Probab. Theory Relat. Fields (2010) 148:37–69
DOI 10.1007/s00440-009-0222-x

Gibbs rapidly samples colorings of $G(n, d/n)$

Elchanan Mossel · Allan Sly

Received: 10 January 2008 / Revised: 29 March 2009 / Published online: 5 May 2009
© The Author(s) 2009. This article is published with open access at Springerlink.com

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 $G(n, d/n)$ 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 $n^{1+\Omega(1/\log \log n)}$ with high probability. High degree nodes pose a technical challenge in proving polynomial time mixing of the dynamics for many models including coloring. Almost all known sufficient conditions in terms of 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 consider sampling q-colorings and show that for every $d < \infty$ there exists $q(d) < \infty$ such that for all $q \geq q(d)$ the mixing time of the Gibbs sampling on $G(n, d/n)$ is polynomial in n with high probability. Our results are the first polynomial time mixing results proven for the coloring model on $G(n, d/n)$ for $d > 1$ where the number of colors does 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

E. Mossel 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 and A. Sly supported by NSF grants DMS-0528488 and DMS-0548249.

E. Mossel (✉) · A. Sly
Department of Statistics, U.C. Berkeley, Berkeley, CA, USA
e-mail: mossel@stat.berkeley.edu

A. Sly
e-mail: sly@stat.berkeley.edu

time is slower than $n\text{polylog}(n)$. In previous work we have shown that similar results hold for the ferromagnetic Ising model. However, the proof for the Ising model crucially relied on monotonicity arguments and the "Weitz tree", both of which have no counterparts in the coloring setting. Our proof presented here exploits in novel ways the local treelike structure of Erdős–Rényi random graphs, block dynamics, spatial decay properties and coupling arguments. Our results give the first polynomial-time algorithm to approximately sample colorings on $G(n, d/n)$ with a constant number of colors. They 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 there exists an $\alpha > 0$ such that every vertex v of the graph has a neighborhood $N(v)$ of radius $O(\log n)$ in which the induced sub-graph is the union of a tree and at most $O(1)$ edges and where each simple path Γ of length $O(\log n)$ satisfies $\sum_{u\in\Gamma}\sum_{v\neq u}\alpha^{d(u,v)} = O(\log n)$. The results also generalize to the hard-core model at low fugacity and to general models of soft constraints at high temperatures.

Keywords Erdős–Rényi random graphs · Gibbs samplers · Glauber dynamics · Mixing time · Colorings

Mathematics Subject Classification (2000) 60J10 · 65C05 · 82C20

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 [7,8,24,25], sampling from the uniform distribution of matchings in a graph [16], or sampling from the uniform distribution of colorings [5,6,12] of a graph. A natural family of approximate sampling techniques is given by Gibbs samplers, also known as the Glauber dynamics or the heat-bath. 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 and the statistical physics perspectives. Approximate convergence in polynomial time, sometimes called *rapid mixing*, is essential in computer science applications. The convergence rate is also of natural interest in physics where the dynamical properties of such distributions are extensively studied, see e.g. [18]. 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 [6–8,12,17,19,24,25] 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 a small number of denser sub-graphs. In a recent work [21], see also [20], we have shown that for any d if $0 \leq \beta < \beta(d)$ is sufficiently small then Gibbs sampling for the *Ising model* on $G(n, d/n)$ rapidly mixes and further that the same result is true in the presence of arbitrary *external field*. The proofs of [21] crucially rely on the monotonicity of the Ising model and on the *"Weitz tree"* [25] which is only defined for two spin models. Thus the proof does not apply to models such as the *hard-core model* or to sampling *uniform coloring*. Other recent work has investigated 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 uniform colorings on $G(n, d/n)$ for $d > 1$ and $O(1)$ colors. This is presented as a major open problem of both [5] and [21].

In this paper we give the first rapid convergence result of Gibbs samplers for the coloring model on Erdős–Rényi random graphs in terms of the average degree and the number of colors only. Our results yields the first FPRAS (Fully Polynomial Randomized Approximation Scheme) for sampling the coloring distribution in this case. Our results are further extended to 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 $O(1)$ edges added and where for each simple path Γ of length $O(\log n)$ it holds that $\sum_{u\in\Gamma}\sum_{v\neq u}\alpha^{d(u,v)} \leq O(\log n)$, where $\alpha > 0$ is some fixed parameter. While the number of colours needed is bounded, we do not attempt to optimize the number or indeed give an explicit bound on it.

Subsequent to completing this work we learned that Efthymiou and Spirakis [9] 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.

Our arguments extend to prove similar results for many other models. In particular, they give an independent proof of rapid mixing for sampling from the Ising model for small inverse temperature β, the hard-core model for small fugacity λ and many other models. Note however, that the result presented here for the Ising model on general graphs are slightly weaker than the result of [21]. Here we require that each $O(\log n)$ radius neighborhood is a tree union a constant number of edges while in [21] an excess of $O(\log n)$ is allowed.

Below we define the coloring model and Gibbs samplers and state our main result for coloring. Some related work and a sketch of the proof are also given in the introduction. The remaining sections give a more detailed proof.

1.1 Models

Our results cover a wide range of graph based distributions including the coloring model, the hardcore model and any model with soft constraints.

Definition 1.1 Let $G = (V, E)$ be a graph and let $\mathcal{C}$ be a set of states/colours with $|\mathcal{C}| = q$. The Hamiltonian is a function $\mathcal{C}^V \to \mathbb{R}$ of the form

$$H(\sigma) = \sum_{u\in V} h(\sigma(u)) + \sum_{(u,v)\in E} g(\sigma(u), \sigma(v)) \tag{1}$$

where $h : \mathcal{C} \to \mathbb{R}$ is the activity function and $g : \mathcal{C}^2 \to \mathbb{R} \cup \{-\infty\}$ is a symmetric interaction function. This defines an interacting particle system which is the distribution on $\sigma \in \mathcal{C}^V$ given by

$$P(\sigma) = \frac{1}{Z} \exp(H(\sigma))$$

where Z is a normalizing constant. We focus our attention on 3 classes of models.

- The *coloring distribution* is the uniform distribution over colorings of G with $g(x, y) = -\infty 1_{\{x=y\}}$ and $h \equiv 0$ so the distribution is given by

$$P(\sigma) = \frac{1}{Z} \prod_{(u,v) \in E} 1_{\{\sigma(u) \neq \sigma(v)\}}. \tag{2}$$

- The *hardcore* model with parameter β is the weighted distribution over independents sets of G given by $\mathcal{C} = \{0, 1\}$ with $h(x) = \beta x$ and $g(x, y) = -\infty 1_{\{x=y=1\}}$ and

$$P(\sigma) = \frac{1}{Z} \exp\left(\beta \sum_{u \in V} \sigma(u)\right) \prod_{(u,v) \in E} 1_{\{\sigma(u)\sigma(v)=0\}} \tag{3}$$

  where σ takes values in $\{0, 1\}^V$ and Z is a normalizing constant.
- If g does not take the value $-\infty$ then we say the model has soft-constraints. This class includes the Ising model.

For $U \subset V$ we let P_U be the colouring model on the subgraph induced by U. Define the *activity free system* $\widehat{P}$ as the distribution with the activity function h set to 0. We define the norm of the Hamiltonian as

$$\|H\| := \max \left\{ \max_{x \in \mathcal{C}} |h(x)|, \max_{x,y \in \mathcal{C}} |g(x, y)| \right\}.$$

1.2 Gibbs sampling

The Gibbs sampler (also Glauber dynamics or heat bath) 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 $G = (V, E)$, a set $\mathcal{C}$ and a Hamiltonian H as in (1), the Gibbs sampler is the discrete time Markov chain on $\mathcal{C}^V$ 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 element $x \in \mathcal{C}$ with probability proportional to

$$\exp\left(h(x) + \sum_{w \in N(v)} g(\sigma(w), x)\right).$$

where $N(v) = \{w \in V : (v, w) \in E\}$. Note that in the case of coloring $\sigma'(v)$ is chosen uniformly from the set $\mathcal{C} \setminus \{\sigma(w) : w \in N(v)\}$.

In the coloring model, it is not completely trivial to find an initial configuration that is a legal coloring. However, for $G(n, d/n)$ finding an initial coloring is easy [23]. It is well known that with high probability if one removes all nodes of large enough degree $D'(d)$ from $G(n, d/n)$ then what remains is a collection of unicyclic components. It is easy to color each unicyclic component with 3 colors and therefore color the graph with $D' + 3$ colors. Similar arguments will allow us to find an initial coloring in the more general setting discussed here. See [10] for a survey of algorithmic results for finding legal coloring in sparse random graphs. For the hard-core model and models with soft constraints, it is trivial to find an initial legal configuration.

While our results are given for the discrete time Gibbs sampler described above, it will at times be convenient to consider the continuous time version of the model. Here sites are updated at rate 1 by independent Poisson clocks. The two chains are closely related, the relaxation time of the continuous time Markov chain is n times the relaxation time of the discrete chain (see e.g. [1]).

We will be interested in the time it takes the dynamics to get close to the distributions (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 $(2e)^{-1}$ from the stationary distribution.

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 $d \geq 1$ is fixed. In the case $d < 1$, it is well known that with high probability all components of $G(n, p)$ are unicyclic and of logarithmic size which implies immediately that the dynamics considered here mix in time polynomial in n.

For a vertex v in $G(n, d/n)$ let $V(v, l) = \{u \in G : d(u, v) \leq l\}$, 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 stated in terms of t and m below.

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

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

For $v \in V$ we let $t(v, l) = t(B(v, l))$.

We call a path $v_1, v_2, \ldots$ *self avoiding* if for all $i \neq j$ it holds that $v_i \neq v_j$.

For $\alpha > 0$ we let the *maximal path α-weight m_α* of a subgraph $H \subset G$ be defined by

$$m_\alpha(H, l) = \max_{\Gamma} \sum_{u \in \Gamma} \sum_{v: u \neq v \in G} \alpha^{d(u,v)}$$

where the maximum is taken over all self-avoiding paths $\Gamma \subset H$ of length at most l.

1.4 Our results

1.4.1 Colouring model

Theorem 1.1 *For all $d \geq 1$ there exists $0 < q(d), C(d) < \infty$ such that for all $q \geq q(d)$ the following holds. Let G be a random graph distributed as $G(n, d/n)$. Then with high probability the mixing time of Gibbs sampling of q-colorings is $O(n^C)$.*

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

Theorem 1.2 *For any $0 < a, \alpha, t, \delta < \infty$ there exists constants $q(a, \alpha, t, \delta)$ and $C = C(a, \alpha, t, \delta)$ such that if $q \geq q(a, \alpha, t, \delta)$ and $G = (V, E)$ is any graph on n vertices satisfying*

$$\forall v \in V, t(v, a \log n) \leq t, \quad m_\alpha(G, a \log n) < \delta \log n, \tag{4}$$

then the mixing time of the Gibbs-sampler of q-colorings of G is $O(n^C)$.

1.4.2 Hardcore model

Theorem 1.3 *For all $d \geq 1$ there exists $C(d), \beta(d) < \infty$ such that for all $\beta \leq \beta(d)$ the following holds. Let G be a random graph distributed as $G(n, d/n)$. Then with high probability the mixing time of Gibbs sampling of the hardcore model with parameter β is $O(n^C)$.*

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

Theorem 1.4 *For any $0 < a, \alpha, t, \delta < \infty$ there exists constants $\beta(a, \alpha, t, \delta)$ and $C = C(a, \alpha, t, \delta)$ such that if $\beta \leq \beta(a, \alpha, t, \delta)$ and $G = (V, E)$ is any graph on n vertices satisfying*

$$\forall v \in V, t(v, a \log n) \leq t, \quad m_\alpha(G, a \log n) < \delta \log n, \tag{5}$$

then the mixing time of the Gibbs-sampler of the hardcore model with parameter β is $O(n^C)$.

1.4.3 Soft constraints

Theorem 1.5 *For all $d \geq 1$ there exists $0 < C(d), H^*(d) < \infty$ such that for all models with $\|H\| \leq H^*(d)$ the following holds. Let G be a random graph distributed as $G(n, d/n)$. Then with high probability the mixing time of Gibbs sampling of the model is $O(n^C)$.*

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

Theorem 1.6 *For any $0 < a, \alpha, t, \delta < \infty$ and all soft constraint models there exists constants $H^*(a, \alpha, t, \delta) > 0$ and $C = C(a, \alpha, t, \delta)$ such that if $\|H\| \leq H^*(a, \alpha, t, \delta)$ and $G = (V, E)$ is any graph on n vertices satisfying*

$$\forall v \in V, t(v, a \log n) \leq t, \quad m_\alpha(G, a \log n) < \delta \log n, \tag{6}$$

then the mixing time of the Gibbs-sampler of the model is $O(n^C)$.

1.5 Related work

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 $q \geq q(d)$ then Gibbs sampling is rapidly mixing [6–8,12,14,17,19,24,25]. For example, it is well known and easy to see that one can take $q(d) = 2d$ (see e.g.[14]). Similarly, results for the Ising model are stated in terms of $\beta < \beta(d)$. The novelty of the result of [21] and the result presented here is that it allows us 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 $q \geq \Omega_d(\log \log n / \log \log \log n)$. Indeed the main open problem of [5] is to determine if one can take q to be a function of d only.

Comparing the results presented here to [20,21] we observe first that there is one sense in which the current results are weaker. In [21] the tree excess t can be of order $O(\log n)$ while for the results presented here t has to be of order $O(1)$. The results of [21] crucially use the fact that the Ising model is attractive (this is a monotonicity property) and that it is a two spin system which allows the use of the "Weitz tree" [25].

We note that for all q and all d the mixing time of Gibbs sampling on $G(n, d/n)$ is with high probability at least $n^{1+\Omega(1/\log \log n)} >> n\text{polylog}(n)$, see [5,21] for details. It is an important challenge to find the critical $q = q(d)$ for rapid mixing. In particular, the question is if the threshold can be formulated in terms of the coloring model on a branching process tree with $Poisson(d)$ degree distribution. One would expect rapid mixing for in the "uniqueness phase", but perhaps even beyond it, see [11,21,22].

1.6 Proof technique

We briefly sketch the main ideas behind the proof focusing on the special case of coloring.

Block dynamics and path coupling. The basic idea of the proof is quite standard. It is based on a combination of *block dynamics*, see e.g. [18], and *path coupling*, see e.g. [3], techniques. We wish to divide the vertex set V of the graph G into disjoint blocks $V_1, \ldots, V_K$ with the following properties:

- There is at most one edge between any pair of blocks.
- For each block V_i and any boundary conditions outside the block, the relaxation time of the dynamics restricted to V_i is polynomial in n.
- If we consider the block dynamics, where we pick a block V_i uniformly at random and update it containing it according to the conditional probability on $V \setminus V_i$, then it has the following property: Given two configurations σ and τ that differ at one vertex v, the updated configurations σ' and τ' may be coupled is such a way that the expected number of differences between them is $1 - \Theta(1/n)$.

The properties above imply a polynomial mixing time for the single site Gibbs-sampling dynamics.

Block decomposition: first attempt. The main task is therefore to show that such a decomposition into blocks exists when (4) holds and q is large enough. A key concept in the construction of the blocks is the notion of *good* vertices. Roughly speaking the blocks are constructed in such a way that the boundary of each block consists of good vertices only.

Good vertices v are vertices that are of degree bounded by c and such that

$$\sum_{u \in G: u \neq v} \alpha^{d(u,v)} \leq \varepsilon. \tag{7}$$

Assume for a moment that all blocks constructed are trees. In this case (7) implies that for a large enough q and given two boundary conditions that differ at one site, it is possible to couple the configurations inside the block with expected hamming distance ε, which implies the desired contraction of the block dynamics. Moreover, in the case where all the blocks are trees, we show that the second condition in (4) together with the small effect of the boundary implies a polynomial relaxation time of the dynamics inside the block.

Cyclic components and skeletons. More work is needed since we may not assume that all blocks are trees. In fact, a crucial step of the construction is to show that there are components $W_1, \ldots, W_r$ that contain all cycles of length $O(\log n)$ and such that all degrees in W_i are bounded, the number of vertices in W_i is $O(\log n)$ and the distance between W_i and W_j is $\Omega(\log n)$. All of the properties above follow from the assumption on the tree excess. We call the components W_i the *skeletons*.

Given the skeletons W_i, we consider two types of blocks: tree blocks and the blocks consisting of W_i and trees attaching to W_i. Using (4) we show that the mixing time of each block is polynomial in n and that the effect of the boundary on each block is small. This allows us to deduce a polynomial mixing time bound.

2 Proof of Theorems 1.1, 1.3 and 1.5

Proof (Theorem 1.1,1.3,1.5) The proofs follow by Lemma 2.1 below and Theorems 1.2, 1.4 and 1.6 respectively. □

Lemma 2.1 *For every $d \geq 1$ there exist $0 < a, \alpha, t, \delta < \infty$ such if G is a random graph distributed according to $G(n, d/n)$ then with high probability $m_\alpha(G, a \log n) \leq \delta \log n$ and for all $v \in V$, $t(v, a \log n) \leq t$.*

Proof It is well known that $G(n, d/n)$ satisfies $t(v, 2a \log n) \leq 1$ for all v with high probability, provided that $a = a(d) > 0$ is sufficiently small, see, e.g. [21]. Next we show that if α is sufficiently small then with high probability for all v_0 and all Γ, a self-avoiding path of length $a \log n$ starting at the vertex v_0, it holds that

$$\sum(\Gamma) := \sum_{u \in \Gamma} \sum_{v: u \neq v \in G} \alpha^{d(u,v)} \leq \delta \log n.$$

Considering the contribution to the sum from $u \notin B(v, 2a \log n)$ we see that

$$\sum(\Gamma) \leq (a \log n) \times n \times \alpha^{a \log n} + \sum_{u \in \Gamma} \sum_{v: u \neq v \in B(v_0, 2a \log n)} \alpha^{d(u,v)}.$$

Note that $(a \log n) \times n \times \alpha^{a \log n} = o(1)$ if $\alpha > 0$ is small enough so that $a \log \alpha + 1 < 0$. In order to bound the first sum we note that

$$\begin{aligned} &\sum_{u \in \Gamma} \sum_{v: u \neq v \in B(v_0, 2a \log n)} \alpha^{d(u,v)} \\ &\quad \leq \sum_{D=1}^{2a \log n} \alpha^D \sum_{v \in B(v_0, 2a \log n)} |\{u \in \Gamma : d(v, u) = D\}|. \end{aligned}$$

Note that for each $v \in B(v_0, 2a \log n)$ the size of the set $\{u \in \Gamma : d(v, u) = D\}$ is at most 4. Indeed suppose that there are five elements $u_1, \ldots, u_5$ in this set. For u_i denote by u'_i the last point on Γ on a shortest path from u_i to v and w_i be the following point. Since Γ is a path it follows that the size of the set $\{u'_i : 1 \leq i \leq 5\}$ is at least 3. Without loss of generality assume that u'_1, u'_2 and u'_3 are distinct. Then removing the edges (u'_1, w_1) and (u'_2, w_2) will maintain the connectivity properties of $B(v_0, 2a \log n)$ contradicting the fact that $t(v_0, 2a \log n) \leq 1$. The argument above implies that

$$\sum_{D=1}^{2a\log n} \alpha^D \sum_{v\in B(v_0,2a\log n)} |\{u \in \Gamma : d(v,u) = D\}|$$
$$\leq 4 \sum_{D=1}^{2a\log n} \alpha^D |\{v \in B(v_0, 2a\log n) : d(v,\Gamma) \leq D\}|.$$

We now use the well known expansion bounds implying that in $G(n, d/n)$ with high probability all connected sets Γ of size at least $a\log n$ have at most $h^D|\Gamma|$ elements at distance at most D from Γ which allows to bound the last sum as

$$4a\log n \sum_{D=1}^{2a\log n} \alpha^D h^D \leq \frac{\delta}{2}\log n,$$

provided α is small enough. Finally, we recall the proof of the expansion bound. Note that it suffices to show that for all connected sets Γ of size at least $a\log n$, the number of elements at distance exactly 1 from the set is bounded by $(h-1)|\Gamma|$. By a first moment calculation, the probability that a set with more neighbors exists is bounded by:

$$\sum_{s=a\log n}^{n} \binom{n}{s} s! \left(\frac{d}{n}\right)^{s-1} P[Bin(s(n-s), d/n) > (h-1)s]$$
$$\leq \sum_{s=a\log n}^{n} nd^{s-1} P[Bin(sn, d/n) > (h-1)s] = o(1),$$

provided h is large enough since by standard large deviation results,

$$\begin{aligned} P[Bin(sn, d/n) > (h-1)s] &\leq E\exp(Bin(sn, d/n) - (h-1)s) \\ &= \left(1 + \frac{d(e-1)}{n}\right)^{sn} \exp(-(h-1)s) \\ &\leq \exp\left(s[d(e-1) - (h-1)]\right). \end{aligned}$$

□

3 Notation and standard background

3.1 Notation

Definition 3.1 Let ∂U denote the *interior boundary* of U:

$$\partial U = \{u \in U : \exists u' \in U^c \text{ s.t. } (u', u) \in E\}.$$

Let $\partial^+ U$ denote the *exterior boundary* of U:

$$\partial^+ U = \{u \in U^c : \exists u' \in U \text{ s.t. } (u', u) \in E\}$$

For $U \subseteq W \subseteq V$ denote the *exterior boundary of W with respect to U*:

$$\partial^+_W U = \{u \in W^c : \exists u' \in U \text{ s.t. } (u', u) \in E\}.$$

If T is a tree rooted at ρ and $u \in T$ then we let T_u denote the subtree of u and all its descendants. Let T_u^+ denote $T_u \cup \partial^+_T T_u$.

Definition 3.2 Define the *α-weight* of a vertex v by $\varphi_\alpha(v) = \sum_{u \neq v} \alpha^{d(v,u)}$. We call v a (c, α, ϵ)-*good* vertex if the degree of v is less than or equal to c and $\varphi_\alpha(v) \leq \epsilon$. If v is not a (c, α, ϵ)-good vertex then it is a (c, α, ϵ)-*bad* vertex. When there is no ambiguity in the parameters (c, α, ϵ) we will simply call vertices good or bad vertices.

3.2 Relaxation and mixing times

Although not necessary for our results, to make use of existing theory it is convenient to make the assumption that the Gibbs sampling is lazy, that is we introduce self-loop probability of a half for all states. 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 $\min\{1 - \lambda_2, 1 - |\lambda_m|\}$ 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

$$\tau = \sup \left\{ \frac{2 \sum_\sigma P(\sigma)(f(\sigma))^2}{\sum_{\sigma \neq \tau} P(\sigma, \tau)(f(\sigma) - f(\tau))^2} \right\} \tag{8}$$

where f is any function on configurations, $P(\sigma, \tau) = P(\sigma) P(\sigma \to \tau)$ 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

$$\tau \leq \tau_{\text{mix}} \leq \tau \left(1 + \frac{1}{2} \log(\min_\sigma P(\sigma))^{-1}\right) \tag{9}$$

where τ_{mix} is the mixing time (see e.g. [1]). In all our examples we have $\log(\min_\sigma P(\sigma))^{-1} = \text{poly}(n)$ 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 $V_1, V_2, \ldots, V_k \subset V$, not necessarily disjoint, 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 $G - \{V_i\}$. In the continuous time version the blocks are updated according to independent rate 1 Poisson clocks. The relaxation time of the continuous time Gibbs sampler can be given in terms of the relaxation time of the continuous time block dynamics and the relaxation times of the continuous time Gibbs sampler on the blocks.

Proposition 3.1 *In the continuous time dynamics 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* [18]

$$\tau \leq \tau_{\text{block}} (\max_i \tau_i) \max_{v \in V} \{\# j : v \in V_j\}. \tag{10}$$

3.3 Canonical paths and conductance

We will use the following conductance result which follows from Cheeger's inequality, see e.g. [15].

Proposition 3.2 *Consider an ergodic reversible Markov chain* X_i *on a discrete space* Ω *where for any two states* $a, b \in \Omega$ *such that* $P(a, b) := P(a)P(a \to b) > 0$ *it holds that* $P(a, b) > \varepsilon$. *Then*

$$\tau_{\text{mix}} \leq 2/\varepsilon^2.$$

We also make use of the method of canonical paths.

Proposition 3.3 *Suppose that for any two states* σ, η *in the state space we have a canonical path* $\gamma_{(\sigma,\eta)} = (\sigma = \sigma^{(0)}, \sigma^{(1)}, \ldots, \sigma^{(k)} = \eta)$ *such that each transitions satisfies* $P(\sigma^{(i)}, \sigma^{(i+1)}) > 0$. *Let L be the length of the longest canonical path between two states and let*

$$\rho = \sup_{(\eta', \eta'')} \sum_{(\sigma,\eta):(\eta',\eta'') \in \gamma_{(\sigma,\eta)}} \frac{P(\sigma)P(\eta)}{P(\eta', \eta'')}$$

where the supremum is over pairs of states η', η'' *with* $P(\eta', \eta'') > 0$ *while the sum is over all pairs of states. Then the relaxation time satisfies*

$$\tau \leq L\rho.$$

3.4 Path coupling

We use the path coupling technique [3] to bound the relaxation time. The proposition below follows from [3] and [4], see also [2]. For two configurations $\sigma, \sigma' \in \mathcal{C}^V$ we denote their *Hamming distance* by $d_H(\sigma, \sigma') = |\{v : \sigma(v) \neq \sigma'(v)\}|$.

Proposition 3.4 *Consider Gibbs sampling on a graph G. Suppose that for any pair of configurations σ_1, σ_2 that differ in one site only, there is a way to couple the dynamics such that if σ_1' and σ_2' denote the configuration after the update then*:

$$E[d_H(\sigma_1', \sigma_2')] \leq 1 - \frac{c}{n}.$$

Then

$$\tau_{\text{mix}} \leq c.$$

4 Block mixing

For the proof we will consider block dynamics where the blocks are in some sense weakly connected. We will bound the relaxation time of the block dynamics in terms of single site dynamics of the sites connecting the blocks as follows. The following Lemma is a generalization of Claim 2.9 of [2] and may be proved in a similar manner. Here we provide an independent proof.

Lemma 4.1 *Let P be any Gibbs measure taking values in $\mathcal{C}$. Let $U \subset V$ and fix some boundary condition η on $\partial^+ U$. Suppose that U is the disjoint union of subsets U_i. Further suppose that for all i there exist $w_i \in U_i$ such that there are no edges between $U - U_i$ and $U_i - \{w_i\}$. Let $W = \cup_i \{w_i\}$. Let $B_i = \partial_U^+ U_i$ and let*

$$p_{w_i}(x) = P_{U_i \cup B_i}(\sigma(w_i) = x | \sigma(B_i) = \eta(B_i)). \tag{11}$$

We define the distribution Q on $\mathcal{C}^W$ by

$$Q(\sigma(W)) = \frac{1}{Z} \widehat{P}_W(\sigma(W)) \prod_i p_{w_i}(\sigma(w_i)) \tag{12}$$

where $\widehat{P}$ is the activity free distribution from Definition 1.1. Then the continuous time relaxation time τ_Q of the Gibbs sampler for Q and the continuous time relaxation time of the block dynamics τ_{block} satisfies $\tau_{\text{block}} \leq \max(1, \tau_Q)$.

Proof Let P^η denote the probability measure on U with boundary conditions η. Then by the Markov property and (12) it follows that $P_W^\eta = Q$. We note furthermore that from the Markov property it follows that the measure P^η satisfies for any i:

$$\begin{aligned} &P^\eta(\sigma(B_i) = \sigma' | \sigma(U \setminus B_i) = \sigma'') \\ &\quad = Q(\sigma(w_i) = \sigma'(w_i) | \sigma(W \setminus \{w_i\}) = \sigma''(W \setminus \{w_i\})) \\ &\qquad \times P^\eta(\sigma(B_i \setminus \{w_i\}) = \sigma'(B_i \setminus \{w_i\}) | \sigma(w_i) = \sigma'(w_i)). \end{aligned} \tag{13}$$

Write σ_t for the state of the block dynamics with blocks B_i and boundary conditions η. Write σ_t' for the state of the single site dynamics for (12). Then assuming that we have $\sigma_0(W) = \sigma'_0$ we obtain by equation (13) that the dynamics on σ and σ' may be coupled in such a way that for all t:

- $\sigma_t(W) = \sigma'_t$.
- If all the blocks (sites) in $\sigma_t(\sigma')$ have been updated at least once then:

$$P(\sigma_t = \sigma^* | \sigma_t(W) = \sigma^{**}) = P^\eta(\sigma = \sigma^* | \sigma(W) = \sigma^{**}).$$

Note that the probability that at least one block has not been updated by time t is at most $|W|e^{-t}$. Let P^t denote the distribution of σ_t and similarly Q^t. Given an optimal coupling between Q^t and Q consider the coupling of P^t to P where given two configurations (σ'_1, σ'_2) distributed according to the coupling, we let σ_1 be distributed according to the conditional distribution given σ'_1 and similarly for σ_2. Moreover by the argument above it follows that we may define σ_1 and σ_2 is such a way that if $\sigma'_1(W) = \sigma'_2(W)$ and all blocks have been updated then $\sigma_1 = \sigma_2$. This implies that

$$d_{TV}(P^t, P^\eta) \leq d_{TV}(Q^t, Q) + |W|e^{-t}.$$

Since the relaxation time measures the exponential rate of convergence to the distribution we conclude that $\tau_{\text{block}} \leq \max(1, \tau_Q)$. □

Our bounds on the relaxations times of trees will be given in terms of their path density defined as follow:

Definition 4.1 For a tree $T \subset G$ rooted at ρ we let the *maximal path density* be defined by

$$m(T, \rho) = \max_\Gamma \sum_{u \in \Gamma} \deg(u)$$

where the maximum is taken over all self-avoiding paths $\Gamma \subset T$ starting at ρ.

4.1 Colouring model

Next we prove two lemmas which will be used together with Lemma 4.1 to prove relaxation bounds below.

Lemma 4.2 *Let W be a star with center v and k leaves. Let*

$$Q(\sigma(W)) = \frac{1}{Z} P_W(\sigma(W)) \prod_{w \in W} p_w(\sigma(w))$$

where the p_w are functions such that for all $w \in W$, $\sum_{x \in \mathcal{C}} p_w(x) = 1$ and for all $w \in W, x \in \mathcal{C}$ either $p_w(x) > (q\delta)^{-1}$ or $p_w(x) = 0$. Further assume that for some $c \leq q - 3$ we have that for all $w \in W - v$, $\#\{x \in \mathcal{C} : p_w(x) = 0\} \leq c$. Then the relaxation time τ of the Gibbs sampler on Q is at most C^k where C is a constant depending only on c, δ, q.

Proof We first show that the chain is ergodic by constructing a path between any two configurations σ and η with $Q(\sigma)$ and $Q(\eta) > 0$. Since for each leaf w there are at least 3 colours x with $p_w(x) > 0$ we can find a colour $x(w)$ such that $p_w(x(w)) > 0$ and $\sigma(v) \neq x(w) \neq \eta(v)$. The path is constructed by changing the states of the leaves to $x(u)$, then changing the state of v to $\eta(v)$, then finally changing the states of the leaves to $\eta(u)$. Now by the hypothesis there are at most q^{k+1} colourings of W so $Z \leq q^{k+1}$ so we have that $Q(\sigma), Q(\eta) > (q^2\delta)^{-(k+1)}$. For two adjacent states σ and σ' with $Q(\sigma), Q(\sigma') > 0$, we have $Q(\sigma \to \sigma') \geq (q\delta(k+1))^{-1}$ and so $Q(\sigma, \sigma') \geq (q^2\delta)^{-(k+1)}(q\delta(k+1))^{-1}$. From Proposition 3.2 it now follows that

$$\tau_2 \leq ((q\delta(k+1))^2(q^2\delta)^{k+1})^4 \leq 4^k q^{20k}\delta^{20k},$$

as needed. □

Similarly, it is easy to see the following lemma.

Lemma 4.3 *Let W be a graph with k vertices of maximum degree d. Let*

$$Q(\sigma(W)) = \frac{1}{Z} P_W(\sigma(W)) \prod_{w \in W} p_w(\sigma(w_i))$$

where the p_w are functions such that for all $w \in W$, $\sum_{x \in \mathcal{C}} p_w(x) = 1$ and for all $w \in W$, $x \in \mathcal{C}$ either $p_w(x) > (q\delta)^{-1}$ or $p_w(x) = 0$. Further, for some $c \leq q - d - 2$ we have that for all $w \in W$, $\#\{x \in \mathcal{C} : p_w(x) = 0\} \leq c$. Then the relaxation time of the Gibbs sampler on Q is at most C^k where C is a constant depending only on c, δ, d and q.

We can now obtain polynomial mixing time results for the treelike blocks that will be used in the construction.

Theorem 4.1 *Let $T \subseteq U \subset V$ such that T is a tree rooted at ρ and so that there are no edges between $T - \{\rho\}$ and $U - T$. Suppose that for all $u \in T$, $\#\{v \in V - U : (v, u) \in E\} < c$ and that for each $u \in T$,*

$$\sup_{\sigma(\partial_U^+ T_u)} \sup_{x \in \mathcal{C}} \sup_{y \in \mathcal{C}: P_{T_u^+}(\sigma(u)=y|\sigma(\partial_U^+ T_u)) \neq 0} \frac{P_{T_u^+}(\sigma(u) = x|\sigma(\partial^+ T))}{P_{T_u^+}(\sigma(u) = y|\sigma(\partial^+ T))} \leq \delta \tag{14}$$

For some $l \geq 1$ assume there are at most l edges between $\{\rho\}$ and $U - T$. Let τ be the relaxation time of the continuous time Gibbs sampler on T. If $q \geq c + l + 2$ then for any boundary condition η on $\partial^+ T$ we have that $\tau \leq C^{m(T,\rho)}$ where $m(T, \rho)$ is the maximal path density on T and where C is a constant depending only on c, δ, q and l.

Proof We proceed by induction on $m(T, \rho)$. If T is a single point then $\tau = 1$ and so $\tau \leq C^{m(T,\rho)}$. Now suppose ρ has children $u_1, \ldots, u_k \in T$. By induction the relaxation time of the Gibbs sampler on T_{u_i}, $\tau_i \leq C^{m(T_{u_i}, u_i)}$ and by the definition of the maximal path density $m(T_{u_i}, u_i) \leq m(T, \rho) - k$. Let τ_{block} denote the block dynamics on T with blocks $\{\{\rho\}, T_{u_1}, \ldots, T_{u_k}\}$. Applying Lemma 4.1 and 4.2 we get that the block dynamics satisfies $\tau_{\text{block}} \leq C^k$. Then by Proposition 3.1 we have that

$$\tau \leq \tau_{\text{block}} \max_i \{1, \tau_i\} \leq C^k C^{m(T,\rho)-k} \leq C^{m(T,\rho)}$$

which completes the result. □

Lemma 4.3 above will be used for the analysis of blocks in the construction that do contain cycles.

4.2 Hardcore model

Lemma 4.4 *Let W be a graph and let*

$$Q(\sigma(W)) = \frac{1}{Z} \widehat{P}_W(\sigma(W)) \prod_{w \in W} p_w(\sigma(w_i))$$

where the p_w are functions such that for some δ and all $w \in W$, $\delta < p_w(0) < 1$ and $p_w(0) + p_w(1) = 1$. Then the relaxation time τ of the Gibbs sampler of Q satisfies $\tau \leq C^{|W|}$ where C depends only on β and δ.

Proof We use the method of canonical paths from Proposition 3.3. Let σ and η be two configurations with $Q(\sigma)$ and $Q(\eta) > 0$. We define the canonical path to be a path which begins from σ, then sequentially changes states of all the vertices to 0 and then sequentially changes the state of $w \in W$ to 1 if $\eta(w) = 1$. Clearly each path is of length at most $2|W|$. Now suppose η', η'' is a step in some path. They must differ at exactly one site $w \in W$ and suppose that $\eta'(w) = 1$ and $\eta''(w) = 0$. If (η', η'') is in the canonical path $\gamma_{(\sigma,\eta)}$ then $\sigma \geq \eta'$ under the canonical partial ordering. Now $P[\eta' \to \eta''] = \frac{p_w(0)}{|W|} \geq \frac{\delta}{|W|}$. Then

$$\begin{aligned}
\sum_{(\sigma,\eta):(\eta',\eta'')\in\gamma_{(\sigma,\eta)}} \frac{P(\sigma)P(\eta)}{P(\eta',\eta'')} &\leq \sum_{\sigma:\sigma\geq\eta'} \frac{P(\sigma)}{P(\eta',\eta'')} \\
&= P[\eta' \to \eta'']^{-1} \sum_{\sigma:\sigma\geq\eta'} \frac{\exp(\beta \sum_u \sigma(u)) \prod_u p_w(\sigma(u))}{\exp(\beta \sum_u \eta'(u)) \prod_u p_w(\eta'(u))} \\
&\leq \frac{|W|}{\delta} (1 + \exp(\max(\beta, 0))\delta^{-1})^{|W|}.
\end{aligned}$$

Similarly the same bound holds for pairs with $\eta'(w) = 0$ and $\eta''(w) = 1$ so $\rho \leq \frac{|W|}{\delta}(1 + \exp(\max(\beta, 0))\delta^{-1})^{|W|}$. From Proposition 3.3 it now follows that

$$\tau_2 \leq \frac{2|W|^2}{\delta}(1 + \exp(\max(\beta, 0))\delta^{-1})^{|W|} \leq 10^{|W|} \exp(\max(\beta, 0)|W|)\delta^{-|W|},$$

as needed. □

Theorem 4.2 *Let $T \subset V$ be a tree rooted at ρ. Then for the continuous time dynamics $\tau \leq C^{m(T,\rho)}$ where $m(T, \rho)$ is the maximal path density on T and where C is a constant depending only on β.*

Proof We proceed by induction on $m(T, \rho)$. If T is a single point then $\tau = 1$ and so $\tau \leq C^{m(T,\rho)}$. Now suppose ρ has children $u_1, \ldots, u_k \in T$. By induction the relaxation time of the Gibbs sampler on T_{u_i} satisfies $\tau_i \leq C^{m(T_{u_i},u_i)}$. By definition of the maximal path density $m(T_{u_i}, u_i) \leq m(T, \rho) - k$. Let τ_{block} denote the block dynamics on T with blocks $\{\{\rho\}, T_{u_1}, \ldots, T_{u_k}\}$. We define $W = \{\rho, u_1, \ldots, u_k\}$ and the distribution Q on $\mathcal{C}^W$ by

$$Q(\sigma(W)) = \frac{1}{Z}\widehat{P}_W(\sigma(W)) \prod_{w \in W} p_{w_i}(\sigma(w_i))$$

and p_{w_i} is as in equation (11). Applying Lemma 4.1 with $W = \{\rho, u_1, \ldots, u_k\}$ implies that $\tau_{\text{block}} = \tau_Q$ where τ_Q is the relaxation time of the Gibbs sampler on the measure Q. In the hardcore model for any vertex v and any boundary condition $\sigma(V - \{v\})$ on $V - \{v\}$ we have that $P(\sigma(v) = 0|\sigma(V - \{v\})) \geq \frac{1}{1+e^{\beta}}$, the probability that the spin at v is 0 given that the spins of all its neighbors are 0, and so each $p_w(0) \geq \frac{1}{1+e^{\beta}}$. It follows that in Lemma 4.4 we can take $\delta = \frac{1}{1+e^{\beta}}$ and so $\tau_{\text{block}} \leq C^k$ for sufficiently large C. Then by Proposition 3.1 we have that

$$\tau \leq \tau_{\text{block}} \max_i \{1, \tau_i\} \leq C^k C^{m(T,\rho)-k} \leq C^{m(T,\rho)}$$

which completes the result. □

4.3 Soft constraint models

For soft constraint models, bounding the mixing time is simplified by the fact that removing an edge adds at most a constant multiplicative factor to the relaxation time.

Theorem 4.3 *Let τ be the relaxation time of the continuous time Gibbs sampler on a tree $T \subset V$. Given arbitrary boundary conditions,*

$$\tau \leq \exp(4\|H\|m(T))$$

where $\|H\|$ is the norm of the Hamiltonian.

Proof We proceed by induction on m with a similar argument to the one used in [21] for the Ising model. 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 model on T' and let τ' be the relaxation time on T'. By equation (8) we have that

$$\tau/\tau' \leq \frac{\max_\sigma P(\sigma)/P'(\sigma)}{\min_{\sigma,\tau} P(\sigma, \tau)/P'(\sigma, \tau)} \leq \exp(4\|H\|k). \tag{15}$$

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 [18] we get that the relaxation time on T' is simply the maximum of the relaxation times on the blocks,

$$\tau' \leq \max\{1, \tau^i\}.$$

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 $m(T^i) \leq m - k$, and therefore for all i it holds that $\tau^i \leq \exp(4\|H\|(m-k))$. This then implies by (15) that $\tau \leq \exp(4\|H\|m)$ as needed. □

5 Correlation decay in tree blocks

In this subsection we prove that if we look at a tree block, all of whose leaves are good, then for large enough q we have the correlation decay property (14).

Definition 5.1 For $0 < \lambda < 1$ and $U \subset V$ define the block boundary weighting as the function defined by:

$$\psi_\lambda(v) = \psi(v) = \sum_{w \in \partial^+ U} \lambda^{d(w,v)},$$

for all $v \in U$.

Lemma 5.1 *If every vertex in $\partial^+ U$ is (c, α, ϵ)-good then for all $\lambda \leq \alpha^2$,*

$$\psi(v) \leq \frac{\epsilon\lambda}{\alpha^2}$$

Proof Let $v \in U$ and let $u \in \partial^+ U$ be an exterior boundary vertex which minimizes the distance to v. Then

$$\psi_{\alpha^2}(v) \leq \sum_{w \in \partial^+ U} \alpha^{(d(v,u)+d(u,w))} \leq \sum_{w \neq u} \alpha^{d(w,u)} = \varphi_\alpha(u) \leq \epsilon. \tag{16}$$

and the result follows since for $\lambda \leq \alpha^2$ we have $\psi_\lambda(v) \leq \frac{\lambda}{\alpha^2}\psi_{\alpha^2}(v)$. □

5.1 Colouring

Lemma 5.2 *Suppose that $T = (V_T, E_T)$ is an induced subgraph of $G = (V, E)$ that is a tree and suppose that for all $v \in V_T$, $\psi(v) \leq 1$. Then there exists a q_λ such that for $q > q_\lambda$ and all $v \in V_T$:*

$$\sup_{\sigma(\partial^+ T)} \sup_{x \in \mathcal{C}} \sup_{y \in \mathcal{C}: P(\sigma(v)=y|\sigma(\partial^+ T)) \neq 0} \frac{P(\sigma(v) = x|\sigma(\partial^+ T))}{P(\sigma(v) = y|\sigma(\partial^+ T))} \leq \exp(\psi(v)) \tag{17}$$

where the supremum is over all boundary conditions $\sigma(\partial^+ T)$ on $\partial^+ T$.

Proof Fix v as the root of the tree. We will prove the result by induction on the size of the tree. When the tree consists of a single vertex v the quantity in the left hand side of (17) is clearly 1.

Let $u_1, \ldots, u_l$ be the children of v in T. Consider the graph $G' = (V', E')$ obtained from G by removing the vertex v and all adjacent edges. Let

$$\delta_i = \sup_{\sigma(\partial_T^+ T_{u_i})} \sup_{x \in \mathcal{C}} \sup_{y \in \mathcal{C}: P_{T_{u_i}^+}(\sigma(u_i)=y|\sigma(\partial_T^+ T_{u_i})) \neq 0} \frac{P_{T_{u_i}^+}(\sigma(u_i) = x|\sigma(\partial_T^+ T_{u_i}))}{P_{T_{u_i}^+}(\sigma(u_i) = y|\sigma(\partial_T^+ T_{u_i}))} \tag{18}$$

For $w' \in T_{u_i}$ write $\psi_i(w') = \sum_{w \in \partial_T^+ T_{u_i}} \lambda^{d(w,w')}$. Note that ψ_i is the function ψ for the subtree T_{u_i} in the graph G'. Note moreover that for all w we have $\psi_i(w) \leq \psi(w)$. By the induction hypothesis we therefore have $\delta_i \leq \exp(\psi_i(u_i))$. Let $d_i = \#\{w \in V' \setminus T_{u_i} : (w, u_i) \in E\}$ and note that there are at least $q - d_i$ elements $y \in \mathcal{C}$ with $P_{T_{u_i}^+}(\sigma(v) = y|\sigma(\partial^+ T_{u_i})) > 0$ so

$$\min_y\{P_{T_{u_i}^+}(\sigma(v) = y|\sigma(\partial^+ T_{u_i})) : P_{T_{u_i}^+}(\sigma(v) = y|\sigma(\partial^+ T_{u_i})) > 0\} \leq \frac{1}{q - d_i}$$

and so by (18) we have

$$\max_y P_{T_{u_i}^+}(\sigma(v) = y|\sigma(\partial^+ T_{u_i})) \leq \frac{\delta_i}{q - d_i}. \tag{19}$$

Since $d_i\lambda \leq \psi_i(u_i) \leq 1$, taking $q > 2/\lambda$ yields $q - d_i > q/2$. When $0 \leq x \leq 1$ we have $e^x - 1 \leq 2x$ so $\delta_i - 1 \leq 2\psi(x)$. And since $\frac{x}{1-x}$ is increasing in x

$$\begin{aligned}
&\sup \frac{1 - P_{T_{u_i}^+}(\sigma(v) = x|\sigma(\partial^+ T_{u_i}))}{1 - P_{T_{u_i}^+}(\sigma(v) = y|\sigma(\partial^+ T_{u_i}))} \\
&= 1 + \sup \frac{P_{T_{u_i}^+}(\sigma(v) = y|\sigma(\partial^+ T_{u_i})) - P_{T_{u_i}^+}(\sigma(v) = x|\sigma(\partial^+ T_{u_i}))}{1 - P_{T_{u_i}^+}(\sigma(v) = y|\sigma(\partial^+ T_{u_i}))} \\
&\leq 1 + \frac{\frac{\delta_i - 1_{\{d_i=0\}}}{q-d_i}}{1 - \frac{\delta_i}{q-d_i}} \text{ (By (19) and since } \tfrac{x}{1-x} \text{ is increasing)} \\
&= 1 + \frac{\delta_i - 1_{\{d_i=0\}}}{q - d_i - \delta_i} \\
&\leq 1 + \frac{\delta_i - 1_{\{d_i=0\}}}{q/2 - e} \text{ (since } \delta_i \leq e \text{ and } q - d_i > q/2) \\
&\leq 1 + \frac{4(\delta_i - 1_{\{d_i=0\}})}{q} \text{ (taking } q \geq 4e)
\end{aligned}$$

$$\leq 1 + \frac{8\psi_i(u_i) + 4d_i}{q} \text{ (since } \delta_i - 1 \leq 2\psi(x))$$
$$\leq \exp\left(\frac{8\psi_i(u_i) + 4d_i}{q}\right)$$

where the supremum is taken over all $x, y \in \mathcal{C}$ and boundary conditions on $\partial^+ T_u$. Now note $\psi(v) \geq \lambda \sum_i \psi_i(u_i)$ (it may be strictly greater due to the contribution of the neighbors of v outside T). Therefore:

$$\sup_{\sigma(\partial^+ T)} \sup_{x\in\mathcal{C}} \sup_{y\in\mathcal{C}:P(\sigma(v)=y|\sigma(\partial^+T))\neq 0} \frac{P(\sigma(v) = x|\sigma(\partial^+ T))}{P(\sigma(v) = y|\sigma(\partial^+ T))}$$
$$= \prod_i \sup \frac{1 - P_{T_{u_i}}(\sigma(v) = x|\sigma(\partial^+ T_{u_i}))}{1 - P_{T_{u_i}}(\sigma(v) = y|\sigma(\partial^+ T_{u_i}))}$$
$$\leq \exp\left(\frac{8\psi_i(u_i) + 4d_i}{q}\right)$$
$$\leq \exp\left(\left[\frac{8}{q\lambda} + \frac{4}{q\lambda^2}\right]\psi(v)\right)$$

which completes the induction provided that q is large enough so that $q \geq \max\left(4e, \frac{8}{\lambda} + \frac{4}{\lambda^2}\right)$. □

The following corollary follows immediately from Lemma 5.2 and Lemma 5.1.

Corollary 5.1 *For all $c, \alpha > 0$ and $\varepsilon > 0$ there exists a q for which the following holds. Let $T \subset V$ be a tree such that every vertex in $\partial^+ T$ is (c, α, ϵ)-good. Then for any $0 < \lambda < 1$ there exists a q_λ such that for $q > q_\lambda$,*

$$\sup_{\sigma(\partial^+ T)} \sup_{x\in\mathcal{C}} \sup_{y\in\mathcal{C}:P(\sigma(v)=y|\sigma(\partial^+T))\neq 0} \frac{P(\sigma(v) = x|\sigma(\partial^+ T))}{P(\sigma(v) = y|\sigma(\partial^+ T))} \leq \exp\left(\sum_{w\in\partial^+ T} \lambda^{d(w,v)}\right)$$

where the supremum is over all boundary conditions $\sigma(\partial^+ U)$ on $\partial^+ U$.

5.2 Hardcore model

Lemma 5.3 *Suppose that $T = (V_T, E_T)$ is an induced subgraph of $G = (V, E)$ that is a tree. For $v \in V_T$ and η a boundary condition on $\partial^+ T$ let P^η denote the measure $P(\sigma(v) = \cdot|\sigma(\partial^+ U))$. Then if $\beta_\lambda = \log \lambda$ then for all $\beta < \beta_\lambda$ and $v \in V_T$:*

$$d_{TV}(P^{\eta^1}, P^{\eta^2}) \leq \psi_\lambda(v) \quad (20)$$

for any two boundary conditions η^1 and η^2 on $\partial^+ T$ where d_{TV} is the total variation distance.

Proof Since the left hand side of equation (20) is bounded by 1 we can assume that $\psi(v) \leq 1$. Fix v as the root of the tree. We will prove the result by induction on the size of the tree. Let $u_1, \ldots, u_l$ be the children of v in U and let $w_1, \ldots, w_m$ be the children of v in $\partial^+ T$. Consider the graph $G' = (V', E')$ obtained from G by removing the vertex v and all adjacent edges and let $P^{\eta}_{T_{u_i}}$ denote $P'(\sigma(u_i) = \cdot|\eta)$. Then

$$d_{TV}(P^{\eta^1}, P^{\eta^2}) = \left|P(\sigma(v) = 0|\eta^1) - P(\sigma(v) = 0|\eta^2)\right|$$

$$= \left|\frac{1}{1 + e^{\beta} \prod_{i=1}^{l} P^{\eta^1}_{T_{u_i}}(0) \prod_{i=1}^{m} 1_{\{\eta^1_{w_i}=0\}}} - \frac{1}{1 + e^{\beta} \prod_{i=1}^{l} P^{\eta^2}_{T_{u_i}}(0) \prod_{i=1}^{m} 1_{\{\eta^2_{w_i}=0\}}}\right|$$

$$\leq e^{\beta} \left|\prod_{i=1}^{l} P^{\eta^1}_{T_{u_i}}(0) \prod_{i=1}^{m} 1_{\{\eta^1_{w_i}=0\}} - \prod_{i=1}^{l} P^{\eta^2}_{T_{u_i}}(0) \prod_{i=1}^{m} 1_{\{\eta^2_{w_i}=0\}}\right|$$

$$\leq \begin{cases} \lambda & m \geq 1 \\ e^{\beta} \left|\prod_{i=1}^{l} P^{\eta^1}_{T_{u_i}}(0) - \prod_{i=1}^{l} P^{\eta^2}_{T_{u_i}}(0)\right| & m = 0 \end{cases} \tag{21}$$

Now if $m \geq 1$ then $\psi(v) \geq \lambda$ so $d_{TV}(P^{\eta^1}, P^{\eta^2}) \leq \psi(v)$. This establishes equation (20) for trees of size 1. We now proceed by induction.

Observe the simple inequality that if $0 \leq x_1, \ldots, x_q \leq 1$ and $0 \leq y_1, \ldots, y_q \leq 1$ then

$$\left|\prod_{l=1}^{q} x_l - \prod_{l=1}^{q} y_l\right| = \left|\sum_{j=1}^{q} (x_j - y_j) \prod_{l=1}^{j-1} x_l \prod_{l=j+1}^{q} y_l\right|$$

$$\leq \sum_{j=1}^{q} \left|x_j - y_j\right|. \tag{22}$$

Applying equation (22) to equation (21) we get that when $m = 0$,

$$d_{TV}(P^{\eta^1}, P^{\eta^2}) \leq e^{\beta} \sum_{i=1}^{l} \left|P^{\eta^1}_{T_{u_i}}(0) - P^{\eta^2}_{T_{u_i}}(0)\right|.$$

By the inductive hypothesis applied to the tree T_{u_i}, we have that

$$|P^{\eta^1}_{T_{u_i}}(0) - P^{\eta^2}_{T_{u_i}}(0)| \leq \sum_{w \in \partial^+ T_{u_i}} \lambda^{d(w,u_i)} = \frac{1}{\lambda} \sum_{w \in \partial^+ T_{u_i}} \lambda^{d(w,v)}$$

so

$$d_{TV}(P^{\eta^1}, P^{\eta^2}) \leq e^{\beta} \sum_{i=1}^{l} |P^{\eta^1}_{T_{u_i}}(0) - P^{\eta^2}_{T_{u_i}}(0)| \leq \psi(v)$$

which completes the induction. □

5.3 Soft constraint models

Lemma 5.4 *Suppose that $T = (V_T, E_T)$ is an induced subgraph of $G = (V, E)$ that is a tree. For $v \in V_T$ and η a boundary condition on $\partial^+ T$ let P^η denote the measure $P(\sigma(v) = \cdot|\sigma(\partial^+ U))$. Then there exists an $H^\lambda > 0$ depending only on λ such that if $\|H\| < H^\lambda$ and $v \in V_T$:*

$$d_{TV}(P^{\eta^1}, P^{\eta^2}) \leq \psi_\lambda(v) \tag{23}$$

for any two boundary conditions η^1 and η^2 on $\partial^+ T$ where d_{TV} is the total variation distance.

Proof Since the left hand side of equation (23) is bounded by 1 we can assume that $\psi(v) \leq 1$. Let $K = 4(e^{\|H\|} - e^{-\|H\|})$. We can take H^λ to be small enough so that $4K < \lambda$ and for $0 \leq x \leq 1/\lambda$ we have $\exp(-xK) \leq 1 - xK/2$ and $\exp(2Kx) \leq 1 + 4Kx$. Fix v as the root of the tree. We will prove the result by induction on the size of the tree. Let $u_1, \ldots, u_l$ be the children of v in U and let $u_{l+1}, \ldots, u_m$ be the children of v in $\partial^+ T$. Consider the graph $G' = (V', E')$ obtained from G by removing the vertex v and all adjacent edges, let P' denote the induced soft constraint model on G' and let $P^\eta_{T_{u_i}}$ denote $P'(\sigma(u_i) = \cdot|\eta)$. Then for all i and $z \in \mathcal{C}$,

$$\begin{aligned}
\frac{\sum_{y_i \in \mathcal{C}} e^{g(z,y_i)} P^{\eta^1}_{T_{u_i}}(y_i)}{\sum_{y_i \in \mathcal{C}} e^{g(z,y_i)} P^{\eta^2}_{T_{u_i}}(y_i)} &= 1 - \frac{\sum_{y_i \in \mathcal{C}} e^{g(z,y_i)} \left(P^{\eta^2}_{T_{u_i}}(y_i) - P^{\eta^1}_{T_{u_i}}(y_i)\right)}{\sum_{y_i \in \mathcal{C}} e^{g(z,y_i)} P^{\eta^2}_{T_{u_i}}(y_i)} \\
&\geq 1 - 2\left(e^{\|H\|} - e^{-\|H\|}\right) d_{TV}\left(P^{\eta^1}_{T_{u_i}}, P^{\eta^2}_{T_{u_i}}\right) \\
&\geq \exp\left(-K d_{TV}\left(P^{\eta^1}_{T_{u_i}}, P^{\eta^2}_{T_{u_i}}\right)\right)
\end{aligned}$$

Similarly we have

$$\frac{\sum_{y_i \in \mathcal{C}} e^{g(z,y_i)} P^{\eta^1}_{T_{u_i}}(y_i)}{\sum_{y_i \in \mathcal{C}} e^{g(z,y_i)} P^{\eta^2}_{T_{u_i}}(y_i)} \leq \exp\left(K d_{TV}\left(P^{\eta^1}_{T_{u_i}}, P^{\eta^2}_{T_{u_i}}\right)\right)$$

Then for each $x \in \mathcal{C}$,

$$\begin{aligned}
&\frac{P^{\eta^1}(v)(x)}{P^{\eta^2}(v)(x)} \\
&= \frac{e^{h(x)} \prod_{i=1}^m \sum_{y_i \in \mathcal{C}} e^{g(x,y_i)} P_{T_{u_i}}^{\eta^1}(y_i)}{\sum_{z \in \mathcal{C}} e^{h(z)} \prod_{i=1}^m \sum_{y_i \in \mathcal{C}} e^{g(z,y_i)} P_{T_{u_i}}^{\eta^1}(y_i)} \Big/ \frac{e^{h(x)} \prod_{i=1}^m \sum_{y_i \in \mathcal{C}} e^{g(x,y_i)} P_{T_{u_i}}^{\eta^2}(y_i)}{\sum_{z \in \mathcal{C}} e^{h(z)} \prod_{i=1}^m \sum_{y_i \in \mathcal{C}} e^{g(z,y_i)} P_{T_{u_i}}^{\eta^2}(y_i)} \\
&= \frac{e^{h(x)} \prod_{i=1}^m \sum_{y_i \in \mathcal{C}} e^{g(x,y_i)} P_{T_{u_i}}^{\eta^1}(y_i)}{e^{h(x)} \prod_{i=1}^m \sum_{y_i \in \mathcal{C}} e^{g(x,y_i)} P_{T_{u_i}}^{\eta^2}(y_i)} \Big/ \frac{\sum_{z \in \mathcal{C}} e^{h(z)} \prod_{i=1}^m \sum_{y_i \in \mathcal{C}} e^{g(z,y_i)} P_{T_{u_i}}^{\eta^1}(y_i)}{\sum_{z \in \mathcal{C}} e^{h(z)} \prod_{i=1}^m \sum_{y_i \in \mathcal{C}} e^{g(z,y_i)} P_{T_{u_i}}^{\eta^2}(y_i)} \\
&\leq \exp\left(2K \sum_{i=1}^m d_{TV}(P_{T_{u_i}}^{\eta^1}, P_{T_{u_i}}^{\eta^2})\right).
\end{aligned}$$

Then

$$\begin{aligned}
d_{TV}(P^{\eta^1}, P^{\eta^2}) &= \sum_{x \in \mathcal{C}} \left| P^{\eta^1}(x) - P^{\eta^2}(x) \right| \\
&= \sum_{x \in \mathcal{C}} P^{\eta^2}(x) \left| \frac{P^{\eta^1}(x)}{P^{\eta^2}(x)} - 1 \right| \\
&\leq \exp\left(2K \sum_{i=1}^m d_{TV}(P_{T_{u_i}}^{\eta^1}, P_{T_{u_i}}^{\eta^2})\right) - 1
\end{aligned}$$

Now suppose that T is a single vertex $\{v\}$ so $u_1, \ldots u_m$ are all in $\partial^+ T$ and so $\psi(v) = m\lambda$. If $m = 0$ then $d_{TV}(P^{\eta^1}, P^{\eta^2}) = \psi(v) = 0$. If $1 \leq m \leq 1/\lambda$ then

$$d_{TV}(P^{\eta^1}, P^{\eta^2}) \leq \exp(2Km) - 1 \leq 4Km \leq \lambda m = \psi(v)$$

while if $m > 1/\lambda$ then $\psi(v) > 1$. So this verifies the case when T is a single point. For the induction step our inductive hypothesis says that

$$d_{TV}\left(P_{T_{u_i}}^{\eta^1}, P_{T_{u_i}}^{\eta^2}\right) \leq \sum_{w \in \partial^+ T_{u_i}} \lambda^{d(w,u_i)} = \frac{1}{\lambda} \sum_{w \in \partial^+ T_{u_i}} \lambda^{d(w,v)}.$$

If $\psi(v) \leq 1$ then $\sum_{i=1}^m d_{TV}\left(P_{T_{u_i}}^{\eta^1}, P_{T_{u_i}}^{\eta^2}\right) \leq \frac{1}{\lambda}$ and so

$$d_{TV}(P^{\eta^1}, P^{\eta^2}) \leq \exp\left(2K \sum_{i=1}^m d_{TV}(P_{T_{u_i}}^{\eta^1}, P_{T_{u_i}}^{\eta^2})\right) - 1 \leq 4K d_{TV}\left(P_{T_{u_i}}^{\eta^1}, P_{T_{u_i}}^{\eta^2}\right) \leq \psi(v)$$

which completes the induction. □

6 Block construction

Lemma 6.1 *For two (c, α, ϵ)-bad points u, u' we define $u \sim u'$ if there is a path $u = u_1, u_2, \ldots, u_k = u'$ such that no two consecutive vertices on the path u_i, u_{i+1} are (c, α, ϵ)-good. Then $\sim$ is an equivalence relation of (c, α, ϵ)-bad vertices in G.*

Proof The relation is clearly reflexive and symmetric. Suppose that there is a path $u \sim u'$ and $u \sim u''$. Then there exist paths $u = v_1, v_2, \ldots, v_k = u'$ and $u = w_1, w_2, \ldots, w_l = u''$ such that no two consecutive vertices are (c, α, ϵ)-good. Let $i = \max(j : v_j \in \{w_1, w_2, \ldots, w_l\})$ and suppose that $v_i = w_j$. Then the path $u' = v_k, v_{k-1}, \ldots, v_i, w_{j+1}, w_{j+2}, \ldots, w_l = u''$ is a path with no two consecutive (c, α, ϵ)-good vertices so $u' \sim u''$. Hence $\sim$ is transitive and is an equivalence relation. □

We now describe our method for partitioning G into smaller blocks for some fixed (c, α, ϵ).

- Two (c, α, ϵ)-bad points u, u' are in the same block if and only if $u \sim u'$.
- A (c, α, ϵ)-good vertex is in the same block as any bad point it is adjacent to.
- A (c, α, ϵ)-good vertex not adjacent to any bad point forms a separate block

By Lemma 6.1 the first point defines a partition of the (c, α, ϵ)-bad vertices. If a good vertex v is adjacent to bad vertices u_1 and u_2 then u_1, v, u_2 has no two consecutive good points so $u_1 \sim u_2$ and hence good points are assigned to exactly one block. Hence this defines a partition of G into blocks whose boundaries are all (c, α, ϵ)-good. We will abuse notation and let $\sim$ denote the equivalence relation on all G for this partition.

Lemma 6.2 *Suppose that G satisfies equation (4). Then for any $0 < L < \infty$ there exists (c, α, ϵ) such that every self-avoiding path $u_1, u_2, \ldots, u_{L \log n}$ contains two consecutive (c, α, ϵ)-good vertices u_i, u_{i+1}.*

Proof We can assume that $L \le a$ and set $\epsilon = \frac{3\delta}{L}$. Then since $\sum_{i=1}^{L \log n} \varphi_\alpha(u_i) < \delta \log n$ at most $\frac{L}{3} \log n$ of the u_i have $\varphi_\alpha(u_i) \ge \epsilon$. If $c = \frac{\epsilon}{\alpha}$ and if $\varphi_\alpha(u_i) < \epsilon$ then

$$\deg(u_i) = \sum_{u:(u,u_i)\in E} \alpha^{d(u,u_i)-1} \le \frac{1}{\alpha} \varphi_\alpha(u_i) < c$$

so u_i is (c, α, ϵ)-good. Since the path $u_1, u_2, \ldots, u_{L \log n}$ contains at least $\frac{2}{3}L \log n (c, \alpha, \epsilon)$-good vertices it must contain two consecutive good vertices. □

The following corollary is immediate from the definition of the equivalence relation.

Corollary 6.1 *Suppose that G satisfies equation (4). Then for any $0 < L < \infty$ there exists (c, α, ϵ) such that if $u \sim v$ then $d(u, v) < L \log n$.*

Our next step is to define a partition of the graph into blocks whose boundaries are good vertices and such that each block is either a tree or a tree plus some bounded number of edges. The decomposition into blocks relies on the following combinatorial lemma.

Lemma 6.3 *Consider a graph $G = (V, E)$ where V is the disjoint union of V_G and V_B. Assume further that for all $v \in V$ it holds that $t(v, a \log n) \leq t$ and that every self avoiding path $u_1, \ldots, u_{L \log n}$ contains two consecutive elements in V_G, where $(20t + 2)L < a$. Then we can partition G into blocks $\{V_j\}$ such there is at most one edge between any two blocks. Moreover, for all j, the diameter of V_j is less than $(20t + 2)L \log n$, it holds that $\partial V_j \subset V_G$, and V_j satisfies one of the following*

- *It is a tree.*
- *There exist vertices w_i and disjoint subsets $U_i \subset V_j$ such that each U_i is a tree of depth at most $2L \log n$, $V_j = \cup_i U_i$ and $w_i \in U_i$, there are no edges between $U_i - w_i$ and $V_j - U_i$. Furthermore the distance between ∂V_j and $W_j = \cup_i w_i$ is at least $L \log n$ and the subgraph W_j has $|W_j| \leq 20tL \log n$ and largest degree at most $2t$.*

Corollary 6.2 *Suppose that G satisfies equation (4). Then there exists $0 < L < \infty$ and (c, α, ϵ) such that we can partition G into blocks $\{V_j\}$ such there is at most one edge between any two blocks. Moreover, for all j, the diameter of V_j is less than $(20t + 2)L \log n$, it holds that $\partial V_j \subset V_G$, and V_j satisfies one of the following*

- *It is a tree.*
- *There exist vertices w_i and disjoint subsets $U_i \subset V_j$ such that each U_i is a tree of depth at most $2L \log n$, $V_j = \cup_i U_i$ and $w_i \in U_i$, there are no edges between $U_i - w_i$ and $V_j - U_i$. Furthermore the distance between ∂V_j and $W_j = \cup_i w_i$ is at least $L \log n$ and the subgraph W_j has $|W_j| \leq 20tL \log n$ and largest degree at most $2t$.*

Proof Letting V_G be the set of good vertices and V_B the set of bad vertices, the proof of the corollary follows from Lemma 6.3 by taking L such that $(20t + 2)L < a$ and choosing (c, α, ϵ) according to Corollary 6.1. □

We now prove Lemma 6.3.

Proof The first step of the proof will be the construction of $W = \cup W_j \subset V$. Beginning with W as the empty set we can add to W in three ways:

- If $u_1, u_2, \ldots, u_m$ is a self-avoiding path of vertices in $V - W$ such that u_1 and u_m are adjacent and $3 \leq m < 5L \log n$ then add $\{u_1, u_2, \ldots, u_m\}$ to W.
- If $u_1, u_2, \ldots, u_m$ is a self-avoiding path in $V - W$ such that both u_1 and u_m are adjacent to W and $2 \leq m < 5L \log n$ then add $\{u_1, u_2, \ldots, u_m\}$ to W.
- If u_1 is adjacent to two vertices in W then add $\{u_1\}$ to W.

The construction of W ends when no more additions are possible.

Claim 1 *W does not depend on the order of the additions.*

Proof Note that if W' and W'' are two different W's obtained for different order of additions then one may add all elements in $W' \setminus W''$ to W' and vice-versa. □

Claim 2 *At each stage of the construction no connected component W_j of W is a tree; each connected component W_j of W has*

$$|W_j| \leq (10Lt(W_j) - 5L)\log n,$$

where $t(W_j)$ is the tree excess of W_j.

Proof We split the additions into three cases. If $u_1, u_2, \ldots, u_m$ is not adjacent to any component of W then this creates a new component W_{new} of W. This must be achieved by an addition of the first type. The new component must contain a loop and have tree excess at least 1 and $|W_{new}|$ is less than $5L\log n$ which is less than $(10Lt(W_{new}) - 5L)\log n$.

Next suppose that an addition $u_1, u_2, \ldots, u_m$ is adjacent to exactly one existing component W_{old} of W. Then the addition forms a new component W_{new} which contains a new loop so $t(W_{new}) \geq t(W_{old}) + 1$. On the other hand

$$|W_{new}| \leq (10Lt(W_{old}) - 5L + 5L)\log n \leq (10Lt(W_{new}) - 5L)\log n.$$

Finally the addition $u_1, u_2, \ldots, u_m$ may be adjacent to two or more components $W_1, \ldots, W_k$ of W and so forms one new component W_{new} from these. Then $t(W_{new}) \geq \sum_{j=1}^{l} t(W_j)$ and

$$|W_{new}| \leq 5L\log n + \sum |W_j| \leq (10Lt(W_{new}) - 5L)\log n.$$

□

Claim 3 *When the construction of W is completed, each component W_j of W is of size at most $20tL\log n$ and tree excess at most t. The distance between two components of W is at least $5L\log n$. All the degrees in W are bounded between 1 and $2t$.*

Proof We have seen that at each of the additions the tree excess of a component increases by at least one. Suppose one of the components of W satisfies $|W_j| > 20tL\log n$. If at some point in the construction the maximum diameter of a component is D then after an addition the new maximum diameter is at most $2D + 5L\log n$. So at some point in the construction there must have been a component W_j with

$$\left(10t - \frac{5}{2}\right) L\log n \leq |W_j| \leq 20tL\log n.$$

Let $v \in W_j$. Then $W_j \subset B(v, 20tL\log n)$ so $t(W_j) \leq t(v, 20tL\log n) \leq t$. Then

$$|W_j| < (10Lt(W_j) - 5L)\log n \leq (10t - 5)L\log n,$$

which is a contradiction. Hence every component of W has size at most $20tL\log n$ and tree excess at most t. By construction all components are separated by distance at least $5L \log n$. Since the tree excess is at most t and by construction W has no leaves the largest degree is at most $2t$. □

As in Lemma 6.1 for $u, u' \in V_B$ we write $u \sim u'$ if there is a path connecting u to u' with no two consecutive vertices belonging to V_G. For each component W_j of W we define V_j as

$$V_j := \{u \in V : \exists u' \in V, u \sim u', d(u', W_j) \leq L\}$$

By construction $W_j \subset V_j$ and if $d(u, W_j) \leq L \log n$ then $u \in V_j$ while if $d(u, w_j) \geq 2L \log n$ then by Corollary 6.1 $u \notin V_j$. It follows that the components V_j are disjoint and are not adjacent. We will show that the components satisfy the hypothesis of the lemma.

Suppose that there exist two self-avoiding paths $u_0, u_1, \ldots, u_l$ and $v_0, v_1, \ldots, v_m$ with $u_l = v_m, u_0, v_0 \in W_j$ and $u_1, \ldots, u_l, v_1, \ldots, v_m \in V_j - W_j$ which are not identical, (i.e. for some $i, u_i \neq v_i$). If $l+m \leq 5L \log n$ then $u_0, u_1, \ldots, u_l, v_0, v_1, \ldots, v_m$ must contain a loop of length less than $5L \log n$ which could be added to W contradicting our assumption. So without loss of generality $l \geq \frac{5}{2}L \log n$. Then there exists u' with $u' \sim u_{\frac{5}{2}L \log n}$ and $d(u', W_j) \leq L \log n$. Then there exists a path in the equivalence class of u' from $u_{\frac{5}{2}L \log n}$ to u' with length at most $L \log n$. Since $d(u', w) \leq L$ for some $w \in W$ there also exists a path from u' to w in $\{u : d(u, W) \leq L\} \subset V_j$ with length at most $L \log n$. Combining these paths there is a path from $u_{\frac{5}{2}L \log n}$ to w in V_j of length at most $2L \log n$. Combining this path with $u_0, u_1, \ldots, u_{\frac{5}{2}L \log n}$ we must have a loop of length at most $\frac{9}{2}L \log n$. But this could be an addition to W which is a contradiction. Hence for each $u \in V_j - W_j$ there is a unique self-avoiding path from u to W_j in $V_j - W_j$. It follows that we can partition V_j into $\{U_i\}$ as required.

Those points in V_B that are not in some V_j can be placed in blocks according to their equivalence class from the relation $\sim$. All such extra blocks are trees of maximum diameter $L \log n$. Finally, vertices $v \in V_G$ belong to the block defined by $u \in V_B$ if (u, v) is an edge E and if no such edge exists v is a separate block. □

7 Block relaxation times

7.1 Colouring model

Lemma 7.1 *Suppose that G satisfies equation (4). For sufficiently large q the relaxation times of the Gibbs sampler on each of the blocks constructed in Lemma 6.3 is bounded by n^C.*

Proof In the blocks V_j which are trees any path is of length at most $20tL \log n$ so

$$m(V_j, v) \leq \frac{1}{\alpha} m_\alpha(V_j, 20tL \log n) \leq \left(1 + \frac{20tL}{a}\right) \frac{\delta}{\alpha} \log n.$$

By Theorem 4.1 and Lemma 5.2 the relaxation time is bounded by n^C.

Now consider a block V_j of the second type. We divide V_j into its sub-blocks U_i. Each U_i is a tree and every $v \in \partial^+_{V_j} U_i$ is (c, α, ϵ)-good. Any path in U_i has length at most $2L \log n$ so

$$m(U_i, w_i) \leq \frac{1}{\alpha} m_\alpha(U_i, 2L \log n) \leq \left(1 + \frac{2L}{a}\right) \frac{\delta}{\alpha} \log n.$$

Then by Theorem 4.1 and Lemma 5.2 the relaxation time of the Gibbs sampler on each U_i is bounded by $n^{C'}$.

Take q sufficiently large so that Lemma 5.2 holds with $\log \lambda < -4/L$. Then for $w_i \in W_j$,

$$\sup_{\sigma(\partial^+_{V_j} U_i)} \sup_{x,y \in \mathcal{C}} \frac{P_{U_i \cup \partial^+_{V_j} U_i}(\sigma(w_i) = x | \sigma(\partial^+_{V_j} U_i))}{P_{U_i \cup \partial^+_{V_j}}(\sigma(w_i) = y | \sigma(\partial^+_{V_j} U_i))} \leq \exp\left(\sum_{v \in \partial^+_{V_j} U_i} \lambda^{d(w_i, v)}\right) \tag{24}$$

$$\leq \exp\left(\sum_{v \in \partial^+_{V_j} U_i} \lambda^{L \log n}\right) \tag{25}$$

$$\leq \exp(n^{-3}) \tag{26}$$

so $P(\sigma(w_i) = x | \sigma(\partial^+_{V_j} U_i)) \geq q^{-1} \exp(-n^{-3})$. Then by Lemmas 4.1 and 4.3 the relaxation time of the block dynamics with blocks $\{U_i\}$ is bounded by $n^{C''}$. Then by Proposition 3.1 we have that the relaxation time of the Gibbs sampler on V_j is bounded by n^C. □

7.2 Hardcore model

Lemma 7.2 *Suppose that G satisfies equation* (4). *For sufficiently small β the relaxation times of the Gibbs sampler on each of the blocks constructed in Lemma 6.3 is bounded by n^C.*

Proof In the blocks V_j which are trees, any path is of length at most $20tL \log n$ so

$$m(V_j, v) \leq \frac{1}{\alpha} m_\alpha(V_j, 20tL \log n) \leq \left(1 + \frac{20tL}{a}\right) \frac{\delta}{\alpha} \log n.$$

By Theorem 4.2 the relaxation time is bounded by n^C.

Now consider a block V_j of the second type. By Lemmas 4.1 and 4.4 the relaxation time of the block dynamics with blocks $\{U_i\}$ is bounded by $n^{C''}$. Then by Proposition 3.1 we have that the relaxation time of the Gibbs sampler on V_j is bounded by n^C. □

7.3 Soft constraints

Lemma 7.3 *Suppose that G satisfies equation* (4). *For small* $\|H\|$ *the relaxation times of the Gibbs sampler on each of the blocks constructed in Lemma 6.3 is bounded by* n^C.

Proof In the blocks V_j which are trees any path is of length at most $20tL\log n$ so

$$m(V_j, v) \leq \frac{1}{\alpha} m_\alpha(V_j, 20tL\log n) \leq \left(1 + \frac{20tL}{a}\right)\frac{\delta}{\alpha}\log n.$$

By Theorem 4.3 the relaxation time is bounded by n^C.

Now consider a block V_j of the second type. Let V_j' be the block obtained by removing each of the edges in the skeleton W_j and let τ' be the relaxation time on V_j'. In the proof of Lemma 4.3 we showed that removing an edge affects the relaxation time by a factor of at most $\exp(4\|H\|)$ so $\tau \leq n^{80\|H\|t}\tau'$. In V_j' each of the trees U_i is separated so τ' is simply the maximum of the relaxation times of the U_i. By Theorem 4.3 the relaxation time is bounded by $n^{C'}$ so each of the U_i are bounded by $n^{C'}$ so $\tau \leq n^C$. □

8 Mixing time of block dynamics

We use the partition from Lemma 6.3 as blocks for the block dynamics of the Gibbs sampler and use path coupling to bound the mixing time of the block dynamics. Let d_H denote the hamming distance of two distributions. Suppose that $T \subset V$ is a tree, let $v \in \partial^+ T$ be (c, α, ϵ)-good and let η, η' be two boundary conditions on $\partial^+ V_j$ which differ only at v and suppose that ρ is the only vertex in T adjacent to v. We must couple two states $\sigma(T), \sigma'(T)$ so that they are distributed as Q and Q' respectively where $Q(\sigma(T)) = P(\sigma(T)|\eta)$ and $Q'(\sigma'(T)) = P(\sigma'(T)|\eta')$. This can be done as follows. Root T at ρ and let $\overleftarrow{u}$ denote the parent of $u \in T$. First couple $\sigma(\rho)$ and $\sigma'(\rho)$ according to their marginal distributions $P(\sigma(\rho)|\eta)$ and $Q'(\sigma'(\rho)|\eta')$ so as to minimize their total variation distance. Proceed inductively down the tree by coupling $\sigma(u)$ and $\sigma'(u)$ according to $P(\sigma(u)|\eta, \sigma(\overleftarrow{u}))$ and $P(\sigma'(u)|\eta, \sigma'(\overleftarrow{u}))$ so as to minimize the total variation distance. When $\sigma(\overleftarrow{u}) = \sigma'(\overleftarrow{u})$ then $\sigma(u) = \sigma'(u)$. We will show that we can bound the expected hamming distance of these coupled distributions.

8.1 Colouring model

Lemma 8.1 *Let T be a tree such that* $\psi(u) = \sum_{w\in\partial^+ T} \lambda^{d(w,u)} < \epsilon$ *for all* $u \in T$. *If* $\delta > 0$ *then for some sufficiently large* $q = q(\delta, \epsilon, \lambda)$, *the above coupling has*

$$E d_H(\sigma(T), \sigma'(T)) \leq \delta.$$

Proof Recalling that v is good, fixing $0 < \gamma < \frac{\delta\epsilon}{\alpha}$ we have that $\varphi_\gamma(v) < \delta$. For all $u \in T$ we have that $\#\{w \in V - T : (w, u) \in E\} \leq \epsilon/\lambda$. By Lemma 5.2 we choose q

large enough so that for each $u \in T$ and $x \in \mathcal{C}$, $P(\sigma(u) = x|\eta) < \gamma/2$. Then

$$d_{TV}(P(\sigma(u) = \cdot|\eta, \sigma(\overleftarrow{u})), P(\sigma(u) = \cdot|\eta, \sigma'(\overleftarrow{u}))) \le 2 \max_x P(\sigma(u) = x|\eta) < \gamma.$$

So given that $\sigma(\overleftarrow{u})$ and $\sigma'(\overleftarrow{u})$ disagree then $\sigma(u)$ and $\sigma'(u)$ disagree with probability at most γ. It follows that the probability that $\sigma(u)$ and $\sigma'(u)$ disagree is at most $\gamma^{d(u,v)}$ and so $Ed_H(\sigma(T), \sigma'(T)) \le \sum_{u \in T} \gamma^{d(u,v)} \le \varphi_\gamma(v) < \delta$ as required. □

Lemma 8.2 *Let V_j be a block constructed from Lemma 6.3. If $v \in \partial^+ V_j$ and η, η' are boundary conditions on $\partial^+ V_j$ which differ only at v then for sufficiently large $q = q(a, \alpha, t, \delta)$ we can couple colourings $\sigma(V_j), \sigma'(V_j)$ distributed as $P(\sigma(V_j)|\eta), P(\sigma'(V_j)|\eta')$ respectively so that*

$$Ed_H(\sigma(V_j), \sigma'(V_j)) \le \delta.$$

Proof The case when V_j is a tree follows by Lemma 8.1 so we consider the blocks of the second type. Let v be adjacent to U_i. If $\sigma^1(W_j)$ and $\sigma^2(W_j)$ are two colourings of W_j then by equation (24)

$$\frac{P(\sigma^1(W_j)|\eta)}{P(\sigma^2(W_j)|\eta)} = \prod_i \frac{P(\sigma^1(w_i)|\eta(\partial^+_{V_j} U_i))}{P(\sigma^2(w_i)|\eta(\partial^+_{V_j} U_i))} \le \prod_i \exp(n^{-3}) \le \exp(n^{-2})$$

and so the total variation distance between $P(\sigma(W_j)|\eta)$ and the free measure on colourings on W_j is $O(n^{-2})$. It follows that we can couple $\sigma(W_j)$ and $\sigma'(W_j)$ so that they agree with probability $1 - O(n^{-2})$. On the event they disagree there are at most $|V_j| \le n$ disagreements so this event contributes $O(n^{-1})$ disagreements to the expected value. So now on the event that $\sigma(W_j) = \sigma'(W_j)$ for all $k \ne i$ we can set $\sigma(U_k - \{w_k\}) = \sigma'(U_k - \{w_k\})$ since they have the same boundary conditions. This just leaves $\sigma(U_i - \{w_i\})$ and $\sigma'(U_i - \{w_i\})$ to be coupled. Now $U_i - \{w_i\}$ is a tree which has every boundary vertex (c, α, ϵ)-good except perhaps w_i. Then repeating the argument of Corollary 5.1 we have that when $\lambda = \alpha^2$

$$\psi(u) \le \lambda + \sum_{u' \in \partial^+ U_i - \{w_i\}} \lambda^{d(u',u)} \le \lambda + \epsilon.$$

Applying Lemma 8.1 to $U_i - \{w_i\}$ completes the result. □

Lemma 8.3 *There exists $\bar{q}$ such that for $q > \bar{q}$ the relaxation time of the discrete time block dynamics with blocks $\{V_j\}$ from Lemma 6.3 is $O(n)$.*

Proof Choose q large enough so that in Lemma 8.2 we can take $\delta < \frac{1}{c}$. By the method of path coupling described in Section 3.4 it is sufficient to show that if σ_0, σ_0' are two colourings with $d_H(\sigma_0, \sigma_0') = 1$ differing only at v then we can couple one step of the block dynamics so that the new pair σ_1, σ_1' has

$$Ed(\sigma_1, \sigma_1') \le 1 - \beta/n$$

for some $\beta > 0$. Let K be the number of blocks. We couple them as follows. If the block V_j chosen by the block dynamics contains v then we set $\sigma(V_j) = \sigma'(V_j)$ and have $d(\sigma_1, \sigma_1') = 1$. If the block chosen is adjacent to v then we couple V_j according to Lemma 8.2. The expected number of new disagreements is at most δ. If V_j neither contains nor is adjacent to v then we set $\sigma(V_j) = \sigma'(V_j)$ and the number of disagreements does not change. Now if v is adjacent to some blocks V_j it must be in the boundary and so therefore must be (c, α, ϵ)-good. Since it has degree at most c it is adjacent to at most c blocks so

$$Ed(\sigma_1, \sigma_1') \leq 1 - \frac{1}{K} + c\frac{\delta}{K} \leq 1 - \beta/n$$

where $\beta = 1 - c\delta$ which completes the proof. □

8.2 Hardcore model

Lemma 8.4 *Let T be a tree such that $\psi(u) = \sum_{w \in \partial^+ T} \lambda^{d(w,u)} < \epsilon$ for all $u \in T$. If $\delta > 0$ then there exists $\beta^* = \beta^*(\delta, \lambda, \epsilon)$ such that if $\beta < \beta^*$, the above coupling has*

$$Ed_H(\sigma(T), \sigma'(T)) \leq \delta.$$

Proof Let $\gamma > 0$ such that $\varphi_\gamma(v) < \delta$. We can choose β small enough so that $\frac{e^\beta}{1+\beta} < \gamma$. For all $u \in T$, $P(\sigma(u) = 1|\eta) \leq P(\sigma(u) = 1|\sigma(V - \{u\}) \equiv 0) = \frac{e^\beta}{1+\beta} < \gamma$. Then

$$\begin{aligned} &d_{TV}(P(\sigma(u) = \cdot|\eta, \sigma(\overleftarrow{u})), P(\sigma(u) = \cdot|\eta, \sigma'(\overleftarrow{u}))) \\ &\quad \leq \left|P(\sigma(u) = 1|\eta, \sigma(\overleftarrow{u})) - P(\sigma(u) = 1|\eta, \sigma'(\overleftarrow{u}))\right| < \gamma. \end{aligned}$$

So given that $\sigma(\overleftarrow{u})$ and $\sigma'(\overleftarrow{u})$ disagree then $\sigma(u)$ and $\sigma'(u)$ disagree with probability at most γ. It follows that the probability that $\sigma(u)$ and $\sigma'(u)$ disagree is at most $\gamma^{d(u,v)}$ and so $Ed_H(\sigma(T), \sigma'(T)) \leq \sum_{u \in T} \gamma^{d(u,v)} \leq \varphi_\gamma(v) < \delta$ as required. □

The following results follow similarly to the colouring model.

Lemma 8.5 *Let V_j be a block constructed from Lemma 6.3. For $\delta > 0$ there exists $\beta^* = \beta^*(a, \alpha, t, \delta)$ such that for $\beta < \beta^*$ if $v \in \partial^+ V_j$ and η, η' are boundary conditions on $\partial^+ V_j$ which differ only at v then we can couple states $\sigma(V_j), \sigma'(V_j)$ distributed as $P(\sigma(V_j)|\eta)$, $P(\sigma'(V_j)|\eta')$ respectively so that*

$$Ed_H(\sigma(V_j), \sigma'(V_j)) \leq \delta.$$

Lemma 8.6 *There exists $\beta^* = \beta^*(a, \alpha, t, \delta)$ such that for $\beta < \beta^*$ the relaxation time of the block dynamics with blocks $\{V_j\}$ from Lemma 6.3 is $O(n)$.*

8.3 Soft constraints model

Lemma 8.7 *Let T be a tree such that $\psi(u) = \sum_{w\in\partial^+T} \lambda^{d(w,u)} < \epsilon$ for all $u \in T$. If $\delta > 0$ then there exists $H^* = H^*(\delta, \lambda, \epsilon) > 0$ such that if $\|H\| < H^*$, the above coupling has*

$$E d_H(\sigma(T), \sigma'(T)) \leq \delta.$$

Proof Let $\gamma > 0$ such that $\varphi_\gamma(v) < \delta$. Repeating the argument of Lemma 5.3 we can choose $\|H\|$ small enough so that

$$d_{TV}(P(\sigma(u) = \cdot|\eta, \sigma(\overleftarrow{u})), P(\sigma(u) = \cdot|\eta, \sigma'(\overleftarrow{u}))) < \gamma.$$

The remainder of the proof follows similarly from Lemma 8.4. □

The following results follow similarly from the colouring model.

Lemma 8.8 *Let V_j be a block constructed from Lemma 6.3. For $\delta > 0$ there exists $H^* = H^*(a, \alpha, t, \delta)$ such that for $\|H\| < H^*$ if $v \in \partial^+ V_j$ and η, η' are boundary conditions on $\partial^+ V_j$ which differ only at v then we can couple states $\sigma(V_j), \sigma'(V_j)$ distributed as $P(\sigma(V_j)|\eta), P(\sigma'(V_j)|\eta')$ respectively so that*

$$E d_H(\sigma(V_j), \sigma'(V_j)) \leq \delta.$$

Lemma 8.9 *There exists $H^* = H^*(a, \alpha, t, \delta)$ such that for $\|H\| < H^*$ the relaxation time of the block dynamics with blocks $\{V_j\}$ from Lemma 6.3 is $O(n)$.*

9 Main results

The main results now follows easily using the block dynamics approach of Proposition 3.1.

Proof (Theorem 1.2) For large enough q, by Lemma 8.3 the relaxation time of the block dynamics of the Gibbs sampler on G with blocks $\{V_j\}$ from Lemma 6.3 is $O(n)$. By Lemma 7.1 the relaxation time of the Gibbs sampler on each block is bounded by $n^{C'}$. Then by Proposition 3.1 we have that the relaxation time is $O(n^{C'+1})$. There are at most q^n colourings of G so $\log(1/\min_\sigma P(\sigma)) \leq n \log q$ so the mixing time of the Gibbs sampler is bounded by $O(n^{C'+2})$ which completes the result. □

The proofs of Theorems 1.4 and 1.6 follow similarly.

Open Access This article is distributed under the terms of the Creative Commons Attribution Noncommercial License which permits any noncommercial use, distribution, and reproduction in any medium, provided the original author(s) and source are credited.

References

1. Aldous, D., Fill, J.A.: 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. Berger, N., Kenyon, C., Mossel, E., Peres, Y.: Glauber dynamics on trees and hyperbolic graphs. Probab. Theory Relat. Fields **131**(3), 311–340 (2005)
3. Bubley, R., Dyer, M.: Path coupling: a technique for proving rapid mixing in Markov chains. In: Proceedings of the 38th Annual Symposium on Foundations of Computer Science (FOCS), pp. 223–231 (1997)
4. Chen, M.F.: Trilogy of couplings and general formulas for lower bound of spectral gap. In: Probability towards 2000, number 128 in Lecture Notes in Statistics (1998)
5. Dyer, M., Flaxman, A.D., Frieze, A.M., Vigoda, E.: Randomly coloring sparse random graphs with fewer colors than the maximum degree. Random Struct. Algorithms **29**(4), 450–465 (2006)
6. Dyer, M., Frieze, A., Hayes, T., Vigoda, E.: Randomly coloring constant degree graphs. In: Proceedings of the 45th Annual IEEE Symposium on Foundations of Computer Science (FOCS'04), pp. 582–589 (2004)
7. Dyer, M., Frieze, A., Jerrum, M.: On counting independent sets in sparse graphs. In: Proceedings of 40th IEEE Sypmosium on Foundations of Computer Science (FOCS), pp. 210–217 (1999)
8. Dyer, M.E., Greenhill, C.S.: On Markov chains for independent sets. J. Algorithms **35**(1), 17–49 (2000)
9. Efthymiou, C., Spirakis, P.G.: Randomly colouring sparse graphs using a constant number of colours (2007, unpublished)
10. Frieze, A.M., McDiarmid, C.: Algorithmic theory of random graphs. Random Struct. Algorithms **10** (1-2), 5–42 (1997)
11. Gershchenfeld, A., Montanari, A.: Reconstruction for models on random graphs. In: Foundations of Computer Science, Annual IEEE Symposium, pp. 194–204. IEEE Computer Society (2007)
12. Goldberg, L.A., Martin, R., Paterson, M.: Strong spatial mixing for lattice graphs with fewer colours. In: Proceedings of the 45th Annual IEEE Symposium on Foundations of Computer Science (FOCS'04), pp. 562–571 (2004)
13. Hayes, T.P.: 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), pp. 39–46 (2006)
14. Jerrum, M.: A very simple algorithm for estimating the number of k-colorings of a low-degree graph. Random Struct. Algorithms **7**, 157–166 (1995)
15. Jerrum, M., Sinclair, A.: Approximating the permanent. SIAM J. Comput. **18**(6), 1149–1178 (1989)
16. Jerrum, M., Sinclair, A., Vigoda, E.: A polynomial-time approximation algorithm for the permanent of a matrix with non-negative entries. J. ACM **51**(4), 671–697 (2004)
17. Kenyon, C., Mossel, E., Peres, Y.: Glauber dynamics on trees and hyperbolic graphs. In: 42nd IEEE Symposium on Foundations of Computer Science (Las Vegas, NV, 2001), pp. 568–578. IEEE Computer Soc., Los Alamitos, CA (2001)
18. Martinelli, F.: Lectures on Glauber dynamics for discrete spin models. In: Lectures on Probability Theory and Statistics (Saint-Flour, 1997), vol. 1717 of Lecture Notes in Mathematics, pp. 93–191. Springer, Berlin (1999)
19. Martinelli, F., Sinclair, A., Weitz, D.: The ising model on trees: boundary conditions and mixing time. In: Proceedings of the Forty Fourth Annual Symposium on Foundations of Computer Science, pp. 628–639 (2003)
20. Mossel, E., Sly, A.: Rapid mixing of gibbs sampling on graphs that are sparse on average. In: Proceedings of the Nineteenth Annual ACM-SIAM Symposium on Discrete Algorithms (SODA), pp. 238–247 (2008)
21. Mossel, E., Sly, A.: Rapid mixing of gibbs sampling on graphs that are sparse on average. Random Struct. Algorithms (2009, to appear)
22. Mossel, E., Weitz, D., Wormald, N.: On the hardness of sampling independent sets beyond the tree threshold. Prob. Theory Relat. Fields **143**, 401–439 (2009)
23. Shamir, E., Upfal, E.: Sequential and distributed graph coloring algorithms with performance analysis in random graph spaces. J. Algorithms **5**(4), 488–501 (1984)
24. Vigoda, E.: A note on the glauber dynamics for sampling independent sets. Electron. J. Comb. **8**(1), 1–8 (2001)
25. Weitz, D.: Counting indpendent sets up to the tree threshold. In: Proceedings of the Thirty-Eighth Annual ACM Symposium on Theory of Computing, pp. 140–149. ACM (2006)