configuration τ on $V \setminus S$, and let Ω_S^τ to be the support of μ_S^τ.

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

$$\mu_S^\tau(f) = \sum_{\sigma\in\Omega_S^\tau} \mu_S^\tau(\sigma) f_\tau(\sigma).$$

Let $\text{Var}_S^\tau(f)$ denote the variance of f_τ with respect to the conditional distribution μ_S^τ:

$$\text{Var}_S^\tau(f) = \mu_S^\tau(f^2) - \left(\mu_S^\tau(f)\right)^2 \quad = \frac{1}{2}\sum_{\sigma,\eta\in\Omega} \frac{\mathbf{1}\{\sigma \setminus S = \tau,\ \eta \setminus S = \tau\}\mu(\sigma)\mu(\eta)}{\left(\sum_{\hat{\sigma}\in\Omega}\mathbf{1}\{\hat{\sigma} \setminus S = \tau\}\mu(\hat{\sigma})\right)^2}\left(f(\sigma) - f(\eta)\right)^2. \tag{4}$$

Furthermore, we let

$$\mu(\text{Var}_S(f)) = \sum_{\tau\in\Omega_{V\setminus S}} \mu_{V\setminus S}(\tau) \cdot \text{Var}_S^\tau(f), \tag{5}$$

i.e., $\mu(\text{Var}_S(f))$ is the average of $\text{Var}_S^\tau(f)$ with respect to 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 $\mu(\text{Var}_v(f))$.

Finally, let

$$\text{Var}(\mu_S(f)) = \mu\left(\left(\mu_S(f)\right)^2\right) - \left(\mu\left(\mu_S(f)\right)\right)^2 = \mathbb{E}_{\tau\sim\mu_{V\setminus S}}\left[\left(\mu_S^\tau(f)\right)^2\right] - \left(\mathbb{E}_{\tau\sim\mu_{V\setminus S}}\left[\mu_S^\tau(f)\right]\right)^2, \tag{6}$$

https://doi.org/10.1017/S0963548324000348 Published online by Cambridge University Press

i.e., $\mathrm{Var}(\mu_S(f))$ is the variance of $\mu_S^\tau(f)$ with respect to τ being distributed as in $\mu_{V\setminus S}(\cdot)$.

When X is the following two-valued random variable:

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

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

$$\mathrm{Var}(X) = p(1-p)(A-B)^2. \tag{7}$$

2.3 Approximate tensorization of variance

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

We begin with the definition of approximate tensorization of variance.

Definition 2.1 (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*

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

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 [6,7]:

$$T_{\mathrm{relax}} \leq Cn. \tag{8}$$

2.4 Decomposition of variance

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

Lemma 2.2. *For any $S \subset V$, for any $f \geq 0$:*

$$\mathrm{Var}(f) = \mu[\mathrm{Var}_S(f)] + \mathrm{Var}(\mu_S(f)), \tag{9}$$

where $\mathrm{Var}(\mu_S(f))$ is defined in Eq. (6).

We utilise the following variance factorisation for product measures, see [[6], Eqn (4.7)].

Lemma 2.3. *Consider $U, W \subset V$ where $\mathrm{dist}(U, W) \geq 2$. Then for all $f \geq 0$ we have:*

$$\mu[\mathrm{Var}_U(\mu_W(f))] \leq \mu[\mathrm{Var}_U(f)], \tag{10}$$

On a first account, the reader might find it challenging to parse the expression $\mu[\mathrm{Var}_U(\mu_W(f))]$. In that respect, (5) and (6) might be useful. Specifically, $\mu[\mathrm{Var}_U(\mu_W(f))]$ is the expectation of $\mathrm{Var}_U^\tau(\mu_W(f))$ with respect to τ being distributed as in $\mu_{\bar{U}\cap\bar{W}}$. Furthermore, $\mathrm{Var}_U^\tau(\mu_W(f))$ corresponds to the variance of $\mu_W^{\tau,\sigma}(f)$ with respect to the configurations τ at $V \setminus (U \cup W)$ and σ at U, while τ is fixed and σ is distributed as in $\mu_U^\tau(\cdot)$.

Proof. We apply [[6], Equation (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 $\mathrm{dist}(U, W) \geq 2$, 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 $S \setminus (U \cup W)$. □

https://doi.org/10.1017/S0963548324000348 Published online by Cambridge University Press

3. Variance factorization

In this section we prove Theorem 1.1, establishing an optimal bound on the relaxation time for the Glauber dynamics on any tree for $\lambda \leq 1.1$. We will prove this by establishing variance tensorization, see Definition 2.1.

Consider a graph $G = (V, E)$ and a collection of fugacities λ_i for each $i \in V$. Throughout this section we will assume that all the fugacities are bounded by 1.1. Consider the following more general definition of the Gibbs distribution μ for the hard-core model, where for an independent set S,

$$\mu(S) \propto \prod_{i \in S} \lambda_i. \tag{11}$$

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

$$\mathrm{Var}(f') \leq \sum_{x \in V'} F(\lambda'_x)\mu'(\mathrm{Var}_x(f')), \tag{12}$$

where $F : \mathbb{R}_{\geq 0} \to \mathbb{R}_{\geq 0}$ is a function to be determined later (in Lemma 3.1). 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 (12) using induction. Let v be a vertex of degree 1 in T' and let u be the unique neighbour of v. Let $T = (V, E)$ be the tree by removing v from T', i.e., T is the induced subgraph of T' 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

$$\mu(S) = \mu'(S) + \mu'(S \cup \{v\}) = \mu'_V(S). \tag{13}$$

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

$$f(S) = \frac{\mu'(S)f'(S) + \mu'(S \cup \{v\})f'(S \cup \{v\})}{\mu'(S) + \mu'(S \cup \{v\})} = \mathbb{E}_{Z \sim \mu'}[f'(Z) \mid Z \cap V = S] = \mu'_v(f')(S). \tag{14}$$

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 2.2 we have that:

$$\mathrm{Var}(f') = \mu'(\mathrm{Var}_v(f')) + \mathrm{Var}(f). \tag{15}$$

For measure μ', when we condition on the configuration at u, the configuration at $V \setminus \{u\}$ is independent of that at $\{v\}$. Hence, from Equation (10) 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 (13) we have $\mu(\mathrm{Var}_x(f)) = \mu'(\mathrm{Var}_x(f))$, the above implies that

$$\mu(\mathrm{Var}_x(f)) \leq \mu'(\mathrm{Var}_x(f')). \tag{16}$$

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

$$F(\lambda_u)\mu(\mathrm{Var}_u(f)) \leq (F(\lambda'_v) - 1)\mu'(\mathrm{Var}_v(f')) + F(\lambda'_u)\mu'(\mathrm{Var}_u(f')). \tag{17}$$

We now utilise the above lemma to prove the main theorem for the relaxation time. We then go back to prove Lemma 3.1.

https://doi.org/10.1017/S0963548324000348 Published online by Cambridge University Press

Proof of Theorem 1.1. Note Equation (17) is equivalent to:

$$\mu'(\mathrm{Var}_v(f')) + F(\lambda_u)\mu(\mathrm{Var}_u(f)) \leq F(\lambda'_v)\mu'(\mathrm{Var}_v(f')) + F(\lambda'_u)\mu'(\mathrm{Var}_u(f')). \tag{18}$$

We can prove variance tensorization by induction as follows:

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

For the first line, we use Equation (15). The second line follows from the inductive hypothesis. For the third line, we use Equation (16) and the fact that $F(\lambda_x) \leq F(\lambda'_x)$, since F is increasing and $\lambda_x \leq \lambda'_x$. The fourth line follows by using Equation (18). □

Our task now is to prove Lemma 3.1. The following technical inequality will be useful.

Lemma 3.2. *Let $p \in [0, 1]$. Suppose $s_1, s_2 > 0$ satisfy $s_1 s_2 \geq 1$. Then for all $A, B, C \in \mathbb{R}$ we have*

$$(C - pA - (1-p)B)^2 \leq (1+s_1)(C-A)^2 + (1-p)^2(1+s_2)(B-A)^2. \tag{19}$$

Proof. Equation (19) is equivalent to

$$2(1-p)(C-A)(A-B) \leq s_1(C-A)^2 + s_2(1-p)^2(B-A)^2. \tag{20}$$

We derive the above by subtracting $(C-A)^2$ and $(1-p)^2(B-A)^2$ from both sides of equation (19) and rearranging.

A simple application of the AM-GM inequality implies that

$$2\sqrt{s_1 s_2}(1-p)(C-A)(A-B) \leq s_1(C-A)^2 + s_2(1-p)^2(B-A)^2.$$

Then, equation (20) follows from the observation that the left-hand side of the above inequality is at least $2(1-p)(C-A)(A-B)$, i.e., since $s_1 s_2 \geq 1$. □

We can now prove the main lemma.

Proof of Lemma 3.1. Our goal is to prove Eq. (17), let us recall its statement:

$$F(\lambda_u)\mu(\mathrm{Var}_u(f)) \leq (F(\lambda'_v) - 1)\mu'(\mathrm{Var}_v(f')) + F(\lambda'_u)\mu'(\mathrm{Var}_u(f')). \tag{17}$$

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 Eq. (17) 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 Eq. (5) and then simplifying we obtain the following (a reader familar with the notation can apply Eq. (7) to skip directly to the last line):

https://doi.org/10.1017/S0963548324000348 Published online by Cambridge University Press

$$
\begin{aligned}
&\mu'(\mathrm{Var}_x(f')) \\
&\quad = \sum_{S\subseteq V'\setminus\{x\}} \mu'_{V'\setminus\{x\}}(S)\cdot \mathrm{Var}_x^S(f_S) \\
&\quad = \sum_{S\subseteq V'\setminus\{x\}} \left(\sum_{T\subseteq\{x\}} \mu'(S\cup T)\right)\left(\frac{1}{2}\sum_{T,U\subseteq\{x\},T\neq U} \mu_x'^S(T)\mu_x'^S(U)(f'(S\cup T)-f'(S\cup U))^2\right) \\
&\quad = \sum_{S\subseteq V'\setminus\{x\}} \left(\sum_{T\subseteq\{x\}} \mu'(S\cup T)\right)\left(\mu_x'^S(x)\mu_x'^S(\emptyset)(f'(S)-f'(S\cup\{x\}))^2\right) \\
&\quad = \sum_{S\subseteq V'\setminus\{x\}} \Big(\mu'(S)+\mu'(S\cup\{x\})\Big)\frac{\mu'(S)\mu'(S\cup\{x\})}{(\mu'(S)+\mu'(S\cup\{x\}))^2}\Big(f'(S)-f'(S\cup\{x\})\Big)^2. \qquad (21)
\end{aligned}
$$

Notice in Eq. (21) that the only $S\subset V'\setminus\{x\}$ which contribute are those where x is unblocked (i.e., no neighbour 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 Eq. (17) (expressed using carefully chosen summations that will allow us to prove (17) 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 (21) (applied to T instead of T'); note that only S where u is unblocked (that is, when no neighbour of u is occupied) contribute to the local variance. Moreover, we can express the expected local variance of f at u in terms of only those S where $u\in S$. In particular, consider an independent set S' where $u\notin S'$. Note that if u is blocked (i.e., $N(u)\cap S'\neq\emptyset$) then the local variance at u is zero for this term. And for those S' where $u\notin S'$ and u is unblocked then the local variance has the same contribution as $S=S'\cup\{u\}$ 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

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

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:

$$
Q_1 = \frac{1+\lambda'_v}{1+\lambda'_u+\lambda'_v}\sum_{S\subseteq V;u\in S}\mu(S)\left(f'(S)-\frac{1}{1+\lambda'_v}f'(S-u)-\frac{\lambda'_v}{1+\lambda'_v}f'(S-u+v)\right)^2. \qquad (22)
$$

We now consider $Q_2 := \mu'(\mathrm{Var}_u(f'))$. As we did for Q_1 we can express Q_2 as a sum over independent sets S where $u\in S$. In this case for an independent set S where $u\in S$, consider $S'=S\setminus\{u\}$ and note that the following hold

$$
\mu'(S') = \mu(S')\frac{1}{1+\lambda'_v} = \mu(S)\frac{1}{\lambda'_u}\frac{1}{1+\lambda'_v},
$$

and

$$
\mu'(S) = \mu(S).
$$

https://doi.org/10.1017/S0963548324000348 Published online by Cambridge University Press

Hence, we have the following:

$$Q_2 := \mu'(\mathrm{Var}_u(f')) = \sum_{S\subseteq V; u\in S} \mu(S)\left(1+\frac{1}{\lambda'_u}\frac{1}{1+\lambda'_v}\right)\frac{\lambda'_u}{1+\lambda'_u}\frac{1}{1+\lambda'_u}\left(f'(S-u)-f'(S)\right)^2$$
$$= \left(\lambda'_u + \frac{1}{1+\lambda'_v}\right)\frac{1}{(1+\lambda'_u)^2}\sum_{S\subseteq V; u\in S}\mu(S)\left(f'(S)-f'(S-u)\right)^2. \tag{23}$$

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 S' from μ'. Only those S' where v is unblocked (that is where u is not in S') contribute to the local variance. We can generate S' by generating S from μ (whether we add or do not add v does not change the contribution to the local variance). As in Eq. (21), we obtain the following:

$$\mu'(\mathrm{Var}_v(f')) = \sum_{S'\subseteq V; u\notin S'} \mu(S')\frac{1}{1+\lambda'_v}\frac{\lambda'_v}{1+\lambda'_v}\left(f'(S'\cup\{v\})-f'(S')\right)^2$$
$$\geq \sum_{\substack{S'\subseteq V; u\notin S'\\ u \text{ is unblocked}}} \mu(S')\frac{1}{1+\lambda'_v}\frac{\lambda'_v}{1+\lambda'_v}\left(f'(S'\cup\{v\})-f'(S')\right)^2$$
$$= \sum_{S\subseteq V; u\in S} \mu(S)\frac{1}{\lambda'_u}\frac{1}{1+\lambda'_v}\frac{\lambda'_v}{1+\lambda'_v}\left(f'(S\cup\{v\}\setminus\{u\})-f'(S\setminus\{u\})\right)^2.$$

Let Q_3 denote the lower bound we obtained above:

$$Q_3 := \frac{1}{\lambda'_u}\frac{1}{1+\lambda'_v}\frac{\lambda'_v}{1+\lambda'_v}\sum_{S\subseteq V; u\in S}\mu(S)\left(f'(S\cup\{v\}\setminus\{u\})-f'(S\setminus\{u\})\right)^2 \geq \mu'(\mathrm{Var}_v(f')). \tag{24}$$

Plugging in (22), (23), (24) we obtain that Eq. (17) follows from the following inequality:

$$F(\lambda_u)Q_1 \leq (F(\lambda'_v)-1)Q_3 + F(\lambda'_u)Q_2. \tag{25}$$

We will establish (25) term-by-term, that is, for each S in the sums of (22), (23), (24). 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

$$\frac{1+\lambda'_v}{1+\lambda'_u+\lambda'_v}\left(C-\frac{1}{1+\lambda'_v}A-\frac{\lambda'_v}{1+\lambda'_v}B\right)^2 F\left(\frac{\lambda'_u}{1+\lambda'_v}\right)$$
$$\leq \frac{1+\lambda'_v+\lambda'_u}{1+\lambda'_v}\frac{1}{(1+\lambda'_u)^2}(C-A)^2F(\lambda'_u) + \frac{1}{\lambda'_u(1+\lambda'_v)^2}(B-A)^2\left(F(\lambda'_v)-1\right). \tag{26}$$

Let $p=1/(1+\lambda'_v)$. We will match (19) to (26), by first dividing both sides of (26) 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 (19) and (26) are equivalent, and hence to prove (26) it is enough to show $s_1s_2\geq 1$. □

Claim 3.3. $s_1s_2\geq 1$.

This completes the proof of the lemma.

https://doi.org/10.1017/S0963548324000348 Published online by Cambridge University Press

We use the following lemma to prove Claim 3.3

Lemma 3.4. *Let* $\alpha = 1.1$ *and* $\beta = 1.3$. *Suppose* $x, y \in (0, \alpha]$ *are such that* $(1+x)y \in [0, \alpha]$. *Then*

$$\big(\exp(\beta xy) - 1\big)\left(\frac{(1+x)}{(1+y)yx^2}\exp(\beta(x-y)) - 1\right) \geq 1.1. \tag{27}$$

Proof. We will show that for $x, y, (1+x)y \in (0, \alpha]$ we have

$$\frac{(1+x)}{(1+y)x}\exp(\beta(x-y)) \geq 1.15, \tag{28}$$

and

$$\big(\exp(\beta xy) - 1\big)\left(\frac{1.15}{xy} - 1\right) \geq 1.1. \tag{29}$$

To see that (28) and (29) imply (27) note

$$\big(\exp(\beta xy) - 1\big)\left(\frac{(1+x)}{(1+y)yx^2}\exp(\beta(x-y)) - 1\right) \geq \big(\exp(\beta xy) - 1\big)\left(\frac{1.15}{yx} - 1\right) \geq 1.1,$$

where the first inequality follows from (28) and the second from (29).

Note that the constraints on x, y imply that $y + xy \leq \alpha$ and $xy \leq \alpha y$. Hence $xy \leq \alpha/(1+1/\alpha)$. To prove (29) it is sufficient to show for $z \in [0, \alpha/(1+1/\alpha)]$

$$\big(\exp(\beta z) - 1\big)\,(1.15 - z) - 1.1z \geq 0. \tag{30}$$

We have

$$\frac{\partial^2}{\partial z^2}\Big(\big(\exp(\beta z) - 1\big)\,(1.15 - z) - 1.1z\Big) = \exp(\beta z)\beta\,(1.15\beta - \beta x - 2) < 0.$$

Hence $\big(\exp(\beta z) - 1\big)\,(1.15 - z) - 1.1z$ is concave and we only need to check (30) for the endpoints of the interval; for $z = 0$ LHS of (30) is zero, for $z = \alpha/(1+1/\alpha)$ the LHS of (30) has value larger than 0.005. This concludes the proof of (29).

To prove (28) note that

$$\frac{\partial}{\partial y}(1+x)\exp(\beta(x-y)) - 1.15(1+y)x = -(1+x)\beta\exp(\beta(x-y)) - 1.15x < 0. \tag{31}$$

Hence we only need to prove (28) for $y = \alpha/(1+x)$; this simplifies to showing

$$\exp(1.3(x - 1.1/(1+x))) \geq \left(\frac{1.15(2.1+x)x}{(1+x)^2}\right). \tag{32}$$

For $x = 0$ and $x = 1.1$ we have that (32) is satisfied (using interval arithmetic). Let

$$Q(x) := 1.3(x - 1.1/(1+x)) - \ln\left(\frac{1.15(2.1+x)x}{(1+x)^2}\right).$$

The critical points of $Q(x)$ are roots of

$$1330x^4 + 5330x^3 + 8290x^2 + 3733x - 2100.$$

Since this polynomial is increasing when x is non-negative, it has exactly one positive real root, which is in the interval $[0.30765554, 0.30765555]$. The value of $Q(x)$ on both endpoints of the interval is at least 0.0032. The derivative of $Q(x)$ (a rational function) is bounded in absolute value by 1 on the interval and hence $Q(x) > 0$ on the entire interval (in particular at the critical point). This proves (32). □

We can now complete the proof of Claim 3.3.

https://doi.org/10.1017/S0963548324000348 Published online by Cambridge University Press

Proof of Claim 3.3. We will use the following substitution to simplify the expression $F(x) = (1+x)^2 H(x)$. Note that $H(x) = 1000\exp(1.3x) = 1000\exp(\beta x)$. In terms of $H(x)$ we have

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

Let $\lambda'_v = x$ and $\lambda'_u = y(1+x)$. We have

$$1+s_1 = \frac{H(y(1+x))}{H(y)} \quad \text{and} \quad 1+s_2 = \frac{1+x}{y(1+y)x^2} \cdot \frac{H(x) - \frac{1}{(1+x)^2}}{H(y)}.$$

Recall, our goal is to show $s_1 s_2 \geq 1$. First we show that for $x, y \in (0, 1.1]$ such that $(1+x)y \leq 1.1$ we have

$$s_2 = \frac{1+x}{y(1+y)x^2} \cdot \frac{H(x) - \frac{1}{(1+x)^2}}{H(y)} - 1 \geq \frac{999}{1000}\left(\frac{1+x}{y(1+y)x^2} \cdot \frac{H(x)}{H(y)} - 1\right). \tag{33}$$

Note that (33) is equivalent to

$$\frac{1+x}{x^2}\exp(1.3x) - \frac{1}{x^2(1+x)} \geq y(1+y)\exp(1.3y). \tag{34}$$

Note that the RHS of (34) is increasing in y and hence it is sufficient to show (34) for the maximal value of $y = 1.1/(1+x)$; this is equivalent to

$$(1+x)^3\exp(1.3x) - (1+x) \geq 1.1x^2(2.1+x)\exp(1.43/(1+x)),$$

the last inequality follows from

$$(1+x)^3\left(1 + 1.3x + \frac{1.3^2}{2}x^2 + \frac{1.3^3}{6}x^3 + \frac{1.3^4}{24}x^4\right) - (1+x) \geq 1.1x^2(2.1+x)\exp(1.43),$$

checked using Sturm sequences. This concludes the proof of (33).

Note that Lemma 3.4 is equivalent to the following

$$s_1\left(\frac{(1+x)}{y(1+y)x^2}\frac{H(x)}{H(y)} - 1\right) = \left(\exp(\beta xy) - 1\right)\left(\frac{(1+x)}{(1+y)yx^2}\exp(\beta(x-y)) - 1\right) \geq 1.1,$$

and hence

$$s_1 \geq 1.1\left(\frac{(1+x)}{y(1+y)x^2}\frac{H(x)}{H(y)} - 1\right)^{-1},$$

which combined with (33) yields $s_1 s_2 \geq 1.1(999/1000) > 1$, concluding the proof. □

4. Spectral independence for trees

Let $G = (V, E)$ be a graph. Suppose we have a set of fugacities λ_i for each $i \in V$, and consider the corresponding Gibbs distribution. Recall that the influence matrix $\mathcal{I}_G$ has u, v entry $\mathrm{Inf}(u \to v)$, where

$$\mathrm{Inf}(u \to v) = \mu(v \in I | u \in I) - \mu(v \in I | u \notin I).$$

Let u, v be any two vertices in a graph. We have

$$\left|\mathrm{Inf}(v \to u)\right| \leq \max\{\mu(u \in I | v \in I), \mu(u \in I | v \notin I)\} \leq \frac{\lambda_u}{1+\lambda_u}. \tag{35}$$

We have

$$\mathrm{Inf}(v \to u)\mu(v \in I)\mu(v \notin I) = \mathrm{Inf}(u \to v)\mu(u \in I)\mu(u \notin I). \tag{36}$$

https://doi.org/10.1017/S0963548324000348 Published online by Cambridge University Press

Let D be diagonal matrix with entries $\mu(v \in I)\mu(v \notin I)$. Equation (36) means that $D\mathcal{I}_G$ is symmetric. That also means that $D^{1/2}\mathcal{I}_G D^{-1/2}$ is symmetric (since it is obtained from $D\mathcal{I}_G$ by multiplying by the same diagonal matrix on the left and on the right). The u, v entry in $D^{1/2}\mathcal{I}_G D^{-1/2}$ is

$$\begin{aligned}\mathrm{Inf}(u \to v)\,(\mu(v \in I)\mu(v \notin I))^{-1/2}\,(\mu(u \in I)\mu(u \notin I))^{1/2} = \\ \mathrm{Inf}(v \to u)\,(\mu(u \in I)\mu(u \notin I))^{-1/2}\,(\mu(v \in I)\mu(v \notin I))^{1/2}\,,\end{aligned} \tag{37}$$

which is equal to $\pm\sqrt{\mathrm{Inf}(u \to v)\mathrm{Inf}(v \to u)}$ (take geometric mean of the sides of (37)). We will call $M = D^{1/2}\mathcal{I}_G D^{-1/2}$ the symmetrised influence matrix (since it is similar to the influence matrix, it has the same spectral radius).

We will prove the following result.

Lemma 4.1. *For any forest T with fugacities in* $[0, 1.3]$ *the spectral radius of the influence matrix of the hard-core model on T is bounded by 10000.*

We will prove Lemma 4.1 by induction; we will prove the following strengthened statement.

Lemma 4.2. *For any forest T with fugacities in* $[0, 1.3]$ *the symmetrised influence matrix M of the hard-core model on T satisfies*

$$M \preceq \mathrm{diag}\Big(\frac{400}{1.3 - \lambda_v^{0.78}},\ v \in V\Big).$$

Proof. Let $T = (V, E)$ be a forest. Let v be a vertex of degree 1 in T. Let u be the neighbour of v in T. Let $T' = (V', E')$ be T with v removed. Let $\lambda'_i = \lambda_i$ for $i \in V' \setminus \{u\}$. Let $\lambda'_u = \lambda_u/(1 + \lambda_v)$. Let μ' be the hard-core model on T' with fugacities $\{\lambda'_i\}_{i \in V'}$. Note that μ' is the same as the marginalisation of μ to V', that is, for an independent set I of T' we have

$$\mu'(I) = \sum_{J \supseteq I} \mu(J), \tag{38}$$

where J ranges over independent sets of T that contain I.

Let M be the symmetrised influence matrix for μ and let M' be the symmetrised influence matrix for μ'. Note that (38) implies that M' is a submatrix of M (removing the column and row of M corresponding to vertex v yields M').

It is standard to show, e.g., see [[1], Lemma B.3], that

$$\mathrm{Inf}(v \to w) = \mathrm{Inf}(v \to u)\mathrm{Inf}(u \to w), \tag{39}$$

and

$$\mathrm{Inf}(w \to v) = \mathrm{Inf}(w \to u)\mathrm{Inf}(u \to v). \tag{40}$$

Let $\tau^2 = \mathrm{Inf}(v \to u)\mathrm{Inf}(u \to v)$. Note that, using (35), we have

$$\tau^2 = \mathrm{Inf}(u \to v)\mathrm{Inf}(v \to u) \le \frac{\lambda_v}{1 + \lambda_v}\frac{\lambda_u}{1 + \lambda_u}. \tag{41}$$

From (39) and (40) we have that M and M' have the following form (Q is a matrix and z is a vector):

$$M' = \begin{pmatrix} Q & z \\ z^T & 0 \end{pmatrix} \quad \text{and} \quad M = \begin{pmatrix} Q & z & \tau z \\ z^T & 0 & \tau \\ \tau z^T & \tau & 0 \end{pmatrix}.$$

Let F be an increasing function on $[0, 1.3]$. We will take $F(x) = 1/H(x)$ where $H(x) = (a - x^b)/400$ and $a = 1.3$, $b = 0.78$. (We will keep a and b as variables to elucidate the choice of a and b in the proof below.) Suppose that we know $M' \preceq \mathrm{diag}(F(\lambda'_i))$ and we want to conclude $M \preceq \mathrm{diag}(F(\lambda_i))$.

https://doi.org/10.1017/S0963548324000348 Published online by Cambridge University Press

Let W be a diagonal matrix with entries $F(\lambda_i)$ for $i \in V \setminus \{u, v\}$. We can restate our goal as follows.

$$\text{KNOW:} \begin{pmatrix} W-Q & -z \\ -z^T & F\left(\frac{\lambda_u}{1+\lambda_v}\right) \end{pmatrix} \succeq 0 \qquad \text{WANT:} \begin{pmatrix} W-Q & -z & -\tau z \\ -z^T & F(\lambda_u) & -\tau \\ -\tau z^T & -\tau & F(\lambda_v) \end{pmatrix} \succeq 0.$$

Since F will be an increasing function, we have $F(\lambda_u/(1+\lambda_v)) < F(\lambda_u)$.

The condition we want is equivalent (by applying row and column operations, specifically subtracting u-th row/column times τ from the v-th row/column) to

$$\begin{pmatrix} W-Q & -z & 0 \\ -z^T & F(\lambda_u) & -\tau(1+F(\lambda_u)) \\ 0 & -\tau(1+F(\lambda_u)) & F(\lambda_v)+2\tau^2+\tau^2 F(\lambda_u) \end{pmatrix} \succeq 0.$$

If we show

$$\begin{pmatrix} F(\lambda_u)-F(\lambda_u/(1+\lambda_v)) & -\tau(1+F(\lambda_u)) \\ -\tau(1+F(\lambda_u)) & F(\lambda_v)+2\tau^2+\tau^2 F(\lambda_u) \end{pmatrix} \succeq 0, \tag{42}$$

then the conclusion will follow (adding the "know" positive semidefinite matrix and (42) (expanded with zeros to match dimensions) yields that the "want" matrix is positive semidefinite). Equation (42) is equivalent to checking if the determinant is positive (since the entry in the first row/column is positive), that is, we need to check

$$(F(\lambda_u)-F(\lambda_u/(1+\lambda_v)))\, F(\lambda_v) > \tau^2\, (1+F(\lambda_u/(1+\lambda_v))(2+F(\lambda_u)))\,.$$

Hence it is sufficient (using (41)) to show

$$\left(F(\lambda_u)-F\left(\frac{\lambda_u}{1+\lambda_v}\right)\right) F(\lambda_v) > \frac{\lambda_v}{1+\lambda_v}\frac{\lambda_u}{1+\lambda_u}\left(1+F\left(\frac{\lambda_u}{1+\lambda_v}\right)(2+F(\lambda_u))\right).$$

Letting $H(x) = 1/F(x)$ the condition becomes the following

$$\left(H\left(\frac{\lambda_u}{1+\lambda_v}\right)-H(\lambda_u)\right) > H(\lambda_v)\frac{\lambda_v}{1+\lambda_v}\frac{\lambda_u}{1+\lambda_u}\left(H\left(\frac{\lambda_u}{1+\lambda_v}\right)H(\lambda_u)+2H(\lambda_u)+1\right). \tag{43}$$

We are going to search for H that is 1) bounded from above by $1/300$, 2) bounded away from 0 on $[0, 1.3]$ and 3) that satisfies

$$\left(H\left(\frac{\lambda_u}{1+\lambda_v}\right)-H(\lambda_u)\right) > H(\lambda_v)\frac{\lambda_v}{1+\lambda_v}\frac{\lambda_u}{1+\lambda_u}(1+1/100). \tag{44}$$

Note that such H will also satisfy (43) since $H\left(\frac{\lambda_u}{1+\lambda_v}\right)H(\lambda_u)+2H(\lambda_u) \le 1/100$ (here the choice of $1/100$ is arbitrary; we just need something sufficiently small).

We will search for H of the following form: $H(x) \propto a - x^b$. Note that (44) is invariant under scaling of H and hence satisfying the upper bound of $1/300$ can be achieved by scaling of H. Ultimately the price we will pay for H being small is that F is big and hence we get a weaker upper bound on the spectral radius of M; we do not optimise the constants at this point. Consider the following function L (obtained as LHS of (44) divided by RHS of (44), excluding the constant term $1+1/100$):

$$L(\lambda_u, \lambda_v) := \frac{\lambda_u^b(1+\lambda_u)}{\lambda_u}\frac{(1+\lambda_v)^b-1}{(a-\lambda_v^b)\lambda_v(1+\lambda_v)^{b-1}}. \tag{45}$$

https://doi.org/10.1017/S0963548324000348 Published online by Cambridge University Press

The minimum of the first term in (45) is attained for $\lambda_u = (1-b)/b$ and hence setting $b = .78$ we have that:

$$\frac{\lambda_u^b(1+\lambda_u)}{\lambda_u} \geq \frac{(1-b)^{b-1}}{b^b} \geq \frac{1.32}{.78}. \tag{46}$$

For the second term in (45) by setting $a = 1.3$ we have the following:

$$\frac{(1+\lambda_v)^b - 1}{(a-\lambda_v^b)\lambda_v(1+\lambda_v)^{b-1}} \geq \frac{(1+\lambda_v)^b - 1}{a\lambda_v(1+\lambda_v)^{b-1}} \geq \frac{b}{a} = \frac{.78}{1.3}. \tag{47}$$

The second inequality in (47) is obtained from the following inequality which is valid for $b \in [0, 1]$ and $x \geq 0$:

$$(1+x)^b - 1 \geq bx(1+x)^{b-1}. \tag{48}$$

For the above, it suffices to prove that

$$(1+x)^b - bx(1+x)^{b-1} \geq 1. \tag{49}$$

We have that for $b \leq 1$:

$$\begin{aligned}(1+x)^b - bx(1+x)^{b-1} &= (1+x)^{b-1}(1-(b-1)x) \\ &\geq (1+x)^{b-1}(1+x)^{-(b-1)} && \text{by Bernoulli's inequality} \\ &\geq 1.\end{aligned}$$

Finally, plugging in (46) and (47) into (45) we obtain:

$$L(\lambda_u, \lambda_v) \geq \frac{1.32}{1.3} > 1.01. \tag{50}$$

Equation (50) implies that (44) is satisfied. Recall that the statement of the lemma assumed that the fugacities are in the interval $[0, 1.3]$. For $\lambda \in [0, 1.3]$ we have

$$H(\lambda) = \frac{a-\lambda^b}{400} \geq \frac{1}{10000} \quad \text{and} \quad H(\lambda) = \frac{a-\lambda^b}{400} \leq \frac{1}{300}. \tag{51}$$

This implies that H satisfies (43) and hence for

$$F(\lambda) = \frac{1}{H(\lambda)} \leq 10000,$$

we have (42) and this completes the proof of Lemma 4.2 by induction. □

We can now complete the proof for spectral independence.

Proof of Theorem 1.2. We apply Theorem 1.12 of [16] which says that if for all pinnings we have η-spectral independence and b-marginally boundedness then the mixing time of the Glauber dynamics is $C(\eta, \Delta, b)n \log n$. A pinning refers to a fixed configuration τ on a subset S of vertices; for the hard-core model a pinning of a tree T corresponds to the hard-core model on a forest which is an induced subgraph. Hence, Lemma 4.1 implies that η-spectral independence holds for all pinnings with $\eta = 10000$. The condition b-marginally boundedness (see Definition 1.9 in [16]) says that for every pinning τ on a subset $S \subset V$, for every vertex $v \notin S$, for every assignment to v denoted as $\sigma(v)$ which has positive probability in the conditional Gibbs distribution μ_τ, then the marginal probability is lower bounded as $\mu_\tau(\sigma(v)) \geq b$. This holds for $b \geq \min\{1, \lambda\}/[\lambda + (1+\lambda)^\Delta]$. Hence, [[16], Theorem 1.12] implies Theorem 1.2. □

5. Proof of Corollary 1.3

We prove Corollary 1.3 by combining the spectral independence result in Lemma 4.1 and Theorem 1.1, while we utilise [[13], Theorem 1.9].

https://doi.org/10.1017/S0963548324000348 Published online by Cambridge University Press

In [13], they introduce the notion of "complete η-spectral independence". Let μ be the hard-core model on graph $G=(V,E)$ with fugacity λ. For $\eta > 0$, complete η-spectral independence for μ corresponds to the following condition: For any induced subgraph G' of G, for the hard-core model μ' on G' such that each vertex $v \in V$ has fugacity $\lambda'_v \leq \lambda$, the corresponding influence matrix $\mathcal{I}_{G'}$ has spectral radius at most η.

Lemma 4.1 is equivalent to complete η-spectral independence for all $\lambda \leq 1.3$ with $\eta \leq 10000$. We can now prove Corollary 1.3.

***Proof of Corollary* 1.3.** For the forest T, let μ be the hard-core model with fugacity $\lambda \leq 1.3$. Also, let γ be the spectral gap for Glauber dynamics on μ. Corollary 1.3 follows by showing that $\gamma = \Omega(n^{-1})$.

First, note that if $\lambda < 1.1$, then Theorem 1.1 already implies $\gamma = \Omega(n^{-1})$. We now focus on the case where $\lambda \in [1.1, 1.3]$.

For the same forest T, let $\widehat{\mu}$ be the hard-core model on T with fugacity $\widehat{\lambda} = 1$. Let $\widehat{\gamma}$ be the spectral gap for Glauber dynamics on $\widehat{\mu}$. From Theorem 1.1, we have that

$$\widehat{\gamma} = \Omega(n^{-1}). \tag{52}$$

Lemma 4.1 and [[9], Theorem 1.9] imply the following relation between γ and $\widehat{\gamma}$: for $\theta = \lambda$ and $\eta \leq 10000$, we have

$$\gamma \geq \left(\frac{\theta}{2}\right)^{2\eta+7} \widehat{\gamma} > \left(\frac{1}{2}\right)^{20007} \widehat{\gamma}. \tag{53}$$

From (52) and (53), we get that $\gamma = \Omega(n^{-1})$, for any $\lambda \in [1.1, 1.3]$.

From the above we have established Corollary 1.3. □

Acknowledgements

We thank Zongchen Chen for pointing out the counterexample to entropy tensorization as discussed in Remark 1.4, and the anonymous referee for pointing out Corollary 1.3.

References

1. Anari, N., Liu, K. and Gharan, S. O. (2020) Spectral independence in high-dimensional expanders and applications to the hardcore model. FOCS 2020: 1319–1330.
2. Berger, N., Kenyon, C., Mossel, E. and Peres, Y. (2005) Glauber dynamics on trees and hyperbolic graphs. *Probab. Theory Relat. Fields* **131**(3) 311–340.
3. Bhatnagar, N., Sly, A. and Tetali, P. (2016) Decay of correlations for the hardcore model on the d-regular random graph. *Electron. J. Probab.* **21** 1–42.
4. Blanca, A., Caputo, P., Chen, Z., Parisi, D., Štefankovič, D. and Vigoda, E. (2022) On Mixing of Markov Chains: Coupling, Spectral Independence, and Entropy Factorization. pp. 3670–3692, In Proceedings of the 33rd Annual ACM-SIAM Symposium on Discrete Algorithms (SODA). SIAM.
5. Brightwell, G. and Winkler, P. (2004) A second threshold for the hard-core model on a Bethe lattice. *Random Struct. Algor.* **24**(3) 303–314.
6. Caputo, P. (2023) Lecture notes on Entropy and Markov Chains. Preprint, available from: http://www.mat.uniroma3.it/users/caputo/entropy.pdf.
7. Caputo, P., Menz, G. and Tetali, P. (2015) Approximate tensorization of entropy at high temperature. Annales de la Faculté des sciences de Toulouse. *Mathématiques* **24**(4) 691–716.
8. Caputo, P. and Parisi, D. (2021) Block factorization of the relative entropy via spatial mixing. *Commun. Math. Phys.* **388**(2) 793–818.
9. Chen, X., Yang, X., Yin, Y., and Zhang, X.. Spectral independence beyond total influence on trees and related graphs. arXiv preprint arXiv:2404.04668, 2024.
10. Chen, Z. (2024) Combinatorial approach for factorization of variance and entropy in spin systems. pp. 4988–5012, In Proceedings of the 35th Annual ACM-SIAM Symposium on Discrete Algorithms (SODA).
11. Chen, Y. and Eldan, R. (2022) Localization schemes: A framework for proving mixing bounds for Markov chains. pp. 110–122, In Proceedings of the 63rd Annual IEEE Symposium on Foundations of Computer Science (FOCS).

https://doi.org/10.1017/S0963548324000348 Published online by Cambridge University Press

12. Chen, X., Feng, W., Yin, Y. and Zhang, X. (2022) Optimal mixing for two-state anti-ferromagnetic spin systems. pp. 588–599, In Proceedings of the 63rd Annual IEEE Symposium on Foundations of Computer Science (FOCS).
13. Chen, X., Feng, W., Yin, Y. and Zhang, X. (2024) Rapid mixing of Glauber dynamics via spectral independence for all degrees. *SIAM J. Comput.* FOCS21-224–FOCS21-298.
14. Chen, Z. and Gu, Y. (2024) Fast sampling of b-matchings and b-edge covers, pp. 4972–4987, In Proceedings of the 35th Annual ACM-SIAM Symposium on Discrete Algorithms (SODA).
15. Chen, Z., Liu, K., Mani, N. and Moitra, A. (2023) Strong spatial mixing for colorings on trees and its algorithmic applications. pp. 810–845, In Proceedings of the 64th Annual IEEE Symposium on Foundations of Computer Science (FOCS).
16. Chen, Z., Liu, K. and Vigoda, E. (2021) Optimal mixing of Glauber dynamics: Entropy factorization via high-dimensional expansion. pp. 1537–1550, In Proceedings of the 53rd Annual ACM Symposium on Theory of Computing (STOC).
17. Chen, Z., Liu, K. and Vigoda, E. (2022) Spectral independence via stability and applications to Holant-type problems. pp. 149–160, In Proceedings of the 63rd Annual IEEE Symposium on Foundations of Computer Science (FOCS).
18. Delcourt, M., Heinrich, M. and Perarnau, G. (2020) The Glauber dynamics for edge-colorings of trees. *Random Struct. Algor.* **57**(4) 1050–1076.
19. Eppstein, D. and Frishberg, D. (2023) Rapid mixing of the hardcore Glauber dynamics and other Markov chains in bounded-treewidth graphs, pp. 30:1–30:13, In Proceedings of the 34th International Symposium on Algorithms and Computation (ISAAC), vol. 283. Schloss Dagstuhl - Leibniz-Zentrum fur Informatik.
20. Efthymiou, C., Hayes, T. P., Štefankovič, D. and Vigoda, E. (2023). Optimal mixing via tensorization for random independent sets on arbitrary trees. arXiv preprint arXiv:2307.07727, version 2.
21. Galanis, A., Štefankovič, D. and Vigoda, E. (2016) Inapproximability of the partition function for the antiferromagnetic Ising and hard-core models. *Comb. Probab. Com.* **25**(4) 500–559.
22. Hayes, T. P. and Sinclair, A. (2007) A general lower bound for mixing of single-site dynamics on graphs. *Ann. Appl. Probab.* **17**(3) 931–952.
23. Liu, K. (2021) From coupling to spectral independence and blackbox comparison with the down-up walk. In *APPROX-RANDOM*, pp. 32:1–32:21.
24. Lucier, B., Molloy, M. and Peres, Y. (2009) The Glauber dynamics for colourings of bounded degree trees. In *Randomization and Approximation Techniques in Computer Science (RANDOM)*, pp. 631–645.
25. Martin, J. B. (2003) Reconstruction thresholds on regular trees. *Discrete Random Walks*, DIMACS Series in Discrete Mathematics and Theoretical Computer Science, pp.191–204.
26. Martinelli, F., Sinclair, A. and Weitz, D. (2003) The Ising model on trees: boundary conditions and mixing time. pp. 628–639, In Proceedings of the 44th Annual IEEE Symposium on Foundations of Computer Science (FOCS).
27. Martinelli, F., Sinclair, A. and Weitz, D. (2004) Fast mixing for independent sets, colorings and other models on trees. pp. 456–465, In Proceedings of the 15th Annual ACM-SIAM Symposium on Discrete Algorithms (SODA). SIAM.
28. Mossel, E. (2004) Survey: Information flow on trees. In *Graphs, Morphisms and Statistical Physics*, DIMACS Series in Discrete Mathematics and Theoretical Computer Science, pp. 155–170.
29. Mossel, E., Weitz, D. and Wormald, N. C. (2007) On the hardness of sampling independent sets beyond the tree threshold. *Probab. Theory Relat. Fields* **143**(3-4) 401–439.
30. Restrepo, R., Štefankovič, D., Vera, J. C., Vigoda, E. and Yang, L. (2014) Phase transition for Glauber dynamics for independent sets on regular trees. *SIAM J. Discrete Math.* **28**(2) 835–861.
31. Sly, A. (2010) Computational transition at the uniqueness threshold. pp. 287–296, In Proceedings of the 51st Annual IEEE Symposium on Foundations of Computer Science (FOCS).
32. Sly, A. and Sun, N. (2014) The computational hardness of counting in two-spin models on d-regular graphs. *Ann. Probab.* **42**(6) 2383–2416.
33. Štefankovič, D., Vempala, S. and Vigoda, E. (2009) Adaptive simulated annealing: A near-optimal connection between sampling and counting. *J. ACM* **56**(3) 1–36.

Cite this article: Efthymiou C, Hayes TP, Štefankovič D, and Vigoda E (2024). Optimal mixing via tensorization for random independent sets on arbitrary trees. *Combinatorics, Probability and Computing*. https://doi.org/10.1017/S0963548324000348

https://doi.org/10.1017/S0963548324000348 Published online by Cambridge University Press