14 Coupling Method, Mixing Times

14.1 Reversible Chains and Metropolis Algorithm

We would like to use Markov chains to sample from a given distribution. The key ingredient is how can we design a simple Markov chain whose stationary distribution is the desired one?

Definition 14.1. (Time-)Reversible Markov Chains

A Markov chain is called *reversible* (or *time-reversible*) if there exists a distribution π s.t.

$$\forall\, x, y \in \Omega, \qquad \pi(x) \cdot \mathbf{P}(x,y) = \pi(y) \cdot \mathbf{P}(y,x)\,.$$

The equation above is called the *detailed balance condition*.

Proposition 14.1.

If π exists, then π is the stationary distribution of $\mathbf{P}$.

Proof. We now verify that π is the stationary distribution:

$$\begin{aligned}(\pi^{\mathsf{T}}\mathbf{P})(y) &= \sum_{x\in\Omega} \pi(x)\cdot\mathbf{P}(x,y)\\ &= \sum_{x\in\Omega} \pi(y)\cdot\mathbf{P}(y,x)\\ &= \pi(y)\cdot\sum_{x\in\Omega}\mathbf{P}(y,x) = \pi(y)\,.\end{aligned}$$

□

Remark 14.1.

Checking the detailed balance condition is usually the simplest way to verify that a particular distribution is stationary. Furthermore, the detailed balance condition implies that for all $x_0, x_1, \ldots, x_n$,

$$\pi(x_0)\mathbf{P}(x_0,x_1)\cdots\mathbf{P}(x_{n-1},x_n) = \pi(x_n)\mathbf{P}(x_n,x_{n-1})\cdots\mathbf{P}(x_1,x_0)\,,$$

namely,

$$\Pr_{X_0\sim\pi}[X_0=x_0, X_1=x_1, \ldots, X_n=x_n] = \Pr_{X_0\sim\pi}[X_0=x_n, X_1=x_{n-1}, \ldots, X_n=x_0]\,.$$

Thus, if a Markov chain $\{X_t\}$ satisfies the detailed balance condition and starts from the stationary distribution, then the distribution of $(X_0, X_1, \ldots, X_n)$ is the same as the distribution of $(X_n, X_{n-1}, \ldots, X_0)$, and that's why the chains satisfying the detailed balance condition are called *time-reversible*.

Example 14.1.
Recall the simple random walk on graphs that we mentioned in the last lecture. Given an undirected graph $G=(V,E)$, we define the random walk as follows. Let $X_0, X_1, \dots X_t, \dots \in V$, and for each X_i, pick a neighbor u of X_i uniformly at random and let $X_{i+1}=u$. It is easy to check that the stationary distribution of this Markov chain is

$$\pi = \begin{pmatrix} d_1/\sum d_k \\ d_2/\sum d_k \\ \vdots \\ d_n/\sum d_k \end{pmatrix}$$

We now verify that this Markov chain is time-reversible:

$$\pi(i)\,\mathbf{P}(i,j) = \frac{d_i}{\sum d_k}\cdot\frac{\mathbb{1}_{[i\sim j]}}{d_i} = \frac{\mathbb{1}_{[i\sim j]}}{\sum d_k} = \frac{d_j}{\sum d_k}\cdot\frac{\mathbb{1}_{[i\sim j]}}{d_j} = \pi(j)\,\mathbf{P}(j,i)\,,$$

where we use $\sim$ to denote the relation of adjacency.

Now we introduce the celebrated Metropolis algorithm.

Let $\Delta = \max_{i\in\Omega}\deg(i)$. Then for all $i\in\Omega$, the Metropolis algorithm (a random walk on Ω) moving from i has two steps:

1. for every neighbor j of i, propose to move to j with probability $1/\Delta$;
2. accept with probability $\min\{\frac{\mu(j)}{\mu(i)},1\}$.

Formally, we define the entries in the transition matrix $\mathbf{P}$ as follows:

$$\mathbf{P}(i,j) = \begin{cases} 0\,, & \text{if } i\nsim j \text{ and } i\neq j\,;\\ \frac{1}{\Delta}\min\{\frac{\mu(j)}{\mu(i)},1\}\,, & \text{if } i\sim j\,;\\ 1-\sum_{i\sim j}\mathbf{P}(i,j)\,, & \text{if } i=j\,. \end{cases}$$

We now verify that μ is indeed the stationary distribution of $\mathbf{P}$. If $i=j$ or $i\nsim j$, it is clear that $\mu(i)\mathbf{P}(i,j)=\mu(j)\mathbf{P}(j,i)$. So we assume that $i\sim j$, and w.l.o.g. we further assume that $\mu(j)\geq\mu(i)$.

Since

$$\mu(i)\mathbf{P}(i,j) = \mu(i)\cdot\frac{1}{\Delta}\cdot\min\Big\{\frac{\mu(j)}{\mu(i)},1\Big\}$$

we obtain that

$$\mu(j)\mathbf{P}(j,i) = \mu(j)\cdot\frac{1}{\Delta}\cdot\frac{\mu(i)}{\mu(j)} = \frac{\mu(i)}{\Delta}$$

and

$$\mu(i)\mathbf{P}(i,j) = \mu(i)\cdot\frac{1}{\Delta} = \frac{\mu(i)}{\Delta}\,.$$

In fact, for most algorithmic applications, the desired distribution μ is unknown or hard to compute, but it is often much easier to calculate $\mu(i)/\mu(j)$, see e.g., sampling proper colorings. The key is that computing $\mu(i)$ directly costs $\Theta(|\Omega|)$ times of calculation while the state distribution of a Markov chain may be sufficiently close to stationary within $o(|\Omega|)$ runs.

14.2 TOTAL VARIATION DISTANCE

Now we are ready to introduce the proof and FTMC, and the analysis of convergence rates.

We first define the distance between two probability distributions that we will use later.

Definition 14.2. Total variation distance

Let Ω be a sample space and $\mu, \nu \in [0,1]^{\Omega}$ be two distributions. The *total variation distance* between two distributions μ and ν on Ω is given by

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

More generally, for continuous probability space,

$$D_{\mathsf{TV}}(\mu,\nu) = \frac{1}{2}\int_{x\in\Omega} |\mu(x) - \nu(x)|\,\mathrm{d}x.$$

Equivalently, we can also define the total variation distance as

$$D_{\mathsf{TV}}(\mu,\nu) \triangleq \max_{A\subseteq\Omega} \mu(A) - \nu(A).$$

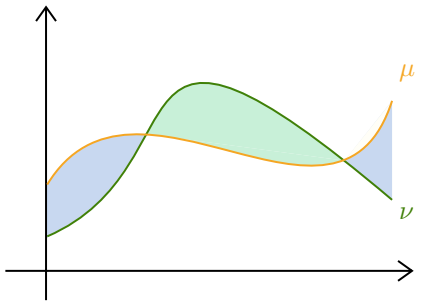


Figure 14.1: The total variation distance between μ and ν

Note that $\int \mu\,\mathrm{d}x = \int \nu\,\mathrm{d}x$. So $\int_{\mu(x)\geq\nu(x)} \mu(x) - \nu(x)\,\mathrm{d}x = \int_{\mu(x)\leq\nu(x)} \nu(x) - \mu(x)\,\mathrm{d}x$. See, for example, Figure 14.1. The area of blue part is equal to the area of green part, and that is why the first definition of the total variation distance has coefficient $1/2$.

Proof. Let $\Omega^+ \subseteq \Omega$ be the set of states such that $\mu(x) \geq \nu(x)$, and let $\Omega^- \subseteq \Omega$ be the set of states such that $\nu(x) > \mu(x)$. It can be easily verified that

$$\max_{A\subseteq\Omega} \mu(A) - \nu(A) = \mu(\Omega^+) - \nu(\Omega^+),$$
$$\max_{A\subseteq\Omega} \nu(A) - \mu(A) = \nu(\Omega^-) - \mu(\Omega^-).$$

By $\mu(\Omega) = \nu(\Omega) = 1$,

$$\mu(\Omega^+) + \mu(\Omega^-) = \nu(\Omega^+) + \nu(\Omega^-) = 1,$$

which implies that

$$\mu(\Omega^+) - \nu(\Omega^+) = \nu(\Omega^-) - \mu(\Omega^-).$$

We derive that

$$\max_{A\subseteq\Omega} |\nu(A) - \mu(A)| = \nu(\Omega^-) - \mu(\Omega^-) = \mu(\Omega^+) - \nu(\Omega^+).$$

Therefore,

$$\begin{aligned} D_{\mathsf{TV}}(\mu,\nu) &= \sum_{x\in\Omega} \frac{1}{2}\left|\mu(x)-\nu(x)\right| \\ &= \frac{1}{2}\Big(\left|\mu(\Omega^+)-\nu(\Omega^+)\right| + \left|\mu(\Omega^-)-\nu(\Omega^-)\right|\Big) \\ &= \max_{A\subseteq\Omega} \left|\nu(A)-\mu(A)\right|. \end{aligned}$$

□

14.3 Coupling Lemma

Although there are many ways to prove the fundamental theorem of Markov chains, we will present one based on the so called coupling method, which will be quite useful not only in the proof of FTMC, but also in the analysis of convergence rates.

The coupling of two distributions is simply a joint distribution of them.

Definition 14.3. Coupling
Let μ and ν be two distributions on the same space Ω. Let ω be a distribution on the space $\Omega\times\Omega$. If $(X,Y)\sim\omega$ satisfies $X\sim\mu$ and $Y\sim\nu$, then ω is called a coupling of μ and ν.

In other words, the marginal probabilities of the disjoint distribution ω are μ and ν respectively. A special case is when x and y are independently. However, in many applications, we want x and y to be correlated while keeping their respect marginal probabilities correct.

Intuitively we can view a coupling as a way to fill in a table with nonnegative reals. For example, let $\Omega=\{1,2,3\}$, $\mu=(1/3,1/3,1/3)^\mathsf{T}$ and $\nu=(1/2,1/4,1/4)^\mathsf{T}$. A coupling $\mathcal{C}$ of μ and ν is a way to fill in the following table with nonnegative reals such that the summations of rows and columns are equal to the corresponding *marginal* probabilities.

μ \ ν	$1/2$	$1/4$	$1/4$
$1/3$	$1/6$	$1/12$	$1/12$
$1/3$	$1/6$	$1/12$	$1/12$
$1/3$	$1/6$	$1/12$	$1/12$

μ \ ν	$1/2$	$1/4$	$1/4$
$1/3$	$1/3$	0	0
$1/3$	$1/12$	$1/4$	0
$1/3$	$1/12$	0	$1/4$

The table defines a joint distribution and the sum of a certain row/column equal to the corresponding marginal probability. It is clear that both table are couplings of the two coins. Among all the possible couplings, sometimes we are interested in the one who is "mostly coupled".

The following theorem reveals the connection between coupling and the total variation distance, and thus is a powerful tool to compute the total variation distance.

Lemma 14.2. Coupling Lemma
Let μ and ν be two distributions on a sample space Ω. Then for any coupling ω of μ and ν it holds that,

$$\Pr_{(X,Y)\sim\omega}[X \neq Y] \geq D_{\mathsf{TV}}(\mu,\nu).$$

And furthermore, there exists a coupling ω^* of μ and ν such that

$$\Pr_{(X,Y)\sim\omega}[X \neq Y] = D_{\mathsf{TV}}(\mu,\nu).$$

Let us prove the coupling lemma. For finite Ω, designing a coupling is equivalent to filling a $\Omega \times \Omega$ matrix in the way that the marginals are correct.

Clearly we have

$$\begin{aligned}\Pr[X=Y] &= \sum_{t\in\Omega}\Pr[X=Y=t] \\ &\leq \sum_{t\in\Omega}\min\{\mu(t),\nu(t)\}.\end{aligned}$$

Thus,

$$\begin{aligned}\Pr[X\neq Y] &\geq 1-\sum_{t\in\Omega}\min\{\mu(t),\nu(t)\} \\ &= \sum_{t\in\Omega}(\mu(t)-\min\{\mu(t),\nu(t)\}) \\ &= \max_{A\subseteq\Omega}\{\mu(A)-\nu(A)\} \\ &= D_{\mathsf{TV}}(\mu,\nu).\end{aligned}$$

To construct ω^* achieving the equality, for every $t \in \Omega$, we let $\Pr_{(X,Y)\sim\omega^*}[X = Y = t] = \min\{\mu(t),\nu(t)\}$. The construction of the off-diagonal entries of ω^* is left as an exercise.

The coupling lemma provides a way to upper bound the distance between two distributions: For any two distributions μ and ν and any coupling ω of μ and ν, an upper bound for $\Pr_{(X,Y)\sim\omega}[X \neq Y]$ is an upper bound for $D_{\mathsf{TV}}(\mu,\nu)$. This is a quite useful approach to bound the total variation distance and we will examine in detail in the next lecture. The coupling lemma also tells us that the upper bound obtained in this way can be tight, as long as you are able to find the optimal coupling.

14.4 Proof of FTMC

We already know that P has a stationary distribution π. What we would like to show is that for all starting distribution μ_0, it holds that

$$\lim_{t\to\infty} D_{\mathsf{TV}}(\mu_t,\pi) = 0\,,$$

where $\mu_t^\intercal = \mu_0^\intercal P^t$.

Suppose that $\{X_t\}$ and $\{Y_t\}$ are two identical Markov chains starting from different distribution, where $Y_0 \sim \pi$ while X_0 is generated from an arbitrary distribution μ_0.

Now we have two sequence of random variables:

$$\begin{array}{ccccccccccccc}
\mu_0 & & \mu_1 & & & & & & \mu_t & & & & \\
\wr & & \wr & & & & & & \wr & & & & \\
X_0 & \to & X_1 & \to & X_2 & \to & \cdots & \to & X_t & \to & X_{t+1} & \to & \cdots \\
\\
Y_0 & \to & Y_1 & \to & Y_2 & \to & \cdots & \to & Y_t & \to & Y_{t+1} & \to & \cdots \\
\wr & & \wr & & & & & & \wr & & & & \\
\pi & & \pi & & & & & & \pi & & & &
\end{array}$$

The coupling lemma establishes the connection between the distance of distributions and the discrepancy of random variables. To show that $D_{\mathsf{TV}}(\mu_t, \pi) \to 0$, it is sufficient to construct a coupling ω_t of μ_t and π and then compute $\Pr_{(X_t,Y_t)\sim\omega_t}[X_t \neq Y_t]$.

Here we give a simple coupling. Let $(X_t, Y_t) \sim \omega_t$ and we construct ω_{t+1}. If $X_t = Y_t$ for some $t \geq 0$, then let $X_{t'} = Y_{t'}$ for all $t' > t$, otherwise X_{t+1} and Y_{t+1} are independent. Namely, $\{X_t\}$ and $\{Y_t\}$ are two independent Markov chains until X_t and Y_t reach the same state for some $t \geq 0$, and once they meet together then they move together forever. The coupling lemma tells us that $D_{\mathsf{TV}}(\mu_t, \pi) \leq \Pr_{(X_t,Y_t)\sim\omega_t}[X_t \neq Y_t]$.

The property of irreducibility implies that

$$\forall\, i, j, \quad \exists n \quad \text{s.t.} \quad P^n(i,j) > 0\,.$$

We claim that combining with aperiodicity,

$$\exists n \quad \text{s.t.} \quad \forall\, i, j, \quad P^n(i,j) > 0\,.$$

Since the state space Ω is finite, it is sufficient to show that

$$\forall\, i, j, \quad \exists t_{i,j} \quad \text{s.t.} \quad \forall\, n > t_{i,j}, \quad P^n(i,j) > 0\,.$$

Suppose that there are s loops of length $c_1, c_2, \ldots, c_s$ starting from and ending at state i. Then we have

$$\gcd(c_1, c_2, \ldots, c_s) = 1\,.$$

Thus, by Bézout's identity, there exists $x_1, x_2, \ldots, x_s \in \mathbb{Z}$ such that

$$c_1 x_1 + c_2 x_2 + \cdots c_s x_s = 1\,.$$

This implies the following lemma.

Lemma 14.3.
For sufficiently large b, there exists $y_1, y_2, \ldots, y_s \in \mathbb{N}$ such that

$$c_1 y_1 + c_2 y_2 + \cdots c_s y_s = b.$$

The claim then follows from the lemma.

Now we know that $\exists n$ s.t. $\forall\, i, j$, $P^n(i,j) > 0$. Then we define

$$\theta \triangleq \min_{x_0, y_0 \in \mathcal{S}} \Pr X_n = Y_n \mid X_0 = x_0, Y_0 = y_0\,.$$

For simplicity, we use $\Pr[x_0, y_0]\cdot$ to denote the conditional probability $\Pr \cdot \mid X_0 = x_0, Y_0 = y_0$ from now on.

Fix $z \in \Omega$. Let

$$\alpha = \min_{w \in \Omega} P^n(w, z) > 0\,,$$

and for any $t \geq 0$ and $z' \in \Omega$,

$$\beta_{t,z'} = \Pr_{x_0,y_0}[X_t = Y_t = z' \wedge X_{t'} \neq Y_{t'} \text{ for all } t' < t].$$

By the Markov property and the independence of $\{X_t\}$ and $\{Y_t\}$ before $X_t = Y_t$, we obtain that

$$\begin{aligned}
&\Pr_{x_0,y_0}[X_n = Y_n] \\
&\geq \Pr_{x_0,y_0}[X_n = Y_n = z] \\
&= \Pr_{x_0,y_0}[X_n = Y_n = z \wedge \forall t < n, X_t \neq Y_t] + \Pr_{x_0,y_0}[X_n = Y_n = z \wedge \exists t < n, X_t = Y_t] \\
&= \left(P^n(x_0,z) \cdot P^n(y_0,z) - \sum_{t=0}^{n-1} \sum_{z'} \beta_{t,z'} \cdot \left(P^{n-t}(z',z)\right)^2\right) + \sum_{t=0}^{n-1} \sum_{z'} \beta_{t,z'} \cdot P^{n-t}(z',z) \\
&\geq P^n(x_0,z) \cdot P^n(y_0,z) \geq \alpha^2.
\end{aligned}$$

Hence $\theta > 0$. By the coupling and the Markov property, we have

$$\begin{aligned}
\Pr_{x_0,y_0}[X_{2n} \neq Y_{2n}] &= \sum_{x_n \neq y_n} \Pr_{x_0,y_0}[X_{2n} \neq Y_{2n}, X_n = x_n, Y_n = y_n] \\
&= \sum_{x_n \neq y_n} \Pr_{x_n,y_n}[X_n \neq Y_n] \cdot \Pr_{x_0,y_0}[X_n = x_n, Y_n = y_n] \\
&\leq (1-\theta) \sum_{x_n \neq y_n} \Pr_{x_0,y_0}[X_n = x_n, Y_n = y_n] \leq (1-\theta)^2,
\end{aligned}$$

and so on ($\Pr_{x_0,y_0}[X_{kn} \neq Y_{kn}] \leq (1-\theta)^k$). It yields directly that

$$\Pr[X_t \neq Y_t] = \sum_{x_0,y_0} \mu_0(x_0) \cdot \pi(y_0) \cdot \Pr_{x_0,y_0}[X_t \neq Y_t] \to 0$$

as $t \to \infty$.

14.5 Mixing Times

In fact, from the proof of FTMC, if we can bound θ (or $\Pr[X_t \neq Y_t]$), then we can show the convergence rate of Markov chains.

We are now ready to study the convergence rate of Markov chains. We start with the notion of *mixing time*. For any $\varepsilon > 0$, the mixing time of a Markov chain P up to error ε is the minimun step t such that if we run the Markov chain from *any* initial distribution, its total variation distance to the stationary distribution is at most ε. Formally,

$$\tau_{\mathsf{mix}}(\varepsilon) := \arg\min_{t \geq 0} \max_{\mu_0} D_{\mathsf{TV}}(\mu_t, \pi) \leq \varepsilon.$$

We usually denote $\tau_{\mathrm{mix}}(1/4)$ by τ_{mix}.

Recalling in our proof of FTMC, we obtain the following inequality

$$D_{\mathsf{TV}}(\mu_t, \pi) \leq \Pr_{(X_t,Y_t) \sim \omega_t}[X_t \neq Y_t].$$

Therefore, if we can construct a coupling ω_t such that $\Pr_{(X_t,Y_t) \sim \omega_t}[X_t \neq Y_t] \leq \varepsilon$, then $\tau_{\mathsf{mix}}(\varepsilon) \leq t$.

In practice, it is sufficient to assume X_t and Y_t are from two arbitrary initial distributions (Why?).

We now introduce some examples.

14.5.1 RANDOM WALKS ON HYPERCUBES

Consider the random walk on the n-cube. The state space $\Omega = \{0,1\}^n$, and we start from a point $X_0 \in \Omega$. In each step,

- With probability $\frac{1}{2}$ do nothing.
- Otherwise, pick $i \in [n]$ uniformly at random and flip $X(i)$.

It's equivalent to the following process:

- Pick $i \in [n], b \in \{0,1\}$ uniformly at random.
- Change $X(i)$ to b.

Now we analyze the mixing time of the process using coupling. The apply the following simple coupling rule:

- We couple two walks X_t and Y_t by choosing the same i, b in every step.

Once a position $i \in [n]$ has been picked, $X_t(i)$ and $Y_t(i)$ will be the same forever. Therefore, the problem again reduces to the coupon collector problem. So we immediately have

$$\tau_{\mathsf{mix}}(\varepsilon) \leq n \log \frac{n}{\varepsilon}.$$

Let's modify the process a bit by changing $\frac{1}{2}$ into $\frac{1}{n+1}$, i.e. w.p. $\frac{1}{n+1}$ do nothing, to make the *lazy* walk more active. Note that we add the lazy move in order to make the chain aperiodic.

Now in this case, we describe another coupling of X_t, Y_t. Without loss of generality, we can reorder the entries of two vectors so that all disagreeing entries come first. Namely there exists an index k such that $X_t(i) \neq Y_t(i)$ if $1 \leq i \leq k$, and $X_t(i) = Y_t(i)$ for $i > k$. Our coupling is as follows:

- If $k = 0$, Y acts the same as X.
- If $k = 1$, Y acts the same as X except when X flips the first entry, Y does nothing and vice versa.
- For $k > 2$, we distinguish between whether X flip indices in $[k]$:
  - If X did nothing or flipped one of $i > k$: Y acts the same.
  - If X flipped $1 \leq i \leq k$: Y flips $(i \bmod k) + 1$, i.e. $1 \mapsto 2, 2 \mapsto 3, \cdots, k-1 \mapsto k, k \mapsto 1$.

It's clear that the above is indeed a coupling. In fact, this coupling acts like a *doubled speed* coupon collector, since in the case $k > 2$ we can always collect two coupons at a time when lady luck is smiling. It is therefore conceivable that

$$\tau_{\mathsf{mix}} \leq \frac{1}{2} n \log n + O(n).$$

The above two examples are easy to analyze since we can reduce the coalesce time of two chains to models we are familiar with. To analyze couplings in general, we often require the coupling enjoy the property that the two chains are expected closer after every step. Therefore we impose a distance $d(\cdot,\cdot)$ between two states and require that

$$\forall t,\ \mathbb{E}[d(X_{t+1}, Y_{t+1}) | (X_t, Y_t)] \leq (1-\alpha) \cdot d(X_t, Y_t).$$

In other words, $\{d(X_t, Y_t)\}_{t\geq 0}$ is a supermartingale.

This is sufficient to imply an upper bound for the mixing time: Without loss of generality, we assume that $\min_{x,y\in\Omega:x\neq y} d(x,y) = 1$ when Ω is finite (why this is WLOG?). By coupling lemma,

$$\begin{aligned} D_{\mathsf{TV}}(X_t, Y_t) &\leq \Pr[X_t \neq Y_t] \\ &= \Pr[d(X_t, Y_t) > 0] \\ &= \Pr[d(X_t, Y_t) \geq 1] \\ &\leq \mathbb{E}[d(X_t, Y_t)] \\ &\leq (1-\alpha)^t \cdot d(X_0, Y_0) \leq \varepsilon. \end{aligned}$$

This implies

$$\tau_{\mathsf{mix}}(\varepsilon) \leq \log \frac{d(X_0, Y_0)}{\varepsilon} \cdot \log \frac{1}{1-\alpha}.$$

14.5.2 Sampling Proper Colorings

Let's consider the problem of sampling proper colorings. Given a graph $G = (V, E)$, we want to color the vertices using q colors under the condition that no two adjacent vertices share the same color. More formally, a coloring of G is a mapping $c : V \mapsto [q]$, and we call it proper iff $\forall \{u, v\} \in E, c(u) \neq c(v)$. The problem is NP-hard in general. However, for $q > \Delta$ there's always at least one suitable solution and can be easily obtained by a greedy algorithm, where Δ is the maximum degree of the graph.

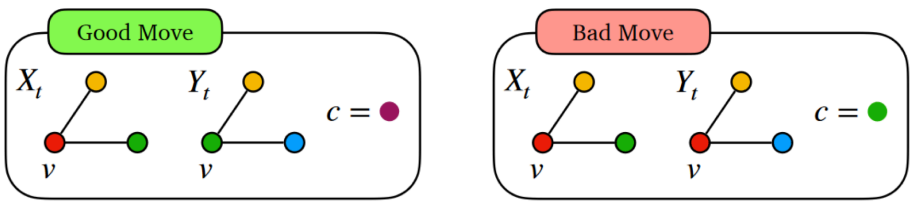
Consider the following Markov chain to sample proper colorings:

- Pick $v \in V$ and $c \in [q]$ uniformly at random.
- Recolor v with c if possible.

The chain is aperiodic since self-loops exist in the walk. For $q \geq \Delta + 2$, the chain is irreducible. The bound $q \geq \Delta + 2$ is tight for irreducibility since when $q = \Delta + 1$, each proper coloring of complete graph is frozen. It is still an open problem if the mixing time of the chain is polynomial in the size of the graph under the condition $q \geq \Delta + 2$. The best bound so far requires that $q \geq 1.809\Delta$. Here, we shall give a rapid mixing proof when $q > 4\Delta$ using the method of coupling.

Suppose X_t, Y_t are two proper colorings. We define the distance $d(X_t, Y_t)$ as their Hamming distance, i.e. the number of vertices colored differently in two colorings. Our coupling of two chains is that we always choose the same v, c in each step. The distance between two colorings can change at most 1 since only v is affected. The possible changes can be divided into two kinds:

- Good move: $X_t(v) \neq Y_t(v)$, and both change into c successfully. It will decrease distance by 1.
- Bad move: $X_t(v) = Y_t(v)$, one succeeds and one fails in the changing. It will increase distance by 1.



Consider the probabilities of two types of moves. For good moves, w.p. $\frac{d(X_t,Y_t)}{n}$, $X_t(v) \neq Y_t(v)$, and there are at least $q-2\Delta$ choices of c to make it a good move. So

$$\Pr[d(X_{t+1},Y_{t+1}) = d(X_t,Y_t)-1] = \Pr_{(v,c)\in V\times[q]}[(v,c) \text{ is a good move}] \geq \frac{d(X_t,Y_t)}{n}\cdot\frac{q-2\Delta}{q}.$$

For bad moves, there exists a neighbor w of v such that its color is different in two colorings, and in one coloring w is of color c. By a counting argument, we have

$$\Pr[d(X_{t+1},Y_{t+1}) = d(X_t,Y_t)+1] = \Pr_{(v,c)\in V\times[q]}[(v,c) \text{ is a bad move}] \leq \frac{\Delta d(X_t,Y_t)}{n}\cdot\frac{2}{q}.$$

Therefore,

$$\begin{aligned}\mathbb{E}[d(X_{t+1},Y_{t+1})|(X_t,Y_t)] &= d(X_t,Y_t) + \Pr[d(X_{t+1},Y_{t+1}) = d(X_t,Y_t)+1] - \Pr[d(X_{t+1},Y_{t+1}) = d(X_t,Y_t)-1] \\ &\leq d(X_t,Y_t) + \frac{\Delta d(X_t,Y_t)}{n}\cdot\frac{2}{q} - \frac{d(X_t,Y_t)}{n}\cdot\frac{q-2\Delta}{q} \\ &\leq d(X_t,Y_t)\Big(1-\frac{q-4\Delta}{nq}\Big).\end{aligned}$$

In the case $q > 4\Delta$,

$$D_{\mathsf{TV}} \leq \left(1-\frac{1}{nq}\right)^t n \leq \varepsilon.$$

The mixing time is therefore bounded by

$$\tau_{\mathsf{mix}}(\varepsilon) \leq nq\log\frac{n}{\varepsilon}.$$