

# Slow mixing of Glauber Dynamics for the hard-core model on regular bipartite graphs

David Galvin\*  
Prasad Tetali†

Appeared 2006

## Abstract

Let  $\Sigma = (V, E)$  be a finite,  $d$ -regular bipartite graph. For any  $\lambda > 0$  let  $\pi_\lambda$  be the probability measure on the independent sets of  $\Sigma$  in which the set  $I$  is chosen with probability proportional to  $\lambda^{|I|}$  ( $\pi_\lambda$  is the *hard-core measure with activity  $\lambda$  on  $\Sigma$* ). We study the Glauber dynamics, or single-site update Markov chain, whose stationary distribution is  $\pi_\lambda$ . We show that when  $\lambda$  is large enough (as a function of  $d$  and the expansion of subsets of single-parity of  $V$ ) then the convergence to stationarity is exponentially slow in  $|V(\Sigma)|$ . In particular, if  $\Sigma$  is the  $d$ -dimensional hypercube  $\{0, 1\}^d$  we show that for values of  $\lambda$  tending to 0 as  $d$  grows, the convergence to stationarity is exponentially slow in the volume of the cube. The proof combines a conductance argument with combinatorial enumeration methods.

## 1 Introduction and statement of the result

Let  $\Sigma = (V, E)$  be a simple, loopless, finite graph on vertex set  $V$  and edge set  $E$ . (For graph theory basics, see e.g. [3], [7].) Write  $\mathcal{I}(\Sigma)$  for the set of independent sets (sets of vertices spanning no edges) in  $V$ . For  $\lambda > 0$  we define the *hard-core measure with activity  $\lambda$  on  $\mathcal{I}(\Sigma)$*  by

$$\pi_\lambda(\{I\}) = \frac{\lambda^{|I|}}{Z_\lambda(\Sigma)} \quad \text{for } I \in \mathcal{I} \quad (1)$$

---

Key words: Mixing time, hard-core model, conductance, Glauber dynamics, discrete hypercube.

\*Institute for Advanced Study, Einstein Drive, Princeton, NJ 08540; galvin@ias.edu. Research supported in part by NSF grant DMS-0111298.

†School of Mathematics & College of Computing, Georgia Institute of Technology, Atlanta, GA 30332-0160. Research supported in part by NSF grant DMS-0100289.

where  $Z_\lambda(\Sigma) = \sum_{I \in \mathcal{I}} \lambda^{|I|}$  is the appropriate normalizing constant. We will often write  $w_\lambda(I)$  for  $\lambda^{|I|}$  and, for  $\mathcal{J} \subseteq \mathcal{I}$ ,  $w_\lambda(\mathcal{J})$  for  $\sum_{J \in \mathcal{J}} w_\lambda(J)$ .

The hard-core measure originally arose in statistical physics (see e.g. [8, 1]) where it serves as a simple mathematical model of a gas with particles of non-negligible size. The vertices of  $\Sigma$  we think of as sites that may or may not be occupied by particles; the rule of occupation is that adjacent sites may not be simultaneously occupied. The activity parameter  $\lambda$  measures the likelihood of a site being occupied.

The measure also has a natural interpretation in the context of communications networks (see e.g. [14]). Here the vertices of  $\Sigma$  are thought of as locations from which “calls” can be made; when a call is made, the call location is connected to all its neighbours, and throughout its duration, no call may be placed from any of the neighbours. Thus at any given time, the collection of locations from which calls are being made is exactly an independent set in the graph. If calls are attempted independently at each vertex as a Poisson process of rate  $\lambda$  and have independent exponential mean 1 lengths, it can be shown that the long-run stationary distribution of this process is the hard-core measure on  $\Sigma$ .

Our particular focus in this paper is the mixing time of the Glauber dynamics, or single-site update Markov chain, for this model. The measure  $\pi_\lambda$  can be realized as the stationary distribution of a certain Markov chain. Specifically, consider the chain  $\mathcal{M}_\lambda = \mathcal{M}_\lambda(\Sigma)$  on state space  $\mathcal{I}(\Sigma)$  with transition probabilities  $P_\lambda(I, J)$ ,  $I, J \in \mathcal{I}(\Sigma)$ , given by

$$P_\lambda(I, J) = \begin{cases} 0 & \text{if } |I \Delta J| > 1 \\ \frac{1}{|V|} \frac{\lambda}{1+\lambda} & \text{if } |I \Delta J| = 1, I \subseteq J \\ \frac{1}{|V|} \frac{1}{1+\lambda} & \text{if } |I \Delta J| = 1, J \subseteq I \\ 1 - \sum_{I \neq J' \in \mathcal{I}(\Sigma)} P_\lambda(I, J') & \text{if } I = J \end{cases}$$

Underpinning the definition of  $\mathcal{M}_\lambda$  is the following dynamical process, known as the *Glauber dynamics* on  $\mathcal{I}(\Sigma)$ . From an independent set  $I$ , the process follows three steps. The first step is to choose a vertex  $v$  uniformly from  $V$ . The second step is to “add”  $v$  to  $I$  with probability  $\lambda/(1 + \lambda)$ , and “remove” it with probability  $1/(1 + \lambda)$ ; that is, to set

$$I' = \begin{cases} I \cup \{v\} & \text{with probability } \frac{\lambda}{1+\lambda} \\ I \setminus \{v\} & \text{with probability } \frac{1}{1+\lambda}. \end{cases}$$

The third step is to move to  $I'$  if it is a valid independent set, and stay at  $I$  otherwise.

It is readily checked that  $\mathcal{M}_\lambda$  is an ergodic, aperiodic, time reversible Markov chain with (unique) stationary distribution  $\pi_\lambda$ . A natural question to ask about  $\mathcal{M}_\lambda$  is how quickly it converges to its stationary distribution. To make this question precise, we need a few definitions.

Let  $\mathcal{M}$  be an ergodic Markov chain on state space  $\Omega$ , with transition probabilities  $P : \Omega^2 \rightarrow [0, 1]$ . For a state  $\omega_0 \in \Omega$ , denote by  $P^t(\omega_0, \cdot)$  the distribution of the state at time  $t$ , given that the initial state is  $\omega_0$ , and denote by  $\pi$  the stationary distribution.

Define the *mixing time* of  $\mathcal{M}$  by

$$\tau_{\mathcal{M}} = \max_{\omega_0 \in \Omega} \min \left\{ t_0 : \frac{1}{2} \sum_{\omega \in \Omega} |P^t(\omega_0, \omega) - \pi(\omega)| \leq \frac{1}{e} \quad \forall t > t_0 \right\}.$$

The mixing time of  $\mathcal{M}$  captures the speed at which the chain converges to its stationary distribution: for every  $\epsilon > 0$ , in order to get a sample from  $\Omega$  which is within  $\epsilon$  of  $\pi$  (in variation distance), it is necessary and sufficient to run the chain from some arbitrarily chosen distribution for some multiple (depending on  $\epsilon$ ) of the mixing time.

Much work has been done on the question of bounding  $\tau_{\mathcal{M}_\lambda}$ . The strongest general result available to date is due to Vigoda [21] who showed that if  $\Sigma$  is any  $n$ -vertex graph with maximum degree  $\Delta$ , then  $\tau_{\mathcal{M}_\lambda}(\Sigma) = O(n \log n)$  whenever  $\lambda < 2/(\Delta - 2)$ . In the other direction, Dyer, Frieze and Jerrum [9] considered the case  $\lambda = 1$  and showed that for each  $\Delta \geq 6$  a random (uniform)  $\Delta$ -regular,  $n$ -vertex bipartite  $\Sigma$  almost surely (with probability tending to 1 as  $n$  tends to infinity) satisfies  $\tau_{\mathcal{M}_1}(\Sigma) \geq 2^{\gamma n}$  for some absolute constant  $\gamma > 0$ .

Here, we continue in the spirit of [9] and construct *explicit* families of graphs for which Glauber dynamics mixes slowly. Specifically, we establish a certain expansion condition in a regular bipartite graph  $\Sigma$  that forces  $\tau_{\mathcal{M}_\lambda(\Sigma)}$  to be (almost) exponential in  $|V|$  provided  $\lambda$  is suitably large (as a function of the expansion). The  $d$ -dimensional hypercube  $\{0, 1\}^d$  satisfies this condition for  $\lambda \geq \omega(d^{-1/4} \log^{3/2} d)$ .

Our work is partly motivated by [5] where a study was made of Glauber dynamics for the hard-core measure on the even discrete torus  $T_{L,d}$ . This is the graph on  $\{0, \dots, L-1\}^d$  (with  $L$  even) in which two strings are adjacent if they differ on only one coordinate, and differ by 1 (mod  $L$ ) on that coordinate. It was shown in [5] that for  $\lambda$  growing exponentially in  $d$  (with a suitably large base),  $\tau_{\mathcal{M}_\lambda(T_{L,d})}$  is exponential in  $cL^{d-1}/\log^2 L$  for some  $c$  that depends on  $d$  but not on  $L$ .

In light of a recent result of Galvin and Kahn [12], we found it tempting to believe that slow mixing on  $T_{L,d}$  should hold for much smaller values of  $\lambda$ ; even for some values of  $\lambda$  tending to 0 as  $d$  grows. The main result of [12] is that the hard-core model on  $\mathbb{Z}^d$  exhibits multiple Gibbs phases for  $\lambda \geq Cd^{-1/4} \log^{3/4} d$  for some large constant  $C$ . Specifically, write  $\mathcal{E}$  and  $\mathcal{O}$  for the sets of even and odd vertices of  $\mathbb{Z}^d$  (defined in the obvious way: a vertex of  $\mathbb{Z}^d$  is *even* if the sum of its coordinates is even). Set

$$\Lambda_M = \Lambda_M^d = [-L, L]^d, \quad \partial\Lambda_M = [-L, L]^d \setminus [-(L-1), L-1]^d.$$

For  $\lambda > 0$ , choose  $\mathbf{I}$  from  $\mathcal{I}(\Lambda_M)$  with  $\Pr(\mathbf{I} = I) \propto \lambda^{|I|}$ . The main result of [12] is that there is a constant  $C$  such that if  $\lambda \geq Cd^{-1/4} \log^{3/4} d$  then

$$\lim_{M \rightarrow \infty} \Pr(\underline{0} \in \mathbf{I} | \mathbf{I} \supseteq \partial\Lambda_M \cap \mathcal{E}) > \lim_{M \rightarrow \infty} \Pr(\underline{0} \in \mathbf{I} | \mathbf{I} \supseteq \partial\Lambda_M \cap \mathcal{O}).$$

Thus, roughly speaking, the influence of the boundary on behavior at the origin persists as the boundary recedes. Informally, this suggests that for  $\lambda$  in this range, the typical independent set chosen according to the hard-core measure is either predominantly odd or predominantly even. Thus there is a highly unlikely “bottleneck” set of balanced independent sets separating the predominantly odd sets from the predominantly even ones. It is the existence of this bottleneck that should cause the conductance of the Glauber dynamics chain to be small (see Section 2), and thus cause its mixing time to be large.

Our main result (Theorem 1.1) provides some support for this belief, verifying it in the case  $L = 2$ ; unfortunately, because of the weak isoperimetry of the torus we cannot hope to use Theorem 1.1 to deal with general  $L$ . (See Remark 1.6 for further discussion of these issues.)

Before stating Theorem 1.1, we establish some notation. From now on,  $\Sigma = (V, E)$  will be a  $d$ -regular, bipartite graph with partition classes  $\mathcal{E}$  and  $\mathcal{O}$ . Set  $N = |V|$  and  $M = |\mathcal{E}| = |\mathcal{O}| (= N/2)$ .

For  $u, v \in V$  we write  $u \sim v$  if there is an edge in  $\Sigma$  joining  $u$  and  $v$ . Set  $N(u) = \{w \in V : w \sim u\}$  ( $N(u)$  is the *neighbourhood* of  $u$ ) and  $N(A) = \cup_{w \in A} N(w)$ . For  $A \subseteq \mathcal{E}$  (or  $\mathcal{O}$ ) set

$$[A] = \{x \in V(\Sigma) : N(x) \subseteq N(A)\};$$

we think of  $[A]$  as an “external closure” of  $A$ . Note that while  $A$  determines  $N(A)$ ,  $N(A)$  determines only  $[A]$ . For this reason, we find it more convenient at some points in the sequel to deal with  $[A]$  rather than with  $A$  itself. Say that  $A$  is *small* if  $|[A]| \leq M/2$ . Define the *bipartite expansion constant* of  $\Sigma$  by

$$\delta(\Sigma) = \min \left\{ \frac{|N(A)| - |[A]|}{|N(A)|} : A \subseteq \mathcal{E} \text{ small or } A \subseteq \mathcal{O} \text{ small, } A \neq \emptyset \right\}.$$

Note that  $0 \leq \delta < 1$ . (The second inequality is obvious. The first follows from regularity, which implies that  $\Sigma$  has a perfect matching, which in turn implies that for all  $A \subseteq \mathcal{E}$  (or  $\mathcal{O}$ ),  $|A| \leq |N(A)|$ .)

All implied constants in  $O$  and  $\Omega$  notation are independent of  $d$ . We use “log” throughout for  $\log_2$  and “ln” for  $\log_e$ . We write  $\exp_2 x$  for  $2^x$ . We always assume that  $d$  is sufficiently large to support our assertions.

Set

$$\alpha(\lambda) = \frac{\log(1 + \lambda)}{44(1 + \log(1 + \lambda)) \log\left(2 + \frac{1}{\log(1 + \lambda)}\right)} \quad (2)$$

and

$$\beta(\lambda) = \frac{\log^2(1 + \lambda)}{\log(1 + \lambda) + \log(d^5/\delta)}.$$

We note for future reference that

$$\lambda \geq \frac{1}{\sqrt{d}} \text{ implies } \frac{1}{44} > \alpha(\lambda) = \Omega\left(\frac{1}{\sqrt{d} \log d}\right). \quad (3)$$

Our main result is

**Theorem 1.1** *Let  $\Sigma$  be a  $d$ -regular, bipartite graph with  $N \geq d^2$  vertices and bipartite expansion constant  $\delta$ . There is a constant  $c > 0$  such that whenever  $\lambda$  and  $\delta$  satisfy*

$$\beta(\lambda) \geq c \max\left\{\frac{\log(d^5/\delta)}{\sqrt{d}}, \frac{\log^2 d}{\delta d}\right\} \quad (4)$$

we have

$$\tau_{\mathcal{M}_\lambda(\Sigma)} \geq \exp_2\{\Omega(N\alpha(\lambda)\beta(\lambda)\delta)\}.$$

**Remark 1.2** *If we add as an additional hypothesis to Theorem 1.1 that  $\Sigma$  has bounded codegree (that is, there is a constant  $\kappa$  independent of  $d$  such that each pair of vertices in  $\Sigma$  has at most  $\kappa$  common neighbours), then we can slightly improve our bound on  $\lambda$  to*

$$\beta(\lambda) \geq c \max\left\{\frac{\log(d^5/\delta)}{\sqrt{d}}, \frac{\log d}{\delta d^2}\right\}. \quad (5)$$

We do not present the more complicated argument here.

Note that since  $\delta < 1$ , we cannot possibly satisfy (4) for  $\lambda \leq 1/\sqrt{d}$ , so we may (and will) assume from here on that  $\lambda \geq 1/\sqrt{d}$ .

A slightly stronger condition that implies (4) is

$$\log(1 + \lambda) \geq c' \max\left\{\frac{\log(d^5/\delta)}{d^{1/4}}, \frac{\log d \sqrt{\log(d^5/\delta)}}{\sqrt{\delta d}}\right\} \quad (6)$$

where the constant  $c'$  depends on  $c$ , from which we can more clearly see the tradeoff between  $\lambda$  and  $\delta$ . From (6) we may also read off the following corollary of Theorem 1.1 addressing Glauber dynamics for sampling a uniform independent set ( $\lambda = 1$ ).

**Corollary 1.3** *Let  $\Sigma$  satisfy the conditions of Theorem 1.1. There is a constant  $c > 0$  such that whenever  $\delta \geq c \log^3 d/d$  we have*

$$\tau_{\mathcal{M}_1(\Sigma)} \geq \exp_2\left\{\Omega\left(\frac{N\delta}{\log d}\right)\right\}.$$

As an application of Theorem 1.1, we consider the case  $\Sigma = Q_d$ , the  $d$ -dimensional hypercube. This is the  $d$ -regular bipartite graph on vertex set  $\{0, 1\}^d$  in which two vertices are adjacent if they differ on exactly one coordinate. The hypercube satisfies  $\delta(Q_d) \geq \Omega(1/\sqrt{d})$  (see, e.g. [16, Lemma 1.3]; this bound can also be derived from isoperimetric inequalities of Bezrukov [2] and Körner and Wei [15]) and so if  $c' > 0$  is a suitably large constant (depending on the constant  $c$  provided by Theorem 1.1) then (4) is satisfied as long as  $\lambda \geq c'd^{-1/4} \log^{3/2} d$ . So the following is a corollary of Theorem 1.1.

**Corollary 1.4** *There are constants  $c, c' > 0$  such that whenever  $\lambda \geq cd^{-1/4} \log^{3/2} d$  we have*

$$\tau_{\mathcal{M}_\lambda(Q_d)} \geq \exp_2 \left\{ \frac{c' 2^d \log^3(1 + \lambda)}{\sqrt{d} (1 + \log(1 + \lambda)) (c_3 \log d + \log(1 + \lambda)) \log \left( 2 + \frac{1}{\log(1 + \lambda)} \right)} \right\}.$$

*In particular,*

$$\tau_{\mathcal{M}_\lambda(Q_d)} \geq \begin{cases} \exp_2 \left\{ \frac{2^d \log^3(1 + \lambda)}{\sqrt{d} \log^2 d} \right\} & \text{if } cd^{-1/4} \log^{3/2} d \leq \lambda \leq O(1), \\ \exp_2 \left\{ \frac{2^d \log^2(1 + \lambda)}{\sqrt{d} \log d} \right\} & \text{if } \Omega(1) \leq \lambda \leq O(d), \\ \exp_2 \left\{ \frac{2^d \log(1 + \lambda)}{\sqrt{d}} \right\} & \text{if } \Omega(d) \leq \lambda. \end{cases}$$

**Remark 1.5** *Using (5) in place of (4) (which we may do, since  $Q_d$  has bounded codegree) we may improve the bound on  $\lambda$  in Corollary 1.4 to  $\lambda \geq cd^{-1/4} \log d$ .*

**Remark 1.6** *Let us return to  $T_{L,d}$ , the even discrete torus. Since  $Q_d$  is easily seen to be isomorphic to  $T_{2,d}$ , Corollary 1.4 gives an exponential lower bound on  $\tau_{\mathcal{M}_\lambda(T_{2,d})}$  for sufficiently large  $d$  whenever  $\lambda \geq \omega(d^{-1/4} \log^{3/2} d)$ . Unfortunately, the best bound we can obtain on the bipartite expansion constant of  $T_{L,d}$  is  $\delta(T_{L,d}) \geq \Omega(1/Ld)$  (see, e.g [12]), so we cannot use Theorem 1.1 to obtain any lower bound on  $\lambda$  independent of  $L$  beyond which  $\tau_{\mathcal{M}_\lambda(T_{L,d})}$  is large for all even  $L \geq 4$  and sufficiently large  $d$ . However, subsequent to the completion of this paper, a strategy specific to the torus has been employed in [11] to show that for all even  $L \geq 4$ ,  $\lambda \geq \omega(d^{-1/4} \log^{3/4} d)$  and  $d$  sufficiently large,*

$$\tau_{\mathcal{M}_\lambda(T_{L,d})} \geq \exp_2 \left\{ \frac{L^{d-1}}{d^3 \log^2 L} \right\}.$$

## 2 Proof of Theorem 1.1

The notion of conductance, introduced in [13], can be used to analyze the behavior of  $\tau_{\mathcal{M}_\lambda}$ . Let  $\mathcal{M}$  be a Markov chain on state space  $\Omega$  with transition matrix  $P$  and

stationary distribution  $\pi$ . For  $\omega_1, \omega_2 \in \Omega$  and  $A, B \subseteq \Omega$ , set

$$Q(\omega_1, \omega_2) = \pi(\omega_1)P(\omega_1, \omega_2) \quad \text{and} \quad Q(A, B) = \sum_{\omega_1 \in A, \omega_2 \in B} Q(\omega_1, \omega_2).$$

For  $\emptyset \neq S \subseteq \Omega$ , define the conductance of  $S$  as

$$\Phi(S) = \frac{Q(S, \Omega \setminus S)}{\pi(S)}.$$

We may interpret  $\Phi(S)$  as the probability under  $\pi$  that the chain escapes from  $S$  in one step, given that it is in  $S$ . Define the conductance of  $\mathcal{M}$  as

$$\Phi_{\mathcal{M}} = \min_{0 < \pi(S) \leq \frac{1}{2}} \Phi(S).$$

We may then bound the mixing time of  $\mathcal{M}$  by

$$\tau_{\mathcal{M}} \geq \left( \frac{1}{2} - \frac{1}{e} \right) \frac{1}{\Phi_{\mathcal{M}}} \quad (7)$$

(see e.g. [9], where the above bound is derived without assuming time-reversibility of the chain  $\mathcal{M}$ ). Thus to show that the mixing time is large, it is enough to exhibit a single  $S$  with small conductance.

Throughout this section we fix  $\Sigma$  satisfying the conditions of Theorem 1.1. Set

$$\mathcal{I}_{\mathcal{E}} = \{I \in \mathcal{I}(\Sigma) : |I \cap \mathcal{E}| > |I \cap \mathcal{O}|\},$$

define  $\mathcal{I}_{\mathcal{O}}$  analogously, and set  $\mathcal{I}_b = \mathcal{I}(\Sigma) \setminus (\mathcal{I}_{\mathcal{E}} \cup \mathcal{I}_{\mathcal{O}})$  ( $\mathcal{I}_b$  is the set of *balanced* independent sets). Without loss of generality, assume  $\pi_{\lambda}(\mathcal{I}_{\mathcal{E}}) \leq 1/2$ . Because Glauber dynamics changes the size of an independent set by at most one at each step, we have that if  $I \in \mathcal{I}_{\mathcal{E}}, J \notin \mathcal{I}_{\mathcal{E}}$  satisfy  $P_{\lambda}(I, J) \neq 0$ , then  $J \in \mathcal{I}_b$ . It follows that

$$\begin{aligned} Q(\mathcal{I}_{\mathcal{E}}, \Omega \setminus \mathcal{I}_{\mathcal{E}}) &= \sum_{I \in \mathcal{I}_{\mathcal{E}}, J \notin \mathcal{I}_{\mathcal{E}}} \pi_{\lambda}(I)P_{\lambda}(I, J) \\ &= \sum_{I \in \mathcal{I}_{\mathcal{E}}, J \notin \mathcal{I}_{\mathcal{E}}} \pi_{\lambda}(J)P_{\lambda}(J, I) \\ &= \sum_{I \in \mathcal{I}_{\mathcal{E}}, J \in \mathcal{I}_b} \pi_{\lambda}(J)P_{\lambda}(J, I) \\ &\leq \pi_{\lambda}(\mathcal{I}_b). \end{aligned} \quad (8)$$

The simplest way to see (8) is to use the fact that  $\mathcal{M}_{\lambda}$  is time-reversible (that is, that  $\pi_{\lambda}(I)P_{\lambda}(I, J) = \pi_{\lambda}(J)P_{\lambda}(J, I)$  for all  $I, J \in \mathcal{I}$ ); but note that more generally if  $\mathcal{M}$  is a (not necessarily time-reversible) Markov chain on finite state space  $\Omega$  with transition matrix  $P$  and stationary distribution  $\pi$  then

$$\sum_{\omega_1 \in S, \omega_2 \notin S} \pi(\omega_1)P(\omega_1, \omega_2) = \sum_{\omega_1 \in S, \omega_2 \notin S} \pi(\omega_2)P(\omega_2, \omega_1)$$

for all  $S \subseteq \Omega$ . Now using the trivial lower bound  $w_\lambda(\mathcal{I}_\mathcal{E}) \geq (1 + \lambda)^M$  (recall that for  $\mathcal{J} \subseteq \mathcal{I}$ ,  $w_\lambda(\mathcal{J}) = \sum_{J \in \mathcal{J}} \lambda^{|J|}$ ) we obtain

$$\Phi_{\mathcal{M}_\lambda} \leq \Phi(\mathcal{I}_\mathcal{E}) \leq \frac{\pi_\lambda(\mathcal{I}_b)}{\pi_\lambda(\mathcal{I}_\mathcal{E})} = \frac{w_\lambda(\mathcal{I}_b)}{w_\lambda(\mathcal{I}_\mathcal{E})} \leq \frac{w_\lambda(\mathcal{I}_b)}{(1 + \lambda)^M}. \quad (9)$$

Thus (recalling (7)) to show that  $\tau_{\mathcal{M}_\lambda}$  is large, it is enough to show that  $w_\lambda(\mathcal{I}_b)$  is small. We may think of  $\mathcal{I}_b$  as a ‘‘bottleneck’’ set through which any run of the chain must pass in order to mix; if the bottleneck has low measure, the mixing time is high.

We will actually consider a larger ‘‘bottleneck’’ set. Set

$$\mathcal{I}^{triv} = \{I \in \mathcal{I} : |I \cap \mathcal{E}|, |I \cap \mathcal{O}| \leq \alpha(\lambda)M\}$$

and

$$\mathcal{I}^{nt} = \{I \in \mathcal{I} : \min\{|I \cap \mathcal{E}|, |I \cap \mathcal{O}|\} \geq \alpha(\lambda)M\},$$

where  $\alpha(\lambda)$  is as defined in (2). Note that  $\mathcal{I}_b \subseteq \mathcal{I}^{triv} \cup \mathcal{I}^{nt}$ . We will show that as long as  $\lambda$  satisfies (4),

$$w_\lambda(\mathcal{I}^{triv} \cup \mathcal{I}^{nt}) \leq (1 + \lambda)^M \exp_2 \{-\Omega(M\alpha(\lambda)\beta(\lambda)\delta)\}, \quad (10)$$

from which Theorem 1.1 follows via (7) and (9).

Dealing with  $w_\lambda(\mathcal{I}^{triv})$  is relatively straightforward. We begin by observing that

$$4\alpha(\lambda) \log \frac{1}{\alpha(\lambda)} \leq \frac{\log(1 + \lambda)}{2(1 + \log(1 + \lambda))}. \quad (11)$$

To see this, first set

$$\gamma(\lambda) = \frac{\log(1 + \lambda)}{1 + \log(1 + \lambda)}.$$

Note that for all  $\lambda > 0$ ,  $0 < \gamma(\lambda) < 1$ . We have  $\alpha(\lambda) = \gamma(\lambda)/(44 \log(1 + 1/\gamma(\lambda)))$  and so (11) is equivalent to

$$\frac{\gamma(\lambda)}{11 \log\left(1 + \frac{1}{\gamma(\lambda)}\right)} \log\left(\frac{44 \log\left(1 + \frac{1}{\gamma(\lambda)}\right)}{\gamma(\lambda)}\right) \leq \frac{\gamma(\lambda)}{2}$$

which is in turn equivalent to

$$44 \log\left(1 + \frac{1}{\gamma(\lambda)}\right) \leq \gamma(\lambda) \left(1 + \frac{1}{\gamma(\lambda)}\right)^{11/2}.$$

That this inequality holds for all  $0 < \gamma(\lambda) < 1$  is a routine calculus exercise. Note also that for  $0 < x < 1/e$ ,

$$x \leq H(x) \leq 2x \log \frac{1}{x} \quad (12)$$



(where recall  $H(x) = -x \log x - (1-x) \log(1-x)$  is the usual binary entropy function). Finally, we use a result concerning the sums of binomial coefficients which follows from the Chernoff bounds [6] (see also [4], p.11):

$$\sum_{i=0}^{\lfloor cN \rfloor} \binom{N}{i} \leq 2^{H(c)N} \quad \text{for } c \leq \frac{1}{2}, \quad (13)$$

where  $\lfloor x \rfloor$  denotes the integer part of  $x$ .

Now with the inequalities justified below, we have

$$\begin{aligned} w_\lambda(\mathcal{I}^{triv}) &\leq \binom{M}{\leq \alpha(\lambda)M}^2 (1+\lambda)^{2\alpha(\lambda)M} \\ &\leq \exp_2 \{2H(\alpha(\lambda))M + 2\alpha(\lambda)M \log(1+\lambda)\} \end{aligned} \quad (14)$$

$$\leq \exp_2 \{2MH(\alpha(\lambda))(1 + \log(1 + \lambda))\} \quad (15)$$

$$\leq \exp_2 \left\{ 4M\alpha(\lambda) \log \frac{1}{\alpha(\lambda)} (1 + \log(1 + \lambda)) \right\} \quad (16)$$

$$\leq \exp_2 \left\{ M \frac{\log(1 + \lambda)}{2} \right\} \quad (17)$$

$$\leq (1 + \lambda)^{\frac{M}{2}}. \quad (18)$$

Here (and throughout) we use  $\binom{n}{\leq k}$  for  $\sum_{i \leq k} \binom{n}{i}$ . In (14), we are using (13), which is applicable by (3). In (15) we are using the first inequality in (12) and in (16) we are using the second (again, both of these are applicable by (3).) Finally in (17) we are using (11).

Bounding  $w_\lambda(\mathcal{I}^{nt})$  requires much more work. We begin by enlarging  $\mathcal{I}^{nt}$  slightly. Say that  $I \in \mathcal{I}(\Sigma)$  is *small on*  $\mathcal{E}$  if  $|\{I \cap \mathcal{E}\}| \leq M/2$  (recall that for  $A \subseteq \mathcal{E}$ ,  $[A] = \{v \in \mathcal{O} : N(v) \subseteq A\}$ ), and set

$$\mathcal{I}_{\mathcal{E}}^{nt} = \{I \in \mathcal{I}^{nt} : I \text{ is small on } \mathcal{E}\}.$$

Define *small on*  $\mathcal{O}$  and  $\mathcal{I}_{\mathcal{O}}^{nt}$  similarly. A simple argument, based on the fact that  $\Sigma$  has a perfect matching, shows that any  $I \in \mathcal{I}(\Sigma)$  must be small on at least one of  $\mathcal{E}$ ,  $\mathcal{O}$ , and so we have

$$w_\lambda(\mathcal{I}^{nt}) \leq 2 \max \{w_\lambda(\mathcal{I}_{\mathcal{E}}^{nt}), w_\lambda(\mathcal{I}_{\mathcal{O}}^{nt})\}.$$

We may assume without loss of generality that

$$w_\lambda(\mathcal{I}_{\mathcal{E}}^{nt}) = \max \{w_\lambda(\mathcal{I}_{\mathcal{E}}^{nt}), w_\lambda(\mathcal{I}_{\mathcal{O}}^{nt})\}$$

so that it is enough to show that

$$w_\lambda(\mathcal{I}_{\mathcal{E}}^{nt}) \leq (1 + \lambda)^M \exp_2 \{-\Omega(M\alpha(\lambda)\beta(\lambda)\delta)\}.$$

For each  $a \geq \alpha(\lambda)M$  and  $g \geq a$  set

$$\mathcal{A}(a, g) = \{A \subseteq \mathcal{E} : |[A]| = a, |N(A)| = g\}$$

and set

$$\mathcal{I}(a, g) = \{I \in \mathcal{I}_{\mathcal{E}}^{nt} : I \cap \mathcal{E} \in \mathcal{A}(a, g)\}.$$

We have

$$\begin{aligned} w_{\lambda}(\mathcal{I}_{\mathcal{E}}^{nt}) &\leq \sum_{a \geq \alpha(\lambda)M, g \geq a} w_{\lambda}(\mathcal{I}(a, g)) \\ &\leq \sum_{a \geq \alpha(\lambda)M, g \geq a} w_{\lambda}(\mathcal{A}(a, g))(1 + \lambda)^{M-g} \\ &\leq (1 + \lambda)^M \sum_{a \geq \alpha(\lambda)M, g \geq a} w_{\lambda}(\mathcal{A}(a, g))(1 + \lambda)^{-g} \\ &\leq (1 + \lambda)^M M^2 \max_{a \geq \alpha(\lambda)M, g \geq a} w_{\lambda}(\mathcal{A}(a, g))(1 + \lambda)^{-g} \end{aligned}$$

The key now is to upper bound  $w_{\lambda}(\mathcal{A}(a, g))$ . The following theorem (whose proof is given in Section 3) is based on ideas of A. Sapozhenko [18, 19].

**Theorem 2.1** *Let  $\Sigma$  be any graph satisfying the assumptions of Theorem 1.1. We have*

$$w_{\lambda}(\mathcal{A}(a, g)) \leq (1 + \lambda)^g \exp_2 \{-\Omega((g - a)\beta(\lambda))\}.$$

for any  $a \geq \alpha(\lambda)M$  and any  $\lambda$  satisfying (4).

For  $a \geq \alpha(\lambda)M$  and  $g \geq a$  we have  $g - a \geq \delta g \geq M\alpha(\lambda)\delta$  and so

$$\begin{aligned} w_{\lambda}(\mathcal{I}_{\mathcal{E}}^{nt}) &\leq (1 + \lambda)^M M^2 \exp_2 \{-\Omega(M\alpha(\lambda)\beta(\lambda)\delta)\} \\ &\leq (1 + \lambda)^M \exp_2 \{-\Omega(M\alpha(\lambda)\beta(\lambda)\delta)\}. \end{aligned} \tag{19}$$

To see that the factor of  $M^2$  may be absorbed into the exponent, note that by hypothesis,  $2M \geq d^2$  and so  $M^2 \leq \exp_2 \{O(M \log d/d^2)\}$ , and that combining (3) and (4) we have  $\alpha(\lambda)\beta(\lambda)\delta \geq \Omega(d^{-3/2} \log d)$ .

Combining (19) and (18) we get (10) and hence Theorem 1.1.

### 3 Proof of Theorem 2.1

For  $u, v \in V$  and  $A, B \subseteq V$  we write  $\nabla(A)$  for the set of edges having one end in  $A$  and (if  $A \cap B = \emptyset$ )  $\nabla(A, B)$  for the set of edges having one end in each of  $A, B$ . We also write  $d_A(v)$  for  $|N(v) \cap A|$ .

Throughout this section, we fix  $\Sigma$  satisfying the assumptions of Theorem 1.1. We also fix  $a$  and  $g$ , but we do not assume  $a \geq \alpha(\lambda)M$ . We write  $\mathcal{A}$  for  $\mathcal{A}(a, g)$ . Given  $A \in \mathcal{A}$  we always write  $G$  for  $N(A)$  and set  $t = g - a$ . Note that for  $A \in \mathcal{A}$ ,

$$|\nabla(G, \mathcal{E} \setminus [A])| = dg - da = td. \quad (20)$$

The proof of Theorem 2.1 involves the idea of “approximation”. We begin with an informal outline. To bound  $w_\lambda(\mathcal{A})$ , we produce a small set  $\mathcal{U}$  with the properties that each  $A \in \mathcal{A}$  is “approximated” (in an appropriate sense) by some  $U \in \mathcal{U}$ , and for each  $U \in \mathcal{U}$ , the total weight of those  $A \in \mathcal{A}$  that could possibly be “approximated” by  $U$  is small. (Each  $U \in \mathcal{U}$  will consist of two parts; one each approximating  $G$  and  $A$ .) The product of the bound on  $|\mathcal{U}|$  and the bound on the weight of those  $A \in \mathcal{A}$  that may be approximated by any  $U$  is then a bound on  $w_\lambda(\mathcal{A})$ . The set  $\mathcal{U}$  is itself produced by an approximation process — we first produce a small set  $\mathcal{V}$  with the property that each  $A \in \mathcal{A}$  is “weakly approximated” (in an appropriate sense) by some  $V \in \mathcal{V}$ , and then show that for each  $V$  there is a small set  $\mathcal{W}(V)$  with the property that for each  $A \in \mathcal{A}$  that is “weakly approximated” by  $V$ , there is a  $W \in \mathcal{W}(V)$  which approximates  $A$ ; we then take  $\mathcal{U} = \cup_{V \in \mathcal{V}} \mathcal{W}(V)$ . (Each  $V \in \mathcal{V}$  will consist of a single part.)

The main inspiration for the proof of Theorem 2.1 is the work of A. Sapozhenko, who, in [19], gave a relatively simple derivation for the asymptotics of the number of independent sets in  $Q_d$  (in the notation of (1), this is the asymptotics of  $Z_\lambda(Q_d)$  with  $\lambda = 1$ ), earlier derived in a more involved way in [16]. Our Lemma 3.3 is a modification of a lemma in [18], and our overall approach is similar to [19]. See e.g. [10] for another recent application of these ideas.

We now begin the formal discussion of Theorem 2.1 by introducing the two notions of approximation that we will use, beginning with the weaker notion. A *covering approximation* for  $A \subseteq \mathcal{E}$  is a set  $F_0 \in 2^{\mathcal{O}}$  satisfying

$$F_0 \subseteq G, \quad N(F_0) \supseteq [A].$$

The second notion of approximation depends on a parameter  $\psi$ ,  $1 \leq \psi \leq d/2$ . A  $\psi$ -*approximation* for  $A \subseteq \mathcal{E}$  is a pair  $(F, S) \in 2^{\mathcal{O}} \times 2^{\mathcal{E}}$  satisfying

$$F \subseteq G, \quad S \supseteq [A], \quad (21)$$

$$d_F(u) \geq d - \psi \quad \forall u \in S \quad (22)$$

and

$$d_{\mathcal{E} \setminus S}(v) \geq d - \psi \quad \forall v \in \mathcal{O} \setminus F. \quad (23)$$

Note that if  $x \in [A]$  then  $N(x) \subseteq G$ , and if  $y \in \mathcal{O} \setminus G$  then  $N(y) \subseteq \mathcal{E} \setminus [A]$ . If we think of  $S$  as “approximate  $[A]$ ” and  $F$  as “approximate  $G$ ”, (22) says that if  $x \in \mathcal{E}$  is in “approximate  $[A]$ ” then almost all of its neighbours are in “approximate  $G$ ”, while

(23) says that if  $y \in \mathcal{O}$  is not in “approximate  $G$ ” then almost all of its neighbours are not in “approximate  $[A]$ ”.

Before continuing, we note a property of  $\psi$ -approximations that will be of use later.

**Lemma 3.1** *If  $(F, S)$  is a  $\psi$ -approximation for  $A \in \mathcal{A}$  then*

$$|S| \leq |F| + \frac{2t\psi}{d-\psi}. \quad (24)$$

*Proof:* Observe that  $|\nabla(S, G)|$  is bounded above by  $d|F| + \psi|G \setminus F|$  and below by  $d|[A]| + (d-\psi)|S \setminus [A]| = d|S| - \psi|S \setminus [A]|$ , giving

$$|S| \leq |F| + \frac{\psi|(G \setminus F) \cup (S \setminus [A])|}{d},$$

and that each  $u \in (G \setminus F) \cup (S \setminus [A])$  contributes at least  $d-\psi$  edges to  $\nabla(G, \mathcal{E} \setminus [A])$ , a set of size  $td$ , giving

$$|(G \setminus F) \cup (S \setminus [A])| \leq \frac{2td}{d-\psi}.$$

These two observations together give (24).  $\square$

There are three parts to the proof of Theorem 2.1. Lemma 3.2 is the first “approximation” step, producing a small family  $\mathcal{V}$  of covering approximations for  $\mathcal{A}$ . Lemma 3.3 is the second “approximation” step, refining the covering approximations to produce a family  $\mathcal{W}$  of  $\psi$ -approximations for  $\mathcal{A}$ . Finally, Lemma 3.4 is the “reconstruction” step, bounding the weight of the set of  $A$ ’s that could possibly be  $\psi$ -approximated by a member of  $\mathcal{W}$ . We now state the three relevant lemmas. We will then derive Theorem 2.1 before turning to the proofs of the approximation and reconstruction lemmas.

**Lemma 3.2** *There is a  $\mathcal{V} = \mathcal{V}(a, g) \subseteq 2^{\mathcal{O}}$  with*

$$|\mathcal{V}| \leq \binom{M}{\leq \frac{2g \log d}{d}}$$

*such that each  $A \in \mathcal{A}$  has a covering approximation in  $\mathcal{V}$ .*

**Lemma 3.3** *For any  $F_0 \in \mathcal{V}$  and  $1 \leq \psi \leq d/2$  there is a  $\mathcal{W} = \mathcal{W}(F_0, \psi, a, g) \subseteq 2^{\mathcal{O}} \times 2^{\mathcal{E}}$  with*

$$|\mathcal{W}| \leq \binom{2g \log d}{\leq \frac{2g}{d}} \binom{2d^3 g \log d}{\leq \frac{2t}{\psi}} \binom{2g \log d}{\leq \frac{td}{(d-\psi)\psi}}$$

*such that any  $A \in \mathcal{A}$  for which  $F_0$  is a covering approximation has a  $\psi$ -approximation in  $\mathcal{W}$ .*

**Lemma 3.4** *Given  $1 \leq \psi \leq d/2$  and  $1 \geq \gamma > \frac{-2\psi}{d-\psi}$ , for each  $(F, S) \in 2^{\mathcal{O}} \times 2^{\mathcal{E}}$  that satisfies (24) we have*

$$\sum w_\lambda(A) \leq \max \left\{ (1 + \lambda)^{g-\gamma t}, \left( \leq \frac{3dg}{\frac{2t\psi}{d-\psi} + \gamma t} \right) (1 + \lambda)^{g-t} \right\} \quad (25)$$

where the sum is over all those  $A$ 's in  $\mathcal{A}$  satisfying  $F \subseteq G$  and  $S \supseteq [A]$ .

Before turning to the proofs of Lemmas 3.2, 3.3 and 3.4, we use them to obtain Theorem 2.1. Throughout, we will use (usually without comment) a simple observation about sums of binomial coefficients: if  $k = o(n)$ , we have

$$\begin{aligned} \sum_{i \leq k} \binom{n}{i} &\leq (1 + O(k/n)) \binom{n}{k} \\ &\leq (1 + O(k/n))(en/k)^k \\ &\leq \exp_2 \{ (1 + o(1))k \log(n/k) \}. \end{aligned} \quad (26)$$

Take  $\psi = \sqrt{d}$  and

$$\gamma = \frac{\log(1 + \lambda) - \frac{\sqrt{d}}{d-\sqrt{d}} \log(d^5/\delta)}{\log(1 + \lambda) + \log(d^5/\delta)}.$$

Note that for this choice of  $\psi$  and  $\gamma$  we have  $\gamma > \frac{-\psi}{d-\psi}$ , and so

$$\log \frac{3d}{\delta \left( \frac{2\psi}{d-\psi} + \gamma \right)} \leq \frac{1}{2} \log(d^5/\delta).$$

The bound in (25) is therefore at most

$$(1 + \lambda)^g \exp_2 \left\{ \max \left\{ -\gamma t \log(1 + \lambda), t \left( \frac{2\psi}{d-\psi} + \gamma \right) \log(d^5/\delta) \right\} \right\}.$$

(Here we have used (26)). For our choice of  $\psi$  and  $\gamma$  this is at most

$$(1 + \lambda)^g \exp_2 \left\{ t \frac{\log(1 + \lambda) \frac{\sqrt{d}}{d-\sqrt{d}} \log(d^5/\delta) - \log^2(1 + \lambda)}{\log(1 + \lambda) + \log(d^5/\delta)} + t \frac{\sqrt{d} \log(d^5/\delta)}{d - \sqrt{d}} \right\},$$

which in turn is at most

$$(1 + \lambda)^g \exp_2 \left\{ O \left( t \frac{\log(d^5/\delta)}{\sqrt{d}} \right) - t \frac{\log^2(1 + \lambda)}{\log(1 + \lambda) + \log(d^5/\delta)} \right\}. \quad (27)$$

The bounds in Lemmata 3.3 and 3.2 are at most

$$\exp_2 \left\{ O \left( g \frac{\log d}{d} + t \frac{\log(d^5/\delta)}{\sqrt{d}} \right) \right\} \quad \text{and} \quad \exp_2 \left\{ O \left( g \frac{\log^2 d}{d} \right) \right\} \quad (28)$$

respectively. For the latter bound, we are using the assumption  $a \geq \alpha(M)$  of Theorem 2.1 and the fact that  $\lambda \geq 1/\sqrt{d}$ , which together imply (via (3)) that

$$\frac{Md}{g \log d} \leq d^{3/2}.$$

Combining (28) with (27), we get

$$w_\lambda(\mathcal{A}) \leq (1 + \lambda)^g \exp_2 \left\{ O \left( g \frac{\log^2 d}{d} + t \frac{\log(d^5/\delta)}{\sqrt{d}} \right) - t \frac{\log^2(1 + \lambda)}{\log(1 + \lambda) + \log(d^5/\delta)} \right\}.$$

Noting that  $t \geq \delta g$  always, we find that if  $\lambda$  satisfies (4) with a suitably large constant  $c$ , then

$$w_\lambda(\mathcal{A}) \leq (1 + \lambda)^g \exp_2 \left\{ -\Omega \left( t \frac{\log^2(1 + \lambda)}{\log(1 + \lambda) + \log(d^5/\delta)} \right) \right\}.$$

and so we get Theorem 2.1.

We now turn to the proofs of Lemmata 3.2, 3.3 and 3.4.

*Proof of Lemma 3.2:* We appeal to a special case of a fundamental result due to Lovász [17] and Stein [20]. For a bipartite graph  $\Gamma$  with bipartition  $P \cup Q$ , we say that  $Q' \subseteq Q$  covers  $P$  if each  $p \in P$  has a neighbour in  $Q'$ .

**Lemma 3.5** *If  $\Gamma$  as above satisfies  $|N(x)| \geq p$  for each  $x \in P$  and  $|N(y)| \leq q$  for each  $y \in Q$ , then  $P$  is covered by some  $Q' \subseteq Q$  with*

$$|Q'| \leq (|Q|/p)(1 + \ln q).$$

Applying the lemma with  $\Gamma$  the subgraph of  $\Sigma$  induced by  $[A] \cup G$ ,  $P = [A]$ ,  $Q = G$  and  $p = q = d$ , we find that each  $A \in \mathcal{A}$  has a covering approximation of size at most  $2g \log d/d$ . Taking  $\mathcal{V}$  to be the set of all subsets of  $\mathcal{O}$  of size at most  $2g \log d/d$ , the lemma follows.  $\square$

*Proof of Lemma 3.3:* We describe an algorithm, which we refer to as the *degree algorithm*, which produces for input  $(F_0, A) \in 2^{\mathcal{O}} \times 2^{\mathcal{E}}$  for which  $F_0$  is a covering approximation of  $A$  (i.e., with  $N(F_0) \supseteq [A]$ ), an output  $(F, S) \in 2^{\mathcal{O}} \times 2^{\mathcal{E}}$  which is a  $\psi$ -approximation for  $A$  (i.e., which satisfies (21), (22) and (23)). The idea for the algorithm is from [18]. To begin, fix a linear ordering  $\ll$  of  $V$ .

**Step 1:** If  $\{u \in [A] : d_{G \setminus F_0}(u) > d/2\} \neq \emptyset$ , pick the smallest (with respect to  $\ll$ )  $u$  in this set and update  $F_0$  by  $F_0 \leftarrow F_0 \cup N(u)$ . Repeat this until  $\{u \in [A] : d_{G \setminus F_0}(u) > d/2\} = \emptyset$ . Then set  $F_1 = F_0$  and  $S_1 = \{u \in \mathcal{E} : d_{F_1}(u) \geq d - d/2\}$  and go to Step 2.

**Step 2:** If  $\{v \in \mathcal{O} \setminus G : d_{S_1}(v) > \psi\} \neq \emptyset$ , pick the smallest (with respect to  $\ll$ )  $v$  in this set and update  $S_1$  by  $S_1 \leftarrow S_1 \setminus N(v)$ . Repeat this until  $\{v \in \mathcal{O} \setminus G : d_{S_1}(v) > \psi\} = \emptyset$ . Then set  $S_2 = S_1$  and  $F_2 = \{v \in \mathcal{O} : d_{S_2}(v) > \psi\}$  and go to Step 3.

**Step 3:** If  $\{w \in [A] : d_{G \setminus F_2}(w) > \psi\} \neq \emptyset$ , pick the smallest (with respect to  $\ll$ )  $w$  in this set and update  $F_2$  by  $F_2 \leftarrow F_2 \cup N(w)$ . Repeat this until  $\{w \in [A] : d_{G \setminus F_2}(w) > \psi\} = \emptyset$ . Then set  $F = F_2$  and  $S = S_2 \cap \{w \in \mathcal{E} : d_F(w) \geq d - \psi\}$  and stop.

**Claim 3.6** *The output of the degree algorithm is a  $\psi$ -approximation for  $A$ .*

*Proof:* To see that  $F \subseteq G$  and  $S \supseteq [A]$ , first observe that  $S_1 \supseteq [A]$  (or Step 1 would not have terminated). We then have  $S_2 \supseteq [A]$  (since Step 2 deletes from  $S_1$  only neighbours of  $\mathcal{O} \setminus G$ ), and  $F_2 \subseteq G$  (or Step 2 would not have terminated). Finally,  $F \subseteq G$  (since the vertices added to  $F_2$  in Step 3 are all in  $G$ ) and  $S \supseteq [A]$  (or Step 3 would not have terminated)

By the definition of  $S$ , (22) is satisfied. To verify (23), note that by definition of  $F_2$ , if  $y \in \mathcal{O} \setminus F_2$  then  $d_{\mathcal{E} \setminus S_2}(y) \geq d - \psi$ . That  $y \in \mathcal{O} \setminus F$  implies  $d_{\mathcal{E} \setminus S}(y) \geq d - \psi$  now follows from the fact that  $F_2 \subseteq F$  and  $S_2 \supseteq S$ .  $\square$

**Remark 3.7** *The alert reader may have noticed that if we replace  $d/2$  by  $\psi$  in Step 1, then the output of Step 2 is already a  $\psi$ -approximation for  $A$ . The three-step algorithm, however, is needed to obtain the right bound on  $\beta(\lambda)$  in Theorem 1.1; see Remark 3.9 following the proof of Claim 3.8.*

**Claim 3.8** *Fix  $F_0 \in \mathcal{V}$ . The degree algorithm has at most*

$$\binom{2g \log d}{\leq \frac{2g}{d}} \binom{2d^3 g \log d}{\leq \frac{2t}{\psi}} \binom{2g \log d}{\leq \frac{td}{(d-\psi)\psi}}$$

*outputs as the input runs over those  $(F_0, A)$  for which  $A \in \mathcal{A}$  and  $F_0$  is a covering approximation for  $A$ .*

Taking  $\mathcal{W}$  to be the set of all possible outputs of the algorithm, the lemma follows.

*Proof of Claim 3.8:* The output of the algorithm is determined by the set of  $u$ 's whose neighbourhoods are added to  $F_0$  in Step 1, the set of  $v$ 's whose neighbourhoods are removed from  $S_1$  in Step 2, and the set of  $w$ 's whose neighbourhoods are added to  $F_2$  in Step 3.

Each iteration in Step 1 removes at least  $d/2$  vertices from  $G \setminus F$ , a set of size at most  $g$ , so there are at most  $2g/d$  iterations. The  $u$ 's in Step 1 are all drawn from  $[A]$  and hence  $N(F_0)$ , a set of size at most  $d|F_0| \leq 2g \log d$ . So the total number of outputs for Step 1 is at most

$$\binom{2g \log d}{\leq \frac{2g}{d}}. \tag{29}$$

At the start of Step 2, each  $x \in S_1 \setminus [A]$  contributes at least  $d/2$  edges to  $\nabla(G, \mathcal{E} \setminus [A])$ , by (20) a set of size  $dt$ , so  $|S_1 \setminus [A]| \leq 2t$ . Each  $v$  used in Step 2 reduces this by at least  $\psi$ , so there are at most  $2t/\psi$  iterations. Each  $v$  is drawn from  $N(S_1)$ , a set which is contained in the fourth neighbourhood of  $F_0$  ( $S_1 \subseteq N(G)$  by construction of  $S_1$ ,  $G = N(A)$  and  $A \subseteq N(F_0)$ ) and so has size at most  $d^4|F_0| \leq 2d^3g \log d$ . So the total number of outputs for Step 2 is

$$\left( \begin{array}{l} 2d^3g \log d \\ \leq \frac{2t}{\psi} \end{array} \right). \quad (30)$$

At the start of Step 3, each  $y \in G \setminus F_2$  contributes at least  $d - \psi$  edges to  $\nabla(G, \mathcal{E} \setminus [A])$ , so  $|G \setminus F_2| \leq dt/(d - \psi)$ . Each  $w$  used in step 3 reduces this by at least  $\psi$ , so there are at most  $dt/((d - \psi)\psi)$  iterations. As in Step 1, the  $w$ 's are all drawn from a set of size at most  $2g \log d$ , so the total number of outputs for Step 1 is at most

$$\left( \begin{array}{l} 2g \log d \\ \leq \frac{td}{(d-\psi)\psi} \end{array} \right). \quad (31)$$

Combining (29), (30) and (31), the claim follows.  $\square$

**Remark 3.9** *The bound in Claim 3.8 is at most  $\exp_2\{O(g \log d/d + t \log(d^5/\delta)/\psi)\}$ . If we replace  $d/2$  by  $\psi$  in Step 1 of the degree algorithm and take the output of Step 2 to be the final output, then the bound in the claim becomes weaker:*

$$\left( \begin{array}{l} 2g \log d \\ \leq \frac{g}{\psi} \end{array} \right) \left( \begin{array}{l} 2gd^3 \log d \\ \leq \frac{td}{(d-\psi)\psi} \end{array} \right) = \exp_2 \{O(g \log d/\psi + t \log(d^5/\delta)/\psi)\}.$$

*(Each iteration of Step 1 now reduces  $G \setminus F$  by at least  $\psi$ ). Using this bound in the proof of Theorem 1.1 instead of the stronger bound given by the three-step degree algorithm would ultimately lead to a weaker bound on  $\beta(\lambda)$  in (4). Step 1 of the degree algorithm may be thought of as an “initialization” which reduces  $|G \setminus F|$  from  $O(g)$  to  $O(t)$  without adding much to the “cost” of the algorithm.*

*Proof of Lemma 3.4:* Say that  $S$  is *small* if  $|S| < g - \gamma t$  and *large* otherwise. We can obtain all  $A \in \mathcal{A}$  for which  $F \subseteq G$  and  $S \supseteq [A]$  as follows.

If  $S$  is small, we specify  $A$  by picking a subset of  $S$ . If  $S$  is large, we first specify  $G$ . Note that by (24) and the definition of large we have in this case that

$$|G \setminus F| < 2t\psi/(d - \psi) + \gamma t \quad \text{and} \quad G \setminus F \subseteq N(S) \setminus F,$$

so we specify  $G$  by picking a subset of  $N(S) \setminus F$  of size at most  $2t\psi/(d - \psi) + \gamma t$  (this is our choice of  $G \setminus F$ ). Then, noting that  $[A]$  is determined by  $G$ , we specify  $A$  by picking a subset of  $[A]$ .



This procedure produces all possible  $A$ 's (and more). We now bound the sum of the weights of the outputs.

If  $S$  is small then the total weight of outputs is at most

$$(1 + \lambda)^{g-\gamma t}. \tag{32}$$

We have

$$|N(S) \setminus F| \leq d|S| \leq dg + \frac{2td\psi}{d-\psi} \leq 3dg$$

so that if  $S$  is large, the total number of possibilities for  $|G \setminus F|$  is at most

$$\left( \begin{array}{c} 3dg \\ \leq \frac{2t\psi}{d-\psi} + \gamma t \end{array} \right)$$

and the total weight of outputs is at most

$$\left( \begin{array}{c} 3dg \\ \leq \frac{2t\psi}{d-\psi} + \gamma t \end{array} \right) (1 + \lambda)^{g-t}. \tag{33}$$

Combining (32) and (33), the lemma follows.  $\square$

*Acknowledgment.* This work originated while the first author was a member of and while the second author was visiting the Theory Group at Microsoft Research in Redmond, Washington. The authors thank Microsoft Research for this support.

## References

- [1] J. van den Berg and J.E. Steif, Percolation and the hard-core lattice gas model, *Stochastic Process. Appl.* **49** (1994), 179-197.
- [2] S. Bezrukov, On minimization of the surrounding of subsets in Hamming space, in *Kombinatorno-Algebraic Methods in Applied Mathematics*, Gorky University Press, 1985. (Russian)
- [3] B. Bollobás, *Modern Graph Theory*, Springer, New York, 1998.
- [4] B. Bollobás, *Random Graphs*, Cambridge University Press, Cambridge, 2001.
- [5] C. Borgs, J. Chayes, A. Frieze, J.H. Kim, P. Tetali, E. Vigoda, V. Vu, Torpid mixing of some Monte Carlo Markov chain algorithms in statistical physics, *Proc. of the IEEE FOCS '99*, 218-229.
- [6] H. Chernoff, A measure of asymptotic efficiency for tests of a hypothesis based on the sum of observations, *Ann. Math. Statistics* **23** (1952), 493-507.
- [7] R. Diestel, *Graph Theory*, Springer, New York, 2000.

- [8] R.L. Dobrushin, The problem of uniqueness of a Gibbs random field and the problem of phase transition, *Funct. Anal. Appl.* **2** (1968), 302-312.
- [9] M. Dyer, A. Frieze and M. Jerrum, On counting independent sets in sparse graphs, *SIAM J. Comput.* **31** (2002), 1527–1541.
- [10] D. Galvin, On homomorphisms from the Hamming cube to  $\mathbb{Z}$ , *Israel J. Math.* **138** (2003), 189–213.
- [11] D. Galvin, Slow mixing of local dynamics for independent sets on the discrete torus, in preparation.
- [12] D. Galvin and J. Kahn, On phase transition in the hard-core model on  $\mathbb{Z}^d$ , *Combin. Probab. Comput.* **13** (2004), 137–164.
- [13] M. Jerrum and A. Sinclair, The Monte Carlo Markov chain method: an approach to approximate counting and integration, in *Approximation Algorithms for NP-hard problems*, PWS, 1996.
- [14] F. Kelly, Loss networks, *Ann. Appl. Probab.* **1** (1991) 319–378.
- [15] J. Körner and V. Wei, Odd and even Hamming spheres also have minimum boundary, *Discrete Math.* **51** (1984), 147–165.
- [16] A. D. Korshunov and A. A. Sapozhenko, The number of binary codes with distance 2, *Problemy Kibernet.* **40** (1983), 111–130. (Russian)
- [17] L. Lovász, On the ratio of optimal integral and fractional covers, *Discrete Math.* **13** (1975) 383–390.
- [18] A. A. Sapozhenko, On the number of connected subsets with given cardinality of the boundary in bipartite graphs, *Metody Diskret. Analiz.* **45** (1987), 42–70. (Russian)
- [19] A. A. Sapozhenko, The number of antichains in ranked partially ordered sets, *Diskret. Mat.* **1** (1989), 74–93. (Russian; translation in *Discrete Math. Appl.* **1** no. 1 (1991), 35–58)
- [20] S. K. Stein, Two combinatorial covering theorems, *J. Combin. Theory Ser. A* **16** (1974), 391–397.
- [21] E. Vigoda, A note on the Glauber dynamics for sampling independent sets, *Electron. J. Combin.* **8** no. 1 (2001), #R8.