6.S891 Lecture 7: Computation Trees and the Correlation Decay Algorithm

Kuikui Liu

September 28, 2023

In this lecture, we demonstrate the correlation decay algorithm for the hardcore model on an arbitrary bounded-degree graph within the uniqueness regime. For reader's convenience, we have reproduced the main theorem we aim to prove.

Theorem 0.1 (Weitz [Wei06])**.** *If $\lambda < \lambda_c(\Delta)$, then* SSM *holds for the hardcore model on any graph of maximum degree Δ. Furthermore, there exists an* FPTAS *for estimating $Z_G(\lambda)$ for every graph $G = (V, E)$ of maximum degree Δ and every $\lambda < \lambda_c(\Delta)$. If $\lambda \leq (1 - \delta)\lambda_c(\Delta)$ for a constant $0 < \delta < 1$, the running time of this algorithm scales as $(n/\epsilon)^{O\left(\frac{1}{\delta}\log\Delta\right)}$, where $0 < \epsilon < 1$ is the estimation error.*

1 Contraction of the Tree Recursion via the Potential Method

Recall the (multivariate) tree recursion for the hardcore model is given by

$$F_d(p_1, \dots, p_d) = \frac{\lambda \prod_{i=1}^d (1 - p_i)}{1 + \lambda \prod_{i=1}^d (1 - p_i)}, \qquad \forall p_1, \dots, p_d \in [0, 1].$$

Furthermore, if $\varphi : [0, 1] \to \mathbb{R}_{\geq 0} \cup \{+\infty\}$ is a smooth strictly increasing potential function, then we define the modified tree recursion for the change of variables $m = \varphi(p)$ by

$$G_d(m_1, \dots, m_d) = \varphi(F_d(\varphi^{-1}(m_1), \dots, \varphi^{-1}(m_d))), \qquad \forall m_1, \dots, m_d \in \mathbb{R}_{\geq 0} \cup \{+\infty\}.$$

Recall that we further write f_d, g_d for their univariate analogs (i.e. when $p_1 = \cdots = p_d = p$ and $m_1 = \cdots = m_d = m$).

In the previous lecture, we saw that SSM for the hardcore Gibbs measure on arbitrary trees of maximum degree Δ holds if we can construct φ satisfying the contraction property $\|\nabla G_d(\boldsymbol{m})\|_1 \leq 1 - O(\delta)$ (plus very mild boundedness assumptions). We now exhibit a good choice of φ due to Li–Lu–Yin [LLY13], which will complete the proof of SSM on trees of maximum degree Δ whenever $\lambda < \lambda_c(\Delta)$.

Define φ such that $\Phi(p) \stackrel{\mathsf{def}}{=} \varphi'(p) = \frac{1}{\sqrt{p}\cdot(1-p)}$.[1] Note φ satisfies all of the required smoothness, monotonicity, and boundedness properties.[2] The main result is the following.

Proposition 1.1 (Contraction for Hardcore Tree Recursion; [LLY13])**.** *Let $\varphi : [0, 1] \to \mathbb{R}_{\geq 0} \cup \{+\infty\}$ be defined implicitly via its derivative $\Phi(p) = \varphi'(p) = \frac{1}{\sqrt{p}(1-p)}$. If $\lambda \leq (1 - \delta) \cdot \lambda_c(\Delta)$, then the modified multivariate tree recursion $G_d = \varphi \circ F_d \circ \varphi^{-1}$ satisfies $\|\nabla G_d(\boldsymbol{m})\|_1 \leq 1 - O(\delta)$ for all $\boldsymbol{m} \in \mathbb{R}^d_{\geq 0}$ and all $1 \leq d \leq \Delta - 1$.*

This result completes the proof that SSM holds on trees of maximum degree Δ whenever $\lambda < \lambda_c(\Delta)$. We establish the desired contractive property in two steps. The first lemma below reduces the multivariate case to the univariate case, which is simpler to analyze. The second lemma is a consequence of our analysis of f_d and its fixed points.

[1] Integrating this, one gets $\varphi(p) = 2\,\mathrm{arctanh}\left(\sqrt{p}\right)$. However, we will not need the explicit formula for φ itself. Its derivative Φ is what matters.

[2] As we mentioned in the previous lecture, we only need the boundedness in the image of a constant number of iterations of F_d, e.g. the interval $\left[\frac{\lambda}{\lambda+(1+\lambda)^d}, \frac{\lambda}{1+\lambda}\right]$ for the hardcore model.

Lemma 1.2. *For every $\boldsymbol{m} \in \mathbb{R}^d_{\geq 0}$, there exists $\overline{m} \in \mathbb{R}_{\geq 0}$ such that*

$$\|\nabla G_d(\boldsymbol{m})\|_1 \leq \|\nabla G_d(\overline{m} \cdot \mathbf{1})\|_1 = |g_d'(\overline{m})|.$$

Lemma 1.3. *Suppose $\lambda \leq (1-\delta) \cdot \lambda_c(\Delta)$. Then for every $m \in \mathbb{R}_{\geq 0}$ and every $1 \leq d \leq \Delta - 1$, we have that $|g_d'(m)| \leq 1 - O(\delta)$.*

These two lemmas combined immediately imply Proposition 1.1. Note that their proofs are completely independent of each other. We prove each in turn.

Proof of Lemma 1.2. By the Chain Rule and the Inverse Function Theorem, we have via calculation that

$$\partial_{m_i} G_d(\boldsymbol{m}) = \frac{\Phi(F_d(\boldsymbol{p}))}{\Phi(p_i)} \cdot \partial_{p_i} F_d(\boldsymbol{p}) = -\sqrt{F_d(\boldsymbol{p})} \cdot \sqrt{p_i},$$

where $m_i = \varphi(p_i)$. As a special case, this also says that $g_d'(m) = -d \cdot \sqrt{f_d(p)} \cdot \sqrt{p}$ where $m = \varphi(p)$. It follows that

$$\begin{aligned}
\|\nabla G_d(\boldsymbol{m})\|_1^2 &= F_d(\boldsymbol{p}) \cdot \left(\sum_{i=1}^d \sqrt{p_i}\right)^2 \\
&\leq d \cdot F_d(\boldsymbol{p}) \cdot \sum_{i=1}^d p_i && \text{(Cauchy–Schwarz)} \\
&= d^2 \cdot F_d(\boldsymbol{p}) \cdot \left(1 - \frac{1}{d}\sum_{i=1}^d (1-p_i)\right) \\
&\leq d^2 \cdot F_d(\boldsymbol{p}) \cdot \left(1 - \prod_{i=1}^d (1-p_i)^{1/d}\right) && \text{(AM-GM Inequality)} \\
&= d^2 \cdot f_d(\overline{p}) \cdot \overline{p} && \text{(Setting } \overline{p} = 1 - \textstyle\prod_{i=1}^d (1-p_i)^{1/d}) \\
&= |g_d'(\overline{m})|^2.
\end{aligned}$$

Taking square roots gives the claim. □

Proof of Lemma 1.3. As we saw in the previous lecture, if $\lambda \leq (1-\delta)\cdot\lambda_c(\Delta)$, then $|f_d'(\hat{p})| \leq 1-O(\delta)$, where $\hat{p} = \hat{p}(\lambda, d)$ is the unique fixed point of f_d. Hence, the crux of the proof will be showing that $|g_d'(m)| \leq \sqrt{|f_d'(\hat{p})|}$ for all $m \in \mathbb{R}_{\geq 0}$. Since $g_d'(m) = -d \cdot \sqrt{f_d(p)} \cdot \sqrt{p}$, where $m = \varphi(p)$ (or $p = \varphi^{-1}(m)$), and $f_d'(p) = -d \cdot \frac{1-f_d(p)}{1-p} \cdot f_d(p)$, the claim is equivalent to showing that for all $p \in [0,1]$, we have $d \cdot f_d(p) \cdot p \leq \hat{p}$.

Let $p^* = p^*(\lambda, d) \in [0,1]$ maximize the function $p \mapsto f_d(p) \cdot p$, whose derivative is given by

$$f_d'(p) \cdot p + f_d(p) = f_d(p) \cdot \left(1 - d \cdot \frac{1-f_d(p)}{1-p} \cdot p\right).$$

Since $p \mapsto f_d(p) \cdot p$ is zero at $p = 0$ and $p = 1$, it must be that $p^* \in (0,1)$ is a stationary point, whence p^* must satisfy

$$\frac{1-p^*}{1-f_d(p^*)} \cdot \frac{1}{p^*} = d.$$

It follows that

$$\sup_{0 \leq p \leq 1} \{d \cdot f_d(p) \cdot p\} = \frac{1-p^*}{1-f_d(p^*)} \cdot f_d(p^*).$$

Note that the function $p \mapsto \frac{1-p}{1-f_d(p)} \cdot f_d(p)$ evaluates to $\hat{p}$ on input $\hat{p}$. Furthermore, both this function as well as the function $p \mapsto \frac{1-p}{1-f_d(p)} \cdot \frac{1}{p}$ are both monotone decreasing in p. Hence, the desired inequality is equivalent to $\frac{1-p^*}{1-f_d(p^*)} \cdot \frac{1}{p^*} \leq \frac{1-\hat{p}}{1-f_d(\hat{p})} \cdot \frac{1}{\hat{p}}$. By definition of p^*, the left-hand side evaluates to d; similarly, by definition of $\hat{p}$, the right-hand side evaluates to $1/\hat{p}$. Hence, the desired inequality is equivalent to $\hat{p} \leq 1/d$, which holds since $\lambda \leq \lambda_c(\Delta)$, as we saw in the previous lecture. □

1.1 Why This Choice of φ?

At the moment, there aren't systematic ways of discovering/designing good potential functions. [LLY13] gave some heuristics for discovering them in the special case of two-spin systems. With the benefit of hindsight, we can provide some intuitive justification for $\Phi(p) = \varphi'(p) = \frac{1}{\sqrt{p(1-p)}}$. Recall our goal is to select φ such that

$$\left\|\nabla G_d(\boldsymbol{m})\right\|_1 \leq 1 - O(\delta), \qquad \forall \boldsymbol{m} \in \mathbb{R}^d_{\geq 0}.$$

Let's see what happens if we didn't use any potential function. Then

$$\left\|\nabla F_d(\boldsymbol{p})\right\|_1 = \sum_{i=1}^{d} \frac{1 - F_d(\boldsymbol{p})}{1 - p_i} \cdot F_d(\boldsymbol{p}).$$

By monotonicity of $F_d(\boldsymbol{p})$ and the fact that $F_d(\boldsymbol{p}) \leq \frac{\lambda}{1+\lambda} \leq O(1/\Delta)$ is always upper bounded by some (small) constant, we have the right-hand side is maximized at $\boldsymbol{p} = \boldsymbol{0}$, yielding $\frac{d\lambda}{(1+\lambda)^2}$. For this to be a contraction, we need $\lambda \leq 1/(d-2)$, which is very far from the true $\lambda_c(d+1)$.

One reason this fails is because we are using a truly worst-case bound of $F_d(\boldsymbol{p}) \leq \frac{\lambda}{1+\lambda}$ everywhere. On the other hand, when we used $\Phi(p) = \frac{1}{\sqrt{p(1-p)}}$, we saw that we got

$$\left\|\nabla G_d(\boldsymbol{m})\right\|_1 = \sum_{i=1}^{d} \sqrt{F_d(\boldsymbol{p})} \cdot \sqrt{p_i}.$$

This is a much more well-behaved quantity because the hardcore model is *antiferromagnetic*. If any p_i is too large, then $F_d(\boldsymbol{p})$ will be small, and vice versa. Thus the terms $\sqrt{F_d(\boldsymbol{p})}$ and $\sqrt{p_i}$ counterbalance each other. A concrete example here is to consider the star on d vertices, where $F_d(\boldsymbol{p})$ gives the marginal of the center vertex, and the $p_1, \ldots, p_d$ gives the marginals of the leaves as isolated vertices.

Historically, it was well-known that one can take advantage of this repulsive nature of the hardcore model by studying two iterations of F_d simultaneously. For instance, this is what we did in the previous lecture in the univariate setting, where we looked at $f_d^{\circ 2}$ and its analytic properties. However, since F_d has d variables, studying two levels at a time yields a recursion with d^2 variables, which becomes unwieldy extremely quickly. At the expense of requiring some creativity in designing φ, the potential method is an extremely useful tool for simplifying these calculations.

2 Algorithms and SSM on General Graphs

So far, we have shown that if $\lambda < \lambda_c(\Delta)$, the strong spatial mixing holds for the hardcore Gibbs measure on arbitrary trees of maximum degree Δ. We now extend this to all graphs of maximum degree Δ, thus establishing that trees, and in particular the infinite d-regular trees, are the worst case. This reduction from general graphs to trees is the main innovation of [Wei06]. Our exposition will differ slightly from that of Weitz's paper, and perhaps is more similar to [LY13]; see Section 3 for further discussion. We first prove strong spatial mixing on general graphs of maximum degree Δ when $\lambda < \lambda_c(\Delta)$. We will then see that the technique also furnishes an efficient approximate counting algorithm.

Right off the bat, the main obstacle we seem to face when analyzing general graphs is we lose our ability to use the tree recursion. This is because once we delete the vertex r in consideration, its neighbors can be remain connected through other paths, and so we don't get any kind of independence or factorization. However, it turns out we can still use the tree recursion F_d to exactly compute the marginals, at the expense of creating an exponentially large tree called the *computation tree*. The nodes of the computation tree will represent "(sub)instances" of the problem of computing the marginal of a vertex. Later on, when we design an efficient algorithm, we will use correlation decay to say that we can truncate this exponentially large computation tree to polynomial size. For now, we focus on (inefficient) exact computation.

Definition 1 (Instance)**.** *An* instance *is a pair (G, r), where $G = (V, E)$ is a graph, and $r \in V$ is a distinguished vertex. For such an instance, let $p_{G,r}$ denote the marginal probability* $\Pr_{I \sim \mu_{G,\lambda}}[r \in I]$ *that r is contained in a random independent set.*

Theorem 2.1 ([Wei06]; see also [LY13]). *Let $G = (V, E)$ be an arbitrary graph, and let $r \in V$ be an arbitrary vertex with neighbors $u_1, \dots, u_d$ ordered arbitrarily. Then*

$$p_{G,r} = F_d(p_{G_1,u_1}, \dots, p_{G_d,u_d}),$$

where for all $k = 1, \dots, d$, the graph G_k is obtained from G by deleting r and $u_1, \dots, u_{k-1}$.

Before we prove this result, let us describe how we will use it. First, observe that Theorem 2.1 immediately yields an (inefficient) recursive algorithm for computing the marginal probability $p_{G,r}$ exactly: When we wish to compute $p_{G,r}$ for an instance (G, r), we recursively call the algorithm on each of the d subinstances $(G_1, u_1), \dots, (G_d, u_d)$ to compute $p_{G_1,u_1}, \dots, p_{G_d,u_d}$, and then apply the multivariate function F_d to compute $p_{G,r}$. The recursion *must* terminate in a finite number of iterations. This is simply because each recursive call is on a graph with strictly fewer vertices, and so we must eventually end up with a graph consisting only of isolated vertices where everything is trivial to compute.

Pictorially, one should imagine that this recursive algorithm traces out a *computation tree* $T_{\mathsf{SAW}}(G, r)$, where each node is labeled by an instance (H, u) for some subgraph H of G and some vertex $u \in H$, and the root is labeled by (G, r). Despite being exponential size in general, this tree has a number of nice properties:

- Since each subinstance (H, u) in $T_{\mathsf{SAW}}(G, r)$ has a distinguished vertex $u \in V$, there is a natural way to identify a subset of vertices $\Lambda \subseteq V$ with a subset of subinstances/vertices Λ_{SAW} in $T_{\mathsf{SAW}}(G, r)$ consisting of all subinstances (H, u) such that $u \in \Lambda$. As a consequence, there is a natural way to "lift" a pinning $\tau : \Lambda \to \{\mathsf{in}, \mathsf{out}\}$ in G to a pinning $\tau_{\mathsf{SAW}} : \Lambda_{\mathsf{SAW}} \to \{\mathsf{in}, \mathsf{out}\}$ in $T_{\mathsf{SAW}}(G, r)$.
- $T_{\mathsf{SAW}}(G, r)$ viewed as a graph is a nice object which has been previously studied extensively in algebraic combinatorics [God93] and statistical physics [SS05]. In particular, the walks in this tree started from the root are in one-to-one correspondence with the *self-avoiding walks* in G, and so $T_{\mathsf{SAW}}(G, r)$ is also called the *self-avoiding walk tree* of G rooted at r (hence the subscript SAW).
- $T_{\mathsf{SAW}}(G, r)$ also has maximum degree Δ by virtue of our identification between subinstances in $T_{\mathsf{SAW}}(G, r)$ and vertices in G. For the same reason, if $\Lambda \subseteq V \setminus \{r\}$, then the shortest path distance of r from Λ in G is equal to the shortest path distance of r from Λ_{SAW} in $T_{\mathsf{SAW}}(G, r)$. In other words, degrees and distances (to r) are preserved.

Combining this gadget, its properties, and the robust contraction property we established in Proposition 1.1, we will be able to establish the desired SSM. As the name computation tree suggests, we will also be able to turn $T_{\mathsf{SAW}}(G, r)$ into an efficient approximate counter once we have decay of correlations. Having all of this motivation in mind, we now prove Theorem 2.1. We give a simplified proof, due to Daniel Lee, which is specialized to the hardcore model. A more generalizable (but notationally more cumbersome) proof is provided in Appendix A.

Proof of Theorem 2.1. Observe that

$$\begin{aligned}
1 &= \Pr_{I\sim\mu_{G,\lambda}}[r \in I] + \Pr_{I\sim\mu_{G,\lambda}}[r \notin I] \\
&= \Pr_{I\sim\mu_{G,\lambda}}[r \in I, u_1, \dots, u_d \notin I] + \Pr_{I\sim\mu_{G,\lambda}}[r \notin I] \\
&= \lambda \cdot \Pr_{I\sim\mu_{G,\lambda}}[r \notin I, u_1, \dots, u_d \notin I] + \Pr_{I\sim\mu_{G,\lambda}}[r \notin I] && \text{(Special to Hardcore Model)} \\
&= \left(\lambda \cdot \Pr_{I\sim\mu_{G-r,\lambda}}[u_1, \dots, u_d \notin I] + 1\right) \cdot \Pr_{I\sim\mu_{G,\lambda}}[r \notin I].
\end{aligned}$$

Rearranging yields

$$p_{G,r} = \frac{\lambda \Pr_{G-r}[u_1, \dots, u_d \leftarrow \mathsf{out}]}{1 + \lambda \Pr_{G-r}[u_1, \dots, u_d \leftarrow \mathsf{out}]}.$$

Writing everything using conditional probabilities and again using the fact that pinning a vertex $u \leftarrow \mathsf{out}$ is equivalent to deleting u from the graph, we obtain

$$\Pr_{G-r}[u_1, \dots, u_d \leftarrow \mathsf{out}] = \prod_{k=1}^{d} \Pr_{G-r}[u_k \leftarrow \mathsf{out} \mid u_1, \dots, u_{k-1} \leftarrow \mathsf{out}]$$
$$= \prod_{k=1}^{d} (1 - p_{G_k, u_k}),$$

from which the claim immediately follows. □

With this massive hammer, we now prove the main theorem of Weitz.

Proof of Theorem 0.1. Let $G = (V, E)$ be an arbitrary graph of maximum degree Δ, and suppose $\lambda \le (1-\delta) \cdot \lambda_c(\Delta)$. We first prove strong spatial mixing. Let $r \in V$, $\Lambda \subseteq V \setminus \{r\}$ and $\tau, \sigma : \Lambda \to \{\mathsf{in}, \mathsf{out}\}$ be arbitrary. Since $\|\mu_r^\tau - \mu_r^\sigma\|_{\mathsf{TV}} = \left|p_{G,r}^\tau - p_{G,r}^\sigma\right|$, we wish to show that these two conditional marginals are within $\lesssim (1-\delta)^{\mathrm{dist}(r, \Lambda_{\tau,\sigma})}$ additive error of each other. Consider the associated computation tree $T = T_{\mathsf{SAW}}(G, r)$, which by construction correctly computes these conditional marginals using the "lifted" pinnings $\tau_{\mathsf{SAW}}, \sigma_{\mathsf{SAW}} : \Lambda_{\mathsf{SAW}} \to \{\mathsf{in}, \mathsf{out}\}$. By Theorem 2.1, this computation is obtained by inductively applying F_d, which by Proposition 1.1, is a contraction globally w.r.t. the metric $\psi(p, q) = |\varphi(p) - \varphi(q)|$. From here, SSM follows by the same argument employed in the preceding lecture to show that contraction implies SSM on trees.

We now describe an FPTAS. If we order the vertices of G arbitrarily as $v_1, \dots, v_n$, then defining $G_0 = G$ and $G_i = G_{i-1} - v_i$, we have $Z_{G_n}(\lambda) = 1$, and so

$$Z_G(\lambda) = \prod_{i=1}^{n} \frac{Z_{G_{i-1}}(\lambda)}{Z_{G_i}(\lambda)} = \prod_{i=1}^{n} \frac{1}{1 - p_{G_{i-1}, v_i}}.$$

Hence, to obtain an $(1 \pm \epsilon)$-multiplicative approximation of $Z_G(\lambda)$, it suffices to obtain $\left(1 \pm \frac{\epsilon}{n}\right)$-multiplicative approximations of each of the marginals $1 - p_{G_{i-1}, v_i}$. Since $p_{G_{i-1}, v_i} \le \frac{\lambda}{1+\lambda}$,[3] it suffices to obtain $\pm \frac{\epsilon}{(1+\lambda)n} = \pm O(\epsilon/n)$ additive approximation to each p_{G_{i-1}, v_i}.

Fix any such instance (G, r). We unroll the recursive computation described by $T_{\mathsf{SAW}}(G, r)$ to depth $L = O\left(\frac{1}{\delta} \log(n/\epsilon)\right)$. Once we hit any subinstance (H, u) at this depth, instead of recursively calling the algorithm to compute $p_{H,u}$ and passing up an exact value for this marginal, we pass up an *arbitrarily* chosen number $\tilde{p}_{H,u} \in [0, 1]$. Applying the tree recursion F_d as in Theorem 2.1, this yields estimates $\tilde{p}_{H,u}$ for all subinstances (H, u) which can be reached from (G, r) in $\le L$ steps in $T_{\mathsf{SAW}}(G, r)$. Let (H, u) be any such instance, and let $(H_1, v_1), \dots, (H_d, v_d)$ be its child instances. Then

$$|\varphi(\tilde{p}_{H,u}) - \varphi(p_{H,u})| = |G_d(\varphi(\tilde{p}_{H_1,v_1}), \dots, \varphi(\tilde{p}_{H_d,v_d}) - G_d(\varphi(p_{H_1,v_1}), \dots, \varphi(p_{H_d,v_d})|$$
$$\le (1 - O(\delta)) \cdot \max_{1 \le i \le d} |\varphi(\tilde{p}_{H_i,v_i}) - \varphi(p_{H_i,v_i})|$$

by the same combination of the Mean Value Theorem and Proposition 1.1. Applying this bound inductively and converting back to total variation in a manner similar to the previous lecture, see that any error incurred at a depth L instance is reduced by a multiplicative factor of $(1 - O(\delta))^L$. Since the error at depth L is bounded, taking the constant in front of $L \le O\left(\frac{1}{\delta} \log(n/\epsilon)\right)$, we obtain our additive approximation $|\tilde{p}_{G,r} - p_{G,r}| \le O(\epsilon/n)$ as desired. □

3 Some Concluding Remarks

Going Beyond the Hardcore Model There is a generalization of Theorem 2.1 to arbitrary q-spin systems, which can be obtained by making straightforward modifications to the proof given in Appendix A. This was first done in [GK12] for colorings, and generalized in [LY13]. However, the resulting computation tree $T_{\mathsf{comp}}(G, r, \mathfrak{c})$ it traces out branches at a factor of dq rather than d, because each subinstance now also depends on the specific color $\mathfrak{c} \in [q]$ whose marginal $p_{G,r}(\mathfrak{c})$

[3] We never proved this, but it isn't hard to see, since the probability that a vertex is in the random independent set is never more than the same probability but w.r.t. a graph where we have deleted all other vertices.

you're trying to compute. This blown up branching number is rather inconvenient, and in many settings, obstructs the reduction from general graphs to trees (at least when the parameters of the q-spin system are close to the boundary of the correlation decay regime).

In our proof above, we used specific properties of the hardcore model to reduce the branching from $2d$ (since $q = 2$) to just d. It turns out, a neat extra trick of Weitz [Wei06] allows one to reduce the branching from $2d$ to d for general 2-spin systems, without appealing to special properties of the hardcore model. This is achieved by replacing computation of the marginals directly with computation of the marginal *ratios* $R_{G,r} = \frac{p_{G,r}}{1-p_{G,r}}$. One can write down a tree recursion for these new variables,[4] and modify the proof of Theorem 2.1 to obtain a similar computation tree, which is the original one considered by Weitz.

Our discussion here naturally raises the following question.

Question 1. *For every q-spin system $\boldsymbol{A}$, does there exist a "natural" family of statistics $S : [0,1]^q \to \mathbb{R}^q_{\geq 0}$ for which we can write down a computation tree for computing $S(\boldsymbol{p}_{G,r})$ which branches at a rate of d instead of dq?*

Again, by "branching" we mean the number of recursively produced subinstances. Interestingly, even though the branching in $T_{\mathsf{comp}}(G, r, \mathfrak{c})$ for computing marginals is dq, each subinstance only passes up the marginal of a specific color, rather than its entire marginal distribution. So in a sense, the total "amount of information" passed up is still the same.

On Connections Between WSM, SSM and Algorithms Weitz's argument shows that weak spatial mixing (or equivalently, $\lambda < \lambda_c(\Delta)$) on the infinite Δ-regular tree is equivalent to strong spatial mixing on all graphs of maximum degree Δ. This equivalence, however, does not hold in general. Indeed, it was already observed by Weitz [Wei06] that for the ferromagnetic Ising model with appropriate (inverse) temperature and (consistent) external fields, weak spatial mixing does not necessarily imply strong spatial mixing, even though the converse obviously holds; a more detailed description of the counterexample is provided in the appendix of [SST14]. Furthermore, it was shown in [LLY13] that for the distribution

$$\mu_{\lambda,\gamma}(I) \propto \lambda^{|S|} \cdot \gamma^{\#\{uv \in E : u,v \notin I\}},$$

which corresponds to the spin system with interaction matrix $A = \begin{bmatrix} 0 & 1 \\ 1 & \gamma \end{bmatrix}$, one can choose λ appropriately such that strong spatial mixing holds for the infinite Δ-regular tree, but fails for the infinite d-regular tree for some $d < \Delta$. This was one motivation for [LLY13] to define "up-to-Δ" uniqueness (as opposed to simply "Δ-uniqueness"), meaning that the derivative of the univariate tree recursion with d children, evaluated at the fixed point, is less than 1 uniformly for all $1 \leq d \leq \Delta - 1$. Allan Sly also constructed an example of a multi-spin system such that the corresponding Gibbs measure on the infinite d-regular tree is unique, but uniqueness fails for some other infinite d-regular graphs [Sly08].

Sharper Analyses of $T_{\mathsf{SAW}}(G, r)$ Note that while every pinning τ on $\Lambda \subseteq V \setminus \{r\}$ can be "lifted" to a pinning τ_{SAW} on Λ_{SAW} in $T = T_{\mathsf{SAW}}(G, r)$, there do exist pinnings in $T_{\mathsf{SAW}}(G, r)$ which cannot be realized as such lifts. Hence, for more specialized classes of graphs G, e.g. tori, strong spatial mixing on $T_{\mathsf{SAW}}(G, r)$ is strictly stronger than strong spatial mixing on G itself. In other words, the region of parameters in which correlation decay holds for G can strictly contain the corresponding region for $T_{\mathsf{SAW}}(G, r)$. We refer interested readers to [Bla+19] and references therein for further discussion of this issue of unfeasible pinnings in $T_{\mathsf{SAW}}(G, r)$, which is in general a very challenging issue to overcome. See also [Res+13] for an analysis which leverages additional combinatorial properties of $T_{\mathsf{SAW}}(G, r)$ when $G = \mathbb{Z}^2$ to improve the known SSM threshold for the hardcore model. We mention the following open problem.

Question 2. *What is the correct threshold for weak/strong spatial mixing for the hardcore model on $\mathbb{Z}^2$? How about other lattices?*

[4] For the hardcore model, the multivariate tree recursion for the ratios is given by $F_d(\boldsymbol{R}) = \lambda \prod_{i=1}^{d} \frac{1}{1+R_i}$.

References

[Bla+19] Antonio Blanca, Yuxuan Chen, David Galvin, Dana Randall, and Prasad Tetali. "Phase Coexistence for the Hard-Core Model on $\mathbb{Z}^2$". In: *Combinatorics, Probability & Computing* 28 (2019), pp. 1–22 (cit. on p. 6).

[GK12] David Gamarnik and Dmitriy Katz. "Correlation decay and deterministic FPTAS for counting colorings of a graph". In: *Journal of Discrete Algorithms* 12 (2012), pp. 29–47 (cit. on pp. 5, 7).

[God93] Christopher David Godsil. *Algebraic Combinatorics*. 1st ed. Chapman & Hall, Inc., 1993 (cit. on p. 4).

[LLY13] Liang Li, Pinyan Lu, and Yitong Yin. "Correlation Decay Up to Uniqueness in Spin Systems". In: *Proceedings of the Twenty-fourth Annual ACM-SIAM Symposium on Discrete Algorithms*. SODA '13. New Orleans, Louisiana: Society for Industrial and Applied Mathematics, 2013, pp. 67–84. ISBN: 978-1-611972-51-1 (cit. on pp. 1, 3, 6).

[LY13] Pinyan Lu and Yitong Yin. "Improved FPTAS for Multi-spin Systems". In: *Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques*. Ed. by Prasad Raghavendra, Sofya Raskhodnikova, Klaus Jansen, and José D. P. Rolim. Berlin, Heidelberg: Springer Berlin Heidelberg, 2013, pp. 639–654 (cit. on pp. 3–5, 7).

[Res+13] Ricardo Restrepo, Jinwoo Shin, Prasad Tetali, Eric Vigoda, and Linji Yang. "Improved mixing condition on the grid for counting and sampling independent sets". In: *Probability Theory and Related Fields* 156.1 (June 2013), pp. 75–99 (cit. on p. 6).

[Sly08] Allan Sly. "Uniqueness thresholds on trees versus graphs". In: *The Annals of Applied Probability* 18.5 (2008) (cit. on p. 6).

[SS05] Alexander D. Scott and Alan D. Sokal. "The Repulsive Lattice Gas, the Independent-Set Polynomial, and the Lovász Local Lemma". In: *Journal of Statistical Physics* 118.5 (2005), pp. 1151–1261 (cit. on p. 4).

[SST14] Alistair Sinclair, Piyush Srivastava, and Marc Thurley. "Approximation Algorithms for Two-State Anti-Ferromagnetic Spin Systems on Bounded Degree Graphs". In: *Journal of Statistical Physics* 155.4 (2014), pp. 666–686 (cit. on p. 6).

[Wei06] Dror Weitz. "Counting Independent Sets Up to the Tree Threshold". In: *Proceedings of the 38th Annual ACM Symposium on Theory of Computing (STOC)*. Seattle, WA, USA, 2006, pp. 140–149 (cit. on pp. 1, 3, 4, 6).

A A More Generalizable Proof of Theorem 2.1

We give a more generic proof which works for an arbitrary q-spin system. This proof is due to Lu–Yin [LY13] (see also [GK12] for colorings). We proceed as in the usual derivation of the tree recursion. Given G and r, we create a new graph $\tilde{G}$, where we have split r into d copies $r_1, \ldots, r_d$, with r_i having unique neighbor u_i and fugacity replaced with $\lambda^{1/d}$. Then

$$p_{G,r} = \frac{Z_G(r \leftarrow \mathsf{in})}{Z_G(r \leftarrow \mathsf{in}) + Z_G(r \leftarrow \mathsf{out})} = \frac{Z_{\tilde{G}}(r_1, \ldots, r_d \leftarrow \mathsf{in})}{Z_{\tilde{G}}(r_1, \ldots, r_d \leftarrow \mathsf{in}) + Z_{\tilde{G}}(r_1, \ldots, r_d \leftarrow \mathsf{out})},$$

where for convenience, we write $Z_G(r \leftarrow \mathsf{in})$ (resp. $Z_G(r \leftarrow \mathsf{out})$) denotes the hardcore partition function where we only sum over independent sets containing (resp. not containing) r. If G were a tree, then $\tilde{G}$ would consist of d many trees which are rooted at each $r_1, \ldots, r_d$ and are not connected to one another. However, since G is some arbitrary graph, we instead employ a telescoping trick.

For each $0 \le k \le d$, define $\tilde{G}_k$ to be the graph obtained from $\tilde{G}$ by deleting $r_{k+1}, \ldots, r_d$ and keeping $r_1, \ldots, r_k$. Then in $\tilde{G}_0$, we have deleted all copies of r, and so $\tilde{G}_0 = G - r$. At the other extreme, we have $\tilde{G}_d = \tilde{G}$. Then

$$Z_{\tilde{G}}(r_1, \ldots, r_d \leftarrow \mathsf{in}) = Z_{\tilde{G}_0} \cdot \prod_{k=1}^{d} \frac{Z_{\tilde{G}_k}(r_1, \ldots, r_k \leftarrow \mathsf{in})}{Z_{\tilde{G}_{k-1}}(r_1, \ldots, r_{k-1} \leftarrow \mathsf{in})}. \tag{1}$$

One could write a similar expression for the case where "in" is replaced by "out". For the hardcore model, this simplifies dramatically to $Z_{\tilde{G}}(r_1, \ldots, r_d \leftarrow \mathsf{in}) = Z_{\tilde{G}_0}$ since pinning all copies $r_1, \ldots, r_d$

to “out” is equivalent to deleting all of them. We now interpret each ratio in Eq. (1) as a probability. Since r_k has u_k as its unique neighbor, pinning $r_k \leftarrow \mathsf{in}$ is equivalent to imposing $u_k \leftarrow \mathsf{out}$. In particular,

$$Z_{\tilde{G}_k}(r_1, \ldots, r_k \leftarrow \mathsf{in}) = \lambda^{1/d} \cdot Z_{\tilde{G}_{k-1}}(r_1, \ldots, r_{k-1} \leftarrow \mathsf{in}, u_k \leftarrow \mathsf{out}),$$

and so

$$\frac{Z_{\tilde{G}_k}(r_1, \ldots, r_k \leftarrow \mathsf{in})}{Z_{\tilde{G}_{k-1}}(r_1, \ldots, r_{k-1} \leftarrow \mathsf{in})} = \lambda^{1/d} \cdot (1 - p_{G_k, u_k}),$$

where G_k is the graph obtained from $\tilde{G}_{k-1}$ by pinning $r_1, \ldots, r_{k-1} \leftarrow \mathsf{in}$, or equivalently, deleting $r_1, \ldots, r_{k-1}, u_1, \ldots, u_{k-1}$. Putting all of these observations together, we have

$$\begin{aligned} Z_{\tilde{G}}(r_1, \ldots, r_d \leftarrow \mathsf{in}) &= Z_{\tilde{G}_0} \cdot \lambda \prod_{i=1}^{d} (1 - p_{G_k, u_k}) \\ Z_{\tilde{G}}(r_1, \ldots, r_d \leftarrow \mathsf{out}) &= Z_{\tilde{G}_0}, \end{aligned}$$

and so the claim follows.