

# Variable Length Path Coupling

Thomas P. Hayes\*      Eric Vigoda†

July 17, 2006

## Abstract

We present a new technique for constructing and analyzing couplings to bound the convergence rate of finite Markov chains. Our main theorem is a generalization of the path coupling theorem of Bubley and Dyer, allowing the defining partial couplings to have length determined by a random stopping time. Unlike the original path coupling theorem, our version can produce multi-step (non-Markovian) couplings. Using our variable length path coupling theorem, we improve the upper bound on the mixing time of the Glauber dynamics for randomly sampling colorings.

## 1 Introduction

### Overview

Analysis of the convergence rate of finite Markov chains has applications in a variety of fields, including Theoretical Computer Science, Statistical Physics, and Probability Theory. The Coupling Method for proving upper bounds on convergence rate dates back to the seminal work of Doeblin [6]. In computer

---

\*Division of Computer Science, University of California, Berkeley. Email: hayest@cs.uchicago.edu. This work was conducted while the author was at the University of Chicago.

†College of Computing, Georgia Institute of Technology. Email: vigoda@cc.gatech.edu. Supported by NSF Grant CCF-0455666. This work was conducted while the author was at the University of Chicago and Issac Newton Institute for Mathematical Sciences.

‡A preliminary version of this paper appeared in Proceedings of the Fifteenth Annual ACM-SIAM Symposium on Discrete Algorithms, 103-110, 2004.

science, coupling has seen many recent applications in the analysis of Markov Chain Monte Carlo algorithms (e.g., [4, 17]).

An important new tool for simplifying and extending the coupling method is the Path Coupling approach of Bubley and Dyer [3] (e.g., see [19]). Briefly, they reduce the problem of proving an upper bound on coupling rate to proving a contraction condition for one step of the evolution of a “partial coupling,” defined on a much smaller subset of pairs of configurations. The simple proof constructs a full coupling via iterated composition of partial couplings.

We present a natural generalization of this technique, which constructs a full coupling via iterated composition of partial couplings *whose lengths may themselves be random variables*. (See Section 3 for a precise statement of our theorem.) The potential usefulness of our result lies in providing a technique for simpler construction and analysis of multi-step (non-Markovian) couplings, which are known to be more powerful than one-step (Markovian) couplings [12, 16, 14]. Moreover, it enables the partial couplings to be variable-length, which may be more natural in some settings. There are several previous works which use a multi-step coupling, see [5, 14], or analyze a single-step coupling over many steps, see [9, 10]. Our main theorem is a general technique which, in many cases, will simplify and improve applications of multi-step couplings and analyses of single-step couplings over many steps.

We use our new technique to analyze a simple Markov chain for randomly sampling  $k$ -colorings of a graph on  $n$  vertices with maximum degree  $\Delta$ . Roughly speaking, we can prove the chain has mixing time  $O(n \log n)$  when  $k \geq 1.953\Delta$  and the girth is at least 5. This improves a result of Dyer et al. [9] which holds when  $k \geq (2 - \epsilon)\Delta$  where  $\epsilon = 8 \times 10^{-7}$ , for triangle-free graphs. More recently, various further improvements have been made (see [11] for a survey).

## Variable Length Path Coupling

Consider a finite ergodic Markov chain with state space  $\Omega$ , transition matrix  $P$  and stationary distribution  $\pi$ . By a “coupling” for this chain, we will mean a joint stochastic process  $(X_t, Y_t)$  on  $\Omega \times \Omega$  such that each of the processes  $(X_t)$  and  $(Y_t)$  is a Markov chain on  $\Omega$  with transition matrix  $P$ . No restrictions are placed on the correlations between  $X_t$  and  $Y_t$ , allowing great variety among possible couplings.

The coupling is said have “coalesced” when  $X_t = Y_t$ . The time until a coupling has coalesced with high probability can be converted to an upper bound on the mixing time for the Markov chain. This is easy to see when one of the initial states, say  $Y_0$ , is drawn from the stationary distribution; then the probability that  $X_t \neq Y_t$  is an upper bound on the distance from stationarity at time  $t$ .

Typically, the probability of coalescence is analyzed by defining a metric  $d : \Omega \times \Omega \rightarrow \{0, 1, 2, \dots, d_{\max}\}$ . For  $\beta < 1$ , suppose there exists a one-step coupling  $(X_0, Y_0) \rightarrow (X_1, Y_1)$  such that

$$\mathbb{E}(d(X_1, Y_1) \mid X_0, Y_0) \leq \beta d(X_0, Y_0), \quad (1)$$

for all  $(X_0, Y_0) \in \Omega \times \Omega$ . Applying this coupling iteratively results in a  $t$ -step coupling, which is a Markov chain on  $\Omega \times \Omega$ , and which has coalescence probability at least  $3/4$  after after  $\lceil \ln(4d_{\max})/(1-\beta) \rceil$  steps, from an arbitrary pair of initial states.

There always exists a coupling whose probability of non-coalescence equals the distance from stationarity (c.f. [12]). However, this coupling is generally *non-Markovian*, by which we mean it is not a Markov chain on  $\Omega \times \Omega$  as in the previous paragraph, and so can only be defined in an inherently multi-step fashion. Kumar and Ramesh [16] gave an interesting example where one-step couplings are insufficient, and Hayes and Vigoda [14] recently illustrated the power of multi-step couplings.

Defining and analyzing a coupling for all pairs  $X_t, Y_t \in \Omega$  is often a difficult task. The path coupling technique simplifies the approach by restricting attention to pairs in a subset  $S \subseteq \Omega \times \Omega$  (assuming the graph  $(\Omega, S)$  is connected). It then suffices to define a one-step coupling such that (1) holds for all  $(X_0, Y_0) \in S$ . Then the path coupling theorem constructs, via simple compositions, a one-step coupling satisfying (1) for all  $X_0, Y_0 \in \Omega$ .

In many settings, it is natural to expect improvements by considering multiple transitions simultaneously. For a fixed length  $\ell$ , the path coupling theorem still applies to the Markov chain defined by  $P^\ell$  (i.e., the  $\ell$ -step evolution of the original chain).

However, we would often like to allow  $\ell$  to be a random variable. For example,  $\ell$  may be a random stopping time (cf. [9]) which depends on the evolution and the initial pair  $(X_t, Y_t)$ . Such an analysis was used by Dyer et al. [9] to prove improved convergence rates of a Markov chain for generating a random  $k$ -coloring.

Our main theorem generalizes the path coupling technique by allowing partial couplings whose length is a random variable. Thus, our main result is a *variable length path coupling* theorem. This work can be viewed as an improvement (and simplification) of a result of Dyer et al. [9, Theorem 2.2].

The proof of our theorem relies on a novel method for composing variable length couplings. Our composition technique can produce multi-step couplings, unlike the methods of Dyer et al. [9], and Buble and Dyer [3].

For proving upper bounds on the coalescence time, one can always avoid the use of a variable length coupling by analyzing a sufficiently large fixed length coupling. However, our approach is often simpler and more natural. This is illustrated in the proof of the following result on randomly sampling graph colorings.

## Applications to Graph Colorings

We apply our variable length path coupling theorem to improve convergence results of a simple Markov chain, known as the *Glauber dynamics*. We analyze the dynamics for randomly sampling proper  $k$ -colorings of an undirected graph  $G = (V, E)$  with  $n$  vertices and maximum degree  $\Delta$ . A proper  $k$ -coloring is an assignment  $\sigma : V \rightarrow [k]$  such that  $\sigma(v) \neq \sigma(w)$  whenever  $\{v, w\} \in E$ .

Transitions of the Glauber dynamics randomly recolor a randomly chosen vertex at each step (see Section 5 for a precise definition). When  $k \geq \Delta + 2$ , the Glauber dynamics is ergodic with unique stationary distribution uniform over (proper)  $k$ -colorings of the input graph. We are interested in bounding the *mixing time*, which is the number of transitions until the chain is close to stationarity (see Section 2 for a formal definition).

The first significant result was by Jerrum [17], who proved the mixing time is  $O(n \log n)$  whenever  $k > 2\Delta$ . Vigoda [19] later proved the mixing time is  $O(n^2)$  whenever  $k > 11\Delta/6$ , via analysis of a different Markov chain. Dyer et al. [9] subsequently proved  $O(n \log n)$  mixing time of the Glauber dynamics for  $k \geq (2 - \epsilon)\Delta$  where  $\epsilon = 8 \times 10^{-7}$ , assuming the input graph is  $\Delta$ -regular with girth  $g \geq 4$  and  $\Delta \geq 14$ . Their result relied on a variable length path coupling theorem which is significantly weaker than our theorem.

Using our new technique, we considerably improve the result of Dyer et al. For any graph with girth  $g \geq 5$  and  $\Delta \geq \Delta_0$ , where  $\Delta_0$  is a sufficiently large constant, we prove the mixing time of the Glauber dynamics is  $O(n \log n)$  whenever  $k \geq 1.953\Delta$ .

There are significantly stronger results known when  $\Delta = \Omega(\log n)$ . Using a multi-step coupling, Hayes and Vigoda [14] recently proved the following result. For all  $\epsilon > 0$ , all graphs with girth  $g \geq 11$  and  $\Delta = \Omega(\log n)$ , all  $k \geq (1 + \epsilon)\Delta$ , the mixing time of the Glauber dynamics is  $O(n \log n)$ . That result builds upon earlier work of Dyer and Frieze [8], Molloy [18], and Hayes [13]. Incidentally, a recent paper of Hayes and Sinclair [15] shows that the Glauber dynamics has mixing time  $\Omega(n \log n)$ , at least when  $\Delta = O(1)$ .

## 2 Preliminaries

Throughout the text, for a finite ergodic Markov chain, we use  $\Omega$  to denote the set of states,  $P$  to denote the transition matrix, and  $\pi$  the stationary distribution.

The total variation distance for a pair of distributions  $\mu$  and  $\nu$  on  $\Omega$  is defined as

$$d_{TV}(\mu, \nu) = \frac{1}{2} \sum_{x \in \Omega} |\mu(x) - \nu(x)|.$$

Our interest is the mixing time  $\tau_{\text{mix}}$  of the chain:

$$\tau_{\text{mix}} = \max_{x \in \Omega} \min\{t : d_{TV}(P^t(x, \cdot), \pi) \leq 1/4\}$$

We use the coupling method to bound the mixing time. For a Markov chain  $(\Omega, P, \pi)$ , a  $t$ -step coupling is defined as follows. For every  $(x_0, y_0) \in \Omega^2$ , let  $(\bar{X}, \bar{Y}) = ((X_0, X_1, \dots, X_t), (Y_0, Y_1, \dots, Y_t))$  be a random variable taking values in  $\Omega^t \times \Omega^t$ . We say  $(\bar{X}, \bar{Y})$  is a valid  $t$ -step coupling if  $X_t$  is distributed according to  $P^t(x_0, \cdot)$  and  $Y_t$  is distributed according to  $P^t(y_0, \cdot)$ . Note, the above definition requires that the distributions at time  $t$  are “faithful copies”, but does not require that the processes  $(X_t)$  and  $(Y_t)$  evolve according to the Markov chain. This definition of coupling suffices for the following result.

A valid coupling gives the following bound, known as the Coupling Inequality, on the convergence rate [6] (or e.g. [1]). For all  $x_0 \in \Omega$ ,

$$d_{TV}(P^t(x_0, \cdot), \pi) \leq \max_{y_0 \in \Omega} \mathbf{Pr}(X_t \neq Y_t)$$

Therefore, by defining a valid  $t$ -step coupling where all initial pairs have coalesced (i.e., are at the same state) with probability at least  $3/4$ , we have proved the mixing time is at most  $t$ .

An important property of couplings is that they “compose”. More precisely, given a valid  $t$ -step coupling  $(\overline{X}, \overline{Y})$  for  $(x_0, y_0)$  and a valid  $t$ -step coupling  $(\overline{Y}, \overline{Z})$  for  $(y_0, z_0)$ , then  $(\overline{X}, \overline{Z})$  is a valid  $t$ -step coupling for  $(x_0, z_0)$  and

$$\Pr(X_t \neq Z_t) \leq \Pr(X_t \neq Y_t) + \Pr(Y_t \neq Z_t).$$

See [2] for a precise proof of the above elementary fact which we refer to as the composition of couplings.

The composition of couplings is a key step in the following lemma, known as the path coupling lemma. The lemma takes a set of couplings for some pairs of initial states, and produces couplings for all pairs of initial states.

**Definition 1.** We call  $S \subseteq \binom{\Omega}{2}$  a *path-generating set* if the graph  $(\Omega, S)$  is connected. (The notation  $\binom{\Omega}{2}$  is the set of unordered pairs  $\{x, y\} \subset \Omega$ .) Note that the graph  $(\Omega, S)$  is an undirected graph whose edges  $S$  might have no relation to the transitions of the Markov chain we are studying. We assume  $S$  has an associated function  $d : S \rightarrow \mathbb{N}^+$ . Extend this to a metric on  $\Omega^2$  by setting  $d(x, y)$  to be the least sum of weights along any path between  $x$  and  $y$  in the weighted graph  $(\Omega, S)$ . Let  $d_{\max}$  denote the maximum over all  $(x, y) \in \Omega^2$  of  $d(x, y)$ .

**Lemma 2** (Bubley and Dyer [3]). *Let  $S$  be a path-generating set. Let  $(\overline{X}, \overline{Y})$  be a valid  $t$ -step coupling for all  $(x_0, y_0) \in S$ . Let*

$$\beta := \max_{(x_0, y_0) \in S} \frac{\mathbb{E}(d(X_t, Y_t) \mid x_0, y_0)}{d(x_0, y_0)}$$

*If  $\beta < 1$ , then the mixing time satisfies*

$$\tau_{\text{mix}} \leq t \left\lceil \frac{\ln(4d_{\max})}{1 - \beta} \right\rceil$$

Traditionally, the path coupling lemma is stated for the case  $t = 1$ , but it extends immediately to any fixed  $t$  by considering the Markov chain with transition matrix  $P^t$ .

Our main result will be an analogue of Lemma 2 which applies to couplings of variable rather than a fixed length. In this case, the definition of what constitutes a valid partial coupling will be less trivial, and the composition along paths will be substantially more involved.

### 3 Variable-length Partial Coupling

#### 3.1 Definition of Valid Variable-length Partial Couplings

Ultimately, we want to produce a valid  $t$ -step coupling of the given Markov chain for arbitrarily large fixed values of  $t$ . To this end, we will require that the given partial coupling meets some basic consistency conditions. Roughly speaking, for every  $t$ , when the variable length  $T$  of our partial coupling exceeds  $t$ , we truncate to the first  $t$  steps, and when  $T$  is less than  $t$ , we extend to  $t$  steps; then we require that the resulting  $t$ -step evolution has the correct marginal distributions.

**Definition 3.** Let  $S$  be a path-generating set. For every  $(x_0, y_0) \in S$ , let  $(\bar{X}, \bar{Y}) = (\bar{X}_{x_0, y_0}, \bar{Y}_{x_0, y_0})$  be a random variable taking values in  $(\Omega \times \Omega)^*$ . Let  $T = |\bar{X}| = |\bar{Y}|$ . Thus,  $\bar{X} = (X_1, \dots, X_T)$  and  $\bar{Y} = (Y_1, \dots, Y_T)$ . We say  $(\bar{X}, \bar{Y})$  is a *variable length partial coupling* when the following holds:

- *Faithful copies:* For all  $t \geq 0$ , for all  $(x_0, y_0) \in S$  define a new random variable  $X^t = X_{x_0, y_0}^t \in \Omega$  by the following experiment. Sample  $(\bar{X}, \bar{Y}) = ((X_1, \dots, X_T), (Y_1, \dots, Y_T))$ . If  $T \geq t$ , then “truncate” by setting  $X^t = X_t$ . Otherwise, “extend” by choosing  $X^t$  from the distribution  $P^{t-T}(X_T, \cdot)$ . We require that, for every  $(x_0, y_0) \in S$ , every  $t$ ,  $X^t$  is distributed according to  $P^t(x_0, \cdot)$ , and  $Y^t$  is distributed according to  $P^t(y_0, \cdot)$ .

**Remark 4. A.** Since our proofs will only examine the distance from stationarity at one time, we do not require that the sequence  $(X_0, X_1, \dots, X_t)$  be a Markov chain with transition matrix  $P$ , only that the distribution of  $X_t$  be correct for every  $t$ . We note however, that when a partial coupling does possess this additional structure, then so will the full coupling which we construct in Section 4.1.

- B.** Note that, in the above definition, although we do not require  $T$  to be a stopping time (that is, measurable with respect to  $(X_1, Y_1), \dots, (X_t, Y_t)$ ), we expect that this does not add any additional power to the approach.
- C.** The definition allows the partial couplings to be multi-step. Specifically, we can allow  $T$  to be a function of the initial states  $(x_0, y_0)$  and the

first  $T$  steps of only one of the chains  $(X_1, \dots, X_T)$ . The evolution of the other chain  $(Y_1, \dots, Y_T)$  is decided upon reaching time  $T$ .

### 3.2 Statement of Results

We can now state our main result.

**Theorem 5.** *For a variable length partial coupling  $(\bar{X}, \bar{Y})$ , let*

$$\beta := \max_{(x_0, y_0) \in S} \frac{\mathbb{E}(d(X_T, Y_T) \mid x_0, y_0)}{d(x_0, y_0)}$$

*If  $\beta < 1$ , then the mixing time for the Markov chain satisfies*

$$\tau_{\text{mix}} \leq M \left\lceil \frac{\ln(4d_{\max})}{1 - \beta} \right\rceil,$$

*where  $M$  denotes the maximum possible length of the partial coupling (infinity if the length is unbounded).*

Even when there is no absolute upper bound on the time  $T$ , Theorem 5 can be applied to a truncated version of the partial coupling, resulting in the following corollary (proved in Section 4.3).

**Corollary 6.** *With the notation in Theorem 5, let  $W$  denote the supremum of  $d(X_t, Y_t)$  over all possible evolutions  $(\bar{X}, \bar{Y}) = ((X_0, \dots, X_T), (Y_0, \dots, Y_T))$  and all  $t \leq T$ . If  $\beta < 1$ , then the mixing time for the Markov chain satisfies*

$$\tau_{\text{mix}} \leq 2M' \left\lceil \frac{\ln(4d_{\max})}{1 - \beta} \right\rceil,$$

*where  $M'$  satisfies  $\Pr(T > M') \leq (1 - \beta)/2W$ .*

**Remark 7.** One can get an easy bound on  $M'$  by Markov's inequality. In particular, we have

$$M' \leq \left\lceil \frac{2W \max_{(x_0, y_0) \in S} \mathbb{E}(T \mid X_0 = x_0, Y_0 = y_0)}{1 - \beta} \right\rceil$$

When the distribution of  $T$  decays exponentially, one can derive better bounds on  $M'$ ; in particular, if  $T$  is stochastically dominated by a geometrically distributed random variable with mean  $\mu$ , then an easy calculation shows

$$M' \leq \left\lceil \mu \ln \left( \frac{2W}{1 - \beta} \right) \right\rceil.$$

## 4 Proof of Theorem 5

Our proof is via the coupling method. Let

$$N := \left\lceil \frac{\ln(4d_{\max})}{1 - \beta} \right\rceil$$

Observe

$$N > \log_{\beta}(1/4d_{\max}) \tag{2}$$

In Section 4.1, we construct an  $MN$ -step coupling, denoted  $g^N$ , and prove that it is valid. Then in Section 4.2 we prove that  $g^N$  coalesces with high probability. The Coupling Inequality establishes the  $MN$  upper bound on mixing time.

### 4.1 Construction of full coupling

The high level idea is to successively run the variable length couplings until coalescence. It is unclear how to do this since the variable length couplings are only defined for pairs in  $S$ . Hence, suppose we start with  $(x_0, y_0) \in S$  and run the variable length coupling ending at  $(X_T, Y_T)$ . If  $(X_T, Y_T) \notin S$  then it is not clear how to proceed further since we have not defined a coupling for such a pair of initial states. Thus, one of our main tasks is to extend the variable length coupling from pairs in  $S$  to a coupling for all pairs in  $\Omega^2$ .

For fixed length couplings it is straightforward to extend a coupling from  $S$  to  $\Omega^2$  (as in the path coupling lemma 2). Hence, a simple way to extend variable length couplings from  $S$  to  $\Omega^2$  is to first convert them into fixed length couplings. We will do this by padding the variable length coupling with a trivial coupling. By appending the trivial coupling after the variable length coupling it will not affect the coupling time, and Definition 3 will ensure that this results in a valid fixed length coupling. Consequently, we will have a fixed length coupling (which will be of length  $M$ ) for all pairs in  $S$ . This can then be extended to all pairs in  $\Omega^2$  in a straightforward manner. Finally, proceeding in an inductive manner we can then construct couplings of length  $iM$  for  $i$  sufficiently large.

We now formalize our construction. Denote the variable-length coupling from the hypothesis of the theorem by  $\mu(x_0, y_0)$ , which is defined for all  $(x_0, y_0) \in S$ . Using this variable-length coupling, we will construct a sequence of fixed-length couplings. For all  $(x_0, y_0) \in \Omega^2$ , all  $1 \leq i \leq N$ , we define a  $(iM)$ -step coupling denoted by  $\nu_i(x_0, y_0)$ .

A basic procedure in our construction is converting variable-length couplings into fixed-length couplings by adding a trivial coupling at the end. From  $(X_t, Y_t) \in \Omega^2$ , the trivial one-step coupling  $(X_t, Y_t) \rightarrow (X_{t+1}, Y_{t+1})$  is defined as follows. Evolve  $X_t \rightarrow X_{t+1}$  according to the Markov chain of interest. If  $X_t = Y_t$ , set  $Y_{t+1} = X_{t+1}$ , otherwise independently evolve  $Y_t \rightarrow Y_{t+1}$ . We call  $t$  steps of this process as the  $t$ -step trivial coupling, denoted, for initial states  $(x_0, y_0) \in \Omega^2$ , as  $\omega_t(x_0, y_0)$ .

By converting our variable-length couplings into fixed-length couplings, we can compose the couplings along a path and define a coupling for all  $(x_0, y_0) \in \Omega^2$ . (Recall the notion of composing couplings from Section 2.) We construct our sequence of fixed-length couplings in an inductive manner which keeps the trivial coupling at the end. The goal is to coalesce (with sufficiently large probability) before we use the trivial coupling.

For  $i = 1$ , we construct  $\nu_1(x_0, y_0)$  using the variable length couplings  $\mu$  and the trivial couplings  $\omega$ . For  $(x_0, y_0) \in S$ , we form  $\nu_1(x_0, y_0)$  by running the variable-length coupling  $\mu(x_0, y_0)$  until time  $T$  (recall  $T \leq M$ ). From the resulting states  $(X_T, Y_T)$ , we then run the trivial coupling  $\omega_t(X_T, Y_T)$  for the final  $t = M - T$  steps. This defines the coupling  $\nu_i(x_0, y_0)$  for all initial pairs  $(x_0, y_0) \in S$ .

We extend the coupling from pairs in  $S$  to all pairs in  $\Omega^2$  by compositions. More precisely, for  $(x_0, y_0) \in \Omega^2 \setminus S$ , the coupling  $\nu_1(x_0, y_0)$  is constructed by considering a shortest path between  $x_0$  and  $y_0$  in the (undirected) graph  $(\Omega, S)$  where edge  $(x, y) \in S$  has weight  $d(x, y)$  (recall Definition 1). Denote a shortest path by  $z^0, z^1, \dots, z^\ell$  where  $z^0 = x_0$  and  $z^\ell = y_0$ . We then construct  $\nu_1(x_0, y_0)$  by composing the couplings along the path, i.e.,

$$\nu_1(x_0, y_0) = \mu(x_0, z^1) \circ \mu(z^1, z^2) \circ \dots \circ \mu(z^{\ell-1}, y_0)$$

Since the couplings on the path are of identical lengths, such a composition is standard (see the discussion in Section 2 and, e.g., see [3, 2]). This completes the definition of the couplings  $\nu_1(x_0, y_0)$  for all  $(x_0, y_0) \in \Omega^2$ .

For  $i > 1$ , we construct  $\nu_i$  inductively, using  $\nu_{i-1}, \mu$ , and  $\omega$ . Suppose  $\nu_{i-1}(x_0, y_0)$  has been defined for all pairs  $(x_0, y_0) \in \Omega^2$ .

For  $(x_0, y_0) \in S$ , we form  $\nu_i(x_0, y_0)$  by running the variable-length coupling  $\mu(x_0, y_0)$  until time  $T$ . From the resulting states  $(X_T, Y_T)$ , we then run the coupling  $\nu_{i-1}(X_T, Y_T)$  for a further  $(i - 1)M$  steps. Finally, from  $(X_\ell, Y_\ell)$ , where  $\ell = T + (i - 1)M$ , we evolve the chains using the trivial coupling  $\omega_t(X_\ell, Y_\ell)$  for the final  $t = M - T$  steps. This defines the coupling  $\nu_i(x_0, y_0)$  for all initial pairs  $(x_0, y_0) \in S$ . We then extend this to all

$(x_0, y_0) \in \Omega^2$  by composing these fixed-length couplings along the shortest path in the graph  $(\Omega, S)$ , as we did earlier in the case  $i = 1$ . This completes the construction of the couplings  $\nu_i$ , and it remains to prove they are valid couplings.

We prove by induction on  $i$  that for all  $1 \leq i \leq N$ ,  $\nu_i$  is a valid coupling. First, consider  $\nu_1(x_0, y_0)$  for  $(x_0, y_0) \in S$ . Observe the random sequence  $(X_T, \dots, X_M)$  which is produced by the trivial coupling  $\omega$ . This sequence is evolving according to the Markov chain of interest, thus  $X_M$  has distribution  $P^{M-T}(X_T, \cdot)$ . Note, we are extending  $X_T$  as in Definition 3, thus the distribution of  $X_M$  is  $P^M(x_0, \cdot)$ . Similarly for  $Y_M$ . Therefore,  $\nu_1(x_0, y_0)$  is a valid coupling for all  $(x_0, y_0) \in S$ . Since the composition of couplings is a well-defined coupling for the same Markov chain, we have that  $\nu_1(x_0, y_0)$  is a valid coupling for all  $(x_0, y_0) \in \Omega^2$ .

Consider  $1 < i \leq N$ . Assume by induction that  $\nu_{i-1}$  is a valid coupling for any pair of initial states in  $\Omega^2$ . Now, consider some  $(x_0, y_0) \in S$ . Focus attention on the random sequence  $(X_1, \dots, X_{iM})$ . The sequence is generated via the variable-length coupling up to time  $T$  and then is being extended via the original Markov chain (since  $\nu_{i-1}$  is valid and so is the trivial coupling). Therefore, again by Definition 3, the distribution of  $X_t$  is  $P^t(x_0, \cdot)$ . The same argument holds for  $(Y_1, \dots, Y_t)$  as well. This proves that  $\nu_i$  is a valid coupling for  $(x_0, y_0) \in S$ , which extends to all  $(x_0, y_0) \in \Omega^2$  as before. This completes the proof that the couplings  $\nu_i$  are valid couplings.

## 4.2 Analysis of full coupling

It remains to bound the probability of coalescence by time  $NM$ . For  $(x_0, y_0) \in \Omega^2$ , let  $p_i(x_0, y_0)$  denote the probability  $\nu_i(x_0, y_0)$  does not coalesce. Also, let

$$p_i(S) = \max_{(x_0, y_0) \in S} p_i(x_0, y_0).$$

Note, the maximum is over initial pairs whose partial coupling is defined. For any  $(x_0, y_0) \in \Omega^2$ , we have

$$p_i(x_0, y_0) \leq d(x_0, y_0)p_i(S), \tag{3}$$

by a union bound. Our goal is to bound  $p_N(S)$ .

We inductively bound  $p_{i+1}(S)$  by conditioning on the distance of  $(X_T, Y_T)$ . Observe  $p_0(S) = 1$ . For  $(x_0, y_0) \in S$ ,  $(x, y) \in \Omega^2$  let us denote

$$Q(x_0, y_0, x, y) = \Pr(X_T = x, Y_T = y \mid X_0 = x_0, Y_0 = y_0)$$

where the coupled evolution of  $X_t$  and  $Y_t$  is according to the variable length coupling  $\mu$ .

With this notation, for  $(x_0, y_0) \in S$ , the coupling  $\nu_{i+1}$  applies the variable length coupling, then applies  $\nu_i$  to the resulting pair of states, and finally uses a trivial coupling for the last  $M - T$  steps. Hence,

$$p_{i+1}(x_0, y_0) \leq \sum_{(x,y) \in \Omega^2} p_i(x, y) Q(x_0, y_0, x, y),$$

where the inequality comes from the fact that coalescence might occur in the final  $M - T$  steps. Continuing we have

$$\begin{aligned} p_{i+1}(x_0, y_0) &\leq \sum_{(x,y) \in \Omega^2} p_i(x, y) Q(x_0, y_0, x, y) \\ &\leq \sum_{(x,y) \in \Omega^2} d(X_T, Y_T) p_i(S) Q(x_0, y_0, x, y), && \text{by inequality (3),} \\ &= p_i(S) \mathbb{E}(d(X_T, Y_T) \mid x_0, y_0) \\ &\leq p_i(S) \beta, && \text{by the definition of } \beta. \end{aligned}$$

Thus  $p_{i+1}(S) \leq \beta p_i(S)$ , and hence by induction,  $p_N(S) \leq \beta^N$ . Applying (2),

$$p_N(S) \leq \beta^N < \frac{1}{4d_{\max}}.$$

Hence, by (3), for any pair  $(x_0, y_0) \in \Omega^2$  and for  $t = NM$ ,

$$\Pr(X_t \neq Y_t) \leq d_{\max} p_N(S) \leq 1/4.$$

Thus the coupling time is at most  $NM$ , and by the Coupling Inequality, this bound also applies to the mixing time.  $\square$

### 4.3 Proof of Corollary 6

Suppose we are given a partial coupling  $(\bar{X}, \bar{Y})$ , and an integer  $M'$ . Let  $T = |\bar{X}| = |\bar{Y}|$ . We will define a “truncated” partial coupling  $(\bar{X}', \bar{Y}')$  where  $|\bar{X}'| = |\bar{Y}'| = \min\{T, M'\}$ . Let

$$(\bar{X}', \bar{Y}') = \begin{cases} (\bar{X}, \bar{Y}) & \text{when } T \leq M' \\ ((X_1, \dots, X_{M'}), (Y_1, \dots, Y_{M'})) & \text{otherwise.} \end{cases}$$

Clearly,  $(\bar{X}', \bar{Y}')$  inherits the property of being a partial coupling from  $(\bar{X}, \bar{Y})$ .

Recall that by definition,

$$\beta := \max_{(x_0, y_0) \in S} \frac{\mathbb{E}(d(X_T, Y_T) \mid x_0, y_0)}{d(x_0, y_0)}, \quad \text{and}$$

$$W := \max_{(x_0, y_0) \in S} \max_{t \leq T} d(X_t, Y_t).$$

Hence the corresponding quantities  $\beta', W'$  for  $(\bar{X}', \bar{Y}')$  satisfy  $W' \leq W$  and

$$\begin{aligned} \beta' &:= \max_{(x_0, y_0) \in S} \frac{\mathbb{E}(d(X_T, Y_T) \mid x_0, y_0)}{d(x_0, y_0)} \\ &\leq \beta + \frac{W \Pr(T > M')}{d(x_0, y_0)} \\ &\leq \beta + W \Pr(T > M') \end{aligned}$$

Assuming  $\Pr(T > M') \leq (1 - \beta)/2W$ , we have  $\beta' \leq (1 + \beta)/2$ . Applying Theorem 5 for the partial coupling  $(\bar{X}', \bar{Y}')$  proves Corollary 6.

## 5 Sampling Colorings

For a graph  $G = (V, E)$  with maximum degree  $\Delta$ , let  $\Omega$  be the set of all assignments  $\sigma : V \rightarrow K$  where  $K = [k] = \{1, \dots, k\}$ . For a coloring  $X$  and vertex  $v$ , let

$$A(X, v) := [k] \setminus X(N(v)),$$

denote the set of available colors for  $v$ .

The Glauber dynamics (heat-bath version) has state space  $\Omega$  and transitions defined as follows. From  $X_t \in \Omega$ ,

- Choose vertex  $z$  uniformly at random from  $V$ .
- Choose  $c$  uniformly at random from the set  $A(X_t, z)$ . That is,  $c$  is a random color not appearing in the neighborhood of  $z$ .
- Set  $X_{t+1}(w) = X_t(w)$  for all  $w \neq z$ .
- Set  $X_{t+1}(z) = c$ .

When  $k \geq \Delta + 2$ , it is straightforward to verify that the Glauber dynamics is ergodic, with the uniform distribution over proper  $k$ -colorings as its unique stationary distribution (e.g., see Jerrum [17]).

In practice, we can consider the chain defined only on proper  $k$ -colorings. The extension to all assignments is for technical reasons, and its mixing time is an upper bound on the mixing time of the chain restricted to proper colorings. In the remainder of the section, we use the term *colorings* to refer to assignments in  $\Omega$ .

We can now state our theorem on the Glauber dynamics.

**Theorem 8.** *There exists  $\Delta_0$  such that, for every graph  $G = (V, E)$  on  $n$  vertices having maximum degree  $\Delta \geq \Delta_0$  and girth  $g \geq 5$ , and for every  $k \geq 1.953\Delta$ , the Glauber dynamics on  $k$ -colorings of  $G$  has mixing time  $O(n \log n)$ .*

We begin with a few definitions. Let  $S$  denote pairs  $X, Y \in \Omega$  which differ at exactly one vertex. For  $X_t, Y_t \in \Omega$ , denote their Hamming distance by

$$H(X_t, Y_t) = |\{z : X_t(z) \neq Y_t(z)\}|.$$

We use the maximal one-step coupling, originally used by Jerrum [17]. More precisely, consider a pair of colorings  $(X_0, Y_0) \in S$ . Let  $v$  be the single vertex of disagreement ( $X_0(v) \neq Y_0(v)$ ).

At every step, both chains update the same vertex  $z$ . The coupling for the update at time 1 is as follows. If  $A(X_0, z) = A(Y_0, z)$ , then both chains choose the same new color  $X_1(z) = Y_1(z)$ . Note, if  $z$  is not adjacent to  $v$  then we have  $A(X_0, z) = A(Y_0, z)$ . For  $z \in N(v)$  we also have  $A(X_0, z) = A(Y_0, z)$  if both of the disagreement colors  $X_0(v), Y_0(v)$  occur on  $N(z) \setminus \{v\}$  (thus,  $X_0(v)$  and  $Y_0(v)$  are not in  $A(Y_0, z)$  and  $A(X_0, z)$ ).

Now suppose  $A(X_0, z) \neq A(Y_0, z)$ . This implies  $z$  is a neighbor of  $v$ . Moreover, we also know  $X_0(v) \in A(Y_0, z)$  and/or  $Y_0(v) \in A(X_0, z)$ . We now consider two cases. If  $|A(X_0, z)| = |A(Y_0, z)|$ , then  $A(X_0, z) \setminus A(Y_0, z) = \{Y_0(v)\}$  and  $A(Y_0, z) \setminus A(X_0, z) = \{X_0(v)\}$ . In this case, the coupling chooses  $X_1(z)$  uniformly at random from  $A(X_0, z)$ , and sets

$$Y_1(z) = \begin{cases} X_0(v) & \text{if } X_1(z) = Y_0(v) \\ X_1(z) & \text{otherwise.} \end{cases}$$

In the second case,  $|A(X_0, z)| \neq |A(Y_0, z)|$ . Without loss of generality, suppose  $|A(X_0, z)|$  is bigger. In this case,  $A(X_0, z) = A(Y_0, z) \cup \{Y_0(v)\}$ . Here

the coupling chooses  $X_1(z)$  uniformly at random from  $A(X_0, z)$ , and sets

$$Y_1(z) = \begin{cases} \text{a uniformly random color from } A(Y_0, z) & \text{if } X_1(v) = Y_0(v) \\ X_1(z) & \text{otherwise.} \end{cases}$$

The motivation behind this definition is that the probability of a second disagreement forming is

$$\sum_{z:A(X_0,z) \neq A(Y_0,z)} \frac{1}{n \max\{|A(X_0, z)|, |A(Y_0, z)|\}}, \quad (4)$$

which is the minimum for any one-step coupling from  $(X_0, Y_0)$ .

We define our variable-length partial coupling by iterating the above one-step partial coupling until the Hamming distance changes; i. e., until the first time  $T > 0$  at which either of the following occurs:

- Vertex  $v$  is selected for recoloring.
- A neighbor  $z$  of  $v$  is selected and recolored differently in the two chains.

Note that this makes sense, because prior to time  $T$ , the Hamming distance remains one, so we can repeatedly apply the partial coupling. The above random variable  $T$  is related to the stopping rule considered by Dyer et al. [9].

It remains to prove  $E(H(X_T, Y_T)) \leq \beta$  for some  $\beta < 1$ .

Our proof will rely on a high-probability lower bound for the number of available colors at a vertex  $w$  after the Glauber dynamics has run for a certain amount of time. This idea was first exploited by Dyer and Frieze [8], to prove that

$$|A(X_t, z)| > (1 - \delta)ke^{-\Delta/k}$$

with high probability (i.e., probability  $\geq 1 - \exp(-\text{poly}(\delta)\Delta)$ ) when  $t$  is sufficiently large as a function of  $\delta$ . For our application, we use an extension of this result with a more explicit (and sharper) trade-off between  $t$  and  $\delta$ .

We now define a function  $L(\alpha, \beta)$  which will be of use in our subsequent analysis of the number of available colors.

**Definition 9.** For all real numbers  $\alpha < \beta$ , define

$$L(\alpha, \beta) = (\beta - \alpha)e^{-(1-\alpha)/(\beta-\alpha)}.$$

**Observation 10.** For any fixed  $\beta > 1$ ,  $L(\alpha, \beta)$  is a differentiable and decreasing function of  $\alpha \in [0, 1]$ , with  $L(0, \beta) = \beta e^{-1/\beta}$  and  $L(1, \beta) = \beta - 1$ . Moreover, for all  $\alpha \in [0, 1]$ ,  $\beta > 1$ , we have

$$\frac{-1}{\beta} \leq \frac{\partial L(\alpha, \beta)}{\partial \alpha} \leq 0.$$

Note that, for  $\beta = k/\Delta$ , we have  $L(0, k/\Delta)\Delta = ke^{-\Delta/k}$ , and  $L(1, k/\Delta)\Delta = k - \Delta$ .

*Proof.* The partial derivative  $\partial L/\partial \alpha$  is easily computed, equalling

$$\partial L/\partial \alpha = -\frac{1-\alpha}{\beta-\alpha} e^{-(1-\alpha)/(\beta-\alpha)} < 0.$$

All the other facts follow easily.  $\square$

**Lemma 11.** For every  $\delta > 0$ , for all sufficiently large  $\Delta \geq \Delta_0(\delta)$ , for every graph  $G = (V, E)$  having girth  $\geq 5$  and maximum degree  $\Delta$ , for  $k > 1.01\Delta$ , for every  $t > 0$ ,  $w \in V$ ,

$$\Pr\left(\frac{|A(X_t, w)|}{\Delta} < L(e^{-t/n}, k/\Delta) - \delta\right) \leq \exp(-\text{poly}(\delta)\Delta).$$

where  $L$  is defined in Definition 9.

Lemma 11 is a generalization of an earlier result of Hayes [13, Lemmas 11, 12], which assumed  $\Delta = \Omega(\log n)$  and  $t/n$  sufficiently large. Although our proof is essentially the same, we include it in Section 5.1 for completeness.

The following corollary says that the result of Lemma 11 holds with high probability for all times  $t \leq T$ .

**Corollary 12.** For every  $\delta > 0$ , for all sufficiently large  $\Delta \geq \Delta_1(\delta)$ , for every graph  $G = (V, E)$  having girth  $\geq 5$  and maximum degree  $\Delta$ , for  $k > 1.01\Delta$ ,

$$\Pr\left((\exists t \leq T, w \in N(v)) \frac{|A(X_t, w)|}{\Delta} < L(e^{-t/n}, k/\Delta) - \delta\right) \leq \exp(-\text{poly}(\delta)\Delta).$$

The quantity  $\text{poly}(\delta)$  hides an exponent which is linear in  $\frac{\Delta}{k-\Delta}$ ; this exponent is constant under our assumption that  $k > 1.01\Delta$ .

*Proof.* Fix a neighbor  $w \in N(v)$ . Let  $t^* = \delta^2 n \Delta / 99$ . Note that

$$\Pr(T > t^*) = (1 - 1/n)^{\lfloor t^* \rfloor} \leq e^{-\lfloor t^* \rfloor / n} < e^{-\delta^2 \Delta / 100},$$

for sufficiently large  $\Delta$ . For each  $t \geq 0$ , define the “bad” event

$$\mathcal{B}_{w,t} := \left\{ \frac{|A(X_t, w)|}{\Delta} < L(e^{-t/n}, k/\Delta) - \frac{\delta}{2} \right\}.$$

Note that for any positive integer  $j$  and time interval  $\{t_1 + 1, \dots, t_2\}$ , the probability that  $j$  or more neighbors of  $w$  are recolored is bounded from above by

$$\binom{t_2 - t_1}{j} \left(\frac{\Delta}{n}\right)^j < \left(\frac{e(t_2 - t_1)\Delta}{jn}\right)^j,$$

since there are only  $\binom{t_2 - t_1}{j}$  ways to choose the first  $j$  times for neighbors to be recolored, and the probability of choosing a neighbor at a particular time is at most  $\Delta/n$ . In particular, as long as  $t_2 - t_1 \leq \frac{jn}{e^2 \Delta}$ , the probability that  $j$  or more neighbors of  $w$  are recolored during the time interval is at most  $e^{-j}$ .

Note that as long as  $|A(X_{t_2}, w)|/\Delta \geq L(e^{-t_2/n}, k/\Delta) - \delta/2$  and fewer than  $\delta(k - e^{-t_2/n} \Delta)/2e$  neighbors of  $w$  are recolored during the interval  $\{t_1 + 1, \dots, t_2\}$ , we must have, for all  $t_1 < t \leq t_2$ ,

$$\begin{aligned} \frac{|A(X_t, w)|}{\Delta} &\geq L(e^{-t_2/n}, k/\Delta) - \frac{\delta}{2} - \frac{\delta(k - e^{-t_2/n} \Delta)}{2e} \\ &\geq L(e^{-t_2/n}, k/\Delta) - \delta \\ &\geq L(e^{-t/n}, k/\Delta) - \delta. \end{aligned}$$

Thus, setting

$$j = \left\lfloor \frac{\delta(k - \Delta)}{2e} \right\rfloor < \frac{\delta \min_t(k - e^{-t/n} \Delta)}{2e},$$

and partitioning  $\{1, \dots, t^*\}$  into at most  $\left\lceil \left\lceil \frac{\delta^2 n \Delta}{100} \right\rceil / \left\lfloor \frac{jn}{e^2 \Delta} \right\rfloor \right\rceil$  intervals of length  $\leq jn/e^2 \Delta$ , a union bound yields the desired result. Note that for sufficiently large  $\Delta_1(\delta)$ , the multiplicative factor from the union bound is dominated by the inverse of the error probability, and so can be dropped.  $\square$

*Proof of Theorem 8.* Let  $\delta > 0$ , let

$$\mathcal{B}_t := \left\{ \exists s < \min\{t, T\} \text{ and } w \in N(v), \frac{|A(X_s, w)|}{\Delta} < L(e^{-s/n}, k/\Delta) - \delta \right\},$$

and let  $\bar{\mathcal{B}}_t$  denote its complement. Also, let  $\mathcal{B} = \cup_t \mathcal{B}_t$  and let  $\bar{\mathcal{B}}$  be the complement of  $\mathcal{B}$ . By Corollary 12,  $\Pr(\mathcal{B}) \leq \exp(-\text{poly}(\delta)\Delta)$ .

Recall that the time  $T$  occurs when either  $v$  is recolored, which always reduces the Hamming distance to zero, or when a new disagreement is formed, increasing the Hamming distance to two. It follows that

$$\mathbb{E}(H(X_T, Y_T)) = 2\Pr(H(X_T, Y_T) = 2).$$

Recall (4), which applies for all  $t \leq T$ . Hence, for all  $t \geq 1$ ,  $x, y \in \Omega$ ,

$$\begin{aligned} \Pr(H(X_t, Y_t) = 2 \mid X_{t-1} = x, Y_{t-1} = y, T \geq t) \\ = \sum_{w \in N(v)} \frac{1}{n \max\{|A(X_{t-1}, w)|, |A(Y_{t-1}, w)|\}}. \end{aligned}$$

Hence,

$$\beta_t := \Pr(H(X_t, Y_t) = 2 \mid \bar{\mathcal{B}}_t, T \geq t) \leq \frac{1}{n(L(e^{-t/n}, k/\Delta) - \delta)}. \quad (5)$$

Decomposing into the “good” and “bad” cases, we have

$$\Pr(H(X_T, Y_T) = 2) \leq \Pr(\mathcal{B}) + \Pr(H(X_T, Y_T) = 2, \bar{\mathcal{B}}).$$

We can expand

$$\begin{aligned} \Pr(H(X_T, Y_T) = 2, \bar{\mathcal{B}}) &= \sum_{t=0}^{\infty} \Pr(H(X_t, Y_t) = 2, T \geq t, \bar{\mathcal{B}}) \\ &\leq \sum_{t=0}^{\infty} \Pr(H(X_t, Y_t) = 2, T \geq t, \bar{\mathcal{B}}_t) \\ &= \sum_{t=0}^{\infty} \Pr(H(X_t, Y_t) = 2 \mid T \geq t, \bar{\mathcal{B}}_t) \Pr(T \geq t, \bar{\mathcal{B}}_t) \\ &= \sum_{t=0}^{\infty} \beta_t \Pr(T \geq t, \bar{\mathcal{B}}_t). \end{aligned}$$

Note that  $T \geq s + 1$  if and only if  $H(X_s, Y_s) = 1$  and  $T \geq s$ . By the definition of our coupling, we have

$$\begin{aligned} \Pr(H(X_s, Y_s) = 1 \mid \bar{\mathcal{B}}_s, T \geq s) &= 1 - \Pr(H(X_s, Y_s) \in \{0, 2\} \mid \bar{\mathcal{B}}_s, T \geq s) \\ &= 1 - \frac{1}{n} - \beta_s. \end{aligned}$$

This allows us to write

$$\begin{aligned} \Pr(T \geq t, \bar{\mathcal{B}}_t) &= \Pr(T \geq 0, \bar{\mathcal{B}}_0) \prod_{s=0}^{t-1} \Pr(T \geq s+1, \bar{\mathcal{B}}_{s+1} \mid T \geq s, \bar{\mathcal{B}}_s) \\ &\leq \prod_{s=0}^{t-1} \Pr(T \geq s+1 \mid T \geq s, \bar{\mathcal{B}}_s) \\ &= \prod_{s=0}^{t-1} \Pr(H(X_s, Y_s) = 1 \mid T \geq s, \bar{\mathcal{B}}_s) \\ &= \prod_{s=0}^{t-1} \left(1 - \frac{1}{n} - \beta_s\right). \end{aligned}$$

Combining the upper bounds just derived, we have

$$\Pr(H(X_T, Y_T) = 2, \bar{\mathcal{B}}) \leq \sum_{t=1}^{\infty} \beta_t \prod_{s=0}^{t-1} \left(1 - \frac{1}{n} - \beta_s\right).$$

It is easily verified that this is an increasing function in each of the variables  $\beta_i$ , where  $0 \leq \beta_i \leq 1 - 1/n$ . Since (5) gives the upper bound

$$\beta_i \leq \frac{1}{n(L(e^{-i/n}, k/\Delta) - \delta)},$$

this implies

$$\begin{aligned} &\Pr(H(X_T, Y_T) = 2, \bar{\mathcal{B}}) \\ &\leq \sum_{t=1}^{\infty} \frac{1}{n(L(e^{-t/n}, k/\Delta) - \delta)} \prod_{s=0}^{t-1} \left(1 - \frac{1}{n} - \frac{1}{n(L(e^{-s/n}, k/\Delta) - \delta)}\right). \end{aligned}$$

When  $n$  is large, the above expression is approximated by the following double integral:

$$\int_0^\infty \frac{1}{L(e^{-x}, k/\Delta) - \delta} \exp\left(-\int_0^x 1 + \frac{1}{L(e^{-y}, k/\Delta) - \delta} dy\right) dx,$$

When  $\delta = 0$  and  $k/\Delta = 1.953$ , numerical computation (using Maple) shows that the above integral is strictly less than  $1/2$ . Hence, for all sufficiently small values of  $\delta$  we have for all sufficiently large values of  $\Delta$  and  $k \geq 1.953\Delta$ ,  $E(H(X_T, Y_T)) < 1 - \delta$ .

Applying Corollary 6 and Remark 7 to this partial coupling completes the proof. Note that in the notation of Remark 7 we have  $\beta = 1 - \delta$ ,  $E(T) \leq n$ ,  $W = 2$ , and  $d_{\max} = n$ .  $\square$

## 5.1 Proof of Lemma 11

Fix  $t$  from the statement of Lemma 11.

We begin with some notation.

**Definition 13.** Let  $\mathcal{V}$  denote the sequence  $v_1, \dots, v_t$  of vertices selected by the Glauber dynamics, and let  $c_1, \dots, c_t$  denote the colors chosen at the corresponding times. For a subset of vertices  $U \subseteq V$ , let

$$\mathcal{H}(U) := \{z \in U : \text{there exists } 1 \leq s \leq t \text{ with } v_s = z\}$$

denote the set of vertices in  $U$  that are recolored at least once. For every vertex  $w \in \mathcal{H}(U)$ , let

$$t_w := \max\{t' : t' \leq t \text{ and } v_{t'} = w\}$$

denote the last time  $w$  is recolored before time  $t$ .

Let

$$R(U) := R(U, t) := \{1 \leq s \leq t : v_s \in U\}$$

denote the set of times when a vertex from  $U$  is chosen for recoloring.

Following [13], we define the following experiment.

**Definition 14.** Consider a triangle-free graph. Fix a vertex  $w \in V$ . Let

$$\mathcal{F} = (\mathcal{V}, (c_t : \forall u \in \mathcal{H}(N(w)), t_u \neq t))$$

denote the outcome of an experiment which reveals the entire sequence of selected vertices, as well as all selected colors *except at times*  $t_u$ . For every  $u \in \mathcal{H}(N(w))$ , let  $r(u)$  denote the range of  $X_{t-1}(u)$ , conditioned on the outcome of  $\mathcal{F}$ . In other words,

$$r(u) := [k] \setminus \bigcup_{s \in [t_u, t]} X_s(N(u)).$$

Since the graph is triangle-free, the ranges  $r(u)$ , for  $u \in \mathcal{H}(N(w))$ , are deterministic functions of  $\mathcal{F}$ .

We will show that, assuming  $w$  and vertices at distance 2 from  $w$  are not recolored too often (which will hold with high probability), the variables  $X_t(z)$ ,  $z \in N(w)$  are, conditioned on  $\mathcal{F}$ , nearly independent and nearly uniformly distributed over their respective conditional ranges. Moreover, we will show that, with high probability over those color choices revealed in  $\mathcal{F}$ , the above conditional ranges are large for all neighbors of  $w$  that have been recolored at least once.

We will need the following lemma, due to Dyer and Frieze [8, Lemma 2.1] (the present formulation appeared in [13, Lemma 17]). We include the proof for completeness.

**Lemma 15** (Dyer and Frieze). *Let  $c_1, \dots, c_s$  be independent random colors, where each  $c_i$  is sampled from a distribution  $D_i$  over a set  $Q$ . Let  $q := |Q|$ . Let*

$$p := \max_{1 \leq i \leq s, c \in Q} \Pr(c_i = c).$$

*Let  $A := Q \setminus (c_1 \cup \dots \cup c_s)$  denote the set of unchosen colors. Then*

$$\mathbb{E}(|A|) \geq q(1-p)^{s/qp} \geq q \left( \frac{1-p}{e} \right)^{s/q}.$$

*and for every  $a > 0$ ,*

$$\Pr(|A| \leq \mathbb{E}(|A|) - a) \leq e^{-a^2/2q}.$$

*Proof.* Let  $\eta_{i,c}$  denote the indicator random variable for the event that  $c_i = c$ . Note that  $|A| = \sum_{c=1}^q \prod_{i=1}^s (1 - \eta_{i,c})$ . By linearity of expectation and

independence of  $c_1, \dots, c_s$ , we have

$$\begin{aligned} \mathbb{E}(|A|) &= \sum_{c=1}^q \prod_{i=1}^s 1 - \mathbb{E}(\eta_{i,c}) \\ &\geq q \prod_{c=1}^q \prod_{i=1}^s (1 - \mathbb{E}(\eta_{i,c}))^{1/q} \end{aligned}$$

where the inequality is an application of the arithmetic mean-geometric mean inequality. The expectations  $\mathbb{E}(\eta_{i,c})$  sum to  $\Delta$  and all lie in the interval  $[0, p]$ ; subject to these assumptions, the product of  $1 - \mathbb{E}(\eta_{i,c})$  is minimized when as many as possible of the  $\mathbb{E}(\eta_{i,c})$  equal  $p$ . Hence,

$$\begin{aligned} \mathbb{E}(|A|) &\geq q(1-p)^{s/pq} \\ &\geq q \left( \frac{1-p}{e} \right)^{s/q} \end{aligned}$$

where the inequality  $(1-p)^{1/p} \geq \frac{1-p}{e}$  holds for all  $0 \leq p \leq 1$ , and can be derived easily from the standard inequality  $1+x \leq e^x$  when  $x = p/(1-p)$ .

The concentration inequality comes from a Chernoff bound, since  $A$  is the sum of the  $q$  negatively associated random variables  $\prod_{i=1}^s (1 - \eta_{i,c})$  (See Dubhashi and Ranjan [7, Proposition 7] for a proof of Chernoff's bound for sums of negatively associated random variables).  $\square$

Next we present a lemma establishing the conditional near-independence and near-uniformity for the colors assigned to  $N(v)$  under suitable hypotheses. Effectively, this is re-proving Lemma 21 of Hayes [13], although we have slightly clarified the notation and proof.

**Lemma 16.** *Let  $G$  be a graph of girth  $\geq 5$  and maximum degree  $\Delta$ . Assume  $k \geq \Delta + 2$ . Let  $w \in V$  and let  $t \geq 1$ . Then,*

$$\Pr \left( \frac{|A(X_t, w)|}{\Delta} \leq (1-p)L \left( 1 - \frac{|\mathcal{H}(N(w))|}{\Delta}, \frac{k}{\Delta} \right) - \delta \mid \mathcal{F} \right) \leq Z e^{-\delta^2 \Delta/2},$$

$$\text{where } Z := \exp \left( \frac{|R(w)|\Delta}{k - \Delta} \right) \quad \text{and} \quad p := \max_{u \in \mathcal{H}(N(w))} \frac{\exp \left( \frac{|R(N(u))|}{k - \Delta} \right)}{|r(u)|}.$$

*Proof.* Conditioned on the vertex sequence  $\mathcal{V}$ , the probability of a particular color sequence  $\theta = (\theta_1, \dots, \theta_t)$ , equals

$$\Pr(\theta \mid \mathcal{V}) = \Pr(\text{for } 1 \leq s \leq t, c_s = \theta_s \mid \mathcal{V}) = \prod_{s=1}^t \frac{1}{|A(z_s, v_s)|}, \quad (6)$$

where  $z_0, z_1, \dots, z_t$  is the sequence of colorings obtained from  $z_0 = X_0$  by recoloring  $z_s(v_s) = \theta_s$  at each timestep  $s$ .

Let  $\Theta$  denote the set of color sequences  $\theta = (\theta_1, \dots, \theta_t)$  which agree with  $\mathcal{F}$ . Note that, for each  $\gamma \in \Theta$ ,

$$\Pr(\gamma \mid \mathcal{F}) = \frac{\Pr(\gamma \mid \mathcal{V})}{\sum_{\theta \in \Theta} \Pr(\theta \mid \mathcal{V})}$$

Note that, in (6), the only terms in the product which are not identical for all  $\theta \in \Theta$ , are those where the index  $s$  is in  $R(N(\mathcal{H}(N(w))))$ ; for all others, the set  $A(z_s, v_s)$  is completely determined by  $\mathcal{F}$ . In fact, the only times  $s$  when  $A(z_s, v_s)$  is not fully determined by  $\mathcal{F}$ , are those times  $s > t_u$ , where  $u \in \mathcal{H}(N(w))$  and  $v_s \in N(u)$ . For  $\theta \in \Theta$ , let  $\mathcal{T}$  denote those times that we recolor  $u \in N(w)$  where  $w \in N(u)$  and  $w$  was recolored at least once, i.e.,

$$\mathcal{T} := R(N(\mathcal{H}(N(w))))$$

Finally, for  $\theta \in \Theta$ , define

$$\alpha(\theta) := \prod_{s \in \mathcal{T}} \frac{1}{|A(z_s, v_s)|}. \quad (7)$$

Then we have, for every  $\gamma \in \Theta$ ,

$$\Pr(\gamma \mid \mathcal{F}) = \frac{\alpha(\gamma)}{\sum_{\theta \in \Theta} \alpha(\theta)}.$$

Next we show how to approximate this distribution by a product distribution. By re-indexing in (7), we can write

$$\alpha(\theta) = Z(\theta)Y(\theta)$$

where

$$\begin{aligned}
Z(\theta) &:= \begin{cases} \prod_{s \in R(w)} \frac{1}{|A(z_s, v_s)|} & \text{if } \mathcal{H}(N(w)) \neq \emptyset \\ 1 & \text{otherwise} \end{cases} \\
Y(\theta) &:= \prod_{u \in \mathcal{H}(N(w))} Y(\theta, u) \\
Y(\theta, u) &:= \prod_{s \in R(N(u) \setminus \{w\})} \frac{1}{|A(z_s, v_s)|}
\end{aligned}$$

Now consider the distribution  $\mu$  over  $\Theta$  defined by

$$\mu(\gamma) := \frac{\prod_{u \in \mathcal{H}(N(w))} Y(\gamma, u)}{\sum_{\theta \in \Theta} \prod_{u \in \mathcal{H}(N(w))} Y(\theta, u)}$$

Because  $G$  has girth  $\geq 5$ , the terms  $Y(\theta, u)$  are completely determined by  $\mathcal{F}$  together with the single color  $z_t(u)$ . This implies that  $\mu$  is a product distribution; and hence for  $\gamma$  sampled according to  $\mu$ , the random variables  $z_t(u), u \in \mathcal{H}(N(w))$  are fully independent.

Again consider  $\gamma \in \Theta$  with resulting colorings  $y_0, \dots, y_t$ , and  $\theta \in \Theta$  with colorings  $z_0, \dots, z_t$ . The assumption that our graph has girth  $\geq 5$  ensures that

$$|A(z_s, v_s)| - |A(y_s, v_s)| \in \{-1, 0, 1\},$$

which implies that

$$\frac{|A(z_s, v_s)|}{|A(y_s, v_s)|} \leq 1 + \frac{1}{k - \Delta} \leq \exp\left(\frac{1}{k - \Delta}\right). \quad (8)$$

Recall  $r(u)$  is the set of consistent colors for  $X_t(u)$ . The upper bound (8) implies that, for  $\theta$  sampled according to  $\mu$ , every color  $z_t(u)$  in  $r(u)$  has conditional probability at most

$$\frac{\exp(|R(N(u) \setminus \{w\})|/(k - \Delta))}{|r(u)|} \leq p$$

to be chosen. Next, we will apply Lemma 15 to the colors  $z_t(u), u \in \mathcal{H}(N(w))$ . We will restrict our attention to colors which do not appear on the unrecolored vertices of  $N(w)$ , so we have  $s = |\mathcal{H}(N(w))|$  and  $q$  in the statement of

Lemma 15 equals  $k - (|N(w)| - s) \geq k - (\Delta - s)$  in our current notation. The conclusion is that under the distribution  $\mu$ ,

$$\begin{aligned} \mathbb{E}_\mu (|A(z_t, w)|) &\geq (k - (\Delta - |\mathcal{H}(N(w))|)) \left( \frac{1-p}{e} \right)^{|\mathcal{H}(N(w))|/(k - (\Delta - |\mathcal{H}(N(w))|))} \\ &\geq (1-p)L \left( 1 - \frac{|\mathcal{H}(N(w))|}{\Delta}, \frac{k}{\Delta} \right) \Delta. \end{aligned}$$

Lemma 15 also tells us the following high-probability bound under the distribution  $\mu$

$$\Pr_\mu \left( \frac{|A(X_t, w)|}{\Delta} < (1-p)L \left( 1 - \frac{|\mathcal{H}(N(w))|}{\Delta}, \frac{k}{\Delta} \right) - \delta \right) \leq e^{-\delta^2 \Delta/2}.$$

In the end we are interested in  $A(X_t, w)$  where the sequence of color choices are chosen according to the distribution  $\Pr(\theta | \mathcal{F})$ . Let

$$Z := \max_{\gamma, \theta \in \Theta} \frac{Z(\gamma)}{Z(\theta)}$$

The quantity  $Z$  bounds the maximum ratio of any event probability under  $\mu$  to its conditional probability given  $\mathcal{F}$ , under the Glauber dynamics. In particular, for the atomic events  $\theta \in \Theta$ ,

$$\Pr(\theta | \mathcal{F}) \leq Z\mu(\theta)$$

Therefore,

$$\Pr \left( \frac{|A(X_t, w)|}{\Delta} < (1-p)L \left( 1 - \frac{|\mathcal{H}(N(w))|}{\Delta}, \frac{k}{\Delta} \right) - \delta \mid \mathcal{F} \right) \leq Z e^{-\delta^2 \Delta/2}.$$

Hence to complete the proof, it remains to prove the bound on  $Z$  which is claimed in the statement of the lemma. For the product indexed over  $s \in R(w)$  in the definition of  $Z(\gamma), Z(\theta)$ , we apply the trivial bounds

$$|A(y_s, v_s)|, |A(z_s, v_s)| \in [k - \Delta, k - 1],$$

to deduce

$$\frac{|A(z_s, v_s)|}{|A(y_s, v_s)|} \leq \frac{k-1}{k-\Delta} = 1 + \frac{\Delta-1}{k-\Delta} \leq \exp \left( \frac{\Delta}{k-\Delta} \right).$$

Hence,

$$Z \leq \exp \left( \frac{|R(w)|\Delta}{k-\Delta} \right). \quad \square$$

Finally we present the main result of the section.

*Proof of Lemma 11.* Assume  $t \leq t^* := n \ln(1/\delta)$ . Since, by Observation 10,  $\frac{\partial L(\alpha, \beta)}{\partial \alpha} \geq -1$  for  $\alpha \in [0, 1], \beta \geq 1$ , it follows that, for all  $t > t^*$ ,

$$L\left(e^{-t/n}, \frac{k}{\Delta}\right) \leq L\left(e^{-t^*/n}, \frac{k}{\Delta}\right) + \delta.$$

Hence this assumption can be made without loss of generality, at the price of at most  $\delta$  additional errors.

Assuming the high-probability event of Lemma 16 holds, all that is required is that  $|\mathcal{H}(N(w))|$  not be too small and  $p$  not be too big. More precisely, if  $|\mathcal{H}(N(w))| \geq (1 - e^{-t/n} - \delta)\Delta$  and  $p \leq \delta\Delta/k$ , then

$$\begin{aligned} & (1-p)L\left(1 - \frac{|\mathcal{H}(N(w))|}{\Delta}, \frac{k}{\Delta}\right) \\ & \geq (1 - \delta\Delta/k)L\left(e^{-t/n} + \delta, \frac{k}{\Delta}\right) && \text{since } L(\alpha, \beta) \text{ is decreasing in } \alpha \\ & \geq (1 - \delta\Delta/k)\left(L\left(e^{-t/n}, \frac{k}{\Delta}\right) - \delta\right) && \text{since } \frac{\partial L(\alpha, \beta)}{\partial \alpha} \geq -1 \\ & \geq L\left(e^{-t/n}, \frac{k}{\Delta}\right) - 2\delta && \text{since } L(\alpha, \beta) \leq \beta. \end{aligned}$$

**Lower bound on  $|\mathcal{H}(N(w))|$ .** Let  $u \in N(w)$ . Let  $X_u$  be a 0–1 random variable that is 1 iff  $u$  is recolored at least once during the  $t$  steps of the Glauber dynamics. Then  $|\mathcal{H}(N(w))| = \sum_{u \in N(w)} X_u$ , and for each  $u$ ,  $\Pr(X_u = 0) = (1 - 1/n)^t \leq \exp(-t/n)$ . Therefore,  $\mathbb{E}(|\mathcal{H}(N(w))|) \geq (1 - \exp(-t/n))\Delta$ . The events  $\{u \in \mathcal{H}(N(w))\}$  for  $u \in N(w)$  are negatively associated, hence Chernoff's bound applies. (see [7, Proposition 7] for this version of the Chernoff bound). Therefore, applying Chernoff's bound, we have

$$\Pr(|\mathcal{H}(N(w))| \leq (1 - \exp(-t/n) - \delta)\Delta) \leq \exp(-\delta^2\Delta/2).$$

**Upper bound on  $p$ .** Fix a vertex  $u \in N(w)$ . Since vertices are chosen uniformly at random, we have, for all  $j \geq 1$ ,

$$\Pr(|R(N(u))| \geq j) \leq \binom{t}{j} \Delta^j n^{-j} \leq \left(\frac{et\Delta}{jn}\right)^j.$$

Let  $j^* := (e^2 t/n + \delta^2)\Delta$ . Note, the probability that  $|R(N(u))| \geq j^*$  is less than  $e^{-\delta^2 \Delta}$ . Hence to prove the bound  $p \leq k/\delta\Delta$ , it suffices to prove that with high probability,

$$|r(u)| \geq \frac{k}{\delta\Delta} \exp\left(\frac{j^*}{k-\Delta}\right) = \frac{k}{\delta\Delta} \exp\left(\left(\frac{e^2 t}{n} + \delta^2\right) \frac{\Delta}{k-\Delta}\right)$$

Note,

$$\frac{k}{\delta\Delta} \exp\left(\left(\frac{e^2 t}{n} + \delta^2\right) \frac{\Delta}{k-\Delta}\right) \leq \text{poly}(1/\delta),$$

because of our assumptions that  $1.01 \leq k/\Delta \leq 3$  and that  $t \leq n \ln(1/\delta)$ .

Consider the distribution of  $|r(u)|$ . This stochastically dominates the distribution of the number of missing colors out of a pool of  $k-\Delta$  colors, after  $|R(N(u))|$  colors are chosen uniformly at random. This is because, whenever  $i$  colors remain available, the conditional probability of hitting another one on the next recoloring is at most  $i/(k-\Delta)$ .

Since we are assuming that  $|R(N(u))| \leq (e^2 t/n + \delta^2)\Delta$ , we may apply Lemma 15 to conclude that the number of remaining colors is at least

$$(k-\Delta) \left(\frac{1 - \frac{1}{k-\Delta}}{e}\right)^{(e^2 t/n + \delta^2)\Delta/(k-\Delta)} - a = \text{poly}(\delta)(k-\Delta) - a,$$

except with failure probability at most  $e^{-a^2/2(k-\Delta)}$ . Since we only need this to be at least  $\text{poly}(1/\delta)$ , we can take  $a = \text{poly}(\delta)(k-\Delta)$ , where the polynomial is half as large as the one being subtracted from. In this case, the failure probability is of the form  $e^{-\text{poly}(\delta)(k-\Delta)}$ . Thus as long as  $\Delta > \Delta_0$  where  $\Delta_0$  is at least a suitable polynomial in  $1/\delta$  (where the degree depends on the desired ratio  $k/\Delta$ ) then  $k-\Delta \gg \text{poly}(1/\delta)$ , and we have shown

$$\Pr(p \geq \delta\Delta/k) \leq e^{-\text{poly}(\delta)(k-\Delta)}.$$

**Upper bound on  $Z$ .** We also need to prove a high-probability upper bound on the probability distortion factor  $Z$  from Lemma 16, which was defined as

$$Z := \exp\left(\frac{|R(w)|\Delta}{k-\Delta}\right)$$

Let  $i \geq 1$ . Since the vertex order is uniformly random, the probability that  $|R(w)| \geq i$  is at most

$$\binom{t}{i} n^{-i} \leq \left(\frac{et}{in}\right)^i.$$

Setting  $Z_0 := \exp\left(\frac{i\Delta}{k-\Delta}\right)$ , we have

$$\Pr(Z \geq Z_0) \leq \left(\frac{et}{in}\right)^i.$$

Finally, assuming that  $Z \leq Z_0$ , Lemma 16 tells us

$$\Pr\left(\frac{|A(X_t, w)|}{\Delta} \leq (1-p)L\left(1 - \frac{|\mathcal{H}(N(w))|}{\Delta}, \frac{k}{\Delta}\right) - \delta \mid \mathcal{F}\right) \leq Z_0 e^{-\delta^2 \Delta/2}.$$

Let  $i = \delta^2(k - \Delta)/4$ , so that  $Z_0 = e^{\delta^2 \Delta/4}$ . Summing up the probabilities that  $p$  is too small, that  $\mathcal{H}(N(w))$  is too small, that  $Z \geq Z_0$ , and that  $|A(X_t, w)|$  is too small, we conclude by a union bound that

$$\begin{aligned} \Pr\left(\frac{|A(X_t, w)|}{\Delta} \leq L\left(e^{-t/n}, \frac{k}{\Delta}\right) - 3\delta\right) &\leq 2e^{-\text{poly}(\delta)(k-\Delta)} + \left(\frac{et}{in}\right)^i + e^{-\delta^2 \Delta/4} \\ &= O\left(e^{-\text{poly}(\delta)(k-\Delta)}\right) \end{aligned}$$

When  $\Delta_0(\delta)$  is sufficiently large, the big-oh in front can become part of the poly in the exponent. Also, the  $k - \Delta$  can be replaced by  $\Delta$ , since  $k \geq 1.01\Delta$ .  $\square$

## References

- [1] D. J. Aldous. Random walks on finite groups and rapidly mixing Markov chains. In *Séminaire de Probabilités XVII*, 243–297. Springer-Verlag, 1983. Lecture Notes in Mathematics 986.
- [2] J. van den Berg and R. Brouwer. Random sampling for the monomer-dimer model on a lattice. *J. Math. Phys.* 41(3):1585–1597, 2000.
- [3] 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 (FOCS)*, 223–231, 1997.
- [4] R. Bubley, M. E. Dyer, and M. R. Jerrum. An elementary analysis of a procedure for sampling points in a convex body. *Random Structures and Algorithms*, 12(3):213–235, 1998.

- [5] A. Czumaj and M. Kutylowski, Delayed path coupling and generating random permutations. *Random Structures and Algorithms*, 17(3-4):238–259, 2000.
- [6] W. Doeblin. Exposé de la theorie des chaînes simples constantes de Markov à un nombre fini d'états. *Rev. Math. Union Interbalkanique*, 2:77–105, 1938.
- [7] D. Dubhashi and D. Ranjan. Balls and bins: a study in negative dependence. *Random Structures and Algorithms*, 13(2):99–124, 1998.
- [8] M. Dyer and A. Frieze. Randomly colouring graphs with lower bounds on girth and maximum degree. *Random Structures and Algorithms*, 23(2):167–179, 2003.
- [9] 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 J. Comput.* 30(6):1962–1975, 2001.
- [10] M. Dyer, C. Greenhill, and M. Molloy. Very rapid mixing of the Glauber dynamics for proper colorings on bounded-degree graphs. *Random Structures and Algorithms* 20(1):98–114, 2002.
- [11] A. Frieze and E. Vigoda. A survey on the use of Markov chains to randomly sample colorings. *Combinatorics, Complexity and Chance*, G. Grimmett and C. McDiarmid eds, Oxford University Press, 2007.
- [12] D. Griffeath, A maximal coupling for Markov chains, *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete*, 31:95–106, 1974/75.
- [13] T. P. Hayes. Randomly coloring graphs with girth at least five. In *Proceedings of the 35th Annual ACM Symposium on Theory of Computer Science (STOC)*, 269-278, 2003.
- [14] T. P. Hayes and E. Vigoda. A non-Markovian coupling for randomly sampling colorings. In *Proceedings of the 44th Annual Symposium on Foundations of Computer Science (FOCS)*, 618-627, 2003.
- [15] T. P. Hayes and A. Sinclair. A general lower bound for mixing of single site dynamics on graphs. In *Proceedings of the 46th Annual Symposium on Foundations of Computer Science (FOCS)*, 511-520, 2005.

- [16] V. S. Anil Kumar and H. Ramesh. Coupling vs. conductance for the Jerrum-Sinclair chain. *Random Structures and Algorithms*, 18(1):1-17, 2001.
- [17] M. R. Jerrum. A very simple algorithm for estimating the number of  $k$ -colorings of a low-degree graph. *Random Structures and Algorithms*, 7(2):157–165, 1995.
- [18] M. Molloy. The Glauber dynamics on colorings of a graph with high girth and maximum degree. *SIAM J. Computing*, 33(3):712-737, 2004.
- [19] E. Vigoda. Improved bounds for sampling colorings. *J. Math. Phys.* 41(3):1555-1569, 2000.