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

David Galvin
Microsoft Research — Theory Group
Prasad Tetali
School of Mathematics & College of Computing
Georgia Institute of Technology

January 2004
Technical Report
MSR-TR-2004-06

Let $\Sigma = (V, E)$ be a finite, d-regular bipartite graph. For any $\lambda > 0$ let π_λ be the probability measure on the independent sets of Σ in which the set I is chosen with probability proportional to $\lambda^{|I|}$ (π_λ is the *hard-core measure with activity* λ *on* Σ). We study the Glauber dynamics, or single-site update Markov chain, whose stationary distribution is π_λ. We show that when λ 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 Σ is the d-dimensional hypercube $\{0, 1\}^d$ we show that for values of λ 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.

Microsoft Research
Microsoft Corporation
One Microsoft Way
Redmond, WA 98052
`http://www.research.microsoft.com`

Key words and phrases: Mixing time, hard-core model, conductance, Glauber dynamics, discrete hypercube. Second author's research supported in part by NSF grant DMS-0100289.

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. [5].) 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* λ on $\mathcal{I}(\Sigma)$ by

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

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. [6, 1]) where it serves as a simple mathematical model of a gas with particles of non-negligible size. The vertices of Σ 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 λ measures the likelihood of a site being occupied.

The measure also has a natural interpretation in the context of communications networks (see e.g. [10]). Here the vertices of Σ 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 λ 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 Σ.

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 π_λ 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

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

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 $\frac{\lambda}{1+\lambda}$, and "remove" it with probability $\frac{1}{1+\lambda}$; 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 π_λ. 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 Ω, with transition probabilities $P : \Omega^2 \to [0, 1]$. 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 ω_0, and denote by π the stationary distribution. Set

$$VAR_{\omega_0}(t) = \frac{1}{2} \sum_{\omega \in \Omega} |P^t(\omega_0, \omega) - \pi(\omega)|$$

(the *variation distance* at time t w.r.t. the initial state ω_0). Define the *mixing time* of $\mathcal{M}$ by

$$\tau_{\mathcal{M}} = \max_{\omega_0 \in \Omega} \min\{t : VAR_{\omega_0}(t') \leq \frac{1}{2e} \ \forall t' > t\}.$$

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 Ω which is within ϵ of π (in variation distance), it is necessary and sufficient to run the chain from some arbitrarily chosen distribution for some multiple (depending on ϵ) of the mixing time.

If $\mathcal{M}$ is time reversible (that is, satisfies the detailed balance equations:

$$\pi(\omega_1)P(\omega_1, \omega_2) = \pi(\omega_2)P(\omega_2, \omega_1) \text{ for all } \omega_1, \omega_2 \in \Omega)$$

then the notion of conductance, introduced in [9], can be used to analyze the behavior of $\tau_{\mathcal{M}}$. 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 $\emptyset \neq S \subset \Omega$, 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 π_λ 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 $\mathcal{M}$ using the bound

$$\tau_{\mathcal{M}} \geq \frac{1}{2\Phi_{\mathcal{M}}} \tag{2}$$

(see e.g. [9]). Thus to show that the mixing time is large, it is enough to exhibit a single S with small conductance.

In [3] a study was made of the Glauber dynamics for the hard-core measure on the discrete torus $T_{L,d}$. This is the graph whose vertex set is the box $[-L, L]^d$ in $\mathbf{Z}^d$, the d-dimensional integer lattice with the usual adjacency, and with opposite faces identified. It was shown, using the conductance method, that for λ growing exponentially with d, the Glauber dynamics mixes exponentially slowly in L^{d-1}, the surface area of the torus.

In light of a recent result of Galvin and Kahn [8], it is tempting to believe that slow mixing on $T_{L,d}$ should hold for much smaller values of λ; even for λ tending to 0 as d grows. The main result of [8] is that the hard-core model on $\mathbf{Z}^d$ exhibits multiple Gibbs phases for $\lambda > Cd^{-1/4}\log^{3/4} d$ for some large constant C. Specifically, write $\mathcal{E}$ and $\mathcal{O}$ for the sets of even and odd vertices of $\mathbf{Z}^d$ (defined in the obvious way: a vertex of $\mathbf{Z}^d$ 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 [8] is:

Theorem 1.1 *There is a constant C such that if $\lambda > Cd^{-1/4}\log^{3/4} d$, then*

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

Thus, roughly speaking, the influence of the boundary on behavior at the origin persists as the boundary recedes. Informally, this suggests that for λ 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, and thus cause its mixing time to be large.

In this paper, by studying a simpler but related graph, we provide some evidence for slow mixing on $\mathbf{Z}^d$ at small values of λ. We give a proof that the Glauber dynamics for the hard-core measure on the d-dimensional Hamming cube $\{0, 1\}^d$ mixes exponentially slowly in the volume of the cube for $\lambda > \omega(\log^{3/2} d/d^{1/4})$.

Our main result actually applies to any regular bipartite graph. Before stating it, 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 $M = |\mathcal{E}| = |\mathcal{O}|$.

For $u, v \in V$ we write $u \sim v$ if there is an edge in Σ 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 $A \subseteq \mathcal{O}$) set

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

and say that such an A is *small* if $|[A]| \leq M/2$. Define the *bipartite expansion constant* of Σ by

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

(Note that if Σ is connected, $0 < \delta < 1$.)

All implied constants in O and Ω 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

$$\alpha(\lambda) = \frac{\log(1+\lambda)}{44\,(1+\log(1+\lambda))\log\left(2+\frac{1}{\log(1+\lambda)}\right)} \tag{3}$$

and

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

We note for future reference that

$$\lambda > \frac{1}{\sqrt{d}} \ \text{ implies } \ \frac{1}{2} > \alpha(\lambda) = \Omega\left(\frac{1}{\sqrt{d}\log d}\right). \tag{4}$$

Our main result is

Theorem 1.2 *Let Σ be a connected, d-regular, bipartite graph with $2M > d^2$ vertices and bipartite expansion constant δ. There is a constant $c_1 > 0$ such that whenever λ satisfies*

$$\beta(\lambda) > c_1 \max\left\{\frac{\log(d^5/\delta)}{\sqrt{d}}, \frac{\log^2 d}{\delta d}\right\} \tag{5}$$

we have

$$\tau_{\mathcal{M}_\lambda(\Sigma)} > \exp_2\left\{\Omega(M\alpha(\lambda)\beta(\lambda)\delta)\right\}.$$

Remark 1. We do not believe that our bound on λ is best possible. In the other direction, Luby and Vigoda [14] have shown that if Σ is any graph (not necessarily bipartite) with maximum degree Δ, then $\tau_{\mathcal{M}_\lambda}(\Sigma)$ is a polynomial in $|V(\Sigma)|$ whenever $\lambda < 2/(\Delta-2)$.

Remark 2. If we add as an additional hypothesis to Theorem 1.2 that Σ has *bounded codegree* (that is, there is a constant κ independent of d such that each pair of vertices in Σ has at most κ common neighbours), then we can slightly improve our bound on λ to

$$\beta(\lambda) > c_1 \max\left\{\frac{\log(d^5/\delta)}{\sqrt{d}}, \frac{\log d}{\delta d^2}\right\}. \tag{6}$$

We do not present the more complicated argument here.

Note that since $\delta < 1$, we cannot possibly satisfy (5) for $\lambda \leq 1/\sqrt{d}$, so we may (and will) assume from here on that $\lambda > 1/\sqrt{d}$.

1.1 Application to the hypercube

As an application of Theorem 1.2, we consider the case $\Sigma = Q_d$, the d-dimensional hypercube. This is the graph on vertex set $\{0,1\}^d$ in which two vertices are adjacent if they differ on exactly one coordinate. The cube is a d-regular bipartite graph; we take $\mathcal{E}$ to be the bipartition class consisting of all those vertices the sum of whose coordinates is even.

Before applying the theorem, we need to know something about isoperimetry in the cube. A *Hamming ball centered at* x_0 in Q_d is any set of vertices B satisfying

$$\{u \in V(Q_d) \ : \ \rho(u, x_0) \leq k\} \subseteq B \subseteq \{u \in V(Q_d) \ : \ \rho(u, x_0) \leq k+1\}$$

for some $k < d$, where ρ is the usual graph distance (which in this case coincides with the Hamming distance on $\{0,1\}^d$). An *even* (resp. *odd*) *Hamming ball* is a set of vertices of the form $B \cap \mathcal{E}$ (resp. $B \cap \mathcal{O}$) for some Hamming ball B. We use the following result of Körner and Wei [11]:

Lemma 1.3 *For every $C \subseteq \mathcal{E}$ (resp. $\mathcal{O}$) and $D \subseteq V(Q_d)$, there exists an even (resp. odd) Hamming ball C' and a set D' such that $|C'| = |C|$, $|D'| = |D|$ and $\rho(C', D') \geq \rho(C, D)$.*

We also use a result concerning the sums of binomial coefficients which follows from the Chernoff bounds [4] (see also [2], p.11):

$$\sum_{i=0}^{[cN]} \binom{N}{i} \leq 2^{H(c)N} \qquad \text{for } c \leq \tfrac{1}{2}, \tag{7}$$

where $H(x) = -x \log x - (1-x) \log(1-x)$ is the usual binary entropy function and $[x]$ denotes the integer part of x.

Claim 1.4 $\delta(Q_d) = \Omega(1/\sqrt{d})$.

Proof: We will show that for any $A \subseteq \mathcal{E}$ (or $\mathcal{O}$) with $|A| \leq M/2$, we have $(|N(A)| - |A|)/|N(A)| = \Omega(1/\sqrt{d})$. Without loss of generality, we may assume that $A \subseteq \mathcal{E}$. Let such an A be given. Applying Lemma 1.3 with $C = A$ and $D = V(Q_d) \setminus (A \cup N(A))$, we find that there exists an even Hamming ball A' with $|A'| = |A|$ and $|N(A)| \geq |N(A')|$. So we may assume that A is an even Hamming ball.

We consider only the case where A is centered at an even vertex, without loss of generality $\underline{0} := \{0, \ldots, 0\}$, the other case being similar. In this case,

$$\{v \in \mathcal{E} \ : \ \rho(v, \underline{0}) \leq k\} \subseteq A \subseteq \{v \in \mathcal{E} \ : \ \rho(v, \underline{0}) \leq k+2\}$$

for some even $k \leq d/2$ (the bound on k coming from the fact that $|A| \leq M/2$). For each $0 \leq i \leq (k+2)/2$, set $B_i = A \cap \{v : \rho(v, \underline{0}) = 2i\}$, and $N^+(B_i) = N(B_i) \cap \{u : \rho(u, \underline{0}) = 2i+1\}$. It is clear that $N(A) = \cup_{0 \leq i \leq (k+2)/2} N^+(B_i)$ and that

$$\cup_{0 \leq i \leq k/2} N^+(B_i) = \{v \in \mathcal{O} \ : \ \rho(v, \underline{0}) \leq k+1\}.$$

Also, observe that for all i

$$\frac{|B_i|}{|N^+(B_i)|} \leq \frac{2i+1}{d-2i}, \tag{8}$$

from which it follows that

$$|N^+(B_{(k+2)/2})| - |B_{(k+2)/2}| \geq \frac{-10}{d-4}|N^+(B_{(k+2)/2})| \geq \frac{-20}{d}\binom{d}{k+3}.$$

Indeed, (8) is an equality except when $i = (k+2)/2$, in which case it follows from the fact that each vertex in $B_{(k+2)/2}$ has exactly $d-(k+2)$ neighbours in $N^+(B_{(k+2)/2})$, and each vertex in $N^+(B_{(k+2)/2})$ has at most $(k+2)+1$ neighbours in $B_{(k+2)/2}$.

We deal first with the case $k \leq d/4$. In this case, (8) gives

$$\frac{|B_i|}{|N^+(B_i)|} \leq \frac{2}{3}$$

and so $(|N(A)| - |A|)/|N(A)| \geq 1/3$.

For $d/4 < k \leq d/2$ and c any constant, we claim that

$$\sum_{i=0}^{k+c} \binom{d}{i} = O\left(\sqrt{d}\binom{d}{k+c}\right), \tag{9}$$

where the constant in the O depends on c. For $k+c \leq d/2$, this follows from (7):

$$\sum_{i=0}^{k+c} \binom{d}{i} \leq 2^{H(\frac{d}{k+c})d} = O\left(\sqrt{d}\binom{d}{k+c}\right),$$

the equality being an easy consequence of Stirling's approximation, while for $k+c \geq d/2$ we have

$$\sum_{i=0}^{k+c} \binom{d}{i} \leq 2^d = O\left(\sqrt{d}\binom{d}{k+c}\right),$$

once again from Stirling's approximation. For $d/4 < k \leq d/2$, we therefore have

$$\begin{aligned}
|N(A)| - |A| &= |N^+(B_{(k+2)/2})| - |B_{(k+2)/2}| + \sum_{i=0}^{k/2} \binom{d}{2i+1} - \binom{d}{2i} \\
&\geq \frac{-20}{d}\binom{d}{k+3} + \sum_{i=0}^{k/2} \binom{d-1}{2i} \frac{d(d-1)}{(2i+1)(d-2i)} \\
&= \Omega\left(2\binom{d-1}{k} - \frac{20}{d}\binom{d}{k+3}\right) \\
&= \Omega\left(\left(1 - \frac{20}{d}\right)\binom{d}{k+3}\right),
\end{aligned} \tag{10}$$

and

$$
\begin{aligned}
|N(A)| &= |N^+(B_{(k+2)/2})| + \sum_{i=0}^{k/2} \binom{d}{2i+1} \\
&\leq \sum_{i=0}^{k+3} \binom{d}{i} = O\left(\sqrt{d}\binom{d}{k+3}\right), \qquad (11)
\end{aligned}
$$

the last equality following from (9). Combining (10) and (11), we may now conclude that

$$
\frac{|N(A)| - |A|}{|N(A)|} = \Omega\left(\frac{1}{\sqrt{d}}\right).
$$

∎

We therefore have that if $c_2 > 0$ is a suitably large constant, then (5) is satisfied as long as $\lambda > c_2 \frac{\log^{3/2} d}{d^{1/4}}$. So a corollary of Theorem 1.2 is

Corollary 1.5 *There are constants $c_2, c_3 > 0$ such that whenever $\lambda > c_2 \frac{\log^{3/2} d}{d^{1/4}}$ we have $\tau_{\mathcal{M}_\lambda(Q_d)}$ at least*

$$
\exp_2\left\{\Omega\left(M \frac{\log^3(1+\lambda)}{\sqrt{d}\,(1+\log(1+\lambda))\,(c_3 \log d + \log(1+\lambda)) \log\left(2 + \frac{1}{\log(1+\lambda)}\right)}\right)\right\}.
$$

In particular,

$$
\tau_{\mathcal{M}_\lambda(Q_d)} \geq \begin{cases} \exp_2\left\{\frac{\log^3(1+\lambda)}{\sqrt{d}\log^2 d}\right\} & \text{if } c_2 \frac{\log^{3/2} d}{d^{1/4}} < \lambda < O(1), \\ \exp_2\left\{\frac{\log^2(1+\lambda)}{\sqrt{d}\log d}\right\} & \text{if } \Omega(1) < \lambda < O(d), \\ \exp_2\left\{\frac{\log(1+\lambda)}{\sqrt{d}}\right\} & \text{if } \Omega(d) < \lambda. \end{cases}
$$

Remark 3. Using (6) in place of (5) (which we may do, since Q_d has bounded codegree) we may improve the bound on λ in Corollary 1.5 to $\lambda > c_4 \frac{\log d}{d^{1/4}}$ for some constant $c_4 > 0$.

2 Proof of Theorem 1.2

Throughout this section we fix Σ satisfying the conditions of Theorem 1.2. 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 $J \in \mathcal{I}_b$. Combining this observation with reversibility, we have

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

and so (using the trivial lower bound $w_\lambda(\mathcal{I}_{\mathcal{E}}) \geq (1+\lambda)^M$)

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

Thus (recalling (2)) 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 (3). Note that $\mathcal{I}_b \subseteq \mathcal{I}^{triv} \cup \mathcal{I}^{nt}$. We will show that as long as λ satisfies (5),

$$w_\lambda(\mathcal{I}^{triv} \cup \mathcal{I}^{nt}) \leq (1+\lambda)^M \exp_2\left\{-\Omega\left(M\alpha(\lambda)\beta(\lambda)\delta\right)\right\}, \tag{12}$$

from which Theorem 1.2 follows via (2).

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

$$4\alpha(\lambda) \log \frac{1}{\alpha(\lambda)} \leq \frac{\log(1+\lambda)}{2\left(1+\log(1+\lambda)\right)}. \tag{13}$$

To see this, first set

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

Note that for all $\lambda > 0$, $0 < \gamma(\lambda) < 1$. Now (13) reduces 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 $0 < x < 1/2$,

$$x \leq H(x) \leq 2x \log \frac{1}{x}, \tag{14}$$

(where recall $H(x) = -x \log x - (1-x)\log(1-x)$). With the inequalities justified below, we have

$$\begin{aligned}
w_\lambda(\mathcal{I}^{triv}) &\leq \binom{M}{\leq \alpha(\lambda)M}^2 (1+\lambda)^{2\alpha(\lambda)M} \\
&\leq \exp_2\{2H(\alpha(\lambda))M + 2\alpha(\lambda)M\log(1+\lambda)\} &(15)\\
&\leq \exp_2\{2MH(\alpha(\lambda))(1+\log(1+\lambda))\} &(16)\\
&\leq \exp_2\left\{4M\alpha(\lambda)\log\frac{1}{\alpha(\lambda)}(1+\log(1+\lambda))\right\} &(17)\\
&\leq \exp_2\left\{M\frac{\log(1+\lambda)}{2}\right\} &(18)\\
&\leq (1+\lambda)^{\frac{M}{2}}. &(19)
\end{aligned}$$

Here (and throughout) we use $\binom{n}{\leq k}$ for $\sum_{i\leq k}\binom{n}{i}$. In (15), we are using (7), which is applicable by (4). In (16) we are using the first inequality in (14) and in (17) we are using the second (again, both of these are applicable by (4).) Finally in (18) we are using (13).

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 $|[I \cap \mathcal{E}]| \leq M/2$, and set

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

Define *small on* $\mathcal{O}$ and $\mathcal{I}^{nt}_{\mathcal{O}}$ similarly. A simple argument, based on the fact that Σ 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

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

We may assume without loss of generality that

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

so that it is enough to show that

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

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

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

We now state the main theorem, whose proof (given in Section 3) is based on ideas of A. Sapozhenko.

Theorem 2.1 *Let Σ be any graph satisfying the assumptions of Theorem 1.2. 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 λ satisfying (5).

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

$$\begin{aligned}
w_\lambda(\mathcal{I}_{\mathcal{E}}^{nt}) &\leq (1+\lambda)^M M^2 \exp_2\{-\Omega(M\alpha(\lambda)\beta(\lambda)\delta)\} \\
&\leq (1+\lambda)^M \exp_2\{-\Omega(M\alpha(\lambda)\beta(\lambda)\delta)\}. \qquad (20)
\end{aligned}$$

To see that the factor of M^2 may be absorbed into the exponent, note that by hypothesis, $2M > d^2$ and so $M^2 \leq \exp_2\{O(M\log d/d^2)\}$, and that combining (4) and (5) we have $\alpha(\lambda)\beta(\lambda)\delta \geq \Omega(\log d/d\sqrt{d})$.

Combining (20) and (19) we get (12) and hence Theorem 1.2.

3 Proof of Theorem 2.1

For $u, v \in V$ and $A, B \subseteq V$ we write $u \sim v$ if there is an edge in Σ joining u and v, $\nabla(A)$ 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.

Throughout this section, we fix Σ satisfying the assumptions of Theorem 1.2. 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 $A \in \mathcal{A}(a,g)$ we always write G for $N(A)$ and set $t = g - a$. For $S, T \subseteq V$ we write $\nabla(S)$ for the set of edges having one end in S and (if $S \cap T = \emptyset$) $\nabla(S,T)$ for the set of edges having one end in each of S, T. Note that for $A \in \mathcal{A}$,

$$|\nabla(G, \mathcal{E} \setminus [A])| = dg - da = td. \qquad (21)$$

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 [16], 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 [12]. Our Lemma 3.3 is a modification of a lemma in [15], and our overall approach is similar to [16]. See e.g. [7] 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

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

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

$$F \subseteq G, \quad S \supseteq [A], \tag{22}$$

$$d_F(u) \geq d - \psi \quad \forall u \in S \tag{23}$$

and

$$d_{\mathcal{E}\setminus S}(v) \geq d - \psi \quad \forall v \in \mathcal{O} \setminus F. \tag{24}$$

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", (23) says that if $x \in \mathcal{E}$ is in "approximate $[A]$" then almost all of its neighbours are in "approximate G", while (24) 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 ψ-approximations that will be of use later.

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

$$|S| \leq |F| + \frac{2t\psi}{d - \psi}. \tag{25}$$

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 (25).

∎

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 ψ-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 ψ-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 ψ-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 (25) we have*

$$\sum w_\lambda(A) \leq \max\left\{(1+\lambda)^{g-\gamma t}, \binom{3dg}{\leq \frac{2t\psi}{d-\psi} + \gamma t}(1+\lambda)^{g-t}\right\} \tag{26}$$

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

$$\begin{aligned}\sum_{i\le k}\binom{n}{i} &\le (1+O(k/n))\binom{n}{k} \\ &\le (1+O(k/n))(en/k)^k \\ &\le \exp_2\{(1+o(1))k\log(n/k)\}. \end{aligned} \tag{27}$$

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 ψ and γ we have $\gamma > \frac{-\psi}{d-\psi}$, and so

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

The bound in (26) 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 (27)). For our choice of ψ and γ this is at most

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

which in turn is at most

$$(1+\lambda)^g\exp_2\left\{O\left(t\frac{\log(d^5/\delta)}{\sqrt{d}}\right) - t\frac{\log^2(1+\lambda)}{\log(1+\lambda)+\log(d^5/\delta)}\right\}. \tag{28}$$

The bounds in Lemmas 3.3 and 3.2 are at most

$$\exp_2\left\{O\left(g\frac{\log d}{d}+t\frac{\log(d^5/\delta)}{\sqrt{d}}\right)\right\} \quad \text{and} \quad \exp_2\left\{O\left(g\frac{\log^2 d}{d}\right)\right\} \tag{29}$$

respectively. For the latter bound, we are using the assumption $a \ge \alpha(M)$ of Theorem 2.1 and the fact that $\lambda > 1/\sqrt{d}$, which together imply (via (4)) that

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

Combining (29) with (28), we get

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

Noting that $t \ge \delta g$ always, we find that if λ satisfies (5) with a suitably large constant c_1, then

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

and so we get Theorem 2.1.

We now turn to the proofs of Lemma 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 [13] and Stein [17]. For a bipartite graph Γ 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* Γ *as above satisfies* $|N(x)| \ge p$ *for each* $x \in P$ *and* $|N(y)| \le q$ *for each* $y \in Q$, *then* P *is covered by some* $Q' \subseteq Q$ *with*

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

Applying the lemma with Γ the subgraph of Σ 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 $\frac{2g\log d}{d}$. Lemma 3.2 follows.

∎

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 (i.e., with $N(F_0) \supseteq [A]$), an output $(F, S) \in 2^{\mathcal{O}} \times 2^{\mathcal{E}}$ which is a ψ-approximation for A (i.e, which satisfies (22), (23) and (24)). The idea for the algorithm is from [15]. To begin, fix a linear ordering $\ll$ of V.

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

Step 2: If $\{v \in \mathcal{O}\setminus G : d_{S_1}(v) > \psi\} \ne \emptyset$, pick the smallest (with respect to $\ll$) v in this set and update S_1 by $S_1 \longleftarrow S_1 \setminus N(v)$. 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 $\{w \in [A] : d_{G\setminus F_2}(w) > \psi\} \ne \emptyset$, pick the smallest (with respect to $\ll$) w in this set and update F_2 by $F_2 \longleftarrow F_2 \cup N(w)$. Repeat this until $\{w \in [A] : d_{G\setminus F_2}(w) > \psi\} = \emptyset$. Then set $F = F_2$ and $S = S_2 \cap \{w \in \mathcal{E} : d_F(w) \ge d - \psi\}$ and stop.

Claim 3.6 *The output of the degree algorithm is a ψ-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, (23) is satisfied. To verify (24), 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 $S_2 \supseteq S$.

∎

Remark 4. The alert reader may have noticed that if we replace $d/2$ by ψ in Step 1, then the output of Step 2 is already a ψ-approximation for A. The three-step algorithm, however, is needed to obtain the right bound on $\beta(\lambda)$ in Theorem 1.2; see the remark following the proof of Claim 3.7.

Claim 3.7 *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 3.7: 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

$$\binom{2g \log d}{\leq \frac{2g}{d}} \tag{30}$$

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 (21) a set of size dt, so $|S_1 \setminus [A]| \leq 2t$. Each v used in Step 2 reduces this by at least ψ, 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 $d^4|F_0| \le 2d^3 g \log d$. So the total number of outputs for Step 2 is

$$\binom{2d^3 g \log d}{\le \frac{2t}{\psi}} \tag{31}$$

At the start of Step 3, each $y \in G \backslash F_2$ contributes at least $d - \psi$ edges to $\nabla(G, \mathcal{E} \backslash [A])$, so $|G \setminus F_2| \le dt/(d-\psi)$. Each w used in step 3 reduces this by at least ψ, 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

$$\binom{2g \log d}{\le \frac{td}{(d-\psi)\psi}}. \tag{32}$$

The claim follows from (30), (31) and (32).

∎

Remark 5. The bound in Claim 3.7 is at most $\exp_2\{O(g \log d/d + t \log(d^5/\delta)/\psi)\}$. If we replace $d/2$ by ψ 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:

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

(Each iteration of Step 1 now reduces $G \setminus F$ by at least ψ). Using this bound in the proof of Theorem 1.2 instead of the stronger bound given by the three-step degree algorithm would ultimately lead to a weaker bound on $\beta(\lambda)$ in (5). Step 1 of the degree algorithm may be though 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 of A by picking a subset of S. If S is large, we first specify G. Note that by (25) 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

$$(1+\lambda)^{g-\gamma t}. \tag{33}$$

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

$$\binom{3dg}{\leq \frac{2t\psi}{d-\psi} + \gamma t}$$

and the total weight of outputs is at most

$$\binom{3dg}{\leq \frac{2t\psi}{d-\psi} + \gamma t}(1+\lambda)^{g-t}. \tag{34}$$

The lemma follows from (33) and (34).

∎

Acknowledgment. This work originated 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, *Stoch. Proc. Appl.* **49** (1994), 179-197.

[2] B. Bollobás, *Random Graphs,* Cambridge University Press, Cambridge, 2001.

[3] 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.

[4] 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.

[5] R. Diestel, *Graph Theory,* Springer, New York, 2000.

[6] R.L. Dobrushin, The problem of uniqueness of a Gibbs random field and the problem of phase transition, *Functional Anal. Appl.* **2** (1968), 302-312.

[7] D. Galvin, On homomorphisms from the Hamming cube to $\mathbf{Z}$, to appear in *Isr. J. Math.*

[8] D. Galvin and J. Kahn, On phase transition in the hard-core model on $\mathbf{Z}^d$, to appear in *Comb. Prob. Comp.*

[9] M. Jerrum and A. Sinclair, The Monte Carlo Markov Chain method: an approach to approximate counting and integration, in *Approximation Alorithms for NP-hard problems*, PWS, 1996.

[10] F. Kelly, Loss networks, *Ann. App. Prob.* **1** (1991) 319–378.

[11] J. Körner and V. Wei, Odd and even Hamming spheres also have minimum boundary, *Discrete Math.* **51** (1984), 147–165.

[12] A. D. Korshunov and A. A. Sapozhenko, The number of binary codes with distance 2, *Problemy Kibernet.* **40** (1983), 111–130. (Russian)

[13] L. Lovász, On the ratio of optimal integral and fractional covers, *Discrete Math.* **13** (1975) 383–390.

[14] M. Luby and E. Vigoda, Fast convergence of the Glauber dynamics for sampling independent sets, *Random Structures and Algorithms* **15** (1999), 229–241.

[15] 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)

[16] 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)

[17] S. K. Stein, Two combinatorial covering theorems, *J. Combinatorial Th. (A)* **16** (1974), 391–397.