Stopping Times, Metrics and Approximate Counting*

M. Bordewich[1], M. Dyer[2] and M. Karpinski[3]

[1] Durham University, Durham DH1 3LE, UK,
m.j.r.bordewich@durham.ac.uk,
[2] Leeds University, Leeds LS2 9JT, UK,
dyer@comp.leeds.ac.uk,
[3] University of Bonn, 53117 Bonn, Germany,
marek@cs.uni-bonn.de

Abstract. In this paper we examine the importance of the choice of metric in path coupling, and its relationship to *stopping time analysis*. We give strong evidence that stopping time analysis is no more powerful than standard path coupling. In particular, we prove a stronger theorem for path coupling with stopping times, using a metric which allows us to analyse a one-step path coupling. This approach provides insight for the design of better metrics for specific problems. We give illustrative applications to hypergraph independent sets and SAT instances, hypergraph colourings and colourings of bipartite graphs, obtaining improved results for all these problems.

1 Introduction

Markov chain algorithms are an important tool in approximate counting [16]. Coupling has a long history in the theory of Markov chains [8], and can be used to obtain quantitative estimates of convergence times [1]. The idea is to arrange the joint evolution of two arbitrary copies of the chain so that they quickly occupy the same state. For all pairs of states, the coupling must specify a distribution on pairs of states so that both marginals give precisely the transition probabilities of the chain. Good couplings are usually not easy to design, but path coupling [6] has recently proved a useful technique for constructing and analysing them. The idea here is to restrict the design of the coupling to pairs of states which are close in some suitable *metric* on the state space, and then (implicitly) obtain the full coupling by composition of these pairs. For example, for independent sets in a graph or hypergraph, the pairs of interest might be independent sets which differ in one vertex (the *change vertex*) and the metric might be Hamming distance.

The limitations of path coupling analysis are always caused by certain "bad" pairs of states. But these pairs may be very unlikely to occur in a typical realisation of the coupling. Consequently, path coupling has been augmented by other techniques, such as *stopping time* analysis. The stopping time approach is

* To appear in Proc. 33rd ICALP (2006).

applicable when the bad pairs have a reasonable probability of becoming less bad as time proceeds. As an illustration, consider the bad pairs for the Glauber dynamics on hypergraph independent sets [3]. These involve almost fully occupied edges containing the change vertex. However, it seems likely that the number of occupied vertices in these edges will be reduced before we must either increase or decrease the distance between the coupled chains. This observation allows a greatly improved analysis [3]. See [3, 11, 14, 18] for some other applications of this technique. General theorems for applying stopping times appear in [3, 14].

The stopping time approach is a multistep analysis, and appears to give a powerful extension of path coupling. However, in this paper we provide strong evidence that the stopping time approach is no more powerful than single-step path coupling. We observe that, in cases where stopping times can be employed to advantage, equally good or better results can be achieved by using a suitably tailored *metric* in the one-step analysis. The intuition behind the choice of metric is precisely that used in the stopping time approach. We will illustrate this with several examples.

In fact, our first example is a proof of a theorem for path coupling using stopping times, relying on a particular choice of metric which enables us to work with the standard one-step path coupling. The resulting theorem is stronger than those in [3, 14]. The proof implies that all results obtained using stopping times can just as well be obtained using standard path coupling and the right choice of metric. This does not immediately imply that we can abandon the analysis of stopping times. Determining the metric used in our proof involves bounding the expected distance at a stopping time. But it does suggest that it may be better to do a one-step analysis using a metric indicated by the stopping time.

With this insight, we revisit the Glauber dynamics for hypergraph independent sets. Equivalently, these are satisfying assignments of monotone SAT formulas, and this relationship is discussed in the full paper [4]. We also revisit hypergraph colourings, analysed in [3] using stopping times. We find that we are able to obtain stronger results than those obtained in [3], using metrics suggested by stopping time considerations but then optimised. The technical advantage arises mainly from the possibility of using simple linearity of expectation where stopping time analysis uses concentration inequalities and union bounds.

We note that this paper does not contain the first uses of "clever" metrics with path coupling. See [7, 17] for examples. But we do give the first general approach to designing a good metric. While there have been instances in the literature of optimising the *chain* [13, 20], the only previous analysis of which we are aware which uses optimisation of the *metric* appeared in [17].

The organisation of the paper is as follows. In section 2 we prove a better stopping time theorem than was previously known, using only standard path coupling. In section 3 we give our improved results for sampling independent sets in hypergraphs. In section 4 we give improved results for sampling colourings of 3-uniform hypergraphs. Finally, in section 5 we give a completely new application, to the "scan" chain for sampling colourings of bipartite graphs. For even relatively small values of Δ, our results improve Vigoda's [20] celebrated $11\Delta/6$ bound on the number of colours required for rapid mixing.

2 Path coupling and stopping times

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 $\mathcal{M}$, i.e. a coupling defined on a path-generating set $S \subseteq \Omega \times \Omega$. See, for example, [12]. We define T_t, a stopping time for the pair $(X_t, Y_t) \in S$, to be the smallest $t' > t$ such that $\mathrm{d}(X_{t'}, Y_{t'}) \neq \mathrm{d}(X_t, Y_t)$. We will define a new metric d' such that contraction in d at T_t implies contraction in d' at every t' with positive probability $T_t = t'$.

Let $\alpha > 0$ be a constant such that $\mathbb{E}[\mathrm{d}(X_{T_t}, Y_{T_t})] \leq \alpha \mathrm{d}(X_t, Y_t)$ for all $(X_t, Y_t) \in S$. If $\alpha < 1$, then for any $(X_t, Y_t) \in S$, we define d' as follows.

$$\mathrm{d}'(X_t, Y_t) = (1-\alpha)\mathrm{d}(X_t, Y_t) + \mathbb{E}[\mathrm{d}(X_{T_t}, Y_{T_t})] \leq \mathrm{d}(X_t, Y_t). \tag{1}$$

The metric is extended in the usual way to pairs $(X_t, Y_t) \notin S$, using shortest paths. See [12]. We will apply path coupling with the metric d' and the original coupling. First we show a contraction property for this metric.

Lemma 1. *If* $\mathbb{E}[\mathrm{d}(X_{T_t}, Y_{T_t})] \leq \alpha \mathrm{d}(X_t, Y_t) < \mathrm{d}(X_t, Y_t)$ *for all* $(X_t, Y_t) \in S$, *then*

$$\mathbb{E}[\mathrm{d}'(X_k, Y_k) \,|\, X_0, Y_0] \leq \big(1 - (1-\alpha)\Pr(T_0 \leq k)\big)\mathrm{d}'(X_0, Y_0).$$

Proof. We prove this by induction on k. It obviously holds for $k = 0$, since $T_0 > 0$. Using $\mathbb{1}_{\mathcal{A}}$ to denote the 0/1 indicator of event $\mathcal{A}$, we may write (1) as

$$\mathrm{d}'(X_0, Y_0) = (1-\alpha)\mathrm{d}(X_0, Y_0) + \mathbb{E}[\mathrm{d}(X_{T_k}, Y_{T_k})\mathbb{1}_{T_0 > k}] + \mathbb{E}[\mathrm{d}(X_{T_0}, Y_{T_0})\mathbb{1}_{T_0 \leq k}], \tag{2}$$

since if $T_0 > k$ then $T_k = T_0$. Similarly, we have that $\mathbb{E}[\mathrm{d}'(X_k, Y_k)]$

$$\begin{aligned}
&= \mathbb{E}[\mathrm{d}'(X_k, Y_k)\mathbb{1}_{T_0 > k}] + \mathbb{E}[\mathrm{d}'(X_k, Y_k)\mathbb{1}_{T_0 \leq k}] \\
&= (1-\alpha)\mathbb{E}[\mathrm{d}(X_k, Y_k)\mathbb{1}_{T_0 > k}] + \mathbb{E}[\mathrm{d}(X_{T_k}, Y_{T_k})\mathbb{1}_{T_0 > k}] + \mathbb{E}[\mathrm{d}'(X_k, Y_k)\mathbb{1}_{T_0 \leq k}]. \\
&= (1-\alpha)\mathbb{E}[\mathrm{d}(X_0, Y_0)\mathbb{1}_{T_0 > k}] + \mathbb{E}[\mathrm{d}(X_{T_k}, Y_{T_k})\mathbb{1}_{T_0 > k}] + \mathbb{E}[\mathrm{d}'(X_k, Y_k)\mathbb{1}_{T_0 \leq k}]. \quad (3)
\end{aligned}$$

Subtracting (2) from (3), we have that $\mathbb{E}[\mathrm{d}'(X_k, Y_k)] - \mathrm{d}'(X_0, Y_0)$

$$= \ -(1-\alpha)\mathbb{E}[\mathrm{d}(X_0, Y_0)\mathbb{1}_{T_0 \leq k}] + \mathbb{E}[(\mathrm{d}'(X_k, Y_k) - \mathrm{d}(X_{T_0}, Y_{T_0}))\mathbb{1}_{T_0 \leq k}].$$

For $T_0 \leq k$, since $k - T_0 \leq k - 1$ the inductive hypothesis implies $\mathbb{E}[\mathrm{d}'(X_k, Y_k) \,|\, X_{T_0}, Y_{T_0}] \leq \mathrm{d}'(X_{T_0}, Y_{T_0}) \leq \mathrm{d}(X_{T_0}, Y_{T_0})$, (if $(X_k, Y_k) \notin S$ this is implied by linearity). Hence

$$\mathbb{E}[\mathrm{d}'(X_k, Y_k)] - \mathrm{d}'(X_0, Y_0) \leq \ -(1-\alpha)\mathbb{E}[\mathrm{d}(X_0, Y_0)\mathbb{1}_{T_0 \leq k}],$$

But now $\mathbb{E}[\mathrm{d}(X_0, Y_0)\mathbb{1}_{T_0 \leq k}] = \Pr(T_0 \leq k)\mathrm{d}(X_0, Y_0) \geq \Pr(T_0 \leq k)\mathrm{d}'(X_0, Y_0)$.

We may now prove the first version of our main result.

Theorem 1. *Let $\mathcal{M}$ be a Markov chain on state space Ω. Let* d *be an integer valued metric on Ω, and let (X_t, Y_t) be a path coupling for $\mathcal{M}$. Let T_t be the above stopping times. Suppose for all $(X_0, Y_0) \in S$ and for some integer k and $p > 0$, that*

(i) $\Pr[T_0 \leq k] \geq p$,
(ii) $\mathbb{E}[\mathrm{d}(X_{T_0}, Y_{T_0})/\mathrm{d}(X_0, Y_0)] \leq \alpha < 1$.

Then the mixing time $\tau(\varepsilon)$ of $\mathcal{M}$ satisfies $\tau(\varepsilon) \leq \frac{k}{p(1-\alpha)} \ln\left(\frac{eD}{\varepsilon(1-\alpha)}\right)$, where $D = \max\{\mathrm{d}(X, Y) : X, Y \in \Omega\}$.

Proof. From Lemma 1, d' contracts by a factor $1-(1-\alpha)p \leq e^{-(1-\alpha)p}$ for every k steps of $\mathcal{M}$. Note also that $\mathrm{d}' \leq D$. It follows that, at time $\tau(\varepsilon)$, we have

$$\Pr(X_\tau \neq Y_\tau) \leq \mathbb{E}[\mathrm{d}(X_\tau, Y_\tau)] \leq \frac{\mathbb{E}[\mathrm{d}'(X_\tau, Y_\tau)]}{1-\alpha} \leq \frac{De^{-(1-\alpha)p\tau/k}}{1-\alpha} \leq \varepsilon,$$

from which the theorem follows. □

If $1-\alpha$ is small compared to ε, it is possible to do better than this. A proof of the following appears in the full paper [4].

Theorem 2. *If $\mathcal{M}$ satisfies the conditions of Theorem 1, the mixing time $\tau(\varepsilon)$ of $\mathcal{M}$ satisfies $\tau(\varepsilon) \leq \frac{k(2-\alpha)}{p(1-\alpha)} \ln\left(\frac{2eD}{\varepsilon}\right)$, where $D = \max\{\mathrm{d}(X, Y) : X, Y \in \Omega\}$.*

Remark 1. One of the most interesting features of these theorems is that their proofs employ only standard path coupling, but with a metric which has some useful properties. Thus, for any problem to which stopping times might be applied, there exists a metric from which the same result could be obtained using one-step path coupling.

Remark 2. We may compare this stopping time theorem with those in [3, 14]. The main result of [14] (Theorem 3) concerns bounded stopping times, where $T_0 \leq M$ for all $(X_0, Y_0) \in S$, and gives a mixing time of $O(M(1-\alpha)^{-1}\log D)$. By setting $k = M$ and $p = 1$ in Theorem 2, we obtain the same mixing time up to minor changes in constants, but with a proof that does not involve defining a multistep coupling. For unbounded mixing times, [14, Corollary 4] gives a bound $O(\mathbb{E}[T](1-\alpha)^{-2}W \log D)$ by truncating the stopping times, where W denotes the maximum of $\mathrm{d}(X_t, Y_t)$ over all $(X_0, Y_0) \in S$ and $t \leq T$. In most applications $\mathbb{E}[T] \leq k/p$, so we obtain an improvement of order $W(1-\alpha)^{-1}$. By comparison with [3], we obtain a more modest improvement, of order $\log W \log(D(1-\alpha)^{-1})/\log D$.

Remark 3. Further improvements to Theorem 2 seem unlikely, other than in constants. The term k/p must be present, since it bounds a single stopping time. A term $1/(1-\alpha)\log(D/\varepsilon) = \Theta(\log_\alpha(D/\varepsilon))$ also seems essential, since it bounds the number of stopping times required.

3 Hypergraph independent sets

We now turn our attention to hypergraph independent sets. These were previously studied in [3]. 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 $\mathcal{H}$. 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.

1. Select a vertex $v \in \mathcal{V}$ uniformly at random,
2. (i) if $v \in X_t$ let $X_{t+1} = X_t \backslash \{v\}$ with probability $1/2$,
   (ii) if $v \notin X_t$ and $X_t \cup \{v\}$ is independent, let $X_{t+1} = X_t \cup \{v\}$ with probability $1/2$,
   (iii) otherwise let $X_{t+1} = X_t$.

This chain is easily shown to be ergodic with uniform stationary 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. The change in the expected *Hamming distance* between X_t and Y_t after one step could be as high as $\frac{\Delta}{2n} - \frac{1}{n}$, and we obtain rapid mixing only in the case $\Delta = 2$.

For $(\sigma, \sigma \cup \{w\}) \in S$, let $E_i(w, \sigma)$ be the set of edges containing w which have i occupied vertices in σ. Using a result like Theorem 1 above, it is shown in [3] that, for the stopping time T given by the first epoch at which the Hamming distance between the coupled chains changes,

$$\mathbb{E}[\mathrm{d}_{\mathrm{Ham}}(X_T, Y_T | X_0 = \sigma, Y_0 = \sigma \cup \{w\})] \leq 2 \sum_{i=0}^{m-2} p_i |E_i| \leq 2p_1\Delta, \tag{4}$$

where the p_i is the probability that $\mathrm{d}(X_T, Y_T) = 2$ if w is in a single edge with i occupied vertices. Since $p_1 < 1/(m-1)$, we obtain rapid mixing when $2\Delta/(m-1) \leq 1$, i.e. when $m \geq 2\Delta + 1$. See [3] for details.

The approach of section 2 would lead us to define a metric for which the distance between σ and $\sigma \cup \{w\}$ is $(1 - 2p_1\Delta) + 2\sum_{i=0}^{m-2} p_i|E_i|$. By Lemma 1, we know that this metric contracts in expectation. However, prompted by the form of this metric, but retaining the freedom to optimise constants, we will instead define the new metric d to be $\mathrm{d}(\sigma, \sigma \cup \{w\}) = \sum_{i=0}^{m-2} c_i|E_i|$, where $0 < c_i \leq 1$ $(0 \leq i \leq m-2)$ are a nondecreasing sequence of constants to be determined. Using this metric, we obtain the following theorem.

Theorem 3. *Let Δ be fixed, and let $\mathcal{H}$ be a hypergraph such that $m \geq \Delta + 2 \geq 5$, or $\Delta = 3$ and $m \geq 2$. Then the Markov chain $\mathcal{M}(\mathcal{H})$ has mixing time $O(n \log n)$.*

Proof. Without loss of generality, we take $c_{m-2} = 1$ and we will define $c_{-1} = c_0, c_{m-1} \geq \Delta + 1$. Note that c_{-1} has no real role in the analysis, and is chosen only for convenience, but c_{m-1} is chosen so that $c_{m-1} - c_{m-2} \geq \Delta \geq \mathrm{d}(\sigma, \sigma')$

for any pair $(\sigma, \sigma') \in S$. We require $c_i > 0$ for all i so that we will always have $\mathrm{d}(\sigma, \sigma') > 0$ if $\sigma \neq \sigma'$.

Now consider the expected change in distance between σ and $\sigma \cup \{w\}$ after one step of the chain.

If w is chosen, then the distance decreases by $\sum_{i=0}^{m-2} c_i|E_i|$. The contribution to the expected change in distance is $-\frac{2}{2n}\sum_{i=0}^{m-2} c_i|E_i|$.

If we insert a vertex v in an edge containing w, then we increase the distance by $(c_{i+1}-c_i) \geq 0$ for each edge in E_i containing v. This holds for $i = 0, \ldots, m-2$, by the choice of $c_{m-1} = \Delta + 1$. Let U be the set of unoccupied neighbours of w, and $\nu_i(v)$ be the number of edges with i occupants containing w and v. Then

$$\sum_{v\in U} \nu_i(v) = \sum_{v\in U}\sum_{e\in E_i} \mathbb{1}_{v\in e} = \sum_{e\in E_i}\sum_{v\in e\cap U} 1 = \sum_{e\in E_i} (m-i-1) = (m-i-1)|E_i|.$$

$$\text{implies that } \sum_{v\in U}\frac{1}{2n}\sum_{i=0}^{m-2} \nu_i(v)(c_{i+1}-c_i) = \frac{1}{2n}\sum_{i=0}^{m-2}(c_{i+1}-c_i)(m-i-1)|E_i|.$$

If we delete a vertex v in an edge containing w, then we decrease the distance by $(c_i - c_{i-1})$ for each edge in E_i containing v. This holds for $i = 0, \ldots, m-2$, by the choice of c_{-1}. Let O be the set of occupied neighbours of w, and $\nu_i(v)$ be the number of edges with i occupants containing w and v. Then a similar argument gives the contribution as

$$-\sum_{v\in O}\frac{1}{2n}\sum_{i=0}^{m-2} \nu_i(v)(c_i - c_{i-1}) = -\frac{1}{2n}\sum_{i=0}^{m-2}(c_i - c_{i-1})i|E_i|.$$

Let $\mathrm{d}_0 = \mathrm{d}(\sigma, \sigma\cup\{w\})$, and let d_1 be the distance after one step of the chain. The change in expected distance $E' = \mathbb{E}[\mathrm{d}_1 - \mathrm{d}_0]$ satisfies

$$\begin{aligned} 2nE' &\leq -2\sum_{i=0}^{m-2} c_i|E_i| + \sum_{i=0}^{m-2}(c_{i+1}-c_i)(m-i-1)|E_i| - \sum_{i=0}^{m-2}(c_i - c_{i-1})i|E_i| \\ &= \sum_{i=0}^{m-2}\left(ic_{i-1} - (m+1)c_i + (m-i-1)c_{i+1}\right)|E_i|. \end{aligned}$$

We require $\mathbb{E}[\mathrm{d}_1 - \mathrm{d}_0] \leq -\gamma$, for some $\gamma \geq 0$, which holds for all possible choices of E_i if and only if $(m-i-1)c_{i+1} - (m+1)c_i + ic_{i-1} \leq -\gamma$ for all $i = 0, 1, \ldots, m-2$. Thus we need a solution to

$$\begin{gathered} ic_{i-1} - (m+1)c_i + (m-i-1)c_{i+1} \leq -\gamma \qquad (i = 0, \ldots, m-2), \qquad (5) \\ 0 = c_{-1} < c_0 \leq c_1 \leq \cdots \leq c_{m-3} \leq c_{m-2} = 1, \quad c_{m-1} \geq \Delta + 1, \quad \gamma \geq 0, \end{gathered}$$

with $\gamma > 0$ if possible. Solving for the optimal solution gives

$$\begin{aligned} c_i &= \frac{\gamma\sum_{j=0}^{i}\binom{m-1}{j} - \frac{m-\Delta-2+\gamma}{m}\sum_{j=0}^{i}\binom{m}{j}}{\binom{m-1}{i}} \qquad (i = 0, \ldots, m-2), \\ \gamma &= \frac{2^m - 1 - m}{(m-2)2^{m-1} + 1}\left(m - \Delta - 2 + \frac{m(m-1)}{2^m - 1 - m}\right). \end{aligned}$$

Let $f(m) = m - 2 + \frac{m(m-1)}{2^m-1-m}$, then we can have $\gamma \geq 0$ if and only if $f(m) \geq \Delta$, and $\gamma > 0$ if and only if $f(m) > \Delta$.

If $m \geq 5$ then $m(m-1)/(2^m - 1 - m) < 1$, so we will have $f(m) > \Delta$ exactly when $m \geq \Delta + 2$. For smaller values of m, $f(2) = 2$, $f(3) = 2\frac{1}{2}$ and $f(3) = 3\frac{1}{11}$.

The new case here is $\Delta = 3, m \geq 4$. In any case for which $f(m) > \Delta$, standard path coupling arguments yield the mixing times claimed since we have contraction in the metric and the minimum distance is at least c_0. Mixing for $\Delta = 3, m \leq 3$ was shown in [13]), so we have mixing for $\Delta = 3$ and every m. □

Remark 4. The independent set problem here has a natural *dual*, that of sampling an *edge cover* from a hypergraph with edge size Δ and degree m. An edge cover is a subset of $\mathcal{E}$ whose union contains V. For the graph case of this sampling problem, with arbitrary m, see [5]. By duality this gives the case $\Delta = 2$ of the independent set problem here.

4 Colouring 3-uniform hypergraphs

In our second application, also from [3], we consider proper colourings of 3-uniform hypergraphs. We again use Glauber dynamics. Our hypergraph $\mathcal{H}$ will have maximum degree Δ, uniform edge size 3, and we will have a set of q colours. For a discussion of the easier problem of colouring hypergraphs with larger edge size see [3]. 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 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

1. selecting a vertex $v \in \mathcal{V}$ and a colour $k \in \{1, 2, \ldots, q\}$ uniformly at random,
2. let X'_t be the colouring obtained by recolouring v colour k
3. if X'_t is a proper colouring let $X_{t+1} = X'_t$
   otherwise let $X_{t+1} = X_t$.

This chain is easily shown to be ergodic with the uniform stationary distribution. For some large enough constant Δ_0, it was shown in [3] to be rapidly mixing for $q > 1.65\Delta$ and $\Delta > \Delta_0$, using a stopping times analysis. Here we improve this result, and simplify the proof, by using a carefully chosen metric which is prompted by the new insight into stopping times analyses. If w is the change vertex, the intuition in [3] was that edges which contain both colours of w are initially "dangerous" but tend to become less so after a time. Thus our metric will be a function of the numbers of edges containing w with various relevant colourings.

Theorem 4. *Let Δ be fixed, and let $\mathcal{H}$ be a 3-uniform hypergraph of maximum degree Δ. Then if $q \geq \lceil \frac{3}{2}\Delta + 1 \rceil$, the Markov chain $\mathcal{C}(\mathcal{H})$ has mixing time $O(n \log n)$.*

Proof. Consider two proper colourings X and Y differing in a single vertex w. Without loss of generality let the change vertex w be coloured 1 in X and 2 in Y.

We will partition the edges $e \in \mathcal{E}$ containing w into four classes E_1, E_2, E_3, E_4, determined by the colouring of $e \setminus \{w\}$, as follows:

$E_1 : \{1,2\}, \quad E_2 : \bigcup_{i>2}\{1,i\} \cup \{2,i\}, \quad E_3 : \bigcup_{i>2}\{i,i\}, \quad E_4 : \bigcup_{2<i<j}\{i,j\}.$

Instead of Hamming distance, we define a metric d by $\mathrm{d}(X,Y) = \sum_{i=1}^4 c_i|E_i|$, where $1 = c_1 \geq c_2 \geq c_3 \geq c_4 > 0$, and for convenience $c_0 = \Delta + 1$. Note that $\mathrm{d}(X,Y) \leq \Delta$ if X, Y have Hamming distance 1. The diameter is therefore at most Δn in the metric d. Arguing as in Section 3, we have

$$\begin{aligned}
nq\mathbb{E}[\mathrm{d}_1 - \mathrm{d}_0] \leq &-(q-|E_3|)\big(c_1|E_1| + c_2|E_2| + c_3|E_3| + c_4|E_4|\big)\\
&+|E_1|\big(\ -2(q-\Delta-1)(c_1-c_2) + 2(c_0-c_1)\big)\\
&+|E_2|\big(\ -(q-\Delta-2)(c_2-c_4) - (c_2-c_3) + (c_0-c_2) + (c_1-c_2)\big)\\
&+|E_3|\big(\ -2(q-\Delta-2)(c_3-c_4) + 4(c_2-c_3)\big)\\
&+|E_4|\big(\ 2(c_3-c_4) + 4(c_2-c_4)\big).
\end{aligned} \tag{6}$$

If we set $c_1 = 1$,

$$c_2 = \frac{2q-2\Delta+1}{2q-\Delta+1},\ c_3 = c_4 = \frac{2q-3\Delta+1}{2q-\Delta+1},\ \gamma = \frac{2q^2 - q(3\Delta-1) - 4\Delta}{2q-\Delta+1}, \tag{7}$$

then (6) yields

$$\mathbb{E}[\mathrm{d}_1] \leq \ \mathrm{d}_0 - \frac{\gamma\Delta}{nq} \leq \Big(1 - \frac{\gamma}{nq}\Big)\mathrm{d}_0. \tag{8}$$

The condition $\gamma \geq 0$ is equivalent to

$$q \ \geq \ \tfrac{3\Delta-1}{4}\Big(1 + \sqrt{1 + \tfrac{32\Delta}{(3\Delta-1)^2}}\Big), \qquad \text{i.e. } \ q \ \geq \lceil \tfrac{3}{2}\Delta \rceil + 1. \tag{9}$$

Note that we have $c_i > 0$ $(i = 1, \ldots, 4)$ under this condition. Note also that $\gamma > 0$ and hence, using (8), the mixing time satisfies

$$\tau(\varepsilon) \ \leq \ \frac{2q^2 - q\Delta + q}{2q^2 - q(3\Delta-1) - 4\Delta}\ n \ln\Big(\frac{\Delta n}{\varepsilon}\Big).$$

5 Colouring bipartite graphs

Let $G = (V, E)$ be a bipartite graph with bipartition V_1, V_2, and maximum degree Δ. For $v \in V$, let $\mathcal{N}(v) = \{w : \{v,w\} \in E\}$ denote the neighbourhood of v, and let $d(v) = |\mathcal{N}(v)|$ be its degree. Let $Q = [q]$ be a colour set, and $X : V \to Q$ be a colouring of G, not necessarily proper. Let $C_X(v) = \{X(w) : w \in \mathcal{N}(v)\}$ be the set of colours occurring in the neighbourhood of v, and $c_X(v)$ denote the size of $C_X(v)$. We consider the Markov chain Multicolour on colourings of G, which in each step picks one side of the bipartition at random, and then recolours every vertex on that side, followed by recolouring every vertex in the other half of the bipartition. If the state of Multicolour at time t is X_t, the state at time $t+1$ is given by

Multicolour

1. choosing $r \in \{1,2\}$ uniformly at random,
2. for each vertex $v \in V_r$,
   (i) choosing a colour $q(v) \in Q \backslash C_{X_t}(v)$ uniformly at random,
   (ii) setting $X_{t+1}(v) = q(v)$. (Heat bath recolouring)
3. for each vertex $v \in V \backslash V_r$,
   (i) choosing a colour $q(v) \in Q \backslash C_{X_{t+1}}(v)$ uniformly at random,
   (ii) setting $X_{t+1}(v) = q(v)$.

Note that the order in which the vertices are processed in steps 2 and 3 is immaterial, and that in step 3, $C_{X_{t+1}}(v)$ is well defined since all of v's neighbours have been recoloured in step 2. We prove the following theorem.

Theorem 5. *The mixing time of* Multicolour *is* $O(\log(n))$ *for* $q > f(\Delta)$, *where* f *is a function such that*
(1) $f(\Delta) \to \beta\Delta$, *as* $\Delta \to \infty$, *where* β *satisfies* $\beta e^{\beta} = 1$,
(3) $f(\Delta) < \lceil 11\Delta/6 \rceil$ *for* $\Delta \geq 31$.
(2) $f(\Delta) \leq \lceil 11\Delta/6 \rceil$ *for* $\Delta \geq 14$.

This chain is a single-site dynamics intermediate between Glauber and Scan (which uses the same vertex update procedure as Glauber, but choses the vertices in a deterministic order). It is easy to see that it is ergodic if $q > \Delta + 1$, and has equilibrium distribution uniform on all proper colourings of G. Observe also that it uses many fewer random bits than Glauber. Indeed the following easy Corollary of Theorem 5 is proved in the full paper [4].

Corollary 1. *The mixing time for* Scan *is at most that for* Multicolour.

To prove Theorem 5 we need the following lemmas, whose proofs are given in [4].

Lemma 2. *For* $1 \leq i \leq \Delta$ *let* S_i *be a subset of* $(Q - q_0)$ *such that* $m_i = |S_i| \geq q - \Delta$. *Let* s_i *be selected uniformly at random from* S_i, *independently for each* i. *Finally let* $C = \{s_i : 1 \leq i \leq \Delta\}$ *and* $c = |C|$. *Then*

$$\mathbb{E}[q - c \mid s_1 = q_1] \geq 1 + (q-2)\left(1 - \frac{1}{q - \Delta}\right)^{\frac{(\Delta-1)(q-\Delta)}{q-2}} = \alpha.$$

Lemma 3. *For* $1 \leq i \leq \Delta$ *let* S_i *be a subset of* $(Q - q_0)$ *such that* $m_i = |S_i| \geq q - \Delta$. *Let* s_i *be selected uniformly at random from* S_i, *independently for each* i. *Finally let* $C = \{s_i : 1 \leq i \leq \Delta\}$ *and* $c = |C|$. *Then*

$$\mathbb{E}\left[\frac{1}{q-c} \mid s_1 = q_1\right] \leq \frac{1}{\alpha}\left(1 + \frac{(q-\alpha-1)(\alpha-1)}{(q-\Delta)(q-2)\alpha}\right) = \alpha'.$$

Proof (Proof of Theorem 5). In the path coupling setting, we will take S to be the set of pairs of colourings which differ at exactly one vertex. Let v be the change vertex for some pair $(X, Y) \in S$, and assume without loss that $v \in V_1$. The distance between X and Y is defined to be $\mathrm{d}(X, Y) = \sum_{w \in \mathcal{N}(v)} \frac{1}{q - c_{X,Y}(w)}$, where $c_{X,Y}(w)$ is taken to be $\min\{c_X(w), c_Y(w)\}$ in the case that they differ. We couple as follows (the usual path coupling for Glauber dynamics). If we are recolouring a vertex which is not a neighbour of v, then the sets of available colours in X and Y are the same, and we use the same colour in both copies of the chain. If we are recolouring a vertex $w \in \mathcal{N}(v)$ then there are three cases:

1. $|\{X(v), Y(v)\} \cap \{X(z) : z \in \mathcal{N}(w)\backslash\{v\}\}| = 2$.
   Colours $X(v)$ and $Y(v)$ are not available for w in either X or Y, the sets of available colours are the same, and we use the same colour in both X, Y.
2. $|\{X(v), Y(v)\} \cap \{X(z) : z \in \mathcal{N}(w)\backslash\{v\}\}| = 1$.
   Without loss assume $X(v)$ is not available to w in either X or Y, and $Y(v)$ is only available in X. For any colour other than $Y(v)$, we couple the same colour for w in X and Y. For $Y(v)$, we couple recolouring w with $Y(v)$ in X by uniformly recolouring w from the available colours in Y.
3. $|\{X(v), Y(v)\} \cap \{X(z) : z \in \mathcal{N}(w)\backslash\{v\}\}| = 0$.
   Here colour $Y(v)$ is only available in chain X, and $X(v)$ in only available in Y. We couple these colours together, and for each other colour available to both X, Y, we recolour w with the same colour.

In case 1, there is no probability of w being coloured differently in the two chains. In the other cases, the probability of disagreement at w is $\frac{1}{q-c_{X,Y}(w)}$.

Let X', Y' be the colourings after recolouring V_r (half a step of Multicolour) and X'', Y'' be the colourings after the full step of Multicolour. We use primes and double primes to denote the quantities in X' and X'' respectively, corresponding to those in X. If we randomly select V_1 to be recoloured first, then the two copies of the chain have coupled in X' and Y' since the vertices in V_1 have the same set of available colours in each chain.

So suppose that we select V_2 to be recoloured first. The only vertices in V_2 that have different sets of available colours are the neighbours of v. Let $\mathcal{N}(v) = \{w_1, \dots, w_k\}$ and consider the path $W_0, W_1, \dots, W_{k+1}$ from X' to Y', where for $1 \le i \le k$, W_i agrees with X' on all vertices except $w_1, \dots, w_i$ which are coloured as in Y', and $W_0 = X'$ and $W_{k+1} = Y'$. Then for $i \le k$ we have

$$\mathrm{d}(W_{i-1}, W_i) = \mathbb{1}_{w_i} \sum_{z \in \mathcal{N}(w_i)} \frac{1}{q - c_{W_{i-1},W_i}(z)} \le \mathbb{1}_{w_i} \sum_{z \in \mathcal{N}(w_i)} \frac{1}{q - c_{W_i}(z)}, \quad (10)$$

where $\mathbb{1}_{w_i}$ indicates whether X' and Y' differ on w_i.

Note that $\Pr[\mathbb{1}_{w_i} = 1] \le \frac{1}{q-c_{X,Y}(w_i)}$. Furthermore, by the construction of the coupling either conditioning on $\mathbb{1}_{w_i} = 1$ is the same as conditioning that $W_{i-1}(w_i) = q_1$, or that $W_i(w_i) = q_1$, for some q_1. We assume without loss that this is W_i. Then for each $z \in \mathcal{N}(w_i) - v$ the selection of colours in $C_{W_i}(z)$ satisfies the conditions of Lemma 3, since we may take $q_0 = X(z)$ and q_1 as above. For v, there is no colour q_0 which is necessarily unavailable for all its neighbours, since some are coloured as in X' and some as in Y'. Hence we use a slightly weaker bound on α and α', given by

$$\alpha_v = (q-1)\left(1 - \frac{1}{q-\Delta}\right)^{\frac{(\Delta-1)(q-\Delta)}{q-1}} \quad \text{and} \quad \alpha'_v = \frac{1}{\alpha_v}\left(1 + \frac{(q-\alpha_v)(\alpha_v)}{(q-\Delta)(q-1)\alpha_v}\right).$$

Hence for $i \leq k$, $\mathbb{E}[\mathrm{d}(W_{i-1}, W_i)] \leq \frac{1}{q - c_{X,Y}(w_i)}((\Delta - 1)\alpha' + \alpha'_v)$. The value of $\mathrm{d}(W_k, W_{k+1})$ is still $\mathrm{d}(X, Y)$ since the vertices in V_1 have not yet been recoloured.

Now we consider the vertices in V_1. We apply the same analysis as above to each path segment W_{i-1}, W_i, but augment the analysis using the fact that at the time a vertex $z \in V_1$ is recoloured, its neighbours (in V_2) will already have been randomly recoloured. Let the neighbours of w_i be $z_1, z_2, \ldots z_l$, and consider the path $Z_0, Z_1, \ldots Z_{l+1}$, where for $1 \leq j \leq l$, Z_j agrees with W_{i-1} on all vertices except $z_1, \ldots, z_j$ which are coloured as in W_i, and $Z_0 = W_{i-1}$ and $Z_{l+1} = W_i$. Arguing as above, for $j \leq l$ we have

$$\mathrm{d}(Z_{j-1}, Z_j) = \mathbb{1}_{z_j} \sum_{w \in \mathcal{N}(z_j)} \frac{1}{q - c_{Z_{i-1}, Z_i}(w)}.$$

But now $\Pr[\mathbb{1}_{z_j} = 1 |\ W_{i-1}, W_i] \leq \frac{1}{q - c_{W_{i-1}, W_i}(z_j)} \mathbb{1}_{w_i}$. This is similar to equation (10), and the same argument gives $\mathbb{E}[\mathbb{1}_{z_j} = 1] \leq \frac{1}{q - c_{X,Y}(w_i)}\alpha'$, for $z_j \neq v$ and $\mathbb{E}[\mathbb{1}_{z_j} = 1] \leq \frac{1}{q - c_{X,Y}(w_i)}\alpha'_v$ if $z_j = v$. Also, since it depends only on the colouring of V_2, we have $\mathrm{d}(Z_l, Z_{l+1}) = \mathrm{d}(W_{i-1}, W_i)$. So

$$\mathbb{E}[\sum_{j=1}^{l+1} \mathrm{d}(Z_{j-1}, Z_j)] \leq \frac{1}{q - c_{X,Y}(w_i)}((\Delta - 1)\alpha' + \alpha'_v)(((\Delta - 1)\alpha' + \alpha'_v) + 1).$$

Finally note that W_k and W_{k+1} differ only in V_1, so after recolouring V_1 they have coupled. Hence

$$\begin{aligned} \mathbb{E}[\mathrm{d}(X'', Y'')] &= \frac{1}{2} \sum_{i=1}^{k} \sum_{j=1}^{l+1} \mathbb{E}[\mathrm{d}(Z_{j-1}, Z_j)] \\ &\leq \sum_{i=1}^{k} \frac{(\Delta - 1)\alpha' + \alpha'_v)(((\Delta - 1)\alpha' + \alpha'_v) + 1}{2(q - c_{X,Y}(w_i))} \\ &= \mathrm{d}(X, Y)((\Delta - 1)\alpha' + \alpha'_v)\frac{(((\Delta - 1)\alpha' + \alpha'_v) + 1)}{2}. \end{aligned}$$

This gives contraction if $((\Delta - 1)\alpha' + \alpha'_v) < 1$. For large Δ, α' and α'_v both approach $\frac{1}{q} e^{\Delta/q}$. Hence we have contraction when $\frac{\Delta}{q} e^{\Delta/q} < 1$. For small Δ, we can compute the smallest integral q giving contraction (see table). If we have contraction, standard path coupling gives the mixing time bounds claimed. $\square$

Δ	q	$\lceil 11\Delta/6 \rceil$	q/Δ	Δ	q	$\lceil 11\Delta/6 \rceil$	q/Δ
22	40	41	1.82	35	63	65	1.80
23	42	43	1.83	40	72	74	1.80
25	46	46	1.84	50	90	92	1.80
30	55	55	1.83	10000	17634	18334	1.76

Minimum values of q for contraction.

Remark 5. Our analysis shows that one-step analysis of a single-site chain on graph colourings need not break down at $q = 2\Delta$ [15, 19]. This apparent boundary seems merely to be an artefact of using Hamming distance.

References

1. D. Aldous, Random walks on finite groups and rapidly mixing Markov chains, in *Séminaire de Probabilités XVII*, Springer Verlag, Berlin, 1983, pp. 243–297.
2. P. Berman, M. Karpinski & A. D. Scott, Approximation hardness of short symmetric instances of MAX-3SAT, *Elec. Coll. on Comp. Compl.*, ECCC TR03-049, 2003.
3. M. Bordewich, M. Dyer & M. Karpinski, Path coupling using stopping times, *Proc. 15th Int. Symp. on Fundamentals of Computation Theory*, Springer Lecture Notes in Computer Science **3623**, pp. 19–31, 2005. (A full version of the paper appears as `http://arxiv.org/abs/math.PR/0501081`.)
4. M. Bordewich, M. Dyer and M. Karpinski, Metric construction, stopping times and path coupling, `http://arxiv.org/abs/math.PR/0511202`, 2005.
5. R. Bubley & M. Dyer, Graph orientations with no sink and an approximation for a hard case of #SAT, in *Proc. 8th ACM-SIAM Symp. on Discrete Algorithms*, SIAM, 1997, pp. 248–257.
6. R. Bubley & M. Dyer, Path coupling: A technique for proving rapid mixing in Markov chains, in *Proc. 38th IEEE Symp. on Foundations of Computer Science*, IEEE, 1997, pp. 223–231.
7. R. Bubley & M. Dyer, Faster random generation of linear extensions, in *Proc. 9th ACM-SIAM Symp. on Discrete Algorithms*, ACM-SIAM, pp. 350–354, 1998.
8. W. Doeblin, Exposé de la théorie des chaînes simples constantes de Markoff à un nombre fini d'états, *Revue Mathématique de l'Union Interbalkanique* **2** (1938), 77–105.
9. M. Dyer & A. Frieze, Randomly colouring graphs with lower bounds on girth and maximum degree, *Random Structures and Algorithms* **23** (2003), 167–179.
10. M. Dyer, A. Frieze, T. Hayes & E. Vigoda, Randomly coloring constant degree graphs, in *Proc. 45th IEEE Symp. on Foundations of Computer Science*, IEEE, 2004, pp. 582–589.
11. M. Dyer, L. Goldberg, C. Greenhill, M. Jerrum & 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.
12. M. Dyer & C. Greenhill, Random walks on combinatorial objects, in *Surveys in Combinatorics* (J. D. Lamb & D. A. Preece, Eds.), London Math. Soc. Lecture Note Series **267**, Cambridge University Press, Cambridge, 1999, pp. 101–136.
13. M. Dyer & C. Greenhill, On Markov chains for independent sets, *Journal of Algorithms* **35** (2000), 17–49.
14. T. Hayes & E. Vigoda, Variable length path coupling, in *Proc. 15th ACM-SIAM Symp. on Discrete Algorithms*, ACM-SIAM, 2004, pp. 103–110.
15. M. Jerrum, A very simple algorithm for estimating the number of k-colorings of a low-degree graph, *Random Structure & Algorithms* **7** (1995), 157–165.
16. M. Jerrum, *Counting, sampling and integrating: algorithms and complexity*, ETH Zürich Lectures in Mathematics, Birkhäuser, Basel, 2003.
17. M. Luby & E. Vigoda, Fast convergence of the Glauber dynamics for sampling independent sets, *Random Structures & Algorithms* **15** (1999), 229–241.
18. M. Mitzenmacher & E. Niklova, Path coupling as a branching process, 2002.
19. J. Salas & A. Sokal, Absence of phase transition for anti-ferromagnetic Potts models via the Dobrushin uniqueness theorem, *J. Stat. Phys.* **86** (1997), 551–579.
20. E. Vigoda, Improved bounds for sampling colorings, *J. Math. Phys.* **41** (1999), 1555–1569.