Strong Spatial Mixing for Binary Markov Random Fields

Jinshan Zhang†, Heng Liang and Fengshan Bai
Department of Mathematical Sciences, Tsinghua University,
Beijing 100084, China

Abstract The remarkable contribution by Weitz gives a general framework to establish the strong spatial mixing property of Gibbs measures. In light of Weitz’s work, we prove the strong spatial mixing for binary Markov random fields under the condition that the ‘external field’ is uniformly large or small by turning them into a corresponding Ising model. Our proof is done through a ‘path’ characterization of the Lipchitz method and recursive formula on trees, which enables us to combine the idea of the self-avoiding tree.

Keywords Gibbs measure · Lipchitz method · Self-avoiding tree · Ising model

1 Introduction

Strong spatial mixing property of Gibbs measures is important in statistical physics. It roughly says that if there is a modification (or perturbation) on the boundary conditions, its influence to the Gibbs measure of a single vertex decays exponentially as the distance to the support of the perturbation (the set of vertices, whose spins are changed) increases. In the classic literatures, it is required that the support of the perturbation has to be a single vertex [3]. Weitz considers the case that the support of perturbation could be a set of vertices with arbitrary size [13]. This generalization is equivalent to the one in [3] where the graph grows sub-exponentially (e.g. integer lattices). In fact, the contribution by Weitz has much wider applications. For example, it provides a natural algorithm to calculate the partition function of Gibbs measures when the strong spatial mixing is met [13]. In this paper, the definition of strong spatial mixing is in the sense of Weitz’s.

Recently the strong spatial mixing is also studied through recursive formula. This approach is introduced by Weitz [13] and Bandyopadhyay and Gamarnik [1] for counting the number of independent sets and colorings. The key point of this method is to build the strong spatial mixing on certain rooted trees. In [13], the equivalence between the marginal probability of a vertex in a general graph \( G \) and that of the root of a tree for hard core model is proved by introducing the self-avoiding tree technique. Using this technique, Weitz gives a remarkable result that hard-core model on bounded degree trees exhibits strong spatial mixing, which resolves a long standing open problem in statistical physics. With the motivation of the
construction of the self-avoiding tree, Weitz’s work is generalized to certain Markov random fields models [6, 11] and TP decoding problem [8]. Instead of constructing a self-avoiding tree, Gamarnik et.al. [5] and Bayati et.al. [2] create a computation tree and establish the strong spatial mixing on the corresponding computation tree for list-coloring and matching problems.

In this paper, we consider the binary Markov random fields. Unlike previous results which establish strong spatial mixing on trees and then appeal to graphs of maximum degree, we focus on the graphs that are sparse on average. These are graphs with a bounded average degree along any self-avoiding simple path, where the total degrees along each self-avoiding path (a path with distinct vertices) with length $O(\log n)$ is $O(\log n)$. This graph has been considered by Mossel and Sly for Ising models [10]. Their idea was to bound size of the self-avoid tree with a lemma that was independent of the model. Now for any problem which exhibits strong spatial mixing on the tree, one can apply the same argument and get strong spatial mixing for the bounded average degree graph. We prove this fact for binary Markov random fields by turning them into a corresponding Ising model.

Our main result is as follows. For any ‘inverse temperature’ on the ‘sparse on average’ graph, Gibbs distribution exhibits strong spatial mixing when the ‘external field’ is uniformly larger than $B(d, \alpha_{\text{max}}, \gamma)$ or smaller than $-B(d, \alpha_{\text{min}}, \gamma)$. Here, $d$ is ‘maximum average degree’ and $\alpha_{\text{min}}, \alpha_{\text{max}}, \gamma$ are parameters of the system. To the best of our knowledge, this condition on ‘external field’ is first considered for strong spatial mixing. Our result is proved by combining two basic tools. One is a recursive formula on a tree and the self-avoiding tree technique, which is proposed in [6]. The other is Lipchitz method, which was used in [1, 2, 5]. The novelty of our proof is that we propose a ‘path’ characterization of Lipchitz method, which enables us to give the ‘external field’ condition in terms of ‘maximum average degree’ for the strong spatial mixing.

The remaining part of the paper has the following structure. In Section 2, we present some preliminary definitions and notations. The main results are presented in Section 3. Section 4 is devoted to prove the main theorem. Conclusions and discussions on further works are given in Section 5.

2 Preliminaries

Let $G = (V, E)$ be a finite simple graph with vertices $V = \{1, 2, \cdots, n\}$ and edge set $E$, and let $d(u, v)$ denote the distance (e.g. the length of shortest path) between $u$ and $v$, for any $u, v \in V$. A path $v_1, v_2, \cdots$ is called a self-avoiding path if $v_i \neq v_j$ for any $i \neq j$. The distance between a vertex $v \in V$ and a subset $\Lambda \subset V$ is defined as

$$d(v, \Lambda) = \min\{d(v, u) : u \in \Lambda\}.$$ 

A set of vertices with distance $l$ to the vertex $v$ is denoted by

$$S(G, v, l) = \{u : d(v, u) = l\}.$$
Let $\delta_v$ denote the degree of the vertex $v \in G$ and $\Delta(G) = \max_v \delta_v$. The maximal path density of the graph $G$ is given by

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

where the maximum is taken over all self-avoiding paths $\Gamma$ starting at $v$ with length at most $l$. The maximum average path degree $\delta(G, v, l)$ is defined by

$$\delta(G, v, l) = (m(G, v, l) - \delta_v)/l, \quad l \geq 1.$$

The maximum average degree of $G$ is defined as

$$\Delta(G, l) = \max_{v \in V} \delta(G, v, l).$$

For any order of all the vertices in $G$ given, an associated partial order of $E$ based on the order of $V$ is defined as $(i, j) > (k, l)$ if and only if $(i, j)$ and $(k, l)$ share a common vertex and $i + j > k + l$. In binary Markov random fields (BMRF) on $G$, each vertex $i \in V$ is associated with a random variable $X_i$ on $\Omega = \{\pm 1\}$ (briefly $\pm$).

**Definition 1.** The Gibbs measure of BMRF on $G$ is defined by the joint distribution of the random vector $X = \{X_1, X_2, \cdots, X_n\}$

$$P_G(X = \sigma) = \frac{1}{Z(G)} \exp \left( \sum_{(i, j) \in E} \beta_{ij}(\sigma_i, \sigma_j) + \sum_{i \in V} h_i(\sigma_i) \right),$$

where $\sigma \in \Omega^n$, $h_i : \Omega \to R$ and $\beta_{ij} : \Omega^2 \to R$. Here $Z(G)$ is the partition function of the system.

Note that the Gibbs measure clearly satisfies $\sum_{\sigma \in \Omega^n} P_G(X = \sigma) = 1$. We use the notation $\beta_{ij}(a, b) = \beta_{ji}(b, a)$ here. For any $\Lambda \subset V$, $\sigma_{\Lambda}$ denotes the set $\{\sigma_i, i \in \Lambda\}$. With a little abuse of notation, $\sigma_{\Lambda}$ also denotes the configuration that $i$ is fixed $\sigma_i$, for any $i \in \Lambda$. Let $Z(G, \Phi)$ denote the partition function under the condition $\Phi$. For example, $Z(G, X_1 = +)$ represents the partition function under the condition the random variable $X_1$ on the vertex 1 is fixed to be $+$.

A self-avoiding walk (SAW) is a sequence of moves (on a graph) which does not visit the same point more than once. A crucial concept in proving our results is given as follows, which is introduced in [13].

**Definition 2.** (Self-avoiding tree) The self-avoiding tree $T_{\text{saw}(v)}(G)$ (for simplicity denoted by $T_{\text{saw}(v)}$) corresponding to the vertex $v$ of $G$ is the tree with root $v$ and generated through all the self-avoiding walks originating at $v$. For all the cycles of $G$, a vertex closing a cycle is included as a leaf of the tree. The leaf is assigned to be $+$, if the edge ending the cycle is larger than the edge starting the cycle, and $-$ otherwise.

Figure 1 shows an example of the generation process of the self-avoiding tree. The graph on the left is the self-avoiding tree of the graph on the right. Consider, for example, the path $1, 2, 3, 1$ in the self-avoiding tree. It is clearly generated
Figure 1: The graph with one vertex assigned + (Right) and its corresponding self-avoiding tree $T_{\text{saw}(1)}$ (Left)

through self-avoiding walk from the cycle with vertices 1, 2, 3 in the original graph. The ending vertex 1 in this path is assigned $+$, since the ending edge $(3, 1)$ is larger than the starting edging $(1, 2)$. Look at another path 1, 3, 2, 1. The ending vertex 1 should be assigned $-$ this time, since the ending edge $(2, 1)$ is less than the starting edge $(1, 3)$.

Figure 2: The graph with one vertex assigned + (Right) and its alternative self-avoiding tree $T_{\text{saw}(1)}$ (Left)

Remark: Consider the self-avoiding tree for certain given configuration $\sigma_\Lambda$ of $G$. Since the spin $\sigma_i$ of the vertex $i$ in the set $\Lambda$ is fixed, the subtree below this vertex does not need to be constructed due to the Markov property on trees. Hence these subtrees need not appear in the self-avoiding tree, by [13]. Take Figure 1 as an example again, where the vertex 5 in $G$ is fixed to be $+$. The self-avoiding graph corresponding to the cycle with vertices 2, 4, 5 should be the path 2, 4, 5, 2, and the path 2, 5, 4, 2 (see Figure 1). However, since vertex 5 is fixed to be $+$, we do not need to construct the path below the vertex 5 due to the Markov property on
trees. Hence the graph appearing on the self-avoiding tree (see Figure 2) on the
right corresponding to this cycle are only the path 2, 5 and the path 2, 4, 5 with 5
assigned +. Therefore the self-avoiding tree shown in Figure 1 can be replaced by
that in Figure 2.

To generalize the strong spatial mixing property on trees to general graph, we
need to utilize an important property of the self-avoiding tree. This is one of the
main results of [13], which is also explicitly stated in [6]. For any configuration \( \sigma_\Lambda \)
of \( G, \Lambda \subset V \), \( \sigma_\Lambda \) is also used to denote the configuration of \( T_{saw(v)} \) obtained by
imposing the condition corresponding to \( \sigma_\Lambda \).

**Proposition 1.** For BMRF on \( G = (V, E) \), \( \sigma_\Lambda \) is any configuration on \( G, \Lambda \subset V \) and any vertex \( v \in V \), then
\[
P_G(X_v = +|\sigma_\Lambda) = P_{T_{saw(v)}}(X_v = +|\sigma_\Lambda).
\]

The following properties are useful in obtaining results of the sparse on average
graph. The proof is based on induction and can be found in [10].

**Proposition 2.** Let \( j, l \) be positive integers. Then one has
\[
m(G, v, jl) \leq j \max_{u \in G} \{m(G, u, l) - \delta_u\} + \delta_v
\]
and
\[
|S(T_{saw(v)}, v, l + 1)| \leq \delta_v(\delta(G, v, l) - 1)^l.
\]

**Definition 3.** *(Strong spatial mixing)* The Gibbs distribution of BMRF exhibits
strong spatial mixing if and only if there exist positive numbers \( 0 < a, 0 < b, 0 < c < 1 \)
independent of \( n \), for any vertex \( v \in V \), subset \( \Lambda \subset V \) and any two
configurations \( \sigma_\Lambda \) and \( \eta_\Lambda \) on \( \Lambda \), denoting perturbation set \( \Theta = \{v \in \Lambda : \sigma_v \neq \eta_v\} \) and \( t = d(v, \Theta) \), when \( t = ka \log n + 1 \), \( k = 1, 2, \ldots \), there is
\[
|P_G(X_v = +|\sigma_\Lambda) - P_G(X_v = +|\eta_\Lambda)| \leq f(t),
\]
where decay function \( f(t) = be^t \).

**Remark:** In Definition 3, we see the decay function decreases exponentially with \( t \). Hence we also call the Gibbs distribution satisfying Definition 3 exhibits exponential
strong spatial mixing.

### 3 Main Results

According to the theory of the binary Markov random fields, when \( \beta_{ij}(\sigma_i, \sigma_j) = J_{ij}\sigma_i\sigma_j \) and \( h_i = B_i\sigma_i \) for every edge \((i, j) \in E\) and vertex \( i \in V \), and \( J_{ij} \) is uniformly
positive (or negative) for all \((i, j) \in E\), the BMRF is just the ferromagnetic (or
antiferromagnetic) Ising model. For simplicity, we use the following notations. Let
\[
J_{ij} = \frac{\beta_{ij}(+,-) + \beta_{ij}(-,+)}{4} - \frac{\beta_{ij}(-,-) - \beta_{ij}(+,+)}{4}, \quad \text{and} \quad B_i = \frac{h_i(+) - h_i(-)}{2}
\]
for all edges and vertices. We call $J_{ij}$ and $B_i$ ‘inverse temperature’ and ‘external field’ of BMRF. It is easy to check these two quantities are exactly corresponding to the inverse temperature and external field when the model is restricted to Ising model. Let $J = \max_{(i,j) \in E} |J_{ij}|$, $B_{\min} = \min_{i \in V} B_i$, and $B_{\max} = \max_{i \in V} B_i$. Denote

$$\alpha_{\max} = \max_{(i,j) \in E} \{\beta_{ij}(-,-) - \beta_{ij}(+,-), \beta_{ij}(-,+) - \beta_{ij}(+,+)\}$$

and

$$\alpha_{\min} = \min_{(i,j) \in E} \{\beta_{ij}(-,-) - \beta_{ij}(+,-), \beta_{ij}(-,+) - \beta_{ij}(+,+)\}.$$  

Let

$$\gamma_{ij} = \max_{(i,j) \in E} \{\frac{|b_{ij}c_{ij} - a_{ij}d_{ij}|}{a_{ij}c_{ij}}, \frac{|b_{ij}c_{ij} - a_{ij}d_{ij}|}{b_{ij}d_{ij}}\},$$

and $\gamma = \max_{(i,j) \in E} \{\gamma_{ij}\}$, where

$$a_{ij} = \exp(\beta_{ij}(+,-)), \quad b_{ij} = \exp(\beta_{ij}(+,-)), \quad c_{ij} = \exp(\beta_{ij}(-,+)), \quad d_{ij} = \exp(\beta_{ij}(-,-)).$$

**Theorem 1.** Let $G = (V, E)$ be a graph with $n$ vertices. There exit two positive numbers $a > 0$ and $d > 0$ such that $\Delta(G, a \log n) \leq d$, and $(d-1) \tanh J \geq 1$. Assume

$$B_{\min} > B(d, \alpha_{\max}, \gamma) \quad (3.1)$$

or

$$B_{\max} < -B(d, -\alpha_{\min}, \gamma) \quad (3.2)$$

where

$$B(d, \alpha, \gamma) = \frac{(d-1)\alpha}{2} + \log\left(\sqrt{\gamma(d-1)} + \sqrt{\gamma(d-1) - \frac{4}{\gamma}}\right).$$

Then the Gibbs distribution of BMRF exhibits exponential strong spatial mixing, and the decay function corresponding to the condition (3.1) is

$$f(t) = \frac{\Delta(G)\gamma}{4} \left(\frac{(d-1)\gamma \exp(2B_{\min} - (d-1)\alpha_{\max})}{1 + \exp(2B_{\min} - (d-1)\alpha_{\max})}\right)^{t-1}$$

with

$$\frac{(d-1)\gamma \exp(2B_{\min} - (d-1)\alpha_{\max})}{1 + \exp(2B_{\min} - (d-1)\alpha_{\max})} < 1,$$

and that to the condition (3.2) is

$$f(t) = \frac{\Delta(G)\gamma}{4} \left(\frac{(d-1)\gamma \exp(2B_{\max} - (d-1)\alpha_{\min})}{1 + \exp(2B_{\max} - (d-1)\alpha_{\min})}\right)^{t-1},$$

with

$$\frac{(d-1)\gamma \exp(2B_{\max} - (d-1)\alpha_{\min})}{1 + \exp(2B_{\max} - (d-1)\alpha_{\min})} < 1.$$

In order to help in understanding the results of Theorem 1, we present a simple example. Consider a ferromagnetic Ising model on $G = (V, E)$ with $n$ vertices, where $\beta_{ij}(\sigma_i, \sigma_j) = J\sigma_i\sigma_j$ and $h_i = B\sigma_i$ for any $(i,j) \in E$ and $i \in V$, and $J > 0$. Also assume there exist two positive numbers $a > 0$ and $d > 0$ such that $\Delta(G, a \log n) \leq d$ and $(d-1) \tanh J \geq 1$. Now the parameters in Theorem 1 are: $\alpha_{\max} = -\alpha_{\min} = 2J$ and $\gamma = e^{2J} - e^{-2J}$. Hence we get $B_0 = B(d, \alpha, \gamma) = (d-1)J + \log\left(\frac{(e^{2J} + e^{-2J})}{2}\right)$. By Theorem 1, when
the external field \( B > B_0 \) or \( B < -B_0 \), the Ising model exhibits exponential strong spatial mixing. This is, in some sense, like the property of uniqueness of Gibbs measures on Ising model. As we know the strong spatial mixing implies the uniqueness of Gibbs measures of Ising model [14], we will discuss such relations further later.

**Remark:** In Theorem 1, by the definition of \( \gamma_{ij} \), one has

\[
\gamma_{ij}^2 \geq \left| \left( \frac{b_{ij}}{a_{ij}} - \frac{d_{ij}}{c_{ij}} \right) \left( \frac{c_{ij}}{d_{ij}} - \frac{a_{ij}}{b_{ij}} \right) \right| = \left( e^{2J_{ij}} - e^{-2J_{ij}} \right)^2.
\]

Hence

\[
\gamma_{ij} \geq \left| e^{2J_{ij}} - e^{-2J_{ij}} \right| = \left( e^{J_{ij}} + e^{-J_{ij}} \right)^2 \tanh J_{ij} \geq 4 \tanh J_{ij}.
\]

Therefore, \( \gamma(d-1) - 4 \geq 0 \) under the condition \((d-1) \tanh J \geq 1.\) The case of \((d-1) \tanh J < 1\) is discussed separately in [14] with totally different method.

If the degree of the graph is bounded with maximum degree \( d \), the condition for ‘external field’ can be relaxed to \( B_i > B(d, \alpha_{\max}, \gamma) \) or \( B_i < -B(d, -\alpha_{\min}, \gamma) \) for any \( i \in V \), which does not require that ‘external field’ is uniformly large or uniformly small as in Theorem 1.

**Corollary 1.** Let the degree of the graph \( G = (V, E) \) be bounded with the maximum degree \( \Delta(G) = d \). Assume a BMRF be defined on \( G \) and \( \tanh J(d-1) \geq 1. \) If \( B_i > B(d, \alpha_{\max}, \gamma) \) or \( B_i < -B(d, -\alpha_{\min}, \gamma) \) for any \( i \in V \), the Gibbs distribution exhibits strong spatial mixing.

### 4 Proofs

Theorem 1 is proved with the recursive formula [6]. The main tool dealing with recursive formula is the first order Taylor expansion of multi-variable function in mathematical analysis. It is called Lipchitz method, which was used in [1, 2, 5]. However unlike these former works, we do not maximize the derived function in each step. Instead, we deal with derived functions and the parameters in the last step. Hence, we present a ‘path’ characterization of this method, which enables us to give the ‘external field’ condition in terms of ‘maximum average degree’ for the strong spatial mixing. Lipchitz method is very useful in analyzing property of the recursive functions.

A ‘path’ version of Lipchitz method is presented first. We use the following notations for simplicity. Let \( T = (V, E) \) be a tree rooted at 0 with vertices \( V = \{0, 1, 2, \cdots, n\} \), edge set \( E \) and BMRF on it. For each edge \((i, j) \in E\), recall the notations in Theorem 1,

\[
a_{ij} = e^{\beta_{ij}(\cdot, +)}, \quad b_{ij} = e^{\beta_{ij}(\cdot, -)}, \quad c_{ij} = e^{\beta_{ij}(\cdot, \cdot)}, \quad d_{ij} = e^{\beta_{ij}(\cdot, \cdot)}.\]

Let \( M_{ij} = c_{ij} - d_{ij} \) and \( N_{ij} = a_{ij} - b_{ij} \). Define

\[
f_{ij}(x) = \frac{M_{ij}x + d_{ij}}{N_{ij}x + b_{ij}} \quad \text{and} \quad h_{ij}(x) = \frac{a_{ij}d_{ij} - b_{ij}c_{ij}}{(M_{ij}x + d_{ij})(N_{ij}x + b_{ij})}.
\]

For any \( i \in V \), let \( T_i \) denote the subtree rooted at \( i \) and there is an associated BMRF on \( T_i \) restricted by BMRF on \( T \). Recall that \( B_i = \frac{h_{ij}(+) - h_{ij}(-)}{2} \) is the external field. Denote \( \lambda_i = e^{-2B_i} \), and let \( \Gamma_{ij} \) be the unique self-avoiding path from \( i \) to \( j \) on \( T \).
Lemma 1. For any \((i, j) \in E\), \(\max_{x \in [0, 1]} |h_{ij}(x)| \leq \gamma_{ij}\).

Proof. Since \(M_{ij}x + d_{ij} \geq 0\) and \(N_{ij}x + b_{ij} \geq 0\), \(\forall x \in [0, 1]\), we only need to prove

\[
\min_{x \in [0, 1]} w(x) = \min(a_{ij}c_{ij}, b_{ij}d_{ij}),
\]

where \(w(x) = (M_{ij}x + d_{ij})(N_{ij}x + b_{ij})\). The case \(M_{ij}N_{ij} = 0\) is trivial. Without loss of generality, suppose \(M_{ij}N_{ij} \neq 0\). Note that \(x_i = \frac{d_{ij}N_{ij} + b_{ij}M_{ij}}{2M_{ij}N_{ij}}\) is an extremum of \(w(x)\) on \(R\). There are three cases needed to be discussed.

Case 1. \(M_{ij}N_{ij} < 0\), then \(w(x)\) reaches its minimum at boundary. Then \(\min_{x \in [0, 1]} w(x) = \min(w(0), w(1)) = \min(a_{ij}c_{ij}, b_{ij}d_{ij})\).

Case 2. \(M_{ij} > 0, N_{ij} > 0\), then \(x_i \leq 0\), \(w(x)\) is increasing on \([0, 1]\), then \(\min_{x \in [0, 1]} w(x) = w(0) = b_{ij}d_{ij}\).

Case 3. \(M_{ij} < 0, N_{ij} < 0\), then \(x_i \geq 1\), \(w(x)\) is decreasing on \([0, 1]\), hence \(\min_{x \in [0, 1]} w(x) = w(1) = a_{ij}c_{ij}\). \(\square\)

With Lemma 1, we present a ‘path’ version of Lipchitz approach.

Lemma 2. Let \(\Lambda \subset V\), \(\zeta_{\Lambda}\) and \(\eta_{\Lambda}\) be any two configurations on \(\Lambda\). Let \(\Theta = \{i : \zeta_{i} \neq \eta_{i}, i \in \Lambda\}\), \(t = d(0, \Theta)\) and \(S(T, 0, t) = \{i : d(0, i) = t, i \in T\}\). Then

\[
|P_{T}(X_0 = +|\zeta_{\Lambda}) - P_{T}(X_0 = +|\eta_{\Lambda})| 
\leq \gamma_{T} \sum_{k \in S(T, 0, t)} \prod_{i \neq k} g_{i}(z_{i})(1 - g_{i}(z_{i}))
\]

where \(z_{i}\) are constant vector with elements in \([0, 1]\), and \(g_{i}(x_{i}) = (1 + \lambda_{i} \prod_{(i, j) \in T_{i}} f_{ij}(x_{ij}))^{-1}\), \(x_{i} = (x_{ii_{1}}, x_{ii_{2}}, \ldots, x_{ii_{l_{-1}}})\).

Proof. For any vertex \(i\) in \(T\), recall that \(T_{i}\) denotes the subtree rooted at \(i\) with BMRF induced by \(T\). Let \(P_{T_{i}}^{\zeta_{\Lambda}} \equiv P_{T_{i}}(X_{i} = +|\zeta_{\Lambda})\) and \(P_{T_{i}}^{\eta_{\Lambda}} \equiv P_{T_{i}}(X_{i} = +|\eta_{\Lambda})\), where \(\zeta_{\Lambda}\) is configuration by restriction of \(\zeta_{\Lambda}\) on \(T_{i}\). Let \(\Omega_{T_{i}}\) denote the configuration space in \(T_{i}\) under the condition \(\zeta_{\Lambda}, i = 1, 2, \ldots, n\), \(\Omega_{0}\) denote the configuration space in \(T_{0}\) under the condition \(\zeta_{\Lambda} \cup \{\sigma_{0}\}\). Let \(0, 0_{1}, 0_{2}, \ldots, 0_{q}\) be the neighbors of \(0\), \(q = \delta_{0}\) (the
degree of the root). Now the recursive formula is presented

\[
R_{0}^{\zeta_{\lambda}} = \frac{Z(T_{0}, X_{0} = +, \zeta_{\lambda})}{Z(T_{0}, X_{0} = -, \zeta_{\lambda})}
\]

\[
e^{h_{0}(+)} \sum_{\sigma \in T_{0}} e^{i=1} \sum_{(k,i) \in T_{0}} (\beta_{0}(+ \sigma_{0}) + \beta_{k}(\sigma_{k}, \sigma_{i}) + \sum_{k \in T_{0}} h_{k}(\sigma_{k}))
\]

\[
e^{h_{0}(-)} \sum_{\sigma \in T_{0}} e^{i=1} \sum_{(k,i) \in T_{0}} (\beta_{0}(- \sigma_{0}) + \beta_{k}(\sigma_{k}, \sigma_{i}) + \sum_{k \in T_{0}} h_{k}(\sigma_{k}))
\]

\[
e^{2B_{0}} \prod_{i=1}^{q} a_{00,i} Z(T_{0}, X_{i} = +, \zeta_{\lambda}) + b_{00} Z(T_{0}, X_{i} = -, \zeta_{\lambda})
\]

Then we have the following equality

\[
p_{0}^{\zeta_{\lambda}} = P_{T}(X_{0} = + \mid \zeta_{\lambda})
\]

\[
= \frac{1}{1 + \frac{P_{T}(X_{0} = - \mid \zeta_{\lambda})}{P_{T}(X_{0} = + \mid \zeta_{\lambda})}} = \frac{1}{1 + 1/R_{0}^{\zeta_{\lambda}}}
\]

\[
= 1 + \lambda_{0} \prod_{(0,j) \in T} a_{00,j} R_{0,j}^{\zeta_{\lambda}} + b_{00,j}
\]

\[
= 1 + \lambda_{0} \prod_{(0,j) \in T} a_{00,j} R_{0,j}^{\zeta_{\lambda}} + b_{00,j}
\]

\[
= g_{0}(x_{0}),
\]

where \(x_{0} = (p_{0,1}^{\zeta_{\lambda}}, p_{0,2}^{\zeta_{\lambda}}, \cdots, p_{0,n}^{\zeta_{\lambda}})\). First, note that for any \(x = (x_{1}, x_{2}, \cdots, x_{q})\) and \(y = (y_{1}, y_{2}, \cdots, y_{q})\), the first order Taylor expansion at \(y\) gives that there exists a \(\theta \in [0, 1]\) such that

\[
g_{0}(x) - g_{0}(y) = \nabla g_{0}(y + \theta(x - y))(x - y)^{T},
\]

where \((x - y)^{T}\) denotes the transportation of the vector \((x - y)\). Careful calculations
lead to
\[
\frac{\partial g_0(x)}{\partial x_i} = -\frac{\lambda_0 \prod_{j=1}^{q} f_{00_j}(x_j) \left( \frac{d \log(f_{00_i}(x_i))}{dx_i} \right)}{(1 + \lambda_0 \prod_{j=1}^{q} f_{00_j}(x_j))^2} \\
= -g_0(x)(1 - g_0(x)) \left( \frac{M_{00_i}}{M_{00_i}x_i + d_{00_i}} - \frac{N_{00_i}}{N_{00_i}x_i + b_{00_i}} \right) \\
= g_0(x)(1 - g_0(x)) \left( \frac{a_{00_i}d_{00_i} - b_{00_i}c_{00_i}}{M_{00_i}x_i + d_{00_i}}(N_{00_i}x_i + b_{00_i}) \right) \\
= g_0(x)(1 - g_0(x))h_{00_i}(x_i).
\]

Hence, let \(x_0 = (p_{01}^\Lambda, p_{02}^\Lambda, \cdots, p_{0n_0}^\Lambda)\) and \(y_0 = (p_{01}^N, p_{02}^N, \cdots, p_{0n_0}^N)\), there exists \(\theta_0 \in [0, 1]\) such that
\[
|p_{0j}^\Lambda - p_{0j}^N| \leq \sum_{j=1}^{q} |g_0(z_0)(1 - g_0(z_0))h_{00_j}(x_j)||p_{0j}^\Lambda - p_{0j}^N| \\
\leq \sum_{j=1}^{q} g_0(z_0)(1 - g_0(z_0))\gamma_{00_j}|p_{0j}^\Lambda - p_{0j}^N| \\
\leq \gamma \sum_{j=1}^{q} g_0(z_0)(1 - g_0(z_0))|p_{0j}^\Lambda - p_{0j}^N|, \\
\]
where \(z_0 = x_0 + \theta_0(x_0 - y_0)\) and the second inequality follows from Lemma 1. Now repeat the procedure on the subtree \(T_{0j}\) for \(|p_{0j}^\Lambda - p_{0j}^N|, j = 1, 2, \cdots, q\) and so on. It is shown that the summation is over all the self-avoiding paths starting at the root 0. For each path \(\Gamma\), if the end point of \(\Gamma\) is a leaf \(j\) with \(d(0, j) \leq t - 1\) or there is a vertex \(i\) on \(\Gamma\) with \(d(0, i) \leq t - 1\) being fixed, the contribution of the path to the summation is zero since \(p_{i}^\Lambda - p_{i}^N = p_{j}^\Lambda - p_{j}^N = 0\). Hence the remaining path with length \(t\) is in the set \(\{\Gamma_{0k} : k \in S(T, 0, t)\}\). This completes the proof of lemma 2.

In order to complete the proof of Theorem 1, we need the following lemma.

**Lemma 3.** Let \(\mu_i \geq 0, i = 1, 2, \cdots, n\). Then
\[
\prod_{i=1}^{n} (1 + \mu_i) \geq (1 + \sqrt[n]{\prod_{i=1}^{n} \mu_i})^n.
\]
Proof. Consider
\[
\prod_{i=1}^{n} (1 + \mu_i) = 1 + \sum_{k=1}^{n} \left( \sum_{i_1 < i_2 < \cdots < i_k} \prod_{j=1}^{k} \mu_{ij} \right) \\
\geq 1 + \sum_{k=1}^{n} \left( C_n^k \left( \prod_{i=1}^{n} \mu_i \right) \frac{c_k-1}{c_n} \right) \\
= 1 + \sum_{k=1}^{n} \left( C_n^k \left( \prod_{i=1}^{n} \mu_i \right) \frac{c_k}{c_n} \right) \\
= (1 + \sqrt[n]{\prod_{i=1}^{n} \mu_i})^n;
\]
where \(C_n^k = \frac{n!}{k!(n-k)!}\). The first inequality uses the arithmetic-geometric average inequality. \(\square\)

With Lemma 2 and 3, the Theorem 1 is proved as follows.

**Proof of Theorem 1.** Following the notations of Lemma 2, let \(s = |S(T, 0, t)|\), we have
\[
|p_0^{\gamma^T} - p_0^{\gamma^T}| \leq \gamma^T \sum_{k \in S(T, 0, t)} \prod_{i \in \Gamma_{0k} \setminus i \neq k} g_i(z_i)(1 - g_i(z_i)) \\
\leq s \gamma^T \max_{k \in S(T, 0, t)} \prod_{i \in \Gamma_{0k} \setminus i \neq k} g_i(z_i)(1 - g_i(z_i)) \\
\leq s \gamma^T \frac{4}{2} \max_{(0, 0) \in T} \prod_{k \in S(T, 0, t)} g_k(z_k)(1 - g_k(z_k)).
\]

For each \(\Gamma_{0j, k}\), where \((0, 0) \in T\), \(k \in S(T, 0, t)\),
\[
\prod_{i \in \Gamma_{0j, k} \setminus i \neq k} g_i(z_i)(1 - g_i(z_i)) \\
= \prod_{i \in \Gamma_{0j, k} \setminus i \neq k} \lambda_i \prod_{(i, i_1) \in T_i} f_{i_l i_1}(z_{i_l i_1}) \prod_{(i, i_2) \in T_i} f_{i_l i_2}(z_{i_l i_2})^2 \\
\leq \left( \frac{r_{jk}}{(1 + r_{jk})^2} \right)^{t-1},
\]
where \(r_{jk} = \prod_{(i, i_1) \in T_i} \lambda_i \prod_{(i, i_2) \in T_i} f_{i_l i_2}(z_{i_l i_2})^{1/(t-1)}\) and the inequality above follows from Lemma 3. A simple calculation gives that \(e^{a_{\min}} \leq f_{ij}(x) \leq e^{a_{\max}}\) for any \((i, j) \in T\).
Hence,
\[
e^{\alpha_{\min}(\delta(T,0,t-1) - 1)} \leq \left( \prod_{i \in T_{0,k}} \prod_{(i,j) \in T_i} f_{ii}(z_{ij}) \right)^{1/(t-1)} \leq e^{\alpha_{\max}(\delta(T,0,t-1) - 1)}.
\]

Now we prove the exponential strong spatial mixing under the assumption of Theorem 1. Suppose $\Gamma$ is a self-avoiding path of $G$. Note that each self-avoiding path on $T$ without its ending point is also a self-avoiding path on $G$. From proposition 1, it is known $0$ is a vertex of $G$ and let $\tilde{p}_0^\Lambda = P_G(X_0 = +|\zeta_\Lambda)$. By proposition 2, we know $\delta(T,0,t-1) \leq \Delta(G, t) \leq d$ when $t = ka \log n + 1$, $k = 1, 2, \cdots$. If $B_{\min} > B(d, \alpha_{\max}, \gamma)$,
\[
\frac{\gamma(d-1)\exp(2B_{\min} - \alpha_{\max}(d-1))}{(1 + \exp(2B_{\min} - \alpha_{\max}(d-1)))^2} < 1.
\]
By proposition 2, there is $s \leq \delta_0(d-1)^{t-1}$. Note that $\left( \prod_{i \in T_{0,k}} \lambda_i \right)^{1/(t-1)} \leq e^{-2B_{\min}}$, we have
\[
|\tilde{p}_0^\Lambda - \tilde{p}_0^\eta| = |\tilde{p}_0^\Lambda - p_0^\eta| \leq s \frac{\gamma}{4} \left( 1 + r_{jk} \right)^{t-1} \leq \delta_0 \gamma \left( \frac{\gamma(d-1)\exp(2B_{\min} - \alpha_{\max}(d-1))}{(1 + \exp(2B_{\min} - \alpha_{\max}(d-1)))^2} \right)^{t-1},
\]
where the first equality follows from the proposition 1. The similar case holds for $B_{\max} < -B(d, -\alpha_{\min}, \gamma)$. This completes the proof. □

**Remark:** It is not difficult to apply Theorem 1 to Erdős-Rényi random graph $G(n, d/n)$, where each potential edge is chosen independently with probability $d/n$ [10]. A proof is sketched here. By Theorem 1, what we need to show is, for $G(n, d/n)$, there exist two positive numbers $a$ and $c$ such that $\Delta(G, a \log n) \leq c$ with a high probability (e.g. with the probability $1 - o(1)$ as $n$ goes to infinity). By Lemma 1 in [10], we know that, for $0 < a < \frac{1}{2 \log d}$, there exists some $c(a, d)$ such that $m(G, v, a \log n) \leq c \log n$ for all $v \in G(n, d/n)$ with a high probability. Hence by the definition of $\delta(G, v, l)$, for all $v \in G$, with a high probability, $\delta(G, v, a \log n) = (m(G, v, a \log n) - \delta_0) / a \log n \leq c / a$, then $\Delta(G, a \log n) = \max_v \delta(G, v, a \log n) \leq c / a$. Therefore the results of Theorem 1 is true for the random graph $G(n, d/n)$.

Now we proceed to prove Corollary 1.

**Proof of Corollary 1.** Following the previous notations, by the formula (1) in Lemma 2, we have
\[
|p_0^\zeta - p_0^\eta| \leq \gamma \sum_{j=1}^q g_0(z_0)(1 - g_0(z_0))|p_0^\zeta - p_0^\eta|.
\]
Without loss of generality, suppose the degree of 0 is $d - 1$. Then

$$|p_0^\Lambda - p_0^{\Omega_\Lambda}| \leq \gamma(d - 1) \max_{i \in T}(g_0(z_i)(1 - g_0(z_i))) \max_{j \in \{1, 2, \ldots, d - 1\}} |p_0^\Lambda - p_0^{\Omega_j}|.$$

If $B_i > B(d, \alpha_{\max}, \gamma)$ or $B_i < -B(d, -\alpha_{\min}, \gamma)$, $\gamma(d - 1)g_i(z_i)(1 - g_i(z_i)) < 1$ for any $i \in T$. Hence by induction on the height $t$, there is

$$|p_0^\Lambda - p_0^{\Omega_\Lambda}| \leq (\gamma(d - 1) \max_{i \in T}(g_0(z_i)(1 - g_0(z_i))))^t.$$

Since the degree of 0 is at most $d$, we have

$$|p_0^\Lambda - p_0^{\Omega_\Lambda}| \leq \gamma d g_0(z_0)(1 - g_0(z_0))(\gamma(d - 1) \max_{i \in T}(g_0(z_i)(1 - g_0(z_i))))^{t-1}
\leq \frac{\gamma d}{4} (\gamma(d - 1) \max_{i \in T}(g_0(z_i)(1 - g_0(z_i))))^{t-1}.$$

Figure 3: The red line and the blue line denote the “external field” curves for uniqueness of Gibbs measures and for strong spatial mixing on three regular trees $T$ respectively, where $\beta_{ij}(\sigma_i, \sigma_j) = J \sigma_i \sigma_j$ and $h_i(\sigma_i) = B \sigma_i$, for any $i \in T$ and $(i, j) \in T$.

The proof is completed by applying proposition 1. □

Remark: We would emphasize here that the tighter bound of $f_{ij}(x)$ is the key to improve the result since a better bound of $f_{ij}(x)$ will give a better bound for $g_i(x)$. We do not optimize the parameter here. As we know strong spatial mixing implies uniqueness of Gibbs measures[14]. We are not sure whether Lipchitz method can make $B(d, \alpha_{\max}, \gamma)$ or $-B(d, -\alpha_{\min}, \gamma)$ optimally approximate the critical points of the ‘external field’ for the uniqueness of Gibbs measures of BMRF. Note that the
critical points of ‘external field’ for ferromagnetic and antiferromagnetic Ising model are different on an infinite $d$ regular tree [4]). We do not expect the critical external field for Ising model on the $d$ regular tree for uniqueness of Gibbs measures is the optimal external field for strong spatial mixing. The intuition for this is that the uniqueness of Gibbs measures on the tree is equivalent to weak spatial mixing (see [3, 13] for definitions) in some sense [9]. If some configurations are close to the root (note that some configurations may be at the height 2 or 3, see Figure 1 and 2, when self-avoiding tree is constructed), the perturbation of the boundary condition changes the Gibbs measures at the root radically. More precisely, strong spatial mixing can be deducted to the weak spatial mixing by removing the support of unmodified boundary configuration and changing the external field of some vertices (see Lemma 2 in [14]). Hence the critical external field condition for weak spatial mixing does not hold for strong spatial mixing. Figure 3 illustrates the curve of external field under our condition for strong spatial mixing and the critical external field for uniqueness of Gibbs measures on the infinite $d$ regular tree, where $d = 3$.

5 Conclusions and Discussion

The Gibbs distribution on a graph $G = (V, E)$ with ‘maximum average degree’ $d$ is considered in this paper. The (exponential) strong spatial mixing is proved for such systems, when the ‘external field’ $B_i$ is uniformly larger than $B(d, \sigma_{\max}, \gamma)$ or smaller than $-B(d, -\sigma_{\min}, \gamma)$. Here $B(d, \sigma, \gamma)$ is a function with parameter $d, \sigma, \gamma$.

For future work, it should be possible to improve the condition on ‘external field’. We have emphasized the essential points in the last remark of section 4. However, we believe that some different methods other than Lipchitz method are required. The fixed point method in [7] might be a promising approach.

The case of non-binary is the most challenging question for sure. Nair and Tetali [11] present a similar construction of self-avoiding tree for general Markov random fields (multi-spin systems). We hope our work on BMRF could be a step to the case of general Markov random fields.

Acknowledgment

The authors would like to thank the anonymous referees for their very helpful comments and suggestions.

References


