Path Coupling Using Stopping Times and Counting Independent Sets and Colourings in Hypergraphs

Magnus Bordewich*, Martin Dyer* and Marek Karpinski†

August 17, 2006

Abstract

We analyse the mixing time of Markov chains using path coupling with stopping times. We apply this approach to two hypergraph problems. We show that the Glauber dynamics for independent sets in a hypergraph mixes rapidly as long as the maximum degree Δ of a vertex and the minimum size m of an edge satisfy $m \geq 2\Delta+1$. We also show that the Glauber dynamics for proper q-colourings of a hypergraph mixes rapidly if $m \geq 4$ and $q > \Delta$, and if $m = 3$ and $q \geq 1.65\Delta$. We give related results on the hardness of exact and approximate counting for both problems.

1 Introduction

We develop a new approach to using stopping times in conjunction with path coupling to bound the convergence of time of Markov chains. Our main interest is in applying these results to randomised approximate counting. For an introduction, see [21]. To illustrate our methods, we consider approximation of the numbers of independent sets and q-colourings in hypergraphs with upper-bounded degree, and lower-bounded edge size. These problems in hypergraphs are of interest in their own right but, while approximate optimisation has received attention [7, 6, 18, 22], there has been surprisingly little work on approximate counting.

Our results are achieved by considering, in the path coupling setting, the stopping time at which the distance between two coupled chains first changes. The first application of stopping times to path coupling was by Dyer, Goldberg, Greenhill, Jerrum and Mitzenmacher [10]. Their analysis was later improved by Hayes and Vigoda [17], using a method closely related to that developed in this paper. Mitzenmacher and Niklova [24] had earlier stated a similar theorem, but could not provide a conclusive proof. Their approach had many simliarities to our Theorem 2.1, but conditioning problems necessitate a proof along somewhat different lines.

Our main technical result, Theorem 2.1, shows that if the expected distance between the two chains has decreased at this stopping time, then the chain mixes rapidly. This also follows from [17, Corollary 4]. However we give a simpler proof than that of [17], and our Theorem 2.1 will usually give a moderate improvement in the bound on mixing time in comparison with [17, Corollary 4]. See Remark 2.3 below.

*School of Computing, University of Leeds, Leeds LS2 9JT, UK. Email: {dyer,magnusb}@comp.leeds.ac.uk.
†Dept. of Computer Science, University of Bonn, 53117 Bonn, Germany. Email: marek@cs.uni-bonn.de.

The problem of approximately counting independent sets in graphs has been widely studied, see for example [9, 12, 23, 25, 29], but the only previous work on the approximate counting of independent sets in *hypergraphs* seems to be that of Dyer and Greenhill [12]. They showed rapid mixing to the uniform distribution of a simple Markov chain on independent sets in a hypergraph with maximum degree 3 and maximum edge size 3. However, this was the only interesting case resolved. Their results imply rapid mixing only for $m \leq \Delta/(\Delta - 2)$, which gives $m \leq 3$ when $\Delta = 3$ and $m \leq 2$ when $\Delta \geq 4$. In Theorem 3.1 we prove rapid mixing of the *Glauber dynamics* for any hypergraph such that $m \geq 2\Delta + 1$, where m is the smallest edge size and Δ is the maximum degree. This is a marked improvement for large m. More generally, we consider the *hardcore distribution* on independent sets with *fugacity* λ. (See, for example, [12, 23, 29].) In [12], it is proved that rapid mixing occurs if $\lambda \leq m/((m-1)\Delta - m)$. Here we improve this considerably for larger values of m, to $\lambda \leq (m-1)/2\Delta$. We also give proofs that computing the number of independent sets in hypergraphs is #P-complete except in trivial cases, and that there can be no approximation for the number of independent sets in a hypergraphs if the minimum edge size is at most logarithmic in Δ. It may be noted that our upper and lower bounds are exponentially different. We have no strong belief that either is close to the threshold at which approximate counting is possible, if such a threshold exists.

Counting q-colourings of hypergraphs was considered by Bubley [2], who showed that the Glauber dynamics was rapidly mixing if $q \geq 2\Delta$, generalising a result of Jerrum [20] and Salas and Sokal [27] for graphs. Much work has been done on improving this result for graph colourings, see [8] and its references, but little attention appears to have been given to the hypergraph case. Here we prove rapid mixing of Glauber dynamics for proper colourings of hypergraphs if $m \geq 4$, $q > \Delta$, and if $m = 3$, $q \geq 1.65\Delta$. For a precise statement of our result see Theorem 5.2. Again we give proofs that computing the number of colourings in hypergraphs is #P-complete except in trivial cases, and that there can be no approximation for the number of colourings of hypergraphs if $q \leq e^{-m/(m-1)}\Delta^{1/(m-1)}$. Again, there is a considerable discrepancy between the upper and lower bounds for large m.

The paper is organised as follows. Section 1.1 give a brief review of the path coupling technique and Section 1.2 gives an intuitive motivation for the stopping time approach of this paper. Section 2 contains the full description and proof of Theorem 2.1 for path coupling with stopping times. We apply this to hypergraph independent sets in Section 3. Section 4 contains the hardness proofs for the independent sets problem. Section 5 contains analysis of the Glauber dynamics for hypergraph colouring. Finally, Section 6 contains the hardness results for counting colourings in hypergraphs.

1.1 A review of path coupling

Let Ω be a finite set and let $\mathcal{M}$ be a Markov chain with state space Ω, transition matrix P and unique stationary distribution π. In order for a Markov chain to be useful for almost uniform sampling or approximate counting, it must converge quickly towards its stationary distribution π. We make this notion more precise below. If the initial state of the Markov chain is x then the distribution of the chain at time t is given by $P_x^t(y) = P^t(x, y)$. The *total variation distance* of the Markov chain from π at time t with initial state x, is defined by

$$\mathrm{d_{TV}}(P_x^t, \pi) = \frac{1}{2} \sum_{y \in \Omega} |P^t(x, y) - \pi(y)|.$$

Let $\tau_x(\varepsilon)$ denote the least value such that $\mathrm{d_{TV}}(P_x^t, \pi) \leq \varepsilon$ for all $t \geq \tau_x(\varepsilon)$. The *mixing time* of $\mathcal{M}$, denoted by $\tau(\varepsilon)$, is defined by $\tau(\varepsilon) = \max\{\tau_x(\varepsilon) : x \in \Omega\}$. A Markov chain is said to be *rapidly* *mixing* if the mixing time is bounded above by some polynomial in n and $\log(\varepsilon^{-1})$, where n is a measure of the size of the elements in Ω.

One method for proving that a Markov chain is rapidly mixing is *coupling*. A coupling for $\mathcal{M}$ is a stochastic process (X_t, Y_t) on $\Omega \times \Omega$ such that each of (X_t) and (Y_t), considered marginally, is a faithful copy of $\mathcal{M}$. The *Coupling Lemma* [**?**] states that

$$\mathrm{d_{TV}}(P_x^t, \pi) \leq \Pr[X_t \neq Y_t],$$

i.e. the total variation distance for $\mathcal{M}$ at time t is bounded by the probability that the process has not coupled. A very useful extension of coupling is the *path coupling* method [3]. In this, one need only define a coupling on a subset S of $\Omega \times \Omega$. Let $\mathcal{G} = (\Omega, S)$ be the corresponding digraph. Relative to S, a *path coupling* $\mathcal{P}$ for $\mathcal{M}$ is specified by giving only the distributions

$$\Pr(X_{t+1} = x', Y_{t+1} = y' \mid X_t = x, Y_t = y) \qquad (x, y) \in S.$$

The coupling $\mathcal{P}$ is then formed by composing these distributions along paths in $\mathcal{G}$. See [**?**], for example, for details. Thus, for each pair $(x, y) \in \Omega^2$, we specify a path $x = z_0, z_1, \ldots, z_r = y$ and compose the couplings for (z_{i-1}, z_i) $(i = 1, \ldots, r)$ to give the coupling for x, y.

The path coupling method requires a metric d on Ω^2, which arises from a *path metric* for $\mathcal{G}$, i.e. for all $(x, y) \in \Omega^2$, there is a path $x = z_0, z_1, \ldots, z_r = y$ in $\mathcal{G}$ such that $\mathrm{d}(x, y) = \sum_{i=1}^{r} \mathrm{d}(z_{i-1}, z_i)$. The triangle inequality implies that such a path must be a *shortest path* in $\mathcal{G}$ under the edge weighting given by d. Therefore, we will call such a path a *geodesic*. The paths used to construct the path coupling $\mathcal{P}$ are then chosen to be geodesics. Note that $\mathcal{P}$ may not be unique if geodesics are not unique in $\mathcal{G}$.

A judicious choice of S can greatly simplify the proof of rapid mixing of a Markov chain by coupling. The pairs in S (i.e. edges of $\mathcal{G}$) need not be transitions of $\mathcal{M}$, and vice versa.

For a path coupling $\mathcal{P}$ with respect to $S \subseteq \Omega^2$, we define

$$\beta(\mathcal{M}, \mathcal{P}) \;=\; \max_{(x,y)\in S} \mathbf{E}_{\mathcal{P}}[\mathrm{d}(X_{t+1}, Y_{t+1})/\mathrm{d}(x, y) \mid X_t = x, Y_t = y].$$

The main path coupling method is then given by the following theorem [11].

Theorem 1.1. *Let* d *be an integer valued metric defined on* $\Omega \times \Omega$ *which takes values in* $\{0, \ldots, D\}$. *Let* S *be a subset of* $\Omega \times \Omega$ *such that for all* $(X_t, Y_t) \in \Omega \times \Omega$ *there exists a path* $X_t = Z_0, Z_1, \ldots, Z_r = Y_t$ *between* X_t *and* Y_t *such that* $(Z_l, Z_{l+1}) \in S$ *for* $0 \leq l < r$ *and* $\sum_{l=0}^{r-1} \mathrm{d}(Z_l, Z_{l+1}) = \mathrm{d}(X_t, Y_t)$.

Define a coupling $\mathcal{P} : (X, Y) \mapsto (X', Y')$ *of the Markov chain* $\mathcal{M}$ *on all pairs* $(X, Y) \in S$. *Suppose there exists* $\beta \leq 1$ *such that* $\mathbf{E}[\mathrm{d}(X', Y')] \leq \beta \mathrm{d}(X, Y)$ *for all* $(X, Y) \in S$.

(i) If $\beta < 1$ *then the mixing time* $\tau(\varepsilon)$ *of* $\mathcal{M}$ *satisfies* $\tau(\varepsilon) \leq \dfrac{\log(D\varepsilon^{-1})}{1 - \beta}$.

(ii) If $\beta = 1$ *and* $\Pr[(X_{t+1}, Y_{t+1}) \neq (X_t, Y_t)] \geq \gamma$ *for some* $\gamma > 0$ *and all* t *then*

$$\tau(\varepsilon) \leq \left\lceil \frac{eD^2}{\gamma} \right\rceil \lceil \log(\varepsilon^{-1}) \rceil.$$

1.2 Intuition

Let $\mathcal{H} = (\mathcal{V}, \mathcal{E})$ be a hypergraph of maximum degree Δ and minimum edge size m. A subset $S \subseteq \mathcal{V}$ of the vertices is *independent* if no edge is a subset of S. Let $\Omega(\mathcal{H})$ be the set of all independent sets of $\mathcal{H}$. Let λ be the *fugacity*, which weights independent sets. (See [12].) The most important case is $\lambda = 1$, which weights all independent sets equally and gives rise to the uniform distribution on all independent sets. We define the Markov chain $\mathcal{M}(\mathcal{H})$ with state space $\Omega(\mathcal{H})$ by the following transition process (*Glauber dynamics*). If the state of $\mathcal{M}$ at time t is X_t, the state at $t+1$ is determined by the following procedure.

(i) Select a vertex $v \in \mathcal{V}$ uniformly at random,

(ii) (a) if $v \in X_t$ let $X_{t+1} = X_t \backslash \{v\}$ with probability $1/(1+\lambda)$,

(b) if $v \notin X_t$ and $X_t \cup \{v\}$ is independent, let $X_{t+1} = X_t \cup \{v\}$ with probability $\lambda/(1+\lambda)$,

(c) otherwise let $X_{t+1} = X_t$.

This chain is easily shown to be ergodic with stationary probability proportional to $\lambda^{|I|}$ for each independent set $I \subseteq \mathcal{V}$. In particular, $\lambda = 1$ gives the uniform distribution. The natural coupling for this chain is the "identity" coupling, the same transition is attempted in both copies of the chain. If we try to apply standard path coupling to this chain, we immediately run into difficulties. Consider two chains X_t and Y_t such that $Y_t = X_t \cup \{w\}$, where $w \notin X_t$ (the *change vertex*) is of degree Δ. An edge $e \in \mathcal{E}$ is *critical* in Y_t if it has only one vertex $z \in \mathcal{V}$ which is not in Y_t, and we call z *critical for* e. If each of the edges through w is critical for Y_t, then there are Δ choices of v in the transition which can be added in X_t but not in Y_t. Thus, if $\lambda = 1$, the change in the expected Hamming distance between X_t and Y_t after one step could be as high as $\frac{\Delta}{2n} - \frac{1}{n}$. Thus we obtain rapid mixing only in the case $\Delta = 2$. This case has some intrinsic interest, since the complement of an independent set corresponds, under hypergraph duality, to an *edge cover* [15] in a graph. Thus we may uniformly generate edge covers, but the scope for unmodified path coupling is obviously severely limited.

The insight on which this paper is based is as follows. Although in one step it could be more likely that a *bad vertex* (increasing Hamming distance) is chosen than a *good vertex* (decreasing Hamming distance), it is even more likely that one of the other vertices in an edge containing w is chosen and removed from the independent set. Once the edge has two unoccupied vertices other than w, then any vertex in that edge can be added in both chains. This observation enables us to show that, if T is defined to be the stopping time at which the distance between X_t and Y_t first changes, the expected distance between X_T and Y_T will be less than 1. Theorem 2.1 below shows that under these circumstances path coupling can easily be adapted to prove rapid mixing.

Having established this general result, we use it to prove that $\mathcal{M}(\mathcal{H})$ is rapidly mixing for hypergraphs with $m \geq 2\lambda\Delta + 1$.

2 Path coupling using a stopping time

First we prove the main result discussed above.

Theorem 2.1. *Let $\mathcal{M}$ be a Markov chain on state space Ω. Let* d *be an integer valued metric on $\Omega \times \Omega$, and let (X_t, Y_t) be a path coupling for $\mathcal{M}$, where S is the set of pairs of states (X, Y) such* *that* $\mathrm{d}(X,Y)=1$*. For any pair of initial states* $(X_0,Y_0)\in S$ *let* T *be the stopping time given by the minimum* t *such that* $\mathrm{d}(X_t,Y_t)\neq 1$*. Suppose, for some* $p>0$*, that*

(i) $\Pr(T=t\,|\,T\geq t,(X_{t-1},Y_{t-1}))\geq p$*, independently for each* t*,*

(ii) $\mathbf{E}[\mathrm{d}(X_T,Y_T)]\leq\alpha<1$*.*

Then $\mathcal{M}$ *mixes rapidly. In particular the mixing time* $\tau(\varepsilon)$ *of* $\mathcal{M}$ *satisfies*

$$\tau(\varepsilon)\ \leq\ \frac{1}{p}\,\frac{3}{1-\alpha}\,\ln(eD_2)\ln\Big(\frac{2D_1}{\varepsilon(1-\alpha)}\Big),$$

where $D_1=\max\{\mathrm{d}(X,Y):X,Y\in\Omega\}$ *and* $D_2=\max\{\mathrm{d}(X_T,Y_T):X_0,Y_0\in\Omega,\ \mathrm{d}(X_0,Y_0)=1\}$*.*

Proof. Consider the following game. In each round (time step) a gambler either finishes the game, by winning £1 or losing some amount £$(l-1)$, or continues the game for another round. If he loses £$(l-1)$ in a game, he starts l separate (but possibly dependent) games simultaneously in an effort to win back his money. If he has several games going and loses £$(l-1)$ in one game at a certain time, he starts l more games, while continuing with the others that did not conclude. We know that the probability he finishes a game in a given step is at least p, and the expected winnings in each game is at least $1-\alpha$. The question is: does his return have positive expectation at any fixed time? We will show that it does. But first a justification for our interest in this game.

Each game represents a single step on the path between two states of the coupled Markov chain. We start with X_0 and Y_0 differing at a single vertex. The first game is won if the first time the distance between the coupled chains changes is in convergence. The game is lost if the distance increases to l. At that point we consider the distance l path X_t to Y_t, and the l games played represent the l steps in the path. Although these games are clearly dependent, they each satisfy the conditions given. The gambler's return at time t is one minus the length of the path at time t, so a positive expected return corresponds to an expected path length less than one. We will show that the expected path length is sufficiently small to ensure coupling.

First note that the gambler's return at time t is one minus the number of games active at time t. For the initial game we define the *level* to be zero, for any other possible game we define the level to be one greater than the level of the game whose loss precipitated it. We define the random variables $M_k, l_{j,k}$ and $I_{j,k}(t)$ as follows. M_k is the number of games at level k that are played, $l_{j,k}$, for $j=1\dots M_k$, is the number of games in level $k+1$ which are started as a result of the outcome of game j in level k, and $I_{j,k}(t)$ is an indicator function which takes the value 1 if game j in level k is active at time t, and 0 otherwise. Note that since each game takes an indeterminate number of steps, the levels and times steps are not in synch, indeed the levels of different games active at a given time may not agree. Let $N(t)$ be the number of games active at time t. Then, by linearity of expectation,

$$\mathbf{E}[N(t)]=\sum_{k=0}^{\infty}\mathbf{E}\left[\sum_{j=1}^{M_k}I_{j,k}(t)\right], \tag{1}$$

where the equality holds if the infinite sum converges, which will follow from the bounds below. We will bound this sum in two parts, splitting it at a point $k=K$ to be determined. For $k\leq K$ we observe that $M_k\leq D_2^k$. We label all games that might be played in the first K levels in such a way that there are exactly D_2^k games in level k. For $k>0$ and $j=(j_1-1)D_2+j_2$ with $j_1\in\{1,\dots,D_2^{k-1}\}$ and $j_2\in\{1,\dots,D_2\}$, the j^{th} game on level k is the j_2^{th} game spawned by the j_1^{th} game on level $k-1$.

It will be played if the j_1^{th} game on level $k-1$ is played and loses, and the random variable l resulting from the loss is at least j_2. Otherwise it will not be played. Now let $v_{j,k}$ denote the j^{th} game on level k. It has a unique path of ancestors in levels $0,\dots,k-1$. Let $i_{j,k}$ be the indicator variable for the event that $v_{j,k}$ or some ancestor of $v_{j,k}$ is active at time t. Since the number of active games in levels up to K at time t is at most the number of games in level K with an active ancestor, by linearity of expectation we obtain $\sum_{k=0}^{K}\mathbf{E}\left[\sum_{j=1}^{M_k} I_{j,k}(t)\right] \le \sum_{j=1}^{D_2^K}\mathbf{E}[i_{j,K}(t)]$. The probability that a game completes in any given step is at least p. If we select a time $t \ge (K+1)/p$ then the probability that at most K games are complete is clearly maximised by taking this probability to be exactly p in all games. Thus, for $t \ge (K+1)/p$,

$$\sum_{k=0}^{K}\mathbf{E}\left[\sum_{j=1}^{M_k} I_{j,k}(t)\right] \le \sum_{j=1}^{D_2^K}\mathbf{E}[i_{j,K}(t)] \le D_2^K \Pr(\mathrm{Bin}(t,p)\le K). \tag{2}$$

On the other hand, for $k > K$ we observe that

$$\mathbf{E}\Bigg[\sum_{j=1}^{M_k} I_{j,k}(t)\Bigg] \le \ \mathbf{E}[M_k] \ = \ \mathbf{E}_{M_{k-1}}\big[\mathbf{E}[M_k|M_{k-1}]\big] \ = \ \mathbf{E}_{M_{k-1}}\big[\mathbf{E}[\sum_{j=1}^{M_{k-1}} l_{j,k-1}|M_{k-1}]\big]$$

where $\mathbf{E}_{M_{k-1}}$ indicates an expectation taken only over the random variable M_{k-1}. Since for any starting conditions $\mathbf{E}[l_{j,k-1}] \le \alpha$, we may apply this bound even when conditioning on M_{k-1}. So

$$\mathbf{E}\left[\sum_{j=1}^{M_k} I_{j,k}(t)\right] \ \le \ \mathbf{E}[\alpha M_{k-1}] \ \le \ \alpha^k, \tag{3}$$

using linearity of expectation, induction and $\mathbf{E}[M_1] \le \alpha$. Putting (2) and (3) together we get, for $t \ge (K+1)/p$,

$$\begin{aligned}\mathbf{E}[N(t)] &\le \ D_2^K \Pr(\mathrm{Bin}(t,p)\le K) \ + \sum_{k=K+1}^{\infty}\alpha^k \\ &= \ D_2^K \Pr(\mathrm{Bin}(t,p)\le K) \ + \ \frac{\alpha^{K+1}}{1-\alpha}.\end{aligned} \tag{4}$$

We now set $K = \lfloor(\ln\alpha)^{-1}\ln(\frac{\varepsilon(1-\alpha)}{2D_1})\rfloor$, hence the final term is at most $\varepsilon/2D_1$. Taking a time $\tau = c/p$, by Chernoff's bound (see, for example, [19, Theorem 2.1]),

$$\begin{aligned}\mathbf{E}[N(\tau)] &\le \ {D_2}^K\sum_{k=0}^{K}\binom{\tau}{k}p^k(1-p)^{\tau-k} + \frac{\varepsilon}{2D_1} \\ &\le \ e^{K\ln D_2 - \frac{(c-K)^2}{2c}} + \frac{\varepsilon}{2D_1} \\ &\le \ e^{K\ln D_2 + K - c/2} + \frac{\varepsilon}{2D_1}.\end{aligned}$$

Choosing $c = 2K\ln(eD_2) + 2\ln\frac{2D_1}{\varepsilon}$, we obtain $\mathbf{E}[N(\tau)] < \frac{\varepsilon}{D_1}$, where $\tau = \left\lceil\frac{3\ln(eD_2)}{p(1-\alpha)}\ln\left(\frac{2D_1}{\varepsilon(1-\alpha)}\right)\right\rceil$. We conclude that the gambler's expected return at time τ is positive. More importantly, for any initial states $X_0, Y_0 \in \Omega$, the expected distance at time τ is at most ε by linearity of expectations, and so the probability that the chain has not coupled is at most ε. The mixing time claimed now follows by standard arguments. See, for example, [21]. □

Remark 2.2. The assumption that the stopping time occurs when the distance changes is not essential. We clearly cannot dispense with assumption (ii), or we cannot bound mixing time. Assumption (i) may appear a restriction, but appears to be naturally satisfied in most applications. It seems more natural than the assumption of bounded stopping time, used in [17]. Assumption (i) can easily be replaced by something weaker, for example by allowing p to vary with time rather than remain constant. Provided $p \neq 0$ sufficiently often, a similar proof will be valid.

Remark 2.3. Let $\gamma = 1/(1-\alpha)$. It seems likely that D_2 will be small in comparison to γ in most applications, so we might suppose $D_2 < \gamma < D_1$. The mixing time bound from Theorem 2.1 can then be written $O(p^{-1}\gamma \log D_2 \log(D_1/\varepsilon))$. We may compare this with the bound which can be derived using [17, Corollary 4]. This can be written in similar form as $O(p^{-1}\gamma \log \gamma \log(D_1/\varepsilon))$. In such cases we obtain a reduction in the estimate of mixing time by a factor $\log \gamma / \log D_2$. In the applications in Section 3, for example, we have $D_2 = 2$ and $\gamma = \Omega(\Delta)$, so the improvement is $\Omega(\log \Delta)$.

Remark 2.4. The reason for our improvement on the result of [17] is that the use of an upper bound on the stopping time, as is done in [17], will usually underestimate the number of stopping times which occur in a long interval, and hence overestimate the mixing time.

3 Hypergraph independent sets

We now use the approach of path coupling via stopping times to prove that the chain discussed in Section 1.2 is rapidly mixing. The metric used in path coupling analyses throughout the paper will be Hamming distance between the coupled chains. We prove the following theorem.

Theorem 3.1. *Let λ, Δ be fixed, and let $\mathcal{H}$ be a hypergraph such that $m \geq 2\lambda\Delta + 1$. Then the Markov chain $\mathcal{M}(\mathcal{H})$ has mixing time $O(n \log n)$.*

Although we have focused here on the relationship between mixing time and n, the proof gives an explicit bound on the mixing time in terms of λ, Δ and n. Before commencing the proof itself, we analyse the stopping time T for this problem.

3.1 Edge Process

Let X_t and Y_t be copies of $\mathcal{M}$ which we wish to couple, with $Y_0 = X_0 \cup \{w\}$. Let e be any edge containing w. Without loss of generality we assume that this edge has the minimum size ($|e| = m$), it will transpire that this is the worst case. We consider only the times at which some vertex in e is chosen. The progress of the coupling on e can then be modelled by the following “game”. We will call the unoccupied vertices in e (excluding w) *units*. At a typical step of the game we have k units, and we either win the game, win a unit, keep the same state or lose a unit. These events happen with the following probabilities: we win the game with probability $1/m$, win a unit with probability $(m-k-1)/(1+\lambda)m$, lose a unit with probability $\lambda k/(1+\lambda)m$ and stay in the same state otherwise. If ever $k = 0$, we are bankrupt and we lose the game. Winning the game models the “good event” that the vertex v is chosen and the two chains couple. Losing the game models the “bad event” that the coupling increases the distance to 2. We wish to know the probability that the game ends in bankruptcy. We are most interested in the case where $k = 1$ initially, which models e being critical. Note that the process on hypergraph independent sets is dominated by our model, since in the hypergraph process we can always delete (win or win a unit in the game), but we may not be able to insert (lose or lose a unit in the game) because the chosen vertex is critical in some other edge.

Let p_k denote the probability that a game is lost, given that we start with k units. By conditioning on the outcome of the first step being a win, win a unit, lose a unit or staying the same, for p_1 we have

$$p_1 = \frac{1}{m}(0) + \frac{(m-2)}{(1+\lambda)m}p_2 + \frac{\lambda}{(1+\lambda)m}(1) + \left(1 - \frac{1}{m} - \frac{(m-2)}{(1+\lambda)m} - \frac{\lambda}{(1+\lambda)m}\right)p_1 \tag{5}$$

and for $p_k, k > 1$ we have

$$p_k = \frac{1}{m}(0) + \frac{(m-k-1)}{(1+\lambda)m}p_{k+1} + \frac{\lambda k}{(1+\lambda)m}p_{k-1} + \left(1 - \frac{1}{m} - \frac{(m-k-1)}{(1+\lambda)m} - \frac{\lambda k}{(1+\lambda)m}\right)p_k \tag{6}$$

Equations (5) and (6) simplify to the following system of simultaneous equations.

$$\begin{aligned}(m-1+2\lambda)p_1 - (m-2)p_2 &= \lambda \\ -k\lambda p_{k-1} + (m-k+(k+1)\lambda)p_k - (m-k-1)p_{k+1} &= 0 \qquad (k = 2, 3, \ldots, m-1)\end{aligned} \tag{7}$$

Replacing the k^{th} equation in (7) with the sum of the equations in (7) from the k^{th} onwards, for $k = 1, \ldots, m-1$, gives

$$\begin{aligned}(m-1)p_1 + m\lambda p_{m-1} &= \lambda \\ -k\lambda p_{k-1} + (m-k)p_k + m\lambda p_{m-1} &= 0 \qquad (k = 2, 3, \ldots, m-1).\end{aligned} \tag{8}$$

Now (8) is equivalent to (7), since we have simply multiplied the coefficient matrix of (7) on the left by an upper triangular matrix with all entries 1. This transformation is clearly nonsingular. We will show by induction that (8) has solution

$$p_k = \frac{\lambda^k - \sum_{i=1}^{k}\binom{m}{i}p_{m-1}\lambda^{k-i+1}}{\binom{m-1}{k}} \qquad (k = 1, 2, \ldots, m-1). \tag{9}$$

When $k = 1$, the first equation in (8) is clearly satisfied by (9). Assume by induction that (9) is true for p_{k-1}, with $k \geq 2$. Then

$$\begin{aligned}p_k &= \frac{\lambda k}{m-k}\,p_{k-1} - \frac{\lambda m}{m-k}\,p_{m-1} \\ &= \frac{\lambda k}{m-k}\,\frac{\lambda^{k-1} - \sum_{i=1}^{k-1}\binom{m}{i}\lambda^{k-i}p_{m-1}}{\binom{m-1}{k-1}} - \frac{\lambda m}{m-k}\,p_{m-1} \\ &= \frac{\lambda^k - \sum_{i=1}^{k-1}\binom{m}{i}\lambda^{k-i+1}p_{m-1}}{\binom{m-1}{k}} - \frac{\binom{m}{k}}{\binom{m-1}{k}}\,\lambda p_{m-1} \\ &= \frac{\lambda^k - \sum_{i=1}^{k}\binom{m}{i}\lambda^{k-i+1}p_{m-1}}{\binom{m-1}{k}},\end{aligned}$$

continuing the induction. In particular, when $k = m-1$ we obtain

$$\begin{aligned}p_{m-1} &= \frac{\lambda^{m-1} - \sum_{i=1}^{m-1}\binom{m}{i}\lambda^{m-i}p_{m-1}}{\binom{m-1}{m-1}} = \lambda^{m-1} - \big((1+\lambda)^m - 1 - \lambda^m\big)p_{m-1}, \\ \text{i.e.}\quad p_{m-1} &= \frac{\lambda^{m-1}}{(1+\lambda)^m - \lambda^m}.\end{aligned} \tag{10}$$

Using (10), (9) can be rewritten

$$p_k \ = \ \frac{1}{\binom{m-1}{k}}\left(\lambda^k - \frac{\sum_{i=1}^{k}\binom{m}{i}\lambda^{m+k-i}}{(1+\lambda)^m - \lambda^m}\right) \qquad (k = 1, 2, \ldots, m-1). \tag{11}$$

In particular

$$p_1 = \frac{\lambda}{m-1}\left(1 - \frac{m\lambda^{m-1}}{(1+\lambda)^m - \lambda^m}\right). \tag{12}$$

Note that the equation for p_1 given above is decreasing in m, and hence by the same analysis any edge of larger size has a smaller value of p_1. A proof of this observation is contained in Appendix C.

3.2 The expected distance between X_T and Y_T

The stopping time for the pair of chains X_t and Y_t will be when the distance between them changes, in other words either a good or bad event occurs. The probability that we observe the bad event on a particular edge e with $w \in e$ is at most p_k as calculated above. Let ξ_t denote the number of empty vertices in e at time t when the process is started with $\xi_0 = k$. Now ξ_t can never reach 0 without first reaching $k-1$ and, since the process is Markovian, it follows that

$$p_k = \Pr(\exists t : \xi_t = 0 | \xi_0 = k) = \Pr(\exists s : \xi_s = k-1 | \xi_0 = k)\Pr(\exists t : \xi_t = 0 | \exists s : \xi_s = k-1) < p_{k-1}.$$

Since w is in at most Δ edges, the probability that we observe the bad event on any edge is at most Δp_1. The probability that the stopping time ends with the good event is therefore at least $1 - \Delta p_1$. The path coupling calculation is then

$$\mathbf{E}[\mathrm{d}(X_T, Y_T)] \leq 2\Delta p_1.$$

This is required to be less than 1 in order to apply Theorem 2.1. If $m \geq 2\lambda\Delta + 1$, then by (12) and the remark following it,

$$2\Delta p_1 \leq 1 - \frac{(2\lambda\Delta + 1)\lambda^{2\lambda\Delta}}{(1+\lambda)^{2\lambda\Delta+1} - \lambda^{2\lambda\Delta+1}}.$$

Proof of Theorem 3.1. The above work puts us in a position to apply Theorem 2.1. Let $m \geq 2\lambda\Delta + 1$. Then for $\mathcal{M}(\mathcal{H})$ we have

(i) $\Pr(\mathrm{d}(X_t, Y_t) \neq 1 | \mathrm{d}(X_{t-1}, Y_{t-1}) = 1) \geq \frac{1}{n}$ for all t, and

(ii) $\mathbf{E}[\mathrm{d}(X_T, Y_T)] \ < \ 1 \ - \ \dfrac{(2\lambda\Delta + 1)\lambda^{2\lambda\Delta}}{(1+\lambda)^{2\lambda\Delta+1} - \lambda^{2\lambda\Delta+1}}.$

Also for $\mathcal{M}(\mathcal{H})$ we have $D_1 = n$ and $D_2 = 2$. Hence by Theorem 2.1, $\mathcal{M}(\mathcal{H})$ mixes in time

$$\tau(\varepsilon) \leq 6n\frac{(1+\lambda)^{2\lambda\Delta+1} - \lambda^{2\lambda\Delta+1}}{(2\lambda\Delta + 1)\lambda^{2\lambda\Delta}} \ln\left(2n\varepsilon^{-1}\frac{(1+\lambda)^{2\lambda\Delta+1} - \lambda^{2\lambda\Delta+1}}{(2\lambda\Delta + 1)\lambda^{2\lambda\Delta}}\right).$$

This is $O(n \log n)$ for fixed λ, Δ. □

Remark 3.2. In the most important case, $\lambda = 1$, we require $m \geq 2\Delta + 1$. This does not include the case $m = 3$, $\Delta = 3$ considered in [12]. We have attempted to improve the bound by employing the chain proposed by Dyer and Greenhill in [12, Section 4]. However, this gives only a marginal improvement. For large $\lambda\Delta$, we obtain convergence for $m \geq 2\lambda\Delta + \frac{1}{2} + o(1)$. For $\lambda = 1$, this gives a better bound on mixing time for $m = 2\Delta + 1$, with dependence on Δ similar to Remark 3.3 below, but does not even achieve mixing for $m = 2\Delta$. We omit the details in order to deal with the Glauber dynamics, and to simplify the analysis.

Remark 3.3. The terms in the running time which are exponential in λ, Δ would disappear if we instead took graphs for which $m \geq 2\lambda\Delta + 2$. In this case the running time would be

$$\tau(\varepsilon) \leq 6(2\lambda\Delta + 1)n \ln(2n\varepsilon^{-1}(2\lambda\Delta + 1)) \leq 12(2\lambda\Delta + 1)n \ln(n\varepsilon^{-1}).$$

Furthermore, if we took graphs such that $m - 1 > (2 + \delta)\lambda\Delta$, for some $\delta > 0$, then the running time would no longer depend on λ, Δ at all, but would be $\tau(\varepsilon) \leq c_\delta n \ln(n\varepsilon^{-1})$ for some constant c_δ.

Remark 3.4. It seems that path coupling cannot show anything better than m linear in $\lambda\Delta$. Consider the following specific example. Suppose the initial configuration has edges $\{w, v_1, \ldots, v_{m-2}, x_i\}$ for $i = 1, \ldots, \Delta$, with $w, v_1, \ldots, v_{m-2} \in X_0$, $x_1, \ldots, x_\Delta \notin X_0$ and w the change vertex. Consider the first step where any vertex changes state. Let $\mu = (1 + \lambda)(m - 1 + \Delta)$. The good event occurs with probability $(1 + \lambda)/\mu$, insertion of a critical vertex with probability $\lambda\Delta/\mu$, and deletion of a non-critical vertex with probability $(m - 2)/\mu$. We therefore need $(m - 2) + (1 + \lambda) \geq \lambda\Delta$, i.e. $m \geq \lambda(\Delta - 1) + 1$, to show convergence by path coupling.

Remark 3.5. It seems we could improve our bound $m \geq 2\lambda\Delta + 1$ for rapid mixing of the Glauber dynamics somewhat if we could analyse the process on all edges simultaneously. Examination of the extreme cases, where all edges adjacent to w are otherwise independent, or where they are dependent except for one vertex (as in Remark 3.4), suggests that improvement to $(1 + o(1))\lambda\Delta$ may be possible, where the $o(1)$ is relative to $\lambda\Delta$. However, the analysis in the general case seems difficult, since edges can intersect arbitrarily.

4 Hardness results for independent sets

We have established that the number of independent sets of a hypergraph can be approximated efficiently using the Markov Chain Monte Carlo technique for hypergraphs with edge size linear in Δ. We show next that exact counting is unlikely to be possible, and that our approximation scheme cannot be extended to cover all hypergraphs with edge size $\Omega(\log \Delta)$.

4.1 #P-completeness

We show that the exact counting problem is #P-Complete except in trivial cases.

Theorem 4.1. *Let $\mathcal{G}(m, \Delta)$ be the class of hypergraphs with minimum edge size $m \geq 3$ and maximum degree Δ. Computing the number of independent sets of hypergraphs in $\mathcal{G}(m, \Delta)$ is #P-complete if $\Delta \geq 2$. If $\Delta \leq 1$, it is in P.*

Proof. The cases $\Delta = 0, 1$ are trivially in P. As discussed in Section 1.2, independent sets in a hypergraph with $\Delta = 2$ correspond to edge covers in a graph (of minimum degree m). Counting these is #P-complete, even for graphs with arbitrarily large minimum degree. This is stated in [3] but without proof, so we provide a proof in Appendix A. We now consider $\Delta \geq 3$. (The case $m = \Delta = 3$ is discussed in [12].) Take a graph $G = (V, E)$, and construct a hypergraph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ by "extending" each edge $e = \{v_1, v_2\} \in E$ to an edge $e^+ = \{v_1, u_1^e, \ldots, u_{m-2}^e, v_2\} \in \mathcal{E}$. Observe that, for each vertex subset U of G and edge $e \in E$, there are $2^{m-2} - 1$ independent assignments to $u_1^e, \ldots, u_{m-2}^e$ if $v_1, v_2 \in U$ and 2^{m-2} otherwise. This is equivalent to evaluating the partition function of a *weighted H-colouring* problem [5, 13] on G, with weight matrix

$$A = \begin{bmatrix} 2^{m-2} & 2^{m-2} \\ 2^{m-2} & 2^{m-2} - 1 \end{bmatrix}.$$

The #P-completeness of H-colouring with this weight matrix follows either directly from [5] or indirectly from [13, Corollary 3.2]. The degree bound $\Delta = 3$ follows from [13, Theorem 5.1], on noting that A is nonsingular. □

4.2 Approximation hardness

We now show that unless NP=RP, there can be no *fpras* for the number of independent sets of all hypergraphs with edge size $\Omega(\ln \Delta)$.

Let $G = (V, E)$, with $|V| = n$, be a graph with maximum degree Δ and N_i independent sets of size i ($i = 0, 2, \ldots, n$). For $\lambda > 0$ let $Z_G(\lambda) = \sum_{i=0}^{n} N_i \lambda^i$ define the *hard core partition function*. The following is a combination of results in Luby and Vigoda [23] and Berman and Karpinski [1].

Theorem 4.2. *If $\lambda > 694/\Delta$, there is no* fpras *for $Z_G(\lambda)$ unless NP=RP.*

Proof. Let ε be a constant such that the size of the largest independent set in a graph of maximum degree 4 cannot be approximated to within a ratio $(1+\varepsilon)$ unless NP = RP. Berman and Karpinski [1] show that $\varepsilon \geq 1/49$. Luby and Vigoda [23, Theorem 4] prove the hardness of approximating $Z_G(\lambda)$ if $\lambda > c/\Delta$ for any $c > 20 \ln 2\,(1+\varepsilon)/\varepsilon$.[1] Together, these two results give the theorem. □

We note that Theorem 4.2 could probably be strengthened using the approach of [9]. However, this has yet to be done.

Theorem 4.3. *Unless NP=RP, there is no* fpras *for counting independent sets in hypergraphs with maximum degree Δ and minimum edge size $m < 2\log_2(1 + \Delta/694) - 1 = \Omega(\ln \Delta)$.*

Proof. Given a graph $G = (V, E)$ with maximum degree Δ, we construct a hypergraph $\mathcal{H} = (\mathcal{V}, \mathcal{E})$ as follows. Let $k = \lceil m/2 \rceil$. For each $v \in V$, let $W_v = \{w_{v1}, w_{v2}, \ldots, w_{vk}\}$ and $\mathcal{V} = \bigcup_{v \in V} W_v$. For each edge $e = \{u, v\} \in E$, let $S_e = W_u \cup W_v$, and let $\mathcal{E} = \{S_e : e \in E\}$. It is clear that $\mathcal{H}$ has maximum vertex degree Δ and every edge has size $2k \geq m$.

An independent set $\mathcal{I}$ in $\mathcal{H}$ corresponds to a unique independent set I in G as follows. If $W_v \subseteq \mathcal{I}$, then $v \in I$, otherwise $v \notin I$. Clearly $\mathcal{I}$ independent in $\mathcal{H}$ implies I independent in G. Note that for each $v \notin I$, there are $(2^k - 1)$ possible subsets of W_v which may be in $\mathcal{I}$. Thus, if $\mathcal{N}$ is the number of independent sets in $\mathcal{H}$,

$$\mathcal{N} = \sum_{i=0}^{n} N_i (2^k - 1)^{n-i} = (2^k - 1)^n \sum_{i=0}^{n} N_i (2^k - 1)^{-i} = (2^k - 1)^n Z_G(1/(2^k - 1)).$$

[1] The expression in [23] omits the $\ln 2$ term

Thus approximating $\mathcal{N}$ is equivalent to approximating $Z_G(\lambda)$ with $\lambda = 1/(2^k - 1)$. But, by Theorem 4.2, this will be hard if $1/(2^k - 1) > 694/\Delta$. This gives $k < \log_2(1 + \Delta/694)$, which holds whenever $m < 2\log_2(1 + \Delta/694) - 1$. □

5 Hypergraph colouring

We now consider Glauber dynamics on the set of proper colourings of a hypergraph. Again our hypergraph $\mathcal{H}$ will have maximum degree Δ, minimum edge size m, and we will have a set of q colours. A colouring of the vertices of $\mathcal{H}$ is proper if no edge is monochromatic. Let $\Omega(\mathcal{H})$ be the set of all (proper and improper) q-colourings of $\mathcal{H}$ and let $\Omega'(\mathcal{H})$ be the set of proper q-colourings of $\mathcal{H}$. We define the Markov chain $\mathcal{C}(\mathcal{H})$ with state space $\Omega(\mathcal{H})$ by the following transition process. If the state of $\mathcal{C}$ at time t is X_t, the state at $t+1$ is determined by

(i) selecting a vertex $v \in \mathcal{V}$ and a colour $k \in \{1, 2, \ldots, q\}$ uniformly at random,

(ii) let X'_t be the colouring obtained by recolouring v colour k

(iii) if no edge containing v is monochromatic in X'_t let $X_{t+1} = X'_t$
otherwise let $X_{t+1} = X_t$.

It is easily shown that the stationary distribution of this chain is uniform on $\Omega'(\mathcal{H})$, in particular every improper colouring has zero probability in the stationary distribution. Moreover, when restricted to the state space of proper colourings the chain is ergodic for $q \geq \Delta + 1$ and $m \geq 3$. For $m \geq 4$ this is a consequence of the path coupling contraction shown below, for $m \geq 3$ it is easily shown using standard arguments, since each edge can only block a colour for one vertex. For graphs, $m = 2$, ergodicity requires $q \geq \Delta + 2$. In the analyses that follow we will use Hamming distance as the metric on $\Omega(\mathcal{H})$ and take S to be the set of pairs of colourings at distance one. Again we will use Theorem 2.1 to prove rapid mixing of this chain under certain conditions, however first we will examine the chain using standard path coupling techniques.

Theorem 5.1. *For $m \geq 4$, $q > \Delta$, the Markov chain $\mathcal{C}(\mathcal{H})$ mixes in time $\tau(\varepsilon) \leq nq\ln(n\varepsilon^{-1})$.*

Proof. Suppose that two copies of $\mathcal{C}(\mathcal{H})$, X_t and Y_t say, start at distance one apart, i.e. X_0 and Y_0 differ in only one vertex w. We assume that w is coloured blue in X_0 and red in Y_0. Suppose that the number of colours available for recolouring w is $q - k$, then the probability of the two copies of the chain coupling in one step is $\frac{q-k}{nq}$. The distance between the two chains can only increase (to 2) if we select a vertex v and recolour it with a colour that is permitted in one copy of the chain only. For this to happen, there must be an edge containing v and w such that the other vertices in this edge are all either red and we have chosen red for v, or blue and we have chosen blue for v. Hence, since $m \geq 4$, there can be at most one vertex on each edge, and one colour for that vertex, such that the chains diverge if we select that vertex and colour. Furthermore, for each of the k unavailable colours there must be an edge containing w which, apart from w itself, is monochromatic in the forbidden colour, so on these edges there are no vertices whose selection can cause the chains to diverge (this holds for improper edges containing w also). Hence the probability that the distance increases to 2 in one step is at most $\frac{\Delta-k}{nq}$. The path coupling calculation is therefore

$$\mathbf{E}[\mathrm{d}(X_1, Y_1)] \leq 1 - \frac{q-k}{nq} + \frac{\Delta - k}{nq}.$$

If $q \geq \Delta+1$ then $\mathbf{E}[\mathrm{d}(X_1, Y_1)] \leq 1-1/nq$, and therefore by the path coupling theorem (Theorem 1.1) the mixing time is

$$\tau(\varepsilon) \leq nq \ln(n\varepsilon^{-1}).$$

□

This analysis leaves little room for improvement in the case $m \geq 4$, indeed it is not clear whether the Markov chain described is even ergodic for $q \leq \Delta$. The following simple construction does show that the chain is not in general ergodic if $q \leq \frac{\Delta}{m} + 1$. Take integers q, Δ and m such that $q = \frac{\Delta}{m}+1$, and form a hypergraph $\mathcal{H}$ on $q(m-1)$ vertices. We will partition the vertices into q parts $\mathcal{V} = \mathcal{V}_1 \cup \mathcal{V}_2 \cup \ldots \cup \mathcal{V}_q$, each of size $m-1$. Then the edge set of $\mathcal{H}$ is $\mathcal{E} = \{\{v\} \cup \mathcal{V}_j : v \in \mathcal{V}, v \notin \mathcal{V}_j\}$. The degree of each vertex is $(q-1) + (q-1)(m-1) = \Delta$. If we now colour each group $\mathcal{V}_j$ a different colour, we obtain $q!$ distinct colourings, but for each of these the Markov chain is frozen (no transition is valid).

The case $m = 2$ is graph colouring and has been extensively studied. See, for example, [8]. This leaves the case $m = 3$, hypergraphs with 3 vertices in each edge. The standard path coupling argument, as in Theorem 5.1, only shows rapid mixing for $q > 2\Delta$, since there may be two vertices in each edge that can be selected and lead to a divergence of the two chains. This occurs if, of the two vertices in an edge which are not w, one is coloured red and the other blue. However, we can do better using Theorem 2.1.

Theorem 5.2. *There exists a Δ^* such that, if $\mathcal{H}$ is a 3-uniform hypergraph with maximum degree $\Delta > \Delta^*$ and $q \geq 1.65\Delta$, the Markov chain $\mathcal{C}(\mathcal{H})$ mixes rapidly, i.e. $\tau(\varepsilon) \leq O\big(nq \ln(n/\varepsilon)\big)$.*

Proof. We choose Δ^* large enough that all the approximations below are valid. We will use $\approx$ to imply equality up to a factor $1 + o_\Delta(1)$ for all times t until the stopping time, which we will bound below. We couple two copies of this chain using the identity coupling. Let X_t and Y_t be two copies of $\mathcal{C}(\mathcal{H})$ such that X_0 and Y_0 differ only at a single vertex w. If w lies in fewer than 0.825Δ edges, then q is more than twice the degree of w and we may obtain contraction in one step as discussed above. Hence we make Δ^* large enough for the approximations to hold for degrees at least 0.825Δ, and since q is certainly at least 1.65 times the degree of w, we may assume without loss that w lies in exactly Δ edges. As before, we will examine the stopping time T at which $\mathrm{d}(X_T, Y_T) \neq 1$ for the first time, and show that $\mathbf{E}[\mathrm{d}(X_T, Y_T)] < 1$. We assume that w is coloured blue in X_0 and red in Y_0. We will call any other colour *neutral*. Let $\Gamma(w)$ denote the set of vertices of $\mathcal{H}$ that share an edge with w, and let $M = |\{w\} \cup \Gamma(w)| = |\Gamma(w)| + 1$. For any $v \in \Gamma(w)$, d_v will denote the number of edges containing v and w. Note that $\sum_{v \in \Gamma(w)} d_v = 2\Delta$, and $|\Gamma(w)| \leq 2\Delta$. We also assume that $|\Gamma(w)| \geq \Delta/4$, since otherwise the standard arguments will give contraction in one step as follows. As in Theorem 5.1, the probability of coupling in one step is at least $\frac{q-\Delta}{nq}$ whereas the probability of a new vertex of disagreement being created is at most $\frac{2|\Gamma(w)|}{nq}$. Thus we expect contraction when $|\Gamma(w)| < (q-\Delta)/2$, which certainly follows if $M \leq \Delta/4$ since $q - \Delta > 0.65\Delta$.

We will consider only transitions in which either w or a vertex in $\Gamma(w)$ is selected, since any transition involving any other vertex cannot change the distance between X and Y. Hence in each step we choose a vertex uniformly at random from $\{w\} \cup \Gamma(w)$ with probability $1/M$. Since w always has at least $(q - \Delta)$ valid colour choices, the probability that the process has not coupled or diverged by time 100Δ is at most

$$\left(1 - \frac{q-\Delta}{Mq}\right)^{100\Delta} \leq \left(1 - \frac{0.65\Delta}{(2\Delta+1) \times 1.65\Delta}\right)^{100\Delta} \leq \left(1 - \frac{1}{5.25\Delta}\right)^{100\Delta} \leq e^{-19} < 10^{-8}, \quad (13)$$

for large enough Δ, so we will only consider times up to $T' = \min\{T, 100\Delta\}$.

We will call any edge adjacent to w for which $e \setminus \{w\}$ is monochromatic *blocking*. Let ρ_t denote the number of blocking edges at time t. Any edge which contains w and another vertex which is coloured either red or blue will be called *dangerous*; let Δ_t denote the number of dangerous edges at time t. Any other (non-blocking, non-dangerous) edge containing w we will call *safe*. Let σ_t denote the number of safe edges at time t. The total number of blocking edges created by time T', ν_1 say, is negligible since

$$\Pr(\nu_1 \geq 200) \;\leq\; \binom{100\Delta}{200}\Big(\frac{1}{q}\Big)^{200} \;\leq\; \Big(\frac{e}{3.3}\Big)^{200} \;<\; 10^{-16}. \tag{14}$$

Since 200 is negligible by comparison with q, these edges do not affect the estimates below. Thus we may assume, within the bounds of our approximation, that no blocking edges are created by time T'.

Similarly, let ν_2 be the total number of times that vertices are recoloured red or blue by time T'. Then ν_2 satisfies

$$\Pr(\nu_2 \geq 400) \;\leq\; \binom{100\Delta}{400}\Big(\frac{2}{q}\Big)^{400} \;\leq\; \Big(\frac{e}{3.3}\Big)^{400} \;<\; 10^{-32}. \tag{15}$$

However, we must also consider the number of dangerous edges that such a recolouring could create. If $v \in \Gamma(w)$ is recoloured red or blue, this could be as large as d_v. Thus the total number of dangerous edges created by time T', ν_3 say, satisfies

$$\mathbf{E}[\nu_3] \;\leq\; 100\Delta \sum_{v\in\Gamma(v)} \frac{2d_v}{Mq} \;=\; \frac{400\Delta^2}{Mq} \;<\; 1000. \tag{16}$$

Hence, using Markov's inequality, $\Pr(\nu_3 > \ln\Delta) = o(1)$. Again, this does not affect our estimates, so we may assume that all red and blue vertices are present initially.

Let S_t denote the event that $T = t$ and $\mathrm{d}(X_T, Y_T) = 0$, which we will call *success*. The event $\mathrm{d}(X_T, Y_T) = 2$ we will call *failure*. The probability that failure occurs (by time T') on an edge which is not initially dangerous is at most

$$\Delta\binom{100\Delta}{2}\frac{2}{(Mq)^2} \;=\; O\Big(\frac{1}{\Delta}\Big). \tag{17}$$

Thus we may disregard safe edges, since they simply reduce the "effective" degree. Blocking edges also reduce the effective degree, but in addition they reduce the number of colours available to recolour w. We have already shown that we can assume that Δ_t is nonincreasing with t. Note that $\Delta_t = \Delta - \rho_t - \sigma_t$. Let q_t be the number of colours available for recolouring w at time t. Note that q_t is nondecreasing with t, since we may assume blocking edges are not created, and $q_t \geq q - \rho_t$.

Suppose we can show success occurs with probability greater than $1/2 + \kappa$ whenever $q \geq c\Delta$ and $\rho_0 = 0 = \sigma_0$, where κ is the sum of the probabilities given in Equations (14),(15) and (17). (Note that we can take Δ large enough that $\kappa < 0.001$. We wish to show eventually that $c \leq 1.65$.)

Then for a configuration in which ρ_0 and σ_0 are not necessarily zero, the analysis may be applied to the configuration obtained by deleting the safe and blocking edges at time 0 and using only q_0 colours (i.e. we temporarily consider some edges, but no vertices, as not being in the hypergraph). This gives Δ_0 edges and a smaller (but constant) value of M for the vertices of interest. This is valid since, everything that happens on any other vertices we know will not lead to failure (with probability at least $1 - \kappa$), and we may also assume that w has at least q_0 valid colour choices at each step and each vertex other than w has at least $q - \Delta$ valid colour choices at each step (where these choices could be determined by an adversary which is trying to maximise the failure probability). Since $c\Delta_0 - 0 \leq c\Delta - \rho_0 = q_0$ and $(c-1)\Delta_0 \leq (c-1)\Delta = q - \Delta$, the edge deletion does not increase the minimum number of colour choices at any vertex. Thus it doesn't increase the estimate of the success probability.

So we have done no harm by removing the safe and blocking edges except that we have to run the process starting with q_0 colours rather than q. Since $q \geq c\Delta$ implies $q_0 \geq c\Delta_0$, we still have sufficient colours for the analysis.

There is one more point to observe. The analysis requires that Δ be large. However Δ_0, the effective Δ, could be small. But suppose $\Delta_0 < (c-1)\Delta$, so $\rho_0 + \sigma_0 > (2-c)\Delta$. If $q_0 > 2\Delta_0$, we will use standard path coupling, as discussed above. Otherwise $q - \rho_0 \leq 2(\Delta - \rho_0 - \sigma_0)$, i.e. $q \leq 2\Delta - \rho_0 - 2\sigma_0$. Thus $q \leq 2\Delta - (\rho_0 + \sigma_0) < c\Delta$, a contradiction. Thus, if $q_0 \leq 2\Delta_0$, we may assume $\Delta_0 \geq (c-1)\Delta$. Then $\Delta_0 > (c-1)\Delta^*$ for $\Delta > \Delta^*$, so we simply take Δ^* to be larger.

Therefore, we now assume that Δ is large and the initial configuration has no blocking or safe edges. Thus the probability that the two chains couple in any one step is always $\frac{q}{Mq} = \frac{1}{M}$. For each $v \in \Gamma(w)$, let $\beta_{v,t}$ be an indicator variable which takes value 1 if $T > t$ and v is either red or blue after t steps of the chain, and takes value 0 otherwise. Note that if $T > t$ then X_t and Y_t agree on $\Gamma(w)$ so $\beta_{v,t}$ is well defined. Also, from above, we can assume that $\beta_{v,t}$ is nonincreasing with t. We describe a choice of vertex $v \in \Gamma(w)$ and colour $c \in \{\text{red,blue}\}$ at step t as "bad" if v and w lie in a dangerous edge whose third vertex is currently coloured c. Let B_t denote the number of bad choices at time t. The probability of failure in step t is therefore $\frac{B_t}{Mq}$. For each $v \in \Gamma(w)$ let d_v be the number of edges which contain both v and w. Then $B_t \leq \sum_{v \in \Gamma(w)} d_v \beta_{v,t}$.

$$\begin{aligned}
\mathbf{E}[\Pr(S_t)] = \; & \mathbf{E}\Big[\prod_{j=0}^{t-1}\Big(1 - \frac{1}{M} - \frac{B_j}{Mq}\Big)\frac{1}{M}\Big] \;\approx\; \frac{1}{M}\mathbf{E}\Big[e^{-\sum_{j=0}^{t-1}(\frac{1}{M}+\frac{B_j}{Mq})}\Big] \\
\geq \; & \frac{1}{M}\mathbf{E}\Big[e^{-\sum_{j=0}^{t-1}(\frac{1}{M}+\frac{\sum_{v\in\Gamma(w)} d_v\beta_{v,j}}{Mq})}\Big] \;=\; \frac{e^{-t/M}}{M}\mathbf{E}\Big[e^{-\frac{1}{Mq}\sum_{v\in\Gamma(w)} d_v \sum_{j=0}^{t-1}\beta_{v,j}}\Big] \\
\geq \; & \frac{e^{-t/M}}{M} e^{-\frac{1}{Mq}\sum_{v\in\Gamma(w)} d_v \mathbf{E}\big[\sum_{j=0}^{t-1}\beta_{v,j}\big]}, \qquad (18)
\end{aligned}$$

using Jensen's inequality. Let $Z_{v,t} = \sum_{j=0}^{t-1}\beta_{v,j}$, so we want to estimate $\mathbf{E}\big[Z_{v,t}\big]$. Let

$$t_v = \max\{t : \beta_{v,t} = 1 \text{ and } t < 100\Delta\} + 1,$$

then $Z_{v,t} = \min\{t_v, t\}$. Now, if v starts out either red or blue, the probability that it is recoloured to a neutral colour in any step is at least $(q - \Delta - 2)/Mq$. Also, the probability that it becomes red or blue before time 100Δ is at most $200\Delta/Mq$. Hence

$$\Pr(t_v > t) \leq \Big(1 - \frac{q-\Delta-2}{Mq}\Big)^t + \frac{200\Delta}{Mq} \;\approx\; e^{-\frac{q-\Delta}{Mq}t} \;=\; e^{-\gamma t},$$

where $\gamma = (q-\Delta)/Mq = \Theta(1/\Delta)$, since the second term is $O(1/\Delta)$ and small compared to the first, which is $\Omega(1)$ for $t \leq 100\Delta$. Thus t_v is dominated by an exponentially distributed random variable t'_v with rate γ, and $Z_{v,t}$ is dominated by $Z'_{v,t} = \min\{t'_v, t\}$. Thus

$$\mathbf{E}[Z_{v,t}] \;\leq\; \mathbf{E}[Z'_{v,t}] \;=\; \mathbf{E}[\min\{t'_v, t\}] \approx \int_0^t e^{-\gamma x}dx \;=\; (1 - e^{-\gamma t})/\gamma.$$

Therefore

$$\sum_{v\in\Gamma(w)} d_v\mathbf{E}[Z_{v,t}] \;\le\; \sum_{v\in\Gamma(w)} d_v(1-e^{-\gamma t})/\gamma \;\le\; 2\Delta(1-e^{-\gamma t})/\gamma,$$

and hence

$$\frac{1}{Mq}\sum_{v\in\Gamma(w)} d_v e\Big[\sum_{j=0}^{t-1}\beta_{v,j}\Big] \;\le\; \frac{2\Delta(1-e^{-\gamma t})}{Mq\gamma} \;=\; \frac{2\Delta}{q-\Delta}\Big(1-e^{-\frac{q-\Delta}{Mq}t}\Big).$$

Hence, for large enough Δ and $t \le 100\Delta$, from (18) we have

$$\mathbf{E}[\Pr(S_t)] \;\ge\; \frac{1}{M}e^{-\frac{t}{M}-\frac{2\Delta}{q-\Delta}(1-e^{-\frac{q-\Delta}{Mq}t})}.$$

Finally, noting that $\Pr(\mathrm{d}(X_T,Y_T)=0)=\sum_{t=0}^{\infty}\Pr(S_t)$ by linearity of expectation, we have

$$\begin{aligned}\Pr(\mathrm{d}(X_T,Y_T)=0) \;&\ge\; \int_0^{100\Delta}\frac{1}{M}e^{-\frac{t}{M}-\frac{2\Delta}{q-\Delta}(1-e^{-\frac{q-\Delta}{Mq}t})}dt\\ &=\; \int_0^{\frac{100\Delta}{M}} e^{-z-\frac{2\Delta}{q-\Delta}(1-e^{-\frac{q-\Delta}{q}z})}dz\\ &\ge\; \int_0^{49.99} e^{-z-\frac{2\Delta}{q-\Delta}(1-e^{-\frac{q-\Delta}{q}z})}dz\end{aligned}$$

since $M \le 2\Delta+1$ and Δ is large. If we substitute $q=1.65\Delta$, we see that

$$\Pr(\mathrm{d}(X_T,Y_T)=0) \;\ge\; \int_0^{49.99} e^{-z-3.077(1-e^{-0.3940z})}dz \;>\; 0.5003.$$

Since $\mathrm{d}(X_T,Y_T)\in\{0,2\}$, it follows that $\mathbf{E}[\mathrm{d}(X_T,Y_T)] < 0.9994$ and we can apply Theorem 2.1. This yields the claimed result. □

Remark 5.3. If we let $\beta=(q-\Delta)/q$ then, as $\Delta^*\to\infty$, the analysis can be tightened slightly to work for $\beta>\beta^*$, where β^* is the root of the equation

$$\int_0^{\infty} e^{-z-\frac{2(1-\beta)}{\beta}(1-e^{-\beta z})}dz \;=\; \tfrac{1}{2}.$$

The integral can be expanded, by parts integration, as an infinite series to give an alternative equation

$$\sum_{i=0}^{\infty}\frac{(-2)^i(1-\beta)^i}{\prod_{j=0}^{i}(1+j\beta)} \;=\; \tfrac{1}{2}.$$

This has root $\beta^*=0.392729$, giving $q>1.64671\Delta$.

Remark 5.4. A route to improving our bound on q would be to consider the changes in the numbers of colours available at each vertex of $\Gamma(w)$ during the process. We make the pessimistic assumption that this is always $q-\Delta$ but, while this could be true initially, we would expect more colours to become available later on. A proper analysis of this effect seems more difficult, however, because $\Theta(\Delta^2)$ vertices are now involved, and the edges containing them may intersect.

6 Hardness results for colouring

6.1 #P-completeness

Again we show that exact counting is #P-complete except in the few cases where it is clearly in P. Let $\mathcal{G}(m, \Delta)$ be as in Theorem 4.1.

Theorem 6.1. *For $m \geq 3$, computing the number of q-colourings of hypergraphs in $\mathcal{G}(m, \Delta)$ is #P-complete if $\Delta, q > 1$. If $\Delta \leq 1$ or $q \leq 1$ it is in P.*

Proof. The cases $\Delta \leq 1$, $q \leq 1$ are trivially in P. The case $\Delta = 2$ corresponds to counting edge q-colourings of graphs (of minimum degree m) in which no vertex is monochromatic. We call an edge colouring with no monochromatic vertex a *weak edge colouring*. Counting weak edge colourings is #P-complete for graphs of arbitrarily large minimum degree. We give a proof in Appendix B.

For $\Delta \geq 3$, $q \geq 2$, we use the construction from the proof of Theorem 4.1. For each colouring $X : V \to \{1, 2, \ldots, q\}$ of G and edge $e \in E$, there are $q^{m-2} - 1$ permitted colourings of $u_1^e, \ldots, u_{m-2}^e$ if $X(v_1) = X(v_2)$ and q^{m-2} otherwise. The corresponding H-colouring problem has the following $q \times q$ weight matrix:

$$A = \begin{bmatrix} q^{m-2} - 1 & q^{m-2} & \cdots & q^{m-2} \\ q^{m-2} & q^{m-2} - 1 & \cdots & q^{m-2} \\ \vdots & \vdots & \ddots & \vdots \\ q^{m-2} & q^{m-2} & \cdots & q^{m-2} - 1 \end{bmatrix}.$$

The #P-completeness of H-colouring with this weight matrix, and the bound $\Delta = 3$, follow as in Theorem 4.1, since A is again nonsingular. □

6.2 Hardness of Approximation

Again let $\mathcal{G}(m, \Delta)$ be as defined in Theorem 4.1. Our result, Corollary 6.3, follows directly from the following NP-completeness proof.

Theorem 6.2. *Determining whether a hypergraph in $\mathcal{G}(m, \Delta)$ has any q-colouring is NP-complete for any $m > 1$ and $q > 2$ such that $\binom{mq-1}{m-1} \leq \Delta$. This holds for all $2 < q \leq e^{-m/(m-1)} \Delta^{1/(m-1)}$.*

Proof. If $m = 2$, this is graph colouring, and the result follows from [14, Theorem 1.4]. (See also [26].) For $m \geq 3$, we use the following reduction from graph colouring. Let m, Δ and $q > 2$ such that $\binom{mq-1}{m-1} \leq \Delta$ be given. Let $\Delta_G = \lceil 4q/3 \rceil$, then determining whether a graph $G = (V, E)$ with maximum degree Δ_G is q-colourable is NP-complete, since $2 < q \leq 3\Delta_G/4$ [14]. For each edge $e = \{v_1, v_2\} \in E$, let $S_i^e = \{u_{i1}^e, u_{i2}^e, \ldots, u_{im}^e\}$ $(i = 1, 2, \ldots, q)$ and $\mathcal{V}_0^e = \bigcup_{i=1}^q S_i^e$. Let $\mathcal{E}_0^e$ comprise all subsets of $\mathcal{V}_0^e$ of size m other than S_i^e $(i = 1, 2, \ldots, q)$. We claim that any proper q-colouring of the hypergraph $\mathcal{H}_0^e = (\mathcal{V}_0^e, \mathcal{E}_0^e)$ must assign the same colour to all $u_{ij}^e \in S_i^e$ $(j = 1, 2, \ldots, m)$, and a different colour for each $i = 1, \ldots, q$. The claim holds since there must be some colour class of size at least m, since there are q colours and mq vertices. If there was a colour class of size greater than m, at least one of its subsets of size m would be a monochromatic edge. Thus there must be exactly q colour classes, each of size m. If these are not the S_i^e $(i = 1, 2, \ldots, q)$, again there is a monochromatic subset of size m which is an edge. Clearly, by symmetry, any assignment of the q colours to the q classes S_i^e is permissible.

Let $\mathcal{V}^e = \mathcal{V}_0^e \cup \{v_1, v_2\}$, and add the edges $\{v_1, u_{i2}^e, \ldots, u_{im}^e\}$ $(i = 1, \ldots, \lfloor q/2 \rfloor)$ and $\{v_2, u_{i2}^e, \ldots, u_{im}^e\}$ $(i = \lfloor q/2 \rfloor + 1, \ldots, q)$ to $\mathcal{E}_0^e$ to give $\mathcal{E}^e$. We claim that, in any proper q-colouring of the hypergraph $\mathcal{H}^e = (\mathcal{V}^e, \mathcal{E}^e)$, v_1 and v_2 must receive different colours. The claim holds since v_1 can have any colour different from all S_i^e $(i = 1, 2, \ldots, \lfloor q/2 \rfloor)$, and v_2 any colour different from all S_i^e $(i = \lfloor q/2 \rfloor + 1, \ldots, q)$. But these permitted colour sets for v_1 and v_2 are disjoint. Also, given any colours for v_1 and v_2, there are $\lfloor q/2 \rfloor \lceil q/2 \rceil (q-2)! > 0$ colourings of $\mathcal{H}^e$. Thus we may use $\mathcal{H}^e$ to simulate the edge $e \in E$. Thus we set $\mathcal{V} = \bigcup_{e \in E} \mathcal{V}^e$, $\mathcal{E} = \bigcup_{e \in E} \mathcal{E}^e$ and consider the hypergraph $\mathcal{H} = (\mathcal{V}, \mathcal{E})$. Then $\mathcal{H}$ is q-colourable if and only if G is q-colourable.

The maximum degree in $\mathcal{H}$ of any u_{ij} is $\binom{mq-1}{m-1}$. The degree in $\mathcal{H}$ of each $v \in V$ is at most $\Delta_G \lceil q/2 \rceil \leq (4q+2)(q+1)/6 < 2q^2$. Thus the maximum degree of $\mathcal{H}$ is $\binom{mq-1}{m-1} \leq \Delta$. So $\mathcal{H}$ is a hypergraph in the class $\mathcal{G}(m, \Delta)$, and the main statement of the theorem follows. Note that

$$\binom{mq-1}{m-1} \;=\; \frac{1}{q}\binom{mq}{m} \;<\; \frac{(eq)^m}{q} \;=\; e^m q^{m-1},$$

so $\binom{mq-1}{m-1} \leq \Delta$ whenever $e^m q^{m-1} \leq \Delta$, i.e. $q \leq e^{-m/(m-1)} \Delta^{1/(m-1)}$. □

Corollary 6.3. *Unless NP=RP, there is no* fpras *for counting q-colourings of a hypergraphs with maximum degree Δ and minimum edge size $m > 1$ if $q > 2$ and $\binom{mq-1}{m-1} \leq \Delta$. In particular this holds for all $2 < q \leq e^{-m/(m-1)} \Delta^{1/(m-1)}$.*

Proof. We cannot tell if there is *any* colouring for q in this range, so there can be no *fpras*. □

Remark 6.4. It is clearly a weakness that our lower bound for approximate counting is based entirely on an NP-completeness result. However, we note that the same situation pertains for graph colouring, which has been the subject of more intensive study.

7 Conclusions

We have presented an approach to the analysis of path coupling with stopping times which improves on the method of [17] in most applications. Our method may itself permit further development.

We apply the method to independent sets and q-colourings in hypergraphs with maximum degree Δ and minimum edge size m. In the case of independent sets, there seems scope for improving the bound $m \geq 2\Delta + 1$, but anything better than $m \geq \Delta + o(\Delta)$ would seem to require new methods. For colourings, there is probably little improvement possible in our result $q > \Delta$ for $m \geq 4$, but many questions remain for $m \leq \Delta$. For example, even the ergodicity of the Glauber (or any other) dynamics is not clearly established. For the most interesting case, $m = 3$, the bound $q > 1.65\Delta$ (for large Δ) can almost certainly be reduced, but substantial improvement may prove difficult.

Our #P-completeness results seem best possible for both of the problems we consider. On the other hand, our lower bounds for hardness of approximate counting seem very weak in both cases, and are far from our upper bounds. These lower bounds can probably be improved, but we have no plausible conjecture as to what may be the truth.

Acknowledgments

We are grateful to Tom Hayes for commenting on an earlier draft of this paper, and to Mary Cryan for useful discussions at an early stage of this work.

References

[1] P. Berman and M. Karpinski, Improved approximation lower bounds on small occurrence optimization, *Electronic Colloquium on Computational Complexity* **10** (2003), Technical Report TR03-008.

[2] R. Bubley, *Randomized algorithms: approximation, generation and counting*, Springer-Verlag, London, 2001.

[3] R. Bubley and M. Dyer, Graph orientations with no sink and an approximation for a hard case of #SAT, in *Proc. 8th Annual ACM-SIAM Symposium on Discrete Algorithms (SODA 1997)*, SIAM, 1997, pp. 248–257.

[4] R. Bubley, M. Dyer, C. Greenhill, and M. Jerrum, On approximately counting colourings of small degree graphs, *SIAM Journal on Computing* **29** (1999), 387–400.

[5] A. Bulatov and M. Grohe, The complexity of partition functions, in *Proc. 31st International Colloquium on Automata, Languages and Programming (ICALP 2004)*, Springer, 2004, pp. 294–306.

[6] I. Dinur, V. Guruswami, S. Khot and O. Regev, A new multilayered PCP and the hardness of hypergraph vertex cover, in *Proc. 35th ACM Symposium on Theory of Computing (STOC 2003)*, ACM, 2003, pp. 595–601.

[7] I. Dinur, O. Regev and C. Smyth, The hardness of 3-uniform hypergraph coloring, in *Proc. 43rd Symposium on Foundations of Computer Science (FOCS 2002)*, IEEE, 2002, pp. 33–42

[8] M. Dyer, A. Frieze, T. Hayes and E. Vigoda, Randomly coloring constant degree graphs, in *Proc. 45th Annual IEEE Symposium on Foundations of Computer Science (FOCS 2004)*, IEEE, 2004, pp. 582–589.

[9] M. Dyer, A. Frieze and M. Jerrum, On counting independent sets in sparse graphs, *SIAM Journal on Computing* **31** (2002), 1527–1541.

[10] M. Dyer, L. Goldberg, C. Greenhill, M. Jerrum and M. Mitzenmacher, An extension of path coupling and its application to the Glauber dynamics for graph colorings, *SIAM Journal on Computing* **30** (2001), 1962–1975.

[11] M. Dyer and C. Greenhill, A more rapidly mixing Markov chains for graph colouring, *Random Structures and Algorithms* **13** (1998), 210–217.

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

[13] M. Dyer and C. Greenhill, The complexity of counting graph homomorphisms, *Random Structures and Algorithms*, **17** (2000), 260–289. See also Corrigendum, *Random Structures and Algorithms* **25** (2004), 346–352.

[14] T. Emden-Weinert, S. Hougardy and B. Kreuter, Uniquely colourable graphs and the hardness of colouring graphs of large girth, *Combinatorics, Probability and Computing* **7** (1998), 375–386.

[15] M. Garey and D. Johnson, *Computer and intractability*, W. H. Freeman and Company, 1979.

[16] C. Greenhill, The complexity of counting colourings and independent sets in sparse graphs and hypergraphs, *Computational Complexity* **9** (2000), 52–73.

[17] T. Hayes and E. Vigoda, Variable length path coupling, in *Proc. 15th Annual ACM-SIAM Symposium on Discrete Algorithms (SODA 2004)*, SIAM, 2004, pp. 103–110.

[18] T. Hofmeister and H. Lefmann, Approximating maximum independent sets in uniform hypergraphs, *Proc. 23rd International Symposium on Mathematical Foundations of Computer Science (MFCS 1998)*, Lecture Notes in Computer Science **1450**, Springer, 1998, pp. 562–570.

[19] S. Janson, T. Łuczak and A. Ruciński, *Random graphs*, Wiley-Interscience, New York, 2000.

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

[21] M. Jerrum, *Counting, sampling and integrating: algorithms and complexity*, ETH Zürich Lectures in Mathematics, Birkhäuser, Basel, 2003.

[22] M. Krivelevich, R. Nathaniel and B. Sudakov, Approximating coloring and maximum independent sets in 3-uniform hypergraphs, in *Proc. 12th Annual ACM-SIAM Symposium on Discrete Algorithms, (SODA 2001)*, SIAM, 2001, pp. 327–328.

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

[24] M. Mitzenmacher and E. Nikova, Path coupling as a branching process, unpublished manuscript, 2002.

[25] M. Molloy, Very rapidly mixing Markov chains for 2Δ-coloring and for independent sets in a graph with maximum degree 4, *Random Structures and Algorithms* **18** (2001), 101–115.

[26] M. Molloy and B. Reed, Colouring graphs when the number of colours is nearly the maximum degree, in *Proc. 33rd Annual ACM Symposium on Theory of Computing (STOC 2001)*, ACM, 2001, pp. 462–470.

[27] J. Salas and A. Sokal, Absence of phase transition for anti-ferromagnetic Potts models via the Dobrushin uniqueness theorem, *Journal of Statistical Physics* **86** (1997), 551–579.

[28] N. Sloane, Sequence A006129, *The on-line encyclopedia of integer sequences*, 2004. Published at `http://www.research.att.com/~njas/sequences/`.

[29] E. Vigoda, A note on the Glauber dynamics for sampling independent sets, *The Electronic Journal of Combinatorics* **8**, R8(1), 2001.

Appendices

A Edge cover is #P-complete

Proof. Recall that for a graph $G(V,E)$ a subset U of the edge set E is an *edge cover* if every vertex of G is incident with at least one edge in U. Let $\mathcal{G}_{\Delta_0}$ be the class of graphs of minimum degree Δ_0. We prove that computing the number of edge covers of a graph is #P-complete, even for the class $\mathcal{G}_{\Delta_0}$, by reduction from counting independent sets. We use methods similar to Bubley and Dyer [3], where this result was claimed without proof. We assume we have an oracle for computing the number of edge covers of graphs in $\mathcal{G}_{\Delta_0}$. Let $\mathcal{G}$ be the class of all 3-regular graphs, for this class counting independent sets is #P-complete [16, Theorem 3.1]. Let $G = (V,E)$, with $I_j(G)$ independent sets of size j ($j = 0, 1, \ldots, n$). Form G' by subdividing each edge $e \in E$ with a new vertex u_e. Let $U = \{u_e : e \in E\}$. Let $N_i(G')$ be the number of edge subsets in G' which leave exactly i vertices in V uncovered, but no vertex in U. In particular, $N_0(G')$ is the number of edge covers of G'. Observe that the uncovered vertices in G' must form an independent set in G.

Let $W \subset V$ be an independent set of size j in G. The number of edge subsets of G' that do not cover W, do cover all members of U and may or may not cover the vertices in $V - W$ may be counted as follows. The $3j$ edges around members of W cannot be in the edge subset. Thus the $3j$ edges adjacent to these must be in. The remaining $3n - 6j$ edges are paired up according to the edge of G they came from. Each pair must have at least one member in the edge subset (in order to cover U), thus there are three possible configurations for the pair. Hence the number of such edge subsets of G' is $3^{(3n-6j)/2}$. It follows that

$$3^{3(n-2j)/2} I_j(G) \;=\; \sum_{i=j}^{n} \binom{i}{j} N_i(G').$$

Thus, if we can determine the $N_i(G')$, we can determine the number of independent sets of all sizes in G. Let $N_{ij}(G')$ be the number of edge subsets of G' in which i vertices in V and j in U are uncovered ($i = 0, \ldots, n$, $j = 0, \ldots, 3n/2$). Then $N_i(G') = N_{i0}(G')$. We attach a copy K_m^v of K_m to each vertex $v \in V$ by identifying v with a vertex in K_m, and a copy K_k^u of K_k to each vertex $u \in U$ similarly. Call the resulting graph G_{mk}. Let M_m be the number of edge covers of K_m, then M_{m-1} is the number of edge sets in K_m which leave a fixed vertex uncovered. We can show by inclusion-exclusion that

$$M_m \;=\; \sum_{i=0}^{m} (-1)^i \binom{m}{i} 2^{\binom{m-i}{2}}.$$

(See [28].) It is easy to show that that M_m/M_{m-1} is a rapidly increasing sequence (in fact $M_m/M_{m-1} \approx 2^{m-1}$ for large m), and hence has a different value for every value of m. Now we count the number of edge covers in G'_{mk}. First note that each edge cover of G'_{mk} induces an edge subset of G'. To extend a subset of G' which leaves i vertices in V and j in U uncovered to an edge cover of G'_{mk}, we must select an edge cover for i copies of K_m, but either an edge cover or a cover missing one fixed vertex for the remaining $n - i$ copies of K_m. Likewise we must select edge covers in j copies of K_k and either edge covers or covers missing one fixed vertex in $3n/2 - j$

copies of K_k. Thus

$$N_0(G'_{mk}) = \sum_{i=0}^{n}\sum_{j=0}^{3n/2} M_m^i (M_m + M_{m-1})^{n-i} M_k^j (M_k + M_{k-1})^{3n/2-j} N_{ij}(G'). \tag{19}$$

$$= M_m^n M_k^{3n/2} \sum_{i=0}^{n} \left(1 + \frac{M_{m-1}}{M_m}\right)^{n-i} \sum_{j=0}^{3n/2} \left(1 + \frac{M_{k-1}}{M_k}\right)^{3n/2-j} N_{ij}(G').$$

By choosing any $5n/2 + 2$ distinct integers greater than Δ_0, assigning $n+1$ of these to be values of m and $3n/2+1$ to be values of k, and using the oracle for $N_0(G'_{mk})$, we can determine all the $N_{ij}(G')$ by interpolation, and hence all the $N_i(G')$. From these, we can determine all the $I_j(G)$, and hence $\sum_{j=1}^{n} I_j(G)$, the total number of independent sets in G. □

B Weak edge colouring is #P-complete

Proof. We use the same notation and construction as in Appendix A. Recall that an edge colouring of a graph is a *weak edge colouring* if there is no vertex whose incident edges all have the same colour (a monochromatic vertex). We prove that counting weak edge colourings is #P-complete, even for the class $\mathcal{G}_{\Delta_0}$. We assume an oracle for counting weak edge colourings of graphs in $\mathcal{G}_{\Delta_0}$. Note that for the class $\mathcal{G}$ vertex q-colouring is #P-complete [16, Theorem 2.2]. Let $N_i(G')$ denote the number of edge colourings of G' with i monochromatic vertices, so $N_0(G')$ is the number of weak edge colourings of G'. Let $N_{ij}(G')$ be the number edge colourings of G' in which i vertices in V and j in U are monochromatic ($i = 0, \ldots, n$, $j = 0, \ldots, 3n/2$). Now observe that $N_{n0}(G')$ is equal to the number of proper *vertex* q-colourings of G, $Q_q(G)$ say. In every colouring counted in $N_{n0}(G')$, every vertex is monochromatic and adjacent vertices receive different colours. Again we attach a copy K_m^v of K_m to each vertex $v \in V$, and a copy K_k^u of K_k to each vertex $u \in U$, to give G_{mk}. Let M_m be the number of weak colourings of K_m, and M'_m the number of edge colourings of K_m with a given monochromatic vertex. Now we have

$$M_m = q^{\binom{m}{2}} + q\sum_{i=1}^{m}(-1)^i \binom{m}{i} q^{\binom{m-i}{2}}, \qquad M'_m = q\sum_{i=0}^{m-1}(-1)^i \binom{m-1}{i} q^{\binom{m-i-1}{2}}.$$

Again the sequence M_m/M'_m increases rapidly ($M_m/M'_m \approx q^{m-2}$ for large m), and takes a different value for every m when $q \geq 2$. Now, as in (19),

$$N_0(G'_{mk}) = M_m^n M_k^{3n/2} \sum_{i=0}^{n} \left(1 + \frac{M'_m}{M_m}\right)^{n-i} \sum_{j=0}^{3n/2} \left(1 + \frac{M'_k}{M_k}\right)^{3n/2-j} N_{ij}(G').$$

Hence, choosing $5n/2 + 2$ distinct integers greater than Δ_0, assigning $n+1$ of these to be values of m and $3n/2+1$ to be values of k, and calling to the oracle, we can determine all the $N_{ij}(G')$ by interpolation. In particular, we can determine $N_{n0}(G') = Q_q(G)$. □

C The probability p_1 is decreasing in m

Proof. If λ is constant, we have

$$p_1(m) = \frac{\lambda}{m-1}\left(1 - \frac{m\lambda^{m-1}}{(1+\lambda)^m - \lambda^m}\right) = \frac{1}{m-1}\left(\lambda - \frac{m}{(1+1/\lambda)^m - 1}\right), \text{ for } m > 1.$$

Let $\alpha = \ln(1 + 1/\lambda) > 0$, so $\lambda = 1/(e^\alpha - 1)$, and let $x = \alpha m > \alpha$. Then

$$p_1(m) = p_1(x/\alpha) = \frac{\alpha}{x-\alpha}\left(\frac{1}{e^\alpha - 1} - \frac{x}{\alpha(e^x - 1)}\right) = \frac{1}{x-\alpha}\left(\frac{\alpha}{e^\alpha - 1} - \frac{x}{e^x - 1}\right), \text{ for } x > \alpha.$$

Since $\alpha > 0$ it is enough to show that $\hat{p}_1(x) = p_1(x/\alpha)$ is decreasing in x. Let $f(x) = x/(e^x - 1)$, so that $\hat{p}_1(x) = \big(f(\alpha) - f(x)\big)/(x - \alpha)$. Then

$$\frac{d}{dx}\ln\hat{p}_1(x) = \frac{-f'(x)}{f(\alpha) - f(x)} - \frac{1}{x-\alpha} \leq 0 \;\text{ if and only if }\; f(\alpha) \geq f(x) + (\alpha - x)f'(x), \tag{20}$$

provided $f(x) < f(\alpha)$. Note that (20) is the subgradient inequality for $f(\alpha)$ at x, and hence follows if $f(x)$ is convex. Thus it is sufficient to show that $f(x)$ is strictly decreasing and convex, so that $f(x) < f(\alpha)$ and $f(\alpha) \geq f(x) + (\alpha - x)f'(x)$. We may prove this as follows.

$$\begin{aligned} f(x) &= \frac{x}{e^x - 1}, \\ \text{so } f'(x) &= \frac{e^x - 1 - xe^x}{(e^x-1)^2} = \frac{e^x(1 - x - e^{-x})}{(e^x-1)^2} < 0 \;\text{ since }\; 1 - x < e^{-x} \;\text{ for all }\; x > 0. \\ \text{Also } f''(x) &= \frac{-x(e^x-1)e^x - 2(e^x - 1 - xe^x)e^x}{(e^x-1)^3} = \frac{e^x}{(e^x-1)^3}\big((x-2)e^x + x + 2\big) > 0, \end{aligned}$$

provided $g(x) = (x-2)e^x + x + 2 > 0$. Note that $g(0) = 0$, and

$$g'(x) = 1 + (x-1)e^x = e^x\big(e^{-x} - (1 - x)\big) > 0, \text{ again using } 1 - x < e^{-x}.$$

Therefore $g(x) > 0$ for all $x > 0$, as required. □