Rapid Mixing of Gibbs Sampling on Graphs that are Sparse on Average

Elchanan Mossel* Allan Sly †

Abstract

Gibbs sampling also known as Glauber dynamics is a popular technique for sampling high dimensional distributions defined on graphs. Of special interest is the behavior of Gibbs sampling on the Erdős-Rényi random graph $G(n, d/n)$, where each edge is chosen independently with probability d/n and d is fixed. While the average degree in $G(n, d/n)$ is $d(1 - o(1))$, it contains many nodes of degree of order $\log n / \log \log n$.

The existence of nodes of almost logarithmic degrees implies that for many natural distributions defined on $G(n, p)$ such as uniform coloring (with a constant number of colors) or the Ising model at any fixed inverse temperature β, the mixing time of Gibbs sampling is at least $n^{1+\Omega(1/\log \log n)}$. Recall that the Ising model with inverse temperature β defined on a graph $G = (V, E)$ is the distribution over $\{\pm\}^V$ given by $P(\sigma) = \frac{1}{Z}\exp(\beta \sum_{(v,u)\in E} \sigma(v)\sigma(u))$. High degree nodes pose a technical challenge in proving polynomial time mixing of the dynamics for many models including the Ising model and coloring. Almost all known sufficient conditions in terms of β or number of colors needed for rapid mixing of Gibbs samplers are stated in terms of the maximum degree of the underlying graph.

In this work we show that for every $d < \infty$ and the Ising model defined on $G(n, d/n)$, there exists a $\beta_d > 0$, such that for all $\beta < \beta_d$ with probability going to 1 as $n \to \infty$, the mixing time of the dynamics on $G(n, d/n)$ is polynomial in n. Our results are the first polynomial time mixing results proven for a natural model on $G(n, d/n)$ for $d > 1$ where the parameters of the model do not depend on n. They also provide a rare example where one can prove a polynomial time mixing of Gibbs sampler in a situation where the actual mixing time is slower than $n\mathrm{polylog}(n)$. Our proof exploits in novel ways the local treelike structure of Erdős-Rényi random graphs, comparison and block dynamics arguments and a recent result of Weitz.

Our results extend to much more general families of graphs which are sparse in some average sense and to much more general interactions. In particular, they apply to any graph for which every vertex v of the graph has a neighborhood $N(v)$ of radius $O(\log n)$ in which the induced sub-graph is a tree union at most $O(\log n)$ edges and where for each simple path in $N(v)$ the sum of the vertex degrees along the path is $O(\log n)$. Moreover, our result apply also in the case of arbitrary external fields and provide the first FPRAS for sampling the Ising distribution in this case. We finally present a non Markov Chain algorithm for sampling the distribution which is effective for a wider range of parameters. In particular, for $G(n, d/n)$ it applies for all external fields and $\beta < \beta_d$, where $d \tanh(\beta_d) = 1$ is the critical point for decay of correlation for the Ising model on $G(n, d/n)$.

Keywords: Erdős-Rényi Random Graphs, Gibbs Samplers, Glauber Dynamics, Mixing Time, Ising model.

*Email: mossel@stat.berkeley.edu. Departments of Statistics and Computer Science, U.C. Berkeley. Supported by an Alfred Sloan fellowship in Mathematics and by NSF grants DMS-0528488, DMS-0548249 (CAREER) by DOD ONR grant N0014-07-1-05-06.

†Email: sly@stat.berkeley.edu Dept. of Statistics, U.C. Berkeley.

1 Introduction

Efficient approximate sampling from Gibbs distributions is a central challenge of randomized algorithms. Examples include sampling from the uniform distribution over independent sets of a graph [23, 22, 4, 7], sampling from the uniform distribution of matchings in a graph [15], or sampling from the uniform distribution of colorings [12, 3, 5] of a graph. A natural family of approximate sampling techniques is given by Gibbs samplers, also known as Glauber dynamics. These are reversible Markov chains that have the desired distribution as their stationary distribution and where at each step the status of one vertex is updated. It is typically easy to establish that the chains will eventually converge to the desired distribution.

Studying the convergence rate of the dynamics is interesting from both the theoretical computer science the statistical physics perspectives. Approximate convergence in polynomial time, sometimes call *rapid mixing*, is essential in computer science applications. The convergence rate is also of natural interest in the physics where the dynamical properties of such distributions are extensively studied, see e.g. [17]. Much recent work has been devoted to determining sufficient and necessary conditions for rapid convergence of Gibbs samplers. A common feature to most of this work [23, 22, 4, 7, 12, 3, 2, 18] is that the conditions for convergence are stated in terms of the maximal degree of the underlying graph. In particular, these results do not allow for the analysis of the mixing rate of Gibbs samplers on the Erdős-Rényi random graph, which is sparse on average, but has rare denser sub-graphs. Recent work has been invested in showing how to relax statements so that they do not involve maximal degrees [5, 13], but the results are not strong enough to imply rapid mixing of Gibbs sampling for the Ising model on $G(n, d/n)$ for $d > 1$ and any $\beta > 0$ or for sampling uniform colorings from $G(n, d/n)$ for $d > 1$ and $1000d$ colors. The second challenge is presented as the major open problem of [5].

In this paper we give the first rapid convergence result of Gibbs samplers for the Ising model on Erdős-Rényi random graphs in terms of the average degree and β only. Our results hold for the Ising model allowing different interactions and arbitrary external fields. We note that there is an FPRAS that samples from the Ising model on any graph [14] as long as all the interactions are positive and the external field is the same for all vertices. However, these results do not provide a FPRAS in the case where different nodes have different external fields as we do here.

Our results are further extended to much more general families of graphs that are "tree-like" and "sparse on average". These are graph where every vertex has a radius $O(\log n)$ neighborhood which is a tree with at most $O(\log n)$ edges added and where for each simple path in the neighborhood, the sum of degrees along the path is $O(\log n)$. An important open problem [5] is to establish similar conditions for other models defined on graphs, such as the uniform distribution over colorings.

Below we define the Ising model and Gibbs samplers and state our main result. Some related work and a sketch of the proof are also given in the introduction. Section 2 gives a more detailed proof though we have not tried to optimize any of the parameters in proofs below. The complete proofs can be found in the full paper at `http://arxiv.org/abs/0704.3603.`

1.1 The Ising Model The Ising model is perhaps the simplest model defined on graphs. This model defines a distribution on labelings of the vertices of the graph by $+$ and $-$. The Ising model has various natural generalizations including the uniform distribution over colorings. The Ising model with varying parameters is of use in a variety of areas of machine learning, most notably in vision, see e.g. [9].

DEFINITION 1.1. *The Ising model on a (weighted) graph G with inverse temperature β is a distribution on configurations $\{\pm\}^V$ such that*

$$P(\sigma) = \frac{1}{Z(\beta)} \exp(\beta \sum_{\{v,u\} \in E} \sigma(v)\sigma(u)) \tag{1.1}$$

where $Z(\beta)$ is a normalizing constant.

More generally, we will be interested in Ising models defined by:

$$P(\sigma) = \frac{1}{Z(\beta)} \exp(\sum_{\{v,u\} \in E} \beta_{u,v}\sigma(v)\sigma(u) + \sum_v h_v \sigma(v)), \tag{1.2}$$

where h_v are arbitrary and where $\beta_{u,v} \geq 0$ for all u and v. In the more general case we will write $\beta = \max_{u,v} \beta_{u,v}$.

1.2 Gibbs Sampling The Gibbs sampler is a Markov chain on configurations where a configuration σ is updated by choosing a vertex v uniformly at random and assigning it a spin according to the Gibbs distribution conditional on the spins on $G - \{v\}$.

DEFINITION 1.2. *Given a graph $G = (V, E)$ and an inverse temperature β, Gibbs sampler is the discrete time Markov chain on $\{\pm\}^V$ where given the current configuration σ the next configuration σ' is obtained by choosing a vertex v in V uniformly at random and*

- *Letting $\sigma'(w) = \sigma(w)$ for all $w \neq v$.*
- *$\sigma'(v)$ is assigned the spin $+$ with probability*

$$\frac{1}{1 + \exp(-2h_v - 2\sum_{u:(v,u)\in E} \beta_{u,v}\sigma(u))}.$$

We will be interested in the time it takes the dynamics to get close to the distributions (1.1) and (1.2). The *mixing time* τ_{mix} of the chain is defined as the number of steps needed in order to guarantee that the chain, starting from an arbitrary state, is within total variation distance $1/2e$ from the stationary distribution. We will bound the mixing time by the relaxation time defined below.

It is well known that Gibbs sampling is a reversible Markov chain with stationary distribution P. Let $1 = \lambda_1 > \lambda_2 \geq \ldots \geq \lambda_m \geq -1$ denote the eigenvalues of the transition matrix of Gibbs sampling. The *spectral gap* is denoted by $\max\{1 - \lambda_2, 1 - |\lambda_m|\}$ and the *relaxation time* τ is the inverse of the spectral gap. The relaxation time can be given in terms of the Dirichlet form of the Markov chain by the equation

$$\tau = \sup\left\{\frac{2\sum_\sigma P(\sigma)(f(\sigma))^2}{\sum_{\sigma\neq\tau} Q(\sigma,\tau)(f(\sigma) - f(\tau))^2}\right\} \tag{1.3}$$

where f is any function on configurations, $Q(\sigma, \tau) = P(\sigma)P(\sigma \to \tau)$ and $P(\sigma \to \tau)$ is transition probability from σ to τ. We use the result that the for reversible Markov chains the relaxation time satisfies

$$\tau \leq \tau_{mix} \leq \tau\left(1 + \frac{1}{2}\log(\min_\sigma P(\sigma))^{-1}\right) \tag{1.4}$$

where τ_{mix} is the mixing time (see e.g. [1]) and so by bounding the relaxation time we can bound the mixing time up to a polynomial factor.

For our proofs it will be useful to use the notion of *block dynamics*. The Gibbs sampler can be generalized to update blocks of vertices rather than individual vertices. For blocks $V_1, V_2, \ldots, V_k \subset V$ with $V = \cup_i V_i$ the block dynamics of the Gibbs sampler updates a configuration σ by choosing a block V_i uniformly at random and assigning the spins in V_i according to the Gibbs distribution conditional on the spins on $G - \{V_i\}$. The relaxation time of of the Gibbs sampler can be given in terms of the relaxation time of the block dynamics and the relaxation times of the Gibbs sampler on the blocks.

PROPOSITION 1.1. *If τ_{block} is the relaxation time of the block dynamics and τ_i is the maximum the relaxation time on V_i given any boundary condition from $G - \{V_i\}$ then by Proposition 3.4 of [17]*

$$\tau \le \tau_{block}(\max_i \tau_i) \max_{v \in V}\{\#j : v \in V_j\}. \tag{1.5}$$

1.3 Erdős-Rényi Random Graphs and Other Models of graphs The Erdős-Rényi random graph $G(n, p)$, is the graph with n vertices V and random edges E where each potential edge $(u, v) \in V \times V$ is chosen independently with probability p. We take $p = d/n$ where $d \ge 1$ is fixed. In the case $d < 1$, it is well known that with high probability all components of $G(n, p)$ are of logarithmic size which implies immediately that the dynamics mix in polynomial time for all β.

For a vertex v in $G(n, d/n)$ let $V(v, l) = \{u \in G : d(u, v) \le l\}$, the set of vertices within distance l of v, let $S(v, l) = \{u \in G : d(u, v) = l\}$, let $E(v, l) = \{(u, w) \in G : u, w \in V(v, l)\}$ and let $B(v, l)$ be the graph $(V(v, l), E(v, l))$.

Our results only require some simple features of the neighborhoods of all vertices in the graph.

DEFINITION 1.3. *Let $G = (V, E)$ be a graph and v a vertex in G. Let $t(G)$ denote the* tree access *of G, i.e.,*

$$t(G) = |E| - |V| + 1.$$

We call a path $v_1, v_2, \ldots$ self avoiding *if for all $i \ne j$ it holds that $v_i \ne v_j$. We let the* maximal path density *m be defined by*

$$m(G, v, l) = \max_\Gamma \sum_{u \in \Gamma} d_u$$

where the maximum is taken over all self-avoiding paths Γ starting at v with length at most l and d_u is the degree of node u. We write $t(v, l)$ for $t(B(v, l))$ and $m(v, l)$ for $m(B(v, l), v, l)$.

1.4 Our Result

THEOREM 1.1. *Let G be a random graph distributed as $G(n, d/n)$. When*

$$\tanh(\beta) < \frac{1}{e^2 d},$$

there exists constant a $C = C(\beta)$ such that the mixing time of the Glauber dynamics is $O(n^C)$ with high probability as n goes to ∞. The result holds for the homogeneous model (1.1) and for the inhomogeneous model (1.2) provided $|h_v| \le 100\beta n$ for all v.

The theorem above may be viewed as a special case of the more general result.

THEOREM 1.2. *Let $G = (V, E)$ be any graph on n vertices satisfying the following properties. There exist $a > 0, 0 < b < \infty$ and $0 < c < \infty$ such that for all $v \in V$ it holds that*

$$t(v, a \log n) \le b \log n, \quad m(v, a \log n) \le c \log n.$$

Then if

$$\tanh(\beta) < \frac{a}{e^{1/a}(c - a)},$$

there exists constant a $C = C(a, b, c, \beta)$ such that the mixing time of the Glauber dynamics is $O(n^C)$. The result holds for the homogeneous model (1.1) and for the inhomogeneous model (1.2) provided $|h_v| \le 100\beta n$ for all v.

REMARK 1.1. *The condition that $|h_v| \le 100\beta n$ for all v will be needed in the proof of the result in the general case (1.2). However, we note that given Theorem 1.2 as a black box, it is easy to extend the result and provide an efficient sampling algorithm in the general case without any bounds on the h_v. In the case where some of the vertices v satisfy $|h_v| \ge 10\beta n$, it is easy to see that the target distribution satisfies except with exponentially small probability that $\sigma_v = +$ for all v with $h_v > 10\beta n$ and $\sigma_v = -$ for all v with $h_v < -10\beta n$. Thus we may set $\sigma_v = +$ when $h_v > 10\beta n$ and $\sigma_v = -$ when $h_v < 10\beta n$ and consider the dynamics where these values are fixed.Doing so will the effectively restrict the dynamics to the graph spanned by the remaining vertices and will modify the values of h_v for the remaining vertices; however, it is easy to see that all remaining vertices will have $|h_v| \le 100\beta n$. It is also easy to verify that if the original graph satisfied the hypothesis of Theorem 1.2 then so does the restricted one. Therefore we obtain an efficient sampling procedure for the desired distribution.*

1.5 Related Work and Open Problems Most results for mixing rates of Gibbs samplers are stated in terms of the maximal degree. Thus for sampling uniform colorings, the result are of the form: for every graph where all degrees are at most d if the number of colors q satisfies $q \ge q(d)$ then Gibbs sampling is rapidly mixing [23, 22, 4, 7, 12, 3, 2, 18]. For example, it is well known and easy to see that one can take $q(d) = 2d$. Similarly, results for the Ising model are stated in terms of $\beta < \beta(d)$. The novelty of the result presented here is that it allows to study graphs where the average degree is small while some degrees may be large.

Previous attempts at studying this problem for sampling uniform colorings yielded weaker results. In [5] it is shown that Gibbs sampling rapidly mixes on $G(n, d/n)$ if $q = \Omega_d((\log n)^\alpha)$ where $\alpha < 1$ and that a variant of the algorithm rapidly mixes if $q \ge \Omega_d(\log \log n / \log \log \log n)$. Indeed the main open problem of [5] is to determine if one can take q to be a function of d only. Our result here provides a positive answer to the analogous question for the Ising model. We further note that other results where the conditions on degree are relaxed [13] do not apply in our setting.

Subsequent to completing this work we have answered the open problem of [5] and have shown in [19] that the Glauber dynamics mixes in polynomial time for a fixed number of colors $q = q(d)$ which does not depend on n. This required a different approach, based on a novel partitioning of the graph into tree-like blocks in which we could bound the spectral gap of the block dynamics. The results also generalize to the hard-core model at low fugacity and to general models of soft constraints at high temperatures. Spirakis and Eythymiou [8] independently have also produced a scheme for approximately sampling from the random coloring distribution in polynomial time. They take a different approach, instead of sampling using MCMC they assign colours to vertices one at a time by calculating the conditional marginal distributions making use of the decay in correlation on the graph.

The following propositions are easy and well known. They show that for $d > 1$ and large β the mixing time is exponential in n and that for all $d > 0$ and $\beta > 0$ the mixing time is more than $n\mathrm{polylog}(n)$. Their proof is omitted from this extended abstract

PROPOSITION 1.2. *If $d > 0$ and $\beta > 0$ then with high probability the mixing time of the dynamics on $G(n, d/n)$ is at least $n^{1+\Omega(1/\log\log n)}$.*

PROPOSITION 1.3. *If $d > 1$ then there exists β'_d such that if $\beta > \beta'_d$ then the with probability going to 1, the mixing time of the dynamics on $G(n, d/n)$ is $\exp(\Omega(n))$.*

It is natural to conjecture that properties of the Ising model on the branching process with $Poisson(d)$ offspring distribution determines the mixing time of the dynamics on $G(n, d/n)$. In particular, it is natural to conjecture that the critical point for *uniqueness* of Gibbs measures plays a fundamental role [10, 21] as results of similar flavor were recently obtained for the hard-core model on random bipartite d *regular* graphs [20].

CONJECTURE 1.1. *If $d\tanh(\beta) > 1$ then with probability going to 1 as $n \to \infty$ over $G(n, d/n)$ the mixing time of Gibbs sampler is $\exp(\Omega(n))$. If $d > 1$ and $d\tanh(\beta) < 1$ then with probability going to 1 as $n \to \infty$ over $G(n, d/n)$ the mixing time of the Gibbs sampler is polynomial in n.*

After proposing the conjecture we have learned that a recent result of Antoine Gerschenfeld and Andrea Montanari implies exponential slow mixing for the Ising model on $G(n, d/n)$ when $d\tanh(\beta) > 1$ [11].

1.6 Proof Technique Our proof follows the following main steps.

- Analysis of the mixing time for Gibbs sampling on trees of varying degrees. We find a bound on the mixing time on trees in terms of the maximal sum of degrees along any simple path from the root. This implies that for *all* β if we consider a tree where each node has number of descendants that has Poisson distribution with parameter $d-1$ then with high probability the mixing time of Gibbs sampling on the tree is polynomial in its size. The motivation for this step is that we are looking at tree-like graphs Note however, that the results established here hold for all β, while rapid mixing for $G(n, d/n)$ does not hold for all β. Our analysis here holds for all boundary conditions and all external fields on the tree.

- We next use standard comparison arguments to extend the result above to case where the graph is a tree with a few edges added. Note that with high probability for all $v \in G(n, d/n)$ the induced subgraph $B(v, \frac{1}{2}\log_d n)$ on all vertices of distance at most $\frac{1}{2}\log_d n$ from v is a tree with at most a few edges added. (Note this still holds for all β).

- We next consider the effect of the boundary on the root of the tree. We show that for tree of $a\log n$ levels, the total variation distance of the conditional distribution at the root given all $+$ boundary conditions and all $-$ boundary conditions is $n^{-1-\Omega(1)}$ with probability $1 - n^{-1-\Omega(1)}$ provided $\beta < \beta_d$ is sufficiently small (this is the only step where the fact that β is small is used).

- Using the construction of Weitz [23] and a Lemma from [2] we show that the spatial decay established in the previous step also holds with probability $1 - o(1)$ for all neighborhoods $B(v, a\log n)$ in the graph.

- The remaining steps use the fact that a strong enough decay of correlation inside blocks each of which is rapidly mixing implies that the dynamics on the full graph is rapidly mixing. This idea is taken from [6].

- In order to show rapid mixing it suffices to exhibit a coupling of the dynamics starting at all $+$ and all $-$ that couples with probability at least $1/2$ in polynomial time. We show that the monotone coupling (where the configuration started at $-$ is always "below" the configuration started at $+$) satisfies this by showing that for each v in polynomial time the two configurations at v coupled except with probability $n^{-1}/(2e)$.

- In order to establish the later fact, it suffices to show that running the dynamics on $B(v, a\log n)$ starting at all $+$ and all $+$ boundary conditions and the dynamics starting at all $-$ and all $-$ will couple at v except with probability $n^{-1}/(2e)$ within polynomial time.

- The final fact then follows from the fact that the dynamics inside $B(v, a\log n)$ have polynomial mixing time and that the stationary distributions in $B(v, \frac{1}{2}\log_d n)$ given $+$ and $-$ boundary conditions agree at v with probability at least $1 - n^{-1}/(4e)$.

We note that the decay of correlation on the self-avoiding tree defined by Weitz that we prove here allows a different sampling scheme from the target distribution. Indeed, this decay of correlation implies that given any assignment to a subset of the vertices S and any $v \notin S$ we may calculate using the Weitz tree of radius $a\log n$ in polynomial time the conditional probability that $\sigma(v) = +$ up to an additive error of $n^{-1}/100$. It is easy to see that this allow sampling the distribution in polynomial time. More specifically, consider the following algorithm from [23].

ALGORITHM 1.1. Fix a radius parameter r and label the vertices $v_1, \ldots, v_n$. Then the algorithm approximately samples from $P(\sigma)$ by assigning the spins of v_i sequentially. Repeating from $1 \le i \le n$:

- In step i construct $T^r_{SAW}(v_i)$, the tree of self-avoiding walks truncated at distance r from v_i.

- Calculate
$$p_i = P_{T^r_{SAW}}(\sigma_{v_i} = + | \sigma_{\{v_1,\ldots,v_{i-1}\}}, \tau_{A-V_{i-1}}).$$
(The boundary conditions at the tree can be chosen arbitrarily; in particular, one may calculate p_i with no boundary conditions).

- Fix $\sigma_{v_i} = X_{v_i}$ where X_{v_i} is a random variable with $p_i = P(X_{v_i} = +) = 1 - P(X_{v_i} = -)$.

Then we prove that:

THEOREM 1.3. *Let G be a random graph distributed as $G(n, d/n)$. When*
$$\tanh(\beta) < \frac{1}{d},$$
for any $\gamma > 0$ there exist constants $r = r(d, \beta, \gamma)$ and $C = C(d, \beta, \gamma)$ such that with high probability Algorithm 1.1, with parameter $r\log n$, has running time $O(n^C)$ and output distribution Q with $d_{TV}(P, Q) < n^{-\gamma}$. The result holds for the homogeneous model (1.1) and for the inhomogeneous model (1.2).

THEOREM 1.4. *Let $G = (V, E)$ be any graph on n vertices satisfying the following properties. There exist $a > 0, 0 < b < \infty$ such that for all $v \in V$,*

$$|V_{T_{SAW}(v)}(v, a\log n)| \le b^{a\log n} \tag{1.6}$$

where $V_{T_{SAW}(v)}(v, r) = \{u \in T_{SAW}(v) : d(u, v) \le r\}$. When
$$\tanh(\beta) < \frac{1}{b},$$
for any $\gamma > 0$ there exist constants $r = r(a, b, \beta, \gamma)$ and $C = C(a, b, \beta, \gamma)$ such that Algorithm 1.1, with parameter $r\log n$, has running time $O(n^C)$ and output distribution Q with $d_{TV}(P, Q) < n^{-\gamma}$. The result holds for the homogeneous model (1.1) and for the inhomogeneous model (1.2).

The proof of theorems 1.4 and 1.3 can be found in the long version of the paper at `http://arxiv.org/abs/0704.3603.`

1.7 Acknowledgment E.M. thanks Andrea Montanari and Alistair Sinclair for interesting related discussions.

2 Proofs

Recall that the local neighborhood of a vertex in $G(n, d/n)$ looks like a branching process tree. In the first step of the proof we bound the relaxation time on a tree generated by a Galton-Watson branching process. More generally, we show that trees that are not too dense have polynomial mixing time.

DEFINITION 2.1. *Let T be a finite rooted tree. We define $m(T) = max_\Gamma \sum_{v\in\Gamma} d_v$ where the maximum is taken over all simple paths Γ emanating from the root and d_v is the degree of node v.*

THEOREM 2.1. *Let τ be the relaxation time of the Glauber dynamics on T where $0 \le \beta_{u,v} \le \beta$ for all u and v and given arbitrary boundary conditions and external field. Then*
$$\tau \le \exp(4\beta m(T)).$$

Proof. We proceed by induction on m with a similar argument to the one used in [2] for a regular tree. Note that if $m = 0$ the claim holds true since $\tau = 1$. For the general case, let v be the root of T, and denote its children by $u_1, \ldots, u_k$ and denote the subtree of the descendants of u_i by T^i. Now let T' be the tree obtained by removing the k edges from v to the u_i, let P' be the Ising model on T' and let τ' be the relaxation time on T'. By equation (1.3) we have that

$$\tau/\tau' \le \frac{\max_\sigma P(\sigma)/P'(\sigma)}{\min_{\sigma,\tau} Q(\sigma,\tau)/Q'(\sigma,\tau)} \le \exp(4\beta k). \tag{2.7}$$

Now we divide T' into $k+1$ blocks $\{\{v\}, \{T^1\}, \ldots, \{T^k\}\}$. Since these blocks are not connected to each other the mixing time of the block dynamics is simply 1. By applying Proposition 3.4 of [17] we get that the relaxation time on T' is simply the maximum of the relaxation times on the blocks,
$$\tau' \le \max\{1, \tau^i\}.$$

where τ^i is the relaxation time on T^i. Note that by the definition of m, it follows that the value of m for each of the subtrees T^i satisfies $m(T^i) \le m-k$, and therefore for all i it holds that $\tau^i \le \exp(4\beta(m-k))$. This then implies by (2.7) that $\tau \le \exp(4\beta m)$ as needed. □

For the applications considered for random and sparse graphs, it is not always the case that the neighborhood of a vertex is a tree, instead it is sometimes a tree with a small number of edges added. Using standard comparison arguments we show that the mixing time of a graph that is a tree with a few edges added is still polynomial. We also show that with high probability for the $G(n, d/n)$ the neighborhoods of all vertices are tree-like.

PROPOSITION 2.1. *Let G be a graph on r vertices with $r+s-1$ edges that has a spanning tree T with $m(T) = m$. Then the mixing time τ of the Glauber dynamics on G with any boundary conditions and external fields satisfies:*

$$\tau \le \exp(4\beta(m+s)).$$

Proof. Repeating the argument of Theorem 2.1, these edges of $G \backslash T$ costs at most $\exp(4\beta s)$ to the relaxation time. The proof follows. □

2.1 Some properties of Galton Watson Trees Here we prove a couple of useful properties for Galton Watson trees that will be used below. We let T be the tree generated by a Galton-Watson branching process with offspring distribution N such that for all t, $E\exp(tN) < \infty$ and such that $E(N) = d$. Of particular interest to us would be the Poisson distribution with mean d which has

$$E\exp(tN) = \exp(d(e^t - 1)).$$

We let T_n denote the first n levels of T. We let $M(n)$ denote the value of m for $T(n)$ and $\tau(n)$ the supremum of the mixing times given any boundary conditions and external fields assuming that $\beta = \sup \beta_{u,v}$. We denote by Z_n the number of descendants at level n.

THEOREM 2.2. *Under the assumptions above we have:*

- *There exists a positive function $c(t)$ such that for all t and all n:*
$$E[\exp(tM(n))] \le \exp(c(t)n).$$
- *Then $E\tau(n) \le C(\beta)^n$ for some $C(\beta) < \infty$ depending on $\beta = \sup \beta_{u,v}$ only.*
- *If N is the Poisson distribution with mean d then for all $t > 0$,*
$$\sup_n E[\exp(tZ_n d^{-n})] < \infty.$$

Proof. Let K denote the degree of the root of T_n and for $1 \le i \le K$ let $M_i(n-1)$ denote the value of m for the sub-tree of T_n rooted at the i'th child. Then:

$$\begin{aligned}
&E[\exp(tM(n))] \\
&= E[\max(1, \max_{1\le i\le K} \exp(t(M_i(n-1)+K)))] \\
&\le E[(1+\exp(tK))\sum_{i=1}^{K} \exp(tM_i(n-1))] \\
&= E[(1+K\exp(tK))]E[\exp(tM(n-1))].
\end{aligned}$$

and so the result follows by induction provided that $c(t)$ is large enough so that

$$\exp(c(t)) \ge E(1+K\exp(tK)).$$

For the second statement of the theorem, note that by the previous theorem we have that

$$E\tau(n) \le E[\exp(4\beta M(n))],$$

where $M(n)$ is the random value of m for the tree T_n so if $C(\beta) = \exp(c(4\beta))$ then $E\tau(n) \le C(\beta)^n$.

For the last part of the theorem, let N_i be independent copies of N and note that

$$\begin{aligned}
&E\exp(tZ_{n+1}) \\
&= E\exp(\sum_{i=0}^{Z_n} td^{-(n+1)}N_i) \\
&= E[E[\exp(\sum_{i=0}^{Z_n} td^{-(n+1)}N_i | Z_n]] \\
&= E[(E[\exp(td^{-n+1}N)])^{Z_n}] \\
&= E\exp(\log(E\exp(td^{-(n+1)}N))Z_n)
\end{aligned} \tag{2.8}$$

which recursively relates the exponential moments of Z_{n+1} to the exponential moments of Z_n. In particular since all the exponential moments of Z_1 exist, $E\exp(tZ_n) < \infty$ for all t and n. When $0 < s \le 1$

$$E\exp(sN) = \exp(d(e^s - 1)) \le \exp(sd(1+s)). \tag{2.9}$$

Now fix a t and let $t_n = t\exp(2t\sum_{i=n+1}^{\infty} d^{-i})$. For some sufficiently large j we have that $\exp(2t\sum_{i=n+1}^{\infty} d^{-i}) < 2$ and $t_n d^{-(n+1)} < 1$ for all $n \ge j$. Then for $n \ge j$ by equations (2.8) and (2.9),

$$\begin{aligned}
&E\exp(t_{n+1}Z_{n+1}d^{-(n+1)}) \\
&= E\exp(\log(E\exp(t_{n+1}d^{-(n+1)}N_i))Z_n) \\
&\le E\exp(t_{n+1}(1+t_{n+1}d^{-(n+1)})Z_n d^{-n}) \\
&\le E\exp(t_{n+1}(1+2td^{-(n+1)})Z_n d^{-n}) \\
&\le E\exp(t_n Z_n d^{-n})
\end{aligned}$$

and so

$$\sup_{n\geq j} E\exp(tZ_n d^{-n}) \leq \sup_{n\geq j} E\exp(t_n Z_n d^{-n}) = E\exp(t_j Z_j d^{-j}) < \infty$$

which completes the result. □

When the branching process is super-critical, the number of vertices is $O((EW)^n)$ and the result above gives that the mixing time is polynomial in the number of vertices on Galton Watson branching process with high probability. We remark that all our bounds here are increasing in the degrees of the vertices so if a random tree T is stochastically dominated by a Galton-Watson branching process then the same bound applies.

2.2 Structure of Random Graphs

LEMMA 2.1. *Let G be a random graph distributed as $G(n,d/n)$. For $0 < a < \frac{1}{2\log d}$ there exists some $c(a,d)$ such that with high probability, for all $v \in G$, $m(v,a\log n) \leq c\log n$. There exists $k = k(a,d) > 0$ such that with high probability, for all $v \in G$, $t(v,a\log n) \leq k$.*

Proof. We construct a spanning tree $T(v,l)$ of $B(v,l)$ in a standard manner. Take some arbitrary ordering of the vertices of G. Start with the vertex v and attach it to all its neighbors in G. Now take the minimal vertex in $S(v,1)$, according to the ordering, and attach it to all its neighbors in G which are not already in the graph. Repeat this for each of the vertices in $S(v,1)$ in increasing order. Repeat this for $S(v,2)$ and continue until $S(v,l-1)$ which completes $T(v,l)$. By construction this is a spanning tree for $B(v,l)$. The construction can be viewed as a breadth first search of $B(v,l)$ starting from v and exploring according to our ordering.

By a standard argument $T(v,a\log n)$ is stochastically dominated by a Galton-Watson branching process with offspring distribution Poisson(d). Then by repeating the argument of Theorem 2.2 for some δ,

$$E\exp(m(T(v,a\log n),v,a\log n)) \leq \delta^{a\log n}$$

and so, $P(m(T(v,a\log n),v,a\log n)) > (a\delta+2)\log n) = O(n^{-2})$ which implies that with high probability $m(T(v,a\log n),v,a\log n)) > (a\delta+2)\log n$ for all v.

If Z_l are the number of offspring in generation l of a Galton-Watson branching process with offspring distribution Poisson(d) by Theorem 2.2 we have that $\sup_l E\exp(Z_l/d^l) < \infty$ and since

$$P(|S(v,l)| > 3d^l\log n) \leq P(\exp(Z_l/d^l) > n^3) \leq n^{-3}E\exp(Z_l/d^l),$$

it follows by a union bound over all $v \in G$ and $1 \leq l \leq a\log n$ we have with high probability for all v, $|B(v,a\log n)| \leq 3(1-d^{-1})n^{a\log d}\log n$.

In the construction of $T(v,a\log n)$ there may be some edges in $B(v,a\log n)$ which are not explored and so are not in $T(v,a\log n)$. Each edge between $u,w \in V(v,a\log n)$ which is not explored in the construction of $T(v,a\log n)$ then is present in $B(v,a\log n)$ independently of $T(v,a\log n)$ with probability d/n. There are at most $(3(1-d^{-1})n^{a\log d}\log n)^2$ unexplored edges. Now when $k > 1/(1-2a\log d)$,

$$P(\text{Binomial}((3(1-d^{-1})n^{a\log d}\log n)^2, d/n) > k) = O(n^{k(2a\log d-1)}(\log n)^{2k}) = n^{-1-\Omega(1)}$$

so by a union bound with high probability we have $t(v,a\log n) \leq k$. Now a self-avoiding path in $B(v,a\log n)$ can traverse each of these k edges at most once so this path can be split into at most $k+1$ self-avoiding paths in $T(v,a\log n)$ and hence with high probability $m(v,l) \leq c\log n$ where $c = (k+1)(a\delta+2)$. □

LEMMA 2.2. *When $0 < a < \frac{1}{2\log d}$ with high probability for all $v \in G$,*

$$|V_{T_{SAW}(v)}(v,a\log n)| \leq O(n^{a\log d}\log n)$$

where $V_{T_{SAW}(v)}(v,r) = \{u \in T_{SAW}(v) : d(u,v) \leq r\}$.

Proof. We now count the number of self-avoiding walks of length at most $a\log n$ in $B(v,a\log n)$. By the proof of Lemma 2.7 we have that with high probability for all v, $|B(v,a\log n)| \leq 3(1-d^{-1})n^{a\log d}\log n$ and $t(v,a\log n) \leq k$. Let $e_1,\ldots,e_{t(v,a\log n)}$ denote the edges in $B(v,a\log n)$ which are not in $T(v,a\log n)$. Now every vertex in $u' \in T_{SAW}(v)$ corresponds to a unique self avoiding walk in $B(v,a\log n)$ from v to u. A self-avoiding walk in $B(v,a\log n)$ passes through each edge at most most once so in particular it passes through each of the e_i at most once. So a path which begins at v traverses through some sequence $e_{i_1},\ldots,e_{i_l}$ in particular directions and then ends at u is otherwise uniquely defined since the intermediate steps are paths in $T(v,a\log n)$ which are unique. There are at most $k(k!)$ sequences $e_{i_1},\ldots,e_{i_l}$, there are 2^k choices of directions to travel through them, and at most $3(1-d^{-1})n^{a\log d}\log n$ possible terminal vertices in $B(v,a\log n)$ so $|V_{T_{SAW}(v)}(v,a\log n)| \leq 3(1-d^{-1})2^k k(k!)n^{a\log d}\log n$. □

2.3 Spatial decay of correlation for tree-like neighborhoods

PROPOSITION 2.2. *Let T be a tree such that $m(v,a) \leq m$. Then $S|(v,a)| \leq (\frac{m-a+1}{a})^a$.*

Proof. By a simple induction on the height of the tree $|S(v,a)|$ must be maximized by a spherically symmetric tree, that is one where the degrees of the vertices depend only on their distance to v. If d_i is the degree of a vertex distance i from v then by the arithmetic-geometric inequality

$$|S(v,a)| = d_0 \prod_{i=1}^{a-1}(d_i - 1) \le ((\sum_{i=0}^{a-1} d_i - (a-1))/a)^a$$
$$\le \left(\frac{m-a+1}{a}\right)^a .$$

□

LEMMA 2.3. *If T is a tree, P is the Ising model with arbitrary external field (including $h_v = \pm\infty$ meaning that v has to take $\pm$) and $\beta_{u,v} \le \beta$ for all u, v and r is the number of vertices in $S(v,l)$ then*

$$\frac{P(\sigma_v = +|\sigma_{S(v,l)} \equiv +)}{P(\sigma_v = +|\sigma_{S(v,l)} \equiv -)} \le \exp(2\beta r(\tanh\beta)^l).$$

Proof. The case of no external field and $\beta_{u,v} \equiv \beta$ follows from the second equation in the proof of Theorem 2.1 in [16].

It is a well known and easy to see that on trees when there is no external field $P(\sigma_v = +|\sigma_\Lambda \equiv +)$ is monotone in $\beta_{u,v}$ since the recursions used to calculate $P(\sigma_v = +|\sigma_\Lambda \equiv +)$ are monotone in $\beta_{u,v}$. Consequently Lemma 2.3 holds as long as all the external fields are 0. The fact that the lemma holds for arbitrary external fields follows from Lemma 4.1 of [2] which we state below. □

LEMMA 2.4. *If T is a tree, P is the inhomogeneous Ising model with interaction $\beta_{u,v}$ with external field and $\widetilde{P}$ is the inhomogeneous Ising model with the same interactions $\beta_{u,v}$ but with no external field then*

$$\frac{P(\sigma_v = +|\sigma_\Lambda \equiv +)}{P(\sigma_v = +|\sigma_\Lambda \equiv -)} \le \frac{\widetilde{P}(\sigma_v = +|\sigma_\Lambda \equiv +)}{\widetilde{P}(\sigma_v = +|\sigma_\Lambda \equiv -)}.$$

Now $B(v, a\log n)$ is not in general a tree so we use the self-avoiding tree construction of Weitz [23] to reduce the problem to one on a tree. The tree of self-avoiding walks, which we denote $T_{saw}(v, a\log n)$, is the tree of paths in $B(v, a\log n)$ starting from v and and not intersecting themselves, except at the terminal vertex of the path. Threw this construction each vertex in $T_{saw}(v, a\log n)$ can be identified with a vertex in G which gives a natural way to relate $\Lambda \subset V$ and a configuration σ_Λ to the corresponding $\Lambda' \subset T_{saw}(v, a\log n)$ and configuration $\sigma_{\Lambda'}$. Furthermore if $A, B \subset V$ then $d(A,B) = d(A', B')$. Then Theorem 3.1 of [23] gives the following result. Each vertex (edge) of T_{saw} corresponds to a vertex (edge) of G and P_T is as the Ising model on T_{saw} defined by taking the corresponding external field and interactions.

LEMMA 2.5. *For a graph G and $v \in G$ there exists $A \subset T_{saw}$ and some configuration τ_A on A such that,*

$$P_G(\sigma_v = +|\sigma_\Lambda) = P_T(\sigma_v = +|\sigma_{\Lambda'}, \tau_{A-\Lambda'}).$$

The set A corresponds to the terminal vertices of path which returns to a vertex already visited by the path.

LEMMA 2.6. *Suppose that a, b, c, β satisfy the hypothesis of Theorem 1.2. Then,*

$$\max_{v\in G} P(\sigma_v = +|\sigma_{\Lambda_n} \equiv +) - P(\sigma_v = +|\sigma_{\Lambda_n} \equiv -) = o(\frac{1}{n})$$

where $\Lambda_n \subseteq S(v, a\log n)$.

Proof. Apply Lemma 2.5 we have that if $\Lambda = S(v, a\log n)$ then

$$\frac{P_G(\sigma_v = +|\sigma_\Lambda \equiv +)}{P_G(\sigma_v = +|\sigma_\Lambda \equiv -)} = \frac{P_T(\sigma_v = +|\sigma_{\Lambda'} \equiv +, \tau_{A-\Lambda'})}{P_T(\sigma_v = +|\sigma_{\Lambda'} \equiv -, \tau_{A-\Lambda'})}.$$

Conditioning on τ_A is equivalent to setting the external field at $u \in A$ to sign$(\tau_v)\infty$. Let $\widetilde{P}_T$ be the Ising model on T_{saw} with no external field. Then by Lemma 2.4

$$\frac{P_T(\sigma_v = +|\sigma_{\Lambda'} \equiv +, \tau_{A-\Lambda'})}{P_T(\sigma_v = +|\sigma_{\Lambda'} \equiv -, \tau_{A-\Lambda'})} \le \frac{\widetilde{P}_T(\sigma_v = +|\sigma_{\Lambda'} \equiv +)}{\widetilde{P}_T(\sigma_v = +|\sigma_{\Lambda'} \equiv -)}$$
$$\le \frac{\widetilde{P}_T(\sigma_v = +|\sigma_{S_{saw}(v,a\log n)} \equiv +)}{\widetilde{P}_T(\sigma_v = +|\sigma_{S_{saw}(v,a\log n)} \equiv -)}$$

where $S_{saw}(v, a\log n) = \{u \in T_{saw}(v, a\log n) : d(u,v) = a\log n\}$. Now suppose $v = u_1, u_2, \ldots, u_k$ is a non-repeating walk in T_{saw} and let $u_1', u_2', \ldots, u_k'$ be the corresponding walk in G. Then from the construction of T_{saw} either $u_1', u_2', \ldots, u_k'$ is a non-repeating walk in G or for some $j < k$, $u_j' = u_k'$ in which case u_k is a leaf of T_{saw} and so has degree 1. It also follows from the construction of T_{saw} that the degree of u_i is less than or equal to the degree of u_i' and so we have that $m(v, a\log n) \le m(T_{saw}, v, a\log n) + 1$. The by Proposition 2.2 $|S_{saw}(v, a\log n)| \le (\frac{c\log n - a\log n + 2}{a\log n})^{a\log n} = O(n^{a\log((c-a)/a)}) = o(n^{-1}(\tanh\beta)^{-a\log n})$. Applying Lemma 2.3 completes the result. □

2.4 Proof of the Main Result

Proof. (Theorem 1.2) Let $X_t^\pm$, denote the Gibbs sampler started from all $\pm$, coupled together on G. Fix some vertex $v \in G$. Define four new chains Q_t^+, Q_t^-, Z_t^+ and Z_t^-. These chains run the Glauber dynamics and are coupled with X_t^+ and X_t^- inside $B(v, a\log n)$ but are fixed (i.e. do not update) outside $B(v, a\log n)$. They are given the following initial conditions.

- Q_t^+ starts from all $+$ configuration (and therefore has all $+$ boundary conditions during the dynamics).

- Q_t^- starts from all $-$ configuration (and therefore has all $-$ boundary conditions during the dynamics).

- Z_t^+ starts from all $+$ configuration outside $B(v, a\log n)$ and Z_0^+ is distributed according to the stationary distribution inside $B(v, a\log n)$ given the all $+$ boundary condition (therefore Z_t^+ will have this distribution for all t).

- Z_t^- starts from all $-$ configuration outside $B(v, a\log n)$ and is distributed according to the stationary distribution inside $B(v, a\log n)$ given the all $-$ boundary condition (therefore Z_t^- will have this distribution for all t).

We can initialize Z_t^+ and Z_t^- so that $Z_0^+ \succcurlyeq Z_0^-$ and by monotonicity we have $Z_t^+ \succcurlyeq Z_t^-$ for all t. We also have that $Q_t^+ \succcurlyeq X_t^+ \succcurlyeq X_t^- \succcurlyeq Q_t^-$.

By Lemma 2.6 we have

$$P(Z_t^+(v) \neq Z_t^-(v)) \leq o(n^{-1}),$$

for all t. By Proposition 2.1 the Gibbs sampler on $B(v, a\log n)$ has relaxation time $\tau \leq \exp(4\beta(a+b)\log n)$ and so has mixing $\tau_{mix} \leq \tau(1 + \frac{1}{2}\log(\min_\sigma P(\sigma))^{-1})$. Each vertex has degree at most $c\log n$ so

$$\log(\min_\sigma P(\sigma))^{-1} \leq (\beta|E|)\sum_u |h_u| \leq (100cn^2\beta^2\log n)$$

so $\tau_{mix} \leq O(n^{3+4(a+b)\beta})$. This implies that for $C = 5 + 4(a+b)\beta$ we have that with high probability after $t = O(n^C)$ steps the Gibbs sampler on all G has chosen every vertex at least n^2 times and so

$$P(Q_t^+(v) \neq Z_t^+(v)) = P(Q_t^-(v) \neq Z_t^-(v)) \leq o(n^{-1}).$$

It follows that $P(Q_t^+(v) \neq Q_t^-(v)) \leq o(n^{-1})$ and hence $P(X_t^+(v) \neq X_t^-(v)) \leq o(n^{-1})$ for all v. By a union bound $P(X_t^+ \neq X_t^-) \leq o(1)$ so the mixing time is bounded by $O(n^C)$ as required. □

The following lemma puts random graphs into the setting of Theorem 1.2, the proof is omitted.

LEMMA 2.7. *Let G be a random graph distributed as $G(n, d/n)$. For $0 < a < \frac{1}{2\log d}$ there exists some $c(a,d)$ such that with high probability, for all $v \in G$, $m(v, a\log n) \leq c\log n$. There exists $k = k(a,d) > 0$ such that with high probability, for all $v \in G$, $t(v, a\log n) \leq k$.*

Proof. (Theorem 1.1) By Lemma 2.7 with high probability a random graph satisfies the hypothesis of Theorem 1.2 for small enough β. To prove the result when $\tanh(\beta) < \frac{1}{e^2 d}$ the only modification to the proof of Theorem 1.2 needed is to show that with high probability when $-1/(\log(d\tanh(\beta))) < a < (2\log d)^{-1}$ we still have $P(Z_t^+(v) \neq Z_t^-(v)) \leq o(n^{-1})$. This is done with a modification of Lemma 2.6 to establish decay of correlation on the graph. □

References

[1] D. Aldous and J. A. Fill. Reversible Markov chains and random walks on graphs. book in preparation. Current version online at http://stat-www.berkeley.edu/users/aldous/book.html.

[2] N. Berger, C. Kenyon, E. Mossel, and Y. Peres. Glauber dynamics on trees and hyperbolic graphs. *Probab. Theory Related Fields*, 131(3):311–340, 2005.

[3] M. Dyer, A. Frieze, T. Hayes, and E. Vigoda. Randomly coloring constant degree graphs. In *Proceedings of the 45th Annual IEEE Symposium on Foundations of Computer Science (FOCS'04)*, pages 582–589, 2004.

[4] M. Dyer, A. Frieze, and M. Jerrum. On counting independent sets in sparse graphs. In *Proceedings of 40th IEEE Sypm. on Foundations of Computer Science (FOCS)*, pages 210–217. 1999.

[5] M. Dyer, A.D. Flaxman, A.M. Frieze, and E. Vigoda. Randomly coloring sparse random graphs with fewer colors than the maximum degree. *Random Struct. Algorithms*, 29(4):450–465, 2006.

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

[7] M. Dyer and C.S. Greenhill. On markov chains for independent sets. *J. Algorithms*, 35(1):17–49, 2000.

[8] C. Eythymiou and P.G. Spirakis. Randomly colouring sparse graphs using a constant number of colours. Preprint., 2007.

[9] S. Geman and D. Geman. Stochastic relaxation, gibbs distributions, and the bayesian restoration of images. *IEEE Trans. Pattern Anal. Mach. Intell*, pages 721–741, 1984.

[10] H. O. Georgii. *Gibbs measures and phase transitions*, volume 9 of *de Gruyter Studies in Mathematics*. Walter de Gruyter & Co., Berlin, 1988.

[11] A. Gershchenfeld and A. Montanari. Reconstruction for models on random graphs. At arXiv:0704.3293. To Appear at FOCS 2007.

[12] L. A. Goldberg, R. Martin, and M Paterson. Strong spatial mixing for lattice graphs with fewer colours. In *Proceedings of the 45th Annual IEEE Symposium on Foundations of Computer Science (FOCS'04)*, pages 562–571. 2004.

[13] T. P. Hayes. A simple condition implying rapid mixing of single-site dynamics on spin systems. In *Proceedings of the 47th Annual IEEE Symposium on Foundations of Computer Science (FOCS 2006)*, pages 39–46, 2006.

[14] M. Jerrum and A. Sinclair. Polynomial-time approximation algorithms for ising model (extended abstract). In *Automata, Languages and Programming*, pages 462–475, 1990.

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

[16] R. Lyons. The Ising model and percolation on trees and tree-like graphs. *Comm. Math. Phys.*, 125(2):337–353, 1989.

[17] F. Martinelli. Lectures on Glauber dynamics for discrete spin models. In *Lectures on probability theory and statistics (Saint-Flour, 1997)*, volume 1717 of *Lecture Notes in Math.*, pages 93–191. Springer, Berlin, 1999.

[18] F. Martinelli, A. Sinclair, and D. Weitz. The ising model on trees: Boundary conditions and mixing time. In *Proceedings of the Forty Fourth Annual Symposium on Foundations of Computer Science*, pages 628–639, 2003.

[19] E. Mossel and A. Sly. Gibbs rapidly samples colorings of g(n,d/n). Availibale at Arxiv http://front.math.ucdavis.edu/0707.3241, 2007.

[20] E. Mossel, D. Weitz, and N. Wormald. On the hardness of sampling independent sets beyond the tree threshold. Submitted, 2007.

[21] R. Pemantle and J. E. Steif. Robust phase transitions for Heisenberg and other models on general trees. *Ann. Probab.*, 27(2):876–912, 1999.

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

[23] D. Weitz. Counting indpendent sets up to the tree threshold. In *Proceedings of the thirty-eighth annual ACM symposium on Theory of computing*, pages 140–149. ACM, 2006.