A simple condition implying rapid mixing of single-site dynamics on spin systems

Thomas P. Hayes*
University of California Berkeley
Computer Science Division
hayest@cs.berkeley.edu

Abstract

Spin systems *are a general way to describe local interactions between nodes in a graph. In statistical mechanics, spin systems are often used as a model for physical systems. In computer science, they comprise an important class of families of combinatorial objects, for which approximate counting and sampling algorithms remain an elusive goal.*

The Dobrushin condition *states that every row sum of the "influence matrix" for a spin system is less than* $1 - \varepsilon$*, where* $\varepsilon > 0$*. This criterion implies rapid convergence (*$O(n \log n)$ *mixing time) of the single-site (Glauber) dynamics for a spin system, as well as uniqueness of the Gibbs measure. The dual criterion that every column sum of the influence matrix is less than* $1 - \varepsilon$ *has also been shown to imply the same conclusions.*

We examine a common generalization of these conditions, namely that the maximum eigenvalue of the influence matrix is less than $1 - \varepsilon$*. Our main result is that this criterion implies* $O(n \log n)$ *mixing time for the Glauber dynamics.*

As applications, we consider the Ising model, hard-core lattice gas model, and graph colorings, relating the mixing time of the Glauber dynamics to the maximum eigenvalue for the adjacency matrix of the graph. For the special case of planar graphs, this leads to improved bounds on mixing time with quite simple proofs.

1 Introduction

Let $G = (V, E)$ be a finite graph and Q a finite set of *spins*. In many applications of both theoretical and practical interest, one is interested in sampling random assignments of spins to vertices according to some target distribution π whose dependencies are "local" in terms of the graph G. By "local," we mean that for any vertex v and assignment of spins to $V \setminus \{v\}$, the marginal distribution of the spin assigned to v under π is a function only of the spins at neighbors of v. Following the terminology of *spin systems* from statistical physics, we will refer to elements of V as *sites*, elements of Q as *spins*, and elements of Q^V as *configurations*. Letting $n = |V|$, our goal is to sample configurations from π (or a close approximation to π) in time $\mathrm{poly}(n)$, that is, polylogarithmic in the size of the configuration space.

EXAMPLE: THE (FERROMAGNETIC) ISING MODEL. Here $Q = \{-1, 1\}$ is the set of spins. The target (Gibbs) distribution π assigns higher weights to configurations where adjacent spins are equal. More precisely, for all $\sigma \in Q^V$,

$$\pi(\sigma) \propto \exp\left(-\beta \sum_{(i,j)\in E} \sigma(i)\sigma(j)\right)$$

where $0 \leq \beta \leq \infty$ is the *inverse temperature*. At lower temperatures (larger β), spins at a site are more strongly influenced by the spins at their neighbors.

EXAMPLE: THE HARD CORE MODEL (INDEPENDENT SETS). Here again there are only two spins, $Q = \{0, 1\}$. We say a site is *occupied* if its spin is 1, and *unoccupied* if 0. This time there is a *hard constraint*: configurations with adjacent sites occupied have probability zero. For valid configurations σ (with no two adjacent sites occupied),

$$\pi(\sigma) \propto \lambda^{\sum_{i\in V} \sigma(i)}$$

where $\lambda > 0$ is the *fugacity*.

EXAMPLE: GRAPH COLORINGS. Here $Q = \{1, \ldots, k\}$, where k is at least the chromatic number of G. A configuration $\sigma : V \to Q$ is considered a *proper coloring* if no two adjacent spins (colors) are equal. Our target distribution π is the uniform distribution over proper colorings.

The *heat bath single-site dynamics* (also known as the *Gibbs sampler* or *heat bath Glauber dynamics*) is undoubtedly among the simplest possible algorithms for sampling from a spin system. Starting from an arbitrary initial configuration, repeatedly choose a site v at random, each time

*Supported by an NSF Postdoctoral Fellowship and by NSF grant CCR-0121555.

replacing the spin at v with one sampled from π conditioned on the spins currently assigned to the neighbors of v. This is a Markov chain over the set of configurations, and under modest assumptions, its distribution is guaranteed to converge to π as the number of steps tends to infinity. Thus, to approximately sample from π, one only needs to simulate a sufficient number of steps of the above Markov chain, outputting the configuration obtained.

Clearly, the efficiency of the above Markov chain Monte Carlo algorithm depends on knowing the rate of convergence to equilibrium, or "mixing time" of the Glauber dynamics. This is the number of steps required until the distribution of the Markov chain is close (in total variation distance) to the stationary distribution π. In recent years, much progress has been made in deriving upper bounds on the mixing time. In many cases, it has been possible to show that the mixing time is as fast as $O(n \log n)$. Celebrated examples include the 2-dimensional Ising model above the critical temperature [14], graph colorings with sufficiently many colors [12], and the hard core model (independent sets) at sufficiently low densities [19]. An upper bound of $O(n \log n)$ arises naturally from various techniques for bounding the mixing time, such as coupling [2] and the log-Sobolev constant [4], and, at least among physicists, is generally taken as the criterion for "rapid mixing" (rather than the weaker notion of being bounded by a polynomial in n, which is more common in computer science). In fact, $O(n \log n)$ mixing is the fastest possible for any reversible single-site dynamics, at least in the case of bounded-degree graphs, as shown recently by Hayes and Sinclair [10].

For adjacent sites i and j, the *influence of j on i* is the maximum amount (in variation distance) that the marginal distribution of the spin at i can ever change due to changing the spin at j. (For non-adjacent sites, the influence is zero.) A more formal description is given in Definition 4. Due to the local nature of their definition, the influences are often easy to compute exactly. By the *influence matrix*, we will mean the $n \times n$ matrix whose (i, j) entry is the influence of i on j.

In 1968, Dobrushin [5], studying spin systems on countably infinitely many sites, considered the case when no site has combined influence on its neighbors greater than $1 - \varepsilon$, for some $\varepsilon > 0$, *i. e.,* every row sum of the influence matrix is at most $1 - \varepsilon$. Under this condition, he proved "uniqueness of the Gibbs measure," which roughly stated says that there is asymptotically no correlation between the spin at a site v and the spins at sites at distance d from v, as d tends to infinity.

As subsequent work would show [6, 8, 11, 1, 17], the same "Dobrushin condition" implies several other important properties. In particular, for graphs on n sites, the Dobrushin condition implies optimal mixing time for the single-site heat bath dynamics, namely $O(n \log n)$ steps. This bound on mixing time also applies to heat bath "block dynamics," which update the spins of multiple sites per step, as well as to single-site "systematic scan" dynamics, which update vertices in a fixed (non-random) order. (See Section 2 for definitions.)

For a discussion of some further important consequences of Dobrushin's condition (also called the "completely analytic" setting), the reader is referred to the excellent survey by Weitz [20], and also the seminal papers by Dobrushin [5], Dobrushin and Shlosman [7, 6, 8], and Stroock and Zegarlinski [17].

Various extensions to the Dobrushin condition are known that imply the same conclusions. For instance, Dobrushin and Shlosman [7, 6] showed that the same consequences follow when every *column* sum of the influence matrix is less than $1 - \varepsilon$. They also generalized the condition to variants of the influence matrix which arise from "weighting" the rows and columns by placing a path metric on the state space. As we shall see in Remark 9, this last idea is closely linked to the present work.

Our main contribution is:

Theorem 1. *When the operator norm (*i. e., *the maximum eigenvalue) of the influence matrix is less than $1 - \varepsilon$, the mixing time for the heat bath Glauber dynamics is $O((n \log n)/\varepsilon)$.*

The same applies to the mixing time of the "systematic scan" dynamics, where the vertex order is fixed rather than uniformly random.

An interesting feature of this result is that the condition that the operator norm of the influence matrix is less than $1 - \varepsilon$ is *not* sufficient to imply uniqueness of Gibbs measure. This fact is demonstrated by two of the three examples presented in the second part of the paper.

In the second part of the paper, we consider applications to three important spin systems: the Ising model, the hard core model (independent sets), and k-colorings of a finite graph. For these systems, among others, the influence matrix is bounded above by a small multiple of the adjacency matrix for the underlying interaction graph G. This observation allows us to extend the class of graphs for which rapid mixing is known to hold.

In particular, we show the mixing time for Glauber dynamics on k-colorings of *planar* graphs of maximum degree Δ is $O(n \log n)$, assuming $k \geq \Delta + O(\sqrt{\Delta})$. This is in fact the first polynomial upper bound on the mixing time for this class of graphs and number of colors. For Δ-regular trees, optimal mixing was already known assuming $k \geq \Delta + 2$ (see [15, Theorem 1.5]). However, polynomial-time mixing for planar graphs was only known under the same hypotheses as for general graphs, namely $k \geq 11\Delta/6$ (see [18]).

For the Ising model on planar graphs, we prove optimal mixing when the inverse temperature is $\beta = O(1/\sqrt{\Delta})$.

Previously, this was known for tree; see [15, Theorem 1.1.] for the exact threshold. For general graphs, the stronger assumption $\beta = O(1/\Delta)$ is required.

In the case of the hard core model (independent sets) with fugacity λ, we can prove optimal mixing for the Glauber dynamics on planar graphs of maximum degree Δ, assuming $\lambda = O(1/\sqrt{\Delta})$. Previously, this was known for trees (see [15, Theorem 1.2]), but the stronger assumption $\lambda = O(1/\Delta)$ was required for more general graphs.

2 Main Results

Let Q be a set of "spins" and let $G = (V, E)$ be a finite graph. Let $\Omega \subseteq Q^V$ be the set of "feasible spin configurations". Let π be a probability distribution on Ω. We will assume throughout that $\varepsilon, \delta > 0$.

Definition 2. The "heat bath Glauber dynamics" is a Markov chain on state space Ω, defined as follows. Suppose the current state is $X_t \in \Omega$. Choose a site $i(t+1)$ uniformly at random from $[n]$. Sample X_{t+1} from the conditional distribution of π, given that $X_{t+1}(j) = X_t(j)$ for all $j \neq i(t+1)$. Thus the spin of $i(t+1)$ is rechosen randomly, and all other spins remain the same. This defines the next state X_{t+1}.

Assuming the state space is connected under the above transitions, the distribution of the Glauber dynamics will converge to π in the limit as the number of steps tends to infinity. To measure the rate of this convergence, we will use the following standard notion of mixing time.

Definition 3. Let $(X_t), t \geq 0$ be the Glauber dynamics for target distribution π, starting from an initial configuration X_0. Let μ_t denote the distribution of X_t. Let $\delta > 0$. The *mixing time* $\tau(\delta)$ for the dynamics is defined by

$$\tau(\delta) := \max_{X_0} \min\{t : d_{\mathrm{TV}}(\mu_t, \pi) \leq \delta\},$$

where $d_{\mathrm{TV}}(\cdot, \cdot)$ denotes the total variation distance.

Definition 4. For $x \in \Omega$, and $i \in V$, let $\mu_i(x, \cdot)$ denote the marginal distribution of the spin of i, for configurations sampled from π conditional on agreeing with x at all other sites. For sites $i, j \in V$, the *influence of j on i* is defined as

$$\rho_{i,j} := \max_{(x,y) \in S_j} d_{\mathrm{TV}}(\mu_i(x, \cdot), \mu_i(y, \cdot)),$$

where S_j denotes the set of all pairs of feasible configurations $(x, y) \in \Omega^2$ such that x and y agree on all sites except j. Let $R = (\rho_{i,j})$ be the $|V| \times |V|$ *influence matrix*. We remark that this matrix has also been referred to as a matrix of "dependencies" or "interdependences" in the literature.

Remark 5. Since R is a square matrix with non-negative entries, its operator norm, defined as

$$\sup_{\mathbf{v} \neq \mathbf{0}} \frac{\|R\mathbf{v}\|}{\|\mathbf{v}\|},$$

where $\|\cdot\|$ denotes the Euclidean norm, is also equal to its maximum eigenvalue in absolute value. Moreover, this *principal* eigenvalue is non-negative real[1] and has at least one corresponding eigenvector whose entries are all non-negative reals.

The *Dobrushin condition* is that every row sum of R is less than $1 - \varepsilon < 1$. The *Dobrushin-Shlosman condition* is that every column sum of R is less than $1 - \varepsilon < 1$. Either of these conditions implies uniqueness of the Gibbs measure, and also $O(n \log n)$ mixing time for the heat bath Glauber dynamics. We will now show that the rapid mixing result also follows the weaker hypothesis that the principal eigenvalue of R is less than 1.

Theorem 6. *Let $1-\varepsilon$ be the operator norm of R. Then after T steps of heat bath Glauber dynamics, the total variation distance from π satisfies*

$$d_{\mathrm{TV}}(P^T, \pi) \leq \left(1 - \frac{\varepsilon}{n}\right)^T n.$$

Proof. Fix $x, y \in \Omega$. For $t \geq 0$, define $(X_t, Y_t) \in \Omega^2$ by iterating a maximal one-step coupling of the Glauber dynamics, starting from initial condition $X_0 = x, Y_0 = y$. For $t \geq 0$, define the vector $\mathbf{p}_t$ by

$$p_t^j := \mathbf{Pr}\left(X_t(j) \neq Y_t(j)\right).$$

Note that since the Glauber dynamics chooses a random site in each step, this implies

$$\mathbf{p}_{t+1}^j = \frac{n-1}{n}\mathbf{p}_t^j + \frac{1}{n}\mathbf{Pr}\left(X_{t+1}(j) \neq Y_{t+1}(j) \mid j \text{ chosen at time } t\right).$$

Now by the definition of a maximal one-step coupling, and linearity of expectation, it follows that

$$\mathbf{p}_{t+1} \leq A\mathbf{p}_t,$$

where

$$A := \frac{n-1}{n} I + \frac{1}{n} R,$$

where I is the $n \times n$ identity matrix. Note that A has eigenvalues $\frac{n-1}{n} + \frac{\alpha}{n}$, where α runs through the eigenvalues of R. Hence the principal eigenvalue of A is at most $1 - \varepsilon/n$.

[1] in degenerate cases, there may be other eigenvalues with the same absolute value

By induction, we obtain the component-wise inequality

$$\mathbf{p}_T \le A^T \mathbf{p}_0. \tag{1}$$

By a union bound, the ℓ_1 norm of $\mathbf{p}_T$ gives an upper bound on the probability of non-coalescence, thus

$$\begin{aligned}
\mathbf{Pr}\left(X_T \neq Y_T\right) &\le \|\mathbf{p}_T\|_1 \\
&\le \sqrt{n}\|\mathbf{p}_T\|_2 && \text{by Cauchy-Schwarz} \\
&\le \sqrt{n}\|A^T\mathbf{p}_0\|_2 && \text{by (1)} \\
&\le \sqrt{n}\|A\|^T\|\mathbf{p}_0\|_2 \\
&\le n\left(1-\frac{\varepsilon}{n}\right)^T. \quad \square
\end{aligned}$$

Theorem 6 immediately implies the following bound on the mixing time, which in turn implies Theorem 1 from the Introduction.

Corollary 7. *Let $1-\varepsilon$ be the largest eigenvalue of R. Then the mixing time satisfies*

$$\tau(\delta) \le \frac{n}{\varepsilon}\ln\frac{n}{\delta}.$$

In our definition of Glauber dynamics, we assumed that the site order used for updating spins was uniformly random. By contrast, in *systematic scan*, the site order $v_1, v_2, \ldots,$ is deterministic and fixed in advance. A special case of interest is when a fixed permutation of the sites is repeated *ad infinitum*). More generally, we will assume that $I_1, I_2, \ldots,$ is a partition of the positive integers into "rounds," such that for every $t \in I_j, t' \in I_{j+1}$ we have $t < t'$, and for every $j \ge 1$, $v \in V$, there exists $t \in I_j$ such that $v_t = v$. (In the special case when the site order is a repeated permutation, the rounds are all of length n.)

We next show that our generalized Dobrushin condition implies rapid mixing for systematic scan. The approach is similar to earlier proofs for the usual Dobrushin condition (see Dyer, Greenhill, Jerrum [9]).

Theorem 8. *Let $1-\varepsilon$ be the largest eigenvalue of R. Then after T rounds of systematic scan, the total variation distance from π satisfies*

$$d_{\mathrm{TV}}(Q,\pi) \le \left(1-\frac{\varepsilon}{2}\right)^{T+1}\frac{2n}{\varepsilon}.$$

Proof. As before, let $(X_t, Y_t) \in \Omega^2$, $t \ge 0$ be a coupling of two copies of the dynamics starting from arbitrary initial conditions, and let $\mathbf{p}_t$ be defined by

$$p_t^j := \mathbf{Pr}\left(X_t(j) \neq Y_t(j)\right).$$

For each site i, define M_i to be the identity matrix with the i'th row replaced by the i'th row of R. Then, if the site order for the systematic scan is $i(1), \ldots, i(t), \ldots,$ then the disagreement probability vector satisfies the component-wise inequality

$$\mathbf{p}_t \le M_{i(t)}\mathbf{p}_{t-1}.$$

Let S be the following slight perturbation of R:

$$S := R + \frac{\varepsilon}{2n}J,$$

where J is the $n \times n$ all-ones matrix.

Let $\mathbf{x} = (x^1, \ldots, x^n)$ denote the principal eigenvector for S, and let $\alpha \le 1-\varepsilon/2$ be the principal eigenvalue. Observe that the components of $\mathbf{x}$ are all strictly positive; moreover, for every component j,

$$\alpha x^j = (S\mathbf{x})^j \ge \frac{\varepsilon}{2n}(J\mathbf{x})^j = \frac{\varepsilon}{2n}\|\mathbf{x}\|_1. \tag{2}$$

For each time t, we have

$$\begin{aligned}
p_t^{i(t)} &\le \left(M_{i(t)}\mathbf{p}_{t-1}\right)^{i(t)} \\
&= \left(R\mathbf{p}_{t-1}\right)^{i(t)} \\
&< \left(S\mathbf{p}_{t-1}\right)^{i(t)} \\
&\le \left(S\mathbf{x}\right)^{i(t)} \max_j \frac{p_{t-1}^j}{x^j} \\
&\le \alpha x^{i(t)} \max_j \frac{p_{t-1}^j}{x^j}.
\end{aligned}$$

Moreover, for each time t and each $k \neq i(t)$, we have

$$p_t^{i(t)} = p_{t-1}^{i(t)}.$$

From these facts, an easy induction shows that, for t in round I_r, for every site i,

$$\frac{p_t^i}{x^i} \le \begin{cases} \alpha \max_j \frac{p_s^j}{x^j} & \text{if } i \text{ was updated in round } r \text{ by time } t. \\ \max_j \frac{p_s^j}{x^j} & \text{if not.} \end{cases} \tag{3}$$

Here the subscript s denotes the final time of round $r-1$. Thus $\max_j p_t^j/x^j$ decreases by at least a factor of α in each round.

For every site j and time t, define

$$q_t^j := p_t^j/x^j.$$

Note that this definition implies the component-wise inequality

$$\mathbf{p}_t \le (\max_j q_t^j)\mathbf{x}.$$

Applying induction to (3), we know that after T rounds of systematic scan,

$$\max_j q_t^j \le \alpha^T \max_j q_0^j \le \frac{\alpha^T}{x_{\min}}$$

and hence

$$\begin{aligned}\mathbf{Pr}\left(X_t \neq Y_t\right) &\leq \|\mathbf{p}_t\|_1 && \text{by a union bound}\\ &\leq \frac{\alpha^T}{x_{\min}}\,\|\mathbf{x}\|_1 \\ &\leq \alpha^{T+1}\,\frac{2n}{\varepsilon} && \text{by (2)}\end{aligned}$$

□

Remark 9. In the above proof, a key step was to rescale the vector $\mathbf{p}$ by dividing componentwise by $\mathbf{x}$. This is a fairly common theme among coupling proofs in the literature. For instance, in proofs based on path coupling, the key step is often to prove contraction in a scaled version of the Hamming metric on configurations. Finding the right such metric may amount to explicitly computing or approximating the principal eigenvector for the influence matrix. Note that to apply Theorems 6 or 8 one only needs a good bound on the principal eigenvalue, which is hopefully much easier to come by.

3 Applications

In many cases, the influence matrix can be easily bounded by a multiple of the adjacency matrix for G. For instance,

Observation 10. *For the hard core model (independent sets) with fugacity λ, the influence matrix R is related to the adjacency matrix A by*

$$R = \frac{\lambda}{1+\lambda}A.$$

Proof. Let σ be any configuration. There are only two possibilities for the marginal distribution of π at a site i, conditioned on the neighbors agreeing with σ. Either some neighbor is occupied under σ, in which case i is unoccupied with probability 1, or all neighbors are unoccupied, in which case i is occupied with probability $\lambda/(1+\lambda)$, and unoccupied otherwise. Since the variation distance between these marginal distributions is $\lambda/(1+\lambda)$, this is an upper bound on the entries of the influence matrix. This completes the proof, since the influence of i on j is zero except when i and j are neighbors. □

Observation 11. *For the Ising model at inverse temperature $\beta \geq 0$, the influence matrix R is related to the adjacency matrix A by*

$$R \leq \tanh(\beta)A \leq \beta A.$$

Proof. Let σ be any configuration. Suppose exactly r neighbors of site i have spin $+1$, and the remaining $d-r$ have spin -1. Then the marginal distribution of π at i conditioned on the neighbors agreeing with σ is $+1$ with probability

$$\frac{\exp(\beta(2r-d))}{\exp(\beta(2r-d)) + \exp(\beta(d-2r))}.$$

Now consider another distribution τ which agrees with σ except at a single neighbor j. Without loss of generality, it has exactly $r+1$ neighbors of i with spin $+1$. Now a little algebra shows that the variation distance equals

$$\frac{\exp(2\beta) - \exp(-2\beta)}{(y+1/y)(y\exp(2\beta) + 1/(y\exp(2\beta)))},$$

where $y = \exp(\beta(2r-d))$. This attains a maximum value of $\tanh(\beta)$ when d is odd and $r = (d-1)/2$. □

Observation 12. *For the uniform distribution on proper k-colorings of a graph with maximum degree Δ, the influence matrix R is related to the adjacency matrix A by*

$$R \leq \frac{1}{k-\Delta}A.$$

Proof. More precisely, for sites i, j, the influence of j on i satisfies

$$\rho_{i,j} = \frac{A_{i,j}}{k-d_i} \leq \frac{A_{i,j}}{k-\Delta},$$

where d_i denotes the degree of site i.

This value for the influence of j on i is realized by colorings which assign distinct colors to all d_i neighbors of i, and disagree only at j. We leave the details to the interested reader. □

This gives us the following consequences. Since the eigenvalues of the adjacency matrix of a graph G is a graph invariant, we will often simply refer to the "eigenvalues of G."

Proposition 13. *The heat bath Glauber dynamics for the hard core model with fugacity λ on a graph with principal eigenvalue ρ has mixing time*

$$\tau(\delta) \leq \frac{n}{\varepsilon}\log\frac{n}{\delta}$$

whenever $\lambda \leq \frac{1-\varepsilon}{\rho}$.

Proof. By Observation 10, we have

$$\|R\| = \frac{\lambda}{1+\lambda}\|A\| \leq \lambda\rho \leq 1-\varepsilon$$

where A is the adjacency matrix for G. The desired result follows by Corollary 7. □

Proposition 14. *The heat bath Glauber dynamics for the Ising model at inverse temperature $\beta \geq 0$, on a graph with principal eigenvalue ρ, has mixing time*

$$\tau(\delta) \leq \frac{n}{\varepsilon} \log \frac{n}{\delta}$$

whenever $\beta \leq \frac{1-\varepsilon}{\rho}$.

Proof. By Observation 11 and monotonicity of the operator norm as a function of the entries of a non-negative matrix, we have

$$\|R\| \leq \beta \|A\| = \beta\rho \leq 1 - \varepsilon$$

where A is the adjacency matrix for G. The desired result follows by Corollary 7. □

Proposition 15. *The heat bath Glauber dynamics on proper k-colorings of a graph with maximum degree Δ and principal eigenvalue ρ has mixing time*

$$\tau(\delta) \leq \frac{n}{\varepsilon} \log \frac{n}{\delta}$$

whenever $k \geq \Delta + \frac{\rho}{1-\varepsilon}$.

Proof. By Observation 12 and monotonicity of the operator norm as a function of the entries of a non-negative matrix, we have

$$\|R\| \leq \frac{1}{k - \Delta} \|A\| = \frac{1 - \varepsilon}{\rho} \rho$$

where A is the adjacency matrix for G. The desired result follows by Corollary 7. □

Now, the usual applications of the Dobrushin and Dobrushin-Shlosman conditions can be seen as consequences of the basic fact that the principal eigenvalue of a graph is at most the maximum degree of a vertex. However, in general, the principal eigenvalue can be much smaller. In particular, we note that for planar graphs (such as trees) the principal eigenvalue is $O(\sqrt{\Delta})$, where Δ is the maximum degree. This follows as a consequence of the following more general result.

Theorem 16. *Let G be an undirected graph of maximum degree Δ, which admits an orientation with maximum outdegree $d \leq \Delta/2$. Then the maximum eigenvalue of G is at most $2\sqrt{d(\Delta - d)}$.*

Proof. We will use the fact that the principal eigenvalue of any graph G equals the limit

$$\lim_{\ell \to \infty} (W_\ell)^{1/\ell},$$

where W_ℓ denotes the number of walks of length ℓ in G. (This in turn follows from the fact that W_ℓ equals the sum of the entries of A^ℓ, where A is an adjacency matrix for G.) In light of this fact, it will suffice to prove the following upper bound on the number of walks of length ℓ:

$$W_\ell \leq n 2^\ell d^{\ell/2} (\Delta - d)^{\ell/2} \tag{4}$$

Let $s = (v_0, \ldots, v_\ell)$ denote any walk of length ℓ. For each vertex v_i, label its neighbors with $\{1, \ldots, \Delta\}$ so that the out-neighbors from v_i all have labels $\leq d$.

Now, let $F(s)$ denote the set of steps i such that v_{i+1} has an index $\leq d$ among the neighbors of v_i. Clearly, for a given subset $F \subseteq [\ell]$, there are at most $n d^{|F|} (\Delta - d)^{\ell - |F|}$ walks s such that $F(s) = F$.

Summing over all $F \subset [\ell]$ such that $|F| \geq \ell/2$, there can be at most

$$n 2^{\ell - 1} d^{\ell/2} (\Delta - d)^{\ell/2}$$

walks s such that $|F(s)| \geq \ell/2$, assuming ℓ is odd. When ℓ is even, the same upper bound applies, if we only count half of the sequences s such that $|F(s)| = \ell/2$.

Now note that, if we let $R(s) = (v_\ell, \ldots, v_0)$ denote the reverse walk of $s = (v_0, \ldots, v_\ell)$, then $F(s) \cup F(R(s)) = [\ell]$. Hence at least one of $|F(s)|$, $|F(R(s))|$ is at least $\ell/2$, which establishes the desired upper bound (4). □

We remark that as an immediate corollary, an upper bound of $2\sqrt{d(\Delta - d)}$ holds for the principal eigenvalue of any graph, all of whose subgraphs contain at least one vertex of degree $\leq d$.

Corollary 17. *Let G be a graph of maximum degree Δ. If G is a tree (or forest) and $\Delta \geq 2$, then the maximum eigenvalue of G is at most $2\sqrt{\Delta - 1}$. If G is planar and $\Delta \geq 6$, then the maximum eigenvalue is at most $2\sqrt{3(\Delta - 3)}$.*

Proof. For a forest, fix a root for each connected component, and orient each tree towards its root. The first result follows by Theorem 16 applied with $d = 1$.

That every finite planar graph has average degree < 6 is a well-known consequence of Euler's formula $v - e + f - c = 1$, where v is the number of vertices, e the number of edges, f the number of faces, and c the number of connected components. It follows that at least one vertex has degree ≤ 5. This shows that we can apply Theorem 16 with $d = 5$ to obtain a weaker upper bound of $2\sqrt{5(\Delta - 5)}$ on the principal eigenvalue (assuming $\Delta \geq 10$). To obtain the claimed bound, a more clever orientation must be found, with $d = 3$. This can be done recursively and efficiently, e.g. using the algorithm of Schnyder for planar embeddings [16]. □

Combining Corollary 17 with Propositions 13 through 15 immediately yields the following rapid mixing results for planar graphs.

Corollary 18. *Let $\Delta \geq 6$, let G be a planar graph on n vertices of degree $\leq \Delta$. Then the heat bath Glauber dynamics for the hard core (independent sets) model with fugacity*

$\lambda \leq \frac{1-\varepsilon}{2\sqrt{3(\Delta-3)}}$ *has mixing time*

$$\tau(\delta) \leq \frac{n}{\varepsilon} \ln \frac{n}{\delta}.$$

Corollary 19. *Let $\Delta \geq 6$, let G be a planar graph on n vertices of degree $\leq \Delta$. Then the heat bath Glauber dynamics for the Ising model at inverse temperature $\beta \leq \frac{1-\varepsilon}{2\sqrt{3(\Delta-3)}}$ has mixing time*

$$\tau(\delta) \leq \frac{n}{\varepsilon} \ln \frac{n}{\delta}.$$

Corollary 20. *Let $\Delta \geq 6$, let G be a planar graph on n vertices of degree $\leq \Delta$. Then the heat bath Glauber dynamics on k-colorings of G where $k \geq \Delta + \frac{2\sqrt{3(\Delta-3)}}{1-\varepsilon}$ has mixing time*

$$\tau(\delta) \leq \frac{n}{\varepsilon} \ln \frac{n}{\delta}.$$

As far as I know, these are the first rapid mixing results to take advantage of the structure of planar graphs. More attention has been paid to the special case of trees, and may help to put our results into context. In 1999, Brightwell and Winkler [3] proved uniqueness of the Gibbs measure for k-colorings of trees, where $k \geq 1.62\Delta$. In 2002, Jonasson [13] proved that $k \geq \Delta + 2$ is the threshold for uniqueness of the Gibbs measure. In 2004, Martinelli, Sinclair and Weitz [15] proved rapid mixing for the heat bath Glauber dynamics on trees under arbitrary boundary conditions, for $k \geq \Delta + 3$. The methods used in these papers are rather sophisticated and involve lengthy calculations, in contrast to our present approach.

4 Further Discussion

An interesting feature of our result is that, unlike Dobrushin's condition, which implies uniqueness of the Gibbs measure as well as rapid mixing of the Glauber dynamics, our new condition only implies rapid mixing. For instance, in the case of the Ising model on a tree of maximum degree Δ, our condition holds and implies rapid mixing for $\beta = O(1/\sqrt{\Delta})$; on the other hand, it is known that for an infinite Δ-regular tree, the threshold for Gibbs uniqueness is $\Theta(1/\Delta)$. Thus, unlike Dobrushin's condition, our condition can be used to derive rapid mixing in some cases where the Gibbs measure is not unique.

A partial explanation comes from the observation that, while the infinite tree is Δ-regular, and so has operator norm Δ, any finite subtree has average degree less than 2. So, for the finite tree, the many vertices of small degree can be viewed as the cause both for the smaller operator norm and for the rapid mixing.

A somewhat more detailed explanation is that, when the graph is infinite, there are several different operator norms corresponding to different vector norms. For the class ℓ_2 of vectors whose squares of entries have a finite sum, under the Euclidean vector norm, the operator norm equals the supremum of the operator norms on finite subgraphs (e.g. $O(\sqrt{\Delta})$ for planar graphs), and hence can be used to deduce $O(n \log n)$ mixing. For the class ℓ_∞ of vectors with absolutely bounded entries, under the sup norm, the operator norm may be larger (for instance it is Δ for any Δ-regular graph, including trees), and can be used to deduce uniqueness of Gibbs measure. We will discuss this issue in more detail in the final version of this paper.

Acknowledgments

I am grateful to Noga Alon for pointing out the current statement of Theorem 16, and to Alistair Sinclair and Eric Vigoda for helpful comments. I would also like to thank the anonymous referees for several corrections.

References

[1] M. Aizenman and R. Holley. Rapid convergence to equilibrium of stochastic Ising models in the Dobrushin-Shlosman régime. In H. Kesten, editor, *Percolation Theory and Ergodic Theory of Infinite Particle Systems (Minneapolis, Minn., 1984)*, volume 8 of *IMS Volumes in Math. and Appl.*, pages 1–11. Springer, Berlin, 1987.

[2] D. Aldous. Random walks on finite groups and rapidly mixing Markov chains. In *Séminaire do Probabilités XVII*, volume 986 of *Lecture Notes in Mathematics*, pages 243–297. Springer, Berlin, 1983.

[3] G. R. Brightwell and P. Winkler. Random colorings of a cayley tree. In B. Bollobás, editor, *Contemporary Combinatorics*, volume 10 of *Bolyai Society Mathematical Studies*, pages 247–276. Janos Bolyai Math. Soc., Budapest, 2002.

[4] P. Diaconis and L. Saloff-Coste. Logarithmic Sobolev inequalities for finite Markov chains. *Annals of applied probability*, 6:695–750, 1996.

[5] R. L. Dobrushin. Prescribing a system of random variables by the help of conditional distributions. *Theory Prob. and its Applications*, 15:469–497, 1970.

[6] R. L. Dobrushin and S. B. Shlosman. Completely analytical Gibbs fields. In J. Fritz, A. Jaffe, and D. Szasz, editors, *Statistical mechanics and dynamical systems*, pages 371–403. Birkhauser, Boston, 1985.

[7] R. L. Dobrushin and S. B. Shlosman. Constructive criterion for the uniqueness of a Gibbs field. In J. Fritz, A. Jaffe, and D. Szasz, editors, *Statistical mechanics and dynamical systems*, pages 347–370. Birkhauser, Boston, 1985.

[8] R. L. Dobrushin and S. B. Shlosman. Completely analytical interactions: constructive description. *Journal of Statistical Physics*, 46:983–1014, 1987.

[9] M. Dyer, L. A. Goldberg, and M. Jerrum. Dobrushin conditions and systematic scan. *ECCC Report No. 75*, page 16pp, July 2005.

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

[11] R. Holley. Possible rates of convergence in finite range, attractive spin systems. In *Particle systems, random media and large deviations (Brunswick, Maine, 1984)*, volume 41 of *Contemporary Mathematics*, pages 215–234. Amer. Math. Soc., Providence, RI, 1985.

[12] M. R. Jerrum. A very simple algorithm for estimating the number of k-colorings of a low-degree graph. *Random Structures and Algorithms*, 7:157–165, 1995.

[13] J. Jonasson. Uniqueness of uniform random colorings of regular trees. *Statistics & Probability Letters*, 57:243–248, 2002.

[14] F. Martinelli and E. Olivieri. Approach to equilibrium of Glauber dynamics in the one phase region i. the attractive case. *Communications in Mathematical Physics*, 161:447–486, 1994.

[15] F. Martinelli, A. Sinclair, and D. Weitz. Fast mixing for independent sets, colorings and other model on trees. In *Proceedings of the 15th Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pages 449–458. ACM, 2004. Full version to appear in *Random Structures and Algorithms*, 2006.

[16] W. Schnyder. Embedding planar graphs on the grid. In *Proceedings of the 1st Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pages 138–148. ACM, 1990.

[17] D. W. Stroock and B. Zegarlinski. The logarithmic Sobolev inequality for discrete spin systems on a lattice. *Communications in Mathematical Physics*, 149:175–194, 1992.

[18] E. Vigoda. Improved bounds for sampling colorings. *Journal of Mathematical Physics*, 41:1555–1569, 2000.

[19] E. Vigoda. A note on the Glauber dynamics for sampling independent sets. *Electronic Journal of Combinatorics*, 8(1), 2001.

[20] D. Weitz. Combinatorial criteria for uniqueness of Gibbs measures. *Random Structures and Algorithms*, 27(4):445–475, 2005.