An extension of path coupling and its application to the Glauber dynamics for graph colourings *

Martin Dyer†, Leslie Ann Goldberg‡, Catherine Greenhill§,
Mark Jerrum¶ and Michael Mitzenmacher‖

October 24, 2000

Abstract

A new method for analysing the mixing time of Markov chains is described. This method is an extension of path coupling and involves analysing the coupling over multiple steps. The expected behaviour of the coupling at a certain stopping time is used to bound the expected behaviour of the coupling after a fixed number of steps. The new method is applied to analyse the mixing time of the Glauber dynamics for graph colourings. We show that the Glauber dynamics has $O(n \log(n))$ mixing time for triangle-free Δ-regular graphs if k colours are used, where $k \geq (2-\eta)\Delta$, for some small positive constant η. This is the first proof of an optimal upper bound for the mixing time of the Glauber dynamics for some values of k in the range $k \leq 2\Delta$.

1 Introduction

In this paper, a new method for analysing the mixing time of Markov chains is described. This method is a non-trivial extension of path coupling, and applies in situations where

*An earlier version of this paper appeared in the Proceedings of the 11th Annual ACM-SIAM Symposium on Discrete Algorithms, January 2000. This work was supported in part by the EPSRC Research Grant "Sharper Analysis of Randomised Algorithms: a Computational Approach" and by the ESPRIT Projects RAND-APX and ALCOM-FT.

†dyer@scs.leeds.ac.uk, School of Computer Studies, University of Leeds, Leeds LS2 9JT U.K.

‡leslie@dcs.warwick.ac.uk, Department of Computer Science, University of Warwick, Coventry, CV4 7AL, U.K.

§csg@ms.unimelb.edu.au, Department of Mathematics and Statistics, University of Melbourne, Parkville, VIC 3052, AUSTRALIA. Supported by an Australian Research Council Postdoctoral Fellowship.

¶mrj@dcs.ed.ac.uk, Department of Computer Science, University of Edinburgh, Edinburgh ED9 3JZ, U.K.

‖michaelm@eecs.harvard.edu, Computer Science Department, Harvard University, 33 Oxford St, Cambridge MA, 02138 U.S.A. Supported in part by the Alfred P. Sloan Foundation.

path coupling is not enough to prove rapid mixing. We run the path coupling for multiple steps and use the expected behaviour of the coupling at a certain *stopping time* to bound the expected behaviour of the coupling after a fixed number of steps. Standard path coupling is a worst-case analysis, in that it considers the the expected change in the distance between the worst possible pair of states over a single step. However, in a multiple-step analysis, the choice of the initial pair of states is mitigated by the random choices made by the coupling over several steps. Hence, with some constant probability, we are not in the worst case. This is how a multiple-step analysis can improve upon one-step path coupling.

The approach of analyzing the behavior of a Markov chain over several steps has proved worthwhile in other settings. For example, it has been used to prove the stability of randomized bin-packing algorithms [7, 16, 1] and contention resolution protocols [14, 13]. Hence this approach appears to be a natural direction for coupling arguments as well.

Czumaj et al. [8] introduced a framework for multiple-step couplings based on path coupling, which they call *delayed path coupling*. Their "delayed path coupling lemma" [8, Lemma 4.2] (reproduced below as Lemma 2.1) shows how the mixing time of a Markov chain can be bounded above in terms of the behaviour of a coupling over a fixed number of steps. However, the way in which the coupling is analysed over the fixed time interval is not specified, and Czumaj et al. give a few different applications. In some applications, they explicitly construct a non-Markovian coupling over the full time interval. The construction and analysis of such a coupling can be very complicated. However, we use straightforward path coupling to drive our multiple-step coupling, performing most of our analysis at a specially defined stopping time. The next section contains a description of this new method.

We then apply our method to the problem of analysing the mixing time of the Glauber dynamics for graph colourings. A proper k-colouring of a graph $G = (V, E)$ is a labelling of the vertices from a set of colours $C = \{1, \ldots, k\}$ such that no two neighboring vertices have the same colour. We consider the problem of sampling nearly uniformly from the set of all proper k-colourings of a graph of maximum degree Δ. Note that efficiently sampling k-colourings nearly uniformly allows one to approximately count such colourings [15]. This problem is interesting as a fundamental combinatorial problem, and it also relates to several problems in statistical physics; see [15, 23], for more details.

A standard approach to the sampling problem is to design a Markov chain whose stationary distribution is uniform over all proper k-colourings. We can then sample nearly uniformly from all proper k-colourings by running the Markov chain until the distribution of the state is sufficiently near the stationary distribution. For this approach to be efficient, the number of steps for which we must run the Markov chain must be sufficiently small. The number of steps for which we must run the Markov chain is generally called the *mixing time*, and a Markov chain for sampling proper k-colourings is *rapidly mixing* if the mixing time is bounded above by some polynomial in $|V| = n$.

Jerrum [15] (and independently Salas and Sokal [21], using different methods) showed that when $k \geq 2\Delta$, a simple Markov chain is rapidly mixing. This Markov chain is easily

described as follows: choose a vertex v uniformly at random and a colour c uniformly at random; recolour v to colour c if doing so yields a proper colouring. This Markov chain is generally referred to as the Glauber dynamics in the statistical physics literature. Jerrum proved that the Glauber dynamics has $O(n\log(n))$ mixing time for $k > 2\Delta$, while for $k = 2\Delta$ the best known upper bound was $O(n^3)$. We use our new method to show that, for $\Delta \geq 14$, the Glauber dynamics chain has $O(n\log(n))$ mixing time for $k \geq (2-\eta)\Delta$ whenever the graph is triangle-free and Δ-regular, where η is some small, positive constant. It seems to be widely believed that $\Omega(n\log n)$ is a lower bound on the mixing time of the Glauber dynamics; however, we do not know of an existing proof. We present a simple proof of this fact in Theorem 3.1, for the special case of graphs with no edges. Therefore our $O(n\log n)$ bound on the mixing time is optimal. Our main result is the first proof of an optimal upper bound for the mixing time of the Glauber dynamics for some values of k in the range $k \leq 2\Delta$.

The 2Δ barrier has been broken using more complicated chains, but as far as we know this is the first proof that involves direct analysis of the simple Glauber dynamics chain. In [**?**], a rapidly mixing Markov chain was presented for the case $\Delta = 3$, $k = 5$ (and for $\Delta = 4$, $k = 7$ when the graph is triangle-free and 4-regular). The proof involves the analysis of several (in the hundreds for the $\Delta = 3$ case) linear programming problems related to the chain. Using a comparison technique such as [9] one can conclude that the Glauber dynamics is also rapidly mixing for these values of k, Δ. However, applying a comparison technique generally increases the upper bound on the mixing time by several factors of n.

In recent work, Vigoda [23] has proven that $k \geq 11\Delta/6$ is sufficient for rapid mixing, using an entirely different Markov chain (similar to the well-known Swendsen-Wang algorithm [22]). Again, his result implies rapid mixing of the Glauber dynamics for $k \geq 11\Delta/6$, but with an $O(n^2\log n)$ bound on the mixing time. His result clearly dominates ours in terms of the range of k for which rapid mixing is established. However, because our analysis is based directly on the Glauber dynamics chain and achieves an optimal bound, and because we use a new technique based on analysing this chain over multiple steps, our result is of independent interest.

2 Path coupling using stopping times

Before describing the new method we present some standard definitions and notation. 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. 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

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

Following Aldous [3], let $\tau_x(\varepsilon)$ denote the least value T such that $\mathrm{d_{TV}}(P_x^t,\, \pi) \leq \varepsilon$ for all $t \geq T$. 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 be 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 of Ω. Throughout this paper all logarithms are to base e.

There are relatively few methods available to prove that a Markov chain is rapidly mixing. One such method 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), (Y_t), considered marginally, is a faithful copy of $\mathcal{M}$. The moves of the coupling are correlated to encourage the two copies of the Markov chain to *couple*: i.e. to achieve $X_t = Y_t$. This gives a bound on the total variation distance using the *Coupling Lemma* (see for example, Aldous [3]), which states that

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

where $X_0 = x$ and Y_0 is drawn from the stationary distribution π. The following standard result is used to obtain an upper bound on this probability and hence an upper bound for the mixing time (the proof is omitted).

Theorem 2.1 *Let (X_t, Y_t) be a coupling for the Markov chain $\mathcal{M}$ and let ρ be any integer valued metric defined on $\Omega \times \Omega$. Suppose that there exists $\beta \leq 1$ such that $\mathbf{E}[\rho(X_{t+1}, Y_{t+1})] \leq \beta\, \rho(X_t, Y_t)$ for all t, and all $(X_t, Y_t) \in \Omega \times \Omega$. Let D be the maximum value that ρ achieves on $\Omega \times \Omega$. If $\beta < 1$ then the mixing time $\tau(\varepsilon)$ of $\mathcal{M}$ satisfies $\tau(\varepsilon) \leq \log(D\varepsilon^{-1})/(1-\beta)$. If $\beta = 1$ and there exists $\alpha > 0$ such that*

$$\mathrm{Prob}[\rho(X_{t+1}, Y_{t+1}) \neq \rho(X_t, Y_t)] \geq \alpha$$

for all t, and all $(X_t, Y_t) \in \Omega \times \Omega$, then $\tau(\varepsilon) \leq \lceil eD^2/\alpha \rceil \lceil \log(\varepsilon^{-1}) \rceil$.

From now on, assume that all couplings are Markovian unless explicitly stated. The path coupling method, introduced in [4], is a variation of traditional coupling which allows us to restrict our attention to a certain subset S of $\Omega \times \Omega$, where Ω is the state space of a given Markov chain. If we view S as a relation, the transitive closure of S must equal Ω. The rate of convergence of the chain is measured with respect to a (quasi)metric ρ on $\Omega \times \Omega$, which can be defined by lifting a proximity function on S to the whole of $\Omega \times \Omega$ (see [12] for details).

In this section we present a modification of path coupling which involves stopping times. Let (X, Y) be any element of $\Omega \times \Omega$. As for ordinary path coupling, we define a path, or sequence

$$X = Z_0,\ Z_1, \ldots, Z_r = Y$$

between X and Y, where $(Z_\ell, Z_{\ell+1}) \in S$ for $0 \leq \ell < r$, and

$$\sum_{\ell=0}^{r-1} \rho(Z_\ell, Z_{\ell+1}) = \rho(X, Y).$$

In ordinary path coupling we allow the coupling to evolve for *one* step, giving a new path

$$Z_0',\ Z_1', \ldots, Z_r'$$

(for a precise definition of the probability distribution of this new path, see [12]). We then define $(X', Y') = (Z_0', Z_r')$. The path coupling lemma says the following. Let $(X, Y) \mapsto (X', Y')$ be a coupling defined on all pairs in S. Suppose there exists a constant β such that $0 < \beta \leq 1$ and for all $(X, Y) \in S$ we have

$$\mathbf{E}\left[\rho(X', Y')\right] \leq \beta\, \rho(X, Y). \tag{1}$$

Then we can conclude that (1) holds for all $(X, Y) \in \Omega \times \Omega$, and apply Theorem 2.1. Suppose however that the smallest value of β for which (1) holds for all $(X, Y) \in S$ satisfies $\beta > 1$. Then path coupling is not good enough to allow us to apply Theorem 2.1. However, if β is not much larger than 1, and there are some "good" initial pairs $(X, Y) \in S$ where the distance *decreases* after one step (in expected value), then we can try the following approach.

The following lemma is the "delayed path coupling lemma" [8, Lemma 4.2] of Czumaj et al., which shows how the mixing time of a Markov chain may be related to the behaviour of a t-step path coupling (which may be non-Markovian). For completeness, we present a proof.

Lemma 2.1 *Let $S \subseteq \Omega \times \Omega$ be such that the transitive closure of S is the whole of $\Omega \times \Omega$. Let ρ be an integer-valued metric on $\Omega \times \Omega$ which takes values in $\{0, \ldots, D\}$. Given $(X_0, Y_0) \in S$, let (X_0, Y_0), $(X_1, Y_1), \ldots, (X_t, Y_t)$ be the t-step evolution of a (possibly non-Markovian) coupling starting from (X_0, Y_0). Suppose that there exists a constant γ such that $0 < \gamma < 1$ and*

$$\mathbf{E}\left[\rho(X_t, Y_t)\right] \leq \gamma \rho(X_0, Y_0) \tag{2}$$

for all $(X_0, Y_0) \in S$. Then the mixing time $\tau(\varepsilon)$ of $\mathcal{M}$ satisfies

$$\tau(\varepsilon) \leq \frac{\log(D\varepsilon^{-1})}{1 - \gamma} \cdot t.$$

Proof. Using the same argument as the path coupling lemma, we know that (2) holds for all $(X_0, Y_0) \in \Omega \times \Omega$. Run the coupling in epochs of length t. After r epochs, we have

$$\mathbf{E}\left[\rho(X_{rt}, Y_{rt})\right] \leq \gamma^r\, \rho(X_0, Y_0) \leq \gamma^r\, D.$$

If $r \geq \log(D\varepsilon^{-1})/(1 - \gamma)$ then $\mathbf{E}\left[\rho(X_{rt}, Y_{rt})\right] \leq \varepsilon$. This gives an upper bound for the number of epochs required to ensure that the distribution of the chain is at most ε away from stationarity, in terms of total variation distance. Multiplying this number by t, the number of steps per epoch, gives the mixing time of the chain. □

Therefore it suffices to show that $\mathbf{E}\left[\rho(X_t, Y_t)\right] \leq \gamma\, \rho(X_0, Y_0)$ for all $(X_0, Y_0) \in S$, where γ is some positive constant less than 1. The main contribution of this paper is to provide a new approach to bounding $\mathbf{E}\left[\rho(X_t, Y_t)\right]$, which we now describe. Let $(X, Y) \mapsto (X', Y')$ be a (one-step, Markovian) coupling for $\mathcal{M}$ defined on all initial pairs in S; that is, $(X, Y) \in S$ and $(X', Y') \in \Omega \times \Omega$. We will apply this coupling for t

steps, using the path coupling machinery to drive the coupling if the trajectory of the coupling leaves the set S. This gives a multiple-step coupling $\{(X_s, Y_s)\}_{s \geq 0}$. Let T be a stopping time for this coupling, defined in such a way that

$$\rho(X_s, Y_s) = \rho(X_0, Y_0) \text{ for } 0 \leq s < T.$$

Then T is a random variable which depends only on the history of the coupling up to the present time. For example, we could define T to be the first time at which the value of ρ changes.

If $T > t$ then we know that $\rho(X_t, Y_t) = \rho(X_0, Y_0)$. Otherwise, we consider (X_T, Y_T), the state of the coupling at the stopping time T. (The pair (X_T, Y_T) need no longer belong to the set S, but the path coupling machinery drives the coupling for all pairs in $\Omega \times \Omega$.) The analysis gives an upper bound for the quantity

$$\mathbf{E}\left[\rho(X_T, Y_T) \mid T \leq t\right].$$

We hope that this quantity will be smaller than $\mathbf{E}\left[\rho(X_1, Y_1)\right]$, with the following heuristic justification. The analysis of one-step coupling is a worst-case analysis. However, after running the Markov chain for T steps, the effect of the chosen starting state is mitigated to some extent by the random choices made during the running of the coupling. In other words, with some positive probability we are not in the worst case. It is here that we can improve on one-step coupling.

We now show how to relate $\mathbf{E}\left[\rho(X_T, Y_T) \mid T \leq t\right]$ and $\mathbf{E}\left[\rho(X_t, Y_t)\right]$.

Theorem 2.2 *Let $\mathcal{M}$ be a Markov chain with state space Ω. Let ρ be a metric on $\Omega \times \Omega$ and let S be some subset of $\Omega \times \Omega$ such that the transitive closure of S is $\Omega \times \Omega$. Suppose that we have a (one-step, Markovian) coupling $(X, Y) \mapsto (X', Y')$, defined on pairs in S such that*

$$\mathbf{E}\left[\rho(X', Y')\right] \leq \beta\, \rho(X, Y)$$

for some constant β such that $\beta \geq 1$. Let $t > 0$ be a fixed integer. Apply the coupling for t steps from initial state $(X_0, Y_0) \in S$, using the path coupling lemma. Let T be some stopping time for $\{(X_s, Y_s)\}_{s \geq 0}$ such that

$$\rho(X_s, Y_s) = \rho(X_0, Y_0)$$

whenenever $0 \leq s < T$. Then

$$\mathbf{E}\left[\rho(X_t, Y_t)\right] \leq \text{Prob}\left[T > t\right] \cdot \rho(X_0, Y_0) + \text{Prob}\left[T \leq t\right] \cdot \beta^t \cdot \mathbf{E}\left[\rho(X_T, Y_T) \mid T \leq t\right]$$

for all $(X_0, Y_0) \in S$.

Proof. The coupling defined on the set S gives rise to a coupling $(X, Y) \mapsto (X', Y')$ on the entire set $\Omega \times \Omega$ such that $\mathbf{E}\left[\rho(X', Y')\right] \leq \beta\, \rho(X, Y)$ for all $(X, Y) \in \Omega \times \Omega$, by the path coupling lemma [4]. Let (X_0, Y_0), $(X_1, Y_1), \ldots, (X_t, Y_t)$ be the t-step evolution of this coupling from the starting state $(X_0, Y_0) \in S$.

If $T > t$ then $\rho(X_t, Y_t) = \rho(X_0, Y_0)$. Next suppose that $T \leq t$. Then

$$\begin{aligned}
\mathbf{E}\left[\rho(X_t, Y_t) \mid T \leq t\right] &\leq \beta\, \mathbf{E}\left[\rho(X_{t-1}, Y_{t-1}) \mid T \leq t\right] \\
&\leq \mathbf{E}\left[\beta^{t-T} \rho(X_T, Y_T) \mid T \leq t\right] \\
&\leq \beta^t\, \mathbf{E}\left[\rho(X_T, Y_T) \mid T \leq t\right].
\end{aligned}$$

(By replacing $t - T$ by t we are, in effect, assuming that the stopping time occurs at the very beginning of the interval.) This proves the theorem. □

Suppose that S is the set of all pairs (X, Y) with $\rho(X, Y) = 1$. In this case, Theorem 2.2 can be rewritten to assert that

$$\mathbf{E}\left[\rho(X_t, Y_t) - 1\right] \leq \mathrm{Prob}\left[T \leq t\right] \left(\beta^t \cdot \mathbf{E}\left[\rho(X_T, Y_T) \mid T \leq t\right] - 1\right).$$

Combining Lemma 2.1 and Theorem 2.2, we see that γ can be defined to be the maximum of the values

$$1 - \mathrm{Prob}\left[T \leq t\right] \left(1 - \beta^t \cdot \mathbf{E}\left[\rho(X_T, Y_T) \mid T \leq t\right]\right) \tag{3}$$

over all $(X_0, Y_0) \in S$. In order to obtain a good bound on the mixing time of the chain, we aim to show that $\gamma < 1$. Clearly $\gamma < 1$ if

$$\beta^t\, \mathbf{E}\left[\rho(X_T, Y_T) \mid T \leq t\right] < 1$$

for all $(X_0, Y_0) \in S$.

3 Applying the new method to the Glauber dynamics for graph colourings

In this section we illustrate the new method by using it to analyse the mixing time of the Glauber dynamics for graph colourings.

Let $G = (V, E)$ be a given graph and let $\Omega_k(G)$ be the set of all proper k-colourings of G, where C is the set of colours. The Glauber dynamics is a Markov chain on $\Omega_k(G)$ with transitions from the current state according to the following procedure:

- choose $(v, i) \in V \times C$ uniformly at random,
- recolour v with i if this results in v being properly recoloured.

This chain was analysed by Jerrum [15] and independently by Salas and Sokal [21]. They proved that the chain is rapidly mixing for graphs with maximum degree Δ when $k > 2\Delta$. The fact that the chain is also rapidly mixing for $k = 2\Delta$ can be found in [4]. Jerrum showed that the Glauber dynamics has $O(n \log(n))$ mixing time for $k > 2\Delta$, and the best known upper bound when $k = 2\Delta$ was $O(n^3)$.

In Section 3.1 we describe the standard path coupling for this chain. Section 3.2 contains the definition of the stopping time for this coupling, and gives a necessary

condition for the success of the new method. In Section 3.3 we perform the calculations needed to establish the necessary condition. All calculations are combined in Section 3.4 to provide an $O(n\log(n))$ upper bound for the mixing time of the Glauber dynamics for Δ-regular, triangle-free graphs, when $(2-\eta)\Delta \leq k \leq 2\Delta$, where η is a small positive constant.

Before we proceed, we present a proof of the "folklore" result that the mixing time of the Glauber dynamics is bounded below by $\Omega(n\log n)$. Our proof concerns graphs with no edges.

Theorem 3.1 *Let G be the empty graph with n vertices, and let $k \geq 2$. Then*

$$\tau((2e)^{-1}) = \Omega(n\log n).$$

Proof. A *stopping rule* Γ (see [17]) is a map that associates every initial sequence w of Markov chain states with a number $\Gamma[w] \in [0,1]$, which is taken to be the probability that the sequence should continue. We can also think of Γ as a random variable taking values in $\{0,1,2,\ldots\}$, whose distribution only depends on $w_0,\ldots,w_\Gamma$ (and w_Γ is the state where we stop). If w_0 is drawn from the distribution σ and $\mathbf{E}[\Gamma]$ is finite, and the distribution of final states is τ, then the rule is called a stopping rule from σ to τ. It is said to be *optimal* for σ and τ if $\mathbf{E}[\Gamma]$ is minimal. For each $x \in \Omega_k(G)$ let σ_x be the distribution concentrated on the state x. Define τ_2 to be the maximum, over all initial states x, of the expected length of an optimal stopping rule from σ_x to π. Since the Glauber dynamics is time-reversible, a result of Aldous' [2, Lemma 12] applies, showing that

$$\tau((2e)^{-1}) \geq c\tau_2$$

where $c = (1-e^{-1})^2/2$. Now let Γ be the stopping rule which says "stop when you have visited every vertex of G at least once". (It may not be immediately apparent that this rule satisfies the definition of a stopping rule given in [17], since it uses information not encoded in the states of the chain. However, it is routine to formulate an equivalent randomized stopping rule which does fit the definition, see [18, p.89].) Since G has no edges and every vertex has been randomly recoloured, the colouring obtained at time Γ is distributed according to π. Hence Γ is a stopping rule from σ_x to π, for all $x \in \Omega_k(G)$. Let $y \in \Omega_k(G)$ be any colouring of G such that $y(v) \neq x(v)$ for all $v \in V$. Then y is a halting state for this stopping rule (that is, the probability that the process will halt if it reaches y is 1). Since Γ has a halting state it is an optimal stopping rule, using [17, Theorem 5.1]. This shows that $\tau_2 = \mathbf{E}[\Gamma]$. Therefore $\tau((2e)^{-1})$ is bounded below by a constant times the expected number of steps required to visit every vertex at least once, and the latter is $\Theta(n\log n)$ by the well-known coupon collector's lemma (see, for example [19, Section 3.6]). □

3.1 Path coupling for the Glauber dynamics

We now give the standard path coupling analysis of the Glauber dynamics. The proximity function is given by Hamming distance, and we let $\mathcal{S}$ be the set of all pairs with Hamming distance 1. The state space of the Markov chain must be extended to the set of all colourings (including non-proper colourings), in order to be able to form a path of length $H(X,Y)$ between any two colourings $(X,Y) \in \Omega_k(G)$. (This approach is standard, and does not cause any problems, since the non-proper colourings are transient states. The stationary distribution is uniform over all proper colourings, and zero elsewhere. Although the extended chain is no longer reversible, the path coupling lemma still applies. Moreover, the mixing time of the chain on the original state space is bounded above by the mixing time of the chain on the extended state space.)

Consider $(X,Y) \in \mathcal{S}$, so X and Y differ just at a single vertex v. Let $N(v)$ denote the set of neighbours of v in G. We can couple at (X,Y) as follows: choose (u,i) uniformly at random from $V \times C$. If $u = v$ then attempt to recolour v with i in both X and Y. This will either succeed in both, or fail in both. If it succeeds then the Hamming distance decreases by 1. The only other moves which can affect the Hamming distance are when $u = w$ where $w \in N(v)$. In this case, if $i \notin \{X(v), Y(v)\}$ then attempt to recolour w with i in both X and Y. This will either succeed in both or fail in both, and the Hamming distance is unaffected. If $i = X(v)$ then attempt to recolour w with $X(v)$ in X and attempt to recolour w with $Y(v)$ in Y. This will fail in both X and Y, so the Hamming distance is unaffected. Finally, if $i = Y(v)$ then attempt to recolour w with $Y(v)$ in X, and attempt to recolour w with $X(v)$ in Y. This may succeed or fail in either, so the Hamming distance could increase by 1 here. Thus the expected change in the Hamming distance is at most

$$-\frac{(k - |\{X(w) : w \in N(v)\}|)}{kn} + \frac{\Delta}{kn}.$$

In general, we have $|\{X(w) : w \in N(v)\}| \leq \Delta$, so that the expected change in the Hamming distance is at most $-(k - 2\Delta)/(kn)$. This gives nonincreasing Hamming distance for $k \geq 2\Delta$. The aim of the new approach is to show that, with constant positive probability, there are fewer than Δ distinct colours around v, just before the stopping time. This gives nonincreasing Hamming distance for a wider range of k.

3.2 A stopping time for the Glauber dynamics on colourings

For simplicity, assume that the given graph G is Δ-regular and triangle-free. Let η be a small positive constant which we fix later, and suppose that $(2-\eta)\Delta \leq k \leq 2\Delta$. We analyse the mixing time of the Glauber dynamics using our new method, to show that the Glauber dynamics has $O(n \log(n))$ mixing time for this range of k.

Let $(X_0, Y_0) \in S$ be given, so that X_0, Y_0 differ just at a single vertex $v \in V$. Perform the coupling described in Section 3.1 with starting point (X_0, Y_0). Let $Q(X_0, Y_0)$ be the set of all moves which involve v or increase the Hamming distance; that is,

$$Q(X_0, Y_0) = \{(v,i) : i \in C\} \cup \{(w, Y_0(v)) : w \in N(v)\}.$$

Then $Q(X_0, Y_0)$ contains all the choices which may affect the Hamming distance, but also some which will not. Define the random variable T to be the first step at which a pair in $Q(X_0, Y_0)$ is chosen by the coupling. Then T is a stopping time since it depends only on the coupling up to the present time. Now (X_T, Y_T) is the state of the coupling after the Tth step, which we refer to as the state of the coupling at the stopping time. Note that $H(X_s, Y_s) = H(X_0, Y_0) = 1$ for $0 \leq s < T$, by the analysis of Section 3.1. Clearly $|Q(X_0, Y_0)| = k + \Delta$ for all pairs $(X_0, Y_0) \in \mathcal{S}$. Let δ be a positive constant, and assume that δn is an integer. (Since n can grow arbitrarily large, there is not much harm in making this assumption.) An (approximately) optimal value of δ will be fixed later, which will satisfy $\delta < (2 - \eta)/3$. We run the coupling for t steps, where $t = \delta n$.

Let $\mathcal{C}$ be a random variable which denotes the number of colours which occur more than once around v *just before* the stopping time T (that is, after step $T - 1$). In the next section we prove that, when n and Δ are "big enough" and η is "small enough", we have

$$\mathbf{E}\left[\mathcal{C} \mid T \leq \delta n\right] \geq \xi\Delta$$

for some constant ξ such that $\xi \geq 2\eta$. We now show why this is sufficient.

The arguments of Section 3.1 show that the expected value of the Hamming distance after one step of normal path coupling from $(X, Y) \in S$ is at most

$$1 - \frac{k - 2\Delta}{kn} \leq 1 + \frac{\eta\Delta}{kn},$$

since $(2 - \eta)\Delta \leq k \leq 2\Delta$. Next, notice that

$$\begin{aligned}
\mathbf{E}\left[H(X_T, Y_T) - 1 \mid T \leq \delta n\right] &\leq -\frac{k - (\Delta - \mathbf{E}\left[\mathcal{C} \mid T \leq \delta n\right])}{k + \Delta} + \frac{\Delta}{k + \Delta} \\
&\leq -\frac{k - (1 - 2\eta)\Delta}{k + \Delta} + \frac{\Delta}{k + \Delta} \\
&\leq -\frac{\eta}{3}.
\end{aligned}$$

Therefore, using Theorem 2.2 (and in particular the remarks following the theorem),

$$\begin{aligned}
\mathbf{E}\left[H(X_{\delta n}, Y_{\delta n}) - 1\right] &\leq \mathrm{Prob}\left[T \leq \delta n\right]\left(\beta^{\delta n} \cdot \mathbf{E}\left[H(X_T, Y_T) \mid T \leq \delta n\right] - 1\right) \\
&\leq \mathrm{Prob}\left[T \leq \delta n\right]\left(\left(1 + \frac{\eta\Delta}{kn}\right)^{\delta n}\left(1 - \frac{\eta}{3}\right) - 1\right) \\
&\leq \mathrm{Prob}\left[T \leq \delta n\right]\left(e^{\eta\delta/(2-\eta)}e^{-\eta/3} - 1\right). \qquad (4)
\end{aligned}$$

This quantity is nonpositive whenever

$$\frac{\eta\delta}{2 - \eta} - \frac{\eta}{3} \leq 0,$$

and this holds for $\delta \leq (2 - \eta)/3$.

We now calculate a lower bound for $\mathbf{E}\left[T \mid T \leq \delta n\right]$, which is needed in Section 3.3.

Lemma 3.1 *Suppose that* $n \geq \delta^{-1}$ *and* $(2-\eta)\Delta \leq k \leq 2\Delta$, *where* $0 < \eta < 2$. *Let* $\theta = 3/(2-\eta)$. *Then*

$$\mathbf{E}\left[T \mid T \leq \delta n\right] \geq \frac{\delta n}{2}(1-\theta\delta).$$

Proof. Let $q = 1-(k+\Delta)/(kn)$, and let p_s denote the probability that $T = s$. Then $p_s = \text{Prob}\left[T = s\right] = (1-q)q^{s-1}$ and $\text{Prob}\left[T \leq \delta n\right] = 1-q^{\delta n}$. Now $q^{\delta n} \geq 1-(k+\Delta)\delta/k$ since $n \geq \delta^{-1}$. Therefore

$$\begin{aligned}
\mathbf{E}\left[T \mid T \leq \delta n\right] &= \left(1-q^{\delta n}\right)^{-1} \sum_{s=0}^{\delta n} s p_s \\
&\geq \frac{p_{\delta n}\, \delta^2\, n^2}{2\left(1-q^{\delta n}\right)} \\
&> \frac{(1-q)\, q^{\delta n}\, \delta^2\, n^2}{2\left(1-q^{\delta n}\right)} \\
&\geq \frac{(1-q)\left(1-\frac{k+\Delta}{k}\delta\right)}{2\frac{k+\Delta}{k}\delta}\, \delta^2\, n^2 \\
&\geq (1-\theta\delta)\,\frac{\delta n}{2},
\end{aligned}$$

as claimed. □

3.3 The expected number of repeated colours just before the stopping time

Let (X_0, Y_0) be a given pair in $\mathcal{S}$ and let v be the vertex which is coloured differently in X and Y. Let T be the stopping time for the coupling when started at (X_0, Y_0). Denote by $\mathcal{C}$ the number of colours which occur at least twice around v just before the stopping time T. That is,

$$\mathcal{C} = |\left\{i \in C : |\left\{w \in N(v) : X_{T-1}(w) = i\right\}| \geq 2\right\}|.$$

In this section we obtain a lower bound for $\mathbf{E}\left[\mathcal{C} \mid T \leq \delta n\right]$ which holds when Δ and n are both "large enough" and η is "small enough". Specifically, take $\Delta \geq 14$, $n \geq 120$ and $\eta < 1/210$.

Let A_w be defined by

$$A_w = C \setminus \left(\left\{X_0(u) : \{u,w\} \in E\right\} \cup \left\{Y_0(v)\right\}\right)$$

for $w \in N(v)$. Then A_w is the set of colours which are acceptable at w in both X_0 and Y_0. Note that $|A_w| \geq k-\Delta-1$ for all $w \in N(v)$. Next, let

$$B_i = \left\{w \in N(v) : i \in A_w\right\}$$

and let $b_i = |B_i|$ for each $i \in C$. So B_i is the set of vertices $w \in N(v)$ at which i is acceptable in both X_0 and Y_0.

Lemma 3.2 *Assume that $\eta < 1/210$ and $\Delta \geq 14$. Let k satisfy $(2-\eta)\Delta \leq k \leq 2\Delta$. Then there are at least $\lceil k/5 \rceil$ colours i such that $b_i \geq \Delta/3$.*

Proof. Let $Z = |\{(i,w) : i \in A_w\}|$. Now $Z \geq \Delta(k - \Delta - 1)$. For a contradiction, suppose that fewer than $\lceil k/5 \rceil$ colours i have $b_i \geq \Delta/3$. If k is a multiple of 5 then

$$\begin{aligned} Z &\leq \left(\frac{k}{5} - 1\right)\Delta + \left(\frac{4k}{5} + 1\right)\frac{\Delta}{3} \\ &\leq \left(1 - \left(\frac{1}{15} - \frac{1}{3\Delta}\right)\right)\Delta^2 - \Delta \\ &< \Delta(k - \Delta - 1), \end{aligned}$$

giving the desired contradiction. Next, suppose that $k = 5\ell + r$ where $r \in \{1,2,3,4\}$. Then

$$\begin{aligned} Z &\leq \ell\Delta + (k-\ell)\frac{\Delta}{3} \\ &\leq \left(1 - \left(\frac{1}{15} - \frac{15-2r}{15\Delta}\right)\right)\Delta^2 - \Delta \\ &< \Delta(k - \Delta - 1), \end{aligned}$$

since $\eta < 1/210$ and $\Delta \geq 14$. Again, this is a contradiction. □

Using this information we can prove a lower bound for the expected number of repeated colours around v just before the stopping time, given that the stopping time occurs in the first δn steps.

Theorem 3.2 *Suppose that $n \geq 120$, $\Delta \geq 14$, $\eta < 1/210$ and $\delta < (2-\eta)/3$. Also assume that $(2-\eta)\Delta \leq k \leq 2\Delta$. Then*

$$\mathbf{E}\left[\mathcal{C} \mid T \leq \delta n\right] \geq \frac{1}{3840} \cdot \delta^2 \left(1 - \theta\delta\right)^2 \cdot e^{-4\delta} \cdot \Delta,$$

where $\theta = 3/(2-\eta)$.

Proof. By Lemma 3.2, there are at least $\lceil k/5 \rceil$ colours i such that $b_i \geq \Delta/3$. Consider ways in which such a colour i can occur at least twice around v just before the stopping time T. One way in which this can occur is as follows. Suppose that there are exactly two distinct elements $u, w \in B_i$ which were chosen with the colour i during the coupling. That is, (u,i) and (w,i) were both chosen but (q,i) was not chosen for any $q \in B_i \backslash \{u,w\}$. Also suppose that u and w are never chosen at any other time, with any colour, and that no neighbour of u or w is ever chosen with colour i. In this situation, both u and

w end up coloured i. We now analyse the probability that this event occurs, for a given value of T.

We know that T is the first stopping time, so there are $T-1$ steps before the stopping time step. We do not have kn possible choices at each of these $T-1$ steps, but rather $kn-(k+\Delta)$ possibilities. With this in mind, the probability that, say, w is chosen with colour i is given by $1/(kn-(k+\Delta)) \geq 1/(kn)$. There are at least $\Delta^2/24$ choices for the unordered pair $\{u,w\} \subseteq B_i$, since $b_i \geq \Delta/3$ and $\Delta \geq 14$. The probability that both (u,i) and (w,i) are chosen at two distinct times in the first $T-1$ steps is at least $\binom{T-1}{2} \cdot 1/(k^2n^2)$. There are also choices which we have ruled out for all other steps, corresponding to the vertex-colour pairs from the set

$$\{(q,i) : q \in (N(u) \cup N(w) \setminus \{v\}) \cup (B_i \setminus \{u,w\})\} \bigcup \{(u,j),\ (w,j) : j \in C \setminus \{Y(v)\}\}$$

(note that the selection of $j = Y(v)$ is ruled out because s is not a stopping time for $0 \leq s < T$). We have ruled out at most $3\Delta + 2k - 6$ choices at each of $T-3$ steps. Thus we see that

$$\text{Prob}\left[i \text{ is repeated} \mid T,\, b_i \geq \frac{\Delta}{3}\right] \geq \frac{\Delta^2}{24} \cdot \binom{T-1}{2} \cdot \frac{1}{k^2n^2} \cdot \left(1 - \frac{3\Delta + 2k - 6}{kn - (k+\Delta)}\right)^{T-3}.$$

Let $x = 3\Delta + 2k - 6$ and $y = kn - (k+\Delta)$. Then

$$\begin{aligned}
\left(1 - \frac{x}{y}\right)^{T-3} &= \exp\left(-(T-3)\sum_{i=1}^{\infty} \frac{1}{i}\left(\frac{x}{y}\right)^i\right) \\
&= \exp\left(-\frac{Tx}{y} + \sum_{i=1}^{\infty}\left(\frac{3}{i} - \frac{Tx}{(i+1)y}\right)\left(\frac{x}{y}\right)^i\right) \\
&\geq e^{-Tx/y} \\
&\geq e^{-4\delta}.
\end{aligned}$$

The first inequality follows since $3/i \geq Tx/((i+1)y$ for all $i \geq 1$, and the second inequality follows since $4y \geq nx$ (using the definition of x, y and the assumptions of the theorem.) Plugging this back into our calculations, we obtain

$$\text{Prob}\left[i \text{ is repeated} \mid T,\, b_i \geq \frac{\Delta}{3}\right] \geq \frac{\Delta^2}{24} \cdot \binom{T-1}{2} \cdot \frac{1}{k^2n^2} \cdot e^{-4\delta}.$$

Now we shall take expectation with respect to T, conditional on $T \leq \delta n$. Using Lemma 3.1 and the fact that $n \geq 120$, we find that

$$\binom{\mathbf{E}\left[T \mid T \leq \delta n\right] - 1}{2} \geq \frac{\delta^2\, n^2\, (1-\theta\delta)^2}{16}.$$

Applying Jensen's inequality, we obtain

$$\begin{aligned}
\text{Prob}\left[i \text{ is repeated} \mid T \leq \delta n,\, b_i \geq \frac{\Delta}{3}\right] &\geq \frac{\Delta^2}{384} \cdot \delta^2\, n^2\, (1-\theta\delta)^2 \cdot \frac{1}{k^2n^2} \cdot e^{-4\delta} \\
&\geq \frac{1}{768} \cdot \delta^2\, (1-\theta\delta)^2 \cdot e^{-4\delta} \cdot \frac{\Delta}{k}. \qquad (5)
\end{aligned}$$

By summing (5) over the $\lceil k/5 \rceil$ most popular colours, the theorem is proved. □

3.4 The mixing time of the Glauber dynamics

We now calculate an upper bound for the mixing time of the Glauber dynamics, using Lemma 2.1 and Theorem 2.2. Let ξ be defined by

$$\begin{aligned} \xi(\delta, \eta) &= \frac{1}{3840} \cdot \delta^2 (1 - \theta\delta)^2 \cdot e^{-4\delta} \\ &= \frac{1}{3840} \cdot \delta^2 \left(1 - \frac{3\delta}{2 - \eta}\right)^2 \cdot e^{-4\delta}. \end{aligned}$$

Theorem 3.2 shows that $\mathbf{E}[\mathcal{C} \mid T \leq \delta n] \geq \xi\Delta$. Note that ξ is a decreasing function of η. Take $\delta = 1/8$ and $\eta = 8 \times 10^{-7}$. Then $\xi(\delta, \eta) \geq 2\eta$. (These values of δ, η are approximately optimal.) The discussion of Section 3.2 suggested that this condition was sufficient to ensure rapid mixing of the Glauber dynamics. We now give the details.

Theorem 3.3 *Let $n \geq 120$ and $\Delta \geq 14$. Suppose that $(2 - \eta)\Delta \leq k \leq 2\Delta$, where $\eta = 8 \times 10^{-7}$. The mixing time of the Glauber dynamics for graph colourings of Δ-regular, triangle-free graphs is bounded above by*

$$\tau(\varepsilon) \leq 4 \times 10^6 \, n \log(n\varepsilon^{-1}).$$

Proof. Let $\delta = 1/8$, as in the previous section. We bound the mixing time by finding an upper bound on the quantity γ such that

$$H(X_{\delta n}, Y_{\delta n}) \leq \gamma$$

over all initial pairs $(X_0, Y_0) \in S$. Using the remark following Theorem 2.2, we can define γ by (3). Let $q = 1 - (k + \Delta)/(kn)$, as in Lemma 3.1. Then

$$\begin{aligned} \text{Prob}[T \leq \delta n] &= 1 - q^{\delta n} \\ &\geq 1 - \exp\left(-\frac{k + \Delta}{k}\delta\right) \geq 1 - e^{-4\delta/3}. \end{aligned}$$

Using this, with the calculations of (4), we obtain

$$\begin{aligned} \gamma &\leq 1 - \left(1 - e^{-4\delta/3}\right)\left(1 - \exp\left(\frac{\eta\delta}{2 - \eta} - \frac{\eta}{3}\right)\right) \\ &\leq 1 - 3.3 \times 10^{-8}, \end{aligned}$$

substituting $\delta = 1/8$ and $\eta = 8 \times 10^{-7}$. Now applying Lemma 2.1 we find that the mixing time of the Glauber dynamics is bounded above by

$$\begin{aligned} \tau(\varepsilon) &\leq \delta n \cdot \frac{\log(n\varepsilon^{-1})}{1-\gamma} \\ &\leq \frac{10^8}{26.4}\, n \log(n\varepsilon^{-1}) \\ &< 4 \times 10^6\, n \log(n\varepsilon^{-1}). \end{aligned}$$

This bound holds for $(2-\eta)\Delta \leq k \leq 2\Delta$, where $\eta = 8 \times 10^{-7}$, assuming that $\Delta \geq 14$ and $n \geq 120$. □

Vigoda [23] described a new Markov chain for graph colourings which alters the colouring of up to six vertices at each transition. He showed using path coupling that this chain is rapidly mixing for $k \geq 11\Delta/6$. The mixing time of this chain is bounded above by

$$\frac{k}{k - \frac{11}{6}\Delta}\, n \log(n\varepsilon^{-1})$$

for $k > 11\Delta/6$. Vigoda also applies the comparison technique of Diaconis and Saloff-Coste [9] to show that the mixing time of the Glauber dynamics is at most

$$O\left(k \log(k)\, n^2 \log(n)\right)$$

when $k > 11\Delta/6$. In particular, this gives an upper bound of $O(n^2 \log n)$ when $k = 2\Delta$. It seems unlikely that any comparison technique could yield the optimal bound of $O(n \log n)$.

References

[1] S. Albers and M. Mitzenmacher, *Average-case analyses of first fit and random fit bin packing*, in *9th Annual Symposium on Discrete Algorithms*, ACM–SIAM, New York–Philadelphia, 1998 pp. 290–299.

[2] D. J. Aldous, *Some inequalities for reversible Markov chains*, Journal of the London Mathematical Society 2, 25 (1982), pp. 564–576.

[3] D. Aldous, *Random walks on finite groups and rapidly mixing Markov chains*, in A. Dold and B. Eckmann, eds., *Séminaire de Probabilités XVII 1981/1982*, vol. 986 of *Springer-Verlag Lecture Notes in Mathematics*, Springer-Verlag, New York, 1983 pp. 243–297.

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

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

[6] F. R. K. Chung and R. L. Graham, *Random walks on generating sets for finite groups*, Electronic Journal of Combinatorics, (1997), pp. 14 (electronic).

[7] E. G. Coffman, D. S. Johnson, P. W. Shor, and R. R. Weber, *Markov chains, computer proofs and average-case analysis of best fit bin packing*, in *25th Annual Symposium on the Theory of Computing*, ACM, New York, 1993, pp. 412–421.

[8] A. Czumaj, P. Kanarek, M. Kutyłowski, and K. Loryś, *Delayed path coupling and generating random permutations via distributed stochastic processes*, in *10th Annual Symposium on Discrete Algorithms*, ACM–SIAM, New York–Philadelphia, 1999 pp. 271–280.

[9] P. Diaconis and L. Saloff-Coste, *Comparison theorems for reversible Markov chains*, The Annals of Applied Probability, 3 (1993), pp. 696–730.

[10] P. Diaconis and L. Saloff-Coste, *Logarithmic Sobolev inequalities for finite Markov chains*, The Annals of Applied Probability, 6 (1996), pp. 695–750.

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

[12] M. Dyer and C. Greenhill, *Random walks on combinatorial objects*, in J. D. Lamb and D. A. Preece, eds., *Surveys in Combinatorics 1999*, vol. 267 of *London Mathematical Society Lecture Note Series*, Cambridge University Press, Cambridge, 1999, pp. 101–136.

[13] L. A. Goldberg and P. D. MacKenzie, *Analysis of practical backoff protocols for contention resolution with multiple servers*, Journal of Computer and System Sciences, 58 (1999), pp. 232–258.

[14] J. Håstad, T. Leighton, and B. Rogoff, *Analysis of backoff protocols for multiple access channels*, SIAM Journal on Computing, 25 (1996), pp. 740–774.

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

[16] C. Kenyon, A. Sinclair, and Y. Rabani, *Biased random walks, Lyapunov functions, and stochastic analysis of Best Fit bin packing*, Journal of Algorithms, 27 (1998), pp. 218–235.

[17] L. Lovász and P. Winkler, *Efficient stopping rules for Markov chains*, in *27th Annual Symposium on the Theory of Computing*, ACM, New York, 1995, pp. 76–82.

[18] L. Lovász and P. Winkler, *Mixing times*, in *Microsurveys in discrete probability*, vol. 41 of *DIMACS Series in Discrete Mathematics and Theoretical Computer Science*, American Mathematical Society, Providence, RI, 1998, pp. 85–133.

[19] R. Motwani and P. Raghavan, *Randomized Algorithms*, Cambridge University Press, Cambridge, 1995.

[20] D. Randall and P. Tetali, *Analysing Glauber dynamics by comparison of Markov chains*, Journal of Mathematical Physics, 41 (2000), pp 1598–1615.

[21] J. Salas and A. D. Sokal, *Absence of phase transition for antiferromagnetic Potts models via the Dobrushin uniqueness theorem*, Journal of Statistical Physics, 86 (Feb. 1997) 3–4, pp. 551–579.

[22] R. Swendsen and J. S. Wang, *Universal critical dynamics in Monte-Carlo simulation*, Physical Review Letters, 58 (1987), pp. 86–88.

[23] E. Vigoda, *Improved bounds for sampling colorings*, Journal of Mathematical Physics, 41 (2000), pp. 1555–1569.