Uniqueness, Spatial Mixing, and Approximation for Ferromagnetic 2-Spin Systems

HENG GUO, University of Edinburgh, UK

PINYAN LU, Shanghai University of Finance and Economics, China

We give fully polynomial-time approximation schemes (FPTAS) for the partition function of ferromagnetic 2-spin systems in certain parameter regimes. The threshold we obtain is almost tight up to an integrality gap. Our technique is based on the correlation decay framework. The main technical contribution is a new potential function, with which we establish a new kind of spatial mixing.

CCS Concepts: • **Theory of computation** → **Approximation algorithms analysis**;

Additional Key Words and Phrases: Approximate counting, correlation decay, spin systems

17

1 INTRODUCTION

Spin systems model nearest-neighbor interactions. In this article, we study two-state spin systems. An instance is a graph $G = (V, E)$, and a configuration $\sigma : V \to \{0, 1\}$ assigns one of the two spins "0" and "1" to each vertex. The local interaction along an edge is specified by a matrix $\mathbf{A} = \begin{bmatrix} A_{0,0} & A_{0,1} \\ A_{1,0} & A_{1,1} \end{bmatrix}$, where $A_{i,j}$ is the (non-negative) local weight when the two endpoints are assigned i and j, respectively. We study symmetric edge interactions, that is, $A_{0,1} = A_{1,0}$. Normalize $\mathbf{A}$ so that $\mathbf{A} = \begin{bmatrix} \beta & 1 \\ 1 & \gamma \end{bmatrix}$. Moreover, we also consider the external field, specified by a mapping $\pi : V \to \mathbb{R}^+$. When a vertex is assigned "0," we give it a weight $\pi(v)$. For a particular configuration σ, its weight $w(\sigma)$ is a product over all edge interactions and vertex weights, that is

$$w(\sigma) = \beta^{m_0(\sigma)} \gamma^{m_1(\sigma)} \prod_{v \mid \sigma(v)=0} \pi(v),$$

where $m_0(\sigma)$ is the number of $(0, 0)$ edges under the configuration σ and $m_1(\sigma)$ is the number of $(1, 1)$ edges. An important special case is the Ising model, where $\beta = \gamma$. The Gibbs measure is a natural distribution in which each configuration σ is drawn with probability proportional to its weight, that is, $\Pr_{G;\beta,\gamma,\pi}(\sigma) \sim w(\sigma)$. The normalizing factor of the Gibbs measure is called

the partition function, defined by $Z_{\beta,\gamma,\pi}(G) = \sum_{\sigma:V\to\{0,1\}} w(\sigma)$. The partition function encodes rich information regarding the macroscopic behavior of the spin system. We will be interested in the computational complexity of approximating $Z_{\beta,\gamma,\pi}(G)$. We also simply write $Z_{\beta,\gamma,\lambda}(G)$ when the field is uniform, that is, $\pi(v) = \lambda$ for all $v \in V$. A system with uniform fields is specified by the three parameters (β,γ,λ).

Spin systems not only are interesting in statistical physics but also find applications in computer science, under the name of *Markov random fields*. Indeed, a two-state spin system is equivalent to a binary Markov random field. For example, Boltzmann Machines [1] can be viewed as a special case where $\gamma = 1$. Computing the partition function is a central task in statistical inference. According to their physical and computational properties, spin systems can be classified into two families: *ferromagnetic* systems, where the edge interaction is attractive ($\beta\gamma > 1$), and *anti-ferromagnetic* systems, where it is repulsive ($\beta\gamma < 1$).

Recently, beautiful connections have been established regarding three different properties of anti-ferromagnetic two-spin systems: tree uniqueness, spatial mixing, and computational complexity transitions. The uniqueness of Gibbs measures in infinite regular trees[1] of degrees up to Δ implies correlation decay[2] in all graphs of maximum degree Δ, and therefore the existence of fully polynomial-time approximation scheme (FPTAS) for the partition function [14, 15, 20, 23]. However, if the tree uniqueness fails, then long range correlation appears and the partition function has no fully polynomial-time randomized approximation scheme (FPRAS) unless NP = RP [5, 21, 22]. It suggests that tree uniqueness, spatial mixing, and the computational complexity of approximating the partition function, line up perfectly in the anti-ferromagnetic regime.

For ferromagnetic systems, the picture is much less clear. In a seminal paper [11], Jerrum and Sinclair gave an FPRAS for the ferromagnetic Ising model $\beta = \gamma > 1$ with any consistent external field for general graphs without degree bounds. Thus, there is no computational complexity transition of approximating these models, whereas uniqueness and spatial mixing do exhibit phase transition. This is in sharp contrast to anti-ferromagnetic Ising models $\beta = \gamma < 1$, where computational and phase transitions align perfectly. It is not clear at all whether spatial mixing or correlation decay plays any role in the computational complexity.

For more general ferromagnetic two-spin systems with external fields, the threshold of efficiently approximating the partition function is still open. On the complexity side, Goldberg and Jerrum showed that any ferromagnetic two-spin system is no harder than counting independent sets in bipartite graphs (#BIS) [7], which is conjectured to have no FPRAS [4] (the approximation complexity of #BIS is still open). Based on an earlier result [2], Liu, Lu and Zhang showed that approximating the partition function is #BIS-hard if we allow external fields beyond $(\gamma/\beta)^{\frac{\lfloor\Delta_c\rfloor+2}{2}}$, where $\Delta_c = \frac{\sqrt{\beta\gamma}+1}{\sqrt{\beta\gamma}-1}$ [16].[3]

On the algorithmic side, by reducing to the Ising model, an MCMC-based FPRAS is known for the range of $\lambda \le \gamma/\beta$ [16] (improving upon Reference [8]). However, if we apply the correlation decay algorithmic framework to various pairs of parameters (β,γ), it is not hard to get bounds better than γ/β. However, such success for individual problems does not seem to share meaningful inner connections. In particular, it is not clear how far one can push this method, and to the best of our knowledge, no threshold has even been conjectured.

---

[1]This property is called "tree uniqueness" or "uniqueness" for short. See Sections 2.2 and 6.1 for details.
[2]That is, the correlation of any two vertices decay exponentially in distance. This is also known as "spatial mixing."
[3]Here and below, we assume $\beta \le \gamma$ due to symmetry.

1.1 Our Contribution

In this article, we identify a new threshold that almost tightly maps out the boundary of the correlation decay regime, that is, $\lambda_c := (\gamma/\beta)^{\frac{\Delta_c+1}{2}} = (\gamma/\beta)^{\frac{\sqrt{\beta\gamma}}{\sqrt{\beta\gamma}-1}}$. We show that for any $\lambda < \lambda_c$ a variant of spatial mixing holds (Theorem 1.1) for arbitrary trees. An interesting feature of our work is that we do not restrict the degree or the shape of the tree. This is almost tight, since it does not hold if $\lambda > (\gamma/\beta)^{\frac{\lceil\Delta_c\rceil+1}{2}}$. This spatial mixing is weaker than what an algorithm usually requires, but in the regime of $\beta \leq 1$ it implies (and therefore is equivalent to) strong spatial mixing. As an algorithmic consequence, we have FPTAS for all $\beta \leq 1 < \gamma$, $\beta\gamma > 1$, and $\lambda < \lambda_c$ (Theorem 1.2). Recall that if we allow λ beyond $(\gamma/\beta)^{\frac{\lceil\Delta_c\rceil+2}{2}}$, then the problem is #BIS-hard [16]. Hence, only an integral gap remains for the $\beta \leq 1 < \gamma$ case.

Formally, let p_v be the marginal probability of v being assigned "0."

THEOREM 1.1. *Let (β, γ, λ) be a set of parameters of the system such that $\beta\gamma > 1, \beta \leq \gamma$, and $\lambda < \lambda_c$. Let T_v and $T'_{v'}$ be two trees with roots v and v', respectively. If the two trees have the same structure in the first ℓ levels, then $|p_v - p_{v'}| \leq O(\exp(-\ell))$.*

In other words, if we simply truncate a tree at depth ℓ, then the marginal probability of its root will change by only at most $O(\exp(-\ell))$. Surprisingly, if we replace λ_c by its integral counterpart, then this implication no longer holds and there is a counterexample (see Section 5). More precisely, it is no longer true that the uniqueness in infinite regular trees implies correlation decay in graphs or even trees, since our counterexample is an irregular tree. We note that this is in sharp contrast to anti-ferromagnetic systems, where (integral) uniqueness implies correlation decay.

From the computational complexity point of view, we would like to get an FPTAS for the partition function, which requires a condition called strong spatial mixing (SSM). It is stronger than the spatial mixing established in Theorem 1.1 by imposing arbitrary partial configurations. We are able to prove SSM with $\lambda < \lambda_c$ in the range of $\beta \leq 1$. Indeed, if $\beta \leq 1$, then the two versions of spatial mixing are equivalent. Let I be an interval of the form $[\lambda_1, \lambda_2]$ or $(\lambda_1, \lambda_2]$. We consider the following problem.

> **Name** #2SPIN(β, γ, I)
> **Instance** A graph $G = (V, E)$ and a mapping $\pi : V \to \mathbb{R}^+$, such that $\pi(v) \in I$ for any $v \in V$.
> **Output** $Z_{\beta, \gamma, \pi}(G)$.

Then, we have the following theorem.

THEOREM 1.2. *Let (β, γ, λ) be a set of parameters of the system such that $\beta\gamma > 1, \beta \leq 1$ and $\lambda < \lambda_c$. There is an FPTAS for #2SPIN$(\beta, \gamma, (0, \lambda])$.*

Therefore, we get an almost tight dichotomy for ferromagnetic two-spin systems when $\beta \leq 1$, since #2SPIN$(\beta, \gamma, (0, \lambda])$ is #BIS-hard, if λ is larger than the integral counterpart of λ_c [16] (see also Proposition 5.1).

The reason behind λ_c is a nice interplay among uniqueness, spatial mixing, and approximability. We start with some purely mathematical observations on the symmetric tree recursion $f_d(x) = \lambda(\frac{\beta x+1}{x+\gamma})^d$, an increasing function in x. Relax the range of d in $f_d(x)$ to be real numbers. Then Δ_c is the critical (possibly fractional) degree and λ_c is the corresponding critical external field for the recursion to have a unique fixed point. This set of critical parameters enjoys some very nice mathematical properties. For $d = \Delta_c$ and $\lambda = \lambda_c$, the function $f_d(x)$ has a unique fixed point $\widehat{x} = \sqrt{\gamma/\beta}$ and $f'_d(\widehat{x}) = 1$. Moreover, it also satisfies that $f''_d(\widehat{x}) = 0$, which is a necessary condition

for the contraction of the tree recursion. (This is easy to derive using the heuristic of finding potential functions described in Reference [15].) All these nice mathematical properties prove to be useful in our later analysis. For degrees other than Δ_c, their critical external fields are much less convenient— the function $f_d(x)$ has two fixed points: one is crossing and the other is tangent. Moreover, $f_d''(\hat{x}) = 0$ does not necessarily hold.

The proof of Theorem 1.1 uses the potential method to analyze decay of correlation, which is now streamlined (see, e.g., References [15, 19]). The main difficulty is to find a good potential function. In other words, we want to solve a variational problem minimizing the maximum of the decay rate function. The main novelty in our solution is that we restrict variables to the range of $(0, \frac{\lambda}{1+\lambda}]$ and our potential function is well-defined *only* in this range. This is, in fact, necessary, as otherwise the statement does not hold, and is valid for the setting of Theorem 1.1. Our choice leads to a relatively clean and significantly simpler proof, comparing to similar proofs in other settings. In particular, we do not need the "symmetrization" argument (see, e.g., References [15, 19]). We also use a trick of truncating the potential to deal with unbounded degrees (see Equation (8)).

For the range of $\beta > 1$, SSM does not hold even if $\lambda < \lambda_c$. However, we conjecture that Theorem 1.2 can be extended to the $\beta > 1$ range as well, mainly due to Theorem 1.1, which does not require $\beta \le 1$. Moreover, we show that even if $\beta > 1$, the marginal probability in any instance is within the range of $(0, \frac{\lambda}{1+\lambda}]$ given $\lambda < \lambda_c$ (see Proposition 3.8). This seems to imply that the main reason why our algorithm fails is due to pinnings (forcing a vertex to be "0" or "1") in the self-avoiding walk tree construction, whereas in a real instance these pinnings cannot aggregate enough "bad" influence. However, to turn such intuition into an algorithm requires a careful treatment of these pinnings to achieve an FPTAS without SSM. We leave this as an important open problem.

At last, we note that neither λ_c nor its integral counterpart is the exact threshold in each own respect, even if $\beta \le 1$. Strong spatial mixing continues to hold even if $\lambda > \lambda_c$ in a small interval. We give a concrete example to illustrate this fact in Section 4, Proposition 4.1. Moreover, as mentioned earlier, an irregular tree exists where the correlation decay threshold is lower than the threshold for all infinite regular trees. This is discussed in Section 5. It is another important open question to figure out the exact threshold between λ_c and its integral counterpart(s).

2 PRELIMINARIES

An instance of a two-spin system is a graph $G = (V, E)$. A configuration $\sigma : V \to \{0, 1\}$ assigns one of the two spins "0" and "1" to each vertex. We normalize the edge interaction to be $\begin{bmatrix} \beta & 1 \\ 1 & \gamma \end{bmatrix}$, and also consider the external field, specified by a mapping $\pi : V \to \mathbb{R}^+$. When a vertex is assigned "0," we give it a weight $\pi(v)$. All parameters are non-negative. For a particular configuration σ, its weight $w(\sigma)$ is a product over all edge interactions and vertex weights, that is

$$w(\sigma) = \beta^{m_0(\sigma)} \gamma^{m_1(\sigma)} \prod_{v \mid \sigma(v) = 1} \pi(v), \tag{1}$$

where $m_0(\sigma)$ is the number of $(0, 0)$ edges given by the configuration σ and $m_1(\sigma)$ is the number of $(1, 1)$ edges. An important special case is the Ising model, where $\beta = \gamma$. Notice that in the statistic physics literature, parameters are usually chosen to be the logarithms of our parameters above. Different parameterizations do not affect the complexity of the same system.

We also write $\lambda_v := \pi(v)$. If π is a constant function such that $\lambda_v = \lambda > 0$ for all $v \in V$, then we also denote it by λ. We say π has a lower bound (or an upper bound) $\lambda > 0$, if π satisfies the guarantee that $\lambda_v \ge \lambda$ (or $\lambda_v \le \lambda$).

The Gibbs measure is a natural distribution in which each configuration σ is drawn with probability proportional to its weight, that is,

$$\Pr_{G;\beta,\gamma,\pi}(\sigma) \propto w(\sigma). \tag{2}$$

The normalizing factor of the Gibbs measure is called the partition function, defined by $Z_{\beta,\gamma,\pi}(G) = \sum_{\sigma:V\to\{0,1\}} w(\sigma)$. Recall that we are interested in the computational problem #2Spin(β,γ,I), where I is an interval of the form $[\lambda_1,\lambda_2]$ or (λ_1,λ_2), for which $Z_{\beta,\gamma,\pi}(G)$ is the output. When input graphs are restricted to have a degree bound Δ, we write #Δ-2Spin(β,γ,I) to denote the problem. When the field is uniform, that is, λ is the only element in I, we simply write #2Spin(β,γ,λ). Due to Reference [3] and a standard diagonal transformation, for any constant $\lambda > 0$, #2Spin(β,γ,λ) is #P-hard unless $\beta = \gamma = 0$ or $\beta\gamma = 1$.

2.1 The Self-Avoiding Walk Tree

We briefly describe Weitz's algorithm [23]. Our algorithms presented later will follow roughly the same paradigm.

The Gibbs measure defines a marginal distribution of spins for each vertex. Let p_v denote the probability of a vertex v being assigned "0". Since the system is self-reducible, #2Spin(β,γ,λ) is equivalent to computing p_v for any vertex v [12] (for details, see, for example, Lemma 2.6).

Let $\sigma_\Lambda \in \{0,1\}^\Lambda$ be a configuration of $\Lambda \subset V$. We call vertices in Λ *fixed* and other vertices *free*. We use $p_v^{\sigma_\Lambda}$ to denote the marginal probability of v being assigned "0" conditional on the configuration σ_Λ of Λ.

Suppose the instance is a tree T with root v. Let $R_T^{\sigma_\Lambda} := p_v^{\sigma_\Lambda}/(1 - p_v^{\sigma_\Lambda})$ be the ratio between the two probabilities that the root v is 0 and 1, while imposing some condition σ_Λ (with the convention that $R_T^{\sigma_\Lambda} = \infty$ when $p_v^{\sigma_\Lambda} = 1$). Suppose that v has d children $v_i, \ldots v_d$. Let T_i be the subtree with root v_i. Due to the independence of subtrees, it is straightforward to get the following recursion for calculating $R_T^{\sigma_\Lambda}$:

$$R_T^{\sigma_\Lambda} = F_d\left(R_{T_1}^{\sigma_\Lambda}, \ldots, R_{T_d}^{\sigma_\Lambda}\right), \tag{3}$$

where the function $F_d(x_1, \ldots, x_d)$ is defined as

$$F_d(x_1, \ldots, x_d) := \lambda_v \prod_{i=1}^{d} \frac{\beta x_i + 1}{x_i + \gamma}.$$

We allow x_i's to take the value ∞ as in that case the function F_d is clearly well defined. In general, we use capital letters like $F, G, C, \ldots$ to denote multivariate functions and small letters $f, g, c, \ldots$ to denote their symmetric versions, where all variables take the same value. Here, we define $f_d(x) := \lambda\left(\frac{\beta x+1}{x+\gamma}\right)^d$ to be the symmetric version of $F_d(\mathbf{x})$ obtained by setting $x_1 = \cdots = x_d = x$.

Let $G(V, E)$ be a graph. Similarly define $R_{G,v}^{\sigma_\Lambda} := p_v^{\sigma_\Lambda}/(1 - p_v^{\sigma_\Lambda})$. In contrast to the case of trees, there is no easy recursion to calculate $R_{G,v}^{\sigma_\Lambda}$ for a general graph G. This is because of dependencies introduced by cycles. Weitz [23] reduced computing the marginal distribution of v in a general graph G to that in a tree, called the self-avoiding walk (SAW) tree, denoted by $T_{SAW}(G, v)$. To be specific, given a graph $G = (V, E)$ and a vertex $v \in V$, $T_{SAW}(G, v)$ is a tree with root v that enumerates all self-avoiding walks originating from v in G, with additional vertices closing cycles as leaves of the tree. Each vertex in the new vertex set V_{SAW} of $T_{SAW}(G, v)$ corresponds to a vertex in G, but a vertex in G may be mapped to more than one vertices in V_{SAW}. A boundary condition is imposed on leaves in V_{SAW} that close cycles. The imposed colors of such leaves depend on whether the cycle is formed from a small vertex to a large vertex or conversely, where the ordering is arbitrarily chosen in G. Vertex sets $S \subset \Lambda \subset V$ are mapped to, respectively, $S_{SAW} \subset \Lambda_{SAW} \subset V_{SAW}$,

and any configuration $\sigma_\Lambda \in \{0, 1\}^\Lambda$ is mapped to $\sigma_{\Lambda_{\text{SAW}}} \in \{0, 1\}^{\Lambda_{\text{SAW}}}$. With slight abuse of notations, we may write $S = S_{\text{SAW}}$ and $\sigma_\Lambda = \sigma_{\Lambda_{\text{SAW}}}$ when no ambiguity is caused.

PROPOSITION 2.1 (THEOREM 3.1 OF WEITZ [23]). *Let $G = (V, E)$ be a graph, $v \in V$, $\sigma_\Lambda \in \{0, 1\}^\Lambda$ be a configuration on $\Lambda \subset V$, and $S \subset V$. Let $T = T_{\text{SAW}}(G, v)$ be constructed as above. It holds that*

$$R^{\sigma_\Lambda}_{G, v} = R^{\sigma_\Lambda}_T.$$

Moreover, the maximum degree of T is at most the maximum degree of G, $\text{dist}_G(v, S) = \text{dist}_T(v, S_{\text{SAW}})$, and any neighborhood of v in T can be constructed in time proportional to the size of the neighborhood.

The SAW tree construction does not solve a #P-hard problem, since $T_{\text{SAW}}(G, v)$ is potentially exponentially large in size of G. For a polynomial time approximation algorithm, we may run the tree recursion within some polynomial size, or equivalently a logarithmic depth. At the boundary where we stop, we plug in some arbitrary values. The question is then how large is the error due to our random guess. To guarantee the performance of the algorithm, we need the following notion of *strong spatial mixing*.

Definition 2.2. A spin system on a family $\mathcal{G}$ of graphs is said to exhibit *strong spatial mixing* (SSM) if for any graph $G = (V, E) \in \mathcal{G}$, any $v \in V, \Lambda \subset V$ and any $\sigma_\Lambda, \tau_\Lambda \in \{0, 1\}^\Lambda$,

$$|p_v^{\sigma_\Lambda} - p_v^{\tau_\Lambda}| \le \exp(-\Omega(\text{dist}(v, S))), \tag{4}$$

where $S \subset \Lambda$ is the subset on which σ_Λ and τ_Λ differ, and $\text{dist}(v, S)$ is the shortest distance from v to any vertex in S.

Weak spatial mixing is defined similarly by replacing $\text{dist}(v, S)$ with $\text{dist}(v, \Lambda)$ in Equation (4), and it corresponds to the uniqueness condition introduced below in Section 2.2. Spatial mixing properties are also called correlation decay in statistical physics.

If SSM holds, then the error caused by early termination in $T_{\text{SAW}}(G, v)$ and arbitrary boundary values is only exponentially small in the depth. Hence, the algorithm is an FPTAS. In a lot of cases, the existence of an FPTAS boils down to establish SSM.
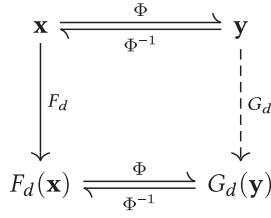
2.2 The Uniqueness Condition in Regular Trees

Let $\mathbb{T}_d$ denote the infinite d-regular tree, also known as the *Bethe lattice* or the *Cayley tree*. If we pick an arbitrary vertex as the root of $\mathbb{T}_d$, then the root has d children and every other vertex has $d - 1$ children. Notice that the difference between $\mathbb{T}_d$ and an infinite $(d - 1)$-ary tree is only the degree of the root. We assume in this subsection that the field is uniform $\lambda > 0$.

The Gibbs measure in $\mathbb{T}_d$ is a probability measure where conditioned on any arbitrary configuration on the boundary of a finite set S, the resulting distribution on S is the same as the one given by Equation (2) on S with the same boundary condition. When the Gibbs measure is unique, we say that *the uniqueness condition* holds in $\mathbb{T}_d$. The tree recursion Equation (3) turns out to be important in analyzing the uniqueness condition. Due to the symmetric structure of $\mathbb{T}_d$, it becomes $R_v = f_{d-1}(R_{v_i})$ (for any vertex v other than the root), where $f_d(x) = \lambda(\frac{\beta x + 1}{x + \gamma})^d$ is the symmetrized version of $F_d(\mathbf{x})$.

For anti-ferromagnetic systems, that is, $\beta\gamma < 1$, there is a unique fixed point to $f_d(x) = x$, denoted by $\widehat{x}$. It has been shown that the Gibbs measure in $\mathbb{T}_d$ is unique if and only if $|f'_{d-1}(\widehat{x})| \le 1$ [6, 13].

In contrast, if $\beta\gamma > 1$, then there may be 1 or 3 positive fixed points such that $x = f_d(x)$. It is known [6, 13] that the Gibbs measure of two-state spin systems in $\mathbb{T}_d$ is unique if and only if there is only one fixed point for $x = f_{d-1}(x)$, or equivalently, for all fixed points $\widehat{x}_d$ of $f_d(x)$, $f'_d(\widehat{x}_d) < 1$.

Fig. 1. Commutative diagram between F_d and G_d.

Let $\Delta_c := \frac{\sqrt{\beta\gamma}+1}{\sqrt{\beta\gamma}-1}$. Then, we have the following result.

PROPOSITION 2.3. *If $\Delta - 1 < \Delta_c$, then the uniqueness condition in $\mathbb{T}_\Delta$ holds regardless of the field.*

Note that the condition $\Delta - 1 < \Delta_c$ matches the exact threshold of fast mixing for Glauber dynamics in the Ising model [18]. In Section 3.1, we will show that, SSM holds and there exists an FPTAS for the partition function, in graphs with degree bound $\Delta < \Delta_c + 1$ and any field $\lambda > 0$. This is Theorem 3.1.

To study general graphs, one needs to consider infinite regular trees of all degrees. If $\beta > 1$ (still assuming $\beta\gamma > 1$ and $\beta \leq \gamma$), then there is no λ such that the uniqueness condition holds in $\mathbb{T}_d$ for all degrees $d \geq 2$. In contrast, let $\lambda_c^{int} := (\gamma/\beta)^{\frac{\lceil\Delta_c\rceil+1}{2}}$ and for $\beta \leq 1$, we have the following.

PROPOSITION 2.4. *Let (β, γ) be two parameters such that $\beta\gamma > 1$.*

- *If $\beta \leq 1 < \gamma$, then uniqueness holds for $\mathbb{T}_d$ with all degrees $d \geq 2$ if and only if $\lambda < \lambda_c^{int}$.*
- *If $\beta, \gamma > 1$, then there is no $\lambda > 0$ such that uniqueness holds for all $\mathbb{T}_d$ with $d \geq 2$.*

However, there exist (β, γ, λ) and an (irregular) tree T such that $\beta\gamma > 1$, $\beta \leq 1 < \gamma$, and $\lambda < \lambda_c^{int}$ and SSM does not hold in T. This is discussed in Section 5. Recall that $\lambda_c := (\gamma/\beta)^{\frac{\Delta_c+1}{2}}$. If we replace λ_c^{int} with $\lambda_c \leq \lambda_c^{int}$ in the condition of Proposition 2.4, that is, $\beta\gamma > 1$, $\beta \leq 1 < \gamma$, and $\lambda < \lambda_c$, then SSM holds in all graphs and an FPTAS exists. This is shown in Section 3.2, Theorem 3.6.

Details and proofs about Propositions 2.3 and 2.4 are given in Section 6.1.

2.3 The Potential Method

We would like to prove the strong spatial mixing in arbitrary trees, sometimes with bounded degree Δ, under certain conditions. This is sufficient for approximation algorithms due to the self-avoiding walk tree construction. Our main technique in the analysis is the potential method. The analysis in this section is a standard routine, with some specialization to ferromagnetic 2-spin models (cf. References [15, 19]). To avoid interrupting the flow, we move all details and proofs to Section 6.2.

Roughly speaking, instead of studying Equation (3) directly, we use a potential function $\Phi(x)$ to map the original recursion to a new domain (see the commutative diagram Figure 1). Morally, we can choose whatever function as the potential function. However, we would like to pick "good" ones to help the analysis of the contraction. Define $\varphi(x) := \Phi'(x)$ and

$$C_{\varphi,d}(\mathbf{x}) := \varphi(F_d(\mathbf{x})) \cdot \sum_{i=1}^{d} \left|\frac{\partial F_d}{\partial x_i}\right| \frac{1}{\varphi(x_i)}.$$

Definition 2.5. Let $\Phi : \mathbb{R}^+ \to \mathbb{R}^+$ be a differentiable and monotonically increasing function. Let $\varphi(x)$ and $C_{\varphi,d}(\mathbf{x})$ be defined as above. Then $\Phi(x)$ is a *good potential function for degree d and field λ* if it satisfies the following conditions:

(1) there exists a constant $C_1, C_2 > 0$ such that $C_1 \leq \varphi(x) \leq C_2$ for all $x \in [\lambda \gamma^{-d}, \lambda \beta^d]$;

(2) there exists a constant $\alpha < 1$ such that $C_{\varphi, d}(\mathbf{x}) \leq \alpha$ for all $x_i \in [\lambda \gamma^{-d}, \lambda \beta^d]$.

We say $\Phi(x)$ is a good potential function for d and field π, if $\Phi(x)$ is a good potential function for d and any λ in the codomain of π,

In Definition 2.5, Condition 1 is rather easy to satisfy. The crux is in fact Condition 2. We call α in Condition 2 the amortized contraction ratio of $\Phi(x)$. It has the following algorithmic implication. The proof is based on establishing strong spatial mixing.

LEMMA 2.6. *Let (β, γ) be two parameters such that $\beta \gamma > 1$. Let $G = (V, E)$ be a graph with a maximum degree Δ and n many vertices and π be a field on G. Let $\lambda = \max_{v \in V}\{\pi(v)\}$. If there exists a good potential function for π and all $d \in [1, \Delta - 1]$ with contraction ratio $\alpha < 1$, then $Z_{\beta, \gamma, \pi}(G)$ can be approximated deterministically within a relative error ε in time $O(n(\frac{n\lambda}{\varepsilon})^{\frac{\log(\Delta-1)}{-\log \alpha}})$.*

When the degree is unbounded, the SAW tree may grow super polynomially even if the depth is of order $\log n$. We use a refined metric replacing the naive graph distance used in Definition 2.2. Strong spatial mixing under this metric is also called *computationally efficient correlation decay* [14, 15].

Definition 2.7. Let T be a rooted tree and $M > 1$ be a constant. For any vertex v in T, define the *M-based depth* of v, denoted $\ell_M(v)$, such that $\ell_M(v) = 0$ if v is the root, and $\ell_M(v) = \ell_M(u) + \lceil \log_M(d + 1) \rceil$ if v is a child of u and u has degree d.

Let $B(\ell)$ be the set of all vertices whose M-based depths of v is at most ℓ. It is easy to verify inductively such that $|B(\ell)| \leq M^\ell$ in a tree. We then define a slightly stronger notion of potential functions.

Definition 2.8. Let $\Phi : \mathbb{R}^+ \to \mathbb{R}^+$ be a differentiable and monotonically increasing function. Let $\varphi(x)$ and $C_{\varphi, d}(\mathbf{x})$ defined in the same way as in Definition 2.5. Then, $\Phi(x)$ is a *universal potential function* for the field λ if it satisfies the following conditions:

(1) there are two constants $C_1, C_2 > 0$ such that $C_1 \leq \varphi(x) \leq C_2$ for any $x \in (0, \lambda]$;

(2) there exists a constant $\alpha < 1$ such that for all d, $C_{\varphi, d}(\mathbf{x}) \leq \alpha^{\lceil \log_M(d+1) \rceil}$ for all $x_i \in (0, \lambda]$.

We say $\Phi(x)$ is a universal potential function for a field π, if $\Phi(x)$ is a universal potential function for any λ in the codomain of π. We also call α the contraction ratio and call M the base. The following two lemmas show that our main theorems follow from the existence of a universal potential function.

The way we define universal potential functions restricts them to only apply to the range of $(0, \lambda]$. This will be true in our applications (see for example Claim 3.4).

LEMMA 2.9. *Let (β, γ, λ) be three parameters such that $\beta \gamma > 1$, $\beta \leq \gamma$, and $\lambda < \lambda_c$. Let T and T' be two trees that agree on the first ℓ levels with root v and v', respectively. If there exists a universal potential function $\Phi(x)$, then $|p_v - p_{v'}| \leq O(\exp(-\ell))$.*

LEMMA 2.10. *Let (β, γ) be two parameters such that $\beta \gamma > 1$ and $\beta \leq 1 < \gamma$. Let $G = (V, E)$ be a graph with n many vertices and π be a field on G. Let $\lambda = \max_{v \in V}\{\pi(v)\}$. If there exists a universal potential function $\Phi(x)$ for π with contraction ratio $\alpha < 1$ and base M, then $Z_{\beta, \gamma, \pi}(G)$ can be approximated deterministically within a relative error ε in time $O(n^3(\frac{n\lambda}{\varepsilon})^{\frac{\log M}{-\log \alpha}})$.*

3 CORRELATION DECAY BELOW Δ_c OR λ_c

In this section, we show our main results. We will first show a folklore result for bounded degree graphs with a very simple proof. Then, we continue to show the main theorem regarding general

graphs. We carefully choose two appropriate potential functions and then apply Lemma 2.6 or Lemma 2.10.

3.1 Bounded Degree Graphs

We first apply our framework to get FPTAS for graphs with degree bound $\Delta < \Delta_c + 1 = \frac{2\sqrt{\beta\gamma}}{\sqrt{\beta\gamma}-1}$. Correlation decay for graphs with such degree bounds is folklore and can be found in [17] for the Ising model. Algorithmic implications are also shown, e.g., in Reference [24]. As we shall see, the proof is very simple in our framework. Note that λ, Δ, and α are considered constants for the FPTAS.

THEOREM 3.1. *Let (β, γ) be two parameters such that $\beta\gamma > 1$. Let $G = (V, E)$ be a graph with a maximum degree $\Delta < \Delta_c + 1$ and n many vertices, and let π be a field on G. Let $\lambda = \max_{v \in V}\{\pi(v)\}$. Then, $Z_{\beta,\gamma,\pi}(G)$ can be approximated deterministically within a relative error ε in time $O(n(\frac{n\lambda}{\varepsilon})^{\frac{\log(\Delta-1)}{-\log\alpha}})$, where $\alpha = \frac{\Delta-1}{\Delta_c}$.*

PROOF. We choose our potential function to be $\Phi_1(x) = \log x$ such that $\varphi_1(x) := \Phi_1'(x) = \frac{1}{x}$. We verify the conditions of Definition 2.5. Condition 1 is trivial. For Condition 2, we have that for any integer $1 \le d \le \Delta - 1$,

$$C_{\varphi_1, d}(\mathbf{x}) = \varphi_1(F_d(\mathbf{x})) \sum_{i=1}^{d} \frac{\partial F_d}{\partial x_i} \cdot \frac{1}{\varphi_1(x)}$$

$$= \frac{1}{F_d(\mathbf{x})} \sum_{i=1}^{d} F_d(\mathbf{x}) \cdot \frac{\beta\gamma - 1}{(x_i + \beta)(\gamma x_i + 1)} \cdot x_i$$

$$= \sum_{i=1}^{d} \frac{(\beta\gamma - 1)x_i}{(\gamma x_i + 1)(x_i + \beta)} \le \sum_{i=1}^{d} \frac{1}{\Delta_c} = \frac{d}{\Delta_c} \le \frac{\Delta - 1}{\Delta_c} = \alpha,$$

where we used the fact that for any $x > 0$,

$$\frac{(\beta\gamma - 1)x}{(\gamma x + 1)(x + \beta)} \le \frac{1}{\Delta_c}.$$

Hence, $\Phi_1(x)$ is a good potential function for all degrees $d \in [1, \Delta - 1]$ with contraction ratio α. The theorem follows by Lemma 2.6. □

Note that Theorem 3.1 matches the uniqueness condition in Proposition 2.3 and, restricted to the Ising model, the fast mixing bound of Gibbs samplers in Reference [18].

3.2 General Graphs

Now we turn our attention to general graphs without degree bounds.

Recall that $\lambda_c = (\frac{\gamma}{\beta})^{\frac{\Delta_c+1}{2}} = (\frac{\gamma}{\beta})^{\frac{\sqrt{\beta\gamma}}{\sqrt{\beta\gamma}-1}}$. The following two technical lemmas show some important properties regarding the threshold λ_c, which are keys to get our main theorems. In particular, Lemma 3.3 is key to bound the decay ratio. Proofs are given in Section 6.3.

LEMMA 3.2. *Let β, γ be two parameters such that $\beta\gamma > 1$ and $\beta \le \gamma$. For any $0 < x \le \lambda_c$, $\frac{\beta x+1}{x+\gamma} \le 1$.*

LEMMA 3.3. *Let β, γ be two parameters such that $\beta\gamma > 1$ and $\beta \le \gamma$. For any $0 < x \le \lambda_c$, we have*

$$(\beta\gamma - 1)x \log \frac{\lambda_c}{x} \le (\beta x + 1)(x + \gamma) \log \frac{x + \gamma}{\beta x + 1}. \tag{5}$$

In our applications, the quantity x in both lemmas will be the ratio of marginal probabilities in trees, denoted by R_v for a vertex v. To make use of these properties, one key requirement is that $0 < x \le \lambda_c$. This is not necessarily true in trees with pinning (and therefore not true in general SAW trees). Nevertheless, it does hold in trees without pinning.

CLAIM 3.4. *For* (β, γ, λ) *where* $\beta\gamma > 1$, $\beta \le \gamma$, *and* $\lambda < \lambda_c$, $R_v \in (0, \lambda]$ *holds in trees without pinning.*

We prove Claim 3.4 by induction. For any tree T_v, if v is the only vertex, then $R_v = \lambda$ and the base case holds. Given Lemma 3.2 and $\lambda < \lambda_c$, the inductive step to show Claim 3.4 follows from the standard tree recursion Equation (3).

In addition, it also holds when $\beta \le 1$, in trees even with pinning (but not counting the pinned vertices). This includes the SAW tree construction as special cases. To see that, for any vertex v, if one of v's child, say u, is pinned to 0 (or 1), then we can just remove u and replace the field λ_v on v with $\lambda'_v = \lambda_v \beta$ (or $\lambda'_v = \lambda_v/\gamma$), without affecting the marginal probability of v and any other vertices.[4] By our assumptions $\lambda_v < \lambda_c$ and $\beta \le 1 < \gamma$, we have that $\lambda'_v < \lambda_c$ as well. Hence, after removing all pinned vertices, we still have that $\lambda_v \le \lambda_c$ for all $v \in V$. This reduces to Claim 3.4.

Indeed, both of Theorem 1.1 and 1.2 can be generalized to the setting where vertices may have different external fields as long as they are all below λ_c, as follows.

THEOREM 3.5. *Let* (β, γ) *be two parameters such that* $\beta\gamma > 1$, $\beta \le \gamma$, *and* $\lambda < \lambda_c$. *Let* T_v *and* $T'_{v'}$ *be two trees with roots* v *and* v', *respectively. Let* $\lambda = \max_{u \in T_v \cup T'_{v'}}\{\pi(u)\}$. *If* $\lambda < \lambda_c$ *and in the first* ℓ *levels,* T_v *and* $T'_{v'}$ *have the same structure and external fields for corresponding pairs of vertices, then* $|p_v - p_{v'}| \le O(\exp(-\ell))$.

THEOREM 3.6. *Let* (β, γ) *be two parameters such that* $\beta\gamma > 1$ *and* $\beta \le 1 < \gamma$. *Let* $G = (V, E)$ *be a graph with* n *many vertices, and let* π *be a field on* G. *Let* $\lambda = \max_{v \in V}\{\pi(v)\}$. *If* $\lambda < \lambda_c$, *then* $Z_{\beta, \gamma, \pi}(G)$ *can be approximated deterministically within a relative error* ε *in time* $O(n(\frac{n\lambda}{\varepsilon})^{\frac{\log M}{-\log \alpha}})$, *where* $M > 1$ *and* $\alpha < 1$ *are two constants depending on* (β, γ, λ).

To show Theorems 3.5 and 3.6, we will apply Lemmas 2.9 and 2.10. Essentially, we only need to show the existence of a universal potential function.

Let $g_\lambda(x) := \frac{(\beta\gamma-1)x \log \frac{\lambda}{x}}{(\beta x+1)(x+\gamma) \log \frac{x+\gamma}{\beta x+1}}$. By Lemma 3.3, $g_{\lambda_c}(x) \le 1$. For $\lambda < \lambda_c$, note that $\lim_{x \to 0} g_\lambda(x) = 0$. Hence, there exists $0 < \varepsilon < \lambda$ and $0 < \delta < 1$ such that if $0 < x < \varepsilon$, $g_\lambda(x) < \delta$. Moreover, if $\varepsilon \le x \le \lambda$, then $\frac{g_\lambda(x)}{g_{\lambda_c}(x)} = \frac{\log \lambda - \log x}{\log \lambda_c - \log x} \le \frac{\log \lambda - \log \varepsilon}{\log \lambda_c - \log \varepsilon}$. Let

$$\alpha_\lambda := \max\left\{\delta, \frac{\log \lambda - \log \varepsilon}{\log \lambda_c - \log \varepsilon}\right\} < 1.$$

Then, we have just shown the following lemma.

LEMMA 3.7. *Let* β, γ *be two parameters such that* $\beta\gamma > 1$ *and* $\beta \le \gamma$. *If* $\lambda < \lambda_c$, *then* $g_\lambda(x) \le \alpha_\lambda$ *for any* $0 < x \le \lambda$, *where* $\alpha_\lambda < 1$ *is defined above.*

Let $t := \frac{\alpha_\lambda \gamma}{\beta\gamma-1} \log \frac{\lambda+\gamma}{\beta\lambda+1}$ so that for any $0 < x \le \lambda$,

$$t < \frac{\alpha_\lambda(\beta x + 1)(x + \gamma)}{\beta\gamma - 1} \log \frac{x + \gamma}{\beta x + 1}, \tag{6}$$

---

[4]We may need to repeat this step for d times, giving rise to the interval in Definition 2.5.

since $(\beta x + 1)(x + \gamma) > \gamma$ and $\log \frac{x+\gamma}{\beta x+1} \geq \log \frac{\lambda+\gamma}{\beta\lambda+1}$. We define $\varphi_2(x) := \min\{\frac{1}{t}, \frac{1}{x \log \frac{\lambda}{x}}\}$. To be more specific, note that $x \log \frac{\lambda}{x} \leq \frac{\lambda}{e}$ for any $0 < x \leq \lambda$. If $t \geq \frac{\lambda}{e}$, then $\frac{1}{x \log \frac{\lambda}{x}} \geq \frac{1}{t}$ for any $0 < x \leq \lambda$, implying that

$$\varphi_2(x) = \frac{1}{t}. \tag{7}$$

Otherwise, $t < \frac{\lambda}{e}$, and there are two roots to $x \log \frac{\lambda}{x} = t$ in $(0, \lambda]$. Denote them by x_0 and x_1. Then, we have that

$$\varphi_2(x) = \begin{cases} \frac{1}{t} & 0 \leq x < x_0; \\ \frac{1}{x \log \frac{\lambda}{x}} & x_0 \leq x < x_1; \\ \frac{1}{t} & x_1 \leq x < \lambda. \end{cases} \tag{8}$$

We define $\Phi_2(x) := \int_0^x \varphi_2(y)dy$ so that $\Phi_2'(x) = \varphi_2(x)$. Our choice of $\varphi_2(x)$ ensures that for any $0 < x \leq \lambda$,

$$\varphi_2(x)x \log \frac{\lambda}{x} \leq 1. \tag{9}$$

Moreover, we claim that

$$\frac{\beta\gamma - 1}{(\beta x + 1)(x + \gamma)} \cdot \frac{1}{\varphi_2(x)} \leq \alpha_\lambda \log \frac{x + \gamma}{\beta x + 1}. \tag{10}$$

This is because if $\varphi_2(x) = \frac{1}{x \log \frac{\lambda}{x}}$, then by Lemmas 3.2 and 3.7,

$$\frac{\beta\gamma - 1}{(\beta x + 1)(x + \gamma)} \cdot x \log \frac{\lambda}{x} \leq \alpha_\lambda \log \frac{x + \gamma}{\beta x + 1},$$

which implies Equation (10). Otherwise, $\varphi_2(x) = 1/t$, and Equation (10) follows from Equation (6). Now, we are ready to prove Theorems 3.5 and 3.6.

PROOF OF THEOREMS 3.5 AND 3.6. We claim that $\Phi_2(x)$ is a universal potential function for any field π with an upper bound λ, with contraction ratio α_λ given above and base M that will be determined shortly. Theorems 3.5 and 3.6 follow from $\Phi_2(x)$ combined with Lemmas 2.9 and 2.10, respectively. We verify the two conditions in Definition 2.8.

For Condition 1, it is easy to see that in case Equation (7), $\varphi_2(x) = \frac{1}{t}$ for any $x \in (0, \lambda]$, and in case Equation (8), $\frac{e}{\lambda} \leq \varphi_2(x) \leq \frac{1}{t}$ for any $x \in (0, \lambda]$.

For Condition 2, we have that

$$C_{\varphi_2, d}(\mathbf{x}) = \varphi_2(F_d(\mathbf{x})) \sum_{i=1}^d \frac{\partial F_d}{\partial x_i} \cdot \frac{1}{\varphi_2(x_i)}$$

$$= \varphi_2(F_d(\mathbf{x}))F_d(\mathbf{x}) \sum_{i=1}^d \frac{\beta\gamma - 1}{(\beta x_i + 1)(x_i + \gamma)} \cdot \frac{1}{\varphi_2(x_i)}$$

$$\leq \varphi_2(F_d(\mathbf{x}))F_d(\mathbf{x}) \sum_{i=1}^d \alpha_\lambda \log \frac{x_i + \gamma}{\beta x_i + 1} \qquad \text{(by Equation (10))}$$

$$= \alpha_\lambda \varphi_2(F_d(\mathbf{x}))F_d(\mathbf{x}) \log \frac{\lambda}{F_d(\mathbf{x})} \tag{11}$$

$$\leq \alpha_\lambda. \qquad \text{(by Equation (9))}$$

Moreover, $F_d(\mathbf{x}) < \lambda(\frac{\beta\lambda+1}{\lambda+\gamma})^d$ for any $x_i \in (0, \lambda]$. We have an alternative bound that

$$C_{\varphi_2, d}(\mathbf{x}) \leq \frac{\alpha_\lambda}{t} F_d(\mathbf{x}) \log \frac{\lambda}{F_d(\mathbf{x})} \qquad\qquad \text{(by Equation (11) and } \varphi_2(x) \leq 1/t)$$

$$\leq \frac{\alpha_\lambda \lambda}{t} \left(\frac{\beta\lambda+1}{\lambda+\gamma} \right)^d d \log \frac{\lambda+\gamma}{\beta\lambda+1}.$$

Since $\frac{\beta\lambda+1}{\lambda+\gamma} < 1$ by Lemma 3.2, the right-hand side decreases exponentially in d. Therefore, there exists a sufficiently large integer M such that for any $1 \leq d < M$, $C_{\varphi_2, d}(\mathbf{x}) \leq \alpha_\lambda \leq \alpha_\lambda^{\lceil \log_M(d+1) \rceil}$, and for any $d \geq M$, $C_{\varphi_2, d}(\mathbf{x}) \leq \alpha_\lambda^{\lceil \log_M(d+1) \rceil}$. This verifies Condition 2. $\square$

3.3 Heuristics Behind $\Phi_2(x)$

The most intricate part of our proofs of Theorems 3.5 and 3.6 is the choice of the potential function $\Phi_2(x)$ given by Equation (8). Here, we give a brief heuristic of deriving it. It is more of an "educated guess" than a rigorous argument.

We want to pick $\Phi_2(x)$ such that Condition 2 holds. In particular, we want

$$\varphi_2(F_d(\mathbf{x})) \sum_{i=1}^d \frac{\partial F_d}{\partial x_i} \cdot \frac{1}{\varphi_2(x_i)} < 1.$$

It is fair to assume that the left-hand side of the equation above takes its maximum when all x_i's are equal. Hence, we hope the following to hold:

$$\frac{\varphi_2(f_d(x)) f_d'(x)}{\varphi_2(x)} < 1, \tag{12}$$

where $f_d(x) = \lambda(\frac{\beta x+1}{x+\gamma})^d$ is the symmetrized version of $F_d(\mathbf{x})$. We will use $z := f_d(x)$ to simplify notation. Since we want Equation (12) to hold for all degrees d, we hope to eliminate d from the left-hand side of Equation (12). Notice that $\varphi_2(x)$ should be independent from d. Therefore, we take the derivative of $\varphi_2(f_d(x)) f_d'(x)$ against d and get

$$\frac{\partial \varphi_2(f_d(x)) f_d'(x)}{\partial d} = \frac{\beta\gamma - 1}{(\beta x+1)(x+\gamma)} \left(\varphi_2(z)z + \varphi_2(z)z \log \frac{z}{\lambda} + \varphi_2'(z)z^2 \log \frac{z}{\lambda} \right)$$

$$= \frac{(\beta\gamma-1)z\varphi_2(z)}{(\beta x+1)(x+\gamma)} \left(1 + \log \frac{z}{\lambda} + (\log \varphi_2(z))' z \log \frac{z}{\lambda} \right).$$

We may achieve our goal of eliminating d by imposing the sum in the last parenthesis to be 0, namely,

$$(\log \varphi_2(z))' = -\frac{1}{z} - \frac{1}{z \log \frac{z}{\lambda}}$$

$$= -(\log z)' - \left(\log \log \frac{\lambda}{z} \right)'. \tag{13}$$

From Equation (13), it is easy to see that $\varphi_2(z) = \frac{1}{z \log \frac{\lambda}{z}}$ satisfies our need. To get the full definition of Equation (8), we apply a thresholding trick to bound $\varphi_2(z)$ away from 0.

3.4 Discussion of the $\beta > 1$ Case

We cannot combine conditions of Theorems 3.5 and 3.6 together to have an FPTAS. In particular, when $\beta > 1$ strong spatial mixing fails for any λ even if $\lambda < \lambda_c$. To see this, given a Δ-ary tree T, we can append t many children to every vertex in T to get a new tree T' and impose a partial configuration σ where all these new children are pinned to 0. Effectively, the tree T' is equivalent to T where every vertex has a new external field of $\lambda\beta^t$, which is larger than λ_c^{int} if t is sufficiently large regardless of λ. Then by Proposition 2.4, long-range correlation exists in T' with the partial configuration σ, and strong spatial mixing fails.

However, it is easy to see from the proof that, Theorem 3.5 can be generalized to allow a partial configuration σ on some subset Λ where the marginal probability of every vertex $v \in \Lambda$ satisfies $p_v^\sigma \leq \frac{\lambda_c}{\lambda_c+1}$. This is not the case for the SAW tree that our algorithm relies on when $\beta > 1$. However, the following observation shows that if $\lambda_v \leq \lambda_c \leq \frac{\gamma-1}{\beta-1}$, then the marginal probability of any instance G satisfies this requirement. Thus, it seems the only piece missing to obtain an algorithm is to design a better recursion tree instead of the SAW tree.

PROPOSITION 3.8. *Let (β, γ) be two parameters such that $1 \leq \beta \leq \gamma$ and $\beta\gamma > 1$. Let $\lambda \leq \frac{\gamma-1}{\beta-1}$ be another parameter. For any graph $G = (V, E)$, if $\pi(v) \leq \lambda$ for all $v \in V$, then $p_v \leq \frac{\lambda}{\lambda+1}$.*

To prove this proposition, we need to use the random cluster formulation of two-spin models. Let G be a graph and $e = (v_1, v_2)$ be one of its edges. Let G^+ be the graph where the edge e is contracted, and G^- be the graph where e is removed. Moreover, in G^+, we assign $\pi^+(\widetilde{v}) = \lambda_{v_1}\lambda_{v_2}\frac{\beta-1}{\gamma-1}$, where $\widetilde{v}$ is the vertex obtained from contacting e. Then, we have that

$$Z(G) = Z(G^-) + (\gamma - 1)Z(G^+), \tag{14}$$

where we write $Z(G)$ instead of $Z_{\beta,\gamma,\pi}(G)$ to simplify the notation. To show the equation above, we only need a simple adapation of the random cluster formulation of the Ising model to the two-spin setting.

PROOF OF PROPOSITION 3.8. Suppose $G = (V, E)$ where $|V| = n$ and $|E| = m$. We show the claim by inducting on (m, n). Clearly the statement holds when $m = 0$ or $n = 1$. Hence, we may assume the claim holds for (m', n) where $m' < m$ as well as (m', n') where $n' < n$, and show that the claim holds for (m, n).

Pick an arbitrary edge $e = (v_1, v_2)$ in G. Let G^+ and G^- be as in the random cluster formulation. It is easy to see that $\pi(\widetilde{v}) = \lambda_{v_1}\lambda_{v_2}\frac{\beta-1}{\gamma-1} \leq \lambda$. Hence, both G^+ and G^- satisfy the induction hypothesis. It implies that $p_{G^-;v} \leq \frac{\lambda}{\lambda+1}$ for any v, where $p_{G^-;v}$ is the mariginal probability of v in G^-. Moreover, $p_{G^+;v} \leq \frac{\lambda}{\lambda+1}$ for any $v \in V^+$, where V^+ is the vertex set of G^+. Let δ be a mapping $V \to V^+$ such that $\delta(v) = v$ if $v \neq v_1, v_2$ and $\delta(v_1) = \delta(v_2) = \widetilde{v}$. Then, using Equation (14), we have that for any vertex $v \in V$,

$$
\begin{aligned}
p_{G;v} &= \frac{Z^{\sigma(v)=0}(G)}{Z(G)} = \frac{Z^{\sigma(v)=0}(G^-) + (\gamma - 1)Z^{\sigma(\delta(v))=0}(G^+)}{Z(G^-) + (\gamma - 1)Z(G^+)} \\
&= p_{G^-;v} \cdot \frac{Z(G^-)}{Z(G^-) + (\gamma - 1)Z(G^+)} + p_{G^+;\delta(v)} \cdot \frac{(\gamma - 1)Z(G^+)}{Z(G^-) + (\gamma - 1)Z(G^+)} \\
&\leq \frac{\lambda}{\lambda+1} \cdot \frac{Z(G^-)}{Z(G^-) + (\gamma - 1)Z(G^+)} + \frac{\lambda}{\lambda+1} \cdot \frac{(\gamma - 1)Z(G^+)}{Z(G^-) + (\gamma - 1)Z(G^+)} = \frac{\lambda}{\lambda+1},
\end{aligned}
$$

where in the last line we use the induction hypotheses. $\qquad\square$

Proposition 3.8 can be also viewed as a generalization of Griffith's first inequality [9] from the Ising model to general ferromagnetic two-spin systems.

4   CORRELATION DECAY BEYOND λ_c

Let β, γ be two parameters such that $\beta \le 1 < \gamma$ and $\beta\gamma > 1$. In this section, we give an example to show that if Δ_c is not an integer, then correlation decay still holds for a small interval beyond λ_c. To simplify the presentation, we assume that π is a uniform field such that $\pi(v) = \lambda$. Note that the potential function $\varphi_2(x)$ does not extend beyond λ_c.

Let $\beta = 0.6$ and $\gamma = 2$. Then, $\Delta_c = \frac{\sqrt{\beta\gamma}+1}{\sqrt{\beta\gamma}-1} \approx 21.95$ and $\lambda_c = (\gamma/\beta)^{\frac{\Delta_c+1}{2}} < 1002761$. Let $\lambda = 1002762 > \lambda_c$. We will show that #2Spin(β, γ, λ) still has an FPTAS.

Define a constant t as

$$t := \frac{\sqrt{\beta\gamma}+1}{\sqrt{\beta\gamma}-1} \cdot \frac{\log\sqrt{\gamma/\beta}}{\sqrt{\gamma/\beta}+1} - \log\left(1 + \sqrt{\beta/\gamma}\right) \approx 4.24032. \tag{15}$$

We consider a potential function $\Phi_3(x)$ so that $\varphi_3(x) := \frac{1}{x(\log(1+1/x)+t)}$. With this choice,

$$C_{\varphi_3, d}(\mathbf{x}) = \varphi_3(F_d(\mathbf{x})) \sum_{i=1}^{d} \frac{\partial F_d}{\partial x_i} \cdot \frac{1}{\varphi_3(x)}$$

$$= \frac{\beta\gamma - 1}{\log\left(1 + 1/F_d(\mathbf{x})\right) + t} \sum_{i=1}^{d} \frac{x_i\left(\log(1 + 1/x_i) + t\right)}{(\beta x_i + 1)(x_i + \gamma)}.$$

We do a change of variables. Let $r_i = \frac{\beta x_i + 1}{x_i + \gamma}$. Then, $x_i = \frac{\gamma r_i - 1}{\beta - r_i}$, $\beta x_i + 1 = \frac{r_i(\beta\gamma - 1)}{\beta - r_i}$, and $x_i + \gamma = \frac{\beta\gamma - 1}{\beta - r_i}$. Hence,

$$\sum_{i=1}^{d} \frac{x_i(\log(1 + 1/x_i) + t)}{(\beta x_i + 1)(x_i + \gamma)} = \sum_{i=1}^{d} \frac{(\gamma r_i - 1)(\beta - r_i)}{r_i(\beta\gamma - 1)^2} \cdot \left(\log\left(1 + \frac{\beta - r_i}{\gamma r_i - 1}\right) + t\right)$$

$$= \frac{1}{(\beta\gamma - 1)^2} \sum_{i=1}^{d} \left(1 + \beta\gamma - \frac{\beta}{r_i} - \gamma r_i\right)\left(\log\left(1 + \frac{\beta - r_i}{\gamma r_i - 1}\right) + t\right).$$

Furthermore, let $s_i = \log r_i$. As $r_i \in (\frac{1}{\gamma}, \beta)$, $s_i \in (-\log\gamma, \log\beta)$. Let

$$\rho(x) := (1 + \beta\gamma - \beta e^{-x} - \gamma e^x)\left(\log\left(1 + \frac{\beta - e^x}{\gamma e^x - 1}\right) + t\right).$$

Then, $\rho(x)$ is concave for any $x \in (-\log\gamma, \log\beta)$. It can be easily verified, as the second derivative is

$$\rho''(x) = \frac{(\beta + 1)(\beta\gamma - 1)}{\beta - 1 + e^x(\gamma - 1)} + \frac{\beta\gamma - 1}{\gamma - 1} - \frac{\beta\gamma - 1}{e^x\gamma - 1} - \frac{(\beta - 1)(\beta\gamma - 1)^2}{(\gamma - 1)(\beta - 1 + e^x(\gamma - 1))^2}$$

$$- \beta t e^{-x} - \gamma t e^x - e^{-x}\left(\beta + e^{2x}\gamma\right)\log\left(1 + \frac{\beta - e^x}{\gamma e^x - 1}\right).$$

$$\le \gamma(\beta + 1) + \frac{\beta\gamma - 1}{\gamma - 1} - 1 - \frac{\beta - 1}{\gamma - 1} - 2t < -5 < 0, \tag{16}$$

where in the last line we used Equation (15) and the fact that $1/\gamma \leq e^x \leq \beta$. Hence, by concavity, we have that for any $x_i \in (0, \lambda]$,

$$C_{\varphi_3, d}(\mathbf{x}) = \frac{\beta\gamma - 1}{\log\left(1 + 1/F_d(\mathbf{x})\right) + t} \sum_{i=1}^{d} \frac{x_i \left(\log(1 + 1/x_i) + t\right)}{(\beta x_i + 1)(x_i + \gamma)},$$

$$\leq \frac{\beta\gamma - 1}{\log\left(1 + 1/f_d(\widetilde{x})\right) + t} \cdot \frac{d\widetilde{x}\left(\log(1 + \widetilde{x}^{-1}) + t\right)}{(\beta\widetilde{x} + 1)(\widetilde{x} + \gamma)} = c_{\varphi_3, d}(\widetilde{x}), \tag{17}$$

where $\widetilde{x} > 0$ is the unique solution such that $f_d(\widetilde{x}) = F_d(\mathbf{x})$.

Next, we show that there exists an $\alpha < 1$ such that for any integer d and $x > 0$, $c_{\varphi_3, d}(x) < \alpha$. In fact, by Equation (15), our choice of t, it is not hard to show that the maximum of $c_{\varphi_3, d}(x)$ is achieved at $x = \sqrt{\gamma/\beta}$ and $d = \Delta_c$, which is 1 if $\lambda = \lambda_c$ and is larger than 1 if $\lambda > \lambda_c$. However, since the degree d has to be an integer, we can verify that for any integer $1 \leq d \leq 100$, the maximum of $c_{\varphi_3, d}(x)$ is $c_{\varphi_3, 22}(x_{22}) = 0.999983$, where $x_{22} \approx 1.83066$. If $d > 100$, then

$$c_{\varphi_3, d}(x) = \frac{d(\beta\gamma - 1)}{\log\left(1 + 1/f_d(x)\right) + t} \cdot \frac{x\left(\log(1 + x^{-1}) + t\right)}{(\beta x + 1)(x + \gamma)}$$

$$\leq C_0 \cdot C_1 < 1,$$

where $C_0 < 1.07191$ is the maximum of $\frac{x\left(\log(1 + x^{-1}) + t\right)}{(\beta x + 1)(x + \gamma)}$ for any $x > 0$, and $C_1 < 0.481875$ is the maximum of $\frac{d(\beta\gamma - 1)}{\log\left(1 + \lambda^{-1}\beta^{-d}\right) + t}$ for any $d > 100$. Then, due to Equation (17), we have that for any $x_i \in (0, \lambda]$, $C_{\varphi_3, d}(\mathbf{x}) < \alpha = 0.999983 < 1$. This is the counterpart of $C_{\varphi_2, d}(\mathbf{x}) < \alpha_\lambda$ in the proof of Theorem 3.6. To make $\varphi_3(x)$ satisfy Conditions 1 and 2 in Definition 2.8, it is sufficient to do a simple "chop-off" trick to $\varphi_3(x)$ as in Equation (8). We will omit the detail here.

PROPOSITION 4.1. *For $\beta = 0.6$, $\gamma = 2$, and $\lambda = 1002762 > \lambda_c$, #2SPIN($\beta, \gamma, \lambda$) has an FPTAS.*

It is easy to see that the argument above works for any $\beta \leq 1 < \gamma$ and $\beta\gamma > 1$ except Equation (16), the concavity of $\rho(x)$. Indeed, the concavity does not hold if, say, $\beta = 1$ and $\gamma = 2$. Nevertheless, the key point here is that λ_c is not the tight bound for FPTAS. Short of a conjectured optimal bound, we did not try to optimize the potential function nor the applicable range of the proof above.

5 LIMITATIONS OF CORRELATION DECAY

In this section, we discuss some limitations of approximation algorithms for ferromagnetic two-spin models based on correlation decay analysis.

The problem of counting independent sets in bipartite graphs (#BIS) plays an important role in classifying approximate counting complexity. #BIS is not known to have any efficient approximation algorithm, despite many attempts. However, there is no known approximation preserving reduction (AP-reduction) to reduce #BIS from #SAT either. It is conjectured to have intermediate approximation complexity, and in particular, to have no FPRAS [4].

Goldberg and Jerrum [7] showed that for any $\beta\gamma > 1$, approximating #2SPIN($\beta, \gamma, (0, \infty)$) can be reduced to approximating #BIS. This is the (approximation) complexity upper bound of all ferromagnetic two-spin models. In contrast, by Theorem 3.1, #Δ-2SPIN($\beta, \gamma, (0, \infty)$) has an FPTAS, if $\Delta < \Delta_c + 1$. Note that when we write #2SPIN($\beta, \gamma, (0, \infty)$) the field is implicitly assumed to be at most polynomial in size of the graph (or in unary).

We then consider fields with some constant bounds. Recall that $\lambda_c^{int} = (\gamma/\beta)^{\frac{\lceil\Delta_c\rceil+1}{2}}$. Let $\lambda_c^{int'} = (\gamma/\beta)^{\frac{\lfloor\Delta_c\rfloor+2}{2}}$. Then $\lambda_c^{int'} = \lambda_c^{int}$ unless Δ_c is an integer. By reducing to anti-ferromagnetic two-spin

models in bipartite graphs, we have the following hardness result, which is first observed in Reference [16, Theorem 3].

PROPOSITION 5.1. *Let* (β, γ, λ) *be a set of parameters such that* $\beta < \gamma$, $\beta\gamma > 1$, *and* $\lambda > \lambda_c^{int'}$. *Then,* *#2SPIN*$(\beta, \gamma, (0, \lambda])$ *is #BIS-hard.*

The reduction goes as follows. Anti-ferromagnetic Ising models with a constant non-trivial field in bounded degree bipartite graphs are #BIS-hard, if the uniqueness condition fails [2]. Given such an instance, we may first flip the truth table of one side. This effectively results in a ferromagnetic Ising model in the same bipartite graph, with two different fields on each side. By a standard diagonal transformation, we can transform such an Ising model to any ferromagnetic two-spin model, with various local fields depending on the degree. It can be verified that for any $\lambda > \lambda_c^{int'}$, we may pick a field in the anti-ferromagnetic Ising model to start with, such that uniqueness fails and after the transformation, the largest field in use is at most λ.

The hardness bound in Proposition 5.1 matches the failure of uniqueness due to Proposition 2.4, unless Δ_c is an integer. In contrast to Proposition 5.1, Theorem 3.6 implies that if $\beta \leq 1 < \gamma$ and $\lambda < \lambda_c = (\gamma/\beta)^{\frac{\Delta_c+1}{2}}$, then #2SPIN$(\beta, \gamma, (0, \lambda])$ has an FPTAS. Hence, Theorem 3.6 is almost optimal, up to an integrality gap.

We note that λ_c is not the tight bound for FPTAS, as observed in Proposition 4.1. Since the degree d has to be an integer, with an appropriate choice of the potential function, there is a small interval beyond λ_c such that strong spatial mixing still holds. Interestingly, it seems that λ_c^{int} is not the right bound either. Let us make a concrete example. Let $\beta = 1$ and $\gamma = 2$. Then $\Delta_c = \frac{\sqrt{\beta\gamma}+1}{\sqrt{\beta\gamma}-1} = \frac{\sqrt{2}+1}{\sqrt{2}-1} \approx 5.82843$. Hence, $\lambda_c \approx 10.6606$ and $\lambda_c^{int} = (2)^{\frac{6+1}{2}} \approx 11.3137$. However, even if $\lambda < \lambda_c^{int}$, the system may not exhibit spatial mixing, neither in the strong nor in the weak sense.

In fact, even the spatial mixing in the sense of Theorem 1.1 does not necessarily hold if $\lambda < \lambda_c^{int}$. To see this, we take any $\lambda \in [10.9759, 10.9965]$ so that $\lambda_c < \lambda < \lambda_c^{int}$. Consider an infinite tree where at even layers, each vertex has five children, and at odd layers, each vertex has seven children. There are more than one Gibbs measures in this tree. This can be easily verified from the fact that the two layer recursion function $f_5(f_7(x))$ has three fixed points such that $x = f_5(f_7(x))$. In addition, all three fixed points $\widehat{x}_i$ satisfy that $\widehat{x}_i < \lambda_c$ for $i = 1, 2, 3$. Consider a tree T with alternating degrees 6 and 8 of depth 2ℓ (so that the number of children is alternatingly 5 and 7), and another tree T' of the same structure in the first 2ℓ layers as T but with one more layer where each vertex has, say, 50 children. It is not hard to verify that as ℓ increases, the marginal ratio at the root of T converges to $\widehat{x}_3$, but the ratio at the root of T' converges to $\widehat{x}_1$. This example indicates that one should not expect correlation decay algorithms to work all the way up to λ_c^{int}.

At last, if we consider the uniform field case #2SPIN(β, γ, λ), then our tractability results still holds. However, to extend the hardness results as in Proposition 5.1 from an interval of fields to a uniform one, there seems to be some technical difficulty. Suppose we want to construct a combinatorial gadget to effectively realize another field. There is a gap between λ and the next largest possible field to realize. This is why in Reference [16] there are some extra conditions transiting from an interval of fields to the uniform case. The observation above about the failure of SSM in irregular trees may suggest a random bipartite construction of uneven degrees. However, to analyze such a gadget is beyond the scope of the current article.

6 MISSING PROOFS

At last, we gather technical details and proofs that are omitted in Sections 2.2, 2.3, and 3.2.

6.1 Details About the Uniqueness Threshold

We prove Propositions 2.3 and 2.4. Technically, by only considering the symmetric recursion $f_d(x) = \lambda(\frac{\beta x+1}{x+\gamma})^d$, we are implicitly assuming uniform boundary conditions. If there is more than one fixed point for $f_d(x)$, then clearly there are multiple Gibbs measures. Hence, $f_d(x)$ having only one fixed point is a necessary condition for the uniqueness condition in $\mathbb{T}_{d+1}$. Moreover, it is also sufficient. The reason is that the influence on the root of an arbitrary boundary condition is bounded between those of the all "0" and all "1" boundary conditions.

First, do some calculation here. Take the derivative of $f_d(x)$:

$$f_d'(x) = \frac{d(\beta\gamma - 1)f_d(x)}{(\beta x + 1)(x + \gamma)}. \tag{18}$$

Then take the second derivative:

$$\frac{f_d''(x)}{f_d'(x)} = \frac{f_d'(x)}{f_d(x)} - \frac{\beta}{\beta x + 1} - \frac{1}{x + \gamma} = \frac{d(\beta\gamma - 1) - \beta\gamma - 1 - 2\beta x}{(\beta x + 1)(x + \gamma)}.$$

Therefore, at $x^* := \frac{d(\beta\gamma-1)-(\beta\gamma+1)}{2\beta}$, $f_d''(x^*) = 0$. It is easy to see when $d < \frac{\beta\gamma+1}{\beta\gamma-1}$, $f_d''(x) < 0$ for all $x > 0$. So $f_d(x)$ is concave and therefore has only one fixed point.

Since $f_d(x)$ has only one inflection point, there are at most three fixed points. Moreover, the uniqueness condition is equivalent to say that for all fixed points $\widehat{x}_d$ of $f_d(x)$, $f_d'(\widehat{x}_d) < 1$. For a fixed point $\widehat{x}_d$, we plug it in Equation (18):

$$f_d'(\widehat{x}_d) = \frac{d(\beta\gamma - 1)\widehat{x}_d}{(\beta\widehat{x}_d + 1)(\widehat{x}_d + \gamma)}.$$

Recall that $\Delta_c := \frac{\sqrt{\beta\gamma}+1}{\sqrt{\beta\gamma}-1}$. If $d < \Delta_c$, then we have that for any x,

$$(\beta x + 1)(x + \gamma) - d(\beta\gamma - 1)x = \beta x^2 + ((\beta\gamma + 1) - d(\beta\gamma - 1))x + \gamma$$

$$> \beta x^2 + \left(\beta\gamma + 1 - \left(\sqrt{\beta\gamma} + 1\right)^2\right)x + \gamma$$

$$= \left(\sqrt{\beta}x - \sqrt{\gamma}\right)^2 \geq 0.$$

Hence, $(\beta x + 1)(x + \gamma) > d(\beta\gamma - 1)x$. In particular, $f_d'(\widehat{x}_d) < 1$ for any fixed point $\widehat{x}_d$ and the uniqueness condition holds. This proves Proposition 2.3.

To show Proposition 2.4, we may assume that $d \geq \Delta_c$. We may also assume that $\beta \leq \gamma$. The equation $(\beta x + 1)(\gamma + x) = d(\beta\gamma - 1)x$ has two solutions, which are

$$x_0 = x^* - \frac{\sqrt{((\beta\gamma + 1) - d(\beta\gamma - 1))^2 - 4\beta\gamma}}{2\beta}$$

and

$$x_1 = x^* + \frac{\sqrt{((\beta\gamma + 1) - d(\beta\gamma - 1))^2 - 4\beta\gamma}}{2\beta}.$$

Notice that both of them are positive, since $x_0 + x_1 = 2x^* > 0$ and $x_0 x_1 = \gamma/\beta$. As d goes to ∞,

$$x_0 = o(1), \qquad x_1 = 2x^* - o(1) = \frac{d(\beta\gamma - 1) - (\beta\gamma + 1)}{\beta} - o(1). \tag{19}$$

Moreover,

$$\frac{d(\beta\gamma - 1)x}{(\beta x + 1)(\gamma + x)} > 1 \quad \text{if and only if} \quad x_0 < x < x_1. \tag{20}$$

We show that $f_d(x_0) > x_0$ or $f_d(x_1) < x_1$ is equivalent to the uniqueness condition. First, we assume this condition does not hold, that is $f_d(x_0) \leq x_0$ and $f_d(x_1) \geq x_1$. If any of the equation holds, then x_0 or x_1 is a fixed point and the derivative is 1. So, we have non-uniqueness. Otherwise, we have $f_d(x_0) < x_0$ and $f_d(x_1) > x_1$. Since $x_0 < x_1$, there is some fixed point $\widetilde{x}$ satisfying $f_d(\widetilde{x}) = \widetilde{x}$ and $x_0 < \widetilde{x} < x_1$. The second inequality implies that $f'_d(\widetilde{x}) > 1$ via Equation (20) and non-uniqueness holds.

To show the other direction, if $f_d(x_0) > x_0$, then

$$f'_d(x_0) = \frac{d(\beta\gamma - 1)f(x_0)}{(\beta x_0 + 1)(x_0 + \gamma)} > \frac{d(\beta\gamma - 1)x_0}{(\beta x_0 + 1)(x_0 + \gamma)} = 1.$$

Assume for contradiction that $f_d(x)$ has three fixed points, denoted by $\widetilde{x}_0 < \widetilde{x}_1 < \widetilde{x}_2$. Then, the middle fixed point $\widetilde{x}_1$ satisfies $f'_d(\widetilde{x}_1) > 1$. Therefore, $\widetilde{x}_1 > x_0$ by Equation (20) and there are two fixed points larger than x_0. However, for $x_0 < x \leq x^*$, $f'_d(x) > 1$ and $f_d(x_0) > x_0$. Hence, there is no fixed point in this interval. For $x > x^*$, the function is concave and has exactly one fixed point. So there is only 1 fixed point larger than x_0. Contradiction. The case that $f_d(x_1) < x_1$ is similar.

These two conditions could be rewritten as

$$\lambda > \frac{x_0(x_0 + \gamma)^d}{(\beta x_0 + 1)^d} \tag{21}$$

and

$$\lambda < \frac{x_1(x_1 + \gamma)^d}{(\beta x_1 + 1)^d}. \tag{22}$$

Notice that the right-hand side has nothing to do with λ in both Equations (21) and (22).

We want to see how condition Equations (21) and (22) change as d changes. Treat d as a continuous variable. Define

$$g_i(d) := \frac{x_i(x_i + \gamma)^d}{(\beta x_i + 1)^d},$$

where $i = 0, 1$ and x_i is defined above depending on β, γ and d. Take the derivative:

$$\frac{g'_i(d)}{g_i(d)} = \frac{\partial x_i}{\partial d}\left(\frac{1}{x_i} + \frac{d}{x_i + \gamma} - \frac{d\beta}{\beta x_i + 1}\right) + \log(x_i + \gamma) - \log(\beta x_i + 1)$$

$$= \frac{\partial x_i}{\partial d}\left(\frac{1}{x_i} + \frac{d(1 - \beta\gamma)}{(x_i + \gamma)(\beta x_i + 1)}\right) + \log\frac{x_i + \gamma}{\beta x_i + 1}$$

$$= \frac{\partial x_i}{\partial d}\left(\frac{1}{x_i} - \frac{1}{x_i}\right) + \log\frac{x_i + \gamma}{\beta x_i + 1} = \log\frac{x_i + \gamma}{\beta x_i + 1}.$$

If $\beta \leq 1$, then these two functions are increasing in d. Recall that $\Delta_c = \frac{\sqrt{\beta\gamma}+1}{\sqrt{\beta\gamma}-1}$, and $\lambda_c^{int} = g_1(\lceil\Delta_c\rceil) = (\gamma/\beta)^{\frac{\lceil\Delta_c\rceil+1}{2}}$. Thus, if $\lambda < \lambda_c^{int}$, Equation (22) holds for all integers d. However, $x_0 = o(1)$ by Equation (19), and

$$g_0(d) = \frac{x_0(x_0 + \gamma)^d}{(\beta x_0 + 1)^d} = \frac{\gamma}{\beta x_1} \cdot \left(\frac{x_0 + \gamma}{\beta x_0 + 1}\right)^d$$

$$> \frac{\gamma}{2\beta x^*} \cdot \left(\frac{x_0 + \gamma}{\beta x_0 + 1}\right)^d = \frac{\gamma}{d(\beta\gamma - 1) - (\beta\gamma + 1)} \cdot \left(\frac{x_0 + \gamma}{\beta x_0 + 1}\right)^d$$

$$\to \infty \text{ as } d \text{ goes to } \infty.$$

Hence, there is no λ such that Equation (21) holds for all integers d.

If $\beta > 1$, then neither Equation (21) nor Equation (22) can hold for all integers d. Since $x_0 = o(1)$ by Equation (19), similar to the argument above, we have that

$$g_0(d) = \frac{x_0(x_0 + \gamma)^d}{(\beta x_0 + 1)^d} > \frac{\gamma}{d(\beta\gamma - 1) - (\beta\gamma + 1)} \cdot \left(\frac{x_0 + \gamma}{\beta x_0 + 1}\right)^d$$
$$\to \infty \text{ as } d \text{ goes to } \infty,$$

which rules out Equation (21). Ruling out Equation (22) is completely analogous by noticing that $x_1 \to \infty$ as d goes to ∞ by Equation (19), and thus $g_1(d) \to 0$. This proves Proposition 2.4.

6.2 Details About the Potential Method

In this section, we provide missing details and proofs in Section 2.3.

To study correlation decay on trees, we use the standard recursion given in Equation (3). Recall that T is a tree with root v. Vertices $v_1, \ldots, v_d$ are d children of v, and T_i is the subtree rooted by v_i. A configuration σ_Λ is on a subset Λ of vertices, and R_T^σ denote the ratio of marginal probabilities at v given a partial configuration σ on T.

We want to study the influence of another set of vertices, say S, upon v. In particular, we want to study the range of ratios at v over all possible configurations on S. To this end, we define the lower and upper bounds as follows. Notice that as S will be fixed, we may assume that it is a subset of Λ.

Definition 6.1. Let $T, v, \Lambda, \sigma_\Lambda, S, R_T^\sigma$ be as above. Define $R_v := \min_{\tau_\Lambda} R_T^{\tau_\Lambda}$ and $R^v := \max_{\tau_\Lambda} R_T^{\tau_\Lambda}$, where τ_λ can only differ from σ_Λ on S. Define $\delta_v := R^v - R_v$.

Our goal is thus to prove that $\delta_v \le \exp(-\Omega(\text{dist}(v, S)))$. We can recursively calculate R_v and R^v as follows. The base cases are:

(1) $v \in S$, in which case $R_v = 0$ and $R^v = \infty$ and $\delta_v = \infty$;
(2) $v \in \Lambda \setminus S$, i.e., v is fixed to be the same value in all τ_Λ, in which case $R_v = R^v = 0$ (or ∞) if v is fixed to be blue (or green), and $\delta_v = 0$;
(3) $v \notin \Lambda$ and v is the only node of T, in which case $R_v = R^v = \lambda$ and $\delta_v = 0$.

For $v \notin \Lambda$, since F_d is monotonically increasing with respect to any x_i for any $\beta\gamma > 1$,

$$R_v = F_d(R_{v_1}, \ldots, R_{v_d}) \text{ and } R^v = F_d(R^{v_1}, \ldots, R^{v_d}),$$

where R_{v_i} and R^{v_i} are recursively defined lower and upper bounds of $R_{T_i}^{\tau_\Lambda}$ for $1 \le i \le d$.

Our goal is to show that δ_v decays exponentially in the depth of the recursion under certain conditions such as the uniqueness. A straightforward approach would be to prove that δ_v contracts by a constant ratio at each recursion step. This is a sufficient but not necessary condition for the exponential decay. Indeed, there are circumstances that δ_v does not necessarily decay in every step but does decay in the long run. To amortize this behaviour, we use a *potential function* $\Phi(x)$ and show that the correlation of a new recursion decays by a constant ratio.

To be more precise, the potential function $\Phi : \mathbb{R}^+ \to \mathbb{R}^+$ is a differentiable and monotonically increasing function. It maps the domain of the original recursion to a new one. Let $y_i = \Phi(x_i)$. We want to consider the recursion for y_i's. The new recursion function, which is the pullback of F_d, is defined as

$$G_d(y_1, \ldots, y_d) := \Phi(F_d(\Phi^{-1}(x_1), \ldots, \Phi^{-1}(x_d))).$$

The relationship between $F_d(\mathbf{x})$ and $G_d(\mathbf{y})$ is illustrated in Figure 1.

We want to prove Lemmas 2.6 and 2.10. To do so, we also define the upper and lower bounds of y. Define $y_v = \Phi(R_v)$ and accordingly $y_{v_i} = \Phi(R_{v_i})$, for $1 \le i \le d$, as well as $y^v = \Phi(R^v)$ and

$y^{v_i} = \Phi(R^{v_i})$, for $1 \leq i \leq d$. We have that

$$y_v = G_d(y_{v_1}, \ldots, y_{v_d}) \text{ and } y^v = G_d(y^{v_1}, \ldots, y^{v_d}). \tag{23}$$

Let $\varepsilon_v = y^v - y_v$. For a good potential function, exponential decay of ε_v is sufficient to imply that of δ_v.

LEMMA 6.2. *Let* $\Phi(x)$ *be a good potential function for the field* λ *at* v. *Then, there exists a constant* C *such that* $\delta_v \leq C\varepsilon_v$ *for any* $\text{dist}(v, S) \geq 2$.

PROOF. By Equation (23) and the Mean Value Theorem, there exists an $\widetilde{R} \in [R_v, R^v]$ such that

$$\varepsilon_v = \Phi(R^v) - \Phi(R_v) = \Phi'(\widetilde{R}) \cdot \delta_v = \varphi(\widetilde{R}) \cdot \delta_v. \tag{24}$$

Since $\text{dist}(v, S) \geq 2$, we have that $R_v \geq \lambda\gamma^{-d}$ and $R^v \leq \lambda\beta^d$. Hence, $\widetilde{R} \in [\lambda\gamma^{-d}, \lambda\beta^d]$, and by Condition 1 of Definition 2.5, there exists a constant C_1 such that $\varphi(\widetilde{R}) \geq C_1$. Therefore, $\delta_v \leq 1/C_1\varepsilon_v$. □

The next lemma explains Condition 2 of Definition 2.5.

LEMMA 6.3. *Let* $\Phi(x)$ *be a good potential function with contraction ratio* α. *Then,*

$$\varepsilon_v \leq \alpha \max_{1 \leq i \leq d} \{\varepsilon_{v_i}\}.$$

PROOF. First, we use Equation (23):

$$\varepsilon_v = y^v - y_v = G_d(y^{v_1}, \ldots, y^{v_d}) - G_d(y_{v_1}, \ldots, y_{v_d}).$$

Let $\mathbf{y}_1 = (y^{v_1}, \ldots, y^{v_d})$ and $\mathbf{y}_0 = (y_{v_1}, \ldots, y_{v_d})$. Let $\mathbf{z}(t) = t\mathbf{y}_1 + (1 - t)\mathbf{y}_0$ be a linear combination of $\mathbf{y}_0$ and $\mathbf{y}_1$ where $t \in [0, 1]$. Then, we have that

$$\varepsilon_v = G_d(\mathbf{z}(1)) - G_d(\mathbf{z}(0)).$$

By the Mean Value Theorem, there exist $\widetilde{t}$ such that $\varepsilon_v = \frac{d G_d(\mathbf{z}(t))}{d t}|_{t=\widetilde{t}}$. Let $\widetilde{y_i} = \widetilde{t}y^{v_i} + (1 - \widetilde{t})y_{v_i}$ for all $1 \leq i \leq d$. Then, we have that

$$\varepsilon_v = \left| \nabla G_d(\widetilde{y_1}, \ldots, \widetilde{y_d}) \cdot (\varepsilon_{v_1}, \ldots, \varepsilon_{v_d}) \right|. \tag{25}$$

It is straightforward to calculate that

$$\frac{\partial G_d(\mathbf{y})}{\partial y_i} = \frac{\varphi(F_d(\mathbf{R}))}{\varphi(R_i)} \cdot \frac{\partial F_d(\mathbf{R})}{\partial R_i}, \tag{26}$$

where $R_i = \Phi^{-1}(y_i)$ and $\mathbf{y}$ and $\mathbf{R}$ are vectors composed by y_i's and R_i's. Plugging Equation (26) into Equation (25), we get that

$$\varepsilon_v = \varphi(F_d(\widetilde{\mathbf{R}})) \cdot \sum_{i=1}^{d} \left| \frac{\partial F_d}{\partial R_i} \right| \frac{1}{\varphi(\widetilde{R_i})} \cdot \varepsilon_{v_i}$$

$$\leq C_{\varphi, d}(\widetilde{R}_1, \ldots, \widetilde{R}_d) \cdot \max_{1 \leq i \leq d} \{\varepsilon_{v_i}\} \leq \alpha \max_{1 \leq i \leq d} \{\varepsilon_{v_i}\},$$

where $\widetilde{R_i} = \Phi^{-1}(\widetilde{y_i})$, $\widetilde{\mathbf{R}}$ is the vector composed by $\widetilde{R_i}$'s, and in the last line we use Condition 2 of Definition 2.5. □

Note that the two conditions of a good potential function does not necessarily deal with all cases in the tree recursion. At the root, we have one more child than other vertices in a SAW tree. Also, if v has a child $u \in S$, then $\varepsilon_u = \infty$ and the range in both conditions of Definition 2.5 does not apply. To bound the recursion at the root, we have the following straightforward bound of the original recursion.

LEMMA 6.4. *Let (β, γ) be two parameters such that $\beta\gamma > 1$ and $\beta < \gamma$. Let v be a vertex and v_i be its children for $1 \leq i \leq d$. Suppose $\delta_{v_i} \leq C$ for some $C > 0$ and all $1 \leq i \leq d$. Then,*

$$\delta_v \leq d\lambda_v(\beta\gamma - 1)\gamma^{-1}\beta^d C.$$

PROOF. It is easy to see that $\gamma \geq 1$. By the same argument as in Lemma 6.3 and Equation (3), there exists x_i's such that

$$\delta_v = \left|\nabla F_d(x_1, \ldots, x_d) \cdot (\delta_{v_1}, \ldots, \delta_{v_d})\right| \leq C \sum_{i=1}^{d} \left|\frac{\partial F_d(\mathbf{x})}{\partial x_i}\right|,$$

where $\mathbf{x}$ is the vector composed by x_i's. Then, we have that

$$\left|\frac{\partial F_d(\mathbf{x})}{\partial x_i}\right| = \frac{d(\beta\gamma - 1)F_d(\mathbf{x})}{(x_i + \gamma)(\beta x_i + 1)} \leq d\lambda_v(\beta\gamma - 1)\gamma^{-1}\beta^d,$$

where we use the fact that $F_d(\mathbf{x}) \leq \lambda_v \beta^d$ for any $x_i \in [0, \infty)$ and $\beta\gamma > 1$. The lemma follows. □

Now, we are ready to prove Lemma 2.6.

PROOF OF LEMMA 2.6. Given G and a partial configuration σ_Λ on a subset $\Lambda \subseteq V$ of vertices, we first claim that we can approximate $p_v^{\sigma_\Lambda}$ within additive error ε deterministically in time $O(\varepsilon^{\frac{\log \Delta}{\log \alpha}})$. We construct the SAW tree $T = T_{\text{SAW}}(G, v)$. Due to Proposition 2.1, we only need to approximate $p_v^{\sigma_\Lambda}$ in T, with respect to v and an arbitrary vertex set S. We will also use σ_Λ to denote the configuration in T on Λ_{SAW}. Let S be the set of vertices whose distance to v is larger than t, where t is a parameter that we will specify later. Let δ_v be defined as in Definition 6.1 with respect to T, v, Λ, σ_Λ, and S. We want to show that $\delta_v = O(\lambda\alpha^t)$.

The maximum degree of T is at most Δ. Thus the root v has at most Δ children in T, and any other vertex in T has at most $\Delta - 1$ children. Assume v has $k \geq 1$ children, as otherwise we are done. We may also assume that $v \notin S$ and let $t = \text{dist}(v, S) - 1 \geq 1$. We recursively construct a path $u_0 = v, u_1, \ldots, u_l$ of length $l \leq t$ as follows. Given u_i, if there is no child of u_i, then we stop and let $l = i$. Otherwise, u_i has at least one child. If $i = t$, then we stop and let $l = t$. Otherwise, $l < t$ and let u_{i+1} be the child of u_i such that $\varepsilon_{u_{i+1}}$ takes the maximum ε among all children of u_i. In other words, by Lemma 6.3, we have that

$$\varepsilon_{u_i} \leq \alpha\varepsilon_{u_{i+1}}, \tag{27}$$

for all $1 \leq i \leq l - 1$. Notice that Equation (27) may not hold for $i = 0$, since $v = u_0$ has possibly Δ children.

First, we note that for all $1 \leq i \leq l$, $\text{dist}(v, u_i) = i \leq l \leq t$, and therefore $u_i \notin S$. If we met any vertex u_l with no child, then we claim that $\varepsilon_{u_l} = 0$. This is because u_l is either a free vertex with no child or $u_l \in \Lambda$ but $u_l \notin S$. However, since ε_{u_l} takes the maximum ε among all children of u_{l-1}, we have that for all children of u_{i-1}, $\varepsilon = 0$, which implies that $\varepsilon_{u_{i-1}} = 0$. Recursively, we get that $\varepsilon_v = \varepsilon_{u_0} = 0$ and clearly the theorem holds by Equation (24).

Hence, we may assume that $l = t$. Since $u_l \notin S$, we have that $\delta_{u_l} \leq \lambda_{u_l}\beta^{-(\Delta-1)}$ if $\beta > 1$, or $\delta_{u_l} \leq \lambda_{u_l}$ if $\beta \leq 1$. Hence, by Equation (24) and Condition 1 in Definition 2.5, we have that $\varepsilon_{u_l} \leq C_0$ for some constant C_0. Applying Equation (27) inductively, we have that

$$\varepsilon_{u_1} \leq \alpha^l \varepsilon_{u_l} \leq \alpha^t C_0.$$

Hence, by Lemma 6.2, there exists another constant C_1 such that $\delta_{u_1} \leq \alpha^t C_1$. To get a bound on δ_{u_0}, we use Lemma 6.4, which states that

$$\delta_{u_0} \leq d_0\lambda_v(\beta\gamma - 1)\gamma^{-1}\beta^{d_0}\delta_{u_1} \leq d_0\lambda_v(\beta\gamma - 1)\gamma^{-1}\beta^{d_0}\alpha^t C_1 = O(\lambda\alpha^t),$$

where $d_0 \leq \Delta$ is the degree of $v = u_0$.

Hence, the recursive procedure returns R_v and R^v such that $R_v \leq R_T^{\sigma_\Lambda} \leq R^v$, and $R^v - R_v = O(\lambda \alpha^t)$, where $\alpha < 1$ is the contraction ratio. Note that $R_T^{\sigma_\Lambda} = R_{G,v}^{\sigma_\Lambda} = \frac{p_v^{\sigma_\Lambda}}{1 - p_v^{\sigma_\Lambda}}$. Let $p_0 = \frac{R_v}{R_v + 1}$ and $p_1 = \frac{R^v}{R^v + 1}$. Then, $p_0 \leq p_v^{\sigma_\Lambda} \leq p_1$ and

$$p_1 - p_0 = \frac{R^v}{R^v + 1} - \frac{R_v}{R_v + 1} \leq R^v - R_v = O(\lambda \alpha^t). \tag{28}$$

The recursive procedure runs in time $O(\Delta^t)$, since it only needs to construct the first t levels of the self-avoiding walk tree. For any $\varepsilon > 0$, let $t = O(\log_\alpha \varepsilon - \log_\alpha \lambda)$ so that $R^v - R_v < \varepsilon$. This gives an algorithm that approximates $p_v^{\sigma_\Lambda}$ within an additive error ε in time $O((\frac{\varepsilon}{\lambda})^{\frac{\log \Delta}{\log \alpha}})$.

Then, we use self-reducibility to reduce computing $Z_{\beta,\gamma,\pi}(G)$ to computing conditional marginal probabilities. To be specific, let σ be a configuration on a subset of V and τ be sampled according to the Gibbs measure. Let $p_v^\sigma := \Pr(\tau(v) = 1 \mid \sigma)$ be the conditional marginal probability. We can compute $Z_{\beta,\gamma,\pi}(G)$ from p_v^σ by the following standard procedure. Let $v_1, \ldots, v_n$ enumerate vertices in G. For $0 \leq i \leq n$, let σ_i be the configuration fixing the first i vertices $v_1, \ldots, v_i$ as follows: $\sigma_i(v_j) = \sigma_{i-1}(v_j)$ for $1 \leq j \leq i - 1$ and $\sigma_i(v_i)$ is fixed to the spin s so that $p_i := \Pr(\tau(v_i) = s \mid \sigma_{i-1}) \geq 1/3$. This is always possible, because clearly

$$\Pr(\tau(v_i) = 0 \mid \sigma_{i-1}) + \Pr(\tau(v_i) = 1 \mid \sigma_{i-1}) = 1.$$

In particular, $\sigma_n \in \{0,1\}^V$ is a configuration of V. The Gibbs measure of σ_n is $\rho(\sigma_n) = \frac{w(\sigma_n)}{Z_{\beta,\gamma,\pi}(G)}$. However, we can rewrite $\rho(\sigma_n) = p_1 p_2 \cdots p_n$ by conditional probabilities. Thus, $Z_{\beta,\gamma,\pi}(G) = \frac{w(\sigma_n)}{p_1 p_2 \cdots p_n}$. The weight $w(\sigma_n)$ given in Equation (1) can be computed exactly in time polynomial in n. Note that p_i equals to either $p_{v_i}^{\sigma_{i-1}}$ or $1 - p_{v_i}^{\sigma_{i-1}}$. Since we can approximate $p_v^{\sigma_\Lambda}$ within an additive error ε in time $O((\frac{\varepsilon}{\lambda})^{\frac{\log \Delta}{\log \alpha}})$, the configurations σ_i can be efficiently constructed, which guarantees that all p_i's are bounded away from 0. Thus, the product $p_1 p_2 \cdots p_n$ can be approximated within a factor of $(1 \pm n\varepsilon')$ in time $O(n(\frac{\varepsilon'}{\lambda})^{\frac{\log \Delta}{\log \alpha}})$. Now let $\varepsilon' = \frac{\varepsilon}{n}$. We get the claimed FPTAS for $Z_{\beta,\gamma,\pi}(G)$. □

Lemma 2.9 follows almost immediately from Lemmas 6.2, 6.3, and 6.4 as in the proof above. The only issue is that the range of x should be restricted to $(0, \lambda]$. This is guaranteed by Claim 3.4.

Finally, we show Lemma 2.10.

PROOF OF LEMMA 2.10. By the same proof of Lemma 2.6, we only need to approximate the marginal probability at the root v of a tree T. By Condition 2 of Definition 2.8, $C_{\varphi,d}(x_1, \ldots, x_d) < \alpha^{\lceil \log_M(d+1) \rceil}$. Denote by $B(\ell)$ the set of all vertices whose M-based depths of v is at most ℓ in T. Hence, $|B(\ell)| \leq M^\ell$. Let $S = \{u \mid \text{dist}(u, B(\ell)) > 1\}$, which is essentially the same S as in Lemma 2.6, but under a different metric. We can recursively compute upper and lower bounds R^v and R_v of $R_T^{\sigma_\Lambda}$ such that $R_v \leq R_T^{\sigma_\Lambda} \leq R^v$, with the base case that for any vertex $u \in S$ trivial bounds $R_u = 0$ and $R^u = \infty$ are used.

We proceed as in the proof of Lemma 2.6. Without loss of generality, we construct a path $u_0 u_1 \cdots u_k$ in T from the root $u_0 = v$ to a u_k with $\ell_M(u_{k-1}) \leq \ell$ and $\ell_M(u_k) > \ell$. As in the proof of Lemma 6.3, $\varepsilon_{u_j} \leq C_{d_j}^{\varphi}(x_{j,1}, \ldots, x_{j,d_j}) \cdot \varepsilon_{u_{j+1}}$ for all $0 \leq j \leq k - 1$, where d_j is the number of children of u_j and $x_{j,i} \in [0, \infty)$, $1 \leq i \leq d_j$. Hence, we have that

$$\varepsilon_v \leq \varepsilon_{u_k} \cdot \prod_{j=0}^{k-1} \alpha^{\lceil \log_M(d_j+1) \rceil} \leq \varepsilon_{u_k} \cdot \alpha^{\sum_{j=0}^{k-1} \lceil \log_M(d_j+1) \rceil}$$

$$= \varepsilon_{u_k} \cdot \alpha^{\ell_M(u_k)} \leq \varepsilon_{u_k} \cdot \alpha^\ell.$$

Note that $\text{dist}(u_k, B(\ell)) = 1$, and hence $u_k \notin S$. So, $\delta_{u_k} < \lambda_{u_k} \leq \lambda$. By Equation (24), we have that $\varepsilon_{u_k} \leq \varphi(\widetilde{R})\delta_{u_k}$, for some $\widetilde{R} \in [\lambda_{u_k}\gamma^{-d_k}, \lambda_{u_k}\beta^{d_k}]$. Hence, $\varepsilon_{u_k} < C_2\lambda$ by Condition 1 of Definition 2.8, and $\varepsilon_v < \lambda\alpha^{\ell}C_2$. By Equation (24) and Condition 1 of Definition 2.8 again, we have that $\delta_v \leq \lambda\alpha^{\ell}C_2/C_1$.

The rest of the proof goes the same as that of Lemma 2.6. The running time has an extra n^2 factor, since we need to go down two more levels (in the worst case) outside of $B(\ell)$. □

6.3 Proofs of Lemma 3.2 and Lemma 3.3

In this section, we show Lemmas 3.2 and 3.3. We prove Lemma 3.2 first and then use it to show Lemma 3.3.

Proof of Lemma 3.2. It is trivial if $\beta \leq 1$. Now, assume that $\beta > 1$. As $\frac{\beta x+1}{x+\gamma}$ is increasing in x, it is equivalent to show that

$$\frac{\gamma - 1}{\beta - 1} \geq \lambda_c = \left(\frac{\gamma}{\beta}\right)^{\frac{\sqrt{\beta\gamma}}{\sqrt{\beta\gamma}-1}} \qquad \Leftrightarrow \qquad \log(\gamma - 1) - \log(\beta - 1) \geq \frac{\sqrt{\beta\gamma}}{\sqrt{\beta\gamma} - 1} \log\left(\frac{\gamma}{\beta}\right).$$

Let $\gamma = k^2\beta$ with $k \geq 1$. We only need to show that $r(k) \geq 0$ for $k \geq 1$, where $r(k)$ is defined as

$$r(k) := \log(\beta k^2 - 1) - \log(\beta - 1) - \frac{2\beta k}{\beta k - 1}\log k.$$

Since $r(1) = 0$, it is enough to prove that $r(k)$ is increasing for $k \geq 1$. It can be easily verified as

$$r'(k) = \frac{2\beta k}{\beta k^2 - 1} - \frac{2\beta}{\beta k - 1} + \frac{2\beta}{(\beta k - 1)^2}\log k$$

$$= \frac{2\beta}{(\beta k - 1)^2(\beta k^2 - 1)}\left((\beta k^2 - 1)\log k - (k - 1)(\beta k - 1)\right).$$

So, it is sufficient to show that

$$(\beta k^2 - 1)\log k - (k - 1)(\beta k - 1) \geq 0.$$

Since $k \geq 1$, we have that $\log k \geq 1 - \frac{1}{k}$. It implies that

$$(\beta k^2 - 1)\log k - (k - 1)(\beta k - 1) \geq (\beta k^2 - 1)\left(1 - \frac{1}{k}\right) - (k - 1)(\beta k - 1) = \frac{(k - 1)^2}{k} \geq 0.$$

This completes the proof. □

Then, we show Lemma 3.3.

Proof of Lemma 3.3. Let $g(x) := (\beta\gamma - 1)x\log\frac{\lambda_c}{x} - (\beta x + 1)(x + \gamma)\log\frac{x+\gamma}{\beta x+1}$. Hence, it is equivalent to show that $g(x) \leq 0$ for all $0 < x < \lambda_c$. Take the derivative of $g(x)$, and we have that

$$g'(x) = (\beta\gamma - 1)(\log\frac{\lambda_c}{x} - 1) - (2\beta x + \beta\gamma + 1)\log\frac{x + \gamma}{\beta x + 1}$$

$$- (\beta x + 1)(x + \gamma)\left(\frac{1}{x + \gamma} - \frac{\beta}{\beta x + 1}\right)$$

$$= (\beta\gamma - 1)\log\frac{\lambda_c}{x} - (2\beta x + \beta\gamma + 1)\log\frac{x + \gamma}{\beta x + 1}.$$

By direct calculation, $g(\sqrt{\frac{\gamma}{\beta}}) = 0$ and $g'(\sqrt{\frac{\gamma}{\beta}}) = 0$. Then, we prove Equation (5) for the case of $0 < x < \sqrt{\frac{\gamma}{\beta}}$ and $\sqrt{\frac{\gamma}{\beta}} < x < \lambda_c$ separately.

If $0 < x < \sqrt{\frac{\gamma}{\beta}}$, then it is sufficient to verify that $g'(x) > 0$. We only need to show that $g'(x)$ is decreasing, since $g'(\sqrt{\frac{\gamma}{\beta}}) = 0$. It is easily verified by taking the derivative again:

$$g''(x) = -\frac{\beta\gamma - 1}{x} - 2\beta \log \frac{x + \gamma}{\beta x + 1} - (2\beta x + \beta\gamma + 1)\left(\frac{1}{x + \gamma} - \frac{\beta}{\beta x + 1}\right)$$

$$= -2\beta \log \frac{x + \gamma}{\beta x + 1} - (\beta\gamma - 1)\left(\frac{1}{x} - \frac{2\beta x + \beta\gamma + 1}{(x + \gamma)(\beta x + 1)}\right)$$

$$= -2\beta \log \frac{x + \gamma}{\beta x + 1} - (\beta\gamma - 1)\frac{\gamma - \beta x^2}{x(x + \gamma)(\beta x + 1)} < 0,$$

where the last inequality uses the fact that $\frac{x+\gamma}{\beta x+1} \geq 1$ by Lemma 3.2 and $x < \sqrt{\frac{\gamma}{\beta}}$.

If $\sqrt{\frac{\gamma}{\beta}} < x < \lambda_c$, then we show Equation (5) directly. First notice that as $x \neq \sqrt{\frac{\gamma}{\beta}}$,

$$\frac{x}{(\beta x + 1)(x + \gamma)} = \frac{1}{\beta x + \frac{\gamma}{x} + \beta\gamma + 1} < \left(\sqrt{\beta\gamma} + 1\right)^{-2},$$

Given this, to get Equation (5), it is sufficient to show that $h(x) < 0$, where

$$h(x) := \frac{\sqrt{\beta\gamma} - 1}{\sqrt{\beta\gamma} + 1} \log \frac{\lambda_c}{x} - \log \frac{x + \gamma}{\beta x + 1}.$$

In fact, $h(x)$ is a decreasing function as

$$h'(x) = -\frac{\sqrt{\beta\gamma} - 1}{x(\sqrt{\beta\gamma} + 1)} - \frac{1}{x + \gamma} + \frac{\beta}{\beta x + 1}$$

$$= -\frac{(\sqrt{\beta\gamma} - 1)\left((x + \gamma)(\beta x + 1) - (\sqrt{\beta\gamma} + 1)^2 x\right)}{x(\sqrt{\beta\gamma} + 1)(x + \gamma)(\beta x + 1)}$$

$$= -\frac{(\sqrt{\beta\gamma} - 1)\left(\sqrt{\beta}x - \sqrt{\gamma}\right)^2}{x(\sqrt{\beta\gamma} + 1)(x + \gamma)(\beta x + 1)} \leq 0.$$

Notice that $h(\sqrt{\frac{\gamma}{\beta}}) = 0$. It implies that $h(x) < 0$ for all $x > \sqrt{\frac{\gamma}{\beta}}$. This completes the proof. □

ACKNOWLEDGMENTS

REFERENCES

[1] David H. Ackley, Geoffrey E. Hinton, and Terrence J. Sejnowski. 1985. A learning algorithm for Boltzmann machines. *Cogn. Sci.* 9, 1 (1985), 147–169.

[2] Jin-Yi Cai, Andreas Galanis, Leslie Ann Goldberg, Heng Guo, Mark Jerrum, Daniel Štefankovič, and Eric Vigoda. 2016. #BIS-hardness for 2-spin systems on bipartite bounded degree graphs in the tree non-uniqueness region. *J. Comput. Syst. Sci.* 82, 5 (2016), 690–711.

[3] Jin-Yi Cai and Michael Kowalczyk. 2012. Spin systems on k-regular graphs with complex edge functions. *Theor. Comput. Sci.* 461 (2012), 2–16.

[4] Martin E. Dyer, Leslie Ann Goldberg, Catherine S. Greenhill, and Mark Jerrum. 2003. The relative complexity of approximate counting problems. *Algorithmica* 38, 3 (2003), 471–500.

[5] Andreas Galanis, Daniel Štefankovič, and Eric Vigoda. 2016. Inapproximability of the partition function for the anti-ferromagnetic Ising and hard-core models. *Comb. Probab. Comput.* 25, 4 (2016), 500–559.

[6] Hans-Otto Georgii. 2011. *Gibbs Measures and Phase Transitions* (2nd ed.). De Gruyter Studies in Mathematics, Vol. 9. de Gruyter, Berlin.

[7] Leslie Ann Goldberg and Mark Jerrum. 2007. The complexity of ferromagnetic Ising with local fields. *Comb. Probab. Comput.* 16, 1 (2007), 43–61.

[8] Leslie Ann Goldberg, Mark Jerrum, and Mike Paterson. 2003. The computational complexity of two-state spin systems. *Random Struct. Algor.* 23, 2 (2003), 133–154.

[9] Robert B. Griffiths. 1972. Rigorous results and theorems. In *Phase Transitions and Critical Phenomena*, Cyril Domb and Melville S. Green (Eds.). Vol. 1. Academic Press, 7–109.

[10] Heng Guo and Pinyan Lu. 2016. Uniqueness, spatial mixing, and approximation for ferromagnetic 2-spin systems. In *Proceedings of the RANDOM (LIPIcs)*, Vol. 60. 31:1–31:26.

[11] Mark Jerrum and Alistair Sinclair. 1993. Polynomial-time approximation algorithms for the Ising model. *SIAM J. Comput.* 22, 5 (1993), 1087–1116.

[12] Mark Jerrum, Leslie G. Valiant, and Vijay V. Vazirani. 1986. Random generation of combinatorial structures from a uniform distribution. *Theor. Comput. Sci.* 43 (1986), 169–188.

[13] Frank P. Kelly. 1985. Stochastic models of computer communication systems. *J. R. Stat. Soc. Series B Stat. Methodol.* 47, 3 (1985), 379–395.

[14] Liang Li, Pinyan Lu, and Yitong Yin. 2012. Approximate counting via correlation decay in spin systems. In *Proceedings of the SODA*. 922–940.

[15] Liang Li, Pinyan Lu, and Yitong Yin. 2013. Correlation decay up to uniqueness in spin systems. In *Proceedings of the SODA*. 67–84.

[16] Jingcheng Liu, Pinyan Lu, and Chihao Zhang. 2014. The complexity of ferromagnetic two-spin systems with external fields. In *Proceedings of the RANDOM*. 843–856.

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

[18] Elchanan Mossel and Allan Sly. 2013. Exact thresholds for Ising-Gibbs samplers on general graphs. *Ann. Probab.* 41, 1 (2013), 294–328.

[19] Alistair Sinclair, Piyush Srivastava, Daniel Štefankovič, and Yitong Yin. 2017. Spatial mixing and the connective constant: Optimal bounds. *Probab. Theory Related Fields* 168, 1 (2017), 153–197.

[20] Alistair Sinclair, Piyush Srivastava, and Marc Thurley. 2014. Approximation algorithms for two-state anti-ferromagnetic spin systems on bounded degree graphs. *J. Stat. Phys.* 155, 4 (2014), 666–686.

[21] Allan Sly. 2010. Computational transition at the uniqueness threshold. In *Proceedings of the FOCS*. 287–296.

[22] Allan Sly and Nike Sun. 2014. The computational hardness of counting in two-spin models on d-regular graphs. *Ann. Probab.* 42, 6 (2014), 2383–2416.

[23] Dror Weitz. 2006. Counting independent sets up to the tree threshold. In *Proceedings of the STOC*. 140–149.

[24] Jinshan Zhang, Heng Liang, and Fengshan Bai. 2011. Approximating partition functions of the two-state spin system. *Inf. Process. Lett.* 111, 14 (2011), 702–710.