Glauber Dynamics on Trees and Hyperbolic Graphs

Claire Kenyon
LRI, UMR CNRS
Université Paris-Sud, France
kenyon@lri.fr

Elchanan Mossel
Microsoft Research, 1 Microsoft way
Redmond WA 98052, U.S.A.
mossel@microsoft.com

Yuval Peres
University of California, Berkeley
and Hebrew university, Jerusalem
peres@stat.berkeley.edu

Abstract

We study discrete time Glauber dynamics for random configurations with local constraints (e.g. proper coloring, Ising and Potts models) on finite graphs with n vertices and of bounded degree. We show that the relaxation time (defined as the reciprocal of the spectral gap $1-\lambda_2$) for the dynamics on trees and on certain hyperbolic graphs, is polynomial in n. For these hyperbolic graphs, this yields a general polynomial sampling algorithm for random configurations. We then show that if the relaxation time τ_2 satisfies $\tau_2 = O(n)$, then the correlation coefficient, and the mutual information, between any local function (which depends only on the configuration in a fixed window) and the boundary conditions, decays exponentially in the distance between the window and the boundary. For the Ising model on a regular tree, this condition is sharp.

1. Introduction

Context

The method of Markov chain Monte-Carlo (MCMC) is a popular method for sampling from large combinatorial structures, or estimating their cardinality. Two celebrated examples are the MCMC method for approximating the permanent [12, 14], and the MCMC method for sampling uniform coloring (see [28] and the references there). For many sampling problems, it is relatively easy to construct markov chains which have the desired stationary distribution. It is usually harder to estimate the convergence rate to the stationary distribution.

In this paper we focus on one of the most important families of MCMC, known as Glauber dynamics or Gibbs samplers. Glauber dynamics are commonly used to design MCMC's in computer science, see [7, 11, 18, 19, 24, 25, 28].

The main goal of our work is to determine which geometric properties of the underlying graph are most relevant to the mixing rate of the Glauber dynamics. We first describe Glauber dynamics for proper coloring. Let $G = (V, E)$ be a graph. A coloring of V with q colors is proper if no two adjacent vertices are assigned the same color. Glauber dynamics are the following Markov chain on the set of proper colorings: Let $\sigma \in \{1, \dots, q\}^{|V|}$ be a proper coloring. Define τ in the following way: pick uniformly at random one of the vertices v of V. For all $v \neq w \in V$ set $\tau(w) = \sigma(w)$. Let $\tau(v)$ be chosen uniformly at random among all the colors which are not assigned to the neighbors of v. The Markov chain with the above transition rules from σ to τ is the Glauber dynamics for colorings of G.

In general we may want to assign different weights to different colors, and allow a mixture of softcore constraints (where adjacent vertices are allowed to have the same color, with some penalty) and hardcore constraints (where a non-proper coloring has probability 0). A way of doing so is given by particle systems (using the physics terminology).

To define a general particle system [17] on an undirected graph $G = (V, E)$, define a configuration as an element σ of A^V where A is some finite alphabet, and to each edge $(v, w) \in E$, associate a weight function $\lambda_{vw} : A \times A \to \mathbb{R}$. The Gibbs distribution assigns configuration σ probability proportional to $\prod_{\{v,w\}\in E} \lambda_{vw}(\sigma_v, \sigma_w)$. The Ising model (for which $\lambda_{vw}(\sigma_v, \sigma_w) = e^{\beta\sigma_v\sigma_w}$) and the Potts model are examples of such systems; so is the coloring model (for which $\lambda_{vw} = 1_{\sigma_v \neq \sigma_w}$)

On a finite graph, Glauber dynamics is the following reversible Monte-Carlo method for sampling from the particle system. Given the current configuration σ, pick a vertex v uniformly at random, and replace σ_v by a random spin σ'_v chosen according to the Gibbs distribution conditional on the rest of the configuration:

$$\frac{\mathbf{P}[\sigma'_v = i \mid \sigma]}{\mathbf{P}[\sigma'_v = j \mid \sigma]} = \prod_{w:\{v,w\}\in E} \frac{\lambda_{vw}(i,\sigma_w)}{\lambda_{vw}(j,\sigma_w)}.$$

The efficiency of the Glauber dynamics approach to sampling depends on the rate of convergence to the stationary distribution.

In section 2.1, we describe a connection between the geometry of a graph and the mixing time of Glauber dynamics on it. In particular, we show that for balls in hyperbolic tilings, the Glauber dynamics for the Ising model, the Potts model and proper coloring with $\Delta + 2$ colors (where Δ is the maximal degree), have polynomial mixing time. An example of such a graph can be obtained from the binary tree by adding horizontal edges across levels; another example is in Figure 1. In sections 2.2-4 we study Glauber dynamics

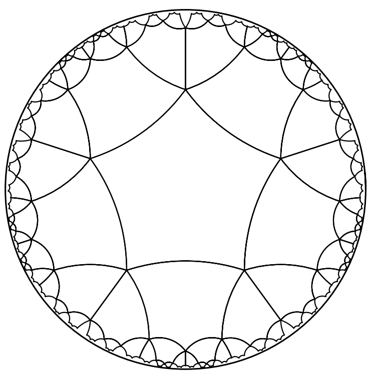

Figure 1. A ball in hyperbolic tiling

for the Ising model on regular trees. Of course, in this case there are alternate, much easier methods to generate a sample from the Gibbs distribution: namely, it suffices to scan the tree top-down once in order to create the σ_v's. Thus the objective of this part is not to obtain the optimal sampling technique for the Ising model on trees, but rather to analyze the Glauber dynamics on trees, as an interesting family of Markov chains which undergoes a phase transition as the temperature varies. The insights obtained from this analysis are useful for other graphs, where there is no better sampling method available. For the trees (1) it follows from the discussion at the first part of the paper that the mixing time is polynomial at all temperatures, and (2) we characterize the range of temperatures for which the inverse spectral gap (which measures the mixing time up to an $O(n)$ factor) is linear. The first fact is slightly surprising, since it is often believed that the two sides of a phase transition should correspond to polynomial versus *exponential* mixing times for the associated dynamics In fact, that belief is not true for the Ising model on trees: here the two sides of the high/intermediate versus low temperature phase transition just correspond to linear versus *superlinear* inverse spectral gap. As a byproduct of the second fact, we exhibit another surprising phenomenon: contrary to common beliefs, there is a range of temperatures in which the inverse spectral gap is linear, even though there are many Gibbs measures on the infinite tree.

In section 5 of the paper we go beyond trees and hyperbolic graphs and study Glauber dynamics for families of finite graphs of bounded degree. We show that if the inverse spectral gap of the Glauber dynamics on the ball centered at ρ grows linearly in the volume of the ball, then the correlation between the state of a vertex ρ and the states of vertices at distance r from ρ, must decay exponentially in r.

Setup

The graphs. Let $G = (V, E)$ be an infinite graph with maximal degree Δ. Let ρ be a distinguished vertex and denote by $G_r = (V_r, E_r)$ the induced graph on $V_r = \{v \in V : \mathrm{dist}(\rho, v) \le r\}$. Let n_r be the number of vertices in G_r. At some parts of the paper we will focus on the case where $G = T = (V, E)$ is the infinite b-ary tree. In these cases, $T_r = (V_r, E_r)$ will denote the r-level b-ary tree.
The Ising model. In the Ising model on G_r at inverse temperature β, every configuration $\sigma \in \{-1, 1\}^{V_r}$ is assigned probability

$$\mu(\sigma) = Z(\beta)^{-1} \exp\left(\beta \sum_{\{v,w\}\in E_r} \sigma_v \sigma_w\right)$$

where $Z(\beta)$ is a normalizing constant. When $G_r = T_r$, this measure has the following equivalent definition [8]: Fix $\epsilon = (1 + e^{2\beta})^{-1}$. Pick a random spin ± 1 uniformly for the root of the tree. Scan the tree top-down, assigning vertex v a spin equal to the spin of its parent with probability $1 - \epsilon$ and opposite with probability ϵ.
Glauber dynamics. Glauber dynamics for the Ising model chooses the new spin σ'_v in such a way that:

$$\frac{\mathbf{P}[\sigma'_v = +1 \mid \sigma]}{\mathbf{P}[\sigma'_v = -1 \mid \sigma]} = \exp\left(2\beta \sum_{w:\{w,v\}\in E_r} \sigma(w)\right).$$

See [17] or [20] for more background.
Mixing times.

Definition 1.1 *For a reversible Markov chain, let* $1 = \lambda_1 > \lambda_2 \ge \ldots \ge \lambda_k \ge -1$ *be the eigenvalues of the transition matrix. The* **spectral gap** *of the chain is defined as* $\max\{1-\lambda_2, 1-|\lambda_k|\}$*, and the* **relaxation time**, τ_2, *is defined as the inverse of the spectral gap.*

Definition 1.2 *For measures μ and ν on the same discrete space, the* **total-variation distance***, $d_V(\mu,\nu)$, between μ and ν is defined as*

$$d_V(\mu,\nu) = \frac{1}{2}\sum_x |\mu(x) - \nu(x)|\,.$$

Definition 1.3 *Consider an ergodic Markov chain $\{X_t\}$ with stationary distribution π on a finite state space. Denote by $\mathbf{P}_x^t$ the law of X_t given $X_0 = x$. The* **mixing time** *of the chain, τ_1, is defined as*

$$\tau_1 = \inf\{t : \sup_{x,y} d_V(\mathbf{P}_x^t, \mathbf{P}_y^t) \le e^{-1}\}.$$

For $t \ge \ln(1/\epsilon)\tau_1$, we have

$$\sup_x d_V(\mathbf{P}_x^t, \pi) \le \sup_{x,y} d_V(\mathbf{P}_x^t, \mathbf{P}_y^t) \le \epsilon.$$

This paper focuses on analyzing the relaxation time τ_2. Using τ_2 one can bound the mixing time τ_1, since every reversible chain with stationary distribution π satisfies (see, e.g., [1]),

$$\frac{1}{2}\tau_2 \le \tau_1 \le \tau_2\left(1 + \frac{1}{2}\log\left((\min_\sigma \pi(\sigma))^{-1}\right)\right). \qquad (1)$$

For the Markov chains studied in this paper, this gives $\frac{1}{2}\tau_2 \le \tau_1 \le O(n)\tau_2$.

Results

Exposure and relaxation time.

Definition 1.4 *The* **exposure** *$\mathcal{E}(G)$ of a graph G is the smallest integer such that there exists a labeling $v_1, \dots, v_n$ of the vertices such that for all $1 \le k \le n$, the number of edges from $\{v_1, \dots, v_k\}$ to $\{v_{k+1}, \dots, v_n\}$, is at most $\mathcal{E}(G)$.*

Remark: The *vertex-separation* of a graph G is defined analogously to the exposure in terms of vertices among $\{v_1, \dots, v_k\}$ that are adjacent to $\{v_{k+1}, \dots, v_n\}$. In [16] it is shown that the vertex-separation of G equals its *path-width*, see [26].

see [26]. Generalizing an argument in [20, Theorem 6.4] for $\mathbf{Z}^d$, (see also [12]), we prove:

Proposition 1.1 *Consider the Ising model on a finite graph G with n vertices and maximal degree Δ. Then the relaxation time of the Glauber dynamics is at most $n^2 e^{(4\mathcal{E}(G)+2\Delta)\beta}$.*

Similarly, for the coloring model on G, if the number of colors q satisfies $q \ge \Delta + 2$, then the relaxation time of the Glauber dynamics is at most $(\Delta+1)n^2(q-1)^{\mathcal{E}(G)+1}$.

Analogous results hold for the independent set and hard core models.

Relaxation time for the Ising model on the tree. The Ising model on the b-ary tree has three different regimes, see [3, 8]. In the high temperature regime, where $1 - 2\epsilon < 1/b$, there is a unique Gibbs measure on the infinite tree, and the expected value of the spin at the root σ_ρ given any boundary conditions $\sigma_{\partial T_r}$ decays exponentially in r. In the intermediate regime, where $1/b < 1 - 2\epsilon < 1/\sqrt{b}$, the exponential decay described above still holds for typical boundary conditions, but not for certain exceptional boundary conditions, such as the all $+$ boundary; consequently, there are infinitely many Gibbs measures on the infinite tree. In the low temperature regime, where $1 - 2\epsilon > 1/\sqrt{b}$, typical boundary conditions impose bias on the expected value of the spin at the root σ_ρ.

Theorem 1.2 *Consider the Ising model on the b-ary tree T_r of height r. Let $\epsilon = (1 + e^{2\beta})^{-1}$. The relaxation time τ_2 for Glauber dynamics on T_r can be bounded as follows:*

1. *The relaxation time is polynomial at all temperatures: $\tau_2 = n_r^{O(\log(1/\epsilon))}$.*
2. **Low temperature regime***.*
   - *(a) If $1 - 2\epsilon \ge 1/\sqrt{b}$ then the relaxation time is superlinear: $\tau_2 = \Omega(n_r^{1+\log_b(b(1-2\epsilon)^2)})$.*
   - *(b) Moreover, the degree of τ_2 tends to infinity as ϵ tends to zero: $\tau_2 = n_r^{\Omega(\log(1/\epsilon))}$.*
3. **Intermediate and high temperature regimes***.* *If $1 - 2\epsilon < 1/\sqrt{b}$ then the relaxation time is linear: $\tau_2 = O(n_r)$.*

In particular we obtain from Equation (1) that in the low temperature region $\tau_1 = n_r^{\Theta(\beta)}$, and in the intermediate and high temperature regions $\tau_1 = O(n_r^2)$. We conjecture that $\tau_1 = O(n_r \log n_r)$ in the intermediate and high temperature regions but can only prove this when $1 - 2\epsilon < (2\sqrt{b})^{-1}$. There is no evidence that there is any qualitative difference in the behavior of Glauber dynamics between the high temperature region (when there is a unique Gibbs measure on the infinite tree) and the intermediate temperature region.

We emphasize that Theorem 1.2 implies that in the intermediate region $1/2 < 1 - 2\epsilon < 1/\sqrt{b}$, the relaxation time is bounded by a constant times the volume n, yet, in the infinite volume there are infinitely many Gibbs measures. This Theorem is perhaps easiest to appreciate when compared to other results on the Gibbs distribution for the Ising model on binary trees, summarized in Table 1.

The proof of the low temperature result is quite general and applies to other models with “soft” constraints, such as Potts models on the tree (see [15, 23] for more details).

Temp.	$1-2\epsilon$	$\sigma_\rho \| \sigma_{\partial T} \equiv +$	$I(\sigma_\rho, \sigma_{\partial T})$	τ_2
high	$< 1/2$	unbiased	$\to 0$	$O(n)$
med.	$\in (\frac{1}{2}, \frac{1}{\sqrt{2}})$	biased	$\to 0$	$O(n)$
low	$> \frac{1}{\sqrt{2}}$	biased	inf> 0	$n^{1+\Omega(1)}$
freeze	$1-o(1)$	biased	$1-o(1)$	$n^{\Theta(\beta)}$

Table 1. The Ising model on binary trees. Here the root is denoted ρ, and the vertices at distance r from the root are denoted ∂T.

Spectral gap and correlations. At infinite temperature, where distinct vertices are independent, the Glauber dynamics on a graph of n vertices reduces to a random walk on a discrete n-dimensional cube, where it is well known that the relaxation time is $\Theta(n)$. Our next result shows that at any temperature where such fast relaxation takes place, a strong form of independence holds. This is well known in $\mathbf{Z}^d$, see [20], but our formulation is valid for any graph of bounded degree. Denote by σ_r the configuration on all vertices at distance r from ρ.

Theorem 1.3 *If G has bounded degree and the relaxation time of the Glauber dynamics satisfies $\tau_2(G_r) = O(n_r)$, then the Gibbs distribution on G_r has the following property. For any fixed finite set of vertices A, there exists $c_A > 0$ such that for r large enough*

$$\mathrm{Cov}(f, g) \le e^{-c_A r} \mathrm{Var}(f) \mathrm{Var}(g)\,, \tag{2}$$

provided that $f(\sigma)$ depends only on σ_A and $g(\sigma)$ depends only on σ_r. Equivalently, there exists $c'_A > 0$ such that

$$I(\sigma_A, \sigma_r) \le e^{-c'_A r}\,, \tag{3}$$

where I denotes mutual information, see [6].

This theorem holds in a very general setting which includes Potts models, random colorings, and other local-interaction models.

Our proof of Theorem 1.3 uses "disagreement percolation" and a coupling argument exploited by van den Berg, see [2], to establish uniqueness of Gibbs measures in $\mathbf{Z}^d$; according to F. Martinelli (personal communication) this kind of argument is originally due to B. Zegarlinski. Note however, that Theorem 1.3 holds also when there are multiple Gibbs measures – as the case of the Ising model in the intermediate regime demonstrates. Moreover, combining Theorem 1.3 and Theorem 1.2, one infers that for $1 - 2\epsilon < 1/\sqrt{b}$, we have $\lim_{r\to\infty} I(\sigma_0, \sigma_r) = 0$. This yields another proof of this fact which was proven before in [3, 9, 8].

Plan of the paper

In section 2 we prove Proposition 1.1 via a canonical path argument, and give the resulting polynomial time upper bound of Theorem 1.2 part 1. We also present a more elementary proof of the upper bound on the relaxation time for the tree, which gives sharper exponents; this proof uses Dirichlet forms to analyze the spectral gap by induction on the height of the tree. In section 3 we sketch a proof of Theorem 1.2 part 2a and present a proof of Theorem 1.2 part 2b. These lower bounds are obtained by finding a low conductance "cut" of the configuration space, using global majority of the boundary spins for the former result, and recursive majority for the latter result. In section 4 we establish the high temperature result, using comparison to block dynamics which are analyzed via path-coupling. Finally, in section 5 we prove Theorem 1.3 by a Peierls argument controlling "paths of disagreement" between two coupled dynamics.

2. Polynomial Upper Bounds

2.1 Exposure and mixing time

We begin by showing how Proposition 1.1 implies the upper bound in Theorem 1.2 part 1. Applying Proposition 1.1 to the b-ary tree with r levels, using the Depth First Search labeling to get an upper bound on the exposure, we see that the relaxation time of the Glauber dynamics is at most

$$C(\epsilon) n_r^{2+2(b-1)\log_b \frac{1-\epsilon}{\epsilon}} = n_r^{O(\log(1/\epsilon))},$$

hence Theorem 1.2 part 1. Similarly, the argument shows that the mixing time for the Glauber dynamics is polynomial in the number of vertices n for other "hyperbolic" graphs (More precisely, our proof applies to balls in infinite planar graphs with positive Cheeger constant and bounded degree; this includes all hyperbolic tilings). For such graphs (as in Figure 1), ordering the vertices in a clockwise manner for a well-chosen geometric embedding yields exposure which is logarithmic in the volume (see [15]). For these graphs, this polynomial mixing is quite surprising. Indeed, it is often believed that long-range correlations imply slow mixing time; yet in these graphs, at low temperature, the correlation between σ_u and σ_v is bounded below, independently of the distance between u and v. Such long range correlations hold for any family of planar graphs with bounded degrees and co-degrees such that that the boundary of each subset containing at most $1/2$ of the vertices is at least logarithmic in the size of the subset (details in [15]).

We now prove Proposition 1.1, following the lines of the proof given in [20, Theorem 6.4] for the Ising model in $\mathbf{Z}^d$, (see also [12]). We first discuss the proof for the Ising model. Let Γ be the graph corresponding to the transitions of the Markov chain on the graph G. Between any two configurations σ and η, we define a "canonical path" $\gamma(\sigma,\eta)$ as follows. Fix an order $<$ on the vertices of G which achieves the exposure. Consider the vertices $v_1 < v_2 < \dots$ at which $\sigma_v \neq \eta_v$.

We define the kth configuration $\sigma^{(k)}$ on the path $\gamma(\sigma,\eta)$ by giving spin σ_v to every vertex labeled $v \le v_k$, spin η_i to every vertex labeled $i > k$, and spin $\sigma_v = \eta_v$ for every unlabeled vertex v. Note that $\sigma^{(0)} = \eta$ and $\sigma^{(d(\sigma,\eta))} = \sigma$. Since $\sigma^{(k-1)}$ and $\sigma^{(k)}$ are identical except for the spin of vertex v_k, they are adjacent in G. This defines $\gamma(\sigma,\eta)$.

Note that there at most $\mathcal{E}(G)$ pairs of adjacent vertices (v_i, v_j) such that $i \le k < j$, hence any configuration on the canonical path between σ and η will have at most $\mathcal{E}(G)$ edges between spins copied from σ and spins copied from η.

Using canonical paths to bound the mixing rate. Let

$$\rho = \sup_e \sum_{\sigma,\eta:\, e\in\gamma(\sigma,\eta)} \frac{\mu[\sigma]\mu[\eta]}{Q(e)},$$

where the supremum is over transitions $e = (u,v)$ between adjacent configurations. Here μ is the stationary measure (i.e. the Gibbs distribution), and for any two adjacent configurations u and v, $Q((u,v)) = \mu[u]\mathbf{P}[u \to v]$. If L is the maximal length of a canonical path, then by the argument in [12, 20], the relaxation time of the Markov chain is at most

$$\tau_2 \le L\rho. \tag{4}$$

Since $L \le n$, it follows that $\tau_2 \le n\rho$, thus it only remains to prove an upper bound on ρ.

Analysis of the canonical path. For each directed edge $\overrightarrow{e}$ in G, we define an injection from canonical paths going through e in the specified direction, and configurations of G. To a canonical path $\gamma(\sigma,\eta)$ going through e, such that $e = (\sigma^{(k-1)}, \sigma^{(k)})$, we associate the configuration φ which has spin η_i for every $i < k$ and spin σ_i for every $i \ge k$. This is an injection.

By the property of our labeling,

$$\mu[\sigma]\mu[\eta] \le \mu[\sigma^{(k-1)}]\mu[\varphi]e^{4\mathcal{E}(G)\beta}. \tag{5}$$

Now a short calculation concludes the proof:

$$\begin{aligned}
\rho &\le \sup_e \sum_{\sigma,\eta \text{ s.t. } e\in\gamma(\sigma,\eta)} \frac{\mu[\sigma]\mu[\eta]}{\mu[\sigma^{(k-1)}]\mathbf{P}[\sigma^{(k-1)} \to \sigma^{(k)}]} \\
&\le e^{4\mathcal{E}(G)\beta} \sup_e \sum_{\varphi} \frac{\mu[\sigma^{(k-1)}]\mu[\varphi]}{\mu[\sigma^{(k-1)}]\mathbf{P}[\sigma^{(k-1)} \to \sigma^{(k)}]} \qquad (6) \\
&\le e^{4\mathcal{E}(G)\beta} n e^{2\Delta\beta} \sum_{\varphi} \mu[\varphi] \le n e^{(4\mathcal{E}(G)+2\Delta)\beta}. \qquad (7)
\end{aligned}$$

The last inequality follows from the fact that the map $\gamma \to \varphi$ is injective and therefore $\sum_\varphi \mu[\varphi] \le 1$.

Paths for coloring. This argument does not directly extend to coloring, as the configurations $\sigma^{(k)}$ in the definition of the path may not be proper colorings. Assume that $q \ge \Delta + 2$ and let $v_1 < v_2 \cdots < v_n$ be an ordering of the vertices of G which achieves the exposure. We construct a path $\gamma(\sigma,\eta)$ such that

$$|\gamma(\sigma,\eta)| \le (\Delta+1)n. \tag{8}$$

Moreover, for all $\tau \in \gamma(\sigma,\eta)$ there exists a k such that

$$\tau_v = \begin{cases} \sigma_v & \text{if } v \le v_k \\ \eta_v & \text{if } v > v_k \text{ and } v \nsim \{v,\dots,v_k\} \end{cases} \tag{9}$$

When estimating ρ, we note that in the right hand side of (5) $e^{4\mathcal{E}(G)\beta}$ is now replaced by 1, as all the legal configurations have the same weight. On the other hand, the map $\gamma \to \varphi$ is not injective. Instead, by (9), there are at most $(q-1)^{\mathcal{E}(G)}$ paths which are mapped to the same coloring. We therefore obtain that for coloring $\rho \le n(q-1)^{\mathcal{E}(G)+1}$ and therefore from (8) and (5),

$$\tau_2 \le (\Delta+1)n^2(q-1)^{\mathcal{E}(G)+1}.$$

The way to construct a path $\gamma(\sigma,\eta)$ satisfying (8) and (9) is by changing the colors of the vertices $v_1,\dots,v_n$ according to their order with some local modifications. Suppose that τ satisfies (9). In order to construct the next configuration we first modify the colors of all w s.t. $(w, v_{k+1}) \in E$, and $\tau_w = \sigma_{v_{k+1}}$. This is possible by the assumption that $q \ge \Delta + 2$. Then we set the next configuration to have color $\sigma_{v_{k+1}}$ at v_{k+1}. □

We continue by analyzing an improved upper bound on relaxation time for the tree. The analysis below yields better exponents for the mixing time, and the proof is simpler. However, the proof below applies to trees only. We note that both the proof above and the proof given below may be adapted to prove polynomial time mixing for Glauber dynamics of any bounded range interacting particle system with "soft" constraints on the tree.

2.2 A recursive argument

It is helpful to refer to figure 2 to follow the proof for the case $b = 2$. Let μ denote the Gibbs measure. To estimate the relaxation time τ_2, our proof uses the characterization in terms of Dirichlet forms (see e.g. [1]):

$$\tau_2 = \sup\left\{ \frac{2\sum_\sigma \mu[\sigma](g(\sigma))^2}{\sum_{\tau\neq\sigma} Q(\sigma,\tau)(g(\sigma)-g(\tau))^2} \;:\; \mu(g) = 0 \right\}, \tag{10}$$

where $Q(\sigma,\tau) = \mu[\sigma]\mathbf{P}[\sigma \to \tau]$.

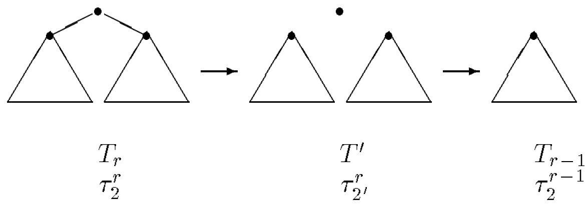

Figure 2. Upper bound on the relaxation time for trees

Our proof is in two steps and uses induction on the height of the tree. Let $T = T_r$ be the b-ary tree of r levels and $\tau_2[r]$ be the relaxation time for the Glauber dynamics on T.

First reduction. Let T' be the graph obtained from T by removing the b edges connecting the root of T to its b children, let $\tau_2'[r]$ be the spectral gap of the Glauber dynamics on T' and denote by μ' the Gibbs measure on T'. From Equation 10, we have

$$\frac{\tau_2[r]}{\tau_2'[r]} \leq \frac{\max_\sigma \mu[\sigma]/\mu'[\sigma]}{\min_{\sigma,\tau} Q(\sigma,\tau)/Q'(\sigma,\tau)}.$$

Note that $\mu[\sigma] \leq (2-2\epsilon)^b \mu'[\sigma]$ for all σ, and a little thought reveals that $Q(\sigma,\tau) \geq (2\epsilon)^b Q'(\sigma,\tau)$ for any adjacent pair σ, τ. We therefore obtain that

$$\tau_2[r] \leq \left(\frac{1-\epsilon}{\epsilon}\right)^b \tau_2'[r]\,. \qquad (11)$$

Second reduction. T' has $b+1$ connected components: one single node, and b copies of T_{r-1}. One step of the Glauber dynamics on T' can be simulated as follows: with probability $1/n$, run one step of the Glauber dynamics on the graph consisting of a single node, and otherwise run a step of the Glauber dynamics on one of the b copies of T_{r-1} (with probability $(n-1)/(bn)$ for each of them).

The relaxation time on T' is easily calculated from the relaxation times of the components:

$$\tau_2'[r] = \max\{\frac{bn}{n-1}\tau_2[r-1], n\}. \qquad (12)$$

Finishing the proof. Combining the two recursion equations (11) and (12), we obtain:

$$\begin{aligned}\tau_2[r] &\leq \frac{n_r}{n_r-1} b \left(\frac{1-\epsilon}{\epsilon}\right)^b \tau_2[r-1] \\ &\leq C b^r \left(\frac{1-\epsilon}{\epsilon}\right)^{br} \\ &\leq C n_r^{1+b\log_b \frac{1-\epsilon}{\epsilon}} .\square\end{aligned}$$

3. Lower Bounds

The superlinear lower bound of Theorem 1.2 part 2a is a direct consequence of the extremal characterization of τ_2 given in equation 10, applied to the particular test function g which sums the spins on the boundary of the tree. This function has Dirichlet form which is $O(1)$, and from the variance given for example in [8], we can deduce its second moment.

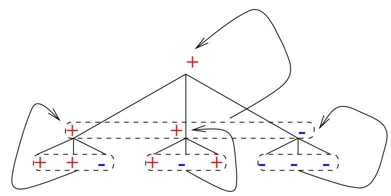

Figure 3. The recursive majority function.

In order to prove the lower bound on the relaxation time for very low temperatures stated in Theorem 1.2 part 2b, we apply (10) to the test function g which is obtained by applying recursive majority to the boundary spins; see [22] for background regarding the recursive-majority function for the Ising model on the tree. For simplicity we consider only the ternary tree T_3, see figure 3. (for other trees and sharper bounds see [15]).). Recursive majority is defined on the configuration space as follows. Given a configuration σ, first label each boundary vertex v by its spin σ_v. Next, inductively label each interior vertex w with the label of the majority of the children w. The value of the recursive majority function g is then the label of the root. We write σ_v for the spin at v and m_v for the recursive majority value at v.

Lemma 3.1 *If u and w are children of the same parent v, then* $\mathbf{P}[m_u \neq m_w] \leq 2\epsilon + 8\epsilon^2$.

Proof:

$$\mathbf{P}[m_u \neq m_w] \leq$$

$$\mathbf{P}[\sigma_u \neq m_u] + \mathbf{P}[\sigma_w \neq m_w] + \mathbf{P}[\sigma_u \neq \sigma_v] + \mathbf{P}[\sigma_w \neq \sigma_v].$$

We will show that recursive majority is highly correlated with spin, *i.e.* if ϵ is small enough (say $\epsilon < 0.01$), then $\mathbf{P}[m_v \neq \sigma_v] \leq 4\epsilon^2$.

The proof is by induction on the distance ℓ from v to the boundary of the tree. For a vertex v at distance ℓ from the boundary of the tree, write $p_\ell = \mathbf{P}[m_v \neq \sigma_v]$. By definition $p_0 = 0 \leq 4\epsilon^2$.

For the induction step, note that if $\sigma_v \neq m_v$ then one of the following events hold:

- At least 2 of the children of v, have different σ value than that of σ_v, or
- One of the children of v has a spin different from the spin at v, and for some other child w we have $m_w \neq \sigma_w$, or
- For at least 2 of the children of v, we have $\sigma_w \neq m_w$.

Summing up the probabilities of these events, we see that $p_\ell \leq 3\epsilon^2 + 6\epsilon p_{\ell-1} + 3p_{\ell-1}^2$. It follows that $p_\ell \leq 4\epsilon^2$, hence the Lemma. □

Proof of Theorem 1.2 part 2b. Let m be the recursive majority function. Then from symmetry $\mathbf{E}[m] = 0$, and $\mathbf{E}[m^2] = 1$. By plugging m in definition (10), we see that

$$\tau_2 \geq \left(\sum_{\sigma,\tau : m[\sigma]=1, m[\tau]=-1} \mu[\sigma]\mathbf{P}[\sigma \to \tau] \right)^{-1} . \quad (13)$$

Observe that if σ, τ are adjacent configurations (i.e., $\mathbf{P}[\sigma \to \tau] > 0$) such that $m(\sigma) = 1$ and $m(\tau) = -1$, then there is a unique vertex v_r on the boundary of the tree where σ and τ differ. Moreover, if $\rho = v_1, \ldots, v_r$ is the path from ρ to v_r, then for σ we have $m(v_1) = \ldots = m(v_r) = 1$ while for τ we have $m(v_1) = \ldots = m(v_r) = -1$. Writing u_i, w_i for the two siblings of v_i for $2 \leq i \leq k$, we see that for all i, for both σ and τ we have $m(u_i) \neq m(v_i)$. Note that these events are independent for different values of i. We therefore obtain that the probability that $v_1, \ldots, v_r$ is such a path is bounded by $(2\epsilon + 8\epsilon^2)^{r-1}$. Since there are 3^r such paths and since $\mathbf{P}[\sigma \to \tau] \leq 3^{-r}$ we obtain that the right term of (13) is bounded below by

$$(2\epsilon + 8\epsilon^2)^{1-r} \geq n^{\Omega(\beta)} .□$$

Noam Berger (personal communication) has refined the recursive argument in order to obtain polynomial mixing time for any (ergodic) particle system on the tree.

4. Higher temperatures

We now prove Theorem 1.2 part 3. Our analysis uses a comparison to block dynamics.

Block dynamics. We view our tree $T = T_r$ as a part of a larger b-ary tree T_* of height $r + 2h$, where the root ρ of T is at level h in T_*. For each vertex v of T_*, consider the subtree of height h rooted at v. A **block** is by definition the intersection of T with such a subtree. At each step of the block dynamics, we pick a block at random, erase all the spins of vertices belonging to the block, and put new spins in, according to the Gibbs distribution conditional on the spins in the rest of T.

A coupling analysis. We use a weighted Hamming metric on configurations,

$$d(\sigma, \eta) = \sum_v \lambda^{|v|} 1(\sigma_v \neq \eta_v),$$

where $|v|$ denotes the distance from vertex v to the root. Let $\theta = 1 - 2\epsilon$ and $\lambda = 1/\sqrt{b}$. Note that $b\lambda\theta < 1$ and $\theta < \lambda$. Starting from two distinct configurations σ and η, our coupling always picks the same block in σ and in η and choose the coupling between the two block moves which minimizes $d(\sigma', \eta')$.

We use path-coupling [4], *i.e.*, we will prove that for every pair of configurations which differ by a single spin, applying one step of the block dynamics will reduce the expected distance between the two configurations.

Let v be the single vertex, such that $\sigma_v \neq \eta_v$. Then $d(\sigma, \eta) = \lambda^{|v|}$. Let B denote the chosen block, and σ', η' be the configurations after the move. There are four situations to consider.

Case 1. If B contains neither v nor any vertex adjacent to v, then $d(\sigma', \eta') = d(\sigma, \eta)$.

Case 2. If B contains v, then $\sigma' = \eta'$ and $d(\sigma', \eta') = 0 = d(\sigma, \eta) - \lambda^{|v|}$. There are h such blocks, corresponding to the h ancestors of v at $1, 2, \ldots, h$ generations above v. (Note that this holds even when v is the root of T or a leaf of T, because of our definition of blocks).

Case 3. If B is rooted at one of v's children, then the conditional probabilities given the outer boundaries of B are not the same since one block has $+1$ above it and the other block has -1 above it. However both blocks have their leaves adjacent to the same boundary configuration. Since conditioning on this lower boundary can only help by Lemma 4.1 below, we bound $d(\sigma', \eta')$ by studying the case where one block is conditioned to having a $+1$ adjacent to the root, the other block is conditioned to having a -1 adjacent to the root, and otherwise the boundary is free. Then the block is simply filled in a top-down manner, every edge is faithful (*i.e.* the spin of the current vertex equals the spin of its parent) with probability θ and cuts information (the spin of the current vertex is a new random spin) with probability $1-\theta$. Coupling these choices for corresponding edges for σ and for η, we see that the distance between σ' and η' will be equal to the weight of the cluster containing v, in expectation $\sum_j \lambda^{|v|+j} b^j \theta^j \leq \lambda^{|v|}/(1 - b\lambda\theta)$. There are b such blocks, corresponding to the b children of v.

Case 4. If B is rooted at v's ancestor exactly $h + 1$ generations above v, then the conditional probabilities are not the same since one block has a leaf v adjacent to a $+1$ and the other block has a leaf adjacent to a -1. There is exactly one such block. Again we appeal to Lemma 4.1 to show that the expected distance is dominated by the size of the θ cluster of w. The expected weight of v's cluster is bounded by summing over the ancestors w of v:

$$\sum_w \theta^{|v|-|w|} \sum_j \lambda^{|w|+j} b^j \theta^j =$$
$$= \frac{\sum_w \lambda^{|w|}\theta^{|v|-|w|}}{1-b\lambda\theta}$$
$$= \frac{\lambda^{|v|}}{(1-\theta\lambda^{-1})(1-b\lambda\theta)}.$$

Overall, the expected change in distance is

$$\mathbf{E}(d(\sigma',\eta') - d(\sigma,\eta)) \le [1ex] \left(\frac{b\lambda^{|v|}}{1-b\lambda\theta} + \frac{\lambda^{|v|}}{(1-\theta\lambda^{-1})(1-b\lambda\theta)} - h\lambda^{|v|}\right)\frac{1}{n+h-1}.$$

If the block height h is a sufficiently large constant, we get that for some positive constant c,

$$\mathbf{E}(d(\sigma',\eta') - d(\sigma,\eta)) \le \frac{-c\lambda^{|v|}}{n} \le \frac{-c}{n} d(\sigma,\eta). \quad (14)$$

Note that $\max d(\sigma,\eta) = \sum_{j\le r} b^j\lambda^j \le \sqrt{n}$. Therefore, by a path-coupling argument (see [4]) we obtain a mixing time of at most $O(n\log n)$ for the blocks dynamics.

Spectral gap of block dynamics. The $(1-c/n)$ contraction at each step of the coupling implies, by an argument from [5] which we now recall, that the spectral gap of the block dynamics is at least c/n. Indeed, let λ_2 be the second largest eigenvalue in absolute value, and f an eigenvector for λ_2. Let $M = \sup_{\sigma,\eta} |f(\sigma) - f(\eta)|/d(\sigma,\eta)$ and denote by $\mathbf{P}$ the transition operator. Then

$$\begin{aligned}
&|\lambda_2|M \\
&= \sup_{\sigma,\eta} \frac{|\mathbf{P}f(\sigma) - \mathbf{P}f(\eta)|}{d(\sigma,\eta)} \quad \text{since } f \text{ eigenvector for } \lambda_2 \\
&\le \sup_{\sigma,\eta} \sum_{\sigma',\eta'} \mathbf{P}[(\sigma,\eta)\to(\sigma',\eta')] \frac{|f(\sigma')-f(\eta')|}{d(\sigma',\eta')} \frac{d(\sigma',\eta')}{d(\sigma,\eta)} \\
&\le \sup_{\sigma,\eta} \sum_{\sigma',\eta'} \mathbf{P}[(\sigma,\eta)\to(\sigma',\eta')] M \frac{d(\sigma',\eta')}{d(\sigma,\eta)} \\
&= M \sup_{\sigma,\eta} \frac{\mathbf{E}[d(\sigma',\eta')]}{d(\sigma,\eta)} \\
&\le (1-c/n)M \text{ by (14).}
\end{aligned}$$

(Here the first inequality is by coupling and the following one is by definition of M). Thus $|\lambda_2|M \le (1-c/n)M$, whence the block dynamics have relaxation time at most $O(n)$.

Relaxation time for single-site dynamics. Since each block update can be simulated by doing a constant number of single-site updates inside the block, and each tree vertex only belongs to a bounded number of blocks, it follows from proposition 3.4 of [20] that the relaxation time of the single-site Glauber dynamics is also $O(n)$. □

We now state and prove the Lemma which was used in the coupling analysis.

Lemma 4.1 *Let T be a finite tree and μ the Gibbs measure for the Ising model on that tree. For a fixed set of vertices A of T,$v \notin A$, and some boundary conditions τ, we consider the following conditional Gibbs measures:*
μ_+: conditioned on $\sigma_v = 1$.
μ_-: conditioned on $\sigma_v = -1$.
$\mu_{+,\tau}$: conditioned on $\sigma_v = 1$ and $\sigma_A = \tau$.
$\mu_{-,\tau}$: conditioned on $\sigma_v = -1$ and $\sigma_A = \tau$.
Let $S = \sum_{w\in T}\sigma_w$. Then one can couple $\mu_{+,\tau}$ and $\mu_{-,\tau}$ in such a way that the expected number of disagreements is $\frac{1}{2}(\mu_{+,\tau}[S] - \mu_{-,\tau}[S])$. Moreover, for all τ,

$$\mu_{+,\tau}[S] - \mu_{-,\tau}[S] \le \mu_+[S] - \mu_-[S].$$

Proof. The first statement follows from the fact that $\mu_{-,\tau}$ is dominated by $\mu_{+,\tau}$ (for definitions and basic properties of domination of measures, see [1, 17]). Therefore using a coupling between these measures which respects the domination we see that the expected number of disagreements is $\frac{1}{2}(\mu_{+,\tau}[S] - \mu_{-,\tau}[S])$. For the second statement, it suffices to show that for all vertices w,

$$\mu_{+,\tau}[\sigma_w] - \mu_{-,\tau}[\sigma_w] \le \mu_+[\sigma_w] - \mu_-[\sigma_w]. \quad (15)$$

Reduction from trees to paths. We first claim that it suffices to prove (15) when the tree T consists of a path $v = v_1,\dots,v_k = w$ where every vertex v_i is connected to a vertex u_i of degree 1 by a bond of interaction-strength γ_i, and where the boundary condition is the configuration of $(u_i)_{i=1}^k$ (see Figure 4). The proof is omitted in this extended abstract.

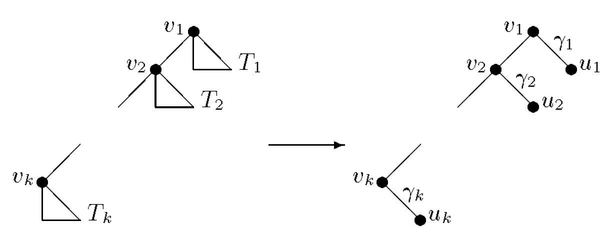

Figure 4. Reduction from trees to paths.

We will now prove the lemma by induction on the length of the path $v_1,\dots,v_k$.

Paths of length 2. Assume $k = 2$. Writing β for the strength of (v_1,v_2) interaction, γ for the strength of the $(w = v_2, u_2)$ interaction,

$$\mu_+[\sigma_w] - \mu_-[\sigma_w] = \frac{e^\beta - e^{-\beta}}{e^\beta + e^{-\beta}} - \frac{e^{-\beta} - e^\beta}{e^{-\beta} + e^\beta} = 2\tanh\beta,$$

and

$$
\begin{aligned}
&\mu_{+,\tau}[\sigma_w] - \mu_{-,\tau}[\sigma_w] \\
&= \frac{e^{\beta+\gamma} - e^{-\beta-\gamma}}{e^{\beta+\gamma} + e^{-\beta-\gamma}} - \frac{e^{-\beta+\gamma} - e^{\beta-\gamma}}{e^{-\beta+\gamma} + e^{\beta-\gamma}} \\
&= \tanh(\beta+\gamma) - \tanh(\gamma-\beta).
\end{aligned}
$$

It therefore suffices to prove that for $\beta > 0$, the function

$$\gamma \mapsto g(\beta,\gamma) = \tanh(\gamma+\beta) - \tanh(\gamma-\beta)$$

has a unique maximum at $\gamma = 0$. Consider the partial derivative,

$$g_\gamma(\beta,\gamma) = \cosh^{-2}(\gamma+\beta) - \cosh^{-2}(\gamma-\beta). \tag{16}$$

Therefore, if $\beta > 0$ and $\gamma > 0$ then $g_\gamma(\beta,\gamma) < 0$ and if $\beta > 0$ and $\gamma < 0$ then $g_\gamma(\beta,\gamma) > 0$. Thus $\gamma = 0$ is the unique maximum and the claim for $k = 2$ follows.

Induction step. We assume that the claim is true for $k-1$ and prove it for k. We denote $v' = v_{k-1}, \mu'_+ = \mu[\cdot|\sigma_{v'} = 1]$ and similarly $\mu'_-, \mu'_{-,\tau}, \mu'_{+,\tau}$. Now,

$$
\begin{aligned}
&\mu_+[\sigma_w] - \mu_-[\sigma_w] \\
&= \big(\mu_+[\sigma_{v'} = 1]\mu'_+[\sigma_w] + \mu_+[\sigma_{v'} = -1]\mu'_-[\sigma_w]\big) \\
&- \big(\mu_-[\sigma_{v'} = 1]\mu'_+[\sigma_w] + \mu_-[\sigma_{v'} = -1]\mu'_-[\sigma_w]\big) \\
&= \frac{1}{2}(\mu_+ - \mu_-)[\sigma_{v'}](\mu'_+ - \mu'_-)[\sigma_w]. \qquad (17)
\end{aligned}
$$

In a similar manner

$$\mu_{+,\tau}[\sigma_w] - \mu_{-,\tau}[\sigma_w] =$$

$$\frac{1}{2}(\mu_{+,\tau} - \mu_{-,\tau})[\sigma_{v'}](\mu'_{+,\tau} - \mu'_{-,\tau})[\sigma_w], \tag{18}$$

and the proof follows since both terms in (17) and (18) are larger for the free measure than for the conditional measure. □

5. Proof of Theorem 1.3

We assume that $\tau_2(G_r) = O(n_r)$. Equivalently, writing λ_2 for the eigenvalue of the dynamics with the second largest absolute value, we assume that $|\lambda_2| \le 1 - cn_r^{-1}$ for some $c > 0$ and all r. Recall that we denoted by σ_r the configuration on all vertices at distance exactly r from ρ.

Mutual information and L^2 estimates. For Markov chains such as $\{\sigma_r\}$, it is generally known [8, 21, 27] that (3) follows from (2), which in turn, is consequence of the following stronger statement:

There exists $c_* > 0$ such that for any vertex set $A \subset G_{r/2}$ and any functions f, g of mean zero,

$$\mathbf{E}(fg) \le e^{-c_* r}(\mathbf{E}(f^2)\mathbf{E}(g^2))^{1/2}, \tag{19}$$

provided that $f(\sigma)$ depends only on σ_A and $g(\sigma)$ depends only on σ_r. We will prove (19) using a coupling argument.

Choose σ drawn from the Gibbs distribution on G_r. Consider a copy of G_r which we denote by G'_r. Run the Glauber dynamics on G_r and G'_r simultaneously, except that on G'_r the boundary variables σ_v for $|v| = r$ are frozen: whenever the dynamics picks such a vertex v, on G'_r the variable labeling it remains fixed. Thus on G'_r the process at all times is at the stationary distribution conditional on σ_r, while on G_r, even given the initial σ_r, the process converges to the (unconditional) stationary distribution.

Initially, the configurations are identical on G_r and on G'_r. We "couple" the dynamics, i.e., we always pick the same site v for G_r and G'_r and, if the neighbors of v have the same spins on G_r and on G'_r then we choose the same new spin for v in G_r and in G'_r. Thus the two processes eventually move apart due to the different behavior on the boundary, which gradually induces different spins further inside the graph.

For a vertex v with $|v| = r$, we define t_v to be the first time at which σ_v was updated. For any $v \in V_r$ with $|v| < r$, we define t_v to be the first time σ_v was updated after $\min_{(w,v)\in E_r} t_w$. Note that at any time $t < t_v$, the labeling of v in G_r and G'_r is identical. Moreover, t_v depends only on the order the vertices are chosen, and is independent of the initial configuration σ. We let $t_A = \min_{v\in A} t_v$.

Given an initial configuration σ, we write X^t for the random configuration after t steps in the G_r dynamics and Y^t for the random configuration after t steps in the G'_r dynamics. We also let $\mathbf{P}^t f(\sigma) = \mathbf{E}[f(X^t)]$, and $Q^t f(\sigma) = \mathbf{E}[f(Y^t)]$. Since for G'_r and all t the process is at the stationary distribution given σ_r, it follows that for all t,

$$\mathbf{E}[fg] = \mathbf{E}[Q^t f \cdot Q^t g] = \mathbf{E}[Q^t f \cdot g].$$

Since Glauber updates cannot increase the L^2 norm, we infer from the coupling above that

$$
\begin{aligned}
&||Q^t f - P^t f||_2^2 \\
&\le 4\mathbf{P}[t \ge t_A]||f||_2^2\,.
\end{aligned}
$$

Therefore, by the Cauchy-Schwarz inequality,

$$\mathbf{E}[(Q^t f - P^t f)g] \le 2\sqrt{\mathbf{P}[t \le t_A]}||f||_2\,||g||_2\,.$$

Since

$$\mathbf{E}[P^t f \cdot g] \le |\lambda_2|^t ||f||_2 ||g||_2,$$

We infer that

$$\mathbf{E}[fg] \le \left(|\lambda_2|^t + 2\sqrt{\mathbf{P}[t \le t_A]}\right)||f||_2||g||_2\,. \tag{20}$$

It remains to bound the two terms in the right-hand side of (20). Recall the hypothesis $\max_{v\in A}|v| \le r/2$ and denote by Δ the maximal degree in G.

We now let $t = c'(r/2)n_r$, where the constant c' will be specified later. We obtain that

$$|\lambda_2|^t \leq (1 - cn_r^{-1})^{c'(r/2)n_r} \leq e^{-rcc'/2}.$$

It remains to bound $\mathbf{P}[t_A \leq t]$. We note that $t_A \leq t$ only if there is some self-avoiding path (sometimes referred to as "path of disagreement") between the set A and the vertices at distance r from ρ, along which the discrepancy between the two distributions has been conveyed in time less than t.

Note that there are at most $|A|(\Delta - 1)^k$ such paths of length k for all $k \geq r/2$. We fix such a path $v_1, \ldots, v_k$ and bound the probability that this path was activated up to time t. This probability is clearly bounded by $\mathbf{P}[\mathsf{Bin}(t, n_r^{-1}) \geq k]$ (think of "success" as an activation of the first non-active element of $v_1, \ldots, v_k$). We let $c' > 0$ be a constant such that for all m and p and for all $k \geq mp/c'$ one has the following tail estimate:

$$\mathbf{P}[\mathsf{Bin}(m, p) \geq k] \leq \Delta^{-2k}.$$

(such a constant exists by standard large deviation estimates, see, e.g., [10, Corollary 2.4]). Thus,

$$\mathbf{P}[\mathsf{Bin}(t, n_r^{-1}) \geq k] \leq \Delta^{-2k}.$$

So summing over all paths we obtain:

$$\mathbf{P}[t_A \leq t] \leq |A| \sum_{k \geq r/2} (\Delta - 1)^k \Delta^{-2k}. \tag{21}$$

Thus both summands in (20) decay exponentially in r, as claimed. □

Acknowledgment. We are grateful to David Aldous, David Levin, Laurent Saloff-Coste and Peter Winkler for useful discussions. The research of Y. Peres was partially supported by NSF grant #DMS-9803597.

References

[1] Aldous, D. and Fill, J. A. (2000) Reversible Markov chains and random walks on graphs, *book in preparation*. Current version available at `www.stat.berkeley.edu/users/aldous/book.html`.

[2] van den Berg, J. (1993) A uniqueness condition for Gibbs measures, with application to the 2-dimensional Ising antiferromagnet. *Comm. Math. Phys.* **152, no. 1**, 161–166.

[3] Bleher, P. M., Ruiz, J. and Zagrebnov V. A. (1995) On the purity of limiting Gibbs state for the Ising model on the Bethe lattice, *J. Stat. Phys* **79**, 473–482.

[4] Bubley, R. and Dyer, M. (1997) Path coupling: a technique for proving rapid mixing in Markov chains. In Proceedings of the 38th Annual Symposium on Foundations of Computer Science (FOCS), 223–231.

[5] Chen, M. F. (1998) Trilogy of couplings and general formulas for lower bound of spectral gap. *Probability towards 2000* Lecture Notes in Statist., **128**, Springer, New York, 123–136.

[6] Cover, T. M. and Thomas, J. A. (1991) *Elements of Information Theory*, Wiley, New York.

[7] Dyer M. and Greenhill C. (2000), 'On Markov chains for independent sets', *J. Algor.* **35,** 17–49.

[8] Evans, W., Kenyon, C., Peres, Y. and Schulman, L. J. (2000) Broadcasting on trees and the Ising Model, *Ann. Appl. Prob.*, **10**, 410–433.

[9] Ioffe, D. (1996). A note on the extremality of the disordered state for the Ising model on the Bethe lattice. *Lett. Math. Phys.* **37**, 137–143.

[10] Janson, S., Luczak T. and Ruciński A. (2000) *Random Graphs*, Wiley, New York.

[11] Jerrum, M. (1995) A very simple algorithm for estimating the number of k-colorings of a low-degree graph. *Rand. Struc. Alg.* **7**, 157–165.

[12] Jerrum, M. and Sinclair, A. (1989). Approximating the permanent. *Siam Jour. Comput.* **18**, 1149–1178.

[13] Jerrum, M. and Sinclair, A. (1993). Polynomial time approximation algorithms for the Ising model. *Siam Jour. Comput.* **22**, 1087–1116.

[14] Jerrum, M., Sinclair, A. and Vigoda, E. (2001). A polynomial-time approximation algorithm for the permanent of a matrix with non-negative entries. Proceedings of the 33*rd Annual ACM Symposium on Theory of Computing*, Crete, Greece.

[15] Kenyon, C., Mossel, E. and Peres, Y. (2001) Glauber dynamics on trees: spectral gap without uniqueness of Gibbs measures, in preparation.

[16] Kinnersley, N. G. (1992) The vertex seperation number of a graph equals its path-width. *Infor. Proc. Lett.*, **42**, 345–350.

[17] Liggett, T. (1985) *Interacting particle systems*, Springer, New York.

[18] Luby, M. and Vigoda, E. (1997) Approximately Counting Up To Four, In proceedings of the 29th Annual Symposium on Theory of Computing (STOC),, 682–687.

[19] Luby, M. and Vigoda, E. (1999). Fast Convergence of the Glauber Dynamics for Sampling Independent Sets, Statistical physics methods in discrete probability, combinatorics and theoretical computer science, *Rand. Struc. Alg.* **15**, 229–241.

[20] Martinelli, F. (1998) Lectures on Glauber dynamics for discrete spin models. *Lectures on probability theory and statistics (Saint-Flour, 1997)* 93–191, *Lecture Notes in Math.* **1717**, Springer, Berlin.

[21] Mossel, E. (2001) Reconstruction on trees: Beating the second eigenvalue, *Ann. Appl. Probab.*, **11 no. 1**,285–300.

[22] Mossel, E. (1998) Recursive reconstruction on periodic trees. *Rand. Struc. Alg.* **13**, 81–97.

[23] Mossel, E. and Peres, Y. (2000) Glauber dynamics and geometry of graphs, in preparation.

[24] Propp, J. and Wilson, D. (1996) Exact Sampling with Coupled Markov Chains and Applications to Statistical Mechanics. *Rand. Struc. Alg.* **9**, 223–252.

[25] Randall, D. and Tetali, P. (2000), Analyzing Glauber dynamics by comparison of Markov chains. *J. of Math. Phys.* **41**, 1598–1615.

[26] N. Robertson and P.D. Seymour (1983), Graph minors. I. Excluding a forest. *J. Comb. Theory Series B* 35, 39–61.

[27] Saloff-Coste, L. (1997) Lectures on finite Markov chains. *Lectures on probability theory and statistics (Saint-Flour, 1996)* 301–413, Lecture Notes in Math. **1665**, Springer, Berlin.

[28] Vigoda, E. (2001). Improved bounds for sampling colorings. Probabilistic techniques in equilibrium and nonequilibrium statistical physics. *J. Math. Phys.* **41**, no. 3, 1555–1569.