On Markov chains for randomly H-colouring a graph

Colin Cooper* Martin Dyer† Alan Frieze‡

September 15, 2000

Abstract

Let $H = (W, F)$ be a graph without multiple edges, but with the possibility of having loops. Let $G = (V, E)$ be a simple graph. A *homomorphism* c is a map $c : V \to W$ with the property that $(v, w) \in E$ implies $(c(v), c(w)) \in F$. We will often refer to $c(v)$ as the *colour* of v and c as an *H-colouring* of G. We consider the problem of choosing a random H-colouring of G by Markov Chain Monte Carlo. The probabilistic model we consider includes random proper colourings, random independent sets and the Widom-Rowlinson and Beach models of Statistical Physics. We prove negative results for uniform sampling and a positive result for weighted sampling when H is a tree.

1 Introduction

We consider a class of graph labellings which are the natural generalisation of well-studied problems such as *proper k-colourings* and *independent sets.*

*School of Mathematical Sciences, University of North London, London N7 8DB, UK

†School of Computer Studies, University of Leeds, Leeds LS2 9JT, UK. Supported by EC Working Group RAND2.

‡Department of Mathematical Sciences, Carnegie Mellon University, Pittsburgh PA15213, USA. Supported by grants from NSF and EPSRC.

These labellings arise as *homomorphisms* from a fixed graph H to the target graph G.

To be precise, let $H = (W, F)$ be a graph without multiple edges, but with the possibility of having loops, with $h = |W|$. Let $G = (V, E)$ be a simple graph, with $n = |V|$. A *homomorphism* c is a map $c : V \to W$ with the property that $(v, w) \in E$ implies $(c(v), c(w)) \in F$. We will often refer to $c(v)$ as the *colour* of v and c as an *H-colouring* of G. As an example, when H is the complete graph K_s with no loops, a homomorphism c defines a *proper* s-colouring of G. As a second example, when H is a single edge (v, w) plus a loop (w, w) then $c^{-1}(v)$ defines an independent set of G and the H-colourings are in 1-1 correspondence with the independent sets of G. A third example is when H is a path of two edges with a loop on all vertices. Then a homomorphism c defines a *Widom-Rowlinson* configuration on G, as studied in statistical physics [2].

Such H-colourings have been considered by several authors. For example, it was shown by Hell and Nešetřil [12] that deciding whether or not there exists at least one H-colouring is NP-complete unless H has at least one loop or is bipartite. Otherwise the decision problem is (trivially) in P. Dyer and Greenhill [8] have recently shown that counting the number of H-colourings is #P-complete unless every component of H is a complete graph with all loops present or a complete bipartite graph with no loops present. Otherwise the counting problem is in P. Moreover, the #P-completeness remains true if G has bounded vertex degrees. This is in contrast to the decision case, which seems rather more complicated in the bounded-degree case. See [11] for further information.

Here, we focus further on the problem of *counting* H-colourings, or more generally determining a "weighted sum" over all configurations, as in the partition functions of statistical physics. (See, for example, [2].) Given the negative results of [8] it is of interest to consider algorithms which try to estimate *approximately* the number of H-colourings of G, in particular using methods of *randomized approximation.* Here the class of H-colouring problems is interesting, both because it contains problems which are important in their own right, but more generally because it provides a "test bed" for existing approaches to randomized approximate counting. In this respect, it is well known that positive results can be obtained for many H if we are able to generate a (nearly) uniform random H-colouring in polynomial time, i.e.

in time polynomial in $n = |V|$. We will not examine here the precise conditions on H needed for this to be true. See [14] for fairly general conditions, or [7] for the weighted case.

There have been several attempts along these lines for specific graphs H, e.g. [13, 17, 15, 9]. In all cases an ergodic Markov chain $\mathcal{M}$ is constructed with state space Ω equal to the set of H-colourings of G. Furthermore, the steady state distribution of $\mathcal{M}$ is uniform. This is usually called the *Markov chain Monte Carlo* (MCMC) method.

In [5], Dyer, Frieze and Jerrum proved two *negative* results for the case of generating a random *independent set* of G by this method. (As previously remarked, this is a special case of H-colouring.) Our first result below is a generalisation of the first theorem given there, and the following definition follows that in [5].

A Markov chain $\mathcal{M}$ on Ω is *d-cautious* if $h(c, c') \leq d$ whenever it is possible to make a transition from c to c' in one step of the chain. Here the quantity $h(c, c') = |\{v \in V : c(v) \neq c'(v)\}|$ is the *Hamming distance* between c and c'. Many chains that have been analysed in this connection are d-cautious with $d = O(1)$, i.e. bounded independently of n.

First observe that if H is not connected, then no d-cautious chain with $d < n$ can be ergodic on Ω, since it cannot move between colourings corresponding to different components of H. This could be overcome simply by running a different chain for each component. But either way we may as well assume H is connected. A similar difficulty arises when H is bipartite with no loops, in which case we may as well assume G is bipartite, since otherwise Ω is empty. Since the bipartition (W_1, W_2) of H can be assigned to the bipartition (V_1, V_2) of G in two ways, no d-cautious chain with $d < n$ can be ergodic on Ω, because it cannot move between these two distinct sets of colourings. Again this could be overcome by considering these two sets separately. So, to handle this triviality, let us call $\mathcal{M}$ *semi-ergodic* if it is ergodic on both $\Omega_1, \Omega_2 \subseteq \Omega$, where

$$\Omega_i = \{c \in \Omega : c^{-1}(V_1) \subseteq W_i,\ c^{-1}(V_2) \subseteq W_{3-i}\} \qquad (i = 1, 2).$$

Then we will show that it is not sufficient to remove this obvious obstruction to rapid mixing. We prove the following negative result, ruling out a large class of Markov chains with which one might try to sample H-colourings:

Theorem 1 *Suppose H is connected and has at least one loop, but is not a complete graph with a loop on every vertex. Then there exist constants $\xi = \xi(H) > 0$, $\zeta = \zeta(H) > 1$, and a positive integer $r_0 = r_0(H)$, such that the following is true:*

For every *ξn-cautious ergodic Markov chain $\mathcal{M}$ with uniform steady state on the set Ω of H-colourings of G, there exist n-vertex r-regular graphs for all $r \geq r_0$ such that the mixing time of $\mathcal{M}$ is at least ζ^n.*

If H is a connected bipartite graph, but not a complete bipartite graph, then the corresponding statement holds for semi-ergodic chains on each of the sets Ω_1, Ω_2 defined above.

Note that the emphasis on either bipartiteness or having at least one loop is to avoid the complexity issue of finding at least one colouring to initiate the chain. Besides, if we cannot determine if the number of colourings is zero or not, we obviously have no hope of computing any finite relative approximation to that number. Note also that there is more than one definition of mixing time of a Markov chain, but we may observe that our theorem seems insensitive to the exact choice of definition.

We prove Theorem 1 in section 2. In section 3, we consider the ergodicity of the Markov chain which "recolours" one vertex at a time. This is often called the *Glauber dynamics.* We give necessary and sufficient conditions for the chain to be connected on *all* graphs G of sufficiently large, but bounded, degree in the case where H has a loop or is bipartite. (Without this assumption, we would be unable to find any H-colouring in the worst case, so ergodicity is a less interesting question.) Our theorem for the case where H has a loop (Theorem 2) was first proved by Brightwell and Winkler [3], as part of a longer theorem concerning Gibbs measures on graphs. We give a proof for completeness and to establish notation for our discussion of the bipartite case.

Positive results which hold for all graphs of large degree seem hard to come by. Nevertheless, in section 4, we construct a sampling algorithm for suitably weighted configurations in the case where H is a tree with a loop on every vertex.

2 Proof of Theorem 1

We consider any Markov chain which does not change the colour of "sufficiently many" vertices at each step, whose asymptotic distribution is the uniform measure on the set of colourings of a suitably chosen graph G. We identify a set of colourings of of G of measure at most $\frac{1}{2}$, such that the "boundary" of this set for the transitions of the chain has very much smaller measure. Thus, if started in such a set, the chain will effectively "get stuck", since it has only a small probability of escaping at each step. Hence it must take a very long time to closely approximate the uniform distribution.

The proof will be split into several cases. In all cases, the "bad" graphs we use are either random r-regular graphs or r-regular bipartite graphs for sufficiently large (but bounded) r.

To make this precise, let $\Omega = \Omega_{H,G}$ denote the set of H-colourings of G. Our proof strategy will be to choose a random graph from some probability distribution and then to identify a subset A of Ω, $|A| \leq \frac{1}{2}|\Omega|$ such that **whp**[1]

$$\frac{|N_d(A)|}{|A|} \leq \lambda^n \tag{1}$$

where $\lambda < 1$ is a constant, dependent only on H, and

$$N_d(A) = \{c \in \Omega \setminus A : \ h(c, c') \leq d = \xi n \text{ for some } c' \in A\}.$$

Then for any d-cautious chain $\mathcal{M}$ with transition matrix P we have that the *conductance* [16] $\Phi_{\mathcal{M}}$ of $\mathcal{M}$ is at most

$$\Phi_A = |A|^{-1} \sum_{x \in A, y \in N_d(A)} P(x, y) \leq |A|^{-1} \sum_{x \in \Omega, y \in N_d(A)} P(x, y) = \frac{|N_d(A)|}{|A|},$$

where the final equality uses the equations satisfied by the (uniform) stationary distribution. It follows from (1) that $\Phi_{\mathcal{M}}$ is exponentially small and hence that the mixing time of $\mathcal{M}$ will be exponentially large.

[1]We use the notation **whp** as shorthand for "with probability tending to 1 as $n \to \infty$".

H has a loop

Let H^o denote the subgraph of H induced by the looped vertices. We consider two cases.

Case 1: H^o contains at least two distinct maximal cliques, both of which contain at least two vertices.

Let L, L' be two such cliques where $2 \leq \ell = |L| \leq |L'|$.

Let $G = G_{r,n}$ be a random r-regular graph with vertex set $V = \{1, 2, \ldots, n\}$. Here r is a constant which is sufficiently large for certain inequalities to be valid. The proof uses other parameters α, β, γ and takes $d = \frac{\beta}{\ell} n$. (To avoid tedious detail, we assume such quantities are integral where convenient.) Rather than define these constants explicitly, we simply give (consistent) inequalities which they must satisfy in order for the proof to be valid. Basically, we need the following quantities to be sufficiently small: α, β/α, γ/α^2.

Now let

$$A = \{c \in \Omega : \ |c^{-1}(v)| \geq (1 - \alpha)n/\ell \text{ for all } v \in L\}.$$

Now **whp** G will have the following two properties:

P1 $I \subseteq V$, I independent, implies that $|I| \leq \frac{2 \ln r}{r} n$.

P2 $S, T \subseteq V$, $S \cap T = \emptyset$, $|S| \geq \frac{1-\alpha-\beta}{\ell} n$, $|T| \geq \gamma n$, implies that there is an edge joining S and T.

Property P1 is proved for example in Frieze and Łuczak [10] and property P2 will be confirmed below.

So let G be an r-regular graph with properties P1 and P2. Now consider $N_d(A)$. It follows that

$$N_d(A) \subseteq B = \{c \in \Omega : \begin{array}{l} \text{(i) } \exists v \in L \text{ s.t. } |c^{-1}(v)| < \frac{1-\alpha}{\ell} n \\ \text{(ii) } w \notin L \text{ implies } |c^{-1}(w)| < \gamma n \\ \text{(iii) } |c^{-1}(L)| \geq (1 - h\gamma)n \qquad \} \end{array}$$

Here property (i) follows from the fact that $N_d(A) \cap A = \emptyset$ and for property (ii) note that if $w \notin L$ then either

(a) w has no loop and so $c^{-1}(w)$ is an independent set. In which case

$$|c^{-1}(w)| \le \frac{2\ln r}{r} n < \gamma n$$

using P1 and assuming r is sufficiently large. Or, alternatively,

(b) w has a loop and there exists $v \in L$ such that v is not adjacent to w. Using P2 and $|c^{-1}(v)| \ge \frac{1-\alpha-\beta}{\ell} n$ we see again that property (ii) must hold.

Property (iii) is a trivial consequence of property (ii).

Now let us estimate $|A|$, $|N_d(A)|$ and see that (1) holds. First of all we see that

$$|A| \ge \tfrac{1}{2}\ell^n \tag{2}$$

since (by concentration of measure) almost all colourings c for which $c(V) \subseteq L$ will be members of A, assuming only that α is a positive constant.

Now consider the size of $B \supseteq N_d(A)$. If $c \in B$ then $t = |c^{-1}(L)| \ge (1-h\gamma)n$. For a fixed t there are at most

$$\binom{n}{t} \ell^{t+1} e^{-\epsilon^2 t/2\ell} \tag{3}$$

ways of defining $c^{-1}(L)$, where ϵ satisfies

$$\frac{1-\alpha}{\ell} n = (1-\epsilon)\frac{1-h\gamma}{\ell}, \qquad \text{i.e.} \quad \epsilon = \frac{\alpha - h\gamma}{1-h\gamma} \ge \tfrac{1}{2}\alpha \tag{4}$$

provided γ is sufficiently small.

The explanation for (3) is as follows: $\binom{n}{t}$ is the number of choices for $X = c^{-1}(L)$. Having chosen this there are ℓ^t ways of colouring X. There are ℓ choices for $v \in L$ with $|c^{-1}(v)| < \frac{1-\alpha}{\ell} n$ and then the final factor $e^{-\epsilon^2 t/\ell}$ is a consequence of the Chernoff bound for the binomial probability that a random ℓ-colouring of X uses v fewer than $\frac{1-\alpha}{\ell} n$ times.

Finally, there are at most $h^{h\gamma n}$ ways of completing a colouring of X to the whole of V. This gives

$$\begin{aligned} |N_d(A)| &\leq \sum_{t=(1-h\gamma)n}^{n} \binom{n}{t} \ell^{t+1} e^{-\epsilon^2 t/2\ell} h^{h\gamma n} \\ &\leq e^{-\epsilon^2(1-h\gamma)n/2\ell} h^{h\gamma n} \sum_{t=(1-h\gamma)n}^{n} \binom{n}{t} \ell^{t+1} \\ &\leq 2\binom{n}{h\gamma n} \ell^{(1-h\gamma)n+1} e^{-\epsilon^2(1-h\gamma)n/2\ell} h^{h\gamma n} \end{aligned}$$

assuming that $h\gamma\ell < (1-h\gamma)/2$, whence the sum telescopes.

Thus

$$\frac{|N_d(A)|}{|A|} \leq ((1+o(1))\psi(h\gamma)\ell^{-h\gamma} e^{-\epsilon^2(1-h\gamma)/2\ell} h^{h\gamma})^n$$

where $\psi(\theta)^{-1} = \theta^\theta(1-\theta)^{1-\theta}$.

Now observe that, if we keep α, β fixed and let γ become small (which requires r to become large), then (1) will hold. (By considering $\gamma = 0$, it is clear that some small enough positive γ will do, by continuity.) Finally, notice that we have not verified that $|A| \leq \frac{1}{2}|\Omega|$. If this is not the case then we replace L by L' to get a new set of colourings A disjoint from that defined w.r.t. L. This completes the proof of Theorem 1 in this case, except to verify property P2.

Proof of property P2: Using the configuration model of a random graph, see for example Bollobás [1].

$$\mathbf{Pr}(\neg\mathrm{P2}) = O\left(\binom{n}{\theta n}\binom{n}{\gamma n}(1-\theta)^{\gamma r n/2}\right) \tag{5}$$

where $\theta = (1-\alpha-\beta)/\ell$. Here $\binom{n}{\theta n}\binom{n}{\gamma n}$ bounds the number of choices for S, T. Clearly the right side of (5) is exponentially small for r sufficiently large.

Case 2: H^o contains a maximal clique L' all of whose neighbours are vertices without loops. (As usual, we say v is a neighbour of L' if v has any neighbour in L'.) It is possible for L' to consist of a single vertex.

Choose $L \subseteq L'$ and $M \subseteq W \setminus L'$ such that the edges between L and M form a maximal complete bipartite subgraph of H. Let $\ell' = |L'| + |M|$, $\ell = |L|$. Thus $\ell \leq \ell'$.

Let $G = G_{r,n,n}$ be a random r-regular bipartite graph with vertex partition $V_1 = \{1, 2, \ldots, n\}$ and $V_2 = \{n+1, n+2, \ldots, 2n\}$.

Now we let

$$A = \{c \in \Omega : \begin{array}{l} \text{(i)} \;\; v \in L \;\; \text{implies} \;\; |c^{-1}(v) \cap V_1| \geq (1-\alpha)n/\ell, \\ \text{(ii)} \; v \in L' \cup M \; \text{implies} \; |c^{-1}(v) \cap V_2| \geq (1-\alpha)n/\ell' \; \} \end{array}$$

Now we assert that for sufficiently large r, **whp** G satisfies

P3 $S \subseteq V_1, |S| \geq \gamma n$, $T \subseteq V_2, |T| \geq \frac{1-\alpha-\beta}{\ell} n$ implies that there is an edge joining S and T.

The proof of this is similar to that for P2 and is omitted.

If $d = \beta n/\ell'$, then $N_d(A) \subseteq B$, where

$$B = \{c \in \Omega : \begin{array}{l} \text{(i)} \; |c^{-1}(L) \cap V_1| \geq (1-h\gamma)n, \; \text{and} \\ \quad\;\; |c^{-1}(L' \cup M) \cap V_2| \geq (1-h\gamma)n, \\ \text{(ii)} \; v \in L \; \text{implies} \; |c^{-1}(v) \cap V_1| \geq (1-\alpha-\beta)n/\ell, \; \text{and} \\ \quad\;\; v \in L' \cup M \; \text{implies} \; |c^{-1}(v) \cap V_2| \geq (1-\alpha-\beta)n/\ell', \\ \text{(iii)} \; \exists v \in L' \cup M \; \text{s.t.} \; |c^{-1}(v) \cap V_2| < (1-\alpha)n/\ell', \\ \quad\;\; \text{or} \;\; \exists v \in L \;\; \text{s.t.} \; |c^{-1}(v) \cap V_1| < (1-\alpha)n/\ell \\ \text{(iv)} \; w \notin L' \cup M \; \text{implies} \; |c^{-1}(w)| < \gamma n \\ \text{(v)} \; v \in M \cup (L' \setminus L) \; \text{implies} \; |c^{-1}(v) \cap V_1| < \gamma n \qquad \} \end{array}$$

Property (ii) follows from the definition of $N_d(A)$. Properties (iv) and (v) then follow from (ii), and the fact that L, M induce a maximal bipartite sub-graph. Property (i) then follows from properties (iv) and (v). Property (iii) follows from the fact that $N_d(A) \cap A = \emptyset$. Arguing similarly to Case 1, we see that

$$\begin{aligned} |A| &\geq \tfrac{1}{2}(\ell'\ell)^n \\ |N_d(A)| &\leq 2 \sum_{t=(1-h\gamma)n}^{n} \binom{n}{t} (\ell'\ell)^n e^{-\epsilon^2(1-h\gamma)n/2\ell'} h^{2h\gamma n}, \end{aligned}$$

where ϵ is as in (4), but with ℓ' replacing ℓ, and $t = |c^{-1}(L)|$. The factor 2 arises from the two cases in property (iii) of B.

Now, by making β small with respect to α, and γ even smaller, we can ensure that (1) holds. If $|A| > \frac{1}{2}|\Omega|$, we can interchange the roles of V_1 and V_2 in the definition of A and obtain a disjoint set of colourings A' for which (1) holds and which satisfies $|A'| \leq \frac{1}{2}|\Omega|$.

Cases 1 and 2 cover all possibilities for H having a loop, except when it is complete with a loop on every vertex.

H is unlooped bipartite

Suppose H has vertex bipartition W_1, W_2 and is not a complete bipartite graph. Without loss of generality, we will consider $\mathcal{M}$ restricted to Ω_1.

Let $L_1, L_1' \subseteq W_1, L_2, L_2' \subseteq W_2$ induce distinct maximal complete bipartite subgraphs of H with vertex bipartitions (L_1, L_2) and (L_1', L_2') respectively. Let $\ell_i = |L_i|$ $(i = 1, 2)$, and $\ell = \ell_1 + \ell_2$. We choose G to be the random r-regular bipartite graph defined in Case 2 above. Now define

$$\begin{aligned} A = \{c \in \Omega_1 : \quad & \text{(i)}\ \ v \in L_1 \text{ implies } |c^{-1}(v) \cap V_1| \geq (1-\alpha)n/\ell_1, \\ & \text{(ii)}\ v \in L_2 \text{ implies } |c^{-1}(v) \cap V_2| \geq (1-\alpha)n/\ell_2\ \}. \end{aligned}$$

We may assume without loss that $|A| \leq \frac{1}{2}|\Omega_1|$, since otherwise we may use A' defined similarly by L_1', L_2', noting that $A \cap A' = \emptyset$. Now, letting $d = \beta n/\ell$, we have $N_d(A) \subseteq B$, where

$$\begin{aligned} B = \{c \in \Omega_1 : \quad & \text{(i)}\ |c^{-1}(L_1) \cap V_1| \geq (1-h\gamma)n, \text{ and} \\ & \quad |c^{-1}(L_2) \cap V_2| \geq (1-h\gamma)n, \\ & \text{(ii)}\ v \in L_1 \text{ implies } |c^{-1}(v) \cap V_1| \geq (1-\alpha-\beta)n/\ell_1, \text{ and} \\ & \quad v \in L_2 \text{ implies } |c^{-1}(v) \cap V_2| \geq (1-\alpha-\beta)n/\ell_2, \\ & \text{(iii)}\ \exists v \in L_1 \text{ s.t. } |c^{-1}(v) \cap V_1| < (1-\alpha)n/\ell_1, \\ & \quad \text{or } \exists v \in L_2 \text{ s.t. } |c^{-1}(v) \cap V_2| < (1-\alpha)n/\ell_2 \\ & \text{(iv)}\ w \notin L_1 \text{ implies } |c^{-1}(w) \cap V_1| < \gamma n, \text{ and} \\ & \quad w \notin L_2 \text{ implies } |c^{-1}(w) \cap V_2| < \gamma n \qquad \} \end{aligned}$$

Property (ii) follows from the definition of $N_d(A)$. Property (iv) then follows from (ii) and the fact that L_1, L_2 induce a maximal complete bipartite subgraph, using P3. Property (i) then follows from property (iv), and property (iii) follows from $N_d(A) \cap A = \emptyset$. Arguing as in Case 2 above, we see that

$$\begin{aligned} |A| &\geq \tfrac{1}{2}(\ell_1\ell_2)^n \\ |N_d(A)| &\leq 2 \sum_{t=(1-h\gamma)n}^{n} \binom{n}{t} (\ell_1\ell_2)^n e^{-\epsilon^2(1-h\gamma)n/2\ell} h^{2h\gamma n}, \end{aligned}$$

where ϵ is again as in (4), and $t = |c^{-1}(L_i)|$ ($i \in \{1, 2\}$). The factor 2 again arises from the two cases in property (iii) of B. The rest of the argument proceeds exactly as in Case 2, and the proof of Theorem 1 is complete. $\square$

3 Connectedness of chain

In this section, we examine the ergodicity of the simplest Markov chain on H-colourings, where we try to re-colour at random a single randomly chosen vertex at each step. This is often called the *Glauber dynamics* and, in our above terminology, is a 1-cautious chain. This chain is known, for example, to be ergodic on the set of proper q-colourings of any graph of maximum degree Δ provided $q \geq \Delta + 2$. We will denote the Glauber dynamics for H-colouring G by $\mathcal{G}_H(G)$. Note that there is little loss in assuming H to be connected. Otherwise we should consider each component separately, since $\mathcal{G}_H(G)$ is clearly disconnected.

We now consider general conditions on $H = (W, F)$ under which $\mathcal{G}_H(G)$ is ergodic for all G. To this end, for $w \in W$, let $N_H(w) = \{w' \in W : \{w, w'\} \in F\}$. Now, if $w_1, w_2 \in W$, we will say that w_2 *dominates* w_1 if $N_H(w_1) \subseteq N_H(w_2)$. If $N_H(w_1) = N_H(w_2)$, we will say w_1 and w_2 are *equivalent*. Consider the following procedure:

CLOSURE(H) : A

(0) $A \leftarrow H$.

(1) **while** there exists w_2 dominating w_1 in A **do** $N_A(w_1) \leftarrow N_A(w_2)$.

The following theorem was first proved by Brightwell and Winkler [3], as part of a longer theorem concerning Gibbs measures on graphs. There an operation similar to that used in line (1) of CLOSURE is called "folding", and any graph that folds to to a single vertex is called "dismantlable". We give a proof here for completeness.

Theorem 2 ([3]) *Suppose H is connected and has a loop. Then $\mathcal{G}_H(G)$ is ergodic for all G if and only if* CLOSURE(H) *is complete, with loops on all vertices. When it is not ergodic for all G, there exists a graph G_0 with $\Delta(G_0) = \max_{w \in W} |N_H(w)|$ for which this remains true.*

Proof Suppose $A = \text{CLOSURE}(H)$ is complete looped. Let s be the number of steps s in the **while** loop of CLOSURE, starting from an arbitrary H-colouring c. If $s = 0$, then the theorem is clearly true. Otherwise, consider w_1 in the first step of CLOSURE. Let A_1 denote A after the execution of $N_A(w_1) \leftarrow N_A(w_2)$. Successively re-colour all vertices of G coloured w_1 with w_2. This is clearly permissible. Since A_1 only differs from H by edges incident with w_1, and we will not use this colour again, the difference between A_1 and H is irrelevant. Now recolour a further $(s-1)$ times, each time "eliminating" one colour. At the termination of CLOSURE, the colours still in use obviously form a looped clique C in A and we observe that if v is such a colour then $N_A(v) = N_H(v)$. Thus C must also be a looped clique in H. Thus we have moved from c to a C-colouring of G, using only steps of $\mathcal{G}_H(G)$. We may move freely between any two C-colourings, and this half of the theorem is proved.

For the converse, suppose A is not complete looped. Since H is a subgraph of A, it suffices to prove the claim for $\mathcal{G}_A(G_0)$, provided that we can exhibit two H-colourings which are disconnected in the chain. The vertices of A are divided into classes by the equivalence relation defined above. Let B be the subgraph of H given by deleting all vertices found to be dominated in CLOSURE, and all but one vertex in each equivalence class among those remaining. There will be exactly one representative of each A-equivalence class in B. By the observation in the first part of the theorem, B is a subgraph of H. Clearly B is unique up to relabelling its vertices with A-equivalent vertices, and has at least two vertices. (Otherwise it follows that $A = H$ is an unlooped independent set.) Let B_1, B_2 be two copies of B, but having all loops deleted, with natural bijections $f_i : B \to B_i$ $(i = 1, 2)$. Now let G_0 be the disjoint union of B_1, B_2, with an additional edge $\{f_1(w), f_2(w)\}$ for each looped vertex w in B. The claimed degree bound on G_0 clearly holds. Let G_0 have the H-colouring c, where $c(f_i(w)) = w$ $(i = 1, 2)$. This is clearly proper.

We claim that in c, if a vertex $f_i(w) \in V$ can be re-coloured with w' in A, then w' is equivalent to w: $f_i(w)$ has a neighbour in G_0 coloured by a representative of each equivalence class to which w is adjacent in A. (Here w is adjacent to itself in A if and only if it has a loop.) Thus w' must also be adjacent to a vertex of A in each equivalence class adjacent to w. It follows that w' must be adjacent in A to *every* vertex in each of these classes. Thus

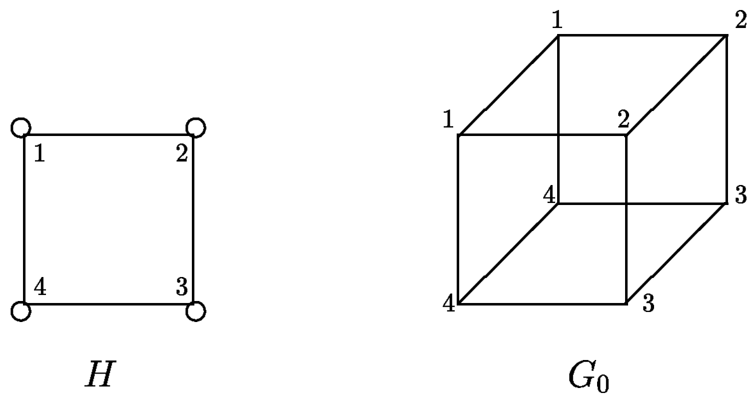

Figure 1: An isolated H-colouring of G_0

w' is adjacent to every vertex to which w is adjacent. It follows that w' is equivalent to w.

Now it follows that, if H has any loop w_0, $\mathcal{G}_A(G_0)$ cannot move from c to the proper H-colouring c_0 such that $c_0(v) = w_0$ for all $v \in V_0$. Thus it cannot be ergodic. $\square$

As discussed in section 1, the reason for non-ergodicity in the bipartite case can simply be that the bipartition (W_1, W_2) of H can be used to colour the bipartition (V_1, V_2) of G in two ways. To exclude this triviality, we called $\mathcal{G}_H(G)$ semi-ergodic provided it is ergodic on both of the colour sets $\Omega_1, \Omega_2 \subseteq \Omega$ defined there. We then have the following extension of Theorem 2.

Theorem 3 *Suppose H is connected, bipartite, and has no loops. Then $\mathcal{G}_H(G)$ is semi-ergodic for all G if and only if* CLOSURE(H) *is a complete unlooped bipartite graph. When it is not semi-ergodic for all G, there exists a graph G_0 with $\Delta(G_0) = \max_{w \in W} |N_H(w)|$ for which this remains true.*

Proof Consider, without loss, Ω_1. The constructions of A and B are exactly as before. Suppose CLOSURE(H) is complete unlooped and bipartite. We reduce any colouring to one in a set of H-colourings between which we can move freely, giving the first part of the theorem. For the converse, suppose CLOSURE(H) is not complete. We let $G_0 = B$ with the same H-colouring c as before. Then we may show that we can only move between colourings which re-colour vertices with equivalent colours. This set is disconnected from the H-colouring c_0 where ${c_0}^{-1}(V_i) = \{w_i\}$ $(i = 1, 2)$, where $\{w_1, w_2\}$ is any edge of H with $w_i \in W_i$ $(i = 1, 2)$. $\square$

The cycle on six vertices (i.e. bipartite 3-colouring) is an example of an unlooped bipartite H whose closure is not complete.

The graph G_0 in the above theorems is a graph of bounded size, but this is not the cause of non-ergodicity. It is easy to construct arbitrarily large connected graphs, with G_0 as a subgraph, which have the same behaviour, by making the maximum degree at most 1 larger.

We have proved nothing for the nonbipartite unlooped case, though we may observe that the theorem of [3] covers this case also. Observe that here CLOSURE(H) will also be a nonbipartite unlooped graph, so we do not have any obvious connectedness property for CLOSURE(H). Thus we might expect that the Glauber dynamics will not be ergodic for all G. Brightwell and Winkler [3] show this intuition to be true, by exhibiting examples with "stuck" colourings as above, using the "weak square" of H. This is certainly interesting from the combinatorial viewpoint but, as regards computation, there is a greater difficulty. Finding a single H-colouring is an NP-Complete problem. Thus, even if we were guaranteed that some chain with a richer set of transitions was ergodic, we might still be unable to find any state at which to start the chain.

Observe finally that ergodicity is not a monotone property of H. Figure 1 shows an H (with CLOSURE(H) $= H$), the graph G_0 and its canonical H-colouring. The Glauber dynamics is "stuck", even though there are more than a thousand other H-colourings of G. However, deleting *any* single edge of H causes it to become ergodic for all G, as does adding either of the two edges "missing" from H.

4 Sampling Algorithm

The recursive characterisation of ergodicity we have proved above seems too weak to guarantee the existence of good sampling algorithms in general. However, there is one case where this recursive stucture is directably employable, when H is a "looped tree". This family has some significance, since it includes both the discrete Widom-Rowlinson and Beach models.

Therefore, let H be a tree with loops on every vertex and let G be the (arbitrary) graph which we wish to H-colour randomly, according to the distribution p_λ defined below. Choose any vertex ρ in H and root the tree at ρ. Then, for vertex x of H, we define its *depth* $d(x)$ to be its edge-distance from ρ in H.

For $\lambda > 0$ we will define a weight function $w : W \to \mathbf{R}$ by $w(x) = \lambda^{d(x)}$. The weight $\mu_\lambda(c)$ of H-colouring c is then defined by

$$\mu_\lambda(c) = \prod_{v \in V} w(c(v)).$$

Let

$$Z_\lambda = Z_{\lambda,H,G} = \sum_{c \in \Omega} \mu_\lambda(c)$$

and define the probability distribution $p_\lambda = p_{\lambda,H,G}$ on Ω by $p_\lambda(c) = \mu_\lambda(c)/Z_\lambda$. This is the (finite-volume) Gibbs distribution for H-colouring G. We will show that for sufficiently small λ it is possible to sample efficiently from p_λ. We employ a simple Markov chain on the set Ω of H-colourings of G, and show that it is rapidly mixing. As discussed in section 3, the simplest, and possibly most important, such chain is the Glauber dynamics, where we change the colour of one vertex at a time. The precise description of the chain in the weighted case is as follows.

Glauber dynamics $\mathcal{G}_H(G)$

Let X denote the current state of the chain, so $X(v)$ is the colour of v. We describe a random move from X to a new state X'.

Step 1: Choose v uniformly at random from V.

Step 2: For all $x \in W$, the vertex set of H, let

$$p_{X,v}(x) = \mathbf{Pr}(c(v) = x \mid c(v') = X(v') \text{ for } v' \neq v).$$

In this expression the unconditional distribution is p_λ and we are simply conditioning on the colour of v, given the colour of v', for all $v' \neq v$.

Step 3: Choose x according to the distribution $p_{X,v}$.

$$X'(v') = \begin{cases} X(v') : & v' \neq v \\ x : & v' = v \end{cases}$$

The steady state distribution for the Glauber dynamics is p_λ, as follows from consideration of "detailed balance". See [7] for further information.

Let us continue by considering $p_{X,v}$ in more detail. Let N_v denote the set of neighbours of v in G, and D_x denote the set of children of x in H. (The *descendants* of x will mean the iterated children.) The distribution $p_{X,v}$ depends only on the colours of N_v. There are several cases to consider:

(a) $X(v') = \rho$ for all $v' \in N_v$.
The possible colours for v are ρ and the children D_ρ of ρ in H.

$$p_{X,v}(x) = \begin{cases} \frac{1}{1+\lambda|D_\rho|} : & x = \rho \\ \frac{\lambda}{1+\lambda|D_\rho|} : & x \in D_\rho \end{cases}$$

(b) $\exists \xi \neq \rho$ such that $X(v') = \xi$ for all $v' \in N_v$.
The possible colours for v are ξ', the parent of ξ in H, ξ and the children D_ξ of ξ.

$$p_{X,v}(x) = \begin{cases} \frac{1}{1+\lambda+\lambda^2|D_\xi|} : x = \xi' \\ \frac{\lambda}{1+\lambda+\lambda^2|D_\xi|} : x = \xi \\ \frac{\lambda^2}{1+\lambda+\lambda^2|D_\xi|} : x \in D_\xi \end{cases}$$

(c) $\exists y, z$ such that y is the parent of z in H and $X(v') \in \{y, z\}$ for all $v' \in N_v$. Furthermore, both colours y, z are used in N_v.

$$p_{X,v}(x) = \begin{cases} \frac{1}{1+\lambda} : & x = y \\ \frac{\lambda}{1+\lambda} : & x = z \end{cases}$$

(d) $\exists x$ such that N_v contains vertices of colour x, y, z where y, z are distinct neighbours of x in H.
The only possible colour for v is x.

We may now state and prove the main result of this section. We use $\mathrm{d_{TV}}$ to denote the variation distance (i.e. half the L_1 norm) between distributions. Also let $\Delta = \max\{\Delta_H, \Delta_G\}$, where Δ_H, Δ_G are the maximum degrees in H, G respectively.

Theorem 4 *Let $\lambda_0 = 1/6\Delta^2$. For any $\epsilon > 0$ and $\lambda \leq \lambda_0$, $O(n \log(n/\epsilon))$ steps of the Glauber dynamics will give a random H-colouring C of G, with distribution $\mathcal{L}(C)$ such that $\mathrm{d_{TV}}(\mathcal{L}, p_\lambda) \leq \epsilon$.*

Proof The theorem is trivial if $\Delta = 1$ and so we assume that $\Delta \geq 2$. We will prove rapid mixing by the method of *path coupling* introduced by Bubley and Dyer [4].

So let $(X_t, Y_t), t = 1, 2, \ldots$ be two (coupled) copies of $\mathcal{G}_H(G)$. In path coupling we try to reduce the expected distance between the copies at each step, in some suitable metric. An obvious candidate is the *Hamming distance* $h(X_t, Y_t) = |\{v : X_t(v) \neq Y_t(v)\}|$. However, Hamming distance does not appear to give the desired result here, as the reader might choose to verify. Therefore, we will replace it by a more useful metric. Choose $\sigma > 0$ and for an edge (x, y) of H, where x is the parent of y, let $\mu_\sigma(x, y) = \sigma^{d(x)}$. If x, y are arbitrary vertices of H, let $\mu_\sigma(x, y)$ be the shortest distance in H between x and y, using μ_σ as edge length. Finally, if X, Y are distinct H-colourings of G, we let $\mu_\sigma(X, Y) = \sum_{v \in V} \mu_\sigma(X(v), Y(v))$. We will use μ_M, where $M = 3\Delta$, to prove the theorem, but we also employ μ_1, the metric based on edge distance in H.

In path coupling, we need analyse only certain pairs of colourings. However, these must be chosen to "respect" the metric. We begin by proving the desired property for the colourings we will use in the proof. See [7] for further details.

Lemma 5 *Let X, Y be H-colourings of G. There exists a sequence $X = Z_0, Z_1, \ldots, Z_\ell = Y$ such that*

(i) Z_i, Z_{i+1} *differ only at a single vertex* $z_i \in V$ $(i = 0, 1, \ldots, \ell - 1)$,

(ii) $Z_i(z_i)$, $Z_{i+1}(z_i)$ *are adjacent in* H $(i = 0, 1, \ldots, \ell - 1)$,

(iii) $\mu_M(X, Y) = \sum_{i=0}^{\ell-1} \mu_M(Z_i, Z_{i+1})$.

Proof We prove this by induction on $\mu_1(X, Y)$. The result is trivial for $\mu_1(X, Y) = 1$. Let

$$d^* = \max\{d(x) : \exists v \in V \text{ s.t. } X(v) \neq Y(v), x \in \{X(v), Y(v)\}\}.$$

Clearly $d^* > 0$. Let $d^* = d(x^*)$, where $x^* = X(v^*) \neq Y(v^*)$. Note that $Y(v^*)$ cannot be a descendant of x^* in H. Let y be the parent of x^* in H. Define X' by

$$X'(v) = \begin{cases} X(v): & v \neq v^* \\ y: & v = v^* \end{cases}$$

Then $\mu_\sigma(X, Y) = \mu_\sigma(X, X') + \mu_\sigma(X', Y)$ for any σ, in particular $\sigma = 1, M$. Furthermore, X' is an H-colouring. This is because no neighbour v of v^* can be coloured with a child x of x^*. If it were then we must have $Y(v) = x$, by the depth property of x^*. But then $Y(v^*), Y(v)$ are not adjacent in H, a contradiction. □

So let us now assume that, for some $w \in V$, we have $X_t(w') = Y_t(w')$ for $w' \neq w$ and $X_t(w) = x$, $Y_t(w) = y$ where x is the parent of y in H. We will write $d = d(x)$ and, unless $x = \rho$, we denote the parent of x by ξ. Using path coupling, it is now enough to couple X_{t+1}, Y_{t+1} so that

$$\mathbf{E}(\mu_M(X_{t+1}, Y_{t+1})) \mid X_t, Y_t) \leq \left(1 - \frac{1}{2n}\right) \mu_M(X_t, Y_t). \tag{6}$$

The diameter of Ω is at most $2n(M^\alpha - 1)/(M-1)$, where $\alpha = \max_{w \in W} d(w)$. The theorem now follows easily from this and (6). (See, for example, [9].)

In our coupling the same vertex v is chosen for X_t, Y_t in Step 1. Unless $v \in N_w$, we will choose the same new colour for v in both X_t, Y_t. This has the correct marginal distributions, and will be our coupling in all cases below where there is no further comment.

Case 1 $v = w$. This occurs with probability n^{-1} and results in $X_{t+1} = Y_{t+1}$.

Case 2 $v \notin N_w \cup \{w\}$. This has probability $1 - \frac{|N_w|+1}{n}$ and results in $\mu_M(X_{t+1}, Y_{t+1}) = \mu_M(X_t, Y_t)$.

Case 3 $v \in N_w$. Then $X_t(v) = Y_t(v)$ is a neighbour of x and y in H and so $X_t(v) \in \{x, y\}$. We now describe a suitable coupling in each of several sub-cases for v, and determine the consequent (conditional) expected change in μ_M, i.e. the quantity

$$E_t = \mathbf{E}(\mu_M(X_{t+1}, Y_{t+1}) - \mu_M(X_t, Y_t) \mid v).$$

(i) In X_t, v has neighbours which are coloured $z, z' \in D_x \cup \{\xi\}$, $z \neq z'$. It follows that $X_t(v) = x$, and we must have $X_{t+1}(v) = Y_{t+1}(v) = x$. Hence, $E_t = 0$.

(ii) In Y_t, v has neighbours which are coloured $z, z' \in D_y \cup \{x\}$, $z \neq z'$. It follows that $X_t(v) = Y_t(v) = y$, and we must have $X_{t+1}(v) = Y_{t+1}(v) = y$. Hence, $E_t = 0$.

(iii) In X_t, v has a neighbour coloured $z \in D_x \setminus \{y\}$, but no neighbour coloured from $(D_x \cup \{\xi\}) \setminus \{z\}$. It follows that $X_t(v) = x$. Then $X_{t+1}(v) \in \{x, z\}$, but $Y_{t+1}(v) = x$. Then the only possible coupling gives $E_t = M^d\lambda/(1 + \lambda) < \lambda M^d$.

(iv) In Y_t, v has a neighbour coloured $z \in D_y$, but no neighbour coloured from $(D_y \cup \{x\}) \setminus \{z\}$. It follows that $X_t(v) = y$. Then $X_{t+1}(v) = y$, but $Y_{t+1}(v) \in \{y, z\}$. The only coupling gives $E_t = M^{d+1}\lambda/(1 + \lambda) < \lambda M^{d+1}$.

(v) In both X_t and Y_t, v has neighbours coloured both x and y, but no neighbour coloured from $D_y \cup \{\xi\} \cup D_x \setminus \{y\}$. Then both $X_{t+1}(v), Y_{t+1}(v) \in \{x, y\}$ and hence we can couple so that $E_t = 0$.

(vi) In X_t, v has only neighbours coloured $x \neq \rho$. Then $X_{t+1}(v) \in \{x, \xi\} \cup D_x$, but $Y_{t+1}(v) \in \{x, y\}$ only. With probability $\beta_x = 1/(1 + \lambda + \lambda^2|D_x|)$, we set $X_{t+1}(v) = \xi$, $Y_{t+1}(v) = x$ and with probability $\beta_y = \lambda/(1+\lambda+\lambda^2|D_x|)$, we set $X_{t+1}(v) = x$, $Y_{t+1}(v) = y$. With probability $1/(1+\lambda) - \beta_x$ we set $X_{t+1}(v) \in D_x$, uniformly at random, and $Y_{t+1}(v) = x$. With probability $\lambda/(1 + \lambda) - \beta_y$, we set $X_{t+1}(v) \in D_x$, uniformly at random, and $Y_{t+1}(v) = y$. This gives

$$E_t < M^{d-1} + (\lambda + 2\lambda^2|D_x|)M^d \leq M^{d-1} + \lambda(1 + 2\lambda\Delta)M^d.$$

(vii) In X_t, v has only neighbours coloured ρ. It follows that $X_t(v) = x = \rho$. Then $X_{t+1}(v) \in \{\rho\} \cup D_\rho$, but $Y_{t+1}(v) \in \{\rho, y\}$. With probability $\beta_\rho = 1/(1 + \lambda|D_\rho|)$, we set $X_{t+1}(v) = Y_{t+1}(v) = \rho$. With probability $1/(1 + \lambda) - \beta_\rho$ we set $X_{t+1} \in D_\rho$, uniformly at random, and $Y_{t+1}(v) = \rho$. With probability $\lambda/(1 + \lambda)$, we set $X_{t+1} \in D_\rho$, uniformly at random, and $Y_{t+1}(v) = y$. This gives

$$E_t < 2\lambda|D_\rho| \leq 2\lambda\Delta = 2\lambda\Delta M^d.$$

(viii) In Y_t, v has only neighbours coloured y. Then $X_{t+1}(v) \in \{x,y\}$, but $Y_{t+1}(v) \in \{x,y\} \cup D_y$. Then we set $X_{t+1}(v) = Y_{t+1}(v) = x$ with probability $\gamma_x = 1/(1+\lambda+\lambda^2|D_y|)$ and $X_{t+1}(v) = Y_{t+1}(v) = y$ with probability $\gamma_y = \lambda/(1+\lambda+\lambda^2|D_y|)$. With probability $1/(1+\lambda) - \gamma_x$ we set $X_{t+1}(v) = x$ and $Y_{t+1}(v) \in D_y$ uniformly at random. With probability $\lambda/(1+\lambda) - \gamma_y$, we set $X_{t+1}(v) = y$ and $Y_{t+1}(v) \in D_y$ uniformly at random. Thus

$$E_t < 2M^{d+1}\lambda^2|D_y| \leq 2M^{d+1}\lambda^2\Delta.$$

(ix) In X_t, v has neighbours coloured from $\{x,\xi\}$. It follows that $X_t(v) = x$. Then $Y_{t+1}(v) = x$, but $X_{t+1}(v) \in \{x,\xi\}$. The only coupling gives

$$E_t = M^{d-1}/(1+\lambda) < M^{d-1}.$$

Suppose now we put $M = 3\Delta$ and $\lambda \leq 1/6\Delta^2$. Then the reader may check that the largest contribution comes from sub-case (iv) above, and is at most $M^d/2\Delta$.

Putting this all together we get that, conditional only on X_t, Y_t,

$$\mathbf{E}(\mu_M(X_{t+1},Y_{t+1}) - \mu_M(X_t,Y_t)) < -\frac{M^d}{n} + \frac{\Delta}{n}\frac{M^d}{2\Delta} = -\frac{M^d}{2n} = -\frac{\mu_M(X_t,Y_t)}{2n}.$$

This verifies (6) and completes the proof of Theorem 4. □

Finally, we might naturally ask how far this result can be extended, i.e. for which H are there *any* sets of positive vertex weights such that random generation is possible ? It is easy to extend the proof above to the case where some of the leaves of H may be unlooped. As well as independent sets, this includes some H studied in [6], such as the "wrench". Beyond this, a reasonable conjecture is that a positive result holds for any H such that, in the notation of section 3, CLOSURE(H) is a complete looped graph. However, we will leave this simply as a question.

References

[1] B. Bollobás, *Random Graphs*, Academic Press, London, 1985.

[2] G. Brightwell, O. Häggström and P. Winkler Nonmonotonic behavior in hard-core and Widom-Rowlinson models, *Journal of Statistical Physics* **94** (1999), 415–435.

[3] G. Brightwell and P. Winkler, Gibbs measures and dismantlable graphs, *Journal of Combinatorial Theory* (Series B) **78** (2000), 141–169.

[4] R. Bubley and M. E. Dyer, Path coupling: a technique for proving rapid mixing in Markov chains, in *38th Annual Symposium on Foundations of Computer Science*, IEEE Press, 1997, pp. 223–231.

[5] M. E. Dyer, A. M. Frieze and M. R. Jerrum, On counting independent sets in sparse graphs, in *40th Annual Symposium on Foundations of Computer Science*, IEEE Press, 1999, pp 210–217.

[6] M. E. Dyer, L. A. Goldberg, C. S. Greenhill, and M. R. Jerrum, On the relative complexity of approximate counting problems, in *Proceedings of RANDOM 2000*, September 2000, to appear.

[7] M. E. Dyer and C.S. Greenhill, Random walks on combinatorial objects, in *Surveys in Combinatorics 1999* (J. D. Lamb and D. A. Preece, Eds.), Cambridge University Press, 1999, pp. 101–136.

[8] M. E. Dyer and C. S. Greenhill, The complexity of counting graph homomorphisms, in *Eleventh Annual ACM/SIAM Symposium on Discrete Algorithms*, ACM/SIAM Press, 2000, pp. 246–255.

[9] M. E. Dyer and C. S. Greenhill, On Markov chains for independent sets, *Journal of Algorithms* **35**, 17–49, 2000.

[10] A. M. Frieze and T. Łuczak, On the independence and chromatic numbers of random regular graphs, *Journal of Combinatorial Theory* Series B **54** (1992), 123–132.

[11] A. Galluccio, P. Hell and J. Nešetřil, On the complexity of H-colouring of bounded degree graphs, Technical Report R42, Istituto di Analisi dei Sistemi ed Informatica del CNR, Rome, November 1998.

[12] P. Hell and J. Nešetřil, On the complexity of H-colouring, *Journal of Combinatorial Theory* (Series B) **48** (1990), 92–110.

[13] M.R. Jerrum, A very simple algorithm for estimating the number of k-colorings of a low-degree graph, *Random Structures and Algorithms* **7** (1995), pp. 157–165.

[14] M.R. Jerrum, L. G. Valiant and V. V. Vazirani, Random generation of combinatorial structures from a uniform distribution. *Theoretical Computer Science* **43** (1986), 169–188.

[15] M. Luby and E. Vigoda, Approximately counting up to four, in *Twenty-Ninth Annual Symposium on Theory of Computing*, ACM Press, 1997, pp. 682–687.

[16] A. Sinclair and M.R. Jerrum, Approximate counting, uniform generation, and rapidly mixing Markov chains, *Information and Computation* **82** (1989), 93–133.

[17] E. Vigoda, Improved bounds for sampling colorings, *40th Annual Symposium on Foundations of Computer Science*, IEEE Press, 1999, pp 51–59.

This research was sponsored in part by National Science Foundation (NSF) grant no. CCR-0122581.