

Fast Sampling via Spectral Independence Beyond Bounded-Degree Graphs

Ivona Bezáková

Department of Computer Science, Rochester Institute of Technology, NY, USA

Andreas Galanis

Department of Computer Science, University of Oxford, UK

Leslie Ann Goldberg

Department of Computer Science, University of Oxford, UK

Daniel Štefankovič

Department of Computer Science, University of Rochester, Rochester, NY, USA

Abstract

Spectral independence is a recently-developed framework for obtaining sharp bounds on the convergence time of the classical Glauber dynamics. This new framework has yielded optimal $O(n \log n)$ sampling algorithms on bounded-degree graphs for a large class of problems throughout the so-called uniqueness regime, including, for example, the problems of sampling independent sets, matchings, and Ising-model configurations.

Our main contribution is to relax the bounded-degree assumption that has so far been important in establishing and applying spectral independence. Previous methods for avoiding degree bounds rely on using L^p -norms to analyse contraction on graphs with bounded connective constant (Sinclair, Srivastava, Yin; FOCS'13). The non-linearity of L^p -norms is an obstacle to applying these results to bound spectral independence. Our solution is to capture the L^p -analysis recursively by amortising over the subtrees of the recurrence used to analyse contraction. Our method generalises previous analyses that applied only to bounded-degree graphs.

As a main application of our techniques, we consider the random graph $G(n, d/n)$, where the previously known algorithms run in time $n^{O(\log d)}$ or applied only to large d . We refine these algorithmic bounds significantly, and develop fast nearly linear algorithms based on Glauber dynamics that apply to all constant d , throughout the uniqueness regime.

1 Introduction

Spectral independence method was introduced by Anari, Liu, and Oveis Gharan [2] as a framework to obtain polynomial bounds on the mixing time of Glauber dynamics. Originally based on a series of works on high-dimensional expansion [17, 9, 21, 18, 1], it has since then been developed further using entropy decay by Chen, Liu, and Vigoda [8] who obtained optimal $O(n \log n)$ mixing results on graphs of bounded maximum degree Δ whenever the

framework applies. This paper focuses on relaxing the bounded-degree assumption of [8], in sparse graphs where the maximum degree is not the right parameter to capture the density of the graph.

As a running example we will use the problem of sampling (weighted) independent sets, also known as the sampling problem from the hard-core model. For a graph $G = (V, E)$, the hard-core model with parameter $\lambda > 0$ specifies a distribution $\mu_{G,\lambda}$ on the set of independent sets of G , where for an independent set I it holds that $\mu_{G,\lambda}(I) = \lambda^{|I|}/Z_{G,\lambda}$ where $Z_{G,\lambda}$ is the partition function of the model (the normalising factor that makes the probabilities add up to 1). For bounded-degree graphs of maximum degree $d + 1$ (where $d \geq 2$ is an integer), it is known that the problems of sampling and approximately counting from this model undergo a computational transition at $\lambda_c(d) = \frac{d^d}{(d-1)^{d+1}}$, the so-called uniqueness threshold [29, 27, 12]: they are poly-time solvable when $\lambda < \lambda_c(d)$, and computationally intractable for $\lambda > \lambda_c(d)$. Despite this clear complexity picture, prior to the introduction of spectral independence, the algorithms for $\lambda < \lambda_c(d)$ were based on elaborate enumeration techniques whose running times scale as $n^{O(\log d)}$ [29, 19, 22, 23]. The analysis of Glauber dynamics¹ using spectral independence in the regime $\lambda < \lambda_c(d)$ yielded initially $n^{O(1)}$ algorithms for any d [2], and then $O(n \log n)$ for bounded-degree graphs [8] (see also [7]). More recently, Chen, Feng, Yin, and Zhang [5] obtained $O(n^2 \log n)$ results for arbitrary graphs $G = (V, E)$ that apply when $\lambda < \lambda_c(\Delta_G - 1)$, where Δ_G is the maximum degree of G (see also [14] for related results when Δ_G grows like $\log n$).

The maximum degree is frequently a bad measure of the density of the graph, especially for graphs with unbounded-degree. One of the most canonical examples is the random graph $G(n, d/n)$ where the maximum degree grows with n but the average degree is d , and therefore one would hope to be able to sample from $\mu_{G,\lambda}$ for λ up to some constant, instead of $\lambda = o(1)$ that the previous results yield. In this direction, [26, 24] obtained an $n^{O(\log d)}$ algorithm based on correlation decay that applies to all $\lambda < \lambda_c(d)$ for all graphs with “connective constant” bounded by d (meaning, roughly, that for all $\ell = \Omega(\log n)$ the number of length- ℓ paths starting from any vertex is bounded by d^ℓ). The result of [24] applies to $G(n, d/n)$ for all $d > 0$. In terms of Glauber dynamics on $G(n, d/n)$, [20] showed an $n^{1+\Omega(1/\log \log n)}$ lower bound on the mixing time in the case of the Ising model; this lower bound actually applies to most well-known models, and in particular rules out $O(n \log n)$ mixing time results for the hard-core model when $\lambda = \Omega(1)$. The mixing-time lower bound on $G(n, d/n)$ has only been matched by complementary fast mixing results in models with strong monotonicity properties, see [20] for the ferromagnetic Ising model and [4] for the random-cluster model. Such monotonicity properties unfortunately do not hold for the hard-core model, and the best known results [10, 11] for Glauber dynamics on $G(n, d/n)$ give an n^C algorithm for $\lambda < 1/d$ and sufficiently large d (where C is a constant depending on d).

Our main contribution is to obtain nearly linear-time algorithms on $G(n, d/n)$, for all of the models considered in [24], i.e., the hard-core model, the monomer-dimer model (weighted matchings), and the antiferromagnetic Ising model. Key to our results are new spectral independence bounds for any $d > 0$ in the regime $\lambda < \lambda_c(d)$ for arbitrary graphs $G = (V, E)$ in terms of their “ d -branching value” (which resembles the connective-constant notion of [24]). To state our main theorem for the hard-core model on $G(n, d/n)$, we first extend the definition

¹ Recall, for a graph G , the Glauber dynamics for the hard-core model iteratively maintains a random independent set $(I_t)_{t \geq 0}$, where at each step t a vertex v is chosen u.a.r. and, if $I_t \cup \{v\}$ is independent, it sets $I_{t+1} = I_t \cup \{v\}$ with probability $\frac{\lambda}{\lambda+1}$, otherwise $I_{t+1} = I_t \setminus \{v\}$. The mixing time is the maximum number (over the starting I_0) of steps t needed to get within total variation distance $1/4$ of $\mu_{G,\lambda}$, see Section 4.1 for the precise definitions.

of $\lambda_c(d)$ to all reals $d > 0$ by setting $\lambda_c(d) = \frac{d^d}{(d-1)^{d+1}}$ for $d > 1$, and $\lambda_c(d) = \infty$ for $d \in (0, 1)$. We use the term “whp over the choice of $G \sim G(n, d/n)$ ” as a shorthand for “as n grows large, with probability $1 - o(1)$ over the choice of $G(n, d/n)$ ”. An ε -sample from a distribution μ supported on a finite set Ω is a random $\sigma \in \Omega$ whose distribution ν satisfies $\|\nu - \mu\|_{\text{TV}} \leq \varepsilon$, where $\|\nu - \sigma\|_{\text{TV}} = \frac{1}{2} \sum_{\sigma \in \Omega} |\nu(\sigma) - \mu(\sigma)|$.

► **Theorem 1.** *Let $d, \lambda > 0$ be such that $\lambda < \lambda_c(d)$. For any arbitrarily small constant $\theta > 0$, there is an algorithm such that, whp over the choice of $G \sim G(n, d/n)$, when the algorithm is given as input the graph G and an arbitrary rational $\varepsilon > 0$, it outputs an ε -sample from $\mu_{G, \lambda}$ in time $n^{1+\theta} \log \frac{1}{\varepsilon}$.*

The reader might wonder why is there no constant in front of the running time (in Theorems 1, 2, and 3) or why is there no requirement that n is sufficiently large? The assumption that n is sufficiently large is taken care of in the whp condition: there is a function $f_{d, \lambda, \theta} : \mathbb{Z} \rightarrow \mathbb{R}$ such that $\lim_{n \rightarrow \infty} f_{d, \lambda, \theta}(n) = 0$ and the “whp” means with probability $\geq 1 - f_{d, \lambda, \theta}(n)$ (the function $f_{d, \lambda, \theta}$ will have value ≥ 1 for small n , making the conclusion trivial for such n). Moreover, the family of $O(n^{1+\theta})$ algorithms from Theorem 1 can be turned into an $n^{1+o(1)}$ algorithm as follows. The function $f_{d, \lambda, \theta}$ is computable (and efficiently invertible), see Remark 35 in the full version for a discussion. Let $n_0 = 1$ and $n_k > n_{k-1}$ be such that $f_{d, \lambda, 1/k}(n) \leq 1/k$ for all $n \geq n_k$. We are going to run the algorithm of Theorem 1 with $\theta = 1/k$ for $n \in \{n_{k-1}, \dots, n_k - 1\}$. Note that the “combined” algorithm succeeds with probability $1 - o(1)$ and runs in time $n^{1+o(1)}$.

We further remark here that the algorithm of Theorem 1 (as well as Theorems 2 and 3 below) can also recognise in time $n^{1+o(1)}$ whether the graph $G \sim G(n, d/n)$ is a “good” graph, i.e., we can formulate graph properties that guarantee the success of the algorithm, are satisfied whp, and are also efficiently verifiable, see Section C.4 in the full version for details.

The key to obtaining Theorem 1 is to bound the spectral independence of $G(n, d/n)$. The main strategy that has been applied so far to bound spectral independence is to adapt suitably correlation decay arguments and, therefore, it is tempting to use the correlation decay analysis of [24]. This poses new challenges in our setting since [24] uses an L^p -norm analysis of correlation decay on trees, and the non-linearity of L^p -norms is an obstacle to converting their analysis into spectral independence bounds (in contrast, for bounded-degree graphs, the L^∞ -norm is used which can be converted to spectral independence bounds using a purely analytic approach, see [7]). Our solution to work around that is to “linearise” the L^p -analysis by taking into account the structural properties of subtrees. This allows us to amortise over the tree-recurrence using appropriate combinatorial information (the d -branching values) and to bound subsequently spectral independence; details are given in Section 3, see Lemmas 10 and 12 (and equation (2) that is at the heart of the argument). Once the spectral independence bound is in place, further care is needed to obtain the fast nearly linear running time, paying special attention to the distribution of high-degree vertices inside $G(n, d/n)$ and to blend this with the entropy-decay tools developed in [8], see Section 4.2 for this part of the argument.

In addition to our result for the hard-core model, we also obtain similar results for the Ising and the matchings models. The configurations of the Ising model on a graph $G = (V, E)$ are assignments $\sigma \in \{0, 1\}^V$ which assign the spins 0 and 1 to the vertices of G . The Ising model with parameter $\beta > 0$ corresponds to a distribution $\mu_{G, \beta}$ on $\{0, 1\}^V$, where for an assignment $\sigma \in \{0, 1\}^V$, it holds that $\mu_{G, \beta}(\sigma) = \beta^{m(\sigma)} / Z_{G, \beta}$ where $m(\sigma)$ is the number of edges whose endpoints have the same spin assignment under σ , and $Z_{G, \beta}$ is the partition function of the model. The model is antiferromagnetic when $\beta \in (0, 1)$, and ferromagnetic otherwise. For $d \geq 1$, let $\beta_c(d) = \frac{d-1}{d+1}$; for $d \in (0, 1)$, let $\beta_c(d) = 0$. It is known that on

bounded-degree graphs of maximum degree $d + 1$ the sampling/counting problem for the antiferromagnetic Ising model undergoes a phase transition at $\beta = \beta_c(d)$, analogous to that for the hard-core model [25, 19, 28, 13].

► **Theorem 2.** *Let $d, \beta > 0$ be such that $\beta \in (\beta_c(d), 1)$. For any constant $\theta > 0$, there is an algorithm such that, whp over the choice of $G \sim G(n, d/n)$, when the algorithm is given as input the graph G and an arbitrary rational $\varepsilon > 0$, it outputs an ε -sample from $\mu_{G, \beta}$ in time $n^{1+\theta} \log \frac{1}{\varepsilon}$.*

For a graph $G = (V, E)$, the matchings model with parameter $\gamma > 0$, also known as the monomer-dimer model, corresponds to a distribution $\mu_{G, \gamma}$ on the set of matchings of G , where for a matching M , it holds that $\mu_{G, \gamma}(M) = \gamma^{|M|} / Z_{G, \gamma}$ where $Z_{G, \gamma}$ is the partition function. For general graphs $G = (V, E)$ and $\gamma = O(1)$, [15, 16] gave an $O(n^2 m \log n)$ algorithm (where $n = |V|, m = |E|$), which was improved for bounded-degree graphs to $O(n \log n)$ in [8] using spectral independence. For $G(n, d/n)$, [24] gave an $O(n^{\log d})$ deterministic algorithm using correlation decay, and [14] showed that Glauber dynamics mixes in $n^{2+o(1)}$ steps in the case that $\gamma = 1$.

► **Theorem 3.** *Let $d, \gamma > 0$. For any constant $\theta > 0$, there is an algorithm such that, whp over the choice of $G \sim G(n, d/n)$, when the algorithm is given as input the graph G and an arbitrary rational $\varepsilon > 0$ outputs an ε -sample from $\mu_{G, \gamma}$ in time $n^{1+\theta} \log \frac{1}{\varepsilon}$.*

In the next section, we give the main ingredients of our algorithm for the hard-core model and we give the proof of Theorem 1. The proofs of Theorems 2 and 3 build on similar ideas, though there are some modifications needed to obtain the required spectral independence bounds. We give their proofs in Section B.3 of the full version.

Before proceeding let us finally mention that, to go beyond the 2-spin models studied here, the main obstacle is to establish the spectral independence bounds for graphs with potentially unbounded degrees. As it is pointed out in [24, Section 7], their correlation-decay analysis in terms of the connective constant using L^p -norms does not extend to other models in a straightforward manner, and hence it is natural to expect that the same is true for spectral independence as well.

2 Proof outline for Theorem 1

Our algorithm for sampling from the hard-core model on a graph $G = (V, E)$ is an adaptation of Glauber dynamics on an appropriate set of “small-degree” vertices U , the details of the algorithm are given in Figure 1. Henceforth, analogously to the Ising model, it will be convenient to view the hard-core model as a 2-spin model supported on $\Omega \subseteq \{0, 1\}^V$, where Ω corresponds to the set of independent sets of G (for an independent set I , we obtain $\sigma \in \{0, 1\}^V$ by setting $\sigma_v = 1$ iff $v \in I$).

Note that for general graphs G , implementing Steps 2 and Steps 3 of the algorithm might be difficult. The following lemma exploits the sparse structure of $G(n, d/n)$ and in particular the fact that high-degree vertices are sparsely scattered. We will use this in the proof of our main theorems to show that the algorithm $\text{SAMPLE}(G, T)$ can be implemented very efficiently for appropriate D , paying only $O(\log n)$ per loop operation in Step 2 and only $O(n \log n)$ in Step 3. The *tree-excess* of a graph $G = (V, E)$ is defined as $|E| - |V| + 1$.

► **Lemma 4.** *Let $d > 0$ be an arbitrary real. There exist constants $D, \ell > 0$ such that the following holds whp over the choice of $G = (V, E) \sim G(n, d/n)$. Each of the connected components of $G[V \setminus U]$, where U is the set of vertices of degree $\leq D$, has size $O(\log n)$ and tree-excess at most ℓ .*

Algorithm $\text{SAMPLE}(G, T)$.

Parameters: $D > 0$ (threshold for small/high degree vertices).

Input: Graph $G = (V, E)$, integer $T \geq 1$ (number of iterations).

1. **Initialisation:** Let U be the set of all vertices with degree $\leq D$.
Let X_0 be the empty independent set on U .
2. **Main loop:** For $t = 1, \dots, T$,
 - Pick a vertex u uniformly at random from U .
 - For every vertex $v \in U \setminus \{u\}$, set $X_t(v) = X_{t-1}(v)$.
 - Sample the spin $X_t(u)$ according to $\mu_{G,\lambda}(\sigma_u \mid \sigma_{U \setminus \{u\}} = X_t(U \setminus \{u\}))$, i.e., update u according to the hard-core distribution on the *whole graph* G , conditioned on the spins of $U \setminus \{u\}$.
3. **Finalisation:** Sample $\sigma \sim \mu_{G,\lambda}(\cdot \mid \sigma_U = X_T)$, i.e., extend X_T to the whole vertex set of G by sampling from $\mu_{G,\lambda}$ conditioned on the configuration on U .

■ **Figure 1** The $\text{SAMPLE}(G, T)$ subroutine for sampling from the hard-core distribution $\mu_{G,\lambda}$. We use the analogue of this algorithm for the Ising model with parameter β (replacing $\mu_{G,\lambda}$ by $\mu_{G,\beta}$). For the monomer-dimer model, the only difference is that the algorithm needs to update (single) edges in F , where F is the set of vertices whose both endpoints lie in U (i.e., degree $\leq D$).

Lemma 4 follows using relatively standard techniques from random graphs and is proved in Section C of the full version. Later, we will establish a more refined version of this property that will allow us to bound the mixing time of the single-site dynamics that we consider (the main loop of $\text{SAMPLE}(G, T)$).

The key ingredient needed to prove our main result is to show that the main loop of our sampling algorithm returns a good sample on the induced hard-core distribution on the set U . More precisely, for a graph $G = (V, E)$ and $U \subseteq V$, we let $\mu_{G,\lambda,U}(\cdot)$ denote the induced distribution on the spins of U , i.e., the marginal distribution $\mu_{G,\lambda}(\sigma_U = \cdot)$.

► **Lemma 5.** *Let $d, \lambda > 0$ be constants such that $\lambda < \lambda_c(d)$. For any arbitrarily small constant $\theta > 0$, there is $D > 0$ such that the following holds whp over the choice of $G \sim G(n, d/n)$.*

Let U be the set of vertices in G of degree $\leq D$. Then, for any $\varepsilon > 0$, for $T = \lceil n^{1+\theta/2} \log \frac{1}{\varepsilon} \rceil$, the main loop of $\text{SAMPLE}(G, T)$ returns a sample X_T from a distribution which is ε -close to $\mu_{G,\lambda,U}$.

We will prove Lemma 5 in Section 4.2. With these two lemmas we are ready to prove Theorem 1.

Proof of Theorem 1. We give first the details for the more interesting case $d \geq 1$. Consider arbitrarily small $\theta > 0$ and D, ℓ as in Lemmas 4 and 5, so that whp G satisfies the properties therein. Let $\varepsilon > 0$ be the desired accuracy for sampling from $\mu_{G,\lambda}$; it is sufficient to consider $\varepsilon < 1/e$. Let U be the set of vertices with degree $\leq D$, and set $T = \lceil n^{1+\theta/2} \log \frac{1}{\varepsilon} \rceil$.

By Lemma 5, whp over the choice of G , the main loop of $\text{SAMPLE}(G, T)$ returns a configuration $X_T : U \rightarrow \{0, 1\}$ that is ε -close to $\mu_{G,\lambda,U}$. Note that each iteration of the main loop of $\text{SAMPLE}(G, T)$ can be implemented in $O(\log n)$ time since $G[V \setminus U]$ has components of size $O(\log n)$ and tree excess at most ℓ . In particular, any vertex $u \in U$ can be adjacent to at most D of these components, and therefore the component of u in $G[(V \setminus U) \cup \{u\}]$ has

size $O(\log n)$ and tree excess at most $k = D \lceil \ell \rceil = O(1)$. We can therefore sample the spin of u under $\mu_{G,\lambda}$ conditioned on the spins of $U \setminus \{u\}$ in time $O(4^k \log n) = O(\log n)$.² Therefore, the main loop of $\text{SAMPLE}(G, T)$ runs in time $O(T \log n)$. Analogously, the finalisation step of $\text{SAMPLE}(G, T)$, i.e., extending the configuration X_T on U to a configuration σ on the whole vertex set V , can be implemented in time $O(n \log n)$ by iterating over the vertices in $V \setminus U$ and using the fact that the components of $G[V \setminus U]$ have excess at most ℓ . Therefore, the overall running time of the algorithm is bounded by $O(T \log n) + O(n \log n)$, which is less than $\lceil n^{1+\theta} \log \frac{1}{\varepsilon} \rceil$ for all sufficiently large n . It remains to note that, since X_T is ε -close to the marginal distribution of $\mu_{G,\lambda}$ on U , and the finalisation step is done perfectly conditioned on the configuration on U , the final configuration σ is ε -close to the distribution $\mu_{G,\lambda}$.

For $d < 1$, whp G consists of tree-like components of size $O(\log n)$, and therefore we can obtain a perfect sample from $\mu_{G,\lambda}$ in time $O(n \log n)$ by going through the vertices one by one and, for each vertex, taking $O(\log n)$ time to compute its marginal, conditioned on the spins already sampled. \blacktriangleleft

3 Spectral independence via branching values

We first introduce the notions of spectral independence and pairwise vertex influences, which we will later use to bound the mixing time of the main loop of $\text{SAMPLE}(G, T)$, i.e., to prove Lemma 5. We will define the terminology in a general way that will be useful both for our analysis of the hard-core model, and for our later analysis of other models.

Let $q \geq 2$ be an integer indicating the number of spins and let V be a set of size n . We will consider distributions μ supported on a set $\Omega \subseteq [q]^V$.³ For $S \subseteq V$, let $\Omega_S = \{\tau \in [q]^S \mid \mu(\sigma_S = \tau) > 0\}$ be the set of all partial configurations on $[q]^S$ that have non-zero marginal under μ . For $\tau \in \Omega_S$, let μ_τ be the conditional distribution on Ω induced by τ , i.e., $\mu_\tau(\cdot) = \mu(\cdot \mid \sigma_S = \tau)$. Let $\mu_{\min} = \min_{\sigma \in \Omega} \mu(\sigma)$.

For $S \subseteq V$ and $\tau \in \Omega_S$, the *influence matrix* conditioned on τ is the matrix Ψ_τ whose rows and columns are indexed by $\tilde{V}_\tau = \{(v, i) \mid v \in V \setminus S, \mu_\tau(\sigma_v = i) > 0\}$, where the entry indexed by $(v, i), (w, k)$ equals $\mu_\tau(\sigma_w = k \mid \sigma_v = i) - \mu_\tau(\sigma_w = k)$ if $v \neq w$, and 0 otherwise. It is a standard fact that the eigenvalues of the matrix Ψ are all real ([2]), and we denote by $\lambda_1(\Psi)$ its largest eigenvalue.

► Definition 6. Let $q \geq 2$ be an integer and V be a set of size $n \geq 2$. Let μ be a distribution supported over $\Omega \subseteq [q]^V$. Let $\eta, b > 0$. We say that μ is η -spectrally independent if for all $S \subset V$ and $\tau \in \Omega_S$, it holds that $\lambda_1(\Psi_\tau) \leq \eta$. We say that μ is b -marginally bounded if for all $S \subset V$, $v \in V \setminus S$, $\tau \in \Omega_S$, and $i \in [q]$, it either holds that $\mu_\tau(\sigma_v = i) = 0$ or else $\mu_\tau(\sigma_v = i) \geq b$.

Following [2, 7], for distributions μ induced by 2-spin systems, we work with the following notion of pairwise vertex-influence, which can be used to bound the spectral independence. For a graph $G = (V, E)$ and $\tau \in \{0, 1\}^S$ for some $S \subset V$, for vertices u, v with $u \in V \setminus S$ and $0 < \mu_\tau(\sigma_u = 1) < 1$, we define the *influence* of u on v (under μ_τ) as

$$\mathcal{I}_G^\tau(u \rightarrow v) = \mu_\tau(\sigma_v = 1 \mid \sigma_u = 1) - \mu_\tau(\sigma_v = 1 \mid \sigma_u = 0).$$

² One “naive” way to do this is by considering a spanning tree and then brute-forcing over all $\leq 4^k$ possibilities for the endpoints of the excess edges (the spins on each edge can be set in at most 4 ways). For each of these, the marginal probability at u and the corresponding partition function can be computed using dynamic programming on the left-over tree.

³ For an integer $k \geq 1$, we denote by $[k]$ the set $\{0, 1, \dots, k-1\}$.

For matchings, we will work with an analogous notion from the perspective of edges (see Section B.2 of the full version). For all these models, spectral independence will be bounded by summing the absolute value of the influences of an arbitrary vertex u to the rest of the graph.

In turn, it has been shown in [7] that summing the influences of a vertex u in a graph G reduces to summing the sum of influences on the self-avoiding walk tree emanating from u , see Lemma 22 in the full version. Therefore, we only need to focus on trees arising as self-avoiding walk trees.

3.1 The branching value

We will need the following notion to capture the growth of the self-avoiding walk tree from a vertex.

► **Definition 7.** *Let $d > 0$ be a real number and $G = (V, E)$ be a graph. For a vertex v in G , the d -branching value S_v equals $\sum_{\ell \geq 0} N_{v,\ell}/d^\ell$, where $N_{v,\ell}$ is the number of (simple) paths with a total of $\ell + 1$ vertices starting from v (for convenience, we set $N_{v,0} = 1$).*

We will show the following lemma in Section C.1 which bounds the d' -branching value of $G(n, d/n)$ for any $d' > d$.

► **Lemma 8.** *Let $d \geq 1$. Then, for every $d' > d$ and $\varepsilon > 0$, whp over the choice of $G \sim G(n, d/n)$, the d' -branching value of every vertex in G is at most $\varepsilon \log n$.*

3.2 Spectral independence for the hard-core model

In this section, we bound the spectral independence of $G(n, d/n)$ in the hard-core model when $\lambda < \lambda_c(d)$. We will need the following technical lemma that can be derived from [24]. The derivation details are similar to an analogous lemma for matchings (cf. Lemma 26 in the full version), which can be found in [3, Lemma 15].

► **Lemma 9 ([24]).** *Let $d > 1$ and $\lambda > 0$ be constants such that $\lambda < \lambda_c(d)$. Let $\chi \in (1, 2)$ be given from $\frac{1}{\chi} = 1 - \frac{d-1}{2} \log\left(1 + \frac{1}{d-1}\right)$ and set $a = \frac{\chi}{\chi-1}$. Consider also the function $\Phi(y) = \frac{1}{\sqrt{y(1+y)}}$ for $y > 0$. Then, there is a constant $0 < \kappa < 1/d$ such that the following holds for any integer $k \geq 1$.*

Let x_1, \dots, x_k be reals and $x = \lambda \prod_{i=1}^k \frac{1}{1+x_i}$. Then $(\Phi(x))^a \sum_{i=1}^k \left(\frac{x}{(1+x_i)\Phi(x_i)}\right)^a \leq \kappa^{a/\chi}$.

We will show the following.

► **Lemma 10.** *Let $d > 1$ and $\lambda > 0$ be constants such that $\lambda < \lambda_c(d)$. Then, there is a constant $\chi > 1$ such that the following holds.*

Let $T = (V, E)$ be a tree rooted at ρ , whose d -branching value is $\leq \alpha$ and whose root has k children. Then, for the hard-core distribution on T with parameter λ , any $S \subseteq V \setminus \{\rho\}$ and $\tau \in \Omega_S$ with $0 < \mu_\tau(\sigma_\rho = 1) < 1$, it holds that

$$\sum_{v \in V} |\mathcal{I}_T^\tau(\rho \rightarrow v)| \leq W_k \alpha^{1/\chi},$$

where $W_k > 0$ is a real depending only on the degree k of the root (and the constants d, λ).

Proof. Let $\kappa \in (0, 1/d)$ and $\chi \in (1, 2)$ be the constants from Lemma 9, and $\Phi(x) = \frac{1}{\sqrt{x(1+x)}}$ be also as in Lemma 9.

We may assume without loss of generality that S is empty (and τ is trivial) by truncating the tree T using the following procedure: just remove vertices $u \in S$ with $\tau_u = 0$, and for $u \in S$ with $\tau_u = 1$ remove u and all of its neighbours. Note that for all the removed vertices v it holds that $\mathcal{I}_T^\tau(\rho \rightarrow v) = 0$, so the removal procedure does not decrease the sum of the absolute influences, while at the same time decreasing the d -branching value of the tree T . Henceforth, we will drop τ and S from notation.

To prove the lemma, we will work inductively on the depth of the tree. To this end, we first define for each vertex u in T the following values α_u and R_u ; the α 's capture a rooted analogue of the branching value of internal vertices within T , while the R 's the marginals of the vertices in the corresponding subtrees. More precisely, if v is a leaf, set $\alpha_v = 1$ and $R_v = \lambda$; otherwise set $\alpha_v = 1 + \frac{1}{d} \sum_{i=1}^k \alpha_{v_i}$ and $R_v = \lambda \prod_{i=1}^k \frac{1}{1+R_{v_i}}$, where v_1, \dots, v_k are the children of v . Note that for the root ρ we have that $\alpha_\rho = S_\rho \leq \alpha$, where S_ρ is the d -branching value of ρ in the tree T . Moreover, if we denote by T_v the subtree of T rooted at v and by u the parent of v in T , then it holds that

$$R_v = \frac{\mu_{T_v, \lambda}(\sigma_v = 1)}{\mu_{T_v, \lambda}(\sigma_v = 0)} \quad \text{and} \quad \mathcal{I}_T(u \rightarrow v) = -\frac{R_v}{R_v + 1}. \quad (1)$$

The first equality is fairly standard and can be proved using induction on the height of the tree, while the second one is [7, Lemma 15] (it also follows directly from the definition of influence and the first equality).

For an integer $h \geq 0$, let $L(h)$ be the nodes at distance h from the root ρ . Let $M_k = \sqrt{1 + (1 + \lambda)^k / \lambda}$, where recall that k is the degree of the root ρ . We will show that

$$\sum_{v \in L(h)} \left(\frac{\alpha_v}{\alpha_\rho} \right)^{1/\chi} \frac{|\mathcal{I}_T(\rho \rightarrow v)|}{R_v \Phi(R_v)} \leq M_k (d\kappa)^{h/\chi}. \quad (2)$$

Since $\alpha_v \geq 1$ for $v \in V$, $\alpha_\rho \leq \alpha$ and $R_v \Phi(R_v) \leq 1$, (2) yields $\sum_{v \in L(h)} |\mathcal{I}_T(\rho \rightarrow v)| \leq M_k \alpha^{1/\chi} (d\kappa)^{h/\chi}$ for all integer $h \geq 0$, and therefore summing over h , we obtain that

$$\sum_{v \in V} |\mathcal{I}_T(\rho \rightarrow v)| \leq M_k \alpha^{1/\chi} \sum_{h \geq 0} (d\kappa)^{h/\chi} \leq \frac{M_k \alpha^{1/\chi}}{1 - (d\kappa)^{1/\chi}},$$

which proves the result with $W_k = \frac{M_k}{1 - (d\kappa)^{1/\chi}}$. So it only remains to prove (2).

We will work inductively. The base case $h = 0$ is equivalent to $M_k \geq 1/(R_\rho \Phi(R_\rho)) = \sqrt{1 + 1/R_\rho}$, which is true since from the recursion for R_ρ we have that $R_\rho \geq \lambda/(1 + \lambda)^k$. For the induction step, consider $v \in L(h - 1)$ and suppose it has $k_v \geq 0$ children, denoted by v_i for $i \in [k_v]$. Then, for each $i \in [k_v]$, since v is on the unique path joining ρ to v_i , it holds that (see [2, Lemma B.2])

$$\mathcal{I}_T(\rho \rightarrow v_i) = \mathcal{I}_T(\rho \rightarrow v) \mathcal{I}_T(v \rightarrow v_i),$$

so we can write

$$\begin{aligned} \sum_{v \in L(h)} \left(\frac{\alpha_v}{\alpha_\rho} \right)^{1/\chi} \frac{|\mathcal{I}_T(\rho \rightarrow v)|}{R_v \Phi(R_v)} &= \\ & \sum_{v \in L(h-1)} \left(\frac{\alpha_v}{\alpha_\rho} \right)^{1/\chi} \frac{|\mathcal{I}_T(\rho \rightarrow v)|}{R_v \Phi(R_v)} \sum_{i \in [k_v]} \left(\frac{\alpha_{v_i}}{\alpha_v} \right)^{1/\chi} R_v \Phi(R_v) \frac{|\mathcal{I}_T(v \rightarrow v_i)|}{R_{v_i} \Phi(R_{v_i})}. \end{aligned} \quad (3)$$

Consider an arbitrary $v \in L(h-1)$. Then, since $\frac{1}{\chi} + \frac{1}{a} = 1$, by Hölder's inequality we have that

$$\sum_{i \in [k_v]} \left(\frac{\alpha_{v_i}}{\alpha_v} \right)^{1/\chi} R_v \Phi(R_v) \frac{|\mathcal{I}_T(v \rightarrow v_i)|}{R_{v_i} \Phi(R_{v_i})} \leq \left(\sum_{i \in [k_v]} \frac{\alpha_{v_i}}{\alpha_v} \right)^{1/\chi} \left((R_v \Phi(R_v))^a \sum_{i \in [k_v]} \left(\frac{|\mathcal{I}_T(v \rightarrow v_i)|}{R_{v_i} \Phi(R_{v_i})} \right)^a \right)^{1/a}. \quad (4)$$

Note that for $x = R_v$ and $x_i = R_{v_i}$, $i \in [k_v]$, we have from (1) that $\frac{|\mathcal{I}_T(v \rightarrow v_i)|}{R_{v_i}} = \frac{1}{1+x_i}$ and $x = \lambda \prod_{i \in [k_v]} \frac{1}{1+x_i}$, so by Lemma 9 we have that

$$\left((R_v \Phi(R_v))^a \sum_{i \in [k_v]} \left(\frac{|\mathcal{I}_T(v \rightarrow v_i)|}{R_{v_i} \Phi(R_{v_i})} \right)^a \right)^{1/a} \leq \kappa^{1/\chi}.$$

By definition of the d -branching value we also have $\alpha_v = 1 + \frac{1}{d} \sum_{i \in [k_v]} \alpha_{v_i} \geq \frac{1}{d} \sum_{i \in [k_v]} \alpha_{v_i}$, so plugging these back into (4) yields

$$\sum_{i \in [k_v]} \left(\frac{\alpha_{v_i}}{\alpha_v} \right)^{1/\chi} R_v \Phi(R_v) \frac{|\mathcal{I}_T(v \rightarrow v_i)|}{R_{v_i} \Phi(R_{v_i})} \leq (d\kappa)^{1/\chi}.$$

In turn, plugging this into (3) and using the induction hypothesis yields (2), finishing the proof. \blacktriangleleft

► **Remark 11.** For simplicity, and since it is not important for our arguments, the constant W_k in the proof depends exponentially on the degree k of the root. With a more careful inductive proof (cf. [7, Proof of Lemma 14]), the dependence on k can be made linear. In either case, because of the high-degree vertices in $G(n, d/n)$, both bounds do not yield sufficiently strong bounds on the spectral independence of the whole distribution $\mu_{G, \lambda}$, and this is one of the reasons that we have to consider the spectral independence on the induced distribution on low-degree vertices.

Recall that for a graph $G = (V, E)$ and $U \subseteq V$, we let $\mu_{G, \lambda, U}(\cdot)$ denote the marginal distribution on the spins of U , i.e., the distribution $\mu_{G, \lambda}(\sigma_U = \cdot)$.

► **Lemma 12.** *Let $d \geq 1$ and $\lambda > 0$ be constants such that $\lambda < \lambda_c(d)$. Then, for any constants $D, \varepsilon > 0$, whp over the choice of $G \sim G(n, d/n)$, the marginal hard-core distribution $\mu_{G, \lambda, U}$, where U is the set of vertices in G with degree $\leq D$, is $(\varepsilon \log n)$ -spectrally independent.*

Proof. Let $D, \varepsilon > 0$ be arbitrary constants, and let $d' > d$ be such that $\lambda < \lambda_c(d')$; such d' exists because the function $\lambda_c(\cdot)$ is continuous in the interval $(1, \infty)$ and $\lambda_c(d) \rightarrow \infty$ for $d \downarrow 1$. Let $\chi \in (1, 2)$ and $W = \max\{W_1, \dots, W_D\}$ where χ and the W_k 's are as in Lemma 10 (corresponding to the constants d', λ). By Lemma 8, whp all of the vertices of the graph $G = (V, E) \sim G(n, d/n)$ have d' -branching value less than $\varepsilon \log n$. We will show that the result holds for all such graphs G (for sufficiently large n).

Let U be the set of vertices in G with degree $\leq D$, and let for convenience $\mu = \mu_{G, \lambda, U}$. Consider arbitrary $S \subset U$ and $\tau \in \Omega_S$. It suffices to bound the largest eigenvalue of the influence matrix Ψ_τ by $\varepsilon \log n$. Analogously to [2, 7], we do this by bounding the absolute-value row sums of Ψ_τ . Recall that the rows and columns of Ψ_τ are indexed by $\tilde{V}_\tau = \{(v, i) \mid v \in U \setminus S, \mu_\tau(\sigma_v = i) > 0\}$, where the entry indexed by $(v, i), (w, k)$ equals $\mu_\tau(\sigma_w = k \mid \sigma_v = i) - \mu_\tau(\sigma_w = k)$ if $v \neq w$, and 0 otherwise. Consider arbitrary $(v, i) \in \tilde{V}_\tau$; our goal is to show

$$\sum_{(w,k) \in \tilde{V}_\tau} |\mu_\tau(\sigma_w = k \mid \sigma_v = i) - \mu_\tau(\sigma_w = k)| \leq \varepsilon \log n. \quad (5)$$

Henceforth, we will also assume that $\mu_\tau(\sigma_v = i) < 1$ (in addition to $\mu_\tau(\sigma_v = i) > 0$), otherwise the sum on the l.h.s. is equal to 0. Then, by the law of total probability, for any $(w, k) \in \tilde{V}_\tau$ we have

$$\begin{aligned} |\mu_\tau(\sigma_w = k \mid \sigma_v = i) - \mu_\tau(\sigma_w = k)| &\leq |\mu_\tau(\sigma_w = k \mid \sigma_v = 1) - \mu_\tau(\sigma_w = k \mid \sigma_v = 0)| \\ &= |\mathcal{I}_G^\tau(v \rightarrow w)|, \end{aligned}$$

where the last equality follows from the fact that μ is the marginal distribution of $\mu_{G,\lambda}$ on U . Therefore, we can bound

$$\sum_{(w,k) \in \tilde{V}_\tau} |\mu_\tau(\sigma_w = k \mid \sigma_v = i) - \mu_\tau(\sigma_w = k)| \leq \sum_{w \in U} |\mathcal{I}_G^\tau(v \rightarrow w)| \leq \sum_{w \in V} |\mathcal{I}_G^\tau(v \rightarrow w)|.$$

By Lemma 22 in the full version, for the self-avoiding walk tree $T = (V_T, E_T)$ from v , there is a subset $Z \subseteq V_T \setminus \{\rho\}$ and a configuration $\phi \in \{0, 1\}^Z$ such that

$$\sum_{w \in V} |\mathcal{I}_G^\tau(v \rightarrow w)| \leq \sum_{w \in V_T} |\mathcal{I}_T^\phi(v \rightarrow w)|,$$

where $\mathcal{I}_T^\phi(v \rightarrow \cdot)$ denotes the influence of v on the vertices of T (in the hard-core distribution $\mu_{T,\lambda}$ conditioned on ϕ). Since the d' -branching value of v (and any other vertex of G) is bounded by $\varepsilon \log n$ and the degree of v is $\leq D$, by Lemma 10 applied to T , we have that

$$\sum_{w \in V_T} |\mathcal{I}_T^\phi(v \rightarrow w)| \leq W(\varepsilon \log n)^{1/\chi}.$$

Since $\chi > 1$, for all sufficiently large n we have that $W(\varepsilon \log n)^{1/\chi} \leq \varepsilon \log n$, which proves (5). \blacktriangleleft

We also record the following corollary of the arguments in Lemma 10.

► Corollary 13. *Let $\lambda > 0$ and $D > 0$ be real numbers. For a graph $G = (V, E)$, let U be the set of vertices in G with degree $\leq D$ and suppose that $|U| \geq 2$. Then, the distribution $\mu := \mu_{G,\lambda,U}$ is b -marginally bounded for $b = \frac{\lambda}{\lambda + (1+\lambda)^D}$.*

Proof. By Lemma 22 in the full version, for any vertex $v \in U$ and any boundary condition τ on (a subset of) $U \setminus \{v\}$, there is a corresponding tree T and a boundary condition ϕ on T such that $\mu_\tau(\sigma_v = \cdot) = \nu_\phi(\sigma_v = \cdot)$. Since v has degree $\leq D$, from the proof of Lemma 10, see in particular equation (1), we have that $\nu_\phi(\sigma_v = \cdot) \geq b$, where b is as in the lemma statement. \blacktriangleleft

4 Entropy factorisation for bounded-degree vertices

In this section, we show how to convert the spectral independence results of the previous section into fast mixing results for Glauber dynamics on the set of small-degree vertices on $G(n, d/n)$. Our strategy here follows the technique of [8], though to obtain nearly linear results we have to pay attention to the connected components induced by high-degree vertices and how these can connect up small-degree vertices.

4.1 Preliminaries

Entropy factorisation for probability distributions

For a real function f on $\Omega \subseteq [q]^V$, we use $\mathbf{E}_\mu(f)$ for the expectation of f with respect to μ and, for $f : \Omega \rightarrow \mathbb{R}_{\geq 0}$, $\text{Ent}_\mu(f) = \mathbf{E}_\mu[f \log f] - \mathbf{E}_\mu(f) \log \mathbf{E}_\mu(f)$, with the convention that $0 \log 0 = 0$. Finally, for $S \subset V$, let $\text{Ent}_\mu^S(f) = \mathbf{E}_{\tau \sim \mu_{V \setminus S}}[\text{Ent}_{\mu_\tau}(f)]$ i.e., $\text{Ent}_\mu^S(f)$ is the expected value of the conditional entropy of f when the assignment outside of S is chosen according to the marginal distribution $\mu_{V \setminus S}$ (the induced distribution of μ on $V \setminus S$). For convenience, when $S = V$, we define $\text{Ent}_\mu^S(f) = \text{Ent}_\mu(f)$. The following inequality of entropy under tensor product is a special case of Shearer's inequalities.

► **Fact 14.** *Let $q, k \geq 2$ be integers and suppose that, for $i \in [k]$, μ_i is a distribution supported over $\Omega_i \subseteq [q]^{V_i}$, where V_1, \dots, V_k are pairwise disjoint sets. Let $\mu = \mu_1 \otimes \dots \otimes \mu_k$ be the product distribution on $\Omega = \Omega_1 \times \dots \times \Omega_k$. Then, for any $f : \Omega \rightarrow \mathbb{R}_{\geq 0}$, it holds that $\text{Ent}_\mu(f) \leq \sum_{i=1}^k \text{Ent}_{\mu_i}^{V_i}(f)$.*

To bound the mixing time of Markov chains such as the Glauber dynamics, we will be interested in establishing inequalities for factorisation of entropy, defined as follows.

► **Definition 15.** *Let $q \geq 2$, $r \geq 1$ be integers and V be a set of size $n \geq r + 1$. Let μ be a distribution supported over $\Omega \subseteq [q]^V$. We say that μ satisfies the r -uniform-block factorisation of entropy with multiplier⁴ C_r if for all $f : \Omega \rightarrow \mathbb{R}_{\geq 0}$ it holds that $\frac{r}{n} \text{Ent}_\mu(f) \leq C_r \frac{1}{\binom{n}{r}} \sum_{S \in \binom{V}{r}} \text{Ent}_\mu^S(f)$.*

The following lemma will be useful to bound the (r -uniform-block) factorisation multiplier for conditional distributions on sets with small cardinality.

► **Lemma 16** ([8, Lemma 4.2]). *Let $q \geq 2$ be an integer and V be a set of size $n \geq 2$. Let μ be a distribution supported over $\Omega \subseteq [q]^V$ which is b -marginally bounded for some $b > 0$. Then, for any $S \subseteq V$ and $\tau \in \Omega_{V \setminus S}$, for $f : \Omega \rightarrow \mathbb{R}_{\geq 0}$, it holds that $\text{Ent}_{\mu_\tau}(f) \leq \frac{2|S|^2 \log(1/b)}{b^{2|S|+2}} \sum_{v \in S} \text{Ent}_{\mu_\tau}^v(f)$.*

The r -uniform-block Glauber dynamics and its mixing time

For an integer $r = 1, \dots, n$, the r -uniform-block Glauber dynamics for μ is a Markov chain $(X_t)_{t \geq 0}$ where $X_0 \in \Omega$ is an arbitrary configuration and, for $t \geq 1$, X_t is obtained from X_{t-1} by first picking a subset $S \in \binom{V}{r}$ of size r uniformly at random and updating the configuration on S according to

$$\mu(\sigma_S = \cdot \mid \sigma_{V \setminus S} = X_{t-1}(V \setminus S)).$$

For $\varepsilon > 0$, the mixing time of the r -uniform-block Glauber dynamics is defined as $T_{\text{mix}}(\varepsilon) = \max_{\sigma \in \Omega} \min \{t \mid X_0 = \sigma, \|\nu_t - \mu\|_{\text{TV}} \leq \varepsilon\}$, where ν_t denotes the distribution of X_t . Note, the case $r = 1$ corresponds to the single-site dynamics, where at every step the spin of a single vertex, chosen u.a.r., is updated conditioned on the spins of the remaining vertices.

⁴ We note that in related works C_r is usually referred to as the ‘‘factorisation constant’’; we deviate from this terminology since for us C_r will depend on n (cf. Corollary 19 and Lemma 21), and referring to it as a constant could cause confusion.

► **Lemma 17** (See, e.g., [8, Lemma 2.6 & Fact 3.5(4)] or [6, Lemma 3.2.6 & Fact 3.4.2]). *Let $q \geq 2$, $r \geq 1$ be integers and V be a set of size $n \geq r + 1$. Let μ be a distribution supported over $\Omega \subseteq [q]^V$ that satisfies the r -uniform-block factorisation of entropy with multiplier C_r . Then, for any $\varepsilon > 0$, the mixing time of the r -uniform-block Glauber dynamics on μ satisfies*

$$T_{\text{mix}}(\varepsilon) \leq \left\lceil C_r \frac{n}{r} \left(\log \log \frac{1}{\mu_{\min}} + \log \frac{1}{2\varepsilon^2} \right) \right\rceil, \text{ where } \mu_{\min} = \min_{\sigma \in \Omega} \mu(\sigma).$$

We remark that to deduce the lemma from [8] or [6], which refer to the so-called ‘‘entropy decay constant κ ’’, one needs to use the equality $C_r \kappa = r/n$ from [8, Lemma 2.6] or [6, Lemma 3.2.6].

From spectral independence to r -uniform-block factorisation multipliers

The following theorem is shown in [8]; while the version that we state here cannot be found verbatim in [8], we explain in the appendix how to combine the results therein to obtain it.

► **Theorem 18** ([8]). *Let $q \geq 2$ be an integer and V be a set of size $n \geq 2$. Let μ be a distribution supported over $\Omega \subseteq [q]^V$ that is η -spectrally independent and b -marginally bounded for $\eta, b > 0$.*

Then, for all integers $r = 1, \dots, n$, the distribution μ satisfies the r -uniform-block factorisation of entropy with multiplier $C_r = \frac{r}{n} \frac{\sum_{k=0}^{n-1} \Gamma_k}{\sum_{k=n-r}^{n-1} \Gamma_k}$, where $\Gamma_k = \prod_{j=0}^{k-1} \alpha_j$ for $k \in [n]^5$ and $\alpha_k = \max \left\{ 0, 1 - \frac{4\eta}{b^2(n-1-k)} \right\}$ for $k \in [n-1]$.

4.2 Entropy factorisation for bounded-degree vertices in the hard-core model

The first step of the analysis of Glauber dynamics for the hard-core model on the set of small-degree vertices will be to employ spectral independence results of Section 3.2 to conclude fast mixing for the r -uniform block Glauber dynamics for $r = \theta|U|$ for any arbitrarily small constant θ . This step will follow by applying the recent technology of entropy factorisation described above.

The second step is the more challenging for us. Here we need to conclude fast mixing for $r = 1$, and in particular prove that $C_1/C_r = n^{o(1)}$. This is done roughly by studying the connected components of G that arise when resampling an r -subset of the low-degree vertices; the factorisation multiplier of these components controls the ratio C_1/C_r . While this resembles the approach of [8], there is a key difference here, in that high-degree vertices are not resampled. This can not only cause potentially large components, but also imposes a deterministic lower bound on components sizes (since a component consisting of high-degree vertices will be deterministically present in the percolated graph consisting of the r -subset of low-degree vertices and all of the high-degree vertices). This lower bound on the component sizes is actually more significant than it might initially seem since the relatively straightforward bound of $\Omega(\log n)$ would unfortunately give a relatively large factorisation multiplier of $n^{\Omega(1)}$ (through Lemma 16). Instead, we need to show that components have size $o(\log n)$, which in turn requires more delicate estimates for the distribution of high-degree vertices in connected sets (see Lemma 20 below).

⁵ We note that for $k = 0$, the product defining Γ_k is empty and therefore $\Gamma_0 = 1$.

We start with the following corollary of Lemma 12, which converts a spectral independence bound into a bound on the factorisation multiplier for the r -uniform-block Glauber dynamics when r scales linearly with small-degree vertices. This is analogous to [8, Lemma 2.4], where they obtain a $2^{O(\eta/b^2)}$ bound on C_r when $r = \Theta(n)$ via Theorem 18 (where η is the spectral independence bound and b is the bound on the marginals). By restricting to small-degree vertices, we obtain that b is a constant, which combined with the bound $\eta = o(\log n)$ from Lemma 12 gives the bound $C_r = n^{o(1)}$, as detailed below. The proof of the corollary is given for completeness in Section D of the full version.

► **Corollary 19.** *Let $d \geq 1$ and $\lambda > 0$ be constants such that $\lambda < \lambda_c(d)$. Then, for any constants $D, \theta > 0$, whp over the choice of $G \sim G(n, d/n)$, the marginal hard-core distribution $\mu_{G, \lambda, U}$, where U is the set of vertices in G with degree $\leq D$, satisfies for any integer $r \in [\theta|U|, |U|]$ the r -uniform block factorisation of entropy with multiplier $C_r \leq n^\theta$.*

Note that the reason that we are able to use the same θ in the bounds for r and C_r is that the bound on C_r is loose (we can obtain a sharper result since we have a bound on the spectral independence of $\varepsilon \log n$ for any $\varepsilon > 0$).

We will now refine the bound of Corollary 19 down to $r = 1$ by exploiting the fact that high-degree vertices are sparsely scattered. In particular, we will need the following lemma which is a refinement of Lemma 4. For a graph $G = (V, E)$, we say that a set $S \subseteq V$ is connected if the induced subgraph $G[S]$ is connected.

► **Lemma 20.** *Let $d > 0$ be an arbitrary real. There exists an $L > 0$ such that for any $\delta \in (0, 1)$, the following holds whp over the choice of $G = (V, E) \sim G(n, d/n)$. For $\Delta = 1/(\delta \log \frac{1}{\delta})$, for all integers $k \geq \delta \log n$ and any $v \in V$, there are $\leq (2e)^{\Delta L k}$ connected sets $S \subseteq V$ containing v with $|S| = k$. Moreover, every such set contains $\geq k/2$ vertices with degree less than $L\Delta$.*

The proof of Lemma 20 is given in Section C.3 of the full version. We are now ready to show the following.

► **Lemma 21.** *Let $d \geq 1$ and $\lambda > 0$ be constants such that $\lambda < \lambda_c(d)$. For any $\theta > 0$, there is a constant $D > 0$ such that whp over the choice of $G \sim G(n, d/n)$, the marginal hard-core distribution $\mu_{G, \lambda, U}$, where U is the set of vertices in G with degree $\leq D$, satisfies the 1-uniform-block factorisation of entropy with multiplier $C_1 \leq n^\theta$.*

Proof. Let $L > 0$ be as in Lemma 20, and consider an arbitrarily small constant $\theta > 0$. Let $\delta \in (0, 1)$ be a sufficiently small constant so that for $D = L/(\delta \log \frac{1}{\delta})$ and $b = \frac{\lambda}{\lambda + (1+\lambda)^D}$ it holds that $\frac{1}{b^{2\delta}} < e^{\theta/4}$; such a constant exists since $b^{2\delta} \rightarrow 1$ as $\delta \downarrow 0$. Let $\Delta = 1/(\delta \log \frac{1}{\delta})$ and $\zeta > 0$ be a small constant so that $2(2e)^{L\Delta} (2\zeta)^{1/2} \leq b^2/2$.

Let U be the vertices in G with degree $\leq D$, and let $r = \lfloor \zeta|U| \rfloor + 1$. Let $\mu = \mu_{G, \lambda, U}$. By Corollary 19, we have that whp over the choice of G , there is $C_r \leq n^{\theta/3}$ such that for every $f : \Omega \rightarrow \mathbb{R}_{\geq 0}$ it holds that

$$\frac{r}{|U|} \text{Ent}_\mu(f) \leq C_r \frac{1}{\binom{|U|}{r}} \sum_{S \in \binom{U}{r}} \text{Ent}_\mu^S(f). \quad (6)$$

For $S \subseteq U$, let $\mathcal{C}'(S)$ denote the collection of the connected components of the graph $G[S \cup (V \setminus U)]$, viewed as vertex sets, and let $\mathcal{C}(S) = \bigcup_{R \in \mathcal{C}'(S)} \{R \cap U\}$ be the restriction of these components to the set U . Note that, for $S \subseteq U$ and $\tau \in \Omega_{U \setminus S}$, μ_τ factorises over the components of $G[S \cup (V \setminus U)]$ and in particular, applying Fact 14, we have that

$$\text{Ent}_\mu^S(f) = \mathbf{E}_{\tau \sim \mu_{U \setminus S}} [\text{Ent}_{\mu_\tau}(f)] \leq \mathbf{E}_{\tau \sim \mu_{U \setminus S}} \left[\sum_{R \in \mathcal{C}(S)} \text{Ent}_{\mu_\tau}^R(f) \right].$$

Using the bound in Lemma 16, we further obtain that

$$\begin{aligned} \text{Ent}_\mu^S(f) &\leq \mathbf{E}_{\tau \sim \mu_{U \setminus S}} \left[\sum_{R \in \mathcal{C}(S)} \frac{2|R|^2 \log(1/b)}{b^{2|R|+2}} \sum_{u \in R} \text{Ent}_{\mu_\tau}^u(f) \right] \\ &= \sum_{R \in \mathcal{C}(S)} \sum_{u \in R} \frac{2|R|^2 \log(1/b)}{b^{2|R|+2}} \text{Ent}_\mu^u(f), \end{aligned}$$

where the last equality follows by linearity of expectation and the fact that $\mathbf{E}_{\tau \sim \mu_{U \setminus S}}[\text{Ent}_{\mu_\tau}^u(f)] = \text{Ent}_\mu^u(f)$. Plugging this bound into (6), we obtain that

$$\text{Ent}_\mu(f) \leq \frac{2C_r \log(1/b)}{b^{2 \binom{|U|-1}{r-1}}} \sum_{S \in \binom{U}{r}} \sum_{R \in \mathcal{C}(S)} \sum_{u \in U} \frac{|R|^2}{b^{2|R|}} \text{Ent}_\mu^u(f).$$

which yields that

$$\text{Ent}_\mu(f) \leq \frac{2C_r \log(1/b)}{b^2} \sum_{u \in U} \text{Ent}_\mu^u(f) \sum_{k=1}^n \frac{k^2}{b^{2k}} \Pr[\mathcal{C}_u(S) = k], \quad (7)$$

where $\Pr[\mathcal{C}_u(S) = k]$ denotes the probability that u belongs to a set of size k in the set $\mathcal{C}(S)$, when we pick S uniformly at random from $\{S \in \binom{U}{r} \mid u \in S\}$. Define analogously $\Pr[\mathcal{C}'_u(S) = k]$ to be the probability that u belongs to a connected component of size k in the set $\mathcal{C}'(S)$. By Lemma 20, whp over $G \sim G(n, d/n)$, for all vertices u and any integer $t \geq \delta \log n$, there are at most $(2e)^{L\Delta t}$ connected sets of size t containing a given vertex u , and each of them contains at least $t/2$ vertices from U . In particular, for any integer $k \geq \delta \log n$, it holds that $\Pr[\mathcal{C}_u(S) = k] \leq \Pr[k \leq \mathcal{C}'_u(S) \leq 2k]$. For all $k \leq 2|U|$, the probability that a specific subset of $k/2$ vertices of U is present in $G[S \cup (V \setminus U)]$ is at most $\frac{\binom{|U| - \lceil k/2 \rceil}{r - \lceil k/2 \rceil}}{\binom{|U| - 1}{r-1}} \leq \left(\frac{r}{|U|}\right)^{k/2} \leq (2\zeta)^{k/2}$. Therefore, for all $k \geq \delta \log n$, by a union bound over the connected sets of size k , we have

$$\Pr[\mathcal{C}'_u(S) = k] \leq (2e)^{L\Delta k} 2^k (2\zeta)^{k/2} = (2(2e)^{L\Delta} (2\zeta)^{1/2})^k \leq (b^2/2)^k,$$

where in the first inequality the first factor is the number of size- k connected sets T of G containing u , the second factor is an upper bound on the number of size $k/2$ subsets W of U that might be included in T and the final factor is the probability that W is included in S . The last inequality is by the choice of ζ . It follows that

$$\Pr[\mathcal{C}_u(S) = k] \leq \Pr[k \leq \mathcal{C}'_u(S) \leq 2k] \leq 2k(b^2/2)^k.$$

From this bound and the inequality $1/b^{2\delta} < e^{\theta/4}$ by the choice of δ , we can split and bound the rightmost sum in (7) by

$$\sum_{k=1}^n \frac{k^2}{b^{2k}} \Pr[\mathcal{C}_u(S) = k] \leq \frac{(\delta \log n)^2}{b^{2\delta \log n}} + \sum_{k \geq \delta \log n} \frac{2k^3}{2^k} \leq n^{\theta/3},$$

where the last inequality holds for all sufficiently large n . In turn, plugging this into (7), we obtain that μ satisfies the 1-uniform block factorisation of entropy with multiplier $C_1 = \frac{2C_r \log(1/b)}{b^2} n^{\theta/3} \leq n^\theta$ for all sufficiently large n (since b is a constant and $C_r \leq n^{\theta/3}$), as needed. \blacktriangleleft

Lemma 5 now follows easily by combining Lemmas 17 and 21. This was the last ingredient needed in the proof of Theorem 1.

References

- 1 V. L. Alev and L. C. Lau. Improved analysis of higher order random walks and applications. In *Proceedings of the 52nd Annual ACM SIGACT Symposium on Theory of Computing*, STOC 2020, pages 1198–1211, 2020.
- 2 N. Anari, K. Liu, and S. Oveis Gharan. Spectral independence in high-dimensional expanders and applications to the hardcore model. In *Proceedings of the 61st Annual Symposium on Foundations of Computer Science (FOCS)*, pages 1319–1330, 2020.
- 3 I. Bezáková, A. Galanis, L. A. Goldberg, and D. Štefankovič. The complexity of approximating the matching polynomial in the complex plane. *ACM Trans. Comput. Theory*, 13(2), 2021.
- 4 A. Blanca and R. Gheissari. Sampling from Potts on random graphs of unbounded degree via random-cluster dynamics. *CoRR*, abs/2107.10246, 2021.
- 5 X. Chen, W. Feng, Y. Yin, and X. Zhang. Rapid mixing of Glauber dynamics via spectral independence for all degrees. *CoRR*, abs/2105.15005, 2021. [arXiv:2105.15005](https://arxiv.org/abs/2105.15005).
- 6 Z. Chen. *Optimal Mixing of Markov Chains for Spin Systems via Spectral Independence*. PhD thesis, Georgia Institute of Technology, 2021.
- 7 Z. Chen, K. Liu, and E. Vigoda. Rapid mixing of Glauber dynamics up to uniqueness via contraction. In *Proceedings of the 61st Annual Symposium on Foundations of Computer Science (FOCS)*, pages 1307–1318, 2020. Full version at [arXiv:2004.09083](https://arxiv.org/abs/2004.09083).
- 8 Z. Chen, K. Liu, and E. Vigoda. Optimal mixing of Glauber dynamics: Entropy factorization via high-dimensional expansion. In *Proceedings of the 53rd Annual ACM SIGACT Symposium on Theory of Computing*, STOC 2021, pages 1537–1550, 2021. Full version at [arXiv:2011.02075](https://arxiv.org/abs/2011.02075).
- 9 I. Dinur and T. Kaufman. High dimensional expanders imply agreement expanders. In *58th Annual IEEE Symposium on Foundations of Computer Science – FOCS 2017*, pages 974–985. IEEE Computer Soc., Los Alamitos, CA, 2017. doi:10.1109/FOCS.2017.94.
- 10 C. Efthymiou. MCMC sampling colourings and independent sets of $G(n, d/n)$ near uniqueness threshold. In *Proceedings of the Twenty-Fifth Annual ACM-SIAM Symposium on Discrete Algorithms, SODA 2014*, pages 305–316, 2014.
- 11 C. Efthymiou, T. P. Hayes, D. Štefankovič, and E. Vigoda. Sampling random colorings of sparse random graphs. In *Proceedings of the Twenty-Ninth Annual ACM-SIAM Symposium on Discrete Algorithms, SODA 2018*, pages 1759–1771, 2018.
- 12 A. Galanis, Q. Ge, D. Štefankovič, E. Vigoda, and L. Yang. Improved inapproximability results for counting independent sets in the hard-core model. *Random Struct. Algorithms*, 45(1):78–110, 2014.
- 13 A. Galanis, D. Štefankovič, and Eric Vigoda. Inapproximability of the partition function for the antiferromagnetic Ising and hard-core models. *Comb. Probab. Comput.*, 25(4):500–559, 2016.
- 14 V. Jain, H. T. Pham, and T. D. Vuong. Spectral independence, coupling with the stationary distribution, and the spectral gap of the Glauber dynamics. *CoRR*, abs/2105.01201, 2021. [arXiv:2105.01201](https://arxiv.org/abs/2105.01201).
- 15 M. Jerrum. *Counting, sampling and integrating: algorithms and complexity*. Springer Science & Business Media, 2003.
- 16 M. Jerrum and A.r Sinclair. Approximating the permanent. *SIAM Journal on Computing*, 18(6):1149–1178, 1989.
- 17 T. Kaufman and D. Mass. High dimensional random walks and colorful expansion. In *8th Innovations in Theoretical Computer Science Conference*, volume 67 of *LIPICs. Leibniz Int. Proc. Inform.*, pages Art. No. 4, 27. Schloss Dagstuhl. Leibniz-Zent. Inform., Wadern, 2017.
- 18 T. Kaufman and I. Oppenheim. High order random walks: beyond spectral gap. *Combinatorica*, 40(2):245–281, 2020. doi:10.1007/s00493-019-3847-0.
- 19 L.Li, P. Lu, and Y. Yin. Correlation decay up to uniqueness in spin dystems. In *Proceedings of the 24th Annual ACM-SIAM Symposium on Discrete Algorithms, SODA 2013*, pages 67–84, 2013.

- 20 E. Mossel and A. Sly. Exact thresholds for Ising–Gibbs samplers on general graphs. *The Annals of Probability*, 41(1):294–328, 2013. doi:10.1214/11-AOP737.
- 21 I. Oppenheim. Local spectral expansion approach to high dimensional expanders Part I: Descent of spectral gaps. *Discrete Comput. Geom.*, 59(2):293–330, 2018. doi:10.1007/s00454-017-9948-x.
- 22 V. Patel and G. Regts. Deterministic polynomial-time approximation algorithms for partition functions and graph polynomials. *SIAM J. Comput.*, 46(6):1893–1919, 2017.
- 23 H. Peters and G. Regts. On a conjecture of Sokal concerning roots of the independence polynomial. *Michigan Mathematical Journal*, 68(1):33–55, 2019.
- 24 A. Sinclair, P. Srivastava, D. Štefankovič, and Y. Yin. Spatial mixing and the connective constant: optimal bounds. *Probability Theory and Related Fields*, 168(1):153–197, 2017.
- 25 A. Sinclair, P. Srivastava, and M. Thurley. Approximation algorithms for two-state anti-ferromagnetic spin systems on bounded degree graphs. In *Proceedings of the Twenty-Third Annual ACM-SIAM Symposium on Discrete Algorithms, SODA 2012*, pages 941–953, 2012.
- 26 A. Sinclair, P. Srivastava, and Y. Yin. Spatial mixing and approximation algorithms for graphs with bounded connective constant. In *54th Annual IEEE Symposium on Foundations of Computer Science, FOCS 2013*, pages 300–309. IEEE Computer Society, 2013.
- 27 A. Sly. Computational transition at the uniqueness threshold. In *Proceedings of the 51th Annual IEEE Symposium on Foundations of Computer Science, FOCS '10*, pages 287–296, 2010.
- 28 A. Sly and N. Sun. The computational hardness of counting in two-spin models on d -regular graphs. In *53rd Annual IEEE Symposium on Foundations of Computer Science, FOCS 2012*, pages 361–369, 2012.
- 29 D. Weitz. Counting independent sets up to the tree threshold. In *Proceedings of the 38th Annual ACM Symposium on Theory of Computing, STOC '06*, pages 140–149, 2006.