# 6.S891 Lecture 3: Mixing Time Bounds via Coupling

Kuikui Liu

September 14, 2023

The goal of this lecture is to introduce the first set of tools for bounding mixing times of Markov chains. The key idea here is to construct good *couplings* of distributions and Markov chains, which give a direct handle on the total variation distance. We will use the coupling method to prove the Fundamental Theorem of Markov Chains from the previous lecture.

## 1 The Coupling Method

We first describe the coupling method, which gives a direct way of upper bounding mixing times. We begin by defining couplings for distributions, and then lift them to couplings of Markov chains.

### 1.1 Coupling for Distributions

**Definition 1** (Coupling)**.** *Let $\mu, \nu$ be probability measures on $\Omega, \Sigma$, respectively. A* coupling *of $\mu, \nu$ is a probability measure $\xi$ on $\Omega \times \Sigma$ such that*

$$\begin{aligned} \mu(x) &= \sum_{y \in \Sigma} \xi(x, y), \qquad \forall x \in \Omega \\ \nu(y) &= \sum_{x \in \Omega} \xi(x, y), \qquad \forall y \in \Sigma. \end{aligned}$$

*In other words, the marginals of $\xi$ on each coordinate are precisely $\mu, \nu$, respectively.*

One can think of a coupling of $\mu, \nu$ as a method for sampling a pair of random variables $(X, Y)$ such that $\mathsf{Law}(X) = \mu$ (ignoring $Y$), and $\mathsf{Law}(Y) = \nu$ (ignoring $X$). Note that couplings always exist, since we always have the *product measure*

$$(\mu \otimes \nu)(x, y) \stackrel{\mathsf{def}}{=} \mu(x) \cdot \nu(y)$$

on $\Omega \times \Sigma$. In this case, we just sample $X \sim \mu, Y \sim \nu$ independently. At the other extreme, if $\mu = \nu$, then we always have the *identity coupling*, where

$$\xi(x, y) = \begin{cases} \mu(x) = \nu(x), & \text{if } x = y \\ 0, & \text{otherwise} \end{cases}.$$

Algorithmically, we just sample $X \sim \mu$ and output $(X, X)$.

The following lemma gives us the connection between couplings and total variation distance, and is key to using couplings towards bounding mixing times.

**Lemma 1.1** (Coupling Lemma)**.** *Let $\mu, \nu$ be two distributions on the same state space $\Omega$. Then*

$$\|\mu - \nu\|_{\mathsf{TV}} = \inf_{\xi} \Pr_{(X,Y) \sim \xi}[X \neq Y],$$

*where the infimum is over all couplings $\xi$ of $\mu$ and $\nu$.*

*Remark* 1. The two characterizations

$$\|\mu - \nu\|_{\mathsf{TV}} = \sup_{f: \Omega \to [0,1]} |\mathbb{E}_{\mu}[f] - \mathbb{E}_{\nu}[f]| = \inf_{\xi} \Pr_{(X,Y) \sim \xi}[X \neq Y],$$

can actually be interpreted through the lens of *linear programming duality*.

*Proof.* We first prove the upper bound. First, since $\xi$ is a coupling, $\xi(x,x) \leq \min\{\mu(x), \nu(x)\}$ for all $x \in \Omega$ so that $\Pr_{(X,Y)\sim\xi}[X = Y] = \sum_{x\in\Omega} \xi(x,x) \leq \sum_{x\in\Omega} \min\{\mu(x), \nu(x)\}$. It follows that

$$\begin{aligned}
\|\mu - \nu\|_{\mathsf{TV}} &= \sum_{x:\mu(x)\geq\nu(x)} (\mu(x) - \nu(x)) \\
&= \sum_{x\in\Omega} (\mu(x) - \min\{\mu(x), \nu(x)\}) \\
&= 1 - \sum_{x\in\Omega} \min\{\mu(x), \nu(x)\} \\
&\leq 1 - \Pr_{(X,Y)\sim\xi}[X = Y] \\
&= \Pr_{(X,Y)\sim\xi}[X \neq Y].
\end{aligned}$$

To prove that we have equality, it suffices to devise a coupling $\xi$ such that the only inequality above is saturated, i.e. we must ensure that $\xi(x,x) = \min\{\mu(x), \nu(x)\}$ for all $x \in \Omega$. At this point, I could just hand you a clean formula and save all of us some pain (see Remark 2). Here, is how one could try to reason about it step by step. Let $A = \{x : \mu(x) > \nu(x)\}$, $B = \{x : \nu(x) > \mu(x)\}$ and $C = \{x : \mu(x) = \nu(x)\}$. Then, as a $\Omega \times \Omega$ matrix with rows summing to $\mu$ and columns summing to $\nu$, we need $\xi$ to have the following block structure:

1. Along the diagonal blocks $A \times A$, $B \times B$ and $C \times C$, we have diagonal entries $\xi(x,x) = \min\{\mu(x), \nu(x)\}$. The off-diagonal entries for these blocks must all be zero since one of the marginal distributions must be saturated.

2. The blocks $A \times C$, $B \times C$, $C \times A$ and $C \times B$ must all be zero both marginals $\mu, \nu$ have already been saturated by the $C \times C$ block. Similarly, the $B \times A$ block must be zero.

$$\xi = \begin{array}{c} \\ \\ \end{array}\begin{array}{ccc} A & B & C \end{array}$$
$$\xi = \begin{pmatrix} \text{diag} & ??? & 0 \\ 0 & \text{diag} & 0 \\ 0 & 0 & \text{diag} \end{pmatrix} \begin{array}{c} A \\ B \\ C \end{array}$$

Thus, the only freedom we have in choosing our optimal coupling is designing the $A \times B$ submatrix of $\xi$ such that

$$\begin{aligned}
\mu(x) - \nu(x) &= \sum_{y\in B} \xi(x,y), \qquad \forall x \in A \\
\nu(y) - \mu(y) &= \sum_{x\in A} \xi(x,y), \qquad \forall y \in B.
\end{aligned}$$

Since $\mu(A) - \nu(A) = \nu(B) - \mu(B)$, such a submatrix is always possible. For instance, one can sort $A$ in increasing order of $\mu(x) - \nu(x)$, sort $B$ analogously, and inductively build a triangular matrix. $\square$

*Remark* 2. Here is a formula for such an optimal coupling. Take $\xi(x,x) = \min\{\mu(x), \nu(x)\}$ for all $x \in \Omega$, and for $x \neq y$, set

$$\xi(x,y) = \frac{(\mu(x) - \xi(x,x)) \cdot (\nu(y) - \xi(y,y))}{1 - \sum_{z\in\Omega} \xi(z,z)}.$$

## 1.2 Coupling for Markov Chains

Now that we have the notion of coupling for distributions, we can define the notion of coupling for Markov chains.

**Definition 2** (Markov Chain Coupling)**.** *Let $\mathsf{P}_X, \mathsf{P}_Y$ be two Markov chains on a common state space $\Omega$. A* coupling *of these two Markov chains is a stochastic process $(X_t, Y_t)_{t=0}^{\infty}$ on $\Omega \times \Omega$ such that for every $t \geq 0$ and every $a, b \in \Omega$,*

$$\begin{aligned}
\Pr[X_{t+1} = b \mid X_t = a] &= \mathsf{P}_X(a \to b) \\
\Pr[Y_{t+1} = b \mid Y_t = a] &= \mathsf{P}_Y(a \to b).
\end{aligned}$$

*In other words, each marginal process* $(X_t)_{t=0}^{\infty}, (Y_t)_{t=0}^{\infty}$ *is a faithful simulation of* $\mathsf{P}_X, \mathsf{P}_Y$*, respectively. We say such a coupling is* Markovian *if the process* $(X_t, Y_t)_{t=0}^{\infty}$ *is a Markov chain, i.e. for every* $t \geq 0$ *and every* $a, b, c \in \Omega$*,*

$$\Pr[X_{t+1} = c \mid X_t = a, Y_t = b] = \mathsf{P}_X(a \to c)$$
$$\Pr[Y_{t+1} = c \mid X_t = a, Y_t = b] = \mathsf{P}_Y(b \to c).$$

In most applications (e.g. bounding mixing times), we'll take $\mathsf{P}_X = \mathsf{P}_Y$, and the coupling will be Markovian. Such couplings are typically easier to analyze. Note that in our couplings, we can always enforce that $X_t = Y_t$ implies that $X_T = Y_T$ for all $T \geq t$ with probability 1; in other words, the moment the two copies of the chain coalesce, they stay stuck together forever. Another way to think of a Markovian coupling is designing a coupling $Q((x, y) \to (\cdot, \cdot))$ of $\mathsf{P}_X(x \to \cdot)$ and $\mathsf{P}_Y(y \to \cdot)$, in the sense of Definition 1, for all pairs $(x, y) \in \Omega \times \Omega$. The matrix $Q \in \mathbb{R}^{\Omega^2 \times \Omega^2}$ describes the transition probabilities of the Markovian coupling $(X_t, Y_t)_{t=0}^{\infty}$.

The following shows the connection between Markov chain couplings and mixing times.

**Lemma 1.2.** *Let* $\mathsf{P}$ *be a Markov chain on* $\Omega$ *with stationary distribution* $\mu$*. Then for every coupling* $(X_t, Y_t)_{t=0}^{\infty}$ *of the Markov chain* $\mathsf{P}$ *(with itself) such that* $X_0 \sim \mu_0$ *and* $Y_0 \sim \mu$*,*

$$\left\|\mu_0 \mathsf{P}^t - \mu\right\|_{\mathsf{TV}} \leq \Pr[X_t \neq Y_t].$$

*Proof.* Since $(X_t, Y_t)$ is a coupling and $X_0 \sim \mu_0, Y_0 \sim \mu$, we have $X_t \sim \mu_0 \mathsf{P}^t$ and $Y_t \sim \mu$. The claim follows immediately via Lemma 1.1. □

*Example* 1 (Hypercube). Consider Glauber dynamics for sampling from the uniform distribution over the discrete hypercube $\{\pm 1\}^n$; this is essentially the lazification of the simple random walk on $\{\pm 1\}^n$ under coordinate flips. In each step, we pick a uniformly random coordinate, and resample a uniformly random $\{\pm 1\}$-assignment for that coordinate. We build a coupling $(X_t, Y_t)_{t=0}^{\infty} \subseteq \{\pm 1\}^n \times \{\pm 1\}^n$ as follows: Regardless of what $X_t, Y_t$ are, in each step, we select the same coordinate $i \in [n]$ and the same $\{\pm 1\}$-assignment $s$, and set $X_{t+1}(i) = Y_{t+1}(i) = s$ (leaving all other coordinates the same).

This is an honest coupling of two copies of the Markov chain with the property that if coordinate $i$ is selected at time step $t$, then $X_T(i) = Y_T(i)$ for all $T \geq t+1$; in particular, once $X_t = Y_t$, then $X_T = Y_T$ for all $T \geq t$. It follows from Lemma 1.2 that for all $t \in \mathbb{N}$ and all $\mu_0$,

$$\left\|\mu_0 \mathsf{P}^t - \mu\right\|_{\mathsf{TV}} \leq \Pr[\exists \text{ unsampled } i \in [n] \text{ after } t \text{ steps}].$$

This is the classical *Coupon Collector Problem*, and it is well-known that the probability on the right-hand side is at most $e^{-c}$ if $t = \frac{1}{2} n \log n + cn$. Hence, this Markov chain has $\epsilon$-mixing time $O(n \log(n/\epsilon))$. In some sense, this is the "gold standard" for mixing of Glauber-like Markov chains.

## 2 Proof of the Fundamental Theorem of Markov Chains

First, as we showed in the previous lecture, we know that at least one stationary distribution $\mu$ exists. Furthermore, if we can prove that $\|\delta_x \mathsf{P}^t - \mu\|_{\mathsf{TV}} \to 0$ as $t \to \infty$ for any starting state $x \in \Omega$, then such a stationary distribution must be unique. Indeed, if $\nu$ is any stationary distribution of $\mathsf{P}$, then

$$\|\nu - \mu\|_{\mathsf{TV}} = \left\|\nu \mathsf{P}^t - \mu\right\|_{\mathsf{TV}} \leq \sum_{x \in \Omega} \nu(x) \cdot \left\|\delta_x \mathsf{P}^t - \mu\right\|_{\mathsf{TV}}$$

holds for every $t \geq 0$. Sending $t \to \infty$ yields $\nu = \mu$.

We show the desired convergence by constructing a coupling. Since $\mathsf{P}$ is ergodic, there exists $t^*$ such that $\epsilon = \min_{x,y} \mathsf{P}^{t^*}(x \to y) > 0$. We can think of $\mathsf{P}^{t^*}$ itself as a Markov chain $Q$ on $\Omega$. We will prove that for every $x, y \in \Omega$, $\|\delta_x Q^t - \delta_y Q^t\|_{\mathsf{TV}} \leq (1-\epsilon)^t$ for all $t \geq 0$. This contraction property immediately implies that $\{\delta_x Q^t\}_{x \in \Omega} = \{\delta_x \mathsf{P}^{t \cdot t^*}\}_{x \in \Omega}$ all converge to the same distribution and hence, so do the distributions $\{\delta_x \mathsf{P}^t\}_{x \in \Omega}$. This contraction follows immediately via the following trivial coupling $(X'_t, Y'_t)$ for $Q$: If $X'_t = Y'_t$, then transition to $X'_{t+1}$ according to $Q$ and set $Y'_{t+1} = Y'_t$. Otherwise, evolve $X'_{t+1}, Y'_{t+1}$ independently. Since $Q(x, y) \geq \epsilon$,

$$\Pr[X'_{t+1} \neq Y'_{t+1} \mid X'_t \neq X'_t] \leq 1 - \epsilon,$$

from which it follows that

$$\Pr[X_t' \neq Y_t' \mid X_0' \neq X_0'] = \prod_{j=0}^{t-1} \Pr[X_{j+1}' \neq Y_{j+1}' \mid X_j' \neq X_j'] \leq (1-\epsilon)^t.$$

The bound $\|\delta_x Q^t - \delta_y Q^t\|_{\mathsf{TV}} \leq (1-\epsilon)^t$ then follows by Lemma 1.2.

# 3 Path Coupling

Constructing good couplings is in general a nontrivial task. The method of *path coupling* greatly simplifies this task. We will leverage the following lemma, which shows how we can compose couplings together. It is a straightforward exercise to verify this lemma.

**Lemma 3.1** (Composition of Couplings)**.** *Let $\mu_1, \mu_2, \mu_3$ be probability measures on $\Omega_1, \Omega_2, \Omega_3$, respectively. Let $\xi_{12}$ be a coupling of $\mu_1, \mu_2$, and let $\xi_{23}$ be a coupling of $\mu_2, \mu_3$. Then the distribution*

$$\xi_{13}(x,z) \stackrel{\mathsf{def}}{=} \sum_{y \in \Omega_2} \frac{\xi_{12}(x,y) \cdot \xi_{23}(y,z)}{\mu_2(y)}$$

*is a coupling of $\mu_1, \mu_3$.*

*Remark* 3. Another way to think of $\xi_{13}$ is as the law of a random pair $(X,Z) \in \Omega_1 \times \Omega_3$ drawn as follows: First, we sample $Y \sim \mu_2$. Then, we sample $X \sim \xi_{12}(\cdot, Y)$ and $Z \sim \xi_{23}(Y, \cdot)$ independently, and output $(X,Z)$.

**Theorem 3.2** (Path Coupling; Bubley–Dyer [BD97a; BD97b])**.** *Let $\mathsf{P}$ be a Markov chain on a finite state space $\Omega$. Let $E \subseteq \binom{\Omega}{2}$ such that the undirected graph $(\Omega, E)$ is connected, and define $\mathrm{dist}(x,y)$ to be the shortest path distance in the graph $(\Omega, E)$. If there exists a coupling of $(X_t, Y_t) \to (X_{t+1}, Y_{t+1})$, for every $(X_t, Y_t) \in E$, such that*

$$\mathbb{E}[\mathrm{dist}(X_{t+1}, Y_{t+1}) \mid X_t, Y_t] \leq (1-\alpha) \cdot \mathrm{dist}(X_t, Y_t), \tag{1}$$

*then*

$$\mathrm{T}_{\mathsf{mix}}(\epsilon; \mathsf{P}) \leq \frac{1}{\alpha} \cdot \log\left(\frac{\mathrm{diam}(\Omega, E)}{\epsilon}\right).$$

*Proof.* The idea is to extend the given (incomplete) coupling into a full one satisfying Eq. (1) for all $(X_t, Y_t)$, not only pairs in $E$. Once we have such a full coupling, then we're done, since for all $t$

$$\begin{aligned}\Pr[X_t \neq Y_t \mid X_0, Y_0] &\leq \mathbb{E}[\mathrm{dist}(X_t, Y_t) \mid X_0, Y_0] \\ &\leq (1-\alpha)^t \cdot \mathrm{diam}(\Omega, E).\end{aligned}$$

This is less than $\epsilon$ if $t \geq \frac{1}{\alpha} \cdot \log\left(\frac{\mathrm{diam}(\Omega,E)}{\epsilon}\right)$.

We perform this extension by composing, in the sense of Lemma 3.1, the given couplings along a shortest path from $X_t$ to $Y_t$. Let $X_t = Z_t^{(0)}, \ldots, Z_t^{(k)} = Y_t$ be such a shortest path from $X_t$ to $Y_t$ in $(\Omega, E)$, where $k = \mathrm{dist}(X_t, Y_t)$. By our hypothesis Eq. (1),

$$\mathbb{E}\left[\mathrm{dist}\left(Z_{t+1}^{(j)}, Z_{t+1}^{(j+1)}\right) \mid Z_t^{(j)}, Z_t^{(j+1)}\right] \leq (1-\alpha) \cdot \mathrm{dist}\left(Z_t^{(j)}, Z_t^{(j+1)}\right).$$

It follows by the Triangle Inequality that

$$\begin{aligned}\mathbb{E}\left[\mathrm{dist}(X_{t+1}, Y_{t+1}) \mid X_t, Y_t\right] &\leq \sum_{j=0}^{k-1} \mathbb{E}\left[\mathrm{dist}\left(Z_{t+1}^{(j)}, Z_{t+1}^{(j+1)}\right) \mid Z_t^{(j)}, Z_t^{(j+1)}\right] \\ &\leq (1-\alpha) \cdot \sum_{j=0}^{k-1} \mathrm{dist}\left(Z_t^{(j)}, Z_t^{(j+1)}\right) \\ &= (1-\alpha) \cdot \mathrm{dist}(X_t, Y_t). && \text{(Shortest Path)}\end{aligned}$$

□

### 3.1 Application: Sampling Proper Colorings

In this section, we present a neat application of path coupling to sampling proper colorings in graphs. Let $G = (V, E)$ be a graph of maximum degree $\Delta$, and let $q \in \mathbb{N}$ be a given number of colors. A *(proper) $q$-coloring* of $G$ is an assignment $\chi : V \to [q]$ such that $\chi(u) \neq \chi(v)$ for all $\{u, v\} \in E$. Let $\mu$ be the uniform distribution over all proper $q$-colorings, which is a subset of $[q]^V$. Further recall that Glauber dynamics is given by the following Markov chain (specialized to colorings in this case): If the current coloring is $\chi : V \to [q]$, then:

- select a uniformly random vertex $v \in V$,
- select a uniformly random color $\mathfrak{c}$ currently available to $v$, i.e. uniformly among $[q] \setminus \{\chi(u) : u \sim v\}$,
- and update $\chi(v) \leftarrow \mathfrak{c}$.

**Fact 3.3.** *If $q \geq \Delta + 2$, then Glauber dynamics is ergodic. Furthermore, if $q = \Delta + 1$, then there exists a graph of maximum degree $\Delta$ such that Glauber dynamics is not connected.*

We prove this in Appendix A. Thus, we will typically require $q \geq \Delta + 2$. Note that in this regime, one can always find a coloring via the simple greedy algorithm. We prove the following.

**Theorem 3.4** ([Jer95])**.** *Suppose $q \geq 2\Delta+1$. Then Glauber dynamics with stationary distribution $\mu$ mixes in $O(\Delta n \log(n/\epsilon))$-steps.*

Before we prove the theorem, we mention some complementary results and conjectures.

**Theorem 3.5** ([Gal+14; GŠV15; GŠV16]; building on [Sly10; SS14])**.** *If $q \leq \Delta$ (even), then there is no* $\mathsf{FPRAS}$ *for approximately counting proper $q$-colorings unless* $\mathsf{NP} = \mathsf{RP}$.

This result says that under standard complexity-theoretic hypotheses, a condition like $q \geq \Delta+1$ is necessary. The following conjecture postulates that this is sharp, i.e. there is a *computation phase transition* at $q = \Delta + 1$.

**Conjecture 1.** *If $q \geq \Delta+1$, then there is an* $\mathsf{FPRAS}$ *for approximately counting proper $q$-colorings. Furthermore, if $q \geq \Delta + 2$, then Glauber dynamics mixes in $O(n \log n)$ steps.*

This is one of the major open problems in the field of approximate counting and sampling.

*Proof of Theorem 3.4.* We use path coupling w.r.t. Hamming distance, where two colorings $\chi, \chi'$ are adjacent if they differ in the color of exactly one vertex, say $w$. We now couple the transitions $\mathsf{P}(\chi \to \cdot)$ and $\mathsf{P}(\chi' \to \cdot)$.

- We select the same uniformly random vertex $v$.
- We now attempt to couple the update colors $\chi(v) \leftarrow \mathfrak{c}, \chi'(v) \leftarrow \mathfrak{c}'$ used. There are a few cases to consider depending on the vertex $v$ chosen.
  1. Suppose $v = w$. Since this disagreeing vertex is unique, $\{\chi(u) : u \sim v\} = \{\chi'(u) : u \sim v\}$, and so we can perfectly couple the update colors, i.e. $\mathfrak{c} = \mathfrak{c}'$ with probability 1. This is the *best* case, since after the update, the Hamming distance decreases by 1 and the two colorings no longer disagree anywhere.
  2. Suppose $v \notin N(w) \cup \{w\}$. Since no neighbor of $v$ is the vertex of disagreement, $\{\chi(u) : u \sim v\} = \{\chi'(u) : u \sim v\}$ still. So we can again perfectly couple $\mathfrak{c}, \mathfrak{c}'$. This is a *good case*, since after the update, the Hamming distance doesn't change.
  3. Suppose $v \in N(w)$. Let $\mathcal{L} = [q] \setminus \{\chi(u) : u \sim v, u \neq w\}$. Then the set of available colors to $v$ w.r.t. $\chi$ is $\mathcal{L} \setminus \{\chi(w)\}$, while the set of available colors w.r.t. $\chi'$ is $\mathcal{L} \setminus \{\chi'(w)\}$. Our goal is to optimally couple $\mathfrak{c} \sim \mathsf{Unif}(\mathcal{L} \setminus \{\chi(w)\})$ and $\mathfrak{c}' \sim \mathsf{Unif}(\mathcal{L} \setminus \{\chi'(w)\})$.
     - This can be done optimally if $\chi(w), \chi'(w) \notin \mathcal{L}$, since both sets are $\mathcal{L}$ itself.
     - If $\chi'(w) \notin \mathcal{L}$ but $\chi(w) \in \mathcal{L}$, then we first sample $\mathfrak{c} \sim \mathsf{Unif}(\mathcal{L} \setminus \{\chi(w)\})$, and with probability $\frac{|\mathcal{L}|-1}{|\mathcal{L}|}$, we take $\mathfrak{c}' = \mathfrak{c}$; with the remaining $\frac{1}{|\mathcal{L}|}$ probability, we take $\mathfrak{c}' = \chi(w)$. We employ essentially the same coupling in the symmetric case $\chi(w) \notin \mathcal{L}, \chi'(w) \in \mathcal{L}$.

– If $\chi(w), \chi'(w) \in \mathcal{L}$, then we first sample $\mathfrak{c}$ uniformly. If $\mathfrak{c} \neq \chi'(w)$, then set $\mathfrak{c}' = \mathfrak{c}$. If $\mathfrak{c} = \chi'(w)$, then set $\mathfrak{c}' = \chi(w)$.

We now analyze contraction. For $\chi_t, \chi'_t$ differing at a single vertex ($\mathrm{dist}(\chi_t, \chi'_t) = 1$), using the fact that $|\mathcal{L}| \geq q - \Delta + 1$ to analyze (3),

$$\mathbb{E}\left[\mathrm{dist}(\chi_{t+1}, \chi'_{t+1}) \mid \chi_t, \chi'_t\right] \leq \underbrace{\frac{n-\Delta-1}{n}}_{\text{Case (2)}} + \underbrace{\frac{\Delta}{n} \cdot 2 \cdot \frac{1}{q-\Delta} + \frac{\Delta}{n} \cdot \frac{q-\Delta-1}{q-\Delta}}_{\text{Case (3)}}$$
$$= 1 - \frac{q-2\Delta}{q-\Delta} \cdot \frac{1}{n}.$$

It follows that we have contraction as long as $q \geq 2\Delta + 1$. Since the diameter of the space of colorings is at most $n$ w.r.t. Hamming distance, and $\frac{q-\Delta}{q-2\Delta} \leq \Delta + 1$ for all $q \geq 2\Delta + 1$, the theorem follows. ☐

## References


[BD97a] R. Bubley and M. Dyer. "Path Coupling: A Technique for Proving Rapid Mixing in Markov Chains". In: *Proceedings of the 38th Annual Symposium on Foundations of Computer Science*. FOCS '97. Washington, DC, USA: IEEE Computer Society, 1997, pp. 223–. ISBN: 0-8186-8197-7 (cit. on p. 4).

[BD97b] R. Bubley and M. E. Dyer. *Path Coupling, Dobrushin Uniqueness, and Approximate Counting*. Tech. rep. University of Leeds, 1997 (cit. on p. 4).

[Gal+14] Andreas Galanis, Qi Ge, Daniel Štefankovič, Eric Vigoda, and Linji Yang. "Improved Inapproximability Results for Counting Independent Sets in the Hard-core Model". In: *Random Struct. Algorithms* 45.1 (Aug. 2014), pp. 78–110. ISSN: 1042-9832 (cit. on p. 5).

[GŠV15] Andreas Galanis, Daniel Štefankovič, and Eric Vigoda. "Inapproximability for Antiferromagnetic Spin Systems in the Tree Nonuniqueness Region". In: *J. ACM* 62.6 (Dec. 2015). ISSN: 0004-5411. DOI: 10.1145/2785964 (cit. on p. 5).

[GŠV16] Andreas Galanis, Daniel Štefankovič, and Eric Vigoda. "Inapproximability of the Partition Function for the Antiferromagnetic Ising and Hard-Core Models". In: *Combinatorics, Probability and Computing* 25.4 (2016), pp. 500–559 (cit. on p. 5).

[Jer95] Mark Jerrum. "A Very Simple Algorithm for Estimating the Number of k-Colorings of a Low-Degree Graph". In: *Random Structures & Algorithms* 7.2 (Sept. 1995), pp. 157–165. ISSN: 1042-9832 (cit. on p. 5).

[Sly10] Allan Sly. "Computational Transition at the Uniqueness Threshold". In: *2010 IEEE 51st Annual Symposium on Foundations of Computer Science*. 2010, pp. 287–296. DOI: 10.1109/FOCS.2010.34 (cit. on p. 5).

[SS14] Allan Sly and Nike Sun. "The Computational Hardness of Counting in Two-Spin Models on $d$-Regular Graphs". In: *The Annals of Probability* 42.6 (2014), pp. 2383–2416 (cit. on p. 5).


# A Unfinished Proofs

*Proof of Fact 3.3.* Aperiodicity follows immediately since $\mathsf{P}^{\mathsf{GD}}(\chi \to \chi) > 0$ for all colorings (the current color $\chi(v)$ assigned to $v$ is always available to $v$). To prove irreducibility, we must show that for every pair of colorings $\chi, \chi'$, there is a sequence of Glauber moves which changes $\chi$ into $\chi'$. Order the vertices $v_1, \ldots, v_n$ arbitrarily. Suppose for some $k \in [n]$, we have that $\chi(v_j) = \chi'(v_j)$ for all $1 \leq j \leq k$. We will update $\chi$ via Glauber moves so that $\chi(v_j) = \chi'(v_j)$ for all $1 \leq j \leq k+1$. Once $k$ reaches $n$ by induction, we'll have $\chi = \chi'$.

Suppose $\chi(v_{k+1}) \neq \chi'(v_{k+1})$. We have two cases:

(A) If $\chi'(v_{k+1})$ is available to $v_{k+1}$ w.r.t. $\chi$ (i.e. no neighbor of $u$ has $\chi(u) = \chi'(v_{k+1})$), then we can simply update $\chi(v_{k+1}) \leftarrow \chi'(v_{k+1})$.

(B) Otherwise, some neighbor $u$ of $v_{k+1}$ satisfies $\chi(u) = \chi'(v_{k+1})$. The high-level idea is to recolor $u$ so that $\chi(u) \neq \chi'(v_{k+1})$, while maintaining our invariant. Once we do this for all such neighbors $u$, $\chi'(v_{k+1})$ becomes available to $v_{k+1}$ and we can reduce to (A).

Since $\chi'(u) \neq \chi'(v_{k+1})$ by the coloring constraint, it must be that $\chi(u) \neq \chi'(u)$. In particular, $u = v_\ell$ for some $\ell > k+1$, so we're freely allowed to change the color of $u$ without violating our invariant that $\chi(v_j) = \chi'(v_j)$ for all $1 \leq j \leq k$. This is true for any such neighbor $u$. Since $q \geq \Delta + 2$, there is always some other color $\mathfrak{c} \neq \chi'(v_{k+1})$ which is available to $u$ w.r.t. $\chi$. Hence, for all $u \sim v_{k+1}$ such that $\chi(u) = \chi'(v_{k+1})$, we can recolor $\chi(u) \leftarrow \mathfrak{c} \neq \chi'(v_{k+1})$. Thus, we've used Glauber moves to change $\chi$ into a coloring such that $\chi(u) \neq \chi'(v_{k+1})$ for all $u \sim v_{k+1}$. Once we have reached such a coloring, $\chi'(v_{k+1})$ becomes available to $v_{k+1}$, so we can update $\chi(v_{k+1})$ and increase $k$.

This shows ergodicity when $q \geq \Delta + 2$. Now suppose $q = \Delta + 1$. Consider the complete graph $K_{\Delta+1}$ on $\Delta + 1$ vertices. This graph has maximum degree $\Delta$. Furthermore, if $\chi$ is a $(\Delta + 1)$-coloring, then all colors in the palette must be used. Hence, for any vertex $v \in V$, the only available color to $v$ is its current color $\chi(v)$. In other words, Glauber dynamics cannot move between the $(\Delta + 1)!$ many possible colorings. $\square$