Convergence of MCMC and Loopy BP in the Tree Uniqueness Region for the Hard-Core Model

Charilaos Efthymiou* Thomas P. Hayes† Daniel Štefankovič‡ Eric Vigoda §
Yitong Yin¶

August 30, 2016

Abstract

We study the hard-core (gas) model defined on independent sets of an input graph where the independent sets are weighted by a parameter (aka fugacity) $\lambda > 0$. For constant Δ, previous work of Weitz (2006) established an FPTAS for the partition function for graphs of maximum degree Δ when $\lambda < \lambda_c(\Delta)$. Sly (2010) showed that there is no FPRAS, unless NP=RP, when $\lambda > \lambda_c(\Delta)$. The threshold $\lambda_c(\Delta)$ is the critical point for the statistical physics phase transition for uniqueness/non-uniqueness on the infinite Δ-regular tree. The running time of Weitz's algorithm is exponential in $\log \Delta$. Here we present an FPRAS for the partition function whose running time is $O^*(n^2)$. We analyze the simple single-site Markov chain known as the Glauber dynamics for sampling from the associated Gibbs distribution. We prove there exists a constant Δ_0 such that for all graphs with maximum degree $\Delta \geq \Delta_0$ and girth ≥ 7 (i.e., no cycles of length ≤ 6), the mixing time of the Glauber dynamics is $O(n \log n)$ when $\lambda < \lambda_c(\Delta)$. Our work complements that of Weitz which applies for small constant Δ whereas our work applies for all Δ at least a sufficiently large constant Δ_0 (this includes Δ depending on $n = |V|$).

Our proof utilizes loopy BP (belief propagation) which is a widely-used algorithm for inference in graphical models. A novel aspect of our work is using the principal eigenvector for the BP operator to design a distance function which contracts in expectation for pairs of states that behave like the BP fixed point. We also prove that the Glauber dynamics behaves locally like loopy BP. As a byproduct we obtain that the Glauber dynamics, after a short burn-in period, converges close to the BP fixed point, and this implies that the fixed point of loopy BP is a close approximation to the Gibbs distribution. Using these connections we establish that loopy BP quickly converges to the Gibbs distribution when the girth ≥ 6 and $\lambda < \lambda_c(\Delta)$.

*Goethe University, Frankfurt am Main, Germany. Email: efthymiou@gmail.com. Research supported by DFG grant EF 103/11.

†Department of Computer Science, University of New Mexico, Albuquerque, NM 87131. Email: hayes@cs.unm.edu.

‡Department of Computer Science, University of Rochester, Rochester, NY 14627. Email: stefanko@cs.rochester.edu. Research supported in part by NSF grant CCF-1318374.

§School of Computer Science, Georgia Institute of Technology, Atlanta GA 30332. Email: ericvigoda@gmail.com. Research supported in part by NSF grants CCF-1217458 and CCF-1555579.

¶State Key Laboratory for Novel Software Technology, Nanjing University, China. Email: yinyt@nju.edu.cn. Research supported by NSFC grants 61272081 and 61321491.

This work was done in part while all of the authors were visiting the Simons Institute for the Theory of Computing.

1 Introduction

Background

The hard-core gas model is a natural combinatorial problem that has played an important role in the design of new approximate counting algorithms and for understanding computational connections to statistical physics phase transitions. For a graph $G = (V, E)$ and a fugacity $\lambda > 0$, the hard-core model is defined on the set Ω of independent sets of G where $\sigma \in \Omega$ has weight $w(\sigma) = \lambda^{|\sigma|}$. The equilibrium state of the system is described by the Gibbs distribution μ in which an independent set σ has probability $\mu(\sigma) = w(\sigma)/Z$. The partition function $Z = \sum_{\sigma\in\Omega} w(\sigma)$.

We study the closely related problems of efficiently approximating the partition function and approximate sampling from the Gibbs distribution. These problems are important for Bayesian inference in graphical models where the Gibbs distribution corresponds to the posterior or likelihood distributions. Common approaches used in practice are Markov Chain Monte Carlo (MCMC) algorithms and message passing algorithms, such as loopy BP (belief propagation), and one of the aims of this paper is to prove fast convergence of these algorithms.

Exact computation of the partition function is #P-complete [37], even for restricted input classes [9], hence the focus is on designing an efficient approximation scheme, either a deterministic FPTAS or randomized FPRAS. The existence of an FPRAS for the partition function is polynomial-time inter-reducible to approximate sampling from the Gibbs distribution.

A beautiful connection has been established: there is a computational phase transition on graphs of maximum degree Δ that coincides with the statistical physics phase transition on Δ-regular trees. The critical point for both of these phase transitions is $\lambda_c(\Delta) := (\Delta-1)^{\Delta-1}/(\Delta-2)^{\Delta}$. In statistical physics, $\lambda_c(\Delta)$ is the critical point for the uniqueness/non-uniqueness phase transition on the infinite Δ-regular tree $\mathbb{T}_\Delta$ [17] (roughly speaking, this is the phase transition for the decay versus persistence of the influence of the leaves on the root). For some basic intuition about the value of this critical point, note its asymptotics $\lambda_c(\Delta) \sim e/(\Delta-2)$ and the following basic property: $\lambda_c(\Delta) > 1$ for $\Delta \leq 5$ and $\lambda_c(\Delta) < 1$ for $\Delta \geq 6$.

Weitz [41] showed, for all constant Δ, an FPTAS for the partition function for all graphs of maximum degree Δ when $\lambda < \lambda_c(\Delta)$. To properly contrast the performance of our algorithm with Weitz's algorithm let us state his result more precisely: for all $\delta > 0$, there exists constant $C = C(\delta)$, for all Δ, all $G = (V, E)$ with maximum degree Δ, all $\lambda < (1-\delta)\lambda_c(\Delta)$, all $\epsilon > 0$, there is a deterministic algorithm to approximate Z within a factor $(1 \pm \epsilon)$ with running time $O\left((n/\epsilon)^{C\log\Delta}\right)$. An important limitation of Weitz's result is the exponential dependence on $\log\Delta$ in the running time. Hence it is polynomial-time only for constant Δ, and even in this case the running time is unsatisfying.

On the other side, Sly [33] (extended in [6, 7, 34, 8]) has established that, unless $NP = RP$, for all $\Delta \geq 3$, there exists $\gamma > 0$, for all $\lambda > \lambda_c(\Delta)$, there is no polynomial-time algorithm for triangle-free Δ-regular graphs to approximate the partition function within a factor $2^{\gamma n}$.

Weitz's algorithm was extremely influential: many works have built upon his algorithmic approach to establish efficient algorithms for a variety of problems (e.g., [28, 31, 18, 19, 32, 38, 20, 30, 21]). One of its key conceptual contributions was showing how decay of correlations properties on a Δ-regular tree are connected to the existence of an efficient algorithm for graphs of maximum degree Δ. We believe our paper enhances this insight by connecting these same decay of correlations properties on a Δ-regular tree to the analysis of widely-used Markov Chain Monte Carlo (MCMC) and message passing algorithms.

Main Results

As mentioned briefly earlier on, there are two widely-used approaches for the associated approximate counting/sampling problems, namely MCMC and message passing approaches. A popular MCMC algorithm is the simple single-site update Markov chain known as the Glauber dynamics. The Glauber dynamics is a Markov chain (X_t) on Ω whose transitions $X_t \to X_{t+1}$ are defined by the following process:

1. Choose v uniformly at random from V.
2. If $N(v) \cap X_t = \emptyset$ then let
$$X_{t+1} = \begin{cases} X_t \cup \{v\} & \text{with probability } \lambda/(1+\lambda) \\ X_t \setminus \{v\} & \text{with probability } 1/(1+\lambda) \end{cases}$$
3. If $N(v) \cap X_t \neq \emptyset$ then let $X_{t+1} = X_t$.

The mixing time T_{mix} is the number of steps to guarantee that the chain is within a specified (total) variation distance of the stationary distribution. In other words, for $\epsilon > 0$,

$$T_{\text{mix}}(\epsilon) = \min\{t : \text{ for all } X_0, d_{\mathsf{TV}}(X_t, \mu) \leq \epsilon\},$$

where $d_{\mathsf{TV}}()$ is the variation distance. We use $T_{\text{mix}} = T_{\text{mix}}(1/4)$ to refer to the mixing time for $\epsilon = 1/4$.

It is natural to conjecture that the Glauber dynamics has mixing time $O(n \log n)$ for all $\lambda < \lambda_c(\Delta)$. Indeed, Weitz's work implies rapid mixing for $\lambda < \lambda_c(\Delta)$ for amenable graphs. On the other hand Mossel et al. in [25] show slow mixing when $\lambda > \lambda_c(\Delta)$ on random regular bipartite graphs. The previously best known results for MCMC algorithms are far from reaching the critical point. It was known that the mixing time of the Glauber dynamics (and other simple, local Markov chains) is $O(n \log n)$ when $\lambda < 2/(\Delta - 2)$ for any graph with maximum degree Δ [5, 22, 39]. In addition, [13] analyzed Δ-regular graphs with $\Delta = \Omega(\log n)$ and presented a polynomial-time simulated annealing algorithm when $\lambda < \lambda_c(\Delta)$.

Here we prove $O(n \log n)$ mixing time up to the critical point when the maximum degree is at least a sufficiently large constant Δ_0, and there are no cycles of length ≤ 6 (i.e., girth ≥ 7).

Theorem 1. *For all $\delta > 0$, there exists $\Delta_0 = \Delta_0(\delta)$ and $C = C(\delta)$, for all graphs $G = (V, E)$ of maximum degree $\Delta \geq \Delta_0$ and girth ≥ 7, all $\lambda < (1 - \delta)\lambda_c(\Delta)$, all $\epsilon > 0$, the mixing time of the Glauber dynamics satisfies:*

$$T_{\text{mix}}(\epsilon) \leq Cn \log(n/\epsilon).$$

Note that Δ and λ can be a function of $n = |V|$. The above sampling result yields (via [35, 15]) an FPRAS for estimating the partition function Z with running time $O^*(n^2)$ where $O^*()$ hides multiplicative $\log n$ factors. The algorithm of Weitz [41] is polynomial-time for small constant Δ, in contrast our algorithm is polynomial-time for all $\Delta > \Delta_0$ for a sufficiently large constant Δ_0.

A family of graphs of particular interest are random Δ-regular graphs and random Δ-regular *bipartite* graphs. These graphs do not satisfy the girth requirements of Theorem 1 but they have few short cycles. Hence, as one would expect the above result extends to these graphs.

Theorem 2. *For all $\delta > 0$, there exists $\Delta_0 = \Delta_0(\delta)$ and $C = C(\delta)$, for all $\Delta \geq \Delta_0$, all $\lambda < (1 - \delta)\lambda_c(\Delta)$, all $\epsilon > 0$, with probability $1 - o(1)$ over the choice of an n-vertex graph G chosen uniformly at random from the set of all Δ-regular (bipartite) graphs, the mixing time of the Glauber dynamics on G satisfies:*

$$T_{\text{mix}}(\epsilon) \leq Cn \log(n/\epsilon).$$

Theorem 2 complements the work in [25] which shows slow mixing for random Δ-regular bipartite graphs when $\lambda > \lambda_c(\Delta)$.

The other widely used approach is BP (belief propagation) based algorithms. BP, introduced by Pearl [27], is a simple recursive scheme designed on trees to correctly compute the marginal distribution for each vertex to be occupied/unoccupied. In particular, consider a rooted tree $T = (V, E)$ where for $v \in V$ its parent is denoted as p and its children are $N(v)$. Let

$$q(v) = \mathbf{Pr}_\mu \left[v \text{ is occupied} \mid p \text{ is unoccupied}\right]$$

denote the probability in the Gibbs distribution that v is occupied conditional on its parent p being unoccupied. It is convenient to work with ratios of the marginals, and hence let $R_{v\to p(v)} = q(v)/(1-q(v))$ denote the ratio of the occupied to unoccupied marginal probabilities. Because T is a tree then it is not difficult to show that this ratio satisfies the following recurrence:

$$R_{v\to p(v)} = \lambda \prod_{w\in N(v)\setminus\{p(v)\}} \frac{1}{1+R_{w\to v}}.$$

This recurrence explains the terminology of BP that $R_{w\to v}$ is a "message" from w to its parent v. Given the messages to v from all of its children then v can send its message to its parent. Finally the root r (with a parent p always fixed to be unoccupied and thus removed) can compute the marginal probability that it is occupied by: $q(r) = R_{r\to p}/(1 + R_{r\to p})$.

The above formulation defines (the sum-product version of) BP a simple, natural algorithm which works efficiently and correctly for trees. For general graphs *loopy BP* implements the above approach, even though there are now cycles and so the algorithm no longer is guaranteed to work correctly. For a graph $G = (V, E)$, for $v \in V$ let $N(v)$ denote the set of all neighbors of v. For each $p \in N(v)$ and time $t \geq 0$ we define a message

$$R^t_{v\to p} = \lambda \prod_{w\in N(v)\setminus\{p\}} \frac{1}{1+R^{t-1}_{w\to v}}.$$

The corresponding estimate of the marginal can be computed from the messages by:

$$q^t(v,p) = \frac{R^t_{v\to p}}{1+R^t_{v\to p}}. \tag{1}$$

Loopy BP is a popular algorithm for estimating marginal probabilities in general graphical models (e.g., see [26]), but there are few results on when loopy BP converges to the Gibbs distribution (e.g., Weiss [40] analyzed graphs with one cycle, and [36, 14, 16] presented various sufficient conditions, see also [2, 29] for analysis of BP variants). We have an approach for analyzing loopy BP and in this project we will prove that loopy BP works well in a broad range of parameters. Its behavior relates to phase transitions in the underlying model, we detail our approach and expected results after formally presenting phase transitions.

We prove that, on any graph with girth ≥ 6 and maximum degree $\Delta \geq \Delta_0$ where Δ_0 is a sufficiently large constant, loopy BP quickly converges to the (marginals of) Gibbs distribution μ. More precisely, $O(1)$ iterations of loopy BP suffices, note each iteration of BP takes $O(n+m)$ time where $n = |V|$ and $m = |E|$.

Theorem 3. *For all $\delta, \epsilon > 0$, there exists $\Delta_0 = \Delta_0(\delta, \epsilon)$ and $C = C(\delta, \epsilon)$, for all graphs $G = (V, E)$ of maximum degree $\Delta \geq \Delta_0$ and girth ≥ 6, all $\lambda < (1-\delta)\lambda_c(\Delta)$, the following holds: for $t \geq C$, for all $v \in V$, $p \in N(v)$,*

$$\left|\frac{q^t(v,p)}{\mu(v \text{ is occupied} \mid p \text{ is unoccupied})} - 1\right| \leq \epsilon$$

where $\mu(\cdot)$ is the Gibbs distribution.

Contributions

Our main conceptual contribution is formally connecting the behavior of BP and the Glauber dynamics. We will analyze the Glauber dynamics using path coupling [1]. In path coupling we need to analyze a pair of *neighboring configurations*, in our setting this is a pair of independent sets X_t, Y_t which differ at exactly one vertex v. The key is to construct a one-step coupling $(X_t, Y_t) \to (X_{t+1}, Y_{t+1})$ and introduce a distance function $\Phi : \Omega \times \Omega \to \mathbf{R}_{\geq 0}$ which "contracts" meaning that the following *path coupling condition* holds for some $\gamma > 0$:

$$\mathbb{E}\left[\Phi(X_{t+1}, Y_{t+1}) \mid X_t, Y_t\right] \leq (1-\gamma)\Phi(X_t, Y_t).$$

We use a simple maximal one-step coupling and hence in our setting the path coupling condition simplifies to:

$$(1-\gamma)\Phi(X_t, Y_t) \geq \sum_{z \in N(v)} \frac{\lambda}{1+\lambda} \mathbf{1}\left\{z \text{ is unblocked in } X_t\right\} \Phi(z),$$

where *unblocked* means that $N(z) \cap X_t = \emptyset$, i.e., all neighbors of z are unoccupied, and we have assumed there are no triangles so as to ignore the possibility that X_t and Y_t differ on the neighborhood of z.

The distance function Φ must satisfy a few basic conditions such as being a path metric, and if $X \neq Y$ then $\Phi(X, Y) \geq 1$ (so that by Markov's inequality $\mathbf{Pr}\left[X_t \neq Y_t\right] \leq \mathbb{E}\left[\Phi(X_t, Y_t)\right]$). A standard choice for the distance function is the Hamming distance. In our setting the Hamming distance does not suffice and our primary challenge is determining a suitable distance function.

We cannot construct a suitable distance function which satisfies the path coupling condition for arbitrary neighboring pairs X_t, Y_t. But, a key insight is that we can show the existence of a suitable Φ when the local neighborhood of the disagreement v behaves like the BP fixpoint. Our construction of this Φ is quite intriguing.

In our proofs it is useful to consider the (unrooted) BP recurrences corresponding to the probability that a vertex is unblocked. This corresponds to the following function $F : [0,1]^V \to [0,1]^V$ which is defined as follows, for any $\omega \in [0,1]^V$ and $z \in V$:

$$F(\omega)(z) = \prod_{y \in N(z)} \frac{1}{1+\lambda\omega(y)}. \tag{2}$$

Also, for some integer $i \geq 0$, let $F^i(\omega) : [0,1]^V \to [0,1]^V$ be the i-iterate of F. This recurrence is closely related to the standard BP operator $R()$ and hence under the hypotheses of our main results, we have that $F()$ has a unique fixed point ω^*, and for any ω, all $z \in V$, $\lim_{i\to\infty} F^i(z) = \omega^*(z)$.

To construct the distance function Φ we start with the Jacobian of this BP operator $F()$. By a suitable matrix diagonalization we obtain the path coupling condition. Since $F()$ converges to a fixed point, and, in fact, it contracts at every level with respect to an appropriately defined potential function, we then know that the Jacobian of the BP operator $F()$ evaluated at its fixed point ω^* has spectral radius < 1 and hence the same holds for the path coupling condition for pairs of states that are BP fixed points. This yields a function Φ that satisfies the following system of inequalities

$$\Phi(v) > \sum_{z \in N(v)} \frac{\lambda\omega^*(z)}{1+\lambda\omega^*(z)} \Phi(z). \tag{3}$$

However for the path coupling condition a stronger version of the above is necessary. More specifically, the sum on the r.h.s. should be appropriately bounded away from $\Phi(v)$, i.e. we need to have

$$(1-\gamma)\Phi(v) > \sum_{z\in N(v)} \frac{\lambda\omega^*(z)}{1+\lambda\omega^*(z)}\Phi(z).$$

Additionally, Φ should be a distance metric, e.g. $\Phi > 0$. It turns out that we use further properties of the distance function Φ, hence we need to explicitly derive a Φ.

There are previous works [11, 12] which utilize the spectral radius of the adjacency matrix of the input graph G to design a suitable distance function for path coupling. In contrast, we use insights from the analysis of the BP operator to derive a suitable distance function. We believe this is a richer connection that can potentially lead to stronger results since it directly relates to convergence properties on the tree. Our approach has the potential to apply for a more general class of spin systems, we comment on this in more detail in the conclusions.

The above argument only implies that we have contraction in the path coupling condition for pairs of configurations which are BP fixed points. A priori we don't even know if the BP fixed points on the tree correspond to the Gibbs distribution on the input graph. We prove that the Glauber dynamics (approximately) satisfies a recurrence that is close to the BP recurrence; this builds upon ideas of Hayes [10] for colorings. This argument requires that there are no cycles of length ≤ 6 for the Glauber dynamics (and no cycles of length ≤ 5 for the direct analysis of the Gibbs distribution). Some local sparsity condition is necessary since if there are many short cycles then the Gibbs distribution no longer behaves similarly to a tree and hence loopy BP may be a poor estimator.

As a consequence of the above relation between BP and the Glauber dynamics, we establish that from an arbitrary initial configuration X_0, after a short burn-in period of $T = O(n\log\Delta)$ steps of the Glauber dynamics the configuration X_T is a close approximation to the BP fixed point. In particular, for any vertex v, the number of unblocked neighbors of v in X_T is $\approx \sum_{z\in N(v)}\omega^*(z)$ with high probability. As is standard for concentration results, our proof of this result necessitates that Δ is at least a sufficiently large constant. Finally we adapt ideas of [4] to utilize these burn-in properties and establish rapid mixing of the Glauber dynamics.

Outline of Paper

In the following section we state results about the convergence of the BP recurrences. We then present in Section 3 our theorem showing the existence of a suitable distance function for path coupling for pairs of states at the BP fixed point. Section 4 sketches the proofs for our local uniformity results that after a burn-in period the Glauber dynamics behaves locally similar to the BP recurrences. Finally, in Section 5 we outline the proof of Theorem 1 of rapid mixing for the Glauber dynamics. The extension to random regular (bipartite) graphs as stated in Theorem 2 is proven in Section F of the appendix. Theorem 3 about the efficiency of loopy BP is proven in Section B of appendix, the key technical results in the proof are sketched in Section 4.

The full proofs of our results are quite lengthy and so we defer many to the appendix.

2 BP Convergence

Here we state several useful results about the convergence of BP to a unique fixed point, and stepwise contraction of BP to the fixed point. The lemmas presented in this section are proved in Section A of the appendix.

Our first lemma (which is proved using ideas from [28, 19, 31]) says that the recurrence for $F()$ defined in (2) has a unique fixed point.

Lemma 4. *For all $\delta > 0$, there exists $\Delta_0 = \Delta_0(\delta)$, for all $G = (V, E)$ of maximum degree $\Delta \geq \Delta_0$, all $\lambda < (1-\delta)\lambda_c(\Delta)$, the function F has a unique fixed point ω^*.*

A critical result for our approach is that the recurrences $F()$ have stepwise contraction to the fixed point ω^*. To obtain contraction we use the following potential function Ψ. Let the function $\Psi : [0,1] \rightarrow \mathbb{R}_{\geq 0}$ be as follows,

$$\Psi(x) = (\sqrt{\lambda})^{-1} \text{arcsinh}\left(\sqrt{\lambda \cdot x}\right). \tag{4}$$

Our main motivation for introducing Ψ is as a normalizing potential function that we use to define the following distance metric, D, on functions $\omega \in [0,1]^V$:

$$D(\omega_1, \omega_2) = \max_{z \in V} |\Psi(\omega_1(z)) - \Psi(\omega_2(z))|.$$

We will also need a variant, $D_{v,R}$, of this metric whose value only depends on the restriction of the function to a ball of radius ℓ around vertex v. For any $v \in V$, integer $\ell \geq 0$, let $B(v, \ell)$ be the set of vertices within distance $\leq \ell$ of v. Moreover, for functions $\omega_1, \omega_2 \in [0,1]^V$, we define:

$$D_{v,\ell}(\omega_1, \omega_2) = \max_{z \in B(v,\ell)} |\Psi(\omega_1(z)) - \Psi(\omega_2(z))|. \tag{5}$$

We can now state the following convergence result for the recurrences, which establishes stepwise contraction.

Lemma 5. *For all $\delta > 0$, there exists $\Delta_0 = \Delta_0(\delta)$, for all $G = (V, E)$ of maximum degree $\Delta \geq \Delta_0$, all $\lambda < (1-\delta)\lambda_c(\Delta)$, for any $\omega \in [0,1]^V$, $v \in V$ and $\ell \geq 1$, we have:*

$$D_{v,\ell-1}(F(\omega), \omega^*) \leq (1 - \delta/6) D_{v,\ell}(\omega, \omega^*).$$

where ω^* is the fixed point of F.

3 Path Coupling Distance Function

We now prove that there exists a suitable distance function Φ for which the path coupling condition holds for configurations that correspond to the fixed points of $F()$.

Theorem 6. *For all $\delta > 0$, there exists $\Delta_0 = \Delta_0(\delta)$, for all $G = (V, E)$ of maximum degree $\Delta \geq \Delta_0$, all $\lambda < (1-\delta)\lambda_c(\Delta)$, there exists $\Phi : V \rightarrow \mathbb{R}_{\geq 0}$ such that for every $v \in V$,*

$$1 \leq \Phi(v) \leq 12, \tag{6}$$

and

$$(1 - \delta/6)\Phi(v) \geq \sum_{u \in N(v)} \frac{\lambda \omega^*(u)}{1 + \lambda \omega^*(u)} \Phi(u), \tag{7}$$

where ω^* is the fixed point of F defined in (2).

Proof. We will prove here that the convergence of BP provides the existence of a distance function Φ satisfying (7). We defer the technical proof of (6) to Section A of the appendix.

The Jacobian J of the BP operator F is given by

$$J(v,u) \;=\; \left|\frac{\partial F(\omega)(v)}{\partial \omega(u)}\right| \;=\; \begin{cases} \frac{\lambda F(\omega)(v)}{1+\lambda\omega(u)} & \text{if } u \in N_v \\ 0 & \text{otherwise} \end{cases}$$

Let $J^* = J|_{\omega=\omega^*}$ denote the Jacobian at the fixed point $\omega = \omega^*$. Let D be the diagonal matrix with $D(v,v) = \omega^*(v)$ and let $\hat{J} = D^{-1}J^*D$.

The path coupling condition (7) is in fact

$$\hat{J}\Phi \leq (1-\delta/6)\Phi. \tag{8}$$

The fact that ω^* is a Jacobian attractive fixpoint implies the existence of a nonnegative Φ with $\hat{J}\Phi < \Phi$. Thus, the theorem would follow immediately if the spectral radius of $\hat{J}$ is $\rho(\hat{J}) \leq 1-\delta/6$ and $\hat{J}$ has a principal eigenvector with each entry from the bounded range $[1,12]$. However, explicitly calculating this principal eigenvector can be challenging on general graphs.

The convergence of BP which is established in Lemmas 4, 5, with respect to the potential function Ψ, guides us to an explicit construction of Φ such that $\hat{J}\Phi < \Phi$. Indeed, let $\Psi'(x) = \frac{1}{2\sqrt{x(1+\lambda x)}}$ denote the derivative of the potential function Ψ. It will follow from the proof of Lemma 5 that:

$$\sum_{u\in N(v)} J^*(v,u)\frac{\Psi'(\omega^*(v))}{\Psi'(\omega^*(u))} \leq 1-\delta/6.$$

This inequality is due to the contraction of the BP system at the fixed point with respect to the potential function Ψ. It is equivalent to the following:

$$\sum_{u\in N(v)} \frac{\hat{J}(v,u)}{\omega^*(u)\Psi'(\omega^*(u))} \leq \frac{1-\delta/6}{\omega^*(v)\Psi'(\omega^*(v))}.$$

Then, (8) is trivially satisfied by choosing Φ such that $\Phi(v) = \frac{1}{2\omega^*(v)\Psi'(\omega^*(v))} = \sqrt{\frac{1+\lambda\omega^*(v)}{\omega^*(v)}}$. In turn we get the path coupling condition (7). The verification of (6) is in Section A of the appendix.

□

4 Local Uniformity for the Glauber Dynamics

We will prove that the Glauber dynamics, after a sufficient burn-in, behaves with high probability locally similar to the BP fixed points. In this section we will formally state some of these "local uniformity" results and sketch the main ideas in their proof. The proofs are quite technical and deferred to Section D of the appendix.

For an independent set σ, for $v \in V$, and $p \in N(v)$ let

$$\mathbf{U}_{v,p}(\sigma) = \mathbf{1}\{\sigma \cap (N(v)\setminus\{p\}) = \emptyset\} \tag{9}$$

be the indicator of whether the children of v leave v unblocked.

We now state our main local uniformity results. We first establish that the Gibbs distribution behaves as in the BP fixpoint, when the girth ≥ 6. We will prove that for any vertex v, the number of unblocked neighbors of v is $\approx \sum_{z\in N(v)}\omega^*(z)$ with high probability. Hence, for $v \in V$ let

$$\mathbf{S}_X(v) = \sum_{z\in N(v)} \mathbf{U}_{z,v}(X),$$

denote the number of unblocked neighbors of v in configuration X.

Theorem 7. *For all $\delta, \epsilon > 0$, there exists $\Delta_0 = \Delta_0(\delta, \epsilon)$ and $C = C(\delta, \epsilon)$, for all graphs $G = (V, E)$ of maximum degree $\Delta \geq \Delta_0$ and girth ≥ 6, all $\lambda < (1 - \delta)\lambda_c(\Delta)$, for all $v \in V$, it holds that:*

$$\mathbf{Pr}_{X\sim\mu}\left[\left|\mathbf{S}_X(v) - \sum_{z\in N(v)} \omega^*(z)\right| \leq \epsilon\Delta\right] \geq 1 - \exp\left(-\Delta/C\right),$$

where ω^* is the fixpoint from Lemma 4.

Theorem 7 will be the key ingredient in the proof of Theorem 3 (to be precise, the upcoming Lemma 9 is the key element in the proofs of Theorems 3 and 7).

For our rapid mixing result (Theorem 2) we need an analogous local uniformity result for the Glauber dynamics. This will require the slightly higher girth requirement ≥ 7 since the grandchildren of a vertex v no longer have a certain conditionally independence and we need the additional girth requirement to derive an approximate version of the conditional independence (this is discussed in more detail in Section C.3 of the appendix).

The path coupling proof weights the vertices according to Φ. Hence, in place of $\mathbf{S}$ we need the following weighted version $\mathbf{W}$. For $v \in V$ and $\Phi : V \to \mathbb{R}_{\geq 0}$ as defined in Theorem 6 let

$$\mathbf{W}_\sigma(v) = \sum_{z\in N(v)} \mathbf{U}_{z,v}(\sigma)\ \Phi(z). \tag{10}$$

We then prove that the Glauber dynamics, after sufficient burn-in, also behaves as in the BP fixpoint with a slightly higher girth requirement ≥ 7. (For path coupling we only need an upper bound on the number of unblocked neighbors, hence we state and prove this simpler form.)

Theorem 8. *For all $\delta, \epsilon > 0$, let $\Delta_0 = \Delta_0(\delta, \epsilon), C = C(\delta, \epsilon)$, for all graphs $G = (V, E)$ of maximum degree $\Delta \geq \Delta_0$ and girth ≥ 7, all $\lambda < (1 - \delta)\lambda_c(\Delta)$, let (X_t) be the Glauber dynamics on the hard-core model. For all $v \in V$, it holds that*

$$\begin{aligned}\mathbf{Pr}\left[(\forall t \in \mathcal{I}) \quad \mathbf{W}_{X_t}(v) < \sum_{z\in N(v)} \omega^*(z)\Phi(z) + \epsilon\Delta\right] \\ \geq \quad 1 - \exp\left(-\Delta/C\right),\end{aligned} \tag{11}$$

where the time interval $\mathcal{I} = [Cn \log \Delta, n \exp\left(\Delta/C\right)]$.

4.1 Proof sketch for local uniformity results

Here we sketch the simpler proof of Theorem 7 of the local uniformity results for the Gibbs distribution. This will illustrate the main conceptual ideas in the proof for the Gibbs distribution, and we will indicate the extra challenge for the analysis of the Glauber dynamics in the proof of Theorem 8. The full proofs for Theorems 7 and 8 are in Section D of the appendix.

Consider a graph $G = (V, E)$. For a vertex v and an independent set σ, consider the following quantity:

$$\mathbf{R}(\sigma, v) = \prod_{z\in N(v)} \left(1 - \frac{\lambda}{1+\lambda}\mathbf{U}_{z,v}(\sigma)\right), \tag{12}$$

where $\mathbf{U}_{z.v}(\sigma)$ is defined in (9) (it is the indicator that the children of z leave it unblocked). The important aspect of this quantity $\mathbf{R}$ is the following qualitative interpretation. Let Y be distributed as in the Gibbs measure w.r.t. G. For triangle-free G we have

$$\mathbf{R}(\sigma, v) = \mathbf{Pr}\left[v \text{ is unblocked} \mid v \notin Y,\ Y(S_2(v)) = \sigma(S_2(v))\right],$$

where $S_2(z)$ are those vertices distance 2 from z and by "$z \notin \sigma$" we mean that z is not occupied. Moreover, conditional on the configuration at z and $S_2(z)$ the neighbors of z are independent in the Gibbs distribution and hence:

$$\mathbf{R}(\sigma, v) = \prod_{z \in N(v)} \mathbf{Pr}\left[z \notin Y \mid v \notin Y,\ Y(S_2(v)) = \sigma(S_2(v))\right]. \tag{13}$$

In the special case where the underlying graph is a *tree* we can extend (13) to the following recursive equations: Let X be distributed as in μ. We have that

$$\mathbf{R}(X, v) = \prod_{z \in N(v)} \left(1 - \frac{\lambda}{1+\lambda}\mathbf{R}(X, z)\right) + O(1/\Delta), \tag{14}$$

For our purpose it turns out that $\mathbf{R}(X, \cdot)$ is an approximate version of $F()$ defined in (2). The error term $O(1/\Delta)$ in (14) is negligible. For understanding $\mathbf{R}(X, \cdot)$ qualitatively, this error term can be completely ignored.

Consider the (BP system of) equations in (14), which is exact on trees. Nothing prevents us from applying (14) on the graph G and get the loopy version of the equations. Now, (14) does not necessarily compute the probability for v to be unblocked. However, we show the following interesting result regarding the quantity $\mathbf{S}_X(v)$, for every $v \in V$. With probability $\geq 1 - \exp(-\Omega(\Delta))$, it holds that

$$\left|\mathbf{S}_X(v) - \sum_{z \in N(v)} \mathbf{R}(X, z)\right| \leq \epsilon \Delta. \tag{15}$$

That is, we can approximate $\mathbf{S}_X(v)$ by using quantities that arise from the loopy BP equations. Still, getting a handle on $\mathbf{R}(X, z)$ in (15) is a non-trivial task. To this end, we show that $X \sim \mu$ satisfies (14) in the following approximate sense:

Lemma 9. *For all $\gamma, \delta > 0$, there exists $\Delta_0, C > 0$, for all graphs $G = (V, E)$ of maximum degree $\Delta \geq \Delta_0$ and girth ≥ 6 all $\lambda < (1-\delta)\lambda_c(\Delta)$ for all $v \in V$ the following is true:*

Let X be distributed as in μ. Then with probability $\geq 1 - \exp(-\Delta/C)$ it holds that

$$\left|\mathbf{R}(X, v) - \prod_{z \in N(v)} \left(1 - \frac{\lambda}{1+\lambda}\mathbf{R}(X, z)\right)\right| < \gamma. \tag{16}$$

We will argue (via (16)) that $\mathbf{R}()$ is an approximate version of $F()$ and then we can apply Lemma 5 to deduce convergence (close) to the fixpoint ω^*. Consequently, we will prove that for every $v \in V$, with probability at least $1 - \exp(-\Omega(\Delta))$, it holds that

$$|\mathbf{R}(X, v) - \omega^*(v)| \leq \epsilon. \tag{17}$$

(See Lemma 16 in Section B.2 of the appendix for a formal statement.) Combining (17) and (15) will finish the proof of Theorem 7. For the detailed proof of Theorem 7 see Section D in the appendix.

4.2 Approximate recurrence - Proof of Lemma 9

Here we prove Lemma 9 which shows that $\mathbf{R}$ satisfies an approximate recurrence similar to loopy BP, this is the main result in the proof of Theorem 7. Before beginning the proof we illustrate the necessity of the girth assumption.

Recall that for triangle-free graphs we have conditional independence in (13) for the neighbors of vertex z. In (15) we need to consider $\sum_{z\in N(v)} \mathbf{R}(X,z)$. To get independence on the grandchildren of v we need to condition on $S_3(v)$, this will require girth ≥ 6, see (18) below.

Proof of Lemma 9. Consider X distributed as in μ. Given some vertex $v \in V$, let $\mathcal{F}$ be the σ-algebra generated by the configuration of v and the vertices at distance ≥ 3 from v.

Note that $\lambda_c(\Delta) \sim e/\Delta$. So, for $\lambda < \lambda_c(\Delta)$ and $\Delta > \Delta_0$ we have $\lambda = O(1/\Delta)$.

Note that $\mathbf{S}_X(v)$ is a function of the configuration at $S_2(v)$. Conditional on $\mathcal{F}$, for any $z, z' \in N(v)$ the configurations at $N(z)\backslash\{v\}$ and $N(z')\backslash\{v\}$ are independent with each other. That is, conditional on $\mathcal{F}$, the quantity $\mathbf{S}_X(v)$ is a sum of $|N(v)|$ many independent random variables in $\{0,1\}$. Then, applying Azuma's inequality (the Lipschitz constant is 1) we get that

$$\mathbf{Pr}\left[\left|\mathbb{E}\left[\mathbf{S}_X(v) \mid \mathcal{F}\right] - \mathbf{S}_X(v)\right| \leq \beta\Delta\right] \geq 1 - 2\exp\left(-\beta^2\Delta/2\right), \tag{18}$$

for any $\beta > 0$.

For $x \in \mathbb{R}_{\geq 0}$, let $f(x) = \exp\left(-\frac{\lambda}{1+\lambda}x\right)$. Since $\lambda \leq e/\Delta$ for $\Delta \geq \Delta_0$, then for $|\gamma| \leq (3e)^{-1}$ it holds that $f(x+\gamma\Delta) \leq 10\gamma$. Using these observations and (18) we get the following: for $0 < \beta < (3e)^{-1}$ it holds that

$$\begin{aligned}\mathbf{Pr}\left[\left|f(\mathbf{S}_X(v)) - f(\mathbb{E}\left[\mathbf{S}_X(v) \mid \mathcal{F}\right])\right| \leq 10\beta\right]& \\ \geq \quad 1 - 2\exp\left(-\beta^2\Delta/2\right).&\end{aligned} \tag{19}$$

Recalling the definition of $\mathbf{R}(X,v)$, we have that

$$\begin{aligned}\mathbf{R}(X,v) &= \prod_{z\in N(v)}\left(1 - \frac{\lambda}{1+\lambda}\mathbf{U}_{z,v}(X)\right) \\ &= \exp\left(-\frac{\lambda}{1+\lambda}\sum_{z\in N(v)}\mathbf{U}_{z,v}(X) + O\left(1/\Delta\right)\right) \\ &= f(\mathbf{S}_X(v)) + O\left(1/\Delta\right),\end{aligned} \tag{20}$$

where the second equality we use the fact that $\lambda = O(1/\Delta)$ and that for $|x| < 1$ we have $1 + x = \exp(x + O(x^2))$; the last equality follows by noting that $f(\mathbf{S}_X(v)) \leq 1$.

We are now going to show that for every $z \in N(v)$ it holds that

$$\left|\mathbb{E}\left[\mathbf{U}_{z,v}(X) \mid \mathcal{F}\right] - \mathbf{R}(X,z)\right| \leq 2\lambda. \tag{21}$$

Before showing that (21) is indeed correct, let us show how we use it to get the lemma.

We have that

$$\begin{aligned}&f(\mathbb{E}\left[\mathbf{S}_X(v) \mid \mathcal{F}\right]) \\ &\quad= \exp\left(-\frac{\lambda}{1+\lambda}\sum_{z\in N(v)}\mathbb{E}\left[\mathbf{U}_{z,v}(X_t) \mid \mathcal{F}\right]\right) \\ &\quad= \exp\left(-\frac{\lambda}{1+\lambda}\sum_{z\in N(v)}\mathbf{R}(X,z)\right) + O(1/\Delta),\end{aligned} \tag{22}$$

where in the first derivation we use linearity of expectation and in the second derivation we use (21) and the fact that $\lambda = O(1/\Delta)$.

The lemma follows by plugging (22) and (20) into (19) and taking sufficiently large Δ.

It remains to show (21). We first get an appropriate upper bound for $\mathbb{E}\left[\mathbf{U}_{z,v}(X) \mid \mathcal{F}\right]$. Using the fact that $\mathbf{U}_{z,w}(X) \le 1$ and $\mathbf{Pr}\left[z \in X | \mathcal{F}\right] \le \lambda$ we have that

$$
\begin{aligned}
&\mathbb{E}\left[\mathbf{U}_{z,v}(X) \mid \mathcal{F}\right] \\
&\quad = \mathbb{E}\left[\mathbf{U}_{z,v}(X) \mid \mathcal{F}, z \in X\right] \cdot \mathbf{Pr}\left[z \in X | \mathcal{F}\right] \\
&\qquad + \mathbb{E}\left[\mathbf{U}_{z,v}(X) \mid \mathcal{F}, z \notin X\right] \cdot \mathbf{Pr}\left[z \notin X | \mathcal{F}\right] \\
&\quad \le \mathbf{Pr}\left[z \in X | \mathcal{F}\right] + \mathbb{E}\left[\mathbf{U}_{z,v}(X) \mid \mathcal{F}, z \notin X\right] \\
&\quad \le \lambda + \mathbb{E}\left[\mathbf{U}_{z,v}(X) \mid \mathcal{F}, z \notin X\right] \\
&\quad = \lambda + \prod_{u \in N(z)\setminus\{v\}} \left(1 - \frac{\lambda}{1+\lambda}\mathbf{U}_{u,z}(X)\right) \qquad (23) \\
&\quad \le 2\lambda + \prod_{u \in N(z)} \left(1 - \frac{\lambda}{1+\lambda}\mathbf{U}_{u,z}(X)\right) \\
&\quad = 2\lambda + \mathbf{R}(X,z), \qquad (24)
\end{aligned}
$$

where (23) uses the fact that given $\mathcal{F}$ the values of $\mathbf{U}_{u,z}(X)$, for $u \in N(z)\setminus\{v\}$ are fully determined. Similarly, we get the lower bound:

$$
\begin{aligned}
\mathbb{E}\left[\mathbf{U}_{z,v}(X) \mid \mathcal{F}\right] &= \mathbb{E}\left[\mathbf{U}_{z,v}(X) \mid \mathcal{F}, z \in X\right] \cdot \mathbf{Pr}\left[z \in X | \mathcal{F}\right] \\
&\quad + \mathbb{E}\left[\mathbf{U}_{z,v}(X) \mid \mathcal{F}, z \notin X\right] \cdot \mathbf{Pr}\left[z \notin X | \mathcal{F}\right] \\
&\ge (1-2\lambda)\mathbb{E}\left[\mathbf{U}_{z,v}(X) \mid \mathcal{F}, z \notin X\right] \\
&\ge (1-2\lambda) \prod_{u \in N(z)\setminus\{w\}} \left(1 - \frac{\lambda}{1+\lambda}\mathbf{U}_{u,z}(X)\right) \\
&\ge (1-2\lambda) \prod_{u \in N(z)} \left(1 - \frac{\lambda}{1+\lambda}\mathbf{U}_{u,z}(X)\right) \\
&= (1-2\lambda)\mathbf{R}(X,z) \\
&\ge \mathbf{R}(X,z) - 2\lambda, \qquad (25)
\end{aligned}
$$

where in the last inequality we use the fact that $\mathbf{R}(X,z) \le 1$.

From (24) and (25) we have proven (21), which completes the proof of the lemma. □

5 Sketch of Rapid Mixing Proof

Theorem 8 tells us that after a burn-in period the Glauber dynamics locally behaves like the BP fixpoints ω^* with high probability (whp). (In this discussion, we use the term whp to refer to events that occur with probability $\ge 1 - \exp(-\Omega(\Delta))$.) Meanwhile Theorem 6 says that there is an appropriate distance function Φ for which path coupling has contraction for pairs of states that behave as in ω^*. The snag in simply combining this pair of results and deducing rapid mixing is that when Δ is constant then there is still a constant fraction of the graph that does not behave like ω^*, and our disagreements in our coupling proof may be biased towards this set. We follow the approach in [4] to overcome this obstacle and complete the proof of Theorem 1. We give a brief sketch of the approach, the details are contained in Section E of the appendix.

The burn-in period for Theorem 8 to apply is $O(n \log \Delta)$ steps from the worst-case initial configuration X_0. In fact, for a "typical" initial configuration only $O(n)$ steps are required as we only need to update $\geq 1-\epsilon$ fraction of the neighbors of every vertex in the local neighborhood of the specified vertex v. The "bad" initial configurations are ones where almost all of the neighbors of v (or many of its grandchildren) are occupied. We call such configurations "heavy" (see Section C.2 of the appendix). We first prove that after $O(n \log \Delta)$ steps a chain is not-heavy in the local neighborhood of v, and this property persists whp (see Lemma 22 in the appendix). Then, only $O(n)$ steps are required for the burn-in period (see Theorem 27 in Section D of the appendix).

Our argument has two stages. We start with a pair of chains X_0, Y_0 that differ at a single vertex v. In the first stage we burn-in for $T_b = O(n \log \Delta)$ steps. After this burn-in period, we have the following properties whp: every vertex in the local neighborhood of v is not-heavy, the number of disagreements is $\leq \text{poly}(\Delta)$, and the disagreements are all in the local neighborhood of v (see Lemma 31, parts 2 and 4, in Section E of the appendix).

In the second stage we have sets of epochs of length $T = O(n)$ steps. For the pair of chains X_{T_b}, Y_{T_b} we apply path coupling again. Now we consider a pair of chains that differ at one vertex z which is not heavy. We look again at the local neighborhood of z (in this case, that means all vertices within distance $\leq \sqrt{\Delta}$ of z). After T steps, whp every vertex in the local neighborhood has the local uniformity properties and the disagreements are contained in this local neighborhood. Then we have contraction in the path coupling condition (by applying Theorem 6), and hence after $O(n)$ further steps the expected Hamming distance is small (see Lemma 32 in the appendix). Combining a sequence of these $O(n)$ length epochs we get that the original pair has is likely to have coupled and we can deduce rapid mixing.

6 Conclusions

The work of Weitz [41] was a notable accomplishment in the field of approximate counting/sampling. However a limitation of his approach is that the running time depends exponentially on $\log \Delta$. It is widely believed that the Glauber dynamics has mixing time $O(n \log n)$ for all G of maximum degree Δ when $\lambda < \lambda_c(\Delta)$. However, until now there was little theoretical work to support this conjecture. We give the first such results which analyze the widely used algorithmic approaches of MCMC and loopy BP.

One appealing feature of our work is that it directly ties together with Weitz's approach: Weitz uses decay of correlations on trees to truncate his self-avoiding walk tree, whereas we use decay of correlations to deduce a contracting metric for the path coupling analysis, at least when the chains are at the BP fixed point. We believe this technique of utilizing the principal eigenvector for the BP operator for the path coupling metric will apply to a general class of spin systems, such as 2-spin antiferromagnetic spin systems (Weitz's algorithm was extended to this class [19]).

We hope that in the future more refined analysis of the local uniformity properties will lead to relaxed girth assumptions. However dealing with very short cycles, such as triangles, will require a new approach since loopy BP no longer seems to be a good estimator of the Gibbs distribution for certain examples.

References

[1] R. Bubley and M. Dyer. Path coupling, Dobrushin uniqueness, and approximate counting. In *Proceedings of the 38th Annual IEEE Symposium on Foundations of Computer Science (FOCS)*, 223–231, 1997.

[2] V. Chandrasekaran, M. Chertkov, D. Gamarnik, D. Shah, and J. Shin. Counting Independent Sets using the Bethe Approximation. *SIAM J. Discrete Math.*, 25(2):1012-1034, 2011.

[3] D. Dubhashi and D. Ranjan. Balls and bins: a study in negative dependence. *Random Structures and Algorithms*, 13:99-124, 1998.

[4] M. E. Dyer, A. M. Frieze, T. P. Hayes, and E. Vigoda. Randomly Coloring Constant Degree Graphs. *Random Structures and Algorithms*, 43(2):181-200, 2013.

[5] M. Dyer and C. Greenhill On Markov chains for independent sets. *J. Algorithms* 35(1):17–49, 2000.

[6] A. Galanis, Q. Ge, D. Štefankovič, E. Vigoda, and L. Yang. Improved inapproximability results for counting independent sets in the hard-core model. *Random Structures and Algorithms*, 45(1):78-110, 2014.

[7] A. Galanis, D. Štefankovič, and E. Vigoda. Inapproximability of the partition function for the antiferromagnetic Ising and hard-core models. To appear in *Combinatorics, Probability & Computing*, 2016.

[8] A. Galanis, D. Štefankovič, and E. Vigoda. Inapproximability for Antiferromagnetic Spin Systems in the Tree Non-Uniqueness Region. *Journal of the ACM*, 62(6):50, 2015.

[9] C. Greenhill. The complexity of counting colourings and independent sets in sparse graphs and hypergraphs. *Computational Complexity*, 9(1):52–72, 2000.

[10] T. P. Hayes. Local uniformity properties for Glauber dynamics on graph colorings. *Random Structures and Algorithms*, 43(2):139-180, 2013.

[11] T. P. Hayes. A simple condition implying rapid mixing of single-site dynamics on spin systems. In *47th Annual IEEE Symposium on Foundations of Computer Science* (FOCS), pages 39-46, 2006.

[12] T. P. Hayes, J. C. Vera, and E. Vigoda. Randomly coloring planar graphs with fewer colors than the maximum degree. *Random Struct. Algorithms* 47(4): 731-759, 2015

[13] T. Hayes and E. Vigoda. Coupling with the Stationary Distribution and Improved Sampling for Colorings and Independent Sets. *Annals of Applied Probability*, 16(4):1297-1318, 2006.

[14] T. Heskes. On the uniqueness of loopy belief propagation fixed points. *Neural Computation*, 16(11):2379-2413, 2004.

[15] M. Huber. Approximation algorithms for the normalizing constant of Gibbs distributions. *Annals of Applied Probability*, 25(2):974-985, 2015.

[16] A. T. Ihler, J. W. Fisher III, and A. S. Willsky. Loopy Belief Propagation: Convergence and Effects of Message Errors. *Journal of Machine Learning Research*, 6:905-936, 2005.

[17] F. P. Kelly. Stochastic Models of Computer Communication Systems. *Journal of the Royal Statistical Society. Series B (Methodological)*, 47(3):379-395, 1985.

[18] L. Li, P. Lu, and Y. Yin. Approximate counting via correlation decay in spin systems. In *Proceedings of the 23rd Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pages 922-940, 2012.

[19] L. Li, P. Lu, and Y. Yin. Correlation Decay up to Uniqueness in Spin Systems. In *Proceedings of the 24th Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pages 67-84, 2013.

[20] C. Lin, J. Liu, and P. Lu. A Simple FPTAS for Counting Edge Covers. In *Proceedings of the 25th Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pages 341-348, 2014.

[21] J. Liu and P. Lu. FPTAS for #BIS with Degree Bounds on One Side. In *Proceedings of the 47th Annual ACM Symposium on Theory of Computing (STOC)*, pages 549-556, 2015.

[22] M. Luby and E. Vigoda. Fast Convergence of the Glauber Dynamics for Sampling Independent Sets. *Random Structures and Algorithms*, 15(3-4):229-241, 1999.

[23] M. Mitzenmacher and E. Upfal. Probability and computing: Randomized algorithms and probabilistic analysis, Cambridge University Press, 2005.

[24] E. Mossel and A. Sly. Exact thresholds for Ising-Gibbs samplers on general graphs. *The Annals of Probability*, 41(1):294-328, 2013.

[25] E. Mossel, D. Weitz, and N. Wormald. On the hardness of sampling independent sets beyond the tree threshold. *Probability Theory and Related Fields*, 143(3-4):401-439, 2009.

[26] K. P. Murphy, Y. Weiss, and M. I. Jordan. Loopy belief propagation for approximate inference: an empirical study. In *Proceedings of the 15th Conference on Uncertainty in Artificial Intelligence (UAI)*, pages 467-475, 1999.

[27] J. Pearl. Reverend Bayes on inference engines: A distributed hierarchical approach. In *Proceedings of the 2nd National Conference on Artificial Intelligence* (AAAI), pages 133-136, 1982.

[28] 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(1-2):75-99, 2013.

[29] J. Shin. The Complexity of Approximating a Bethe Equilibrium. *IEEE Transactions on Information Theory*, 60(7): 3959-3969, 2014.

[30] A. Sinclair, P. Srivastava, D. Štefankovič, and Y. Yin. Spatial mixing and the connective constant: Optimal bounds. In *Proceedings of the 26th Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pages 1549-1563, 2015.

[31] A. Sinclair, P. Srivastava, and M. Thurley. Approximation algorithms for two-state anti-ferromagnetic spin systems on bounded degree graphs. In *Proceedings of the Twenty-third Annual ACM-SIAM Symposium on Discrete Algorithms* (SODA), 2012.

[32] A. Sinclair, P. Srivastava, and Y. Yin. Spatial Mixing and Approximation Algorithms for Graphs with Bounded Connective Constant. In *Proceedings of the 54th Annual IEEE Symposium on Foundations of Computer Science (FOCS)*, pages 300-309, 2013.

[33] A. Sly. Computational Transition at the Uniqueness Threshold. In *Proceedings of the 51st Annual IEEE Symposium on Foundations of Computer Science (FOCS)*, pages 287-296, 2010.

[34] A. Sly and N. Sun. The computational hardness of counting in two-spin models on d-regular graphs. *The Annals of Probability*, 42(6):2383-2416, 2014.

[35] D. Štefankovič, S. Vempala, and E. Vigoda. Adaptive Simulated Annealing: A Near-optimal Connection between Sampling and Counting. *Journal of the ACM*, 56(3):1-36, 2009.

[36] S. Tatikonda and M. Jordan. Loopy belief propagation and Gibbs measures. In *Proceedings of the 18th Conference on Uncertainty in Artificial Intelligence* (UAI), pages 493-500, 2002.

[37] L. G. Valiant. The Complexity of Enumeration and Reliability Problems. *SIAM Journal on Computing*, 8(3):410-421, 1979.

[38] J. C. Vera, E. Vigoda, and L. Yang. Improved Bounds on the Phase Transition for the Hard-Core Model in 2-Dimensions. *SIAM Journal on Discrete Mathematics*, 29(4):1895-1915, 2015.

[39] E. Vigoda. A Note on the Glauber Dynamics for Sampling Independent Sets. *Electronic Journal of Combinatorics*, 8(1): Research paper 8, 2001.

[40] Y. Weiss. Correctness of Local Probability Propagation in Graphical Models with Loops. *Neural Computation*, 12(1):1-41, 2000.

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

[42] N. C. Wormald. The asymptotic distribution of short cycles in random regular graphs. *J. Combinatorial Theory (Series B)*, 31:168-182, 1981.

A BP convergence: Missing proofs in Sections 2 and 3

In this section we prove Lemma 4 about the convergence of recurrence F defined in (2) to a unique fixed point ω^*, Lemma 5 about the contraction of the error at every step with respect to a potential function, and Theorem 6 about the existence of a suitable distance function Φ for path coupling.

The next theorem unifies these key results regarding the convergence of F defined in (2).

Theorem 10. *For all $\delta > 0$, there exists $\Delta_0 = \Delta_0(\delta)$, for all $G = (V, E)$ of maximum degree $\Delta \geq \Delta_0$, all $\lambda < (1-\delta)\lambda_c(\Delta)$, the following hold:*

1. *For any $x_1, x_2 \in [(1+\lambda)^{-\Delta}, 1]$,*
$$\frac{1}{3}|x_1 - x_2| \leq |\Psi(x_1) - \Psi(x_2)| \leq 3|x_1 - x_2|. \tag{26}$$
2. *(Lemma 4) The function F defined in (2) has a unique fixed point ω^*. Moreover, for any initial value $\omega^0 \in [0,1]^V$, denoting by $\omega^i = F^i(\omega)$ the vector after the i-th iterate of F, it holds that*
$$\|\omega^i - \omega^*\|_\infty \leq 3(1-\delta/6)^i.$$
3. *(Lemma 5) for any $\omega \in [0,1]^V$, $v \in V$ and $R \geq 1$, we have:*
$$D_{v,R-1}(F(\omega), \omega^*) \leq (1-\delta/6) D_{v,R}(\omega, \omega^*),$$
where $D_{v,R}$ is as defined in (5).
4. *(Theorem 6) There exist $\Phi : V \to \mathbb{R}_{\geq 0}$ such that for every $v \in V$, $1 \leq \Phi(v) \leq 12$, and*
$$(1-\delta/6)\Phi(v) \geq \sum_{u \in N(v)} \frac{\lambda \omega^*(u)}{1 + \lambda \omega^*(u)} \Phi(u).$$

In part 4 the astute reader may notice that we are considering BP without a parent, and hence each vertex depends on Δ neighbors. Consequently parts of our analysis will consider the tree with branching factor Δ. This is not essential in our proof, but it allows us to consider slightly simpler recurrences. In our setting we have Δ sufficiently large and since $\lambda_c(\Delta) = O(1/\Delta)$ and hence this simplification has no effect on the final result that we prove.

We first analyze the uniqueness regime described in the above Theorem 10.

Let $f_{\lambda,d}(x) = (1+\lambda x)^{-d}$ be the symmetric version of the BP recurrence (2). Let $\hat{x} = \hat{x}(\lambda, d)$ be the unique fixed point of $f_{\lambda,d}(x)$, satisfying $\hat{x}(\lambda, d) = (1 + \lambda \hat{x}(\lambda, d))^{-d}$. We define

$$\alpha(\lambda, d) = \sqrt{\frac{d \cdot \lambda \hat{x}(\lambda, d)}{1 + \lambda \hat{x}(\lambda, d)}}. \tag{27}$$

Proposition 11. *For all $\delta > 0$, there exists $\Delta_0 = \Delta_0(\delta)$, for all $\Delta \geq \Delta_0$, all $\lambda < (1-\delta)\lambda_c(\Delta)$ where $\lambda_c(\Delta) = \frac{(\Delta-1)^{\Delta-1}}{(\Delta-2)^{\Delta}}$, it holds that $\alpha(\lambda, \Delta) \leq 1 - \delta/6$.*

Proof. Let $x_0 = \frac{1-\delta/3}{\lambda(\Delta-1+\delta/3)}$. It is easy to verify that

$$\sqrt{\frac{\Delta \cdot \lambda x_0}{1 + \lambda x_0}} \leq 1 - \delta/6.$$

Note that the function $\sqrt{\frac{\Delta\lambda x}{1+\lambda x}}$ is increasing in x. Since $f(x)$ is increasing in λ, it is easy to verify that $\hat{x}(\lambda, d)$ is increasing in λ. We then show that for all $\Delta \geq \Delta_0$, it holds that $\hat{x}(\lambda_0, \Delta) \leq x_0$ where $\lambda_0 = (1-\delta)\lambda_c(\Delta) = \frac{(1-\delta)(\Delta-1)^{\Delta-1}}{(\Delta-2)^{\Delta}}$, which will prove our proposition.

Since $f_{\lambda_0,\Delta}(x)$ is decreasing in x and $f_{\lambda_0,\Delta}(\hat{x}(\lambda_0, \Delta)) = \hat{x}(\lambda_0, \Delta)$, it is sufficient to show that

$$f_{\lambda_0,\Delta}(x_0) = (1 + \lambda_0 x_0)^{-\Delta} \leq x_0.$$

Note that it holds that

$$\frac{f_{\lambda_0,\Delta}(x_0)}{x_0} = \frac{\lambda_0(\Delta - 1 + \delta/3)}{(1-\delta/3)(1 + \frac{1-\delta/3}{(\Delta-1+\delta/3)})^{\Delta}} = \frac{1-\delta}{1-\delta/3} \cdot \frac{(\Delta-1)^{\Delta}(\Delta-1+\delta/3)^{\Delta}}{(\Delta-2)^{\Delta}\Delta^{\Delta}} \cdot \frac{\Delta-1+\delta/3}{\Delta-1}.$$

Therefore, there is a suitable $\Delta_0 = O(\frac{1}{\delta})$ such that for all $\Delta \geq \Delta_0$,

$$\frac{f_{\lambda_0,\Delta}(x_0)}{x_0} \leq \frac{1-\delta}{1-\delta/3}\left(1 + O\left(\frac{\eta}{\Delta}\right)\right) \mathrm{e}^{\delta/2.99} < 1,$$

which proves the proposition. $\square$

Recall recurrence F as defined in (2). The following proposition was proved implicitly in [19].

Proposition 12 ([19])**.** *Let $G = (V, E)$ be a graph with maximum degree at most Δ. Assume that $\alpha(\lambda, \Delta) \leq 1$. For any $\omega \in [0, 1]^V$, and $v \in V$,*

$$\sqrt{\frac{\lambda F(\omega)(v)}{1 + \lambda F(\omega)(v)}} \sum_{u \in N(v)} \sqrt{\frac{\lambda\omega(u)}{1 + \lambda\omega(u)}} \leq \alpha(\lambda, \Delta),$$

where $\alpha(\lambda, \Delta)$ is defined in (27).

Proof. Let $\bar{\omega} \in [0, 1]$ be that satisfies $1 + \lambda\bar{\omega} = \left(\prod_{u \in N(v)}(1 + \lambda\omega(u))\right)^{\frac{1}{|N(v)|}}$. Denote that $\bar{\nu} = \ln(1 + \lambda\bar{\omega})$ and $\nu(u) = \ln(1 + \lambda\omega(u))$. It then holds that $\bar{\nu} = \frac{1}{|N(v)|}\sum_{u \in N(v)} \nu(u)$. Due to the concavity of $\sqrt{\frac{\mathrm{e}^{\nu}-1}{\mathrm{e}^{\nu}}}$ in ν, by Jensen's inequality:

$$\frac{1}{|N(v)|} \sum_{u \in N(v)} \sqrt{\frac{\lambda\omega(u)}{1 + \lambda\omega(u)}} = \frac{1}{|N(v)|} \sum_{u \in N(v)} \sqrt{\frac{\mathrm{e}^{\nu(u)} - 1}{\mathrm{e}^{\nu(u)}}} \leq \sqrt{\frac{\mathrm{e}^{\bar{\nu}} - 1}{\mathrm{e}^{\bar{\nu}}}} = \sqrt{\frac{\lambda\bar{\omega}}{1 + \lambda\bar{\omega}}}.$$

Therefore,

$$\sqrt{\frac{\lambda F(\omega)(v)}{1 + \lambda F(\omega)(v)}} \sum_{u \in N(v)} \sqrt{\frac{\lambda\omega(u)}{1 + \lambda\omega(u)}} \leq \sqrt{\frac{\lambda d f(\bar{\omega})}{1 + \lambda f(\bar{\omega})} \cdot \frac{\lambda d\bar{\omega}}{1 + \lambda\bar{\omega}}},$$

where $d = |N(v)|$ is the degree of vertex v in G and $f(\bar{\omega}) = (1 + \lambda\bar{\omega})^{-d}$ is the symmetric version of the recursion (2).

Define $\alpha_{\lambda,d}(x) = \sqrt{\frac{\lambda d f(x)}{1+\lambda f(x)} \cdot \frac{\lambda d x}{1+\lambda x}}$ where as before $f(x) = (1 + \lambda x)^{-d}$. The above convexity argument shows that

$$\sqrt{\frac{\lambda F(\omega)(v)}{1 + \lambda F(\omega)(v)}} \sum_{u \in N(v)} \sqrt{\frac{\lambda\omega(u)}{1 + \lambda\omega(u)}} \leq \alpha_{\lambda,d}(x), \text{ for some } x \in [0, 1]. \tag{28}$$

Fixed any λ and d, the critical point of $\alpha_{\lambda,d}(x)$ is achieved at the unique positive $x(\lambda, d)$ satisfying

$$\lambda d x(\lambda, d) = 1 + \lambda f(x(\lambda, d)). \tag{29}$$

It is also easy to verify by checking the derivative $\frac{\mathrm{d}\alpha_{\lambda,d}(x)}{\mathrm{d}x}$ that the maximum of $\alpha_{\lambda,d}(x)$ is achieved at this critical point $x(\lambda, d)$.

Recall that $\hat{x}(\lambda, d)$ is the fixed point satisfying $\hat{x}(\lambda, d) = f(\hat{x}(\lambda, d)) = (1 + \lambda\hat{x}(\lambda, d))^{-d}$, and $\alpha(\lambda, d) = \sqrt{\frac{\lambda d\hat{x}(\lambda,d)}{1+\lambda\hat{x}(\lambda,d)}}$. Under the assumption that $\alpha(\lambda, d) \leq 1$, we must have $\hat{x}(\lambda, d) \leq x(\lambda, d)$. If otherwise $\hat{x}(\lambda, d) > x(\lambda, d)$, then we would have $\lambda d\hat{x}(\lambda, d) > \lambda d x(\lambda, d) = 1 + \lambda f(x(\lambda, d)) > 1 + \lambda f(\hat{x}(\lambda, d)) = 1 + \lambda\hat{x}(\lambda, d)$, contradicting that $\frac{\lambda d\hat{x}(\lambda,d)}{1+\lambda\hat{x}(\lambda,d)} = \alpha(\lambda, d)^2 \leq 1$. Therefore, for any $x \in [0, 1]$, it holds that

$$\begin{aligned}
\alpha_{\lambda,d}(x) &\leq \alpha(d, x(\lambda, d)) \\
&= \sqrt{\frac{\lambda d f(x(\lambda, d))}{1 + \lambda f(x(\lambda, d))} \cdot \frac{\lambda d x(\lambda, d)}{1 + \lambda x(\lambda, d)}} \\
&= \sqrt{\frac{\lambda d f(x(\lambda, d))}{1 + \lambda x(\lambda, d)}} && \text{(due to (29))} \\
&\leq \sqrt{\frac{\lambda d f(\hat{x}(\lambda, d))}{1 + \lambda\hat{x}(\lambda, d)}} && (\hat{x}(\lambda, d) \leq x(\lambda, d)) \\
&= \sqrt{\frac{\lambda d\hat{x}(\lambda, d)}{1 + \lambda\hat{x}(\lambda, d)}} \\
&= \alpha(\lambda, d).
\end{aligned}$$

Finally, it is easy to observe that $\alpha(\lambda, d)$ is increasing in d since $\alpha(\lambda, d)$ is increasing in $\hat{x}(\lambda, d)$ and $\hat{x}(\lambda, d)$ is increasing in d. Therefore, $\alpha(\lambda, d) \leq \alpha(\lambda, \varDelta)$ because $d = |N(v)| \leq \varDelta$. Combined this with (28), the proposition is proved. ☐

We are now ready to prove Theorem 10

Proof of Theorem 10. By Proposition 11, for the regime of λ described in the theorem, it holds that $\alpha(\lambda, \varDelta) < 1 - \delta/6$ where $\varDelta$ is the maximum degree of the graph $G = (V, E)$.

Recall that in Section 2, we introduce the following potential function:

$$\varPsi(x) = (\sqrt{\lambda})^{-1}\mathrm{arcsinh}\left(\sqrt{\lambda \cdot x}\right).$$

We then show that for any $\omega_1, \omega_2 \in [0, 1]^V$, and $v \in V$,

$$|\varPsi(\omega_1(v)) - \varPsi(\omega_2(v))| \leq 1, \tag{30}$$

and

$$|\varPsi(F(\omega_1)(v)) - \varPsi(F(\omega_2)(v))| \leq (1 - \delta/6) \max_{u \in N(v)} |\varPsi(\omega_1(v)) - \varPsi(\omega_2(v))|. \tag{31}$$

We first prove (30). It is easy to see that $\varPsi(x)$ is monotonically increasing for $x \in [0, 1]$, thus $|\varPsi(\omega_1(v)) - \varPsi(\omega_2(v))| \leq \varPsi(1) - \varPsi(0) = \mathrm{arcsinh}(\sqrt{\lambda})/\sqrt{\lambda}$. Observe that $\mathrm{arcsinh}(x) \leq x$ for any $x \geq 0$ and hence $\mathrm{arcsinh}(\sqrt{\lambda})/\sqrt{\lambda} \leq 1$. This proves (30).

We then prove (31). Note that the derivative of the potential function Ψ is $\Psi'(x) = \frac{\mathrm{d}\Psi(x)}{\mathrm{d}x} = \frac{1}{2\sqrt{x(1+\lambda x)}}$. Due to the mean value theorem, there exists an $\tilde{\omega} \in [0,1]^{N(v)}$ such that

$$\begin{aligned}|\Psi(F(\omega_1)(v)) - \Psi(F(\omega_2)(v))| &= \sum_{u \in N(v)} \left| \frac{\partial F(\omega)(v)}{\partial \omega(u)} \right|_{\omega=\tilde{\omega}} \frac{\Psi'(F(\tilde{\omega})(v))}{\Psi'(\tilde{\omega}(u))} \left| |\Psi(\omega_1(u)) - \Psi(\omega_2(u))| \right. \\ &= \sqrt{\frac{\lambda F(\tilde{\omega})(v)}{1+\lambda F(\tilde{\omega})(v)}} \sum_{u \in N(v)} \sqrt{\frac{\lambda \tilde{\omega}(u)}{1+\lambda \tilde{\omega}(u)}} |\Psi(\omega_1(u)) - \Psi(\omega_2(u))| \\ &\leq \left(\sqrt{\frac{\lambda F(\tilde{\omega})(v)}{1+\lambda F(\tilde{\omega})(v)}} \sum_{u \in N(v)} \sqrt{\frac{\lambda \tilde{\omega}(u)}{1+\lambda \tilde{\omega}(u)}} \right) \cdot \max_{u \in N(v)} |\Psi(\omega_1(u)) - \Psi(\omega_2(u))|.\end{aligned}$$

Then (31) is implied by Proposition 11 and Proposition 12.

Next, we prove the statements in the theorem.

1. (Proof of Equation (26)) By the mean value theorem, for any $x_1, x_2 \in [(1+\lambda)^{-\Delta}, 1]$, there exists a mean value $\xi \in [(1+\lambda)^{-\Delta}, 1]$ such that

   $$|\Psi(x_1) - \Psi(x_2)| = \Psi'(\xi)|x_1 - x_2| = \frac{1}{2\sqrt{\xi(1+\lambda\xi)}} |x_1 - x_2|.$$

   For all sufficiently large Δ, it holds that $(1+\lambda)^{-\Delta} > 1/36$ and $\lambda < \lambda_c(\Delta) \leq 0.25$, thus $\xi \in [1/36, 1]$. Therefore, $\frac{1}{2\sqrt{\xi(1+\lambda\xi)}} \geq \frac{1}{2\sqrt{1+\lambda}} > \frac{1}{3}$ and $\frac{1}{2\sqrt{\xi(1+\lambda\xi)}} < \frac{1}{2\sqrt{\xi}} < 3$.

2. (Proof of Lemma 4) Consider the dynamical system defined by $\omega^{(i)} = F(\omega^{(i-1)})$ with arbitrary two initial values $\omega_1^{(0)}, \omega_2^{(0)} \in [0,1]^V$. The derivative of the potential function satisfies that $\Psi'(x) \geq \frac{1}{2\sqrt{1+\lambda}}$ for any $x \in [0,1]$. Due to the mean value theorem, for any $v \in V$, there exists a mean value $\xi \in [0,1]$ such that

   $$\left| \omega_1^{(i)}(v) - \omega_2^{(i)}(v) \right| = \frac{1}{\Psi'(\xi)} \left| \Psi\left(\omega_1^{(i)}(v)\right) - \Psi\left(\omega_2^{(i)}(v)\right) \right| \leq 2\sqrt{1+\lambda} \left| \Psi\left(\omega_1^{(i)}(v)\right) - \Psi\left(\omega_2^{(i)}(v)\right) \right|.$$

   Combined with (30) and (31), we have

   $$\begin{aligned}\left\| \omega_1^{(i)} - \omega_2^{(i)} \right\|_\infty &\leq 2\sqrt{1+\lambda} \left\| \Psi\left(\omega_1^{(i)}\right) - \Psi\left(\omega_2^{(i)}\right) \right\|_\infty \\ &\leq 2(1-\delta/6)^i \sqrt{1+\lambda} \max_{z \in V} \left| \Psi\left(\omega_1^{(0)}(z)\right) - \Psi\left(\omega_2^{(0)}(z)\right) \right| \\ &\leq 2(1-\delta/6)^i \sqrt{1+\lambda},\end{aligned}$$

   which is at most $3(1-\delta/6)^i$ for $\lambda < \lambda_c(\Delta)$ for all sufficiently large Δ.

   Therefore, $\left\| \omega_1^{(i)} - \omega_2^{(i)} \right\|_\infty \to 0$ as $i \to \infty$ for arbitrary initial values $\omega_1^{(0)}, \omega_2^{(0)} \in [0,1]^V$. This shows that the F defined in (2) has a unique fixed point ω^*.

3. (Proof of Lemma 5) According to the definition of $D_{v,R}$ in (5),

$$\begin{aligned} D_{v,R-1}(F(\omega),\omega^*) &= \max_{u\in B(v,R-1)} |\Psi(F(\omega)(u)) - \Psi(\omega^*(u))| \\ &= \max_{u\in B(v,R-1)} |\Psi(F(\omega)(u)) - \Psi(F(\omega^*)(u))| && (\omega^* \text{ is fixed point}) \\ &\le \max_{u\in B(v,R-1)} (1-\delta/6) \max_{z\in N(u)} |\Psi(\omega(z)) - \Psi(\omega^*(z))| && (\text{due to } (31)) \\ &= (1-\delta/6) \max_{u\in B(v,R)} |\Psi(\omega(u)) - \Psi(\omega^*(u))| \\ &= (1-\delta/6)\cdot D_{v,R}(\omega,\omega^*). \end{aligned}$$

4. (Proof of Theorem 6) Due to Propositions 11 and 12, for any $\omega \in [0,1]^V$, and $v \in V$,

$$\sqrt{\frac{\lambda F(\omega)(v)}{1+\lambda F(\omega)(v)}} \sum_{u\in N(v)} \sqrt{\frac{\lambda\omega(u)}{1+\lambda\omega(u)}} \le 1-\delta/6.$$

In particular, this inequality holds for the fixed point ω^* where $F(\omega^*)(v) = \omega^*(v)$. Therefore,

$$\sqrt{\frac{\lambda\omega^*(v)}{1+\lambda\omega^*(v)}} \sum_{u\in N(v)} \sqrt{\frac{\lambda\omega^*(u)}{1+\lambda\omega^*(u)}} \le 1-\delta/6.$$

We construct $\Phi : V \to \mathbb{R}_{\ge 0}$ as $\Phi(v) = \sqrt{\frac{1+\lambda\omega^*(v)}{\omega^*(v)}}$ for every $v \in V$. Then

$$\sum_{u\in N(v)} \frac{\lambda\omega^*(u)}{1+\lambda\omega^*(u)}\Phi(u) \le (1-\delta/6)\Phi(v).$$

We now show that $1 \le \Phi(v) \le 12$. Since $\omega^* \in [0,1]^V$, we have $\Phi(v) = \sqrt{\frac{1+\lambda\omega^*(v)}{\omega^*(v)}} \ge 1$. Meanwhile, it holds that $\omega^*(v) = \prod_{u\in N_v} \frac{1}{1+\lambda\omega^*(u)} \ge (1+\lambda)^{-\Delta}$. By our assumption, $\lambda \le (1-\delta)\lambda_c(\Delta) \le \frac{4}{\Delta-2}$ for all $\Delta \ge 3$. Therefore, $\omega^*(v) \ge (1+\frac{4}{\Delta-2})^{-\Delta} \ge 5^{-3}$ and $\Phi(v) = \sqrt{\frac{1+\lambda\omega^*(v)}{\omega^*(v)}} \le \sqrt{5^3+4} \le 12$.

□

We consider a recurrence which corresponds to the rooted belief propagation. For an undirected graph $G = (V,E)$, let $\bar{E}$ be the set of all orientations of edges in E. The function $H : [0,1]^{\bar{E}} \to [0,1]^{\bar{E}}$ is defined as follows: For any $\omega \in [0,1]^{\bar{E}}$ and $(v,p) \in \bar{E}$,

$$H(\omega)(v,p) = \prod_{u\in N(v)\setminus\{p\}} \frac{1}{1+\lambda\omega(u,v)} \tag{32}$$

With the approach used in the proof of Theorem 10, analyzing the convergence of H is the same as analyzing F on a graph G with maximum degree $\Delta-1$. Recall that $\alpha(\lambda,\Delta)$ is increasing in Δ. The same proof as of Theorem 10 gives us the following corollary.

Corollary 13. *For $G = (V,E)$ and λ assumed by Theorem 10, the function H defined in (32) has a unique fixed point ω^*. Moreover, for any initial value $\omega^0 \in [0,1]^{\bar{E}}$, denoting by $\omega^i = F^i(\omega)$ the vector after the i-th iterate of F, it holds that*

$$\|\omega^i - \omega^*\|_\infty \le 3(1-\delta/6)^i.$$

B Loopy BP: Proof of Theorem 3

Consider the version of Loopy BP defined with the following sequence of messages: For all $t \geq 1$, for $v \in V$:

$$\tilde{R}_v^t = \lambda \prod_{w \in N(v)} \frac{1}{1+\tilde{R}_w^{t-1}}. \tag{33}$$

The system of equations specified by (33) is equivalent to the one in (2) in the following sense: Given any set of initial messages $(\tilde{R}_v^0)_{v\in V} \in \mathbb{R}_{\geq 0}$, it holds that $\tilde{R}_v^t = \lambda F^t(\bar{\omega})(v)$, for appropriate $\bar{\omega}$ which depends on the initial messages, i.e. $(\tilde{R}_v^0)_{v\in V}$. F^t is the t-th iteration of the function F.

Of interest is in the quantity $q^t(v)$, $v \in V$, defined as follows:

$$\tilde{q}^t(v) = \frac{\tilde{R}_v^t}{1+\tilde{R}_v^t}.$$

From Lemma 4, there exists $\tilde{q}^* \in [0,1]^V$ such that $\tilde{q}^t$ converges to $\tilde{q}^*$ as $t \to \infty$, in the sense that $\tilde{q}^t/\tilde{q}^* \to 1$. It is elementary to show that the following holds for any $t > 0$, any $p \in V$ and $v \in N(p)$:

$$\frac{q^t(v,p)}{\mu(v \text{ is occupied} \mid p \text{ is unoccupied})} = \frac{q^t(v,p)}{q^*(v,p)} \frac{q^*(v,p)}{\tilde{q}^*(v)} \cdot \frac{\tilde{q}^*(v)}{\mu(v \text{ is occupied})} \cdot \frac{\mu(v \text{ is occupied})}{\mu(v \text{ is occupied} \mid p \text{ is unoccupied})}.$$

The theorem follows by showing that each of the four ratios on the r.h.s. are sufficiently close to 1. For the first two ratios we use Theorem 14, and for the third one we use the Lemma 15.

Theorem 14. *For all $\delta, \epsilon > 0$, there exists $\Delta_0 = \Delta_0(\delta,\epsilon)$ and $C = C(\delta,\epsilon)$, such that for all $\Delta \geq \Delta_0$, all $\lambda < (1-\delta)\lambda_c(\Delta)$, all graphs G of maximum degree Δ and girth ≥ 6, all $\epsilon > 0$ the following holds:*

There exists $q^* \in [0,1]^E$ such that for $t \geq C$, for all $p \in V$, $v \in N(p)$ we have that

$$\left|\frac{q^t(v,p)}{q^*(v,p)} - 1\right| \leq \epsilon \quad \text{and} \quad \left|\frac{q^*(v,p)}{\tilde{q}^*(v)} - 1\right| \leq \epsilon, \tag{34}$$

where $q^t(v,p)$ is defined in (1).

The proof of Theorem 14 appears in Section B.1.

Lemma 15. *For all $\delta, \epsilon > 0$, there exists $\Delta_0 = \Delta_0(\delta,\epsilon)$ and $C = C(\delta,\epsilon)$, such that for all $\Delta \geq \Delta_0$, all $\lambda < (1-\delta)\lambda_c(\Delta)$, all graphs G of maximum degree Δ and girth ≥ 6, the following holds: Let $\mu(\cdot)$ be the Gibbs distribution, for all $v \in V$ we have*

$$\left|\frac{\tilde{q}^*(v)}{\mu(v \textit{ is occupied})} - 1\right| \leq \epsilon.$$

The proof of Lemma 15 appears in Section B.2.

The theorem follows by showing that

$$\left|\frac{\mu(v \text{ is occupied})}{\mu(v \text{ is occupied} \mid p \text{ is unoccupied})} - 1\right| \leq 10/\Delta.$$

From Bayes' rule we get that $\mu(v \text{ is occupied} \mid p \text{ is unoccupied}) = \frac{\mu(v \text{ is occupied})}{\mu(p \text{ is unoccupied})}$. Using this observation we get that

$$\left|\frac{\mu(v \text{ is occupied})}{\mu(v \text{ is occupied} \mid p \text{ is unoccupied})} - 1\right| = |\mu(p \text{ is unoccupied}) - 1| \leq 10/\Delta.$$

In the last inequality we use the fact that $0 \leq \mu(p \text{ is occupied}) \leq \lambda$.

B.1 Proof of Theorem 14

Proof. Note that by denoting $\omega^t(v,p) = \frac{R^t_{p\to v}}{\lambda}$, we have

$$\omega^{t+1}(v,p) = H(\omega^t)(v,p),$$

where H is as defined in (32). Then the convergence of $q^t(v,p) = \frac{R^t_{p\to v}}{1+R^t_{p\to v}}$ to a unique fixed point q^* follows from Corollary 13. More precisely, there is $\Delta_0 = \Delta_0(\delta)$ and $C = C(\epsilon_0, \delta)$ such that for all $\Delta > \Delta_0$ all $\lambda < (1-\delta)\lambda_c(T_\Delta)$ and all $t > C$,

$$\left|\omega^t(v,p) - \omega^*(v,p)\right| \le \epsilon_0,$$

Note that for all $t > 1$, we have $\omega^t(v,p), \omega^*(v,p) \in [(1+\lambda)^{-\Delta}, 1]$ where $(1+\lambda)^{-\Delta} > 1/36$ for $\lambda < \lambda_c(T_\Delta)$ for all sufficiently large Δ. Then

$$\left|\frac{q^t(v,p)}{q^*(v,p)} - 1\right| = \left|\frac{\omega^t(v,p)}{\omega^*(v,p)} \cdot \frac{1+\omega^*(v,p)}{1+\omega^t(v,p)} - 1\right| = \frac{|\omega^t(v,p) - \omega^*(v,p)|}{\omega^*(v,p)(1+\omega^t(v,p))} \le 36\epsilon_0.$$

By choosing $\epsilon_0 = \frac{\epsilon}{36}$, we have $\left|\frac{q^t(v,p)}{q^*(v,p)} - 1\right| \le \epsilon$.

We then show that there is a $\Delta_0 = O(\frac{1}{\delta\epsilon})$ such that for all $\Delta > \Delta_0$ and all $\lambda < (1-\delta)\lambda_c(T_\Delta)$, the fixed points of the two BPs have $\left|\frac{q^*(v,p)}{\tilde{q}^*(v)} - 1\right| \le \epsilon$

Let $\omega^t(v,p) = \frac{q^t(v,p)}{\lambda(1-q^t(v,p))}$ and $\tilde{\omega}^t(v) = \frac{\tilde{q}^t(v)}{\lambda(1-\tilde{q}^t(v))}$. It follows that

$$\omega^{t+1}(v,p) = \prod_{u\in N(v)\setminus\{p\}} \frac{1}{1+\lambda\omega^t(u,v)} = (1+\lambda\omega^t(p,v)) \prod_{u\in N(v)} \frac{1}{1+\lambda\omega^t(u,v)},$$

$$\tilde{\omega}^{t+1}(v) = \prod_{u\in N(v)} \frac{1}{1+\lambda\tilde{\omega}^t(u)}.$$

We also define

$$\omega^{t+1}(v) = \prod_{u\in N(v)} \frac{1}{1+\lambda\omega^t(u,v)},$$

therefore $\omega^{t+1}(v,p) = (1+\lambda\omega^t(p,v))\omega^{t+1}(v)$. Note that $\omega^t(p,v) \in (0,1]$, thus $|\omega^{t+1}(v,p) - \omega^{t+1}(v)| \le \lambda$. Also recall that $\lambda < \lambda_c(T_\Delta) \le 3/(\Delta-2)$ for all sufficiently large Δ, therefore

$$|\omega^{t+1}(v,p) - \omega^{t+1}(v)| \le 3/(\Delta-2).$$

Let $\Psi(\cdot)$ be as defined in (4). Note for $t > 1$ both $\omega^{t+1}(v,p)$ and $\omega^{t+1}(v)$ are from the range $[(1+\lambda)^{-\Delta}, 1]$. By (26), for $\lambda < \lambda_c(T_\Delta)$ for all sufficiently large Δ, we have

$$|\Psi(\omega^{t+1}(v,p)) - \Psi(\omega^{t+1}(v))| \le 9/(\Delta-2). \tag{35}$$

We assume that $|\Psi(\omega^t(v,p)) - \Psi(\tilde{\omega}^t(v))| \le \epsilon_0$ for all $(v,p) \in E$. Then due to (31),

$$|\Psi(\omega^{t+1}(v)) - \Psi(\tilde{\omega}^{t+1}(v))| \le (1-\delta/6) \cdot \max_{u\in N(v)} |\Psi(\omega^t(u,v)) - \Psi(\tilde{\omega}^t(u))| \le (1-\delta/6)\epsilon_0.$$

Combined with (35), by triangle inequality, we have

$$|\Psi(\omega^{t+1}(v,p)) - \Psi(\tilde{\omega}^{t+1}(v))| \le (1-\delta/6)\epsilon_0 + 9/(\Delta-2),$$

which is at most ϵ_0 as long as $\Delta \geq \Delta_0 \geq \frac{54}{\delta\epsilon_0} + 2$. It means that if $|\Psi(\omega^t(v)) - \Psi(\omega^t(v,p))| \leq \epsilon_0 \leq \frac{54}{\delta(\Delta_0-2)}$, then $|\Psi(\omega^{t+1}(v,p)) - \Psi(\tilde{\omega}^{t+1}(v))| \leq \frac{54}{\delta(\Delta_0-2)}$. Knowing the convergences of $\omega^t(v,p)$ to $\omega^*(v,p)$ and $\tilde{\omega}^t(v)$ to $\omega^*(v)$ as $t \to \infty$, this gives us that

$$|\Psi(\omega^*(v,p)) - \Psi(\tilde{\omega}^*(v))| \leq \frac{54}{\delta(\Delta_0 - 2)}.$$

By (26), it implies $|\omega^*(v,p) - \tilde{\omega}^*(v)| \leq \frac{162}{\delta(\Delta_0-2)}$. Again since $\omega^*(v,p), \tilde{\omega}^*(v) \in [1/36, 1]$ when $\lambda < \lambda_c(T_\Delta)$ for sufficiently large Δ. It holds that

$$\left|\frac{q^*(v,p)}{\tilde{q}^*(v)} - 1\right| = \left|\frac{\omega^*(v,p)}{\tilde{\omega}^*(v)} \cdot \frac{1 + \lambda\tilde{\omega}^*(v)}{1 + \lambda\omega^*(v,p)} - 1\right| \leq \frac{6000}{\delta(\Delta_0 - 2)}.$$

By choosing a suitable $\Delta_0 = O(\frac{1}{\delta\epsilon})$, we can make this error bounded by ϵ. □

B.2 Proof of Lemma 15

Proof. It holds that

$$\left|\frac{\tilde{q}^*(v)}{\mu(v \text{ occupied})} - 1\right| = \left|\frac{q^*(v)}{\frac{\lambda}{1+\lambda}\mathbb{E}\left[\mathbf{R}(X,v)\right]} \frac{\frac{\lambda}{1+\lambda}\mathbb{E}\left[\mathbf{R}(X,v)\right]}{\mu(v \text{ occupied})} - 1\right|, \tag{36}$$

where the expectation in the nominator is w.r.t. the random variable X which is distributed as in μ. For showing the lemma we need to bound appropriately the two ratios on the r.h.s. of (36). For this we use the following two results. The first one is that

$$\left|\frac{\frac{\lambda}{1+\lambda}\mathbb{E}\left[\mathbf{R}(X,v)\right]}{\mu(v \text{ is occupied})} - 1\right| \leq 200e^e\lambda. \tag{37}$$

The second result is Lemma 16.

Lemma 16. *For every $\delta, \theta > 0$, there exists $\Delta_0 = \Delta_0(\delta, \theta)$ and $C > 0$ all $\lambda < (1-\delta)\lambda_c(\Delta)$, and G of maximum degree Δ and girth ≥ 6, the following is true:*

Let X be distributed as the Gibbs distribution. For any $z \in V$, it holds that

$$\mathbf{Pr}\left[|\mathbf{R}(X,z) - \omega^*(z)| \leq \theta\right] \geq 1 - \exp(-\Delta/C),$$

where ω^* is defined in Lemma 4.

The proof of Lemma 16 appears in Section B.3.

Before proving (37), let us show how it implies the lemma, together with Lemma 16. For any independent set σ and any v, it holds that $e^{-e} \leq \omega^*(v), \mathbf{R}(\sigma, v) \leq 1$. Then, Lemma 16 implies that

$$\left|\frac{\omega^*(v)}{\mathbb{E}\left[\mathbf{R}(X,v)\right]} - 1\right| \leq \epsilon/20. \tag{38}$$

Noting that by definition it holds that $\tilde{q}^*(v) = \frac{\lambda\omega^*}{1+\lambda\omega^*}$, we have that

$$\begin{aligned}
\left|\frac{\tilde{q}^*(v)}{\frac{\lambda}{1+\lambda}\mathbb{E}\left[\mathbf{R}(X,v)\right]} - 1\right| &= \left|\frac{1+\lambda}{1+\lambda\omega^*(v)} \frac{\omega^*(v)}{\mathbb{E}\left[\mathbf{R}(X,v)\right]} - 1\right| \\
&\leq \frac{10\lambda}{(1+\lambda\omega^*(v))} \frac{\omega^*(v)}{\mathbb{E}\left[\mathbf{R}(X,v)\right]} + \left|\frac{\omega^*(v)}{\mathbb{E}\left[\mathbf{R}(X,v)\right]} - 1\right| \leq \epsilon/15. \qquad (39)
\end{aligned}$$

In the last inequality we use (38), the fact that $\lambda < 2e/\Delta$ and Δ is sufficiently large. The lemma follows by plugging (37) and (39) into (36). We proceed by showing (37). It holds that

$$\mu(v \text{ is occupied}) = \frac{\lambda}{1+\lambda}\mu(v \text{ is unblocked}) \tag{40}$$

We are going to express $\mu(v \text{ is unblocked})$ it terms of the quantity $\mathbf{R}(\cdot,\cdot)$. For X distributed as in μ it is elementary to verify that

$$\mathbb{E}\left[\mathbf{R}(X,v) \mid v \text{ is unoccupied}\right] = \mu(v \text{ is unblocked}|v \text{ is unoccupied}) \tag{41}$$

Furthermore, it holds that

$$\begin{aligned}
\mathbb{E}\left[\mathbf{R}(X,v)\right] &= \mu(v \text{ occupied}) \cdot \mathbb{E}\left[\mathbf{R}(X,v) \mid v \text{ occupied}\right] + \mu(v \text{ unoccupied}) \cdot \mathbb{E}\left[\mathbf{R}(X,v) \mid w \text{ unoccupied}\right] \\
&\leq \mu(v \text{ occupied}) + \mathbb{E}\left[\mathbf{R}(X,v) \mid v \text{ unoccupied}\right] \qquad [\text{since } 0 < R(X,v) \leq 1] \\
&\leq 2\lambda + \mathbb{E}\left[\mathbf{R}(X,v) \mid v \text{ unoccupied}\right] \qquad [\text{since } \mu(v \text{ occupied}) \leq 2\lambda]
\end{aligned}$$

Since $e^{-e} \leq \mathbf{R}(X,v) \leq 1$, the inequality above yields

$$\mathbb{E}\left[\mathbf{R}(X,v) \mid v \text{ unoccupied}\right] \geq (1-2e^e\lambda)\,\mathbb{E}\left[\mathbf{R}(X,v)\right].$$

Also, using the fact that $\mathbf{R}(X,v) > 0$, we get

$$\mathbb{E}\left[\mathbf{R}(X,v) \mid v \text{ unoccupied}\right] \leq \frac{\mathbb{E}\left[\mathbf{R}(X,v)\right]}{\mu(v \text{ is unoccupied})} \leq (1+5\lambda)\mathbb{E}\left[\mathbf{R}(X,v)\right].$$

In the last inequality we use the fact that $\mu(w \text{ is occupied}) \leq 2\lambda$. From the above two inequalities we get that

$$\left|\mathbb{E}\left[\mathbf{R}(X,w) \mid w \text{ unoccupied}\right] - \mathbb{E}\left[\mathbf{R}(X,w)\right]\right| \leq 10e^e\lambda. \tag{42}$$

In a very similar manner as above, we also get that

$$|\mu(v \text{ is unblocked}|v \text{ is unoccupied}) - \mu(v \text{ is unblocked})| \leq 10e^e\lambda \tag{43}$$

Combining (41), (42), (43) , (40) and using the fact that $e^{-e} \leq \mu(v \text{ is unblocked}), \mathbb{E}\left[\mathbf{R}(X,w)\right]$ we get the following

$$\mu(v \text{ is occupied}) = \frac{\lambda}{1+\lambda}\mathbb{E}\left[\mathbf{R}(X,w)\right](1+50e^e\lambda)\,. \tag{44}$$

Then (37) follows from (44). □

B.3 Proof of Lemma 16

The proof of the lemma is similar to the proof of Lemma 28.

Let some fixed integer $R > 0$ whose value is going to be specified later. R is independent of Δ, the maximum degree of G. For every integer $i \leq R$, we define

$$\beta_i := \max\left|\Psi(\mathbf{R}(X,x)) - \Psi(\omega^*(x))\right|,$$

where Ψ is defined in (4). The maximum is taken over all vertices $x \in B_i(w)$.

An elementary observation is that $\beta_i \leq C_0 = 3$ for every $i \leq R$. To see why this holds, note that for any $z \in V$ and any independent sets σ, it holds that $e^{-e} \leq \mathbf{R}(\sigma,z), \omega^*(z) \leq 1$. Then we get $\beta_i \leq 3$ from (26).

We start by using the fact that $\beta_R \leq C_0$. Then we show that with sufficiently large probability, if $\beta_{i+1} \geq \theta/5$, then $\beta_i \leq (1-\gamma)\beta_{i+1}$ where $0 < \gamma < 1$. Then the lemma follows by taking large R.

For any $i \leq R$, there exists $C_d > 0$ such that with probability at least $1 - \exp(-\varDelta/C_d)$ the following is true: For every vertex $x \in B_i(w)$ it holds that

$$\left| \mathbf{R}(X,x) - \exp\left(-\frac{\lambda}{1+\lambda} \sum_{z \in N(x)} \mathbf{R}(X,z) \right) \right| < \frac{\theta\delta}{40} \tag{45}$$

Note that (45) (that follows from Lemma 9) implies the following.

Fix some $i \leq R$, $z \in B_i(w)$. From the definition of the quantity β_{i+1} we get the following: For any $x \in B_{i+1}(w)$ consider the quantity $\tilde{\omega}(x) = \mathbf{R}(X,x)$. We have that

$$D_{v,i+1}(\tilde{\omega}_s, \omega^*) \leq \beta_{i+1}. \tag{46}$$

We will show that if (45) holds for $\mathbf{R}(X,z)$, where $z \in B_i(w)$, and $\beta_{i+1} \geq \theta/5$, then we have that

$$\left| \varPsi\left(\mathbf{R}(X,z)\right) - \varPsi\left(\omega^*(z)\right) \right| \leq (1-\delta/24)\beta_{i+1}.$$

For proving the above inequality, first note that if $\mathbf{R}(X,z)$ satisfies (45), then (26) implies that

$$\left| \varPsi\left(\mathbf{R}(X,z)\right) - \varPsi\left(\exp\left(-\frac{\lambda}{1+\lambda} \sum_{r \in N(z)} \mathbf{R}(X,r) \right) \right) \right| \leq \frac{\delta\theta}{12}. \tag{47}$$

Furthermore, we have that

$$\begin{aligned}
&\left| \varPsi\left(\mathbf{R}(X,z)\right) - \varPsi\left(\omega^*(z)\right) \right| \\
&\quad \leq \frac{\delta\theta}{12} + \left| \varPsi\left(\exp\left(-\frac{\lambda}{1+\lambda} \sum_{r \in N_z} \mathbf{R}(X,r) \right) \right) - \varPsi\left(\omega^*(z)\right) \right| \qquad \text{[from (47)]} \\
&\quad \leq \frac{\delta\theta}{12} + \left| \varPsi\left(\prod_{r \in N(z)} \left(1 - \frac{\lambda \mathbf{R}(X,r)}{1+\lambda}\right) \right) - \varPsi\left(\omega^*(z)\right) \right| + \\
&\qquad + \left| \varPsi\left(\prod_{r \in N(z)} \left(1 - \frac{\lambda \mathbf{R}(X,r)}{1+\lambda}\right) \right) - \varPsi\left(\exp\left(-\frac{\lambda}{1+\lambda} \sum_{r \in N(z)} \mathbf{R}(X,r) \right) \right) \right|,
\end{aligned} \tag{48}$$

where the last derivation follows from the triangle inequality.

From our assumption about λ and the fact that $\mathbf{R}(X,r) \in [e^{-e}, 1]$, for $r \in N(z)$, we have that

$$\left| \prod_{r \in N(z)} \left(1 - \frac{\lambda \mathbf{R}(X,r)}{1+\lambda}\right) - \exp\left(-\lambda \sum_{r \in N(z)} \frac{\mathbf{R}(X,r)}{1+\lambda} \right) \right| \leq \frac{10}{\varDelta}.$$

The above inequality and (26) imply that

$$\left| \varPsi\left(\prod_{r \in N(z)} \left(1 - \frac{\lambda \mathbf{R}(X,r)}{1+\lambda}\right) \right) - \varPsi\left(\exp\left(-\frac{\lambda}{1+\lambda} \sum_{r \in N(z)} \mathbf{R}(X,r) \right) \right) \right| \leq \frac{30}{\varDelta}.$$

Plugging the inequality above into (48) we get that

$$\begin{aligned} |\Psi(\mathbf{R}(X,z)) - \Psi(\omega^*(z))| &\le \frac{\delta\theta}{12} + \frac{30}{\Delta} + \left| \Psi\left(\prod_{r\in N(z)} \left(1 - \frac{\lambda \mathbf{R}(X,r)}{1+\lambda}\right)\right) - \Psi(\omega^*(z)) \right| \\ &\le \delta\theta/12 + 60/\Delta + D_{v,i}(F(\tilde{\omega}), \omega^*), \end{aligned} \tag{49}$$

where $\tilde{\omega} \in [0,1]^V$ is such that $\tilde{\omega}(z) = \mathbf{R}(X,z)$ for $z \in V$. The function F is defined in (2). Since $\tilde{\omega}$ satisfies (46), Lemma 5 implies that

$$D_{v,i}(F(\tilde{\omega}), \omega^*) \le (1-\delta/6)\beta_{i+1}. \tag{50}$$

Plugging (50) into (49) we get that

$$|\Psi(\mathbf{R}(X_s,z)) - \Psi(\omega^*(z))| \le \delta\theta/12 + 60/\Delta + (1-\delta/6)\beta_{i+1} \le (1-\delta/24)\beta_{i+1}, \tag{51}$$

where the last inequality holds if we have $\beta_{i+1} \ge \theta/5$. Note that (51) holds provided that $\mathbf{R}(X,z)$ satisfies (45). The lemma follows by taking sufficiently large $R = R(\theta)$.

C Basic Properties of Glauber dynamics

C.1 Continuous versus discrete time chains

For many of our results we have a simpler proof when instead of a discrete time Markov chain we consider a continuous time version of the chain. That is, consider the Glauber dynamics where the spin of each vertex is updated according to an independent Poisson clock with rate $1/n$.

We use the following observation, Corollary 5.9 in [23], as a generic tool to argue that typical properties of continuous time chains are typical properties of the discrete time chains too.

Observation 17. *Let (X_t) by any discrete time Markov chain on state space Ω, and let (Y_t) be the corresponding continuous-time chain. Then for any property $P \subset \Omega$ and positive integer t, we have that*

$$\mathbf{Pr}[X_t \notin P] \le e\sqrt{t}\,\mathbf{Pr}[Y_t \notin P].$$

Observation 17 would suffice for our purposes when $\Delta = \Omega(\log n)$, but not for Glauber dynamics on graphs of e.g. constant degree. For the latter case, instead of focusing on specific times t in discrete time, our goal will be to show how events which are rare at a single instant in continuous time must also be rare over a time interval of length $O(n)$ in discrete time, without taking a union bound over all the times in the time interval.

Let the set Ω contain all the independent sets of G. We say that a function $f : \Omega \to \mathbb{R}$ has "total influence" J, if for every independent set $X \in \Omega$ we have

$$\mathbb{E}\left[\left|f(X') - f(X)\right|\right] \le J/n,$$

where X' is the result of one Glauber dynamics update, starting from X.

The next result, Lemma 13 in [10], shows that, for functions f which have Lipschitz constant $O(1/\Delta)$ and total influence $J = O(1)$, in order to prove high-probability bounds for the discrete-time chain that apply for all times in an interval of length $O(n)$, it suffices to be able to prove a similar bound at a single instant in continuous time.

Lemma 18 (Hayes [10]). *Suppose $f : \Omega \to \mathbb{R}$ is a function of independent sets of G and f has Lipschitz constant $\alpha < O(1/\Delta)$ and total influence $J = O(1)$. Let $X_0 = Y_0$ be given and let $(X_t)_{t\geq 0}$ be continuous-time single site dynamics on the hard-core model of G and let $(Y_i)_{i=0,1,2\ldots}$ be the corresponding discrete-time dynamics.*

Suppose that t_0 is a positive integer and S is a measurable set of real numbers, such that for all $t \geq t_0$, $\mathbf{Pr}[f(X_t) \in S] \geq 1 - \exp(-\Omega(\Delta))$. Then, for all $\epsilon \in \Omega(1)$ and all integers $t_1 \geq t_0$, there $t_1 - t_0 = O(n)$ we have that

$$\mathbf{Pr}[(\forall i \in \{t_0, t_0+1, \ldots, t_1\})\ f(Y_i) \in S \pm \epsilon] \geq 1 - \exp(-\Omega(\Delta)),$$

where the hidden constant in Ω notation depends only on the hidden constant in the assumption.

C.2 Basic burn-in properties

Consider a graph $G = (V, E)$. Given some integer $r \geq 0$ and $v \in V$, let $B_r(v)$ be the the ball of radius r, centered at v. Also, let $S_r(v)$ be the sphere of radius r, centered at v. Finally, let $N(v)$ denote the set of vertices which are adjacent to v.

Definition 19. *Let $G = (V, E)$ be a graph of maximum degree Δ and let σ be an independent set of G. For some $\rho > 0$, we say that σ is ρ-heavy for the vertex $v \in V$ if $|B_2(v) \cap \sigma| \geq \rho\Delta$ or $|B_1(v) \cap \sigma| \geq \rho\Delta/\log\Delta$.*

Definition 20. *Let $G = (V, E)$ be a graph of maximum degree Δ. Let σ, τ be independent sets of G. Consider integer $r > 0$ and $v \in V$. If there is a vertex $w \in B_r(v)$ such w is ρ-heavy, then σ is called ρ-suspect for radius r at v. Otherwise, we say that σ is ρ-above suspicion for radius r at v.*

Similarly, for σ, τ such that $\sigma(v) \neq \tau(v)$, we say that v is a ρ-suspect disagreement for radius r if there exists a vertex $w \in B_r(v)$ such that either σ or τ is ρ-heavy at w. Otherwise, we say that v is a ρ-above suspicion disagreement for radius r.

For the purposes of path coupling for every pair of independent sets X, Y we consider shortest paths between X and Y along neighboring independent sets. That is, $X = Z_0 \sim Z_1 \sim \cdots \sim Z_\ell = Y$. This sequence $Z_1, \ldots, Z_\ell$ we call interpolated independent sets for X and Y. A key aspect of the above definitions is that the "niceness" is inherited by interpolated independent sets.

Observation 21. *If X, Y are independent sets, neither of which is ρ-heavy at vertex v, then no interpolated independent set is 2ρ-heavy at v. Likewise, if v is ρ-above suspicion disagreement for radius r, then in every interpolated independent sets v is 2ρ-above suspicion for radius r.*

The following lemma states that (X_t) requires $O(n\log\Delta)$ to burn-in, regardless of X_0.

Lemma 22. *For $\delta > 0$ let $\Delta \geq \Delta_0(\delta)$ and $C_b = C_b(\delta)$. Consider a graph $G = (V, E)$ of maximum degree Δ. Also, let $\lambda \leq (1-\delta)\lambda_c(\Delta)$.*

Let (X_t) be the continuous (or discrete) time Glauber dynamics on the hard-core model with fugacity λ and underlying graph G. Consider $v \in V$ and let $\mathcal{C}_t$ be the event that X_t, is 50-above suspicion for radius $r = \Delta^{9/10}$ for v at time t. Then, for $\mathcal{I} = [10n\log\Delta, n\exp(\Delta/C_b)]$ it holds that

$$\mathbf{Pr}[\cap_{t\in\mathcal{I}}\mathcal{C}_t] \geq 1 - \exp(-\Delta/C_b).$$

Proof. For now, consider the continuous time version of (X_t). Recall that for X_t, the vertex u is not ρ-heavy if both of the following conditions hold

1. $|X_t \cap B_2(u)| \leq \rho\Delta$

2. $|X_t \cap N(u)| \leq \rho\Delta/\log\Delta$.

First we consider a fixed time $t \in \mathcal{I}$. Let $c = t/n$. Note that $c = c(\Delta) \geq 10\log\Delta$. We are going to show that there exists $C' > 0$ such that

$$\mathbf{Pr}\left[\mathcal{C}_t\right] \geq 1 - \exp\left(-\Delta/C'\right). \tag{52}$$

Fix some vertex $u \in B_r(v)$. Let N_0 be the set of vertices in $B_2(u) \cap X_0$ which are not updated during the time period $(0, t]$. That is, for $z \in N_0$ it holds that $X_0(z) = X_t(z)$. Each vertex $z \in B_2(u) \cap X_0$ belongs to N_0 with probability $\exp(-t/n) = e^{-c}$, independently of the other vertices. Since $|B_2(u) \cap X_0| \leq \Delta^2$, it is elementary that the distribution of $|N_0|$ is dominated by $\mathcal{B}(\Delta^2, e^{-c})$, i.e. the binomial with parameters Δ^2 and e^{-c}.

Using Chernoff's bounds we get the following: for $c > 10\log\Delta$ it holds that

$$\mathbf{Pr}\left[N_0 > \Delta/10\right] \leq \exp\left(-\Delta/10\right). \tag{53}$$

Additionally, let $N_1 \subseteq B_2(u)$ contain every vertex u which is updated at least once during the period $(0, t]$. Each vertex $z \in N_1$, which is last updated prior to t at time $s \leq t$, becomes occupied during the update at time s with probability at most $\frac{\lambda}{1+\lambda}$, regardless of $X_s(N(z))$. Then, it is direct that $|X_t \cap N_1|$ is dominated by $\mathcal{B}(N_1, \frac{\lambda}{1+\lambda})$.

Noting that $|N_1| \leq |B_2(u)| \leq \Delta^2$ and $\frac{\lambda}{1+\lambda} < 2e/\Delta$, for $\Delta > \Delta_0$ Chernoff's bound imply that

$$\mathbf{Pr}\left[|N_1 \cap X_t| \geq 15e\Delta\right] \leq \exp\left(-15e\Delta\right). \tag{54}$$

From (53), (54) and a simple union bound, we get that

$$\mathbf{Pr}\left[|X_t \cap B_2(u)| > 42\Delta\right] \leq \exp\left(-\Delta/20\right). \tag{55}$$

Using exactly the same arguments, we also get that

$$\mathbf{Pr}\left[|X_t \cap N(u)| > 42\Delta/\log\Delta\right] \leq \exp\left(-\Delta/20\right). \tag{56}$$

Note that X_0 could be such that $N(u) \cap X_0 = \alpha\Delta$, for some fixed $\alpha > 0$. So as to get $|X_t \cap N(u)| \leq 42\Delta/\log\Delta$ with large probability, we have to ensure that with large probability all the vertices in $N(v)$ are updated at least once. For this reason the burn-in requires at least $10n\log\Delta$ steps.

From (55) and (56) we get the following: For any $\rho > 50$ it holds that

$$\mathbf{Pr}\left[X_t(u) \text{ is not } \rho\text{-heavy}\right] \leq \exp\left(-\Delta/25\right). \tag{57}$$

Then (52) follows by taking a union bound over all the, at most Δ^r vertices in $B_r(v)$. In particular, for $r = \Delta^{9/10}$ and sufficiently large Δ, there exists $C > 0$ such that

$$\mathbf{Pr}\left[\mathcal{C}_t\right] \leq \Delta^r \exp\left(-\Delta/25\right) \leq \exp\left(-\Delta/30\right).$$

The above implies that (52) is indeed true but only for a specific time step $t \in \mathcal{I}$. Now we use a covering argument to deduce the above for the whole interval $\mathcal{I}$.

For sufficiently small $\gamma > 0$, independent of Δ, consider a partition of the time interval $\mathcal{I}$ into subintervals each of length $\frac{\gamma^2}{\Delta}n$, (where the last part can be shorter). We let $T(j)$ be the j-th part in the partition.

Each $z \in B_2(w)$ is updated at least once during the time period $T(j)$ with probability less than $2\frac{\gamma^2}{\Delta}$, independently of the other vertices. Note that $|B_2(w)| \leq \Delta^2$. Clearly, the number of vertices in $B_2(v)$ which are updated during $T(j)$ is dominated by $\mathcal{B}(\Delta^2, 2\gamma^2/\Delta)$. Chernoff bounds imply that with probability at least $1-\exp\left(-20\Delta\gamma^2\right)$, the number of vertices in $B_2(w)$ which are updated during the interval $T(j)$ is at most $20\gamma^2\Delta$. Furthermore, changing any $20\Delta\gamma^2$ variables in $B_2(w)$ can only make the independent set heavier by at most $20\Delta\gamma^2$.

Similarly, we get that with probability at least $1-\exp(-\gamma\Delta)$, the number of vertices in $N(v)$ which are updated during the interval $T(j)$ is at most $\gamma\Delta/\log\Delta$. The change of at most $\gamma\Delta/\log\Delta$ neighbors of v does not change the weight of its neighborhood by more than $\gamma\Delta/\log\Delta$.

From the above arguments we get that the following: We can choose sufficiently large $C_b > 0$ such that for $j \in \{1, 2, \ldots, \lceil\Delta/(\gamma^2)\exp(\Delta/C_b)\rceil\}$ it holds that

$$\mathbf{Pr}\left[\cap_{t\in T(j)}\mathcal{C}_t\right] \geq 1-\exp\left(-100\Delta/C_b\right).$$

The result for continuous time follows by taking a union bound over all the $\left\lceil\Delta/(\gamma^2)\exp(\Delta/C_b)\right\rceil$ many subintervals of $\mathcal{I}$.

For the discrete time case the arguments are very similar. The only extra ingredient we need is that, now, the updates of the vertices are negatively dependent and use [3]. The lemma follows. □

The following lemma states that if (X_t) start from a not so heavy state it only requires $O(n)$ steps to burn in.

Lemma 23. *For $\delta > 0$, let $\Delta \geq \Delta_0(\delta)$ and $C_b = C_b(\delta)$. Consider a graph $G = (V, E)$ of maximum degree Δ. Also, let $\lambda \leq (1-\delta)\lambda_c(\Delta)$.*

Let (X_t) be the continuous (or discrete) time Glauber dynamics on the hard-core model with fugacity λ and underlying graph G. Also, let $\mathcal{C}_t$ be the event that X_t is, are 50-above suspicion for radius $R \leq \Delta^{9/10}$ for v at time t. Assume that X_0 is 400-above suspicion for radius R for v. Then, for $\mathcal{I} = [C_b n, n\exp(\Delta/C_b)]$ we have that

$$\mathbf{Pr}\left[\cap_{t\in\mathcal{I}}\mathcal{C}_t\right] \geq 1-\exp\left[-\Delta/C_b\right].$$

The proof of Lemma 23 is almost identical to the proof of Lemma 22, for this reason we omit it.

C.3 G versus G^* and comparison

Consider G with girth 7. For such graph and some vertex w in G, the radius 3 ball around w is a tree. We let G^*_w be graph that is derived from G by orienting towards w every edge that is within distance 2 from w [1]. For a vertex $x \in G^*_w$, we let $N^*(x) \subseteq N(x)$ contain every z in the neighborhood of x such that either the edge between x, z is unoriented, or the orientation is towards x.

We let the Glauber dynamics (X^*_t) on the hard-core model with underlying graph G^*_w and fugacity λ, be a Markov chain whose transition $X_t \to X_{t+1}$ is defined by the following:

1. Choose u uniformly at random from V.

2. If $N^*(u) \cap X^*_t = \emptyset$, then let

$$X^*_{t+1} = \begin{cases} X^*_t \cup \{u\} & \text{with probability } \lambda/(1+\lambda) \\ X^*_t \setminus \{u\} & \text{with probability } 1/(1+\lambda) \end{cases}$$

3. If $N^*(w) \cap X_t \neq \emptyset$, then let $X^*_{t+1} = X^*_t$.

[1] An edge $\{w_1, w_2\} \in E$ is at distance ℓ from w if the minimum distance between w and any of w_1, w_2 is ℓ.

The state space of (X_t^*) that is implied by the above is a superset of the independent sets of G, since there are pairs of vertices which are adjacent in G while they can both be occupied in X_t^*.

The motivation for using G_w^* and (X_t^*) is better illustrated by considering Lemma 9. In Lemma 9 we establish a recursive relation for $\mathbf{R}()$ for G of girth ≥ 6, in the setting of the Gibbs distribution. An important ingredient in the proof there is that for every vertex x conditioned on the configuration at x and the vertices at distance ≥ 3 from x, the children of x are mutually independent of each other under the Gibbs distribution.

For establishing the uniformity property for Glauber dynamics we need to establish a similar "conditional independence" relation but in the dynamic setting of Markov chains. To obtain this, we will need that G has girth at least 7. Clearly, the conditional independence of Gibbs distribution no longer holds for the Glauber dynamics. To this end we employ the following: Instead of considering G and the standard Glauber dynamics (X_t), we consider G_w^* and the corresponding dynamics (X_t^*).

Using G_w^* and (X_t^*) we get (in the dynamics setting) an effect which is similar to the conditional independence. During the evolution of (X_t^*) the neighbors of w can only exchange information through paths of G_w^* which travel outside the ball of radius 3 around w, i.e. $B_3(w)$. This holds due to the girth assumption for G_w^* and the definition of (X_t^*). In turn this implies that conditional on the configuration of X_t^* outside $B_3(w)$, the (grand)children of w are mutually independent.

The above trick allows to get a recursive relation for $R(X_t^*, w)$ similar to that for the Gibbs distribution. So as to argue that a somehow similar relation holds for the standard dynamics (X_t), we use the following result which states that if (X_t^*) and (X_t) start from the same configuration, then after $O(n)$ the number of disagreements between the two chains is not too large.

Lemma 24. *For $\gamma > 0$, $C_1 > 0$, there exists Δ_0, $C_2 > 0$ such that the following is true: For $w \in V$ consider G_w^* of maximum degree $\Delta > \Delta_0$ and girth at least 7. Also, let (X_t) and (X_t^*) be the continuous time Glauber dynamics on the hard-core model with fugacity $\lambda < (1-\delta)\lambda_c(\Delta)$, underlying graphs G and G_u^*, respectively.*

Assume that (X_t^*) and (X_t) are maximally coupled. Then, if $X_0 = X_0^*$ for X_0 which is 400-above suspicion for radius $R \leq \Delta^{9/10}$, we have that

$$\Pr[\forall s \leq C_1 n,\ \forall u \in V\ |(X_s \oplus X_s^*) \cap B_2(u)| \leq \gamma\Delta] \geq 1 - \exp\left(-\Delta/C_2\right).$$

Before proving Lemma 24 we need to introduce certain notions.

Let us call Z a "generalized Poisson random variable with jumps α and instantaneous rate $r(t)$" if Z is the result of a continuous-time adapted process, which begins at 0 and in each subsequent infinitesimal time interval, samples an increment ∂Z from some distribution over $[0, \alpha]$, having mean $\leq r(t)dt$. Z, the sum of the increments over all times $0 < t < 1$, is a random variable, as is the maximum observed rate, $r^* = \max_{t\in[0,1]} r(t)$.

Remark 25. *In the special case where $\alpha \geq 1$ and the distribution is supported in $\{0, 1\}$ with constant rate $\mu \cdot dt$, Z is a Poisson random variable with mean μ.*

We are going to use the following result, Lemma 12 in [10].

Lemma 26 (Hayes)**.** *Suppose Z is a generalized Poisson random variable with maximum jumps α and maximum observed rate r^*. Then, for every $\mu > 0$, $C > 1$ it holds that*

$$\mathbf{Pr}\left[Z \geq C\mu \text{ and } r^* \leq \mu\right] \leq \exp\left[-\frac{\mu}{\alpha}(C\ln(C) - C + 1)\right] < \left(\frac{e}{C}\right)^{\mu C\alpha}.$$

Proof of Lemma 24. In this proof assume that γC_1 is sufficiently small constant. Also, let $D = \cup_{t\leq C_1 n}(X_t \oplus X_t^*)$, i.e. D denotes the set of all vertices which are disagreeing at least once during the time interval from 0 to $C_1 n$. Given some vertex $u \in V$ let $A_u = \cup_{t \le C_1 n} X_t \cap N(u)$ and $A^*_u = \cap_{t \le C_1 n} X^*_t \cap N(u)$. That is A_u contains every $z \in N(u)$ for which there exists at least one $s < C_1 n$ such that z is occupied in X_s. Similarly for A^*_u. Finally, let the integer $r = \left\lfloor \gamma^5 \frac{\Delta}{\log \Delta} \right\rfloor$.

Let $\mathcal{A}$ denote the event that $\exists s \le C_1 n,\ \exists u \in V\ |(X_s \oplus X^*_s) \cap S_2(u)| \ge \gamma\Delta/2$. Consider the events $\mathcal{B}_1, \mathcal{B}_2$, $\mathcal{B}_3$, $\mathcal{B}_4$ and $\mathcal{B}_5$ be defined as follows: $\mathcal{B}_1$ denotes the event that $D \not\subseteq B_r(w)$. $\mathcal{B}_2$ denotes the event that $|D| \ge \gamma^3\Delta^2$. $\mathcal{B}_3$ denotes the event that the total number of disagreements that appear in $N(u)$, for every $u \in V$, is at most $\gamma^3\Delta$. Finally, $\mathcal{B}_4$ denotes the event that there exists $u \in B_{100}(w)$ such that either $|A(u)| \ge \gamma^3\Delta$ or $|A^*(u)| \ge \gamma^3\Delta$.

Then, the lemma follows by noting the following:

$$\mathbf{Pr}\left[\exists s \le C_1 n,\ \exists u \in V\ |(X_s \oplus X^*_s) \cap B_2(u)| \ge \gamma\Delta\right] \le \mathbf{Pr}\left[\mathcal{A}\right] + \mathbf{Pr}\left[\mathcal{B}_3\right]. \tag{58}$$

The lemma follows by bounding appropriately the probability terms on the r.h.s. of (58).

First consider $\mathbf{Pr}[\mathcal{A}]$. Let $\mathcal{B} = \mathcal{B}_1 \cup \mathcal{B}_2 \cup \mathcal{B}_3 \cup \mathcal{B}_4$. We bound $\mathbf{Pr}[\mathcal{A}]$ by using $\mathcal{B}$ as follows:

$$\begin{aligned}
\mathbf{Pr}\left[\mathcal{A}\right] &= \mathbf{Pr}\left[\mathcal{B}, \mathcal{A}\right] + \mathbf{Pr}\left[\bar{\mathcal{B}}, \mathcal{A}\right] \\
&\le \mathbf{Pr}\left[\mathcal{B}\right] + \mathbf{Pr}\left[\bar{\mathcal{B}}, \mathcal{A}\right] \\
&\le \sum_{i=1}^{4} \mathbf{Pr}\left[\mathcal{B}_i\right] + \mathbf{Pr}\left[\bar{\mathcal{B}}, \mathcal{A}\right],
\end{aligned} \tag{59}$$

where the last inequality follows by applying a simple union bound.

Consider some vertex $u \in V$ and let Z be the total number of disagreements that ever occur in $S_2(u)$ up ot the first time that either $\mathcal{B}$ occurs or up to time $C_1 n$, whichever happens first. If $u \notin B_r(w)$, then Z is always zero since we stop the clock when $D \not\subseteq B_{r-1}(w)$. So our focus is on the case where $u \in B_{r-1}(w)$. For such u the random variable Z follows a generalized Poisson distribution, with jumps of size 1 and maximum observed rate at most $30\gamma^3\Delta dt/n$, over at most $C_1 n$ time units. To see this consider the following.

Given that $\mathcal{B}$ does not occur, disagreements in $S_2(u)$ may be caused due to the following categories of disagreeing edges. Each disagreement in $N(u)$ has at most $\Delta - 1$ disagreeing edges in $S_2(u)$. Since the number of disagreements that appear in $N(u)$ during the time period up to $C_1 n$ is at most $\gamma^3\Delta$, there are at most $\gamma^3\Delta^2$ disagreeing edges incident to $S_2(u)$. On the whole there are at most $\gamma^3\Delta^2$ disagreements from vertices different than those in $N(u)$. Each one of them has at most one neighbor in $S_2(u)$, since the girth is at least 7. That is there are additional $\gamma^3\Delta^2$ many disagreeing edges. Finally, disagreements on $S_2(u)$ may be caused by edges which belong to $G \oplus G^*_w$. There are at most Δ^3 many such edges. Each one of these edges generates disagreements only on the vertex on its tail. Since the out-degree in G^*_w is at most 1, there are Δ^2 disagreeing edges from $G \oplus G^*_w$ which are incident to $S_2(v)$. Additionally, each one of these edges should point to an occupied vertex so as to be disagreeing. Since $\mathcal{B}_4$ does not occur, there at most $2\gamma^3\Delta^2$ edges in $G \oplus G^*_w$ which point to an occupied vertex and have the tail in S_2.

From the above observations, we have that there are at most $10\gamma^3\Delta^2$ disagreeing edges incident to S_2. For the new disagreement to occur in S_2 due to a given such edge, a specific vertex must chosen and should become occupied, which occurs with rate at most $e \cdot dt/(n\Delta)$.

Using Lemma 26, applied with $\mu = 30C_1\gamma^3\Delta$, $\alpha = 1$ and $C = \gamma\Delta/\mu$, we have that

$$\mathbf{Pr}\left[Z \ge \gamma\Delta\right] \le \left(30e\gamma^2 C_1\right)^{\gamma\Delta}.$$

Taking a union bound over the, at most, Δ^r vertices in $B_r(v)$, we get that

$$\mathbf{Pr}\left[\bar{\mathcal{B}}, \mathcal{A}\right] \ \le\ \Delta^r \left(30e\gamma^2 C_1\right)^{\gamma\Delta} \ =\ \exp\left(-\Delta/C_3\right), \tag{60}$$

where $C_3 = C_3(\gamma) > 0$ is a sufficiently large number. In the last derivation we used the fact that $r \leq \frac{\gamma^5 \Delta}{\log \Delta}$.

We proceed by bounding the probability of the events $\mathcal{B}_1$, $\mathcal{B}_2$, $\mathcal{B}_3$ and $\mathcal{B}_4$. The approach is very similar to the proof of Theorem 27 in [10]. We repeat it for the sake of completeness.

Recall that $\mathcal{B}_1$ denotes the event that $D \nsubseteq B_r(w)$. The bound for $\mathbf{Pr}[\mathcal{B}_1]$ uses standard arguments of disagreement percolation. First we observe that every disagreement outside $B_r(w)$ must arise via some path of disagreement which starts within $B_2(w)$. That is we need at least one path of disagreement of length $r-4$. We fix a particular path of length $r-4$ with $B_r(w)$. Let us call it $\mathcal{P}$. We are going to bound the probability that disagreements percolate along $\mathcal{P}$ within $C_1 n$ time units. Let us call this probability ρ.

The number of steps along this path that a disagreement actually percolates is a generalized Poisson random variable with jumps 1 and maximum overall rate at most $C_1 e/\Delta$. This follows by noting that the maximum instantaneous rate is at most $e \cdot dt/(n\Delta)$ integrated over $C_1 n$ time units. We use Lemma 26, to bound the probability for the disagreement to percolate along $\mathcal{P}$, i.e. ρ. Setting $\mu = eC_1/\Delta$, $\alpha = 1$ and $C = (r-4)/\mu$ in Lemma 26 yields the following bound for ρ

$$\rho \leq \left(\frac{e^2 C_1}{\Delta(r-4)}\right)^{r-4}.$$

The above bound holds for any path of length $r-4$ in $B_r(w)$. Taking a union bound over the at most Δ^3 starting point in $B_2(w)$ and the at most Δ^{r-4} paths of length $r-4$ from a given starting point we get that

$$\mathbf{Pr}[\mathcal{B}_1] \leq \Delta^3 \left(\frac{e^2 C_1}{r-4}\right)^{r-4} \leq \exp\left(-\Delta/C_4\right), \tag{61}$$

where $C_4 = C_4(\gamma) > 0$ is a sufficiently large number.

Recall that $\mathcal{B}_2$ denotes the event that $|D| \geq \gamma^3 \Delta^2$. For $\Pr[\mathcal{B}_2]$ we consider the waiting time τ_i for the i'th disagreement, counting from when the $(i-1)$'st disagreement is formed. The event $\mathcal{B}_2$ is equivalent to $\sum_{i=1}^{(\gamma^3\Delta^2)} \tau_i \leq C_1 n$.

Each new disagreement can be attributed to either an edge joining it to an existing disagreement, or to one of the edges in $G \oplus G^*_w$. It follows easily that the total number of such edges is at most $|G \oplus G^*_w| + |(i-1)\Delta| = \Delta^3 + (i-1)\Delta$. Furthermore, for the new disagreement to occur due to a given such edge, a specific vertex must chosen, which occurs with rate at most $e \cdot dt/(n\Delta)$.

The above observations suggest that the waiting time τ_i is stochastically dominated by an exponential distribution with mean $n/[e(\Delta^2+i-1)]$, even conditioning on an arbitrary previous histry $\tau_1, \tau_2, \ldots, \tau_{i-1}$. Therefore, $\sum_i \tau_i$ is stochastically dominated by the sum of independent exponential distributions with mean $n/[e(\Delta^2+i-1)]$.

Applying Corollary 26, from [10] to $\tau_1 + \cdots + \tau_{(\gamma^3\Delta^2)}$, with

$$\mu = \sum_{i=1}^{(\gamma^3\Delta^2)} \frac{n}{e(\Delta^2+i-1)} \geq \int_0^{(\gamma^3\Delta^2)} \frac{n}{e(\Delta^2+x)} dx = \frac{n}{e}\log(1+\gamma^3)$$

and

$$V = \sum_{i=1}^{(\gamma^3\Delta^2)} \frac{n^2}{e^2(\Delta^2+i-1)^2} \leq \int_0^{\infty} \frac{n^2}{e^2(\Delta^2+x-1)^2} dx = \frac{n^2}{e^2(\Delta^2-1)}.$$

All the above yield

$$\Pr[\mathcal{B}_2] \leq \exp\left(-(\mu - C_1 n)^2/(2V)\right) \leq \exp\left(-\Delta^2/C_5\right), \tag{62}$$

where $C_5 = C_5(\gamma) > 0$ is sufficiently large number.

Let Y be the total number of disagreements that ever occur in $N(u)$ up to the first time that either $D \not\subseteq B_{r-1}(w)$ or $|D| > \gamma^3\Delta^2$ occur or time $C_1 n$ whichever happens first. The variable Y follows a generalized Poisson distribution with jumps of size 1. It is direct to check that the maximum observed rate is at most $(\gamma^3\Delta^2 + 2\Delta)e \cdot dt/(\Delta n) \leq 10\gamma^3\Delta dt/n$, integrated over at most $C_1 n$ time units. This is because the clock stops when $|D| \geq \gamma^3\Delta^2$ and since G has girth at least 7 it is only vertex u that is adjacent to more than one element of $N(u)$. Hence there are at most $\gamma^3\Delta^2 + \Delta - 1$ edges joining joining a disagreement with some vertex in $N(u)$ before the clock stops. Furthermore, disagreements on $N(u)$ may also be caused by incident edges which belong to $G \oplus G^*_w$. Each vertex in $v \in N(u)$ is incident to at most one edge which belongs to $G \oplus G^*_w$ and could cause disagreement in v. That is, $N(u)$ has at most at most Δ such edges.

Applying Lemma 26, once more, for Y with $\mu = 10C_1\gamma^3\Delta$, $\alpha = 1$ and $C = \gamma^{3/2}\Delta/\mu$ we get that

$$\mathbf{Pr}\left[Y \geq \gamma^2\Delta\right] \leq \left(\frac{10eC_1\gamma^3\Delta}{\gamma^{3/2}\Delta}\right)^{\gamma^{3/2}\Delta} \leq \left(10eC_1\gamma^{3/2}\right)^{\gamma^{3/2}\Delta}.$$

Taking a union bound over the at most Δ^r vertices in $B_r(w)$ gives an upper bound for the probability the event $\mathcal{B}_3$ happens and at the same time neither $\mathcal{B}_1$ nor $\mathcal{B}_2$ occur. That is

$$\mathbf{Pr}\left[\bar{\mathcal{B}}_1 \text{ and } \bar{\mathcal{B}}_2 \text{ and } \mathcal{B}_3\right] \leq \Delta^r \left(10eC_1\gamma^{3/2}\right)^{\gamma^{3/2}\Delta} \tag{63}$$

Letting $\mathcal{C} = \mathcal{B}_1 \cup \mathcal{B}_2$, we have that

$$\begin{aligned}
\mathbf{Pr}\left[\mathcal{B}_3\right] &= \mathbf{Pr}\left[\mathcal{C}, \mathcal{B}_3\right] + \mathbf{Pr}\left[\bar{\mathcal{C}}, \mathcal{B}_3\right] \\
&\leq \mathbf{Pr}\left[\mathcal{C}\right] + \mathbf{Pr}\left[\bar{\mathcal{C}}, \mathcal{B}_3\right] \\
&\leq \mathbf{Pr}\left[\mathcal{B}_1\right] + \mathbf{Pr}\left[\mathcal{B}_2\right] + \mathbf{Pr}\left[\bar{\mathcal{B}}_1 \text{ and } \bar{\mathcal{B}}_2 \text{ and } \mathcal{B}_3\right] \qquad [\text{union bound for } \mathbf{Pr}\left[\mathcal{C}\right]] \\
&\leq \exp\left(-\Delta/C_6\right),
\end{aligned} \tag{64}$$

where $C_6 = C_6(\gamma) > 0$. In the last inequality we used (63), (62) and (61).

As far as $\mathbf{Pr}\left[\mathcal{B}_4\right]$ is regarded, first recall that $\mathcal{B}_4$ denotes the event that there exists $z \in B_{100}(w)$ such that either $|A(u)| \geq \gamma^3\Delta$ or $|A^*(u)| \geq \gamma^3\Delta$. Fix some vertex $z \in B_{100}(w)$. W.l.o.g. we consider the chain X_t. There are two cases for z. The first one is that z is occupied in X_0. The second one is z is not occupied in X_0. Then probability that the vertex z is updated becomes occupied at least once up to time $C_1 n$ is at most $2C_1 e/\Delta$, regardless of the rest of the vertices.

Fix some vertex $u \in B_{100}(w)$. Let J_u be the number of vertices $z \in N(u)$ which are *unoccupied* in X_0 but they get into A_u. J_u is dominated by the binomial distribution with parameters Δ and $2C_1 e/\Delta$, i.e. $\mathcal{B}(\Delta, 2C_1 e/\Delta)$. Using Chernoff's bounds we get that

$$\mathbf{Pr}\left[J_u \geq \gamma^3\Delta/10\right] \leq \exp\left(-\gamma^3\Delta/10\right).$$

Let L_u be the number of vertices in $z \in N(u)$ which are occupied in X_0. Since we have that X_0 is 400-above suspicious for radius $R \gg 100$ around w and $u \in \mathcal{B}_{100}(w)$, it holds that $L_u \leq 400\Delta/\log\Delta$. Since $|A_u| = J_u + L_u$ we get that $\mathbf{Pr}\left[|A_u| \geq \gamma^3\Delta\right] \leq \exp\left(-\gamma^3\Delta/10\right)$. Taking a union bound over the at most Δ^{100} vertices in $B_{100}(w)$ we get that

$$\mathbf{Pr}\left[\exists u \in B_{100}(w) \ s.t. \ |A_u| \geq \gamma^3\Delta\right] \leq \Delta^{100}\exp\left(-\gamma^3\Delta/10\right) \leq \exp\left(-\gamma^3\Delta/20\right),$$

where the last inequality holds for sufficiently large Δ. Working in the same way we get that

$$\mathbf{Pr}\left[\exists u \in B_{100}(w) \ s.t. \ |A^*_u| \geq \gamma^3\Delta\right] \leq \exp\left(-\gamma^3\Delta/20\right),$$

Combining the two inequalities above, there exists $C_7 = C_7(\gamma) > 0$ such that

$$\mathbf{Pr}\left[\mathcal{B}_4\right] \le \exp\left(-\Delta/C_7\right). \tag{65}$$

Plugging (65), (64), (62), (61) and (60) into (59), we get that

$$\mathbf{Pr}\left[\mathcal{A}\right] \le \exp\left(-\Delta/C_8\right), \tag{66}$$

for appropriate $C_8 > 0$. The lemma follows by plugging (66) and (64) into (58). □

D Proof of Local Uniformity - Proof of Theorems 7, 8

In this section we prove the uniformity results (Theorems 7 and 8) that are presented in Section 4.

D.1 Proof of Theorem 8

In light of Lemma 22, Theorem 8 follows as a corollary from the following result which considers initial state for (X_t) which is not heavy around v.

Theorem 27. *For all $\delta, \epsilon > 0$, let $\Delta_0 = \Delta_0(\delta, \epsilon), C = C(\delta, \epsilon)$. For graph $G = (V, E)$ of maximum degree $\Delta \ge \Delta_0$ and girth ≥ 7, all $\lambda < (1-\delta)\lambda_c(\Delta)$, let (X_t) be the continuous (or discrete) time Glauber dynamics on the hard-core model. If X_0 is 400-above suspicion for radius $R = R(\delta, \epsilon) > 1$ for $v \in V$, it holds that*

$$\mathbf{Pr}\left[(\forall t \in \mathcal{I}) \quad \mathbf{W}_{X_t}(v) < \sum_{z \in N(v)} \omega^*(z)\Phi(z) + \epsilon\Delta\right] \ge 1 - \exp\left(-\Delta/C\right), \tag{67}$$

where the time interval $\mathcal{I} = [Cn, n\exp\left(\Delta/C\right)]$.

We will use Lemmas 22, 23 and 24 to complete the proof of Theorem 27. For an independent set σ of G and $w \in V$, recall that $\mathbf{R}(\sigma, w) = \prod_{z \in N(w)} \left(1 - \frac{\lambda}{1+\lambda}\mathbf{U}_{z,v}(\sigma)\right)$, where $\mathbf{U}_{z,w}(\sigma) = \mathbf{1}\left\{\sigma \cap (N(z) \setminus \{w\}) = \emptyset\right\}$.

The following result which is the Glauber dynamics' version of (17), in Section 4.1.

Lemma 28. *Let $\epsilon > 0$, R, C and λ be as in Theorem 27. Let (X_t) be the continuous time Glauber dynamics on the hard-core model with fugacity λ and underlying graphs G. If X_0 is 400-above suspicion for radius R for $w \in V$, then we have that*

$$\mathbf{Pr}\left[(\forall t \in \mathcal{I}) \quad |\mathbf{R}(X_t, w) - \omega^*(w)| \le \epsilon/10\right] \ge 1 - \exp\left(-20\Delta/C\right), \tag{68}$$

where $\mathcal{I} = [Cn, n\exp\left(\Delta/C\right)]$

The proof of Lemma 28 makes use of the following result, which is the Glauber dynamics' version of Lemma 9, in Section 4.1.

Lemma 29. *For $\delta, \gamma > 0$, let $\Delta_0 = \Delta_0(\delta, \gamma), C = C(\delta, \gamma)$, $\hat{C} = \hat{C}(\delta, \gamma)$. For all graphs $G = (V, E)$ of maximum degree $\Delta \ge \Delta_0$ and girth ≥ 7, all $\lambda < (1-\delta)\lambda_c(\Delta)$, let (X_t) be the continuous time Glauber dynamics on the hard-core model.*

Let X_0 be 400-above suspicion for radius $R \le \Delta^{9/10}$ for $w \in V$. Then, for $x \in B_{R/2}(w)$ and $I = [t_0, t_1]$, where $t_0 = Cn$, it holds that

$$\mathbf{Pr}\left[(\forall t \in I) \quad \left| \mathbf{R}(X_t, x) - \exp\left(-\frac{\lambda}{1+\lambda} \sum_{z \in N(x)} \mathbb{E}_{t_z}\left[\mathbf{R}(X_{t_z}, z)\right]\right)\right| \le \gamma\right] \ge 1 - \left(1 + \frac{t_1 - t_0}{n}\right) \exp\left(-\Delta/\hat{C}\right),$$

where $\mathbb{E}_{t_z}\left[\mathbf{R}(X_{t_z}, z)\right]$ is the expectation w.r.t. random time t_z, the last time that vertex z is updated prior to time t.

Note that $\mathbb{E}_{t_z}\left[\mathbf{R}(X_{t_z}, z)\right] = \exp(-t/n)\ \mathbf{R}(X_0, z) + \int_0^t \mathbf{R}(X_s, z) n \exp\left((s-t)/n\right) ds$. The proof of Lemma 29 appears in Section D.3.

Proof of Lemma 28. Recall that $\mathcal{I} = [Cn, n\exp(\Delta/C]$. Let $R = \lfloor 30\delta^{-1}\log(6\epsilon^{-1})\rfloor$ Assume that C is sufficiently large such that $C = (R+1)C_1$, where C_1 is specified later. Let $T_0 = (R+1)C_1 n$ and $T_1 = \exp(\Delta/C)$. Finally, for $i \le R$ let $\mathcal{I}_i := [T_0 - iC_1 n, T_1]$.

Consider the continuous time Glauber dynamics (X_t). Also, consider times $t \ge T_0 - RC_1 n$. For each such time t and positive integer $i \le R$, we define

$$\alpha_i := \max\left|\Psi(\mathbf{R}(X_t, x)) - \Psi(\omega^*(x))\right|,$$

where Ψ is defined in (4). The maximum is taken over all $t \in \mathcal{I}_i$ and over all vertices $x \in B_i(w)$.

An elementary observation about α_i is that $\alpha_i \le 3$ for every $i \le R$. To see why this holds, note the following: For any $z \in V$ and any independent sets σ, it holds that

$$\mathbf{R}(\sigma, z) \;=\; \prod_{r \in N(z)} \left(1 - \frac{\lambda \cdot \mathbf{U}_{r,z}(\sigma)}{1+\lambda}\right) \;\ge\; (1+\lambda)^{-\Delta} \;\ge\; e^{-\lambda\Delta} \;\ge\; \mathrm{e}^{-\mathrm{e}},$$

where in the last inequality we use the fact that Δ is sufficiently large, i.e. $\Delta > \Delta_0(\epsilon, \delta)$ and $\lambda < e/\Delta$. Furthermore, using the same arguments as above we get that $\omega^*(z) \ge e^{-e}$, as well. Since for any $x \in V$ and any independent sets σ, we have $\mathbf{R}(\sigma, x), \omega^*(x) \in [e^{-e}, 1]$, (26) implies $\alpha_i \le C_0 = 3$, for every $i \le R$.

We prove our result by showing that typically α_0 is very small. Then, the lemma follows by using standard arguments. We use an inductive argument to show that α_0 very small. We start by using the fact that $\alpha_R \le C_0$. Then we show that with sufficiently large probability, if $\alpha_{i+1} \ge \epsilon/20$, then $\alpha_i \le (1-\gamma)\alpha_{i+1}$ where $0 < \gamma < 1$.

For any $i \le R$, we use the fact that there exists $\hat{C} > 0$ such that with probability at least $1 - \exp\left(-\Delta/\hat{C}\right)$ the following is true: For every vertex $z \in B_i(w)$ it holds that

$$(\forall t \in \mathcal{I}_i) \quad \left|\mathbf{R}(X_t, z) - \exp\left(-\frac{\lambda}{1+\lambda}\sum_{r \in N(z)} \tilde{\omega}(r)\right)\right| < \frac{\epsilon\delta}{40} \tag{69}$$

where

$$\tilde{\omega}_t(r) = \exp(-C_1) \cdot \mathbf{R}(X_{t - C_1 n}, r) + \int_{t - C_1 n}^{t} \mathbf{R}(X_s, r) n \exp\left[(s - C_1 n)/n\right] ds. \tag{70}$$

Eq. (69) is implied by Lemmas 23, 29.

Fix some $i \leq R$, $z \in B_i(w)$ and time $s \in \mathcal{I}_i$. We consider X_s by conditioning on $X_{s-C_1 n}$. From the definition of the quantity α_{i+1} we get the following: For any $x \in B_{i+1}(w)$ consider the quantity $\tilde{\omega}_s(x)$. We have that

$$D_{v,i+1}(\tilde{\omega}_s, \omega^*) \leq \alpha_{i+1}. \tag{71}$$

We will show that if (69) holds for $\mathbf{R}(X_s, z)$, where $z \in B_i(w)$, and $\alpha_{i+1} \geq \epsilon/20$, then we have that

$$|\Psi(\mathbf{R}(X_s, z)) - \Psi(\omega^*(z))| \leq (1 - \delta/24)\alpha_{i+1},$$

For proving the above inequality, first note that if $\mathbf{R}(X_s, z)$ satisfies (69), then (26) implies that

$$\left| \Psi(\mathbf{R}(X_s, z)) - \Psi\left(\exp\left(-\frac{\lambda}{1+\lambda}\sum_{r \in N(z)} \tilde{\omega}_s(r)\right)\right)\right| \leq \frac{\delta\epsilon}{12}. \tag{72}$$

Furthermore, we have that

$$\begin{aligned} &|\Psi(\mathbf{R}(X_s, z)) - \Psi(\omega^*(z))| \\ &\quad \leq \frac{\delta\epsilon}{12} + \left| \Psi\left(\exp\left(-\frac{\lambda}{1+\lambda}\sum_{r \in N_z} \tilde{\omega}_s(r)\right)\right) - \Psi(\omega^*(z))\right| \qquad \text{[from (72)]} \\ &\quad \leq \frac{\delta\epsilon}{12} + \left| \Psi\left(\prod_{r \in N(z)}\left(1 - \frac{\lambda\tilde{\omega}_s(r)}{1+\lambda}\right)\right) - \Psi(\omega^*(z))\right| \\ &\qquad + \left| \Psi\left(\prod_{r \in N(z)}\left(1 - \frac{\lambda\tilde{\omega}_s(r)}{1+\lambda}\right)\right) - \Psi\left(\exp\left(-\frac{\lambda}{1+\lambda}\sum_{r \in N(z)} \tilde{\omega}(r)\right)\right)\right|, \end{aligned} \tag{73}$$

where the last derivation follows from the triangle inequality.

From our assumptions about λ, Δ and the fact that $\tilde{\omega}_s(r) \in [e^{-e}, 1]$, for $r \in N(z)$, we have that

$$\left| \prod_{r \in N(z)}\left(1 - \frac{\lambda\tilde{\omega}_s(r)}{1+\lambda}\right) - \exp\left(-\lambda \sum_{r \in N(z)} \frac{\tilde{\omega}_s(r)}{1+\lambda}\right)\right| \leq \frac{10}{\Delta}.$$

The above inequality and (26) imply that

$$\left| \Psi\left(\prod_{r \in N(z)}\left(1 - \frac{\lambda\tilde{\omega}_s(r)}{1+\lambda}\right)\right) - \Psi\left(\exp\left(-\frac{\lambda}{1+\lambda}\sum_{r \in N(z)} \tilde{\omega}_s(r)\right)\right)\right| \leq \frac{30}{\Delta}.$$

Plugging the inequality above into (73) we get that

$$\begin{aligned} |\Psi(\mathbf{R}(X_s, z)) - \Psi(\omega^*(z))| &\leq \frac{\delta\epsilon}{12} + \frac{30}{\Delta} + \left| \Psi\left(\prod_{r \in N(z)}\left(1 - \frac{\lambda\tilde{\omega}_s(r)}{1+\lambda}\right)\right) - \Psi(\omega^*(z))\right| \\ &\leq \frac{\delta\epsilon}{12} + \frac{60}{\Delta} + D_{v,i}(F(\tilde{\omega}), \omega^*), \end{aligned} \tag{74}$$

where the function F is defined in (2). Since $\tilde{\omega_s}$ satisfies (71), Lemma 5 implies that

$$D_{v,i}(F(\tilde{\omega}), \omega^*) \leq (1 - \delta/6)\alpha_{i+1}. \tag{75}$$

Plugging (75) into (74) we get that

$$|\Psi(\mathbf{R}(X_s,z)) - \Psi(\omega^*(z))| \le \frac{\delta\epsilon}{12} + \frac{60}{\Delta} + (1-\delta/6)\alpha_{i+1} \le (1-\delta/24)\alpha_{i+1}, \tag{76}$$

where the last inequality follows if we have $\alpha_{i+1} \ge \epsilon/20$. Note that (76) holds provided that $\mathbf{R}(X_s,z)$ satisfies (69).

So as to bound α_i we have to take the maximum over all times $t \in \mathcal{I}_i$ and vertices $z \in B_i(w)$. So far, i.e. in (76), we only considered a fixed time $s \in \mathcal{I}_i$ and a fixed vertex z.

Consider, now, a partition of $\mathcal{I}_i$ into subintervals each of length $\frac{\epsilon^4\eta}{200\Delta}n$, where the last part can be of smaller length. Let $T(j)$ be the j-th part, for $j \in \{1,\dots,\lceil 200C_1\Delta/(\epsilon^4\eta)\rceil\}$. For some some vertex $x \in V$, each $r \in N(x)$ is updated during the time period $T(j)$ with probability less than $\frac{\epsilon^4\eta}{100\Delta}$, independently of the other vertices.

Chernoff's bounds imply that with probability at least $1-\exp\left(-\Delta\epsilon^3/3\right)$, the number of vertices in $S_2(x)$ which are updated during the interval $T(j)$ is at most $\Delta\epsilon^3/3$. Furthermore, changing any $\Delta\epsilon^2/3$ variables in $S_2(x)$ can only change $\mathbf{R}(X_s,x)$ by at most $\epsilon^2/3$. Consequently, $\Psi(\mathbf{R}(X_s,x))$ can change by only ϵ^2 within $T(j)$. From a union bound over all subintervals $T(j)$ and all vertices $x \in B_i(w)$, there exists sufficiently large $C>0$ such that:

$$\mathbf{Pr}\left[\alpha_i = \max\{3\epsilon^2 + (1-\delta/24)\alpha_{i+1}, \epsilon/20\}\right] \ge 1 - \exp\left(-52\Delta/C\right).$$

The fact that $\alpha_R \le C_0$ and $R = \left\lfloor 20\delta^{-1}\log(6\epsilon^{-1})\right\rfloor$, implies the following: With probability at least $1-\exp\left(-50\Delta/C\right)$ for every $t \in \mathcal{I}$ it holds that $\alpha_0 \le \epsilon/30$. In turn, (26) implies that

$$|\mathbf{R}(X_t,v) - \omega^*(v)| \le \epsilon/11. \tag{77}$$

The lemma follows. ☐

We conclude the technical results for Theorem 27 by proving the following lemma.

Lemma 30. *Let $\epsilon>0$, R, $\mathcal{I}$ and λ be as in Theorem 27. Let (X_t) be the continuous time Glauber dynamics on the hard-core model with fugacity λ and underlying graphs G. Assume that X_0 is 400 above suspicion for v. Then, for any $t \in \mathcal{I}$, any $\gamma>0$, there is $\hat{C}=\hat{C}(\gamma)>0$ such that*

$$\mathbf{Pr}\left[\left|\mathbf{W}(X_t,v) - \sum_{z\in N(v)} \Phi(z)\cdot \mathbb{E}_{t_z}\left[\mathbf{R}(X_{t_z},z)\right]\right| > \gamma\Delta\right] < \exp\left(-\Delta/\hat{C}\right).$$

Recall that $\mathbb{E}_{t_z}\left[\mathbf{R}(X_{t_z},z)\right]$ is the expectation w.r.t. t_z the time when z was last updated prior to time t, i.e. $\mathbb{E}_{t_r}\left[\mathbf{R}(X_{t_z},z)\right] = \exp(-t/n)\mathbf{R}(X_0,z) + \int_0^t \mathbf{R}(X_s,z)n\exp\left[(s-t)/n\right]ds$.

Proof. Consider, first, the graph G_v^* and the dynamics (X_t^*) such that $X_0^*=X_0$. Condition on X_0^* and on X_t^* restricted to $V\setminus B_2(x)$ for all $t\in\mathcal{I}$. Denote this conditional information by $\mathcal{F}$.

First we are going to show that $\mathbb{E}\left[\mathbf{W}(X_t^*,v)\mid\mathcal{F}\right]$ and $\sum_{z\in N(v)}\Phi(z)\cdot\mathbb{E}\left[\mathbf{R}(X_t^*,z)\mid\mathcal{F}\right]$ are very close. From the definition of $\mathbf{W}(X_t^*,v)$ we have that

$$\mathbb{E}\left[W(X_t^*,v)\mid\mathcal{F}\right] = \sum_{z\in N(v)}\Phi(z)\cdot\mathbb{E}\left[\mathbf{U}_{z,v}(X_t^*)\mid\mathcal{F}\right].$$

Let $c>0$ be such that $t/n=c$. For $\zeta>0$ whose value is going to be specified later, let $H(v)\subseteq N(v)$ be such that $z\in H(v)$ is $|N(z)\cap X_0^*|\ge\zeta^{-1}$. In (101) and (102) we have shown that for $z\notin H(v)$ it holds that

$$\left|\mathbb{E}\left[\mathbf{U}_{z,v}(X_t^*)\mid\mathcal{F}\right] - \mathbb{E}_{t_z}\left[\mathbf{R}(X_{t_z}^*,z)\mid\mathcal{F}\right]\right| \le \theta, \tag{78}$$

where $0 < \theta = \theta(c, \zeta) < 20(\zeta e^c)^{-1}$ while (as in we previously defined)

$$\mathbb{E}_{t_z}\left[\mathbf{R}(X^*_{t_z}, z) \mid \mathcal{F}\right] = \exp(-t/n)\mathbf{R}(X^*_0, z) + \int_0^t \mathbf{R}(X^*_s, z) n \exp\left[(s-t)/n\right] ds.$$

Since X^*_0 is 400-above suspicion for radius R around v, it holds that $|H(v)| \le 400\zeta\Delta$. We have that,

$$\begin{aligned}
&\left|\mathbb{E}\left[\mathbf{W}(X^*_t, v) \mid \mathcal{F}\right] - \sum_{z \in N(v)} \Phi(z) \cdot \mathbb{E}_{t_z}\left[\mathbf{R}(X^*_{t_z}, z) \mid \mathcal{F}\right]\right| \\
&\le \left|\mathbb{E}\left[\mathbf{W}(X^*_t, v) \mid \mathcal{F}\right] - \sum_{z \notin H(v)} \Phi(z) \cdot \mathbb{E}_{t_z}\left[\mathbf{R}(X^*_{t_z}, z) \mid \mathcal{F}\right]\right| + \sum_{z \in H(v)} \Phi(z) \cdot \mathbb{E}_{t_z}\left[\mathbf{R}(X^*_{t_z}, z) \mid \mathcal{F}\right] \\
&\le \left|\mathbb{E}\left[\mathbf{W}(X^*_t, v) \mid \mathcal{F}\right] - \sum_{z \notin H(v)} \Phi(z) \cdot \mathbb{E}_{t_z}\left[\mathbf{R}(X^*_{t_z}, z) \mid \mathcal{F}\right]\right| + 5000\zeta\Delta \qquad [\text{since } \max_z \Phi(z) \le 12] \\
&\le (12\theta + 5000\zeta)\,\Delta. \qquad [\text{from } (78)]
\end{aligned} \tag{79}$$

The fact that $\max_z \Phi(z) \le 12$ is from Theorem 6.

We proceed by showing that $W(X^*_t, v)$ is sufficiently well concentrated about its expectation. Conditioning on $\mathcal{F}$ the random variables $\mathbf{U}_{z,v}(X^*_t)$, for $z \in N(v)$, are fully independent. From Chernoff's bounds, there exists appropriate $C_1 > 0$ such that

$$\mathbf{Pr}\left[\left|\mathbf{W}(X^*_t, v) - \mathbb{E}\left[\mathbf{W}(X^*_t, v) \mid \mathcal{F}\right]\right| > \gamma\Delta/100\right] \le \exp\left(-\Delta/C_1\right). \tag{80}$$

From (80) and (79) there exists $C_2 > 0$ such that that

$$\mathbf{Pr}\left[\left|W(X^*_t, v) - \sum_{z \in N(v)} \Phi(z) \cdot \mathbb{E}_{t_z}\left[\mathbf{R}(X^*_{t_z}, z) \mid \mathcal{F}\right]\right| \ge \gamma\Delta/50\right] \le \exp\left(-\Delta/C_2\right). \tag{81}$$

Furthermore, using Lemma 24 with error parameter γ^2, i.e. $|(X^*_t \oplus X_t) \cap B_2(v)| \le \gamma^2\Delta$, we get the following: There exists appropriate $C_3 = C_3(\gamma) > 0$ such that

$$\mathbf{Pr}\left[\left|\mathbf{W}(X^*_t, v) - \mathbf{W}(X_t, v)\right| \le \gamma\Delta/40\right] \ge 1 - \exp\left(-\Delta/C_3\right). \tag{82}$$

Also, (from Lemma 24 again) with probability at least $1 - \exp\left(-\Delta/C_3\right)$ it holds that

$$\left|\int_0^t \mathbf{R}(X_s, z) n \exp\left[(s-t)/n\right] ds - \int_0^t \mathbf{R}(X^*_s, z) n \exp\left[(s-t)/n\right] ds\right| \le \gamma/600, \tag{83}$$

for every $z \in N(v)$. The above follows by using the fact that changing the spin of any $\gamma^2\Delta$ vertices in $X^*_t(B_2(z))$ changes $\mathbf{R}(X^*_t, z)$ by at most $\gamma/1000$.

Noting that $\Phi(z) \le 12$, for any z, the lemma follows by combining (83), (82) and (81). □

D.2 Local Uniformity for the Glauber Dynamics: Proof of Theorem 27

Using Lemmas 28 and 30, in this section we prove Theorem 27. Recall that Theorem 8 follows as a corollary of Theorem 27 and Lemma 22.

Proof of Theorem 27. For a vertex $u \in N(v)$ consider G^*_u. Consider also the continuous time dynamics (X^*_t) such that $X^*_0 = X_0$.

We condition on the restriction of (X^*_t) to $V \setminus B_2(u)$, for every $t \in \mathcal{I}$. We denote this by $\mathcal{F}$. Fix some $t \in \mathcal{I}$. Since $u \in B_R(v)$ and X_0 is 400-above suspicion for radius R around v we get that

$$\begin{aligned}
&\mathbb{E}_s\left[\mathbf{R}(X^*_s, u) \mid \mathcal{F}\right] \\
&= \exp(-t/n)\mathbf{R}(X^*_0, u) + \int_0^t \mathbf{R}(X^*_t, u) n \exp\left((s-t)/n\right) \\
&= \mathbb{E}_s\left[\exp\left(-\frac{\lambda}{1+\lambda}\sum_{z\in N(u)} \mathbf{U}_{z,u}(X^*_s) + O\left(1/\Delta\right)\right) \mid \mathcal{F}\right] \\
&= \exp\left(-\frac{\lambda}{1+\lambda}\sum_{z\in N(u)} \mathbb{E}_s\left[\mathbf{U}_{z,u}(X^*_s) \mid \mathcal{F}\right] + O\left(1/\Delta\right)\right) \qquad \text{[due to conditioning on } \mathcal{F}\text{]} \\
&\leq \exp\left(-\frac{\lambda}{1+\lambda}\sum_{z\in N(u)} \mathbb{E}_s\left[\mathbf{R}(X^*_s, z) \mid \mathcal{F}\right] + \theta\lambda\Delta + O\left(1/\Delta\right)\right), \qquad (84)
\end{aligned}$$

where in the last inequality we use (78). Note that so as apply (78) $X^*_0(u)$ should be sufficiently "light". This is guaranteed from our assumption that $u \in B_R(v)$ and X_0 is 400-above suspicious for radius R from v.

Furthermore, (69) and Lemma 24 imply the following: There exists $C_1 > 0$ such that with probability at least $1 - \exp\left(-\Delta/C_1\right)$, we have that

$$(\forall t \in \mathcal{I}) \qquad \left|\mathbf{R}(X^*_t, u) - \exp\left(-\frac{\lambda}{1+\lambda}\sum_{r\in N^*(u)} \hat{\omega}(r)\right)\right| < \gamma, \qquad (85)$$

where

$$\hat{\omega}(r) = \exp(-t/n)\mathbf{R}(X^*_0, r) + \int_0^t \mathbf{R}(X^*_s, r) n \exp\left[(s-t)/n\right] ds.$$

Note that for every $r \in N^*(u)$ we have $\hat{\omega}(r) = \mathbb{E}_{t_r}\left[\mathbf{R}(X^*_{t_r}, r) \mid \mathcal{F}\right]$. Using this observation, we plug (84) into (85) and get

$$\mathbf{Pr}\left[\left|\mathbf{R}(X^*_t, u) - \hat{\omega}(u)\right| \geq 10e\theta + \gamma\right] \leq \exp\left(-\Delta/C_1\right). \qquad (86)$$

In the above inequality we used the fact that $\lambda\Delta < 2e$.

Consider the continuous time Glauber dynamics (X_t). From Lemma 24 and (86) there exists $C_3 > 0$ such that for X_t the following holds

$$\mathbf{Pr}\left[\left|\mathbf{R}(X_t, u) - \tilde{\omega}(u)\right| \geq 20e\theta + 2\gamma\right] \leq \exp\left(-\Delta/C_3\right), \qquad (87)$$

where

$$\tilde{\omega}(z) = \exp(-t/n)\mathbf{R}(X_0, z) + \int_0^t \mathbf{R}(X_s, z) n \exp\left[(s-t)/n\right] ds.$$

Furthermore a simple union bound over $u \in N(v)$ and (87) gives that

$$\mathbf{Pr}\left[\forall u \in N(v) \quad |\mathbf{R}(X_t,u) - \tilde{\omega}(u)| \geq 20e\theta + 2\gamma\right] \leq \Delta \exp\left(-\Delta/C_3\right). \tag{88}$$

Taking sufficiently small θ, γ in (88) and using Lemma 30 we get that

$$\mathbf{Pr}\left[\left|\mathbf{W}(X_t,v) - \sum_{w \in N(v)} \Phi(z) \cdot \mathbf{R}(X_t,w)\right| > \epsilon\Delta/15\right] \leq \exp\left(-\Delta/C_4\right), \tag{89}$$

for appropriate $C_4 > 0$. Furthermore, applying Lemma 28, for each $w \in N(v)$ and using (89) yields

$$\mathbf{Pr}\left[\left|\mathbf{W}(X_t,v) - \sum_{w \in N(v)} \Phi(w) \cdot \omega^*(w)\right| > \epsilon\Delta/2\right] \leq \exp\left(-\Delta/C_5\right), \tag{90}$$

for appropriate $C_5 > 0$. The above inequality establishes the desired result for a fixed $t \in \mathcal{I}$.

Now we will prove that (90) holds for all $t \in \mathcal{I}$. Consider a partition of the time interval $\mathcal{I}$ into subintervals each of length $\frac{\psi^2}{\Delta}n$, where the last part can be of smaller length. The quantity $\psi > 0$ is going to be specified later. Also, let $T(j)$ be the j-th part.

Each $z \in B_2(v)$ is updated at least once during the time period $T(j)$ with probability less than $2\frac{\psi^2}{\Delta}$, independently of the other vertices. Note that $|B_2(v)| \leq \Delta^2$. Clearly, the number of vertices in $B_2(v)$ which are updated during $T_i(j)$ is dominated by $\mathcal{B}(\Delta^2, 2\psi^2/\Delta)$. Chernoff's bounds imply that with probability at least $1 - \exp\left(-20\Delta\psi^2\right)$, the number of vertices in $B_2(v)$ which are updated during the interval $T(j)$ is at most $20\psi^2\Delta$. Furthermore, changing any $2\Delta\psi^2$ variables in $B_2(v)$ can only change the weighted sum of unblocked vertices in N_v by at most $20C_0\psi^2\Delta$. Taking sufficiently small $\psi > 0$ we get the following:

$$\mathbf{Pr}\left[\left|\mathbf{W}(X_t,v) - \sum_{w \in N(v)} \Phi(w) \cdot \omega^*(w)\right| > \epsilon\Delta\right] \leq \exp\left(-2\Delta/C_b\right). \tag{91}$$

The above completes the proof of Theorem 27 for the case where (X_t) is the continuous time process.

The discrete time result follows by working as follows: instead of $\mathbf{W}(X_t,v)$ we consider the "normalized" variable $\Lambda(X_t,v) = \frac{\mathbf{W}(X_t,v)}{\Delta}$. Rephrasing (90) in terms of $\Lambda(X_t,v)$ we have, for a specific $t \in \mathcal{I}$:

$$\mathbf{Pr}\left[\left|\Lambda(X_t,v) - \Delta^{-1} \sum_{w \in N(v)} \Phi(w) \cdot \omega^*(w)\right| > \epsilon/2\right] \leq \exp\left(-\Delta/C_5\right). \tag{92}$$

Note that $\Lambda(X_t,v)$ satisfies the Lipschitz and total influence conditions of Lemma 18. Hence by Lemma 18 the result for the discrete time process holds. □

D.3 Approximate recurrence for Glauber dynamics - Proof of Lemma 29

Consider G^*_x and let (X^*_t) be the Glauber dynamics on G^*_x with fugacity $\lambda > 0$ and let $X^*_0 = X_0$. Also assume that (X^*_t) and (X_t) are maximally coupled.

Condition on X_0^*, let $\mathcal{F}$ be the σ-algebra generated by X_t^* restricted to $V \setminus B_2(x)$ for all $t \in I$. Fix $t \in I$. Let $c > 0$ be such that $t/n = c$, i.e. c is a large constant. Recalling the definition of $\mathbf{R}(X_t^*, x)$, we have that

$$\begin{aligned}\mathbf{R}(X_t^*, x) &= \prod_{z \in N(x)} \left(1 - \frac{\lambda}{1+\lambda}\mathbf{U}_{z,x}(X_t^*)\right) \\ &= \exp\left(-\frac{\lambda}{1+\lambda}\sum_{z \in N(x)} \mathbf{U}_{z,x}(X_t^*) + O\left(1/\mathit{\Delta}\right)\right). \end{aligned} \tag{93}$$

Let $\mathbf{Q}(X_t^*) = \sum_{z \in N(x)} \mathbf{U}_{z,x}(X_t^*)$. Conditional on $\mathcal{F}$, the quantity $\mathbf{Q}(X_t^*)$ is a sum of $|N(x)|$ many independent random variables in $[0,1]$. Applying Azuma's inequality, for $0 \leq \gamma \leq (3e)^{-1}$, we have

$$\mathbf{Pr}\left[\left|\mathbb{E}\left[\mathbf{Q}(X_t^*) \mid \mathcal{F}\right] - \mathbf{Q}(X_t^*)\right| \geq \gamma \mathit{\Delta}\right] \leq 2\exp\left(-\gamma^2 \mathit{\Delta}/2\right). \tag{94}$$

Combining the fact that $\mathbb{E}\left[\mathbf{Q}(X_t^*) \mid \mathcal{F}\right] = \sum_{z \in N(x)} \mathbb{E}\left[\mathbf{U}_{z,x}(X_t^*) \mid \mathcal{F}\right]$ with (94) and (93) we get that

$$\mathbf{Pr}\left[\left|\mathbf{R}(X_t^*, x) - \exp\left(-\frac{\lambda}{1+\lambda}\sum_{z \in N(x)} \mathbb{E}\left[\mathbf{U}_{z,x}(X_t^*) \mid \mathcal{F}\right]\right)\right| \geq 3\gamma\lambda\mathit{\Delta}\right] \leq 2\exp\left(-\gamma^2\mathit{\Delta}/2\right). \tag{95}$$

For every $z \in N^*(x)$, it holds that

$$\begin{aligned}&\mathbb{E}\left[\mathbf{U}_{z,x}(X_t^*) \mid \mathcal{F}\right] \\ &\quad= \prod_{y \sim N(z)\setminus\{x\}} \mathbb{E}\left[\mathbf{1}\{y \notin X_t^*\} \mid \mathcal{F}\right] \\ &\quad= \prod_{y \sim N(z)\setminus\{x\}} \left(\Pr[t_y = 0] \cdot \mathbf{1}\{y \notin X_0^*\} + \mathbb{E}\left[\mathbf{1}\{y \notin X_t^*\} \cdot \mathbf{1}\{t_y > 0\} \mid \mathcal{F}\right]\right), \end{aligned} \tag{96}$$

where t_y is the time that vertex y is last updated prior to time t and it is defined to be equal to zero if y is not updated prior to t. Note, for any $0 \leq s \leq t$, it holds that $\Pr[t_y \leq s] = e^{-(t-s)/n}$. Also, we have that

$$\begin{aligned}\mathbb{E}\left[\mathbf{1}\{y \notin X_t^*\} \cdot \mathbf{1}\{t_y > 0\} \mid \mathcal{F}\right] &= \mathbb{E}\left[\mathbb{E}\left[\mathbf{1}\{y \notin X_t^*\} \cdot \mathbf{1}\{t_y > 0\} \mid \mathcal{F}, t_y\right] \mid \mathcal{F}\right] \\ &= \int_0^t \left(1 - \frac{\lambda}{1+\lambda}\mathbf{U}_{y,z}(X_s^*)\right) n \exp\left[(s-t)/n\right] ds, \end{aligned} \tag{97}$$

where the last equality follows because we are using G_x^* and (X_t^*). The use of G^* and (X_t^*) ensures that the configuration in $V \setminus B_2(x)$ is never affected by that in $B_2(x)$. For this reason, if y is updated at time $s \in I$, then the probability for it to be occupied, given $\mathcal{F}$, is exactly $\frac{\lambda}{(1+\lambda)}\mathbf{U}_{y,z}(X_s^*)$. That is, the configuration outside $B_2(x)$ does not provide any information for y but the value of $\mathbf{U}_{y,z}(X_s^*)$.

Plugging (97) into (96) we get that

$$\begin{aligned}&\mathbb{E}\left[\mathbf{U}_{z,x}(X_t^*) \mid \mathcal{F}\right] \\ &\quad= \prod_{y \sim N(z)\setminus\{x\}} \left[\exp\left(-t/n\right)\mathbf{1}\{y \notin X_0^*\} - \int_0^t \left(1 - \frac{\lambda}{1+\lambda}\mathbf{U}_{y,z}(X_s^*)\right) n \exp\left[(s-t)/n\right] ds\right] \\ &\quad= \prod_{y \sim N(z)\setminus\{x\}} \left[1 - \exp\left(-t/n\right)\mathbf{1}\{y \in X_0^*\} - \int_0^t \frac{\lambda}{1+\lambda}\mathbf{U}_{y,z}(X_s^*) n \exp\left[(s-t)/n\right] ds\right]. \end{aligned} \tag{98}$$

For appropriate $\zeta \in (0,1)$, which we define later, let $H(x) \subseteq N^*(x)$ be such that $z \in H(x)$ if $|N^*(z) \cap X_0^*| \geq 1/\zeta$.

Noting that each integral in (98) is less than λ, for every $z \notin H(x)$, we get that

$$\mathbb{E}\left[\mathbf{U}_{z,x}(X_t^*) \mid \mathcal{F}\right] = (1+\delta) \prod_{y \in N(z) \backslash \{x\}} \left(1 - \int_0^t \frac{\lambda}{1+\lambda} \mathbf{U}_{y,z}(X_s^*) n \exp\left[(s-t)/n\right] ds\right), \tag{99}$$

where $|\delta| \leq 4(\zeta e^c)^{-1}$.

Recall that for some vertex y in G_x^* we let $\mathbb{E}_{t_y}[\cdot \mid \mathcal{F}]$, denote the expectation w.r.t. t_y, the random time that y is updated prior to time t. It holds that

$$\mathbb{E}_{t_y}[\mathbf{U}_{y,z}(X_{t_y}^*)|\mathcal{F}] = \exp(-t/n)\mathbf{U}_{y,z}(X_0^*) + \int_0^t \mathbf{U}_{y,z}(X_s^*) n \exp\left[(s-t)/n\right] ds.$$

For every $y \in N(z) \backslash \{x\}$, where $z \notin H(x)$ it holds that

$$\begin{aligned} \mathbb{E}_{t_y}\left[\mathbf{U}_{y,z}(X_{t_y}^*) \mid \mathcal{F}\right] - \int_0^t \mathbf{U}_{y,z}(X_s^*) n \exp\left[(s-t)/n\right] ds &= \exp\left(-t/n\right) \mathbf{U}_{y,z}(X_0^*) \\ &\leq \exp\left(-t/n\right) \leq \exp(-c). \end{aligned} \tag{100}$$

Since $\lambda < e/\Delta$, (99) implies that there is a quantity θ, with $0 < \theta \leq 20(\zeta e^c)^{-1}$, such that

$$\begin{aligned} \mathbb{E}\left[\mathbf{U}_{z,x}(X_t^*) \mid \mathcal{F}\right] &\leq \prod_{y \in N(z) \backslash \{x\}} \left(1 - \int_0^t \frac{\lambda}{1+\lambda} \mathbf{U}_{y,z}(X_s^*) n \exp\left[(s-t)/n\right] ds\right) + \theta/2 \\ &\leq \prod_{y \in N(z) \backslash \{x\}} \left(1 - \mathbb{E}_{t_y}\left[\frac{\lambda}{1+\lambda} \mathbf{U}_{y,z}(X_{t_y}^*) \mid \mathcal{F}\right]\right) + \theta \qquad \text{[from (100)]} \\ &= \prod_{y \in N(z) \backslash \{x\}} \left(1 - \mathbb{E}_s\left[\frac{\lambda}{1+\lambda} \mathbf{U}_{y,z}(X_s^*) \mid \mathcal{F}\right]\right) + \theta, \end{aligned}$$

where in the last derivation, we substituted the variables t_y, for $y \in N(z) \setminus \{x\}$, with a new random variable s which follows the same distribution as t_y. Note that the variables t_y are identically distributed.

Given the σ-algebra $\mathcal{F}$, the variables $\mathbf{U}_{y,z}(X_s^*)$, for $y \in N(z) \setminus \{x\}$, are independent with each other, this yields

$$\begin{aligned} \mathbb{E}\left[\mathbf{U}_{z,x}(X_t^*) \mid \mathcal{F}\right] &= \mathbb{E}_s\left[\prod_y \left(1 - \frac{\lambda}{1+\lambda} \mathbf{U}_{y,z}(X_s^*)\right) \mid \mathcal{F}\right] + \theta \\ &= \mathbb{E}_s\left[\mathbf{R}(X_s^*, z) \mid \mathcal{F}\right] + \theta, \end{aligned} \tag{101}$$

where the last derivation follows from the definition of $\mathbf{R}(X_s^*, z)$. In the same manner, we get that

$$\mathbb{E}\left[\mathbf{U}_{z,x}(X_t^*) \mid \mathcal{F}\right] \geq \mathbb{E}_s\left[\mathbf{R}(X_s^*, z) \mid \mathcal{F}\right] - \theta, \tag{102}$$

for every $z \notin H(x)$.

Since X_0^* is 400 above suspicion for radius R, around w and $x \in B_R(w)$, we have that $|H(x)| \leq 400\zeta\Delta$. This observation and (101), (102) (95), yield that there exists $C' > 0$ such that

$$\mathbf{Pr}\left[\left|\mathbf{R}(X_t^*, x) - \exp\left(-\frac{\lambda}{1+\lambda} \sum_{z \in N(x)} \mathbb{E}_s\left[\mathbf{R}(X_s^*, z) \mid \mathcal{F}\right]\right)\right| \geq 7(\theta + 400\zeta + 3\gamma)\right] \leq \exp\left(-C'\Delta\right), \tag{103}$$

where we use the fact $\frac{\lambda}{1+\lambda}\Delta < e$ and θ, ζ, γ are sufficiently small.

So as to get from (X^*_t) to (X_t) we use Lemma 24, with parameter γ^3. That is, we have that

$$\mathbf{Pr}\left[\exists s \in I \ |(X_s \oplus X^*_s) \cap S_2(x)| \geq \gamma^3 \Delta\right] \leq \exp\left(-\Delta/C''\right),$$

for some sufficiently large constant $C'' > 0$. This implies that

$$\mathbf{Pr}\left[\exists t \in I \ |\mathbf{R}(X^*_t, x) - \mathbf{R}(X_t, x)| \geq \gamma^2\right] \leq \exp\left(-\Delta/C''\right), \tag{104}$$

since changing any $\Delta\gamma^3$ variables in $S_2(x)$ can only change $\mathbf{R}(X^*_s, x)$ by at most γ^2.

With the same observation we also get that with probability at least $1 - \exp\left(-\Delta/C''\right)$ it holds that

$$\left| \int_0^t \mathbf{R}(X^*_s, x) n \exp\left[(s-t)/n\right] ds - \int_0^t \mathbf{R}(X_s, x) n \exp\left[(s-t)/n\right] ds \right| \leq 2\gamma^2. \tag{105}$$

Plugging (104), (105) into (103) and taking appropriate γ, ζ the following is true: There exists $\hat{C} > 0$ such that

$$\mathbf{Pr}\left[\left| \mathbf{R}(X_t, x) - \exp\left(-\frac{\lambda}{1+\lambda} \sum_{r \in N(x)} \mathcal{E}(r) \right) \right| \geq \frac{\eta\epsilon}{20 C_0} \right] \leq \exp\left[-\Delta/\hat{C}\right],$$

where

$$\mathcal{E}(r) = \exp[-t/n] \cdot \mathbf{R}(X_0, r) + \int_0^t \mathbf{R}(X_s, r) n \exp\left[(s-t)/n\right] ds.$$

At this point, we remark that the above tail bound holds for a fixed $t \in I$. For our purpose, we need a tail bound which holds for *every* $t \in I$.

Consider a partition of the time interval I into subintervals each of length $\frac{\zeta^3}{200\Delta} n$, where the last part can be of smaller length. Let $T(j)$ be the j-th part. Each $z \in S_2(x)$ is updated during the time period $T(j)$ with probability less than $\frac{\zeta^3}{100\Delta}$, independently of the other vertices.

Note that $|S_2(x)| \leq \Delta^2$. Chernoff's bounds imply that with probability at least $1 - \exp\left(-\Delta\zeta^3\right)$, the number of vertices in $S_2(x)$ which are updated more than once during the time interval $T(j)$ is at most $\zeta^3 \Delta$. Also, changing any $\Delta\zeta^3$ variables in $S_2(x)$ can only change $\mathbf{R}(X_s, x)$ by at most ζ^2.

The lemma follows by taking a union bound over all $T(j)$ for $j \in \{1, \ldots, \lceil 200 |I| \Delta / (\zeta^3) \rceil\}$ and all vertices $z \in B_i(x)$.

D.4 Local Uniformity for the Gibbs Distribution: Proof of Theorem 7

Proof of Theorem 7. Let $\mathcal{F}$ be the σ-algebra generated by the configuration of v and the vertices at distance greater than 2 from x, i.e. $V \setminus B_2(v)$. Conditioning on $\mathcal{F}$, S_v is a sum of $|N(v)|$ many 0-1 independent random variables. From Azuma's inequality, for any fixed $\gamma > 0$, we have that

$$\mathbf{Pr}\left[|S_v - \mathbb{E}\left[S_v \mid \mathcal{F}\right]| > \gamma\Delta\right] \leq 2\exp\left(-\gamma^2\Delta/2\right). \tag{106}$$

Working as in the proof of Theorem 3 (i.e. for (24), (25)) we get the following: For each $z \in N(v)$ it holds that

$$|\mathbb{E}\left[\mathbf{U}_{z,v}(X) \mid \mathcal{F}\right] - \mathbf{R}(X, z)| \leq 10 e^e \lambda.$$

Note that, given $\mathcal{F}$ the quantity $\mathbf{R}(X, z)$ is uniquely specified.

From the above we get that

$$\mathbb{E}\left[S_v \mid \mathcal{F}\right] = \sum_{z \in N(v)} \mathbb{E}\left[\mathbf{U}_{z,v}(X) \mid \mathcal{F}\right] = \sum_{z \in N(v)} \mathbf{R}(X, z) + \zeta, \tag{107}$$

where $|\zeta| \leq e^{5e}$. Furthermore, from Lemma 16 we have that for every $w \in V$ and every $\theta > 0$, there exists $C_0 > 0$ such that

$$\mathbf{Pr}\left[|\mathbf{R}(X, w) - \omega^*(w)| \leq \theta\right] \geq 1 - \exp\left(-\Delta/C_0\right). \tag{108}$$

From (108), (107) and a simple union bound we get the following: for every $\gamma' > 0$, the exists $C_a > 0$ such that

$$\mathbf{Pr}\left[\left|\mathbb{E}\left[S_v \mid \mathcal{F}\right] - \sum_{z \in N(v)} \omega^*(z)\right| \leq \gamma' \Delta\right] \geq 1 - \exp\left(-\Delta/C_a\right). \tag{109}$$

The theorem follows by combining (106) and (109). □

E Rapid Mixing for Glauber dynamics: Proof of Theorem 1

The following lemma considers a worst case of neighboring independent sets. It states some upper bounds on the Hamming distance after Cn and $Cn \log \Delta$ steps in the coupling.

Lemma 31. *For $\delta > 0$, $0 < \epsilon < 1$ and $C > 10$ let $\Delta \geq \Delta_0$. Consider a graph $G = (V, E)$ of maximum degree Δ and let $\lambda \leq (1 - \delta)\lambda_c(\Delta)$. Let $(X_t), (Y_t)$ be the Glauber dynamics on the hard-core model with fugacity λ and underlying graphs G. Assume that the two chains are maximally coupled. Then, the following is true:*

Assume X_0, Y_0 to be such that $X_0 \oplus Y_0 = \{v\}$ and $T = Cn/\epsilon$. Then it holds that

1. $\mathbb{E}\left[|X_T \oplus Y_T|\right] \leq \exp\left(3C/\epsilon\right)$
2. $\mathbb{E}\left[|X_{T \log \Delta} \oplus Y_{T \log \Delta}|\right] \leq \Delta^{3C/\epsilon}$
3. *Let $\mathcal{E}_T$ be the event that at some time $t \leq T$, $|X_t \oplus Y_t| > \Delta^{2/3}$. Then*
$$\mathbb{E}\left[|X_T \oplus Y_T| \cdot \mathbf{1}\{\mathbf{E}_T\}\right] < \exp\left(-\sqrt{\Delta}\right).$$
4. *Let $S_{T \log \Delta}$ denote the set of disagreements of $(X_{T \log \Delta}, Y_{T \log \Delta})$, that are 200-suspect for radius $2\Delta^{3/5}$. Then $\mathbb{E}\left[|S_{T \log \Delta}|\right] \leq \exp\left(-\sqrt{\Delta}\right)$.*

The proof of Lemma 31 appears in Section E.1.

The above lemma, i.e. Lemma 31.4, shows that from a worst case pair of neighboring independent sets, after $O(n \log \Delta)$ steps, all the disagreements are likely to be "nice" in the sense of being above suspicion. The heart of rapid mixing proof will be the following results, which shows that for a pair of neighboring independent sets that are "nice" there is a coupling of $O(n)$ steps of the Glauber dynamics where the expected Hamming distance decreases. Also, at the end of this $O(n)$ step coupling, it is extremely unlikely that there are any disagreements that are not "nice".

Lemma 32. *Let $C' > 10$, $\epsilon > 0$ and $\Delta_0 = \Delta_0(\epsilon)$. For any graph $G = (V, E)$ on n vertices and maximum degree $\Delta > \Delta_0$ and girth $g \geq 7$ the following holds:*

Suppose that X_0, Y_0 differ only at v, while v is 400-above suspicion for R, where $\Delta^{3/5} \leq R \leq 2\Delta^{3/5}$. For $T_m = C'n/\epsilon$ we have that

1. $\mathbb{E}\left[|X_{T_m} \oplus Y_{T_m}|\right] \leq 1/3$

2. *Let $\mathcal{Z}$ denote the event that there exists a 200-suspect disagreement for $R' = R - 2\sqrt{\Delta}$ at time T_m. Then it holds*
$$\mathbf{Pr}\left[\mathcal{Z}\right] \leq 2\exp(-2\sqrt{\Delta}).$$

The proof of Lemma 32 appears in Section E.2.

Proof of Theorem 1. The proof of the theorem is very similar to the proof of Theorem 1 in [4]. In particular, Lemma 31 is analogous to Lemma 10 in [4]. Similarly, Lemma 32 is analogous to Theorem 11 in [4]. Furthermore, working as for Lemma 12 in [4] we get the following result, which ties together Lemma 32 and Lemma 31. It shows that for a worst case initial pairs of independent sets, after $O(n\log\Delta)$ steps, the expected Hamming distance is small.

Let $C' > 10$, $\epsilon > 0$ and $\Delta_0 = \Delta_0(\epsilon)$. For any graph $G = (V, E)$ on n vertices and maximum degree $\Delta > \Delta_0$ and girth $g \geq 7$ the following holds: Let X_0, Y_0 be independent sets which disagree on a single vertex v that is 400-above suspicion for radius $R = 2\Delta^{3/5}$. Let $T = \frac{C' n \log\Delta}{\epsilon}$. Then,
$$\mathbb{E}\left[|X_T \oplus Y_T|\right] \leq 1/\sqrt{\Delta}.$$

In light of the above result, Theorem 1 follows using the same arguments as those for the proof of Theorem 1 in [4].

□

E.1 Proof of Lemma 31

Proof of Lemma 31.1 and 31.2. The treatment for both cases are very similar. Note that each vertex can only become disagreeing at time step t, if it is updated at time t and it is next to a vertex which is also disagreeing. Furthermore, for such vertex the probability to become disagreeing is at most e/Δ. Using the observations and noting that each disagreeing vertex has at most Δ non-disagreeing neighbors we get the following: The expected number of disagreements at each time step increases by a factor which is at most $(1 + \Delta\frac{e}{n\Delta}) \leq \exp(3/n)$.

By using induction, it is straightforward that for any $t \geq 0$ it holds
$$\mathbb{E}\left[X_t \oplus Y_t\right] \leq \exp\left(3t/n\right). \tag{110}$$

Then, the statement 1, follows by plugging into (110) $t = Cn/\epsilon$. The statement 2 follows by plugging into (110) $t = T\log\Delta$. □

Proof of Lemma 31.3. Recall that for any X_t, Y_t, we have that $D_t = \{w : X_t \neq Y_t\}$, while let
$$D_{\leq t} = \bigcup_{t' \leq t} D_{t'}.$$

Also, let $H_{\leq t} = |D_{\leq t}|$. We prove that for any integer $1 \leq \ell \leq n$, for $T = Cn/\epsilon$, it holds that
$$\mathbf{Pr}\left[H_{\leq T} \geq \ell\right] \leq \exp\left(-\ell e^{-6C/\epsilon}\right). \tag{111}$$

For $1 \leq i \leq \ell$, let t_i be the time at which the i'th disagreement is generated (possibly counting the same vertex set multiple times). Denote $t_0 = 0$. Let $\eta_i := t_i - t_{i-1}$ be the waiting time for the formation of the i'th disagreement. Conditioned on the evolution at all times in $[0, t_i]$, the distribution of η_i stochastically dominates a geometric distribution with success probability ρ_i and range $\{1, 2, \ldots\}$, where
$$\rho_i = \frac{e \cdot \min\{i\Delta, n - i\}}{n\Delta}.$$

This is because at all times prior to t_i we have $H_t \leq i$, while the sets $H_{\leq t}$ increases with probability at most ρ_i at each time step, regardless of the history. The quantity $\min\{i\Delta, n-i\}$ in the numerator in the expression for ρ_i is an upper bound on the number of vertices that are non-disagreeing neighbors of the disagreeing vertices. The quantity $e/(n\Delta)$ is an upper bound for the probability of a neighbor of a disagreement to be chosen and become a disagreement itself.

Hence, $\eta_1 + \cdots + \eta_\ell$ stochastically dominates the sum of independent geometrically distributed random variables with success probability $\rho_1, \ldots \rho_\ell$. For any real $x \geq 0$ it holds that

$$\mathbf{Pr}\left[\eta_i \geq x\right] \geq (1-\rho_i)^{\lceil x \rceil - 1} \geq \exp\left[-\frac{\rho_i}{1-\rho_i}x\right] \geq e^{2\rho_i x}.$$

In the above series of inequalities we used that $1 - x > \exp(-\frac{x}{1-x})$ for $0 < x < 1$ and $\rho_i < 1/3$.

The above inequality implies that $\eta_1 + \cdots + \eta_\ell$ dominates the sum of exponential random variables with parameters $2\rho_1, 2\rho_2, \ldots, 2\rho_\ell$. Since $\rho_i \leq i\rho$, where $\rho = \frac{e}{n}$, we have that $\eta_1 + \cdots + \eta_\ell$ stochastically dominates the sum of exponential random variables $\zeta_1, \zeta_2, \ldots, \zeta_\ell$ with parameters $2\rho, 4\rho, \ldots, 2\ell\rho$, respectively.

Consider the problem of collecting ℓ coupons, assuming that each coupon is generated by a Poisson process with rate 2ρ. The time interval between collecting the i'th coupon and the $i+1$'st coupon is exponentially distributed with rate $2(\ell - i)\rho$. Hence the time to collect all ℓ coupons has the same distribution as $\zeta_1 + \zeta_2 + \cdots + \zeta_\ell$. But the event that the total delay is less than T nothing but the intersection of the (independent) events that all coupons are generated in the time interval $[0, T]$. The probability of this event is

$$(1 - \exp^{-2T\rho})^\ell < \exp\left(-\ell \exp\left(-2Ce/\epsilon\right)\right).$$

The above completes the proof of (111). Then we proceed as follows:

$$\begin{aligned}
\mathbb{E}\left[|X_T \oplus Y_T| \cdot \mathbf{1}\{\mathbf{E}_T\}\right] &\leq \mathbb{E}\left[H_{\leq T}\mathbf{1}\{\mathcal{E}_T\}\right] \leq \sum_{\ell=\Delta^{2/3}}^{n} \ell \cdot \mathbf{Pr}\left[H_{\leq T} = \ell\right] \\
&\leq \Delta^{2/3} \cdot \mathbf{Pr}\left[H_{\leq T} \geq \Delta^{2/3}\right] + \sum_{\ell=\Delta^{2/3}+1}^{n} \mathbf{Pr}\left[H_{\leq T} \geq \ell\right] \\
&< \Delta^{2/3} \sum_{\ell=\Delta^{2/3}}^{n} \mathbf{Pr}\left[H_{\leq T} \geq \ell\right] \\
&< \Delta^{2/3} \sum_{\ell=\Delta^{2/3}}^{n} \exp\left(-\ell \exp\left(-6C/\epsilon\right)\right) \qquad \text{[from (111)]} \\
&\leq 2\Delta^{2/3} \exp(-\Delta^{2/3} e^{-6C/\epsilon}) \qquad (112)
\end{aligned}$$

Note that the above quantity is at most $\exp\left(-\sqrt{\Delta}\right)$, for large Δ. This completes the proof. ☐

Proof of Lemma 31.4. For this proof we need to use Lemma 22. We consider the contribution to the expectation $\mathbb{E}\left[|S_{T \log \Delta}|\right]$ from the vertices inside the ball $B_R(v)$ and the vertices outside the ball, i.e. $V \setminus B_R(v)$, where $R = \sqrt{\Delta}$.

First consider the vertices in $B_R(v)$. Lemma 22 implies that for some vertex $w \in B_R(v)$ at time $T' = T \log \Delta \leq \exp(\Delta/C)$ is 50-above suspicion for radius $2\Delta^{3/5}$ with probability at least $1 - \exp(-\Delta/C)$. This observation implies that

$$\mathbb{E}\left[|S_{T \log \Delta} \cap B_R(v)|\right] \leq \exp(-\Delta/C)|B_R(v)| \leq \exp\left(-4\sqrt{\Delta}\right). \qquad (113)$$

To bound the number of disagreements outside $B_R(v)$, we observe that each such disagreement comes from a path of disagreements which starts from v. Such a path of disagreements is of length at least R. This observation implies that $\mathbb{E}\left[|S_{T\log\Delta}\cap\bar{B}_R(v)|\right]$ is upper bounded by the expected number of disagreements that start from v and have length at least R.

Note that there are at most Δ^ℓ many paths of disagreement of length ℓ that start from v. Furthermore, so as a fixed path of length ℓ to become path of disagreement up to time $T\log\Delta$, there should be ℓ updates which turn its vertices into disagreeing. Each vertex is chosen to be updated with probability $1/n$, while it becomes disagreeing with probability at most e/Δ.

All the above imply that

$$\begin{aligned}
\mathbb{E}\left[|S_{T\log\Delta}\cap\bar{B}_R(v)|\right] &\leq \sum_{\ell\geq R}\Delta^\ell\binom{T\log\Delta}{\ell}\left(\frac{e}{n\Delta}\right)^\ell \\
&\leq \sum_{\ell\geq R}\left(\frac{e^2T\log\Delta}{\ell n}\right)^\ell \qquad\qquad [\text{as } \tbinom{n}{s}\leq(ne/s)^s\ ] \\
&\leq \sum_{\ell\geq R}\left(\frac{e^2C\log\Delta}{\ell\epsilon}\right)^\ell \\
&\leq (1/20)^{\sqrt{\Delta}}\leq\exp\left(-10\sqrt{\Delta}\right). \qquad (114)
\end{aligned}$$

Summing the bound of $\mathbb{E}\left[|S_{T\log\Delta}\cap B_R(v)|\right]$ and $\mathbb{E}\left[|S_{T\log\Delta}\cap\bar{B}_R(v)|\right]$ from (113) and (114), respectively gives the desired bound for $\mathbb{E}\left[|S_{T\log\Delta}|\right]$. ☐

E.2 Proof of Lemma 32

Fix v and R as specified in the statement of the theorem. Recall, for X_t,Y_t we let $D_t=\{w: X_t\oplus Y_t\}$ and denote $H(X_t,Y_t)=|D_t|$. That is, $H(X_t,Y_t)$ is the Hamming distance between X_t,Y_t. We let the accumulative difference be

$$D_{\leq t}=\bigcup\nolimits_{t'\leq t}D_t.$$

Also, let $H_{\leq t}=|D_{\leq t}|$. We define the distance between the two chains X_t,Y_t as follows

$$\mathcal{D}(X_t,Y_t)=\sum_{v\in X_t\oplus Y_t}\Phi(v),$$

where $\Phi: V\to[1,12]$ is defined in Theorem 6. The metric $\mathcal{D}(X_t,Y_t)$ generalizes the Hamming metric in the following sense: the disagreement in each vertex v instead of contributing one it contributes $\Phi(v)$. Since $\Phi(v)\geq 1$, for every $v\in V$, for any two X_t,Y_t we always have

$$\mathcal{D}(X_t,Y_t)\geq H(X_t,Y_t). \qquad (115)$$

For proving the lemma we use the following result.

Lemma 33. *For $\delta>0$, let sufficiently small $\epsilon=\epsilon(\delta)$ and $\Delta\geq\Delta_0$. Consider a graph $G=(V,E)$ of maximum degree Δ and let $\lambda\leq(1-\delta)\lambda_c(\Delta)$. Also, let $(X_t),(Y_t)$ be the Glauber dynamics on the hard-core model with fugacity λ and underlying graphs G.*

For some time t, assume that $X_t\oplus Y_t=\{v\}$, for some $v\in V$ such that

$$W_{X_t}(v)\leq\sum_{z\in N(v)}\omega^*(z)\cdot\Phi(z)+\epsilon\Delta, \qquad (116)$$

$W_{X_t}(v)$ is defined in (10). Then, coupling the chains maximally we have that

$$\mathbb{E}\left[\mathcal{D}(X_{t+1}, Y_{t+1}) - \mathcal{D}(X_t, Y_t)\right] < -c/n,$$

for appropriate $c = c(\epsilon, \delta) > 0$.

The proof of Lemma 33 appears in Section E.3.

We start by proving statement 1 of Lemma 32.

Proof of Lemma 32.1. Let

$$T_b = \max\{C_b n, C_a n\},$$

where the quantities C_b, C_a are from Lemma 23 and Theorem 27, respectively.

Since $T_m \leq n \exp\left(\Delta/(C \log \Delta)\right)$, we can apply Theorem 27 to conclude that the desired local uniformity properties holds with high probability for all $t \in I := [T_b, T_m]$.

For $t \geq T_b$ we define the following *bad* events:

- $\mathcal{E}(t)$ denotes the event that at some time $s < t$, it holds $H_s > \Delta^{2/3}$
- $\mathcal{B}_1(t)$ denotes the event that $D_{\leq t} \not\subseteq B_{\sqrt{\Delta}}(v)$
- $\mathcal{B}_2(t)$ denotes the event that there exists a time $T_b \leq \tau \leq t$, $z \in B_{\sqrt{\Delta}}(v)$ such that

$$W_{X_t}(z) > \Theta(z, \epsilon) = \sum_{z \in N(v)} \omega^*(z)\Phi(z) + \epsilon\Delta,$$

  where $\omega^* \in [0,1]^V$ is defined in Lemma 4 and $\Phi : V \to [1, 12]$. is defined in Theorem 6

Also, we let the event

$$\mathcal{B}(t) = \mathcal{B}_1(t) \cup \mathcal{B}_2(t),$$

while we let the "good" event

$$\mathcal{G}(t) = \bar{\mathcal{E}}(t) \cap \bar{\mathcal{B}}(t).$$

We follow the convention that we drop the time t, for all the above events when we are referring to the event at time T_m.

We bound the Hamming distance by conditioning on the above event in the following manner,

$$\begin{aligned}
\mathbb{E}\left[H_{T_m}\right] &= \mathbb{E}\left[H_{T_m}\mathbf{1}\{\mathcal{E}\}\right] + \mathbb{E}\left[H_{T_m}\mathbf{1}\{\bar{\mathcal{E}}\}\mathbf{1}\{\mathcal{B}\}\right] + \mathbb{E}\left[H_{T_m}\mathbf{1}\{\bar{\mathcal{E}}\}\mathbf{1}\{\bar{\mathcal{B}}\}\right] \\
&\leq \mathbb{E}\left[H_{T_m}\mathbf{1}\{\mathcal{E}\}\right] + \Delta^{2/3}\mathbf{Pr}\left[\mathcal{B}\right] + \mathbb{E}\left[H_{T_m}\mathbf{1}\{\mathcal{G}\}\right] \\
&\leq \exp(-\sqrt{\Delta}) + \Delta^{2/3}\mathbf{Pr}\left[\mathcal{B}\right] + \mathbb{E}\left[H_{T_m}\mathbf{1}\{\mathcal{G}\}\right], \qquad (117)
\end{aligned}$$

where in the last inequality we used Lemma 31.3.

For the second term in the (117) we prove the following

$$\mathbf{Pr}\left[\mathcal{B}\right] \leq \exp\left(-\sqrt{\Delta}\right). \qquad (118)$$

Finally, for the third term in the (117) we prove the following

$$\mathbb{E}\left[H_{T_m}\mathbf{1}\{\mathcal{G}\}\right] \leq 1/9. \qquad (119)$$

The part 1 of the theorem follows by plugging into (117), the bounds in (118) and (119). □

Proof of (118). We can bound the probability of the event $\mathcal{B}_1$ by a standard paths of disagreement argument. We are looking at the probability of a path of disagreement of length $\ell = \sqrt{\Delta}$, within $T_m = C'n/\epsilon$ steps, hence:

$$\begin{aligned}\mathbf{Pr}\left[\mathcal{B}_1\right] &\leq \Delta^{\ell}\binom{T_m}{\ell}\left(\frac{e}{n\Delta}\right)^{\ell} \\ &\leq \left(e^2C'/\epsilon\right)^{\ell} \qquad\qquad [\text{as } \tbinom{N}{i} \leq (Ne/i)^i] \\ &\leq \exp\left(-2\sqrt{\Delta}\right). \end{aligned} \tag{120}$$

We can bound the probability of the event $\mathcal{B}_2$ by working as follows: The assumption is that v is 400-above suspicion for radius $R \geq \Delta^{3/5}$. Then, each vertex $z \in B_{\sqrt{\Delta}(v)}$ is 400-above suspicion for the constant radius $R'(\gamma,\delta)$ required for the statement for the hypothesis of Theorem 27. Therefore, in the interval $I = [T_b, T_m]$ the uniformity condition for each vertex z fails with probability at most $\exp(-\Delta/(C\log\Delta))$. More precisely, we have that

$$\mathbf{Pr}\left[\mathcal{B}_2\right] \leq \exp(-\Delta/C)\Delta^{\sqrt{\Delta}+1} \leq \exp\left(-2\sqrt{\Delta}\right). \tag{121}$$

Using a simple union bound, we get that $\mathbf{Pr}\left[\mathcal{B}\right] \leq \mathbf{Pr}\left[\mathcal{B}_1\right] + \mathbf{Pr}\left[\mathcal{B}_2\right]$. Then (118) follows by plugging (120) and (121) into the union bound. □

Proof of (119). Recall that for the two chains X_t, Y_t we defined the following notion of distance

$$\mathcal{D}(X_t, Y_t) = \sum_{v \in X_t \oplus Y_t} \Phi(v).$$

Note that for every $z \in V$ it holds that $1 \leq \Phi(z) \leq 12$. This implies that we always have that $\mathcal{D}(X_t, Y_t) \geq H(X_t, Y_t)$. For showing that (119) indeed holds, it suffices to show that

$$\mathbb{E}\left[\mathcal{D}(X_{T_m}, Y_{T_m})\mathbf{1}\{\mathcal{G}\}\right] \leq 1/9. \tag{122}$$

Let $Q_0 = X_t, Q_1, Q_2, \ldots, Q_h = Y_t$ be a sequence of independent sets where $h = |X_t \oplus Y_t|$ and Q_{i+1} is obtained from Q_i by changing the assignment of one vertex w_i from $X_t(w_i)$ to $Y_t(w_i)$. We maximally couple W_i and W_{i+1} in one step of the Glauber dynamics to obtain W_i' and W_{i+1}'. More precisely, both chains update the spin of the same vertex and maximize the probability of choosing the same new assignment for the chosen vertex.

Consider a pair Q_i, Q_{i+1}. Note that Q_i, Q_{i+1} differ only on the assignment of w_i. With probability $1/n$ both chains update the spin of vertex w_i. Since all the neighbors of w_i have the same spin, with probability 1 we assign the same spin on w_i in both chains. Such an update reduces the distance of the two chains by $\Phi(w_i)$.

Consider now the update of vertex $w \in N(w_i)$. Also, w.l.o.g. assume that $Q_i(w_i)$ is occupied while $Q_{i+1}(w_i)$ is unoccupied. It is direct that the worst case is when w is unblocked in the chain Q_{i+1}. Otherwise, i.e. if w is blocked then with probability 1 we have $Q_{i+1}(w) = Q_i(w) =$"unoccupied", since in Q_i, we have w_i blocked.

Assuming that w_i blocked in the chain Q_i and unblocked in the chain Q_{i+1}, we get $Q_i'(w) \neq Q_{i+1}'(w)$ if the coupling chooses to set w_i occupied in Q_{i+1}'. Otherwise, we have $Q_i'(w) = Q_{i+1}'(w)$. Clearly, the disagreement happens with probability at most $\frac{\lambda}{1+\lambda} < e/\Delta$.

Therefore, given Q_i, Q_{i+1}, we have that

$$\mathbb{E}\left[\mathcal{D}(Q_{i+1}', Q_i') - \mathcal{D}(Q_{i+1}, Q_i)\right] \leq -\frac{\Phi(w_i)}{n} + \frac{e}{n\Delta}\sum_{z \in N(w_i)} \Phi(z). \tag{123}$$

Since we have that $1 \le \Phi(z) \le 12$, for any $z \in V$ and $|N(v)| \le \Delta$, we get the trivial bound that

$$\mathbb{E}\left[\mathcal{D}(Q'_{i+1}, Q'_i) - \mathcal{D}(Q_{i+1}, Q_i)\right] \le 35/n.$$

Therefore,

$$\mathbb{E}\left[\mathcal{D}(X_{t+1}, Y_{t+1})\right] \le (1 + 35/n)\, \mathcal{D}(X_t, Y_t). \tag{124}$$

The above bound is going to be used only for the burn-in phase, i.e. the first T_b steps. We use a significantly better bound for the remaining $T_m - T_b$ steps.

Since the event $\mathcal{G}$ holds, for all $0 \le i \le h$, $z \in B_R(v)$ and all $t \in [T_b, T_m - 1]$, we have that

$$W(Q_i, z) \le \Theta(z, \epsilon) + \Delta^{2/3} \le \Theta(z, 2\epsilon). \tag{125}$$

The first inequality follows from our assumption that both event $\bar{\mathcal{E}}$ and $\bar{\mathcal{B}}_2$ occur. The second follows from the definition of the quantity Θ.

Using Lemma 33 and get the following: For Q_i, Q_{i+1} which satisfy (125) it holds that

$$\mathbb{E}\left[\mathcal{D}(Q'_{i+1}, Q'_i)\right] \le \left(1 - C'/n\right) \mathcal{D}(Q_{i+1}, Q_i),$$

for appropriately chosen C'. The above inequality implies the following: Given X_t, Y_t and assuming that $\mathcal{G}(t)$ holds, we get that

$$\mathbb{E}\left[\mathcal{D}(X_{t+1}, Y_{t+1})\right] \le (1 - C/n)\, \mathcal{D}(X_t, Y_t). \tag{126}$$

Let $t \in [T_b, T_m - 1]$. Then we have that

$$\begin{aligned}
\mathbb{E}\left[\mathcal{D}(X_{t+1}, Y_{t+1})\mathbf{1}\{\mathcal{G}(t)\}\right] &= \mathbb{E}\left[\mathbb{E}\left[\mathcal{D}(X_{t+1}, Y_{t+1})\mathbf{1}\{\mathcal{G}(t)\} \mid X_0, Y_0, \ldots, X_t, Y_t\right]\right] \\
&= \mathbb{E}\left[\mathbb{E}\left[\mathcal{D}(X_{t+1}, Y_{t+1}) \mid X_0, Y_0, \ldots, X_t, Y_t\right]\mathbf{1}\{\mathcal{G}(t)\}\right] \\
&\le (1 - C/n)\, \mathbb{E}\left[\mathcal{D}(X_t, Y_t)\mathbf{1}\{\mathcal{G}(t)\}\right] \\
&\le (1 - C/n)\, \mathbb{E}\left[\mathcal{D}(X_t, Y_t)\mathbf{1}\{\mathcal{G}(t-1)\}\right].
\end{aligned}$$

The frist equality is Fubini's Theorem, the second equality is due to the fact that $X_0, Y_0, \ldots X_t, Y_t$ determine uniquely $\mathcal{G}(t)$ The first inequality uses (126) while the second inequality uses the fact that $\mathcal{G}(t) \subset \mathcal{G}(t-1)$. By induction, it follows that

$$\mathbb{E}\left[\mathcal{D}(X_{T_m}, Y_{T_m})\mathbf{1}\{\mathcal{G}(T_m)\}\right] \le (1 - C/n)^{T_m - T_b}\, \mathbb{E}\left[\mathcal{D}(X_{T_b}, Y_{T_b})\mathbf{1}\{\mathcal{G}(T_b)\}\right].$$

Using the same arguments and (124) for $\mathbb{E}\left[\mathcal{D}(X_{T_b}, Y_{T_b})\mathbf{1}\{\mathcal{G}(T_b)\}\right]$ we get that

$$\mathbb{E}\left[\mathcal{D}(X_{T_m}, Y_{T_m})\mathbf{1}\{\mathcal{G}(T_m)\}\right] \le (1 - C/n)^{T_m - T_b}\, (1 + 35/n)^{T_b}\, \mathcal{D}(X_0, Y_0). \tag{127}$$

The result follows from the choice of constants and noting that $\mathcal{D}(X_0, Y_0) < 12$. ∎

Proof of Lemma 32.2. Recall from the proof of Lemma 32.1 that $\mathcal{B}_1$ is the event that $D_{\le T_m} \not\subseteq B_{\sqrt{\Delta}}(v)$. Also consider $\mathcal{B}'_1$ to be the event that $D_{T_m} \not\subseteq B_{\sqrt{\Delta}(v)}$. Noting that $\mathcal{B}'_1 \subset \mathcal{B}_1$, we get that

$$\mathbf{Pr}\left[\mathcal{B}'_1\right] \le \mathbf{Pr}\left[\mathcal{B}_1\right] \le \exp\left(-\sqrt{\Delta}\right),$$

where the last inequality follows from (120).

We can assume the disagreements are contained in $B_{\sqrt{\Delta}}(v)$. By the hypothesis of Lemma 32, each vertex $w \in B_{\sqrt{\Delta}}(v)$ is 400-above suspicion for radius $R - \sqrt{\Delta}$ in both X_0 and Y_0. Therefore, by Lemma 23, each vertex $w \in B_{\sqrt{\Delta}}(v)$ is 20-above suspicion for radius $R - \sqrt{\Delta} - 2$ in X_{T_m} and Y_{T_m} with probability at least $1 - \exp(-\Delta/C_b)$. Therefore, all $w \in B_{\sqrt{\Delta}}(v)$ is 50-above suspicion for radius $R - \sqrt{\Delta} - 2$ in X_{T_m} and Y_{T_m} with probability at least $1 - \exp(-\Delta/C_b)$. That is, we have proven that all disagreements between X_{T_m} and Y_{T_m} are 50-above suspicion for radius $R - \sqrt{\Delta} - 2$ with probability at least $1 - 2\exp(-\Delta/C_b)$. This proves Lemma 32.2. ∎

E.3 Proof of Lemma 33

Proof of Lemma 33. Let $\Phi_{\max} = \max_{z\in V} \Phi(z)$, where $\Phi : V(G) \to \mathbb{R}_{\geq 0}$, as in Theorem 6. Each vertex $v \in V$ is called a "low degree vertex" if $\deg(v) \leq \hat{\Delta} = \frac{\Delta}{e\cdot\Phi_{\max}}$.

If v is a low degree vertex then the following holds

$$\mathbb{E}\left[\mathcal{D}(X_{t+1}, Y_{t+1}) - \mathcal{D}(X_t, Y_t)\right] \leq -\frac{\Phi(v)}{n} + \frac{1}{n}\sum_{z\in N(v)} \frac{\lambda}{1+\lambda}\Phi(z).$$

We get the inequality above by working as follows: The distance between the two chains changes when we updated either v or some vertex $z \in N(v)$.

With probability $1/n$ the the update involves the vertex v. Since there is no disagreement at the neighborhood of v we can couple X_t and Y_t such that $X_{t+1}(v) = Y_{t+1}(v)$ with probability 1. That is, the distance between the chain decreases by $\Phi(v)$.

We make the (worst case) assumption that all the vertices in $N(v)$ are unblocked and unoccupied. We have a new disagreement between the two chains, i.e. an increase in the distance, only if some vertex $z \in N(v)$ is chosen to be updated and one of the chains sets z occupied. Since $X_t(v) \neq Y_t(v)$ one of the chains cannot set z occupied. Each $z \in N(v)$ is chosen with probability $1/n$ and it is set occupied by one the two chains with probability $\frac{\lambda}{1+\lambda}$. Then, the distance between the chains increases by $\Phi(z)$. Then we get the following

$$\begin{aligned}
\mathbb{E}\left[\mathcal{D}(X_{t+1}, Y_{t+1}) - \mathcal{D}(X_t, Y_t)\right] &\leq -\frac{\Phi(v)}{n} + \frac{1}{n}\sum_{z\in N(v)} \frac{\lambda}{1+\lambda}\Phi(z) \\
&\leq -\frac{1}{n}\left(\Phi(v) - \Phi_{\max}\cdot(1-\delta)\lambda_c(\Delta)\cdot\hat{\Delta}\right) \\
&\leq -\frac{1}{n}\left(\Phi(v) - 1\right) \leq -10/n, \qquad (128)
\end{aligned}$$

where the last inequality follows from the fact that $1 \leq \Phi(v) \leq 12$, for every $v \in V$, $\hat{\Delta} = \frac{\Delta}{e\cdot\Phi_{\max}}$ and $\lambda \leq e/\Delta$. For the case where v is a high degree vertex we have the following

$$\mathbb{E}\left[\mathcal{D}(X_{t+1}, Y_{t+1}) - \mathcal{D}(X_t, Y_t)\right] \leq -\frac{\Phi(v)}{n} + \frac{1}{n}\sum_{z\in N(v)} \frac{\lambda}{1+\lambda}\omega^*(z)\Phi(z) + \frac{1}{n}\frac{\lambda}{1+\lambda}\epsilon\Delta.$$

As before, the interesting cases are those where the update involves the vertex v or $N(v)$. As we argued above when the vertex v is updated the distance between the two chains decreases by $\Phi(v)$.

As far as the neighbors of v are regarded we observe the following: If some $z \in N(v)$ is blocked, then with probability 1 is set unoccupied in both chains. This means that $X_{t+1}(z) = Z_{t+1}(z)$, i.e. the distance between the two chains remains unchanged. If the update involves an unblocked vertex $z \in N(v)$, then with probability $\frac{\lambda}{1+\lambda}$ the vertex z becomes occupied at only one of the two chains and the distance between the chains increases by $\Phi(z)$.

In the inequality above, we use also use the fact that (116) holds for the high degree vertex v. Then we get that

$$\begin{aligned}
\mathbb{E}\left[\mathcal{D}(X_{t+1}, Y_{t+1}) - \mathcal{D}(X_t, Y_t)\right] &\leq -\frac{\Phi(v)}{n} + \frac{1}{n}\frac{\lambda}{1+\lambda}\mathbf{W}_\sigma(v) \\
&\leq -\frac{\Phi(v)}{n} + \frac{1}{n}\sum_{z\in N(v)} \frac{\lambda}{1+\lambda}\omega^*(z,v)\Phi(z) + \frac{1}{n}\frac{\lambda}{1+\lambda}\epsilon\Delta. \\
&\leq -\frac{1}{n}\left(\Phi(v) - \sum_{z\in N(v)} \frac{\lambda}{1+\lambda}\omega^*(z,v)\Phi(z) + e\epsilon\right) \leq -c/n, \qquad (129)
\end{aligned}$$

where the last inequality follows by taking sufficiently small $\epsilon > 0$.

The lemma follows from (128) and (129). □

F Random Regular (Bipartite) Graphs: Proof of Theorem 2

It turns out that the girth restriction of Theorem 1 can be relaxed a bit. The main technical reason why we need girth at least 7 is for establishing what we call "local uniformity property". Roughly speaking, local uniformity amounts to showing that the number of unblocked neighbors of a vertex v is concentrated about the quantity $\sum_{z\in N(v)} \omega^*(z)$, where $\omega^* \in [0,1]^V$ is the fixed points of a BP-like system of equations. In particular, uniformity amounts to showing that the number of unblocked neighbors of v is $\sum_{w\in N(v)} \omega^*(w) \pm \epsilon\Delta$, with probability that tends to 1 as Δ grows.

The analysis of local uniformity could be carried out for graph with short cycles, i.e. cycles of length less than 7. The effect of the short cycles is an increase to the fluctuation of the number of unblocked neighbors of a vertex. However, if the number of such cycles is small, i.e. constant, then the increase in the fluctuation is negligible. That is, the proof of Theorem 1 carries out if, instead of girth at least 7, we have smaller girth but only a *constant* number of cycles of length less than 7 around each vertex v. The above observation leads to the following corollary from Theorem 1.

For some integers $\ell, g \geq 0$, let $\mathcal{G}_n(\ell, g)$ denote all the graphs on n vertices such that each vertex belongs to at most ℓ cycles of length less than g.

Corollary 34. *For all $\delta > 0$, there exists $\Delta_0 = \Delta_0(\delta)$, $\ell = \ell(\delta)$ and $C = C(\delta)$, for all $\Delta \geq \Delta_0$, all $\lambda < (1-\delta)\lambda_c(T_\Delta)$, all graphs $G \in \mathcal{G}(\ell, 7)$ of maximum degree Δ, all $\epsilon > 0$, the mixing time of the Glauber dynamics satisfies:*

$$T_{\mathrm{mix}}(\epsilon) \leq Cn\log(n/\epsilon).$$

Using the above corollary we can show the following rapid mixing result for random regular (bipartite) graphs with sufficiently large degree Δ. The theorem follows by using e.g. the result from [42]. Let G be chosen uniformly at random among all Δ regular (bipartite) graphs with n. Then, with probability that tends to 1 as n tends to infinity it holds that $G \in \mathcal{G}(1, 7)$. Then the theorem follows from Corollary 34.