Lecture Note 17

Instructor: Alistair Sinclair

17.1 Correlation Decay

We now move from MCMC to another technique called correlation decay, which leads to *deterministic* algorithms for approximating partition functions. Take the canonical problem of approximating the partition function of a spin system. Intuitively, if the interaction between neighboring spins is weak enough, then they behave "almost independently"; in particular, the value of two spins that are far apart in the graph should be asymptotically uncorrelated. This should in turn mean that we can approximate the marginal distribution of the spin at any given vertex v by explicit enumeration of the configurations in a small neighborhood around v. And this would enable us to approximate the partition function by pinning that spin value and recursing, as we saw earlier in the course. In the next few lectures, we'll illustrate this with the canonical example of the *hard core model*, or independent set problem.

Given a graph $G = (V, E)$, let

$$Z_G(\lambda) = \sum_{\text{Ind Sets } I} \lambda^{|I|} = \sum_{k \geq 0} a_k \lambda^k,$$

where a_k is the number of independent sets (of vertices) of size k and $\lambda > 0$ is a parameter known as the *fugacity*. Recall also that $Z_G(\lambda)$ corresponds to a Gibbs distribution, in which the probability of an independent set of size k is $\frac{\lambda^k}{Z_G(\lambda)}$.

Let's start by giving a simple reduction to show that this problem is hard to approximate on general graphs G at any fixed value $\lambda > 0$. We reduce the decision problem IndSet to approximating $Z_G(\lambda)$. Given a graph G and integer k_0, IndSet asks whether there is an independent set of size at least k_0 in G.

Given such a graph $G = (V, E)$, we replace each vertex v with a "cloud" C_v of independent vertices of size r; then, for all adjacent pairs $\{v, v'\} \in E$, we connect the corresponding clouds C_v and $C_{v'}$ with a complete bipartite graph $K_{r,r}$. Let G_r denote this blown-up version of G. Note that we can view each independent set I' of G_r as a *witness* for a unique independent set I in G, where $I = \{v \in V : C_v \cap I' \neq \emptyset\}$. Then if I has size k, the total weight of all the witnesses for I is $((1 + \lambda)^r - 1)^k$. (**Exercise:** check this.) As a result, we have

$$Z_{G_r}(\lambda) = \sum_{k \geq 0} a_k \left((1 + \lambda)^r - 1\right)^k.$$

Now we pick r large enough so that

$$\left((1 + \lambda)^r - 1\right)^{k_0} \gg 2^n \left((1 + \lambda)^r - 1\right)^{k_0 - 1}. \tag{17.1}$$

This ensures that the weight of a single independent set of size k_0 in G dominates the aggregated weight of all independent sets of size strictly less than k_0. Now for any fixed $\lambda > 0$, we can ensure (17.1) by setting

$r = cn$ for a constant c (depending on λ), where $n = |V|$ is the number of vertices in G. Hence the reduction to G_r can be computed in polynomial time.

Now we can decide if G contains an independent set of size at least k_0 by computing an approximation to $Z_{G_r}(\lambda)$ and comparing the result with the value $((1+\lambda)^r - 1)^{k_0}$.

We have proved:

Proposition 17.1. *If there exists a FPRAS (resp., and FPTAS) for $Z_G(\lambda)$ for some fixed $\lambda > 0$, then there is a randomized (resp., deterministic) polynomial-time algorithm for IndSet, and hence* RP = NP *(resp.,* P = NP*).*

The hard-core model is therefore uninteresting for general graphs. However, if we restrict attention to graphs of bounded degree (a natural restriction in spin systems, where each spin usually has just a bounded number of neighbors), then a beautiful picture emerges, as embodied in the following theorem.

Theorem 17.2. *For graphs G of max degree $\Delta \geq 3$, there exists a threshold $\lambda_c(\Delta) = \frac{(\Delta-1)^{(\Delta-1)}}{(\Delta-2)^{\Delta}}$ such that*

1. [Weitz [Wei06]] *If $\lambda < \lambda_c(\Delta)$, there exists a FPTAS for approximating $Z_G(\lambda)$.*
2. [Sly [Sly10]] *If $\lambda > \lambda_c(\Delta)$, there is no FPRAS for $Z_G(\lambda)$, unless RP = NP.*

Note that it is natural to expect the problem to get harder as λ gets larger, because typical independent sets have to pack in more vertices, meaning that correlations between distant vertices may exist. Hence we might expect the correlation decay idea sketched earlier to fail for large λ.

What is remarkable about the above theorem is that there is a threshold value for λ so that, when λ exceeds that threshold, not only does the correlation decay method fail, but *no* polynomial time (randomized) algorithm works (unless RP = NP). And, moreover, all the way up to the threshold correlation decay *does* work.

This threshold $\lambda_c(\Delta)$ turns out to be the so-called "uniqueness threshold" for the infinite Δ-regular tree. This means that, in a certain precise sense, trees are the worst case graphs for this problem among all graphs of maximum degree Δ.

The "uniqueness threshold" is the value of λ below which the associated Gibbs measure on the infinite tree is unique. To make this more precise, we introduce the following definition:

Definition 17.3. *Let $S \subseteq V$ be any subset of vertices in the infinite Δ-regular tree, and τ be an arbitrary configuration of "pinned spins" on S (i.e., vertices set to be either in the independent set or not, also called a "boundary condition"). Let p_r^τ denote the probability, under the associated Gibbs measure on independent sets on the tree, that the root is occupied. Then* Weak Spatial Mixing (WSM) *holds if there exists a constant $c > 0$ such that*

$$|p_r^\tau - p_r^{\tau'}| \leq \exp(-c \cdot \mathrm{dist}(r, S))$$

for any two configurations τ, τ' on S.

In other words, WSM says that the effect of any set of fixed spins on the spin at the root (i.e., the probability that the root is in the independent set) decays exponentially to zero with the distance of the fixed spins from the root. How do we construct a Gibbs measure on the infinite tree? Slightly informally, we consider an increasing sequence $T_1, T_2, \ldots, T_\ell, \ldots$ of finite Δ-regular trees of increasing depths ℓ, that converges to the infinite tree. We place some arbitrary fixed boundary condition τ outside T_ℓ (the same for all ℓ: think of this as being inherited from a single fixed configuration on the infinite tree), and look at the sequence of Gibbs distributions on the finite trees T_ℓ. Then WSM ensures that the resulting infinite-volume Gibbs measure, constructed as the limit of the finite Gibbs distributions on the trees T_ℓ, is independent of the boundary condition, and thus unique. (This can all be made precise using appropriate topological notions that we won't get into here.)

Conversely, in the absence of WSM, the boundary condition may have a non-zero effect on the spin at the root even as the depth $\ell \to \infty$. In this case we get multiple Gibbs measures, depending on the boundary condition. Specifically, for the hard-core model, we will get precisely two *extremal* such Gibbs measures, one corresponding to all leaves of the tree at even depths ℓ being occupied, and the other corresponding to all leaves at odd depths being occupied. (Of course, there will also be infinitely many other measures obtained as convex combinations of these two extremal ones.)

This sharp change in the behavior of the Gibbs measure at $\lambda = \lambda_c(\Delta)$ is an example of a *phase transition*: for $\lambda > \lambda_c(\Delta)$, two phases (the "odd phase" and the "even phase") emerge in the model, whereas for $\lambda < \lambda_c(\Delta)$ there is a single phase.

Theorem 17.2 can therefore be seen as a *computational* manifestation of this spatial phase transition. In the next couple of lectures we'll show how to prove both parts of the theorem.

17.2 The uniqueness threshold $\lambda_c(\Delta)$.

Let $d = \Delta - 1$ (so that we're working with the infinite d-ary tree). Then

$$\lambda_c(\Delta) = \frac{d^d}{(d-1)^{d+1}} \sim \frac{e}{\Delta},$$

as $\Delta \to \infty$.

To understand where $\lambda_c(\Delta)$ comes from, we revisit a classical analysis of Spitzer [Spi75] and Kelly [Kel85]. Given an infinite d-ary tree T_v rooted[1] at v, define $p_v := \Pr_{T_v}[\sigma(v) = 0]$ be the probability of v being unoccupied. We can express p_v in terms of restricted partition functions (with obvious notation) as follows[2]:

$$p_v = \frac{Z_{T_v}[\sigma(v) = 0]}{Z_{T_v}[\sigma(v) = 0] + Z_{T_v}[\sigma(v) = 1]}.$$

Then we can use the tree structure to write these partition functions in terms of those of the subtrees rooted at the children $v_1, \ldots, v_d$ of v:

$$p_v = \frac{\Pi_{i=1}^d Z_{T_{v_i}}}{\Pi_{i=1}^d Z_{T_{v_i}} + \lambda \Pi_{i=1}^d Z_{T_{v_i}}[\sigma(v_i) = 0]}.$$

To see this expression, note that the subtrees are independent given the value $\sigma(v)$ at the root, and when $\sigma(v) = 1$ we must have $\sigma(v_i) = 0$ for all i and we also pick up a factor of λ. Dividing all terms by the numerator, this becomes

$$p_v = \frac{1}{1 + \lambda \Pi_{i=1}^d p_{v_i}}. \tag{17.2}$$

Now if we postulate the existence of a translation-invariant Gibbs measure on the infinite tree (i.e., p_v is the same for all vertices v), then this value p_v must be a fixed point of the recurrence

$$f(x) = \frac{1}{1 + \lambda x^d}.$$

[1] Technically, the infinite Δ-regular tree doesn't have a root; however, we will gloss over this minor detail as the error introduced by having a root is negligible.

[2] Again, technically these partition functions are infinite, but we can just think of a very large finite tree and take limits.

Denote the fixed point of $f(x)$ by x^* (i.e., $f(x^*) = x^*$). Note that x^* is unique because $f(x)$ is monotonically decreasing on $[0, 1]$. When λ is small, the (negative) slope of $f(x)$ at x^* is not very steep. This means that, if we iterate the recurrence, we will spiral in to x^*: i.e., x^* is a *stable* fixed point, and the recurrence will converge to it from any initial values at the leaves as the depth of the tree goes to infinity. Conversely, when λ is large the negative slope at x^* becomes so steep that the recurrence spirals away from the fixed point, which is unstable. So in this case the recurrence does not converge. See Figure 17.1 for a sketch of these behaviors.

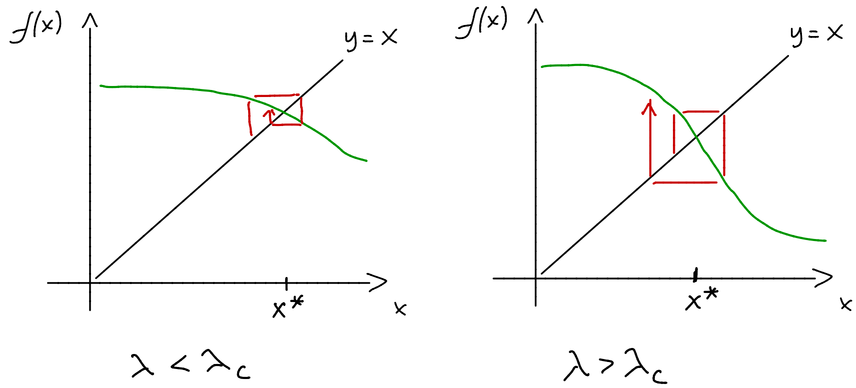

Figure 17.1: Sketch of graph of function f on $[0, 1]$, for small and large values of λ. The red trajectory shows the process of iterating f. For $\lambda < \lambda_c$ the fixed point x^* is stable, while for $\lambda > \lambda_c$ it is not.

It turns out that the feature that characterizes these two different behaviors is the value of the derivative of f at the fixed point: if $f'(x^*) > -1$ then the fixed point is stable, while if $f'(x^*) < -1$ then it is unstable. And the critical value of λ at which this change in the derivative occurs is precisely $\lambda_c(\Delta)$.

One way to see this more rigorously is to consider the two-step recurrence $f \circ f$, depicted in the two regimes in Figure 17.2. For small λ this function has just one fixed point (which is stable and coincides with the unique fixed point for f itself), while for large λ it has three fixed points: the (unstable) fixed point of f, together with two stable fixed points, corresponding to the odd and even Gibbs measures respectively. (In these measures, $f \circ f$ converges to two different possible values according to the initial value at the leaves. These measures are *semi-translation invariant*: all vertices on odd levels look the same, as do all vertices on even levels.) Note that these two behaviors are clearly characterized by whether or not $(f \circ f)'(x^*) < 1$.

To see the earlier claim about the derivative of f itself, note that

$$(f \circ f)'(x) = f'(f(x))f'(x).$$

At the fixed point x^* we get

$$(f \circ f)'(x^*) = f'(x^*)f'(x^*) = f'(x^*)^2.$$

So $(f \circ f)'(x^*) < 1$ is equivalent to $|f'(x^*)| < 1$, i.e., $f'(x^*) > -1$, as claimed above (since $f'(x) < 0$ for all x).

Finally, to see that $\lambda = \lambda_c(\Delta)$ is the threshold for these two regimes, we look at solutions of the following two equations:

$$x^* = f(x^*) \implies x^* = \frac{1}{1 + \lambda x^{*d}} \qquad \text{(the fixed point equation)};$$

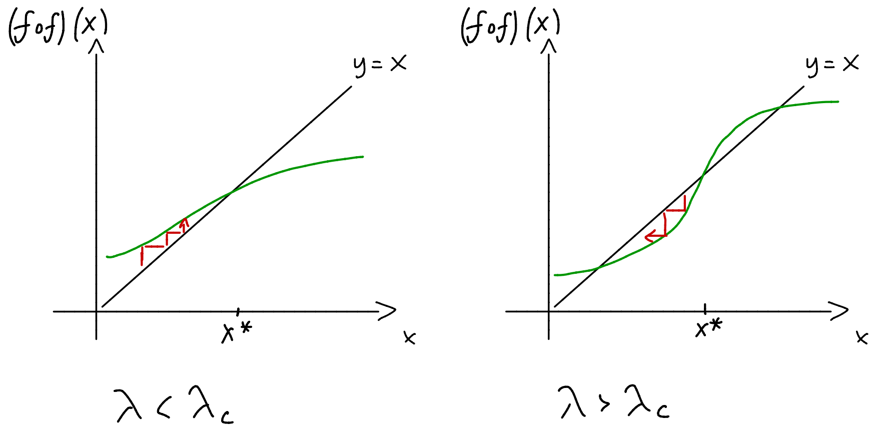

Figure 17.2: Sketch of graph of function $f \circ f$ on $[0, 1]$, for small and large values of λ. When $\lambda < \lambda_c$ there is a single (stable) fixed point, x^*; when $\lambda > \lambda_c$ the fixed point x^* becomes unstable and there are two additional stable fixed points.

and

$$f'(x^*) = -1 \implies \frac{-\lambda d x^{*(d-1)}}{(1 + \lambda x^{*d})^2} = -1 \qquad \text{(derivative condition)}.$$

Plugging the first equation into the second simplifies it to $d(1-x^*) = 1$, so that $x^* = \frac{d-1}{d}$. The first equation then implies that $\lambda = \lambda_c(d) = \frac{d^d}{(d-1)^{d+1}}$. Also, it's easy to verify (**exercise!**) that x^* is monotonically decreasing in λ, and hence that $f'(x^*)$ is monotonically decreasing in λ, which implies that $f'(x^*) > -1$ for $\lambda < \lambda_c(d)$ and $f'(x^*) < -1$ for $\lambda > \lambda_c(d)$, as claimed earlier.

17.3 An FPTAS when $\lambda < \lambda_c(\Delta)$

Now we go back to prove part (i) of Theorem 17.2, i.e., if $\lambda < \lambda_c(\Delta)$ then there exists an FPTAS for approximating $Z_G(\lambda)$. This is a direct result of Weitz's algorithm in [Wei06]. The outline of Weitz's algorithm is as follows:

1. Given a general graph G of maximum degree Δ, and an arbitrary vertex v, we construct the Self-Avoiding Walk Tree of G rooted at v, denoted $T_{SAW}(G, v)$. This tree will have the property that $\Pr[\sigma(v) = 1]$ is the same in G as in the tree! The key to the construction of the tree is the introduction of carefully chosen boundary conditions (fixed spins) on certain leaves.

2. Show that for $\lambda < \lambda_c(\Delta)$, the above correlation decay property (spatial mixing) that holds in the infinite Δ-regular tree continues to hold in this new tree $T_{SAW}(G, v)$. (In the presence of boundary conditions, this is not at all obvious.)

3. Use #1 and #2 to get an FPTAS by truncating T_{SAW} at depth $O(\log n)$, so that the computation of the marginal at the root v can be done exactly on the truncated tree, since it has polynomial size and no cycles. Property #2 then guarantees that the error due to truncation will be small.

We start with the construction of $T_{SAW}(G, v)$. This tree enumerates all *self-avoiding walks* in G starting at v; this means that we enumerate all walks, but we terminate each walk when it revisits a previous vertex (closes a cycle). At that point, we introduce a leaf with a certain boundary condition. (Trivial cycles of length 2 don't count, so we never revisit a vertex immediately after leaving it.)

Here is the rule for assigning boundary conditions. Suppose the tree encounters a cycle starting and ending at u, i.e., $u \to w_1 \to w_2 \to w_3 \cdots \to w_\ell \to u$ where the w_i are distinct vertices. Then we make the second occurrence of u a leaf, and we assign the boundary condition at this leaf according to the order in which we traverse the cycle: assuming some ordering on the neighbors of each vertex, we assign the spin at the leaf to be 0 if $w_1 < w_l$ in this order at u, and 1 otherwise. Since each cycle can be traversed in two directions, each cycle will give rise to two leaves with opposite boundary conditions (corresponding to the two orders of traversal).

We illustrate this construction with a concrete example in Figure 17.3.

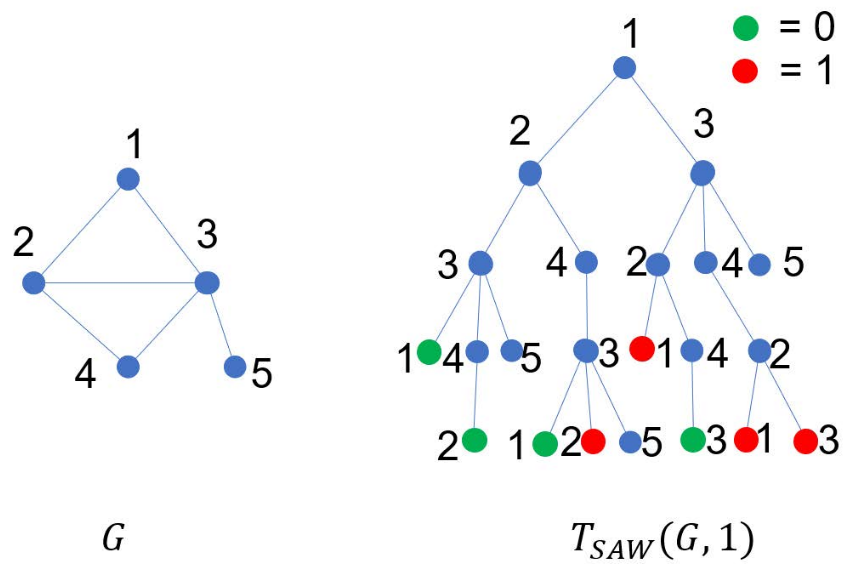

Figure 17.3: Illustration of the construction of the SAW tree $T_{SAW}(G,1)$ from the graph G. The root-leaf paths of $T_{SAW}(G,1)$ correspond precisely to self-avoiding walks in G starting at vertex 1. When a walk terminates (closes a cycle), we assign either a 0 (green) or a 1 (red) boundary condition as specified by the rule in the text. Vertices that are leaves in G (like vertex 5) become leaves in the tree with no boundary condition.

Now we are ready to prove that with this construction, the probability of the root vertex being occupied in the tree is the same as that in the graph, which is the key guarantee in point #1 of the outline above.

Claim 17.4. $\Pr_G[\sigma(v) = 0] = \Pr_{T_{SAW}(G,v)}[\sigma(v) = 0]$.

Proof. Throughout the proof, we denote the SAW tree $T_{SAW}(G,v)$ by T_v, and the subtree rooted at any interior vertex u by T_u. Also, to simplify the algebra we will work with occupation ratios rather than probabilities. Denote

$$R_v = \frac{\Pr[\sigma(v) = 1]}{\Pr[\sigma(v) = 0]}.$$

It suffices to prove that $R_v^G = R_v^{T_v}$, since this clearly implies the claim in the theorem.

For the SAW tree, following essentially the same computation that led us to (17.2), we can write the following recurrence (**exercise!**) for the occupation ratio at the root in terms of the occupation ratios of the roots of the subtrees (denoted now by u_i):

$$R_v^{T_v} = \frac{Z_{T_v}[\sigma(v)=1]}{Z_{T_v}[\sigma(v)=0]} = \lambda \prod_{i=1}^{d} \frac{1}{1+R_{u_i}^{T_{u_i}}}. \tag{17.3}$$

Turning now to G, to get an expression for R_v^G we first split the vertex v into d copies, $v_1, v_2, \cdots, v_d$, where each v_i is connected only to u_i. This gives us a new graph G' in which all the v_i are leaves. Now we can write the occupation ratio in G as:

$$R_v^G = \frac{\Pr_G[\sigma(v)=1]}{\Pr_G[\sigma(v)=0]} = \lambda^{-(d-1)} \frac{\Pr_{G'}[\sigma(v_1)=\cdots=\sigma(v_d)=1]}{\Pr_{G'}[\sigma(v_1)=\cdots=\sigma(v_d)=0]}.$$

The factor $\lambda^{-(d-1)}$ comes from the fact that v being occupied in G gives only a factor of λ, while all the v_i being occupied in G' gives a factor of λ^d. We can further rewrite the above formula via a cascading product of conditional probabilities:

$$\begin{aligned} R_v^G = \lambda^{-(d-1)} & \frac{\Pr_{G'}[\sigma(v_1)=1|\sigma(v_2)=\cdots=\sigma(v_d)=0]}{\Pr_{G'}[\sigma(v_1)=0|\sigma(v_2)=\cdots=\sigma(v_d)=0]} \times \frac{\Pr_{G'}[\sigma(v_2)=1|\sigma(v_1)=1,\sigma(v_3)=\cdots=\sigma(v_d)=0]}{\Pr_{G'}[\sigma(v_2)=0|\sigma(v_1)=1,\sigma(v_3)=\cdots=\sigma(v_d)=0]} \\ & \times\cdots\times \frac{\Pr_{G'}[\sigma(v_d)=1|\sigma(v_1)=\cdots=\sigma(v_{d-1})=1]}{\Pr_{G'}[\sigma(v_d)=0|\sigma(v_1)=\cdots=\sigma(v_{d-1})=1]}. \end{aligned} \tag{17.4}$$

(To see this, use the fact that $\frac{\Pr[AB]}{\Pr[CD]} = \frac{\Pr[A|D]}{\Pr[C|D]} \cdot \frac{\Pr[B|A]}{\Pr[D|A]}$ for any four events A, B, C, D.) Now (17.4) can be written as

$$R_v^G = \lambda^{-(d-1)} \prod_{i=1}^{d} R_{v_i}^{G',\tau_i}, \tag{17.5}$$

where R_v^{G',τ_i} denotes the occupation ratio of v_i in G' with the boundary condition τ_i given by

$$\tau_i(v_j) = \begin{cases} 1 & j<i; \\ 0 & j>i. \end{cases} \tag{17.6}$$

Now since v_i is a leaf in G', we can write its occupation ratio just in terms of that of its single neighbor u_i, i.e.,

$$R_{v_i}^{G',\tau_i} = \frac{\lambda Z_{G'-v_i}^{\tau_i}[\sigma(u_i)=0]}{Z_{G'-v_i}^{\tau_i}[\sigma(u_i)=0] + Z_{G'-v_i}^{\tau_i}[\sigma(u_i)=1]} = \frac{\lambda}{1+R_{u_i}^{G'-v_i,\tau_i}}.$$

Plugging in the above formula to (17.5), we have

$$R_v^G = \lambda^{-(d-1)} \prod_{i=1}^{d} \frac{\lambda}{1+R_{u_i}^{G'-v_i,\tau_i}} = \lambda \prod_{i=1}^{d} \frac{1}{1+R_{u_i}^{G'-v_i,\tau_i}}. \tag{17.7}$$

But now comparing the expressions in (17.7) and (17.3), we see that they are identical in form! Moreover, it is not hard to check (**exercise!**) that the SAW tree of the graph $G'-v_i$ rooted at u_i, with boundary conditions τ_i at the leaves v_j, is precisely the same as the subtree T_{u_i} of T_v. (In particular, the boundary condition τ_i defined in (17.6) corresponds precisely to the rule for closing cycles at v specified in the construction of the SAW tree.) Thus inductively we can argue that $R_{u_i}^{G'-v_i,\tau_i} = R_{u_i}^{T_{u_i}}$, implying that indeed $R_v^G = R_v^{T_v}$, as required. □

Exercise: You are strongly encouraged to hand-turn the above proof on the toy example in Figure 17.3 to get a feel for how it works.

17.4 Strong spatial mixing

Given a graph G of maximum degree Δ, Weitz's algorithm constructs the *self-avoiding walk tree* $T_{SAW}(G, v)$ rooted at a vertex v, with the property that the occupation probability of v under the Gibbs distribution on G is exactly the same as the occupation probability of the root in the SAW tree. This means that we can compute this probability in the tree rather than in the graph, which we can do using a simple recursive procedure.

However, the SAW tree will in general have exponential size (compared to the size of G itself), so we can't work with the entire tree. Instead, we want to *truncate* the tree at depth $O(\log n)$, so that its size will be $\Delta^{O(\log n)}$, which is polynomial in n, so we can do the recursive computation on it. This will only work if the error introduced by the truncation is small. To prove that this is the case, we can appeal to the spatial mixing property (because we are assuming $\lambda < \lambda_c(\Delta)$), which ensures exponential decay of correlations and hence that the error at depth $O(\log n)$ is inverse polynomial.

There is, however, one crucial issue we have to address: we saw that $\lambda < \lambda_c(\Delta)$ implies exponential decay of correlations *in the* Δ*-regular tree without boundary conditions*. However, in the SAW tree there are boundary conditions, and some of them may be very close to the root (due to short cycles in G), so they will still be present after we truncate the tree.

To address this issue, we need a stronger notion of decay of corrections that holds even in the presence of boundary conditions close to the root.

Definition 17.5. *Let $S \subset V$ be any subset of vertices of the infinite tree, and τ be any fixed configuration on S. Let p_r^τ denote the probability, in the Gibbs measure of the hard-core model on the tree, that the root is occupied, given the fixed boundary condition τ. Then* Strong Spatial Mixing (SSM) *holds if there exists a constant $c > 0$ s.t.*

$$|p_r^\tau - p_r^{\tau'}| \le \exp(-c \cdot dist(r, R)).$$

for any two configurations τ, τ' on S that agree on $S \setminus R$.

This definition is almost identical to that of Weak Spatial Mixing above, except that SSM requires the correlation to decay exponentially only with the distance to the *disagreeing portion* of the boundary conditions, even if τ, τ' have (the same) fixed spins close to the root. It's clear from the above definition that SSM implies WSM, since $\text{dist}(r, S) \le \text{dist}(r, R)$. But the converse is not true in general. Even though it may seem that the presence of additional (fixed) boundary conditions can only make the root less sensitive to distant boundary conditions, this is not the case, e.g., for the ferromagnetic Ising model, where it is possible to construct examples in which WSM holds but SSM does not. (Essentially the examples use boundary conditions to "shift" the occupation probability of the root into a range of values where it is *more* sensitive to the spins at the distant leaves; see, e.g., [SST14].)

The following theorem states that, in the case of the hard-core model, WSM in fact does imply SSM.

Theorem 17.6. *For the hard-core model on T_Δ, SSM holds for all $\lambda < \lambda_c(\Delta)$.*

Before we prove this key theorem, let's complete the description and analysis of the algorithm assuming that SSM holds. Here is the algorithm:

1. Given G and a vertex v of G, construct the first ℓ levels of $T_{SAW}(G, v)$, where $\ell = \frac{1}{c}\ln(\frac{5n}{\varepsilon})$, with c being the constant in the definition of SSM.
2. Assign unoccupancy probabilities to the leaves as follows. Note that there are two kinds of leaves: those that are created by truncation, and those that were leaves in the original tree (above the point where we do the truncation).

- The leaf nodes that aren't created by truncation either already have a boundary condition, or they are free (in which case they were leaves in G). If a leaf is occupied we assign it value 0, and if it's unoccupied we assign it 1. (Recall that we're computing the probabilities that vertices are *un*occupied.) If a leaf is free we assign it the value $\frac{1}{1+\lambda}$, which is the probability that a vertex with no occupied neighbors is unoccupied. (This is the appropriate value to assign to a leaf, since it has no children.) We note also for future reference that the leaves with boundary conditions can be removed as follows: an unoccupied leaf can just be removed; an occupied leaf can be removed, together with its parent. This leaves just free leaves, each of which gets the same value $\frac{1}{1+\lambda}$. However, the shape of the tree is now highly non-uniform, reflecting the effect of the boundary conditions.
- The leaves that were created artificially by truncation in step 1 are assigned an arbitrary value in the range $[\frac{1}{1+\lambda}, 1]$, which is the range of legal unoccupation probabilities.

3. Use the tree recurrence on the SAW tree $T = T_{SAW}(G, v)$ to compute $\Pr_T[\sigma(v) = 0]$, the unoccupation probability of the root v.
4. Fix $\sigma(v) = 0$, and repeat the above for another vertex v' until we exhaust all vertices in G. (Note that the fixed spin $\sigma(v) = 0$ is just another boundary condition, so the algorithm works as before.)

In this way, we end up with an estimate of the probability that all vertices are unoccupied, which is just the probability $\pi[\emptyset]$ of the empty set in the Gibbs measure. (The empty set is always an independent set.) But since $\pi[\emptyset] = 1/Z_G(\lambda)$, taking the reciprocal of our estimate immediately gives us an estimate for $Z_G(\lambda)$ with the same (multiplicative) accuracy. Furthermore, in order to get a $1 \pm \varepsilon$ estimate of $Z_G(\lambda)$, we need to get a $1 \pm \frac{\varepsilon}{n}$ estimate of each marginal $\Pr_G[\sigma(v) = 0]$. If we can show that we can obtain each such estimate in polynomial time, then we will have an FPTAS. Note that this algorithm is deterministic.

Claim 17.7. *The running time of the above procedure is polynomial in n and ε^{-1}.*

Proof. The size of the the tree is

$$O(\Delta^l) = O\left(\Delta^{\frac{1}{c}\ln(\frac{5n}{\varepsilon})}\right) = O\left((\frac{5n}{\varepsilon})^{\frac{1}{c}\ln\Delta}\right)$$

Clearly, this is polynomial in n and $\frac{1}{\varepsilon}$ for any fixed Δ (though the degree of the polynomial increases with Δ). $\square$

Claim 17.8. *The accuracy for each marginal is $1 \pm \frac{\varepsilon}{n}$.*

Proof. Redefine p_v^τ to denote the *un*occupation probability of the root v with boundary condition τ. For any two sets of boundary conditions τ, τ' that differ only on the *truncated* leaves, and are legal on those leaves, by SSM we have

$$|p_v^\tau - p_v^{\tau'}| \le \exp(-c \cdot l) = \exp\left(-c \cdot \frac{1}{c} \cdot \ln(\frac{5n}{\varepsilon})\right) = \frac{\varepsilon}{5n}.$$

But since the minimum unoccupation probability for any vertex is $\frac{1}{1+\lambda}$, we have

$$p_v^\tau, p_v^{\tau'} \ge \frac{1}{1+\lambda} \ge \frac{1}{5},$$

using the fact that $\lambda < \lambda_c(\Delta) \le \lambda_c(3) = 4$. Hence

$$\left|\frac{p_v^\tau}{p_v^{\tau'}} - 1\right| \le \frac{\varepsilon}{n}.$$

But now we're done because if *any* pair of legal boundary conditions give estimates that are within $\le 1+\frac{\varepsilon}{n}$ of each other, then the estimate obtained with any arbitrary legal boundary condition (which is what the algorithm outputs) must be within $\le 1 \pm \frac{\varepsilon}{n}$ of the true value. $\square$

17.5 Proof of Theorem 17.6

In this section, we will prove Theorem 17.6, namely that on the Δ-regular tree, for $\lambda < \lambda_c(\Delta)$, SSM holds. This is the main technical meat in the analysis of Weitz's algorithm. Weitz's original proof is ingenious but tailored to the hard-core model. Instead we use technology that has since proved useful in proving the SSM property for a wider range of models. This particular version is due to Restrepo *et al.* [RST$^+$13] and is typical of these arguments.

Let $d = \Delta - 1$, so that we are working on the d-ary tree. Consider an arbitrary node v in the tree, along with its children $\{v_1, \dots, v_d\}$. Let p_v denote the probability that v is unoccupied. (We suppress the boundary condition τ for ease of notation.) Following the same calculation as earlier, we can write the following recurrence for p_v:

$$p_v = \frac{1}{1 + \lambda \prod_{i=1}^{d} p_{v_i}} =: f(p_{v_1}, \dots, p_{v_d}). \tag{17.8}$$

We want to show that the distance between any two valid computations on the tree decreases at a uniform rate under this recurrence, i.e., for some $\gamma > 0$:

$$|p_v - p'_v| \le (1-\gamma) \max_i |p_{v_i} - p'_{v_i}|. \tag{17.9}$$

This will imply that

$$|p_r - p'_r| \le (1-\gamma)^l \max_{\text{Leaves } L_i} |p_{L_i} - p'_{L_i}|,$$

which is exactly what we need to show for SSM.

Unfortunately, however, the stepwise decay property (17.9) doesn't hold in general for a uniform γ. But we can get around this by considering instead the decay of distance under some function of p_v.

Definition 17.9. *A* message *(or* potential*) is a continuously differentiable function $\phi : [\frac{1}{1+\lambda}, 1] \to \mathbb{R}$ with positive derivative. Hence ϕ is increasing and invertible on its range, and ϕ^{-1} is also continuously differentiable with positive derivative.*

The choice of an appropriate message is problem-dependent, and not very well understood. For this application to the hard-core model, the following message works:

$$\phi(x) := \frac{1}{s} \log \frac{x}{s-x}, \quad \text{where } s := \frac{d+1}{d}.$$

Note: A plausible derivation of this message is as a modification of the simpler message $\phi(x) = \log \frac{x}{1-x}$ that arises naturally in the analysis of the ferromagnetic Ising model with zero field.

Abbreviating p_{v_i} to p_i and writing $m = \varphi(p_v)$ and $m_i = \phi(p_i)$, we can turn the recurrence (17.8) into a recurrence on messages, as follows:

$$m = \phi\Big(f\big(\phi^{-1}(m_1), \dots, \phi^{-1}(m_d)\big)\Big) =: F(m_1, \dots, m_d).$$

We will write $\mathbf{m}$ to denote the vector of values $(m_1, \dots, m_d)$.

Claim 17.10. $\exists \gamma > 0$ *s.t.* $\forall \mathbf{m}, \mathbf{m}' \in (\phi[\frac{1}{1+\lambda}, 1])^d$,

$$|F(\mathbf{m}) - F(\mathbf{m}')| \le (1-\gamma)\|\mathbf{m} - \mathbf{m}'\|_\infty$$

First we will show that proving Claim 17.10 is sufficient to prove Theorem 17.6. Then we will go back and prove the Claim.

Lemma 17.11. *Claim 17.10* $\Longrightarrow$ *Theorem 17.6*

Proof. Assuming Claim 17.10 is true, we have for any l,

$$|m_r - m_r'| \le (1-\gamma)^l \max_{\text{leaves } L_i \text{ at depth } l} |m_{L_i} - m_{L_i}'|. \tag{17.10}$$

We just have to translate this to a similar statement with m replaced by p. Since ϕ is monotone,

$$\begin{aligned}
\max |m_{L_i} - m'_{L_i}| &\le \phi(1) - \phi\Big(\frac{1}{1+\lambda}\Big) \\
&= \frac{1}{s}\Big[\log \frac{1}{s-1} - \log \frac{(1+\lambda)^{-1}}{s-(1+\lambda)^{-1}}\Big] \\
&= \frac{1}{s}\log\Big(\frac{s+s\lambda-1}{s-1}\Big) \\
&\le c \log d,
\end{aligned} \tag{17.11}$$

for a constant c, since $\lambda \le 4$ and $s = \frac{d+1}{d}$. Next, by the mean value theorem applied to ϕ:

$$|m_v - m_v'| = |\phi(p_v) - \phi(p_v')| \ge |p_v - p_v'| \inf_{p \in [\frac{1}{1+\lambda}, 1]} \phi'(p). \tag{17.12}$$

But it is simple to compute

$$\phi'(x) = \frac{1}{x(s-x)} \ge \frac{4}{s^2} \ge \frac{16}{9},$$

using the fact that $s = \frac{d+1}{d} \le \frac{3}{2}$. Hence (17.12) implies that $|p_v - p_v'| \le |m_v - m_v'|$, and therefore by (17.10) and (17.11) we get

$$|p_r - p_r'| \le |m_r - m_r'| \le (1-\gamma)^l \cdot c \log d \le \exp(-c'l),$$

as required. $\square$

All that remains to prove Theorem 17.6 is to prove Claim 17.10.

Proof of Claim 17.10. By the multivariate mean value theorem:

$$|F(\mathbf{m}) - F(\mathbf{m}')| \le \sup_{\mathbf{m}} \|\nabla F(m_1, \ldots, m_d)\|_1 \cdot \|\mathbf{m} - \mathbf{m}'\|_\infty$$

where $\nabla F = \Big(\frac{\partial F}{\partial m_i}\Big)_i$. So it is sufficient to prove that $\sup_{\mathbf{m}} \|\nabla F(\mathbf{m})\|_1 \le 1 - \gamma$, where

$$\|\nabla F(\mathbf{m})\|_1 = \sum_{i=1}^d \Big|\frac{\partial F}{\partial m_i}\Big|.$$

Recall that $F = \phi \circ f \circ \phi^{-1}$, where

$$\phi(x) = \frac{1}{s}\log\frac{x}{s-x}, \quad \phi^{-1}(y) = \frac{ye^{sy}}{e^{sy}+1}, \quad f(x_1, \ldots, x_d) = \frac{1}{1+\lambda\prod_i x_i}.$$

Now by direct calculation using the chain rule, we see [**exercise!**] that

$$\frac{\partial F}{\partial m_i} = \frac{1-p}{s-p}(s-p_i),$$

where $p_i = \phi^{-1}(m_i)$ and $p := 1/(1+\lambda\prod_i p_i)$.

Thus

$$\begin{aligned}
\|\nabla F(\mathbf{m})\|_1 &= \sum_{i=1}^{d} \frac{1-p}{s-p}(s-p_i) \\
&\le \frac{1-p}{s-p}\cdot d\cdot\Big[s-\Big(\prod_{i=1}^{d} p_i\Big)^{1/d}\Big] \\
&= \frac{1-p}{s-p}\cdot d\cdot\Big[s-\Big(\frac{1-p}{\lambda p}\Big)^{1/d}\Big]. \qquad (17.13)
\end{aligned}$$

The second line follows from the AM-GM inequality applied to the p_i

We want an upper bound for the expression in (17.13), which is provided by the following fact.

Proposition 17.12. *The function*

$$h(x) = \frac{(1-x)\Big[1+\frac{1}{d}-\Big(\frac{1-x}{\lambda x}\Big)^{1/d}\Big]}{1+\frac{1}{d}-x}$$

satisfies $\max_{x\in[0,1]} h(x) \le \frac{w}{1+w}$ *where* w *is the unique solution to* $w(1+w)^d = \lambda$.

Note that (17.13) is precisely $d\cdot h(p)$.

Proof. Write

$$h(x) = \frac{(1-x)(1+\frac{1}{d}-\Phi(x))}{1+\frac{1}{d}-x},$$

where $\Phi(x) := (\frac{1-x}{\lambda x})^{1/d}$. Now note that $\Phi'(x) = \frac{-\Phi(x)}{dx(1-x)}$, so Φ is monotonically decreasing on $[0,1]$, going from $+\infty$ to 0 in this interval. Thus it has a unique fixed point, which one can check [**exercise!**] is $x^* = \frac{1}{1+w}$ with w defined as above. Moreover, $\Phi(x) > x$ iff $x < x^*$. But now we can compute the derivative of h as

$$h'(x) = \frac{(1+\frac{1}{d})(\Phi(x)-x)}{dx(1+\frac{1}{d}-x)^2} \qquad \begin{cases} >0 & \text{for } x<x^*; \\ <0 & \text{for } x>x^*. \end{cases}$$

Hence $h(x)$ is maximized at x^*, and $h(x^*) = \frac{(1-x^*)(1+\frac{1}{d}-\Phi(x^*))}{1+\frac{1}{d}-x^*} = 1-x^* = \frac{w}{1+w}$, as required. □

Now, if we set $\lambda = \lambda_c(d) = \frac{d^d}{(d-1)^{d+1}}$, then it's easy to check [**exercise!**] that $w = \frac{1}{d-1}$, and hence $h(x) \le \frac{w}{1+w} = \frac{1}{d}$, so $d\cdot h(x)\le 1$.

But one can also readily check [**exercise!**] that $w = w(\lambda)$ is monotonically strictly increasing with λ, and of course $\frac{w}{1+w}$ is increasing with w. Hence for any $\lambda < \lambda_c(d)$, there exists $\gamma > 0$ such that

$$\sup_{\mathbf{m}} \|\nabla F(\mathbf{m})\|_1 \le d\sup_x h(x) \le 1-\gamma.$$

This completes the proof of Claim 17.10, and hence also of Theorem 17.6. □

17.6 Concluding remarks

1. An essentially equivalent version of the SAW tree construction (but not the correlation decay algorithm) appeared earlier in the context of matchings in the work of Godsil [God81].

2. Weitz's algorithm is deterministic, yielding an FPTAS for the partition function. It's interesting to note, however, that the SSM property that underlies the algorithm indirectly implies that MCMC also works for this problem, yielding an FPRAS when $\lambda < \lambda_c(\Delta)$, on all graphs of maximum degree Δ provided the graph is *amenable*, in the sense that the number of vertices in a ball of radius r around a vertex grows only polynomially with r. (This applies, e.g., to lattices $\mathbb{Z}^d$, but not to tree-like graphs.) This follows from a generic implication that, on amenable graphs, SSM implies $O(n \log n)$ mixing time of the Glauber dynamics [DSVW04]. Note that, for the hard-core model, the SSM property we have proved for trees applies also to general graphs of maximum degree Δ via the SAW tree construction. For more refined results about Glauber dynamics for the hard-core model, see [ALO20, CLV21].

3. For some special families of graphs, one may be able to do better than for arbitrary graphs of maximum degree Δ by exploiting additional structure of the SAW tree, notably the fact that its "average degree" (suitably defined) may be significantly less than Δ. An important example is the hard-core model on the Cartesian lattice $\mathbb{Z}^d$. Here one can show, e.g., that the Weitz algorithm works on (arbitrarily large square regions of) $\mathbb{Z}^2$ for $\lambda < 2.08$, which is much larger than $\lambda_c(4) = 1.69$, with analogous (though decreasing) improvements for larger dimensions d [SSŠY17]. It is notable that these lower bounds on the uniqueness threshold[3], which are currently the best known, have come from algorithmic investigations in Computer Science rather than from Physics arguments.

4. An analogous algorithmic result holds also for the *antiferromagnetic Ising model* [SST14] (and indeed for any antiferromagnetic spin system with two spin values; "antiferromagnetic" here means roughly that the potential function gives higher weight to neighboring spins that disagree than agree—this is true for the hard-core model due to the hard constraint forbidding adjacent +1 spins). Recall that the antiferromagnetic Ising model on graph $G = (V, E)$ has partition function

$$Z_G(\lambda, \mu) = \sum_{S \subseteq V} \lambda^{|E(S,\bar{S})|} \mu^{|S|},$$

   where $\lambda = \exp(2\beta) \geq 1$ (nearest-neighbor interaction) and $\mu = \exp(\beta h)$ (external field). The uniqueness diagram for the Ising model on the Δ-regular tree is sketched in Figure 17.4. (This picture can be verified via the same kind of recurrence analysis we used for the hard-core model.) Note in particular that, for $1 \leq \lambda \leq \lambda_c(\Delta)$, where $\lambda_c(\Delta) = \frac{\Delta}{\Delta - 2}$ is the threshold for the model with zero field ($\mu = 1$), we have a unique Gibbs measure for all external fields; and for $\lambda > \lambda_c(\Delta)$ there exists a critical field $\mu_c(\lambda, \Delta) > 0$ such that uniqueness occurs for $|\log \mu| \geq \log \mu_c(\lambda, \Delta)$. I.e., as the interaction parameter λ gets stronger, we need a larger field to break the separation of the system into $+$ and $-$ phases. Within the uniqueness region, by definition we get WSM on the tree, and the same arguments as those above yield an FPTAS. Specifically:

   - Weitz's SAW tree construction still works in exactly the same way, since it depends only on the fact that there are two spin values.
   - An analogous stepwise decay to that in the proof of Theorem 17.6 can be used to prove that WSM implies SSM, though the details are different: namely, the recurrence is different and one needs a different message function.

[3] Strictly speaking, we should insert the caveat that a single uniqueness threshold is not known to exist for the hard core model on $\mathbb{Z}^d$; this is because the property of uniqueness is not known to be monotone in λ (though it is conjectured to be so)—so it's conceivable that there are multiple thresholds.

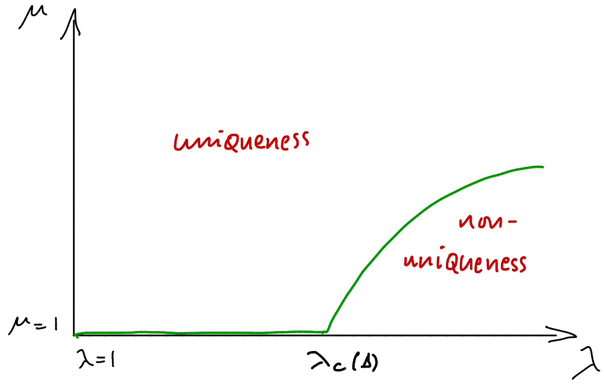

Figure 17.4: Phase diagram for the anti-ferromagnetic Ising model on the Δ-regular tree.

And as in the case of the hard-core model, there are complementary negative results showing that no FPRAS can exist outside the uniqueness region unless NP=RP [SS12, GGŠ+14]. (We'll be seeing a proof of this for the hard-core model in the next lecture.)

5. The situation for *ferromagnetic* 2-spin systems is much murkier than the clean picture we have seen for the antiferromagnetic case. First, as we saw in an earlier lecture, there is an FPRAS (based on MCMC) for *any* graph, at *all* values of the parameters [JS93]. Hence we can't expect any negative complexity theoretic results in this case. On the other hand, it is still an open problem to find a *deterministic* approximation algorithm (based on correlation decay or any other technique) throughout this range. There is still a uniqueness threshold on the Δ-regular tree here (exactly analogous to the antiferromagnetic case—indeed the same phase diagram as in Figure 17.4 holds here as the tree is bipartite), so we might expect Weitz's algorithm to work at least in this regime. But even this is not known, since while the SAW tree construction works (as it does for any 2-spin system), as noted earlier the implication that WSM implies SSM no longer holds in the ferromagnetic case. On the other hand, if we don't restrict to graphs of bounded degree, then correlation decay algorithms are known essentially up to the uniqueness threshold for all ferromagnetic 2-spin systems [LLY13, GL18]. And this positive result is complemented by a #BIS-hardness result for these models on the other side of this threshold (excluding the special case of the ferromagnetic Ising model, which as we've seen is tractable everywhere) [LLZ14].

6. It is an open problem to extend Weitz's SAW tree framework to spin systems with more than two spins (such as colorings or the Potts model). Some attempts in this direction (which fall quite far short of what can be done with other methods, such as MCMC) can be found in [GK12, GKM15]. The second part (Strong Spatial Mixing) has recently been addressed in [CLMM23], but the self-avoiding-walk tree construction in the form presented here doesn't work with more than two spins anywhere close to the uniqueness threshold.

References

[ALO20] N. Anari, K. Liu, and S. Oveis Gharan. Spectral independence in high-dimensional expanders and applications to the hardcore model. *Proceedings of the 61st Annual IEEE Symposium on Foundations of Computer Science (FOCS)*, 2020.

[CLMM23] Z. Chen, K. Liu, N. Mani, and A. Moitra. Strong spatial mixing for colorings on trees and its algorithmic applications. ArXiv preprint 2304.01954. 2023.

[CLV21] Z. Chen, K. Liu, and E. Vigoda. Optimal mixing of Glauber dynamics: Entropy factorization via high-dimensional expansion. *Proceedings of the 53rd Annual ACM Symposium on Theory of Computing (STOC)*, 2021.

[DSVW04] M. Dyer, A. Sinclair, E. Vigoda, and D. Weitz. Mixing in time and space for lattice spin systems. *Random Structures & Algorithms*, 24:461–479, 2004.

[GGŠ+14] A. Galanis, Q. Ge, D. Štefankovič, E. Vigoda, and Y. Yin. Improved inapproximability results for counting independent sets in the hard-core model. *Random Structures & Algorithms*, 45:78–110, 2014.

[GK12] D. Gamarnik and D. Katz. Correlation decay and deterministic fptas for counting colorings of a graph. *Journal of Discrete Algorithms*, pages 29–47, 2012.

[GKM15] D. Gamarnik, D. Katz, and S. Misra. Strong spatial mixing of list coloring of graphs. *Random Structures & Algorithms*, pages 599–613, 2015.

[GL18] H. Guo and P. Lu. Uniqueness, spatial mixing and approximation for ferromagnetic 2-spin systems. *ACM Transactions on Computing Theory*, 10:17, 2018.

[God81] C.D. Godsil. Matchings and walks in graphs. *Journal of Graph Theory*, 5:285–297, 1981.

[JS93] M. Jerrum and A. Sinclair. Polynomial-time approximation algorithms for the Ising model. *SIAM Journal on Computing*, 22:1087–1116, 1993.

[Kel85] F.P. Kelly. Stochastic models of computer communication systems. *Journal of the Royal Statistical Society B*, 47:379–395, 1985.

[LLY13] L. Li, P. Lu, and Y. Yin. Correlation decay up to uniqueness in spin systems. *Proceedings of the 24th Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pages 67–84, 2013.

[LLZ14] J. Liu, P. Lu, and C. Zhang. The complexity of ferromagnetic two-spin systems with external fields. *Proceedings of RANDOM*, pages 843–856, 2014.

[RST+13] R. Restrepo, J. Shin, P. Tetali, E. Vigoda, and L. Yang. Improved mixing condition on the grid for counting and sampling independent sets. *Probability Theory and Related Fields*, 156:75–99, 2013.

[Sly10] A. Sly. Computational transition at the uniqueness threshold. *Proceedings of the 51st Annual IEEE Symposium on Foundations of Computer Science (FOCS)*, pages 287–296, 2010.

[Spi75] F. Spitzer. Markov random fields on an infinite tree. *Annals of Probability*, 3:387–398, 1975.

[SS12] A. Sly and N. Sun. The computational hardness of counting in two-spin models on d-regular graphs. *Prodeedings of the 53rd Annual IEEE Symposium on Foundations of Computer Science (FOCS)*, pages 361–369, 2012.

[SSŠY17] A. Sinclair, P. Srivastava, D. Štefankovič, and Y. Yin. Spatial mixing and the connective constant: Optimal bounds. *Probability Theory and Related Fields*, 168:153–197, 2017.

[SST14] A. Sinclair, P. Srivastava, and M. Thurley. Approximation algorithms for two-state anti-ferromagnetic spin systems on bounded degree graphs. *Journal of Statistical Physics*, 155:666–686, 2014.

[Wei06] D. Weitz. Counting independent sets up to the tree threshold. *Proceedings of the 38th Annual ACM Symposium on Theory of Computing (STOC)*, pages 140–149, 2006.