Bounding mixing time using coupling

Definition

Consider an MC (Z_t) with state space Ω and transition matrix P.
A Markovian coupling for (Z_t) is an MC (X_t, Y_t) on $\Omega \times \Omega$, with transition probabilities defined by

$$\Pr[X_{t+1} = x' \mid X_t = x, Y_t = y] = P(x, x'),$$

$$\Pr[Y_{t+1} = y' \mid X_t = x, Y_t = y] = P(y, y').$$

Equivalently, if $\widehat{P} : \Omega^2 \to \Omega^2$ denotes the transition matrix of the coupling,

$$\sum_{y' \in \Omega} \widehat{P}((x, y), (x', y')) = P(x, x'),$$

$$\sum_{x' \in \Omega} \widehat{P}((x, y), (x', y')) = P(y, y').$$

Example 9

Consider a slightly different Markov chain for the problem of q colorings.

1. Choose a vertex v u.a.r.
2. Choose a color $c \in Q \setminus X_t(N(v))$ u.a.r.
3. Recolor v with c and leave all the other colored vertices the same.

Example 9

Consider a slightly different Markov chain for the problem of q colorings.

1. Choose a vertex v u.a.r.
2. Choose a color $c \in Q \setminus X_t(N(v))$ u.a.r.
3. Recolor v with c and leave all the other colored vertices the same.

We will discuss some possible transitions of a coupling (X_t, Y_t) on Ω^2.

Example 9

- Suppose $Q = \{0, 1, ..., 6\}$.
- At t-th step, the same vertex v is chosen u.a.r. for both transitions.
- Suppose that a vertex v has been chosen for which $X_t(N(v)) = \{3, 6\}$ and $Y_t(N(v)) = \{4, 5, 6\}$ hold.
- So, the legal colors for v in X_{t+1} and Y_{t+1} are $c_x \in \{0, 1, 2, 4, 5\}$ and $c_y = \{0, 1, 2, 3\}$, repsectively.

Example 9

- So, the legal colors for v in X_{t+1} and Y_{t+1} are $c_x \in \{0, 1, 2, 4, 5\}$ and $c_y = \{0, 1, 2, 3\}$, repsectively.
  1. First option: $\Pr(c_x, c_y) = \frac{1}{5} \cdot \frac{1}{4} = \frac{1}{20}$.

Example 9

- So, the legal colors for v in X_{t+1} and Y_{t+1} are $c_x \in \{0, 1, 2, 4, 5\}$ and $c_y = \{0, 1, 2, 3\}$, repsectively.
  1. First option: $\Pr(c_x, c_y) = \frac{1}{5} \cdot \frac{1}{4} = \frac{1}{20}$.

$$\Pr[X_{t+1}(v) = 1] = \sum_{c_y} \Pr(1, c_y) =$$
$$\Pr(1,0) + \Pr(1,1) + \Pr(1,2) + \Pr(1,3) = \frac{1}{5}$$

Example 9

- So, the legal colors for v in X_{t+1} and Y_{t+1} are $c_x \in \{0, 1, 2, 4, 5\}$ and $c_y = \{0, 1, 2, 3\}$, repsectively.
  1. First option: $\Pr(c_x, c_y) = \frac{1}{5} \cdot \frac{1}{4} = \frac{1}{20}$.

     $$\Pr[X_{t+1}(v) = 1] = \sum_{c_y} \Pr(1, c_y) =$$
     $$\Pr(1, 0) + \Pr(1, 1) + \Pr(1, 2) + \Pr(1, 3) = \frac{1}{5}$$

  2. Second option:

     $$\Pr(0, 0) = \Pr(1, 1) = \Pr(2, 2) = \frac{1}{5}$$
     $$\Pr(4, 3) = \frac{1}{5}$$
     $$\Pr(5, c_y) = \frac{1}{20} \text{ for every } c_y \in \{0, 1, 2, 3\}$$

Example 9

- So, the legal colors for v in X_{t+1} and Y_{t+1} are $c_x \in \{0, 1, 2, 4, 5\}$ and $c_y = \{0, 1, 2, 3\}$, repsectively.
  1. First option: $\Pr(c_x, c_y) = \frac{1}{5} \cdot \frac{1}{4} = \frac{1}{20}$.

     $$\Pr[X_{t+1}(v) = 1] = \sum_{c_y} \Pr(1, c_y) =$$
     $$\Pr(1,0) + \Pr(1,1) + \Pr(1,2) + \Pr(1,3) = \tfrac{1}{5}$$

  2. Second option:

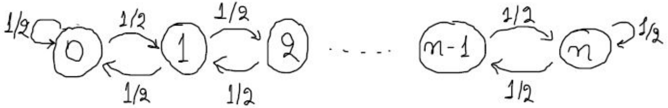
     $$\Pr(0,0) = \Pr(1,1) = \Pr(2,2) = \frac{1}{5}$$
     $$\Pr(4,3) = \frac{1}{5}$$
     $$\Pr(5, c_y) = \frac{1}{20} \text{ for every } c_y \in \{0, 1, 2, 3\}$$

     $$\Pr[Y_{t+1}(v) = 0] = \sum_{c_x} \Pr(c_x, 0) = \Pr(0,0) + \Pr(5,0) = \frac{1}{5} + \frac{1}{20} = \frac{1}{4}$$

Example 10

Simple random walk on $\{0, 1, ..., n\}$

- The transition graph of (Z_t) is the following.



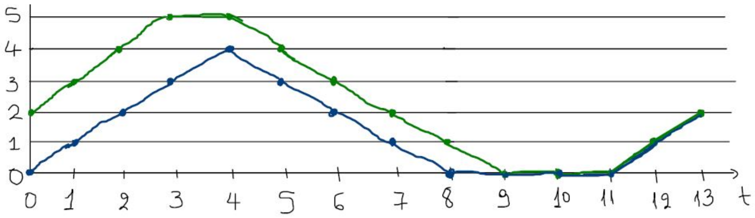
- Add either $+1$ or -1, each with probability $1/2$, to the current state if possible.

- Do nothing if attempt to add either -1 to 0, or $+1$ to n.

Example 10

A coupling (X_t, Y_t) for (Z_t) starting in (x, y):

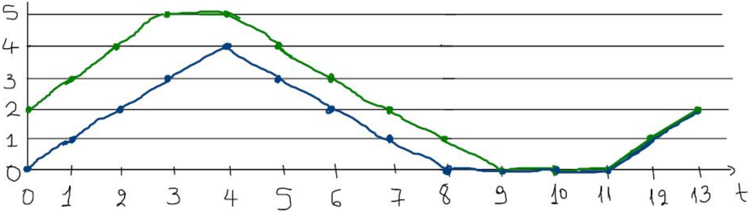
- $X_0 = x, \; Y_0 = y$.
- At the (t+1)-th step, choose $b_{t+1} \in \{-1, 1\}$ u.a.r.
- Attempt to add b_{t+1} to both X_t and Y_t.



Example 10

A coupling (X_t, Y_t) for (Z_t) starting in (x, y):

- $X_0 = x,\ Y_0 = y$.
- At the (t+1)-th step, choose $b_{t+1} \in \{-1, 1\}$ u.a.r.
- Attempt to add b_{t+1} to both X_t and Y_t.



Note: We can modify any coupling so that the chains stay together after the first time they meet.

Coupling lemma

Let (X_t, Y_t) be any coupling for (Z_t) on Ω. Suppose $t : [0,1] \to \mathbb{N}$ is a function satisfying the condition: for all $x, y \in \Omega$ and all $\varepsilon > 0$

$$\Pr[X_{t(\varepsilon)} \neq Y_{t(\varepsilon)} \mid X_0 = x, Y_0 = y] \leq \varepsilon.$$

Then the mixing time $\tau(\varepsilon)$ of (Z_t) is bounded by $t(\varepsilon)$.

Coupling lemma

Let (X_t, Y_t) be any coupling for (Z_t) on Ω. Suppose $t : [0,1] \to \mathbb{N}$ is a function satisfying the condition: for all $x, y \in \Omega$ and all $\varepsilon > 0$

$$\Pr[X_{t(\varepsilon)} \neq Y_{t(\varepsilon)} \mid X_0 = x, Y_0 = y] \leq \varepsilon.$$

Then the mixing time $\tau(\varepsilon)$ of (Z_t) is bounded by $t(\varepsilon)$.

Proof.

- Let P be the transition matrix of (Z_t). Let $A \subseteq \Omega$ be arbitrary.
- Let $x \in \Omega$ be fixed, and Y_0 be chosen according to the stationary distribution π of (Z_t).

Coupling lemma

Let (X_t, Y_t) be any coupling for (Z_t) on Ω. Suppose $t : [0,1] \to \mathbb{N}$ is a function satisfying the condition: for all $x, y \in \Omega$ and all $\varepsilon > 0$

$$\Pr[X_{t(\varepsilon)} \neq Y_{t(\varepsilon)} \mid X_0 = x, Y_0 = y] \leq \varepsilon.$$

Then the mixing time $\tau(\varepsilon)$ of (Z_t) is bounded by $t(\varepsilon)$.

Proof.

- Let P be the transition matrix of (Z_t). Let $A \subseteq \Omega$ be arbitrary.
- Let $x \in \Omega$ be fixed, and Y_0 be chosen according to the stationary distribution π of (Z_t).
- For any $\varepsilon \in (0,1)$ and the corresponding $t = t(\varepsilon)$, we are going to prove that
  1. $P^t(y, A) - P^t(x, A) \leq \varepsilon$ for any $x, y \in \Omega$.
  2. $\pi(A) - P^t(x, A) \leq \varepsilon$ for any $x \in \Omega$.

Coupling lemma

Let (X_t, Y_t) be any coupling for (Z_t) on Ω. Suppose $t : [0,1] \to \mathbb{N}$ is a function satisfying the condition: for all $x, y \in \Omega$ and all $\varepsilon > 0$

$$\Pr[X_{t(\varepsilon)} \neq Y_{t(\varepsilon)} \mid X_0 = x, Y_0 = y] \leq \varepsilon.$$

Then the mixing time $\tau(\varepsilon)$ of (Z_t) is bounded by $t(\varepsilon)$.

Proof.

- Let P be the transition matrix of (Z_t). Let $A \subseteq \Omega$ be arbitrary.
- Let $x \in \Omega$ be fixed, and Y_0 be chosen according to the stationary distribution π of (Z_t).
- For any $\varepsilon \in (0,1)$ and the corresponding $t = t(\varepsilon)$, we are going to prove that
  1. $P^t(y, A) - P^t(x, A) \leq \varepsilon$ for any $x, y \in \Omega$.
  2. $\pi(A) - P^t(x, A) \leq \varepsilon$ for any $x \in \Omega$.
- Since A is arbitrary, by the definition of total variation distance $||\mu - \nu||_{TV} = \max_{A \subseteq \Omega} |\mu(A) - \nu(A)|$ we have that

$$||P^t(x, \cdot) - \pi||_{TV} \leq \varepsilon.$$

Proof cont. For any $\varepsilon \in (0,1)$ and the corresponding $t = t(\varepsilon)$,

$$\begin{aligned}
P^t(x, A) &= \Pr[X_t \in A] \\
&\geq \Pr[X_t = Y_t \wedge Y_t \in A] \\
&= 1 - \Pr[X_t \neq Y_t \vee Y_t \notin A] \\
&\geq 1 - (\Pr[X_t \neq Y_t] + \Pr[Y_t \notin A]) \\
&\geq \Pr(Y_t \in A) - \varepsilon \\
&= P^t(y, A) - \varepsilon \\
&= \pi(A) - \varepsilon.
\end{aligned}$$

□

Bounding the mixing time of the MC

Theorem

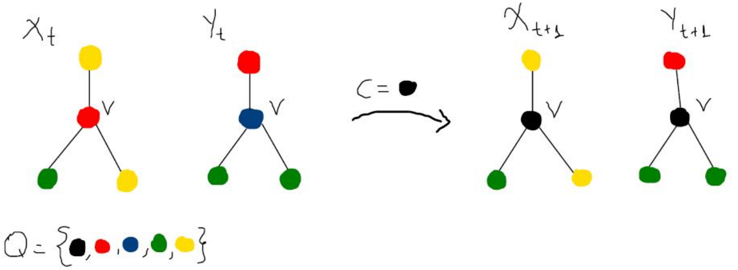
The mixing time of the above MC is $\tau_{mix} = \mathcal{O}(n \log n)$ *for* $q \geq 4\Delta + 1$.

Proof.

- We choose arbitrary colorings X_0 and Y_0 of G.
- We couple (X_t, Y_t) by picking the same vertex v and color c u.a.r. at all times t.
- We denote by D_t be the number of vertices on which the colorings X_t and Y_t disagree.

Proof cont. There are three types of possible moves: good moves, bad moves, and neutral moves.

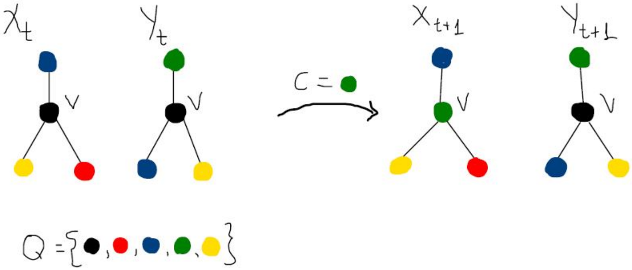
1. Good moves ($D_{t+1} = D_t - 1$): v has different colors in X_t and Y_t, and c does not appear in the neighborhood of v in either X_t or Y_t.



$$\Pr[D_{t+1} = D_t - 1] \geq \frac{D_t}{n} \cdot \frac{q - 2\Delta}{q}$$

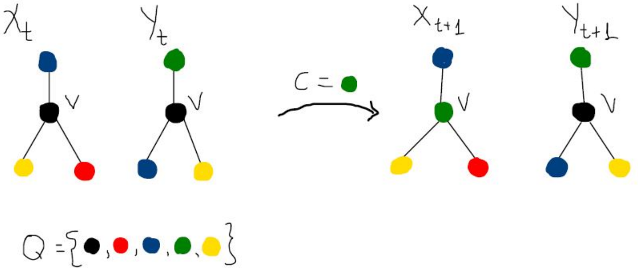
Proof cont.

2. Bad moves ($D_{t+1} = D_t + 1$): v has the same color in X_t and Y_t, and c appears among the neighbors of v in exactly one of X_t or Y_t.



Proof cont.

2. Bad moves ($D_{t+1} = D_t + 1$): v has the same color in X_t and Y_t, and c appears among the neighbors of v in exactly one of X_t or Y_t.



- v is a neighbor of a disagreement vertex u and c is the color of u in one of the chains.
- The disagreement vertices have at most $D_t \cdot \Delta$ neighbors, and for any such neighbor there are at most 2 bad colors.

$$\Pr[D_{t+1} = D_t + 1] \leq \frac{D_t \cdot \Delta}{n} \cdot \frac{2}{q}$$

Proof cont.

3. Neutral moves ($D_{t+1} = D_t$): In any other move D_t remains invariant.

Proof cont.

3. Neutral moves ($D_{t+1} = D_t$): In any other move D_t remains invariant.

$$\mathbb{E}[D_{t+1} \mid D_t] = (D_t - 1) \cdot \Pr[D_{t+1} = D_t - 1] + (D_t + 1) \cdot \Pr[D_{t+1} = D_t + 1] + D_t \cdot (1 - \Pr[D_{t+1} = D_t + 1] - \Pr[D_{t+1} = D_t - 1])$$

Proof cont.

3. Neutral moves ($D_{t+1} = D_t$): In any other move D_t remains invariant.

$$\begin{aligned}
\mathbb{E}[D_{t+1} \mid D_t] =& (D_t - 1) \cdot \Pr[D_{t+1} = D_t - 1] + (D_t + 1) \cdot \Pr[D_{t+1} = D_t + 1] \\
&+ D_t \cdot (1 - \Pr[D_{t+1} = D_t + 1] - \Pr[D_{t+1} = D_t - 1]) \\
=& D_t - \Pr[D_{t+1} = D_t - 1] + \Pr[D_{t+1} = D_t + 1] \\
\leq& D_t - \frac{D_t(q - 2\Delta)}{qn} + \frac{2D_t\Delta}{nq} \\
=& D_t\left(1 - \frac{q - 4\Delta}{qn}\right)
\end{aligned}$$

Proof cont.

3. Neutral moves ($D_{t+1} = D_t$): In any other move D_t remains invariant.

$$\begin{aligned}
\mathbb{E}[D_{t+1} \mid D_t] =& (D_t - 1) \cdot \Pr[D_{t+1} = D_t - 1] + (D_t + 1) \cdot \Pr[D_{t+1} = D_t + 1] \\
&+ D_t \cdot (1 - \Pr[D_{t+1} = D_t + 1] - \Pr[D_{t+1} = D_t - 1]) \\
=& D_t - \Pr[D_{t+1} = D_t - 1] + \Pr[D_{t+1} = D_t + 1] \\
\leq& D_t - \frac{D_t(q - 2\Delta)}{qn} + \frac{2D_t\Delta}{nq} \\
=& D_t\left(1 - \frac{q - 4\Delta}{qn}\right)
\end{aligned}$$

where $0 < 1 - \frac{q-4\Delta}{qn} < 1$, since $q > 4\Delta$.

Proof cont. By taking expectation on both sides and iterating, we have that

$$\begin{aligned}
\mathbb{E}[D_t \mid D_0] &\le D_0\Big(1 - \frac{q-4\Delta}{qn}\Big)^t \\
&\le n\Big(1 - \frac{q-4\Delta}{qn}\Big)^t \\
&\le n\exp\Big(-\frac{q-4\Delta}{qn}\cdot t\Big) && \text{since } (1-x)^n \le e^{-nx} \\
&\le \varepsilon && \text{when } t \ge \frac{q}{q-4\Delta} n(\log n + \log \varepsilon^{-1})
\end{aligned}$$

Proof cont.

- By Markov's inequality $\Pr[X \geq a] \leq \frac{\mathbb{E}[X]}{a}$, we have that

$$\Pr[X_t \neq Y_t \mid (X_0, Y_0)] = \Pr[D_t \geq 1 \mid D_0] \leq \mathbb{E}[D_t \mid D_0]$$
$$\leq n \exp\left(-\frac{q - 4\Delta}{qn} \cdot t\right) \leq \varepsilon$$

for $t \geq \frac{q}{q-4\Delta} n(\log n + \log \varepsilon^{-1})$.

Proof cont.

- By Markov's inequality $\Pr[X \geq a] \leq \frac{\mathbb{E}[X]}{a}$, we have that

$$\Pr[X_t \neq Y_t \mid (X_0, Y_0)] = \Pr[D_t \geq 1 \mid D_0] \leq \mathbb{E}[D_t \mid D_0]$$
$$\leq n \exp\left(-\frac{q - 4\Delta}{qn} \cdot t\right) \leq \varepsilon$$

for $t \geq \frac{q}{q-4\Delta} n(\log n + \log \varepsilon^{-1})$.

- By the Coupling lemma, the following holds for mixing time of the Markov chain

$$\tau(\varepsilon) = \frac{q}{q - 4\Delta} n(\log n + \log \varepsilon^{-1})$$

$$\tau_{mix} = \mathcal{O}\left(\frac{q}{q - 4\Delta} n \log n\right)$$

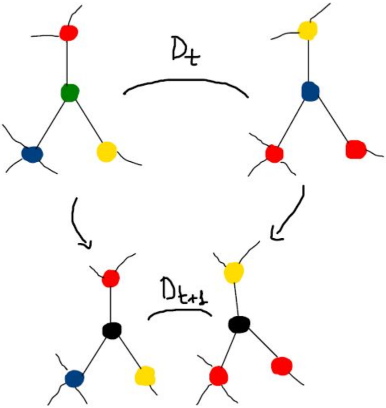
for $q \geq 4\Delta + 1$.

□

Contraction in D_t

We showed contraction in one step: for some $\alpha > 0$

$$\mathbb{E}[D_{t+1} \mid D_t] \leq D_t e^{-\alpha} \Rightarrow t_{mix}(\varepsilon) \leq \frac{\log n + \log \varepsilon^{-1}}{\alpha}$$



Lemma

Let Z_t be an MC on Ω and let $d : \Omega \times \Omega \to \mathbb{N}$ be a metric. Suppose that there is a coupling (X_t, Y_t) such that for all $x, y \in \Omega$

$$\mathbb{E}[d(X_{t+1}, Y_{t+1}) \mid X_t = x, Y_t = y] \leq (1-\alpha)d(x,y) \text{ for } \alpha < 1.$$

Then, $\tau(\varepsilon) \leq \alpha^{-1} \log \frac{D}{\varepsilon}$, where D is the diameter of Ω under d.

The case of $q > 2\Delta$

- The metric d does not need to be defined on $\Omega \times \Omega$, but can be extended.
- Using path coupling, we are going to prove the following theorem.

Theorem

Let G have max degree Δ. If $q > 2\Delta$, the mixing time of the Metropolis chain on colorings is

$$t_{mix}(\varepsilon) \leq \left\lceil \left(\frac{q}{q - 2\Delta}\right) n(\log n + \log \varepsilon^{-1}) \right\rceil.$$

Path coupling (Bubley & Dyer 1997)

- We define a **connected graph** (Ω, E_0).
- **Length function** $\ell : E_0 \to [1, \infty)$.
- A **path** from x_0 to x_r is $\xi = (x_0, x_1, ..., x_r)$ such that $(x_{i-1}, x_i) \in E_0$.
- The **length of path** ξ is defined as $\ell(\xi) := \sum_{i=1}^{r} \ell(x_{i-1}, x_i)$.
- We are considering the **shortest path metric** ρ on Ω

$$\rho(x, y) := \min\{\ell(\xi) \mid \xi \text{ is a path between } x, y\}.$$

  - ρ satisfies triangle inequality since

$$\text{shortest path}(x, y) \leq \text{shortest path}(x, z) + \text{shortest path}(z, y).$$

Theorem

Let Z_t be an MC on Ω and let $\rho : \Omega \times \Omega \to \mathbb{N}$ be the shortest path metric. Suppose that there exists a coupling (X_t, Y_t) defined for all adjacent pair of states in the graph (Ω, E_0) such that for all adjacent X_t, Y_t

$$\mathbb{E}[\rho(X_{t+1}, Y_{t+1}) \mid X_t, Y_t] \leq (1-\alpha)\rho(X_t, Y_t) \text{ for } \alpha < 1.$$

Theorem

Let Z_t be an MC on Ω and let $\rho : \Omega \times \Omega \to \mathbb{N}$ be the shortest path metric. Suppose that there exists a coupling (X_t, Y_t) defined for all adjacent pair of states in the graph (Ω, E_0) such that for all adjacent X_t, Y_t

$$\mathbb{E}[\rho(X_{t+1}, Y_{t+1}) \mid X_t, Y_t] \leq (1-\alpha)\rho(X_t, Y_t) \text{ for } \alpha < 1.$$

Then this coupling can be extended to a coupling between all pairs of states that also satisfies the above inequality, so

$$\tau_{mix}(\varepsilon) \leq \frac{\log D + \log \varepsilon^{-1}}{\alpha}$$

where $D = \max_{x,y} \rho(x, y)$.

Definition of the MC on extended state space

- Let $\tilde{\Omega}$ be the set of all colorings of G (both proper and improper ones).
- Metropolis MC:
  - Select $v \in V$ and $c \in Q$ u.a.r.
  - If c is allowed at v, update.
- The stationary distribution of this MC defined on $\tilde{\Omega}$ is still the uniform distribution on Ω, i.e. the set of proper colorings.

Path coupling for colorings

- Define $(\tilde{\Omega}, E_0)$ such that $(x, y) \in E_0$ iff they differ at one vertex.
- Let $\ell(x, y) = 1$ for $(x, y) \in E_0$.
- Note that the diameter of $\tilde{\Omega}$ under ρ is n.

Definition of the coupling

- We need to define the coupling only for adjacent states in $\tilde{\Omega}$.

Definition of the coupling

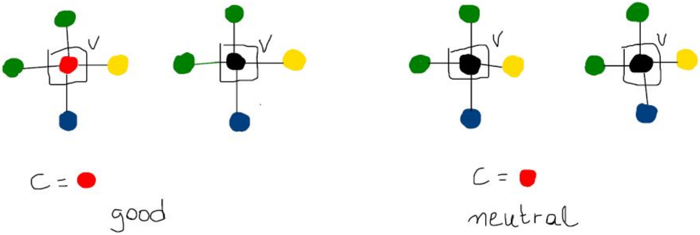
- We need to define the coupling only for adjacent states in $\tilde{\Omega}$.
- Let u be the vertex that has different colors in X_t and Y_t.

Definition of the coupling

- We need to define the coupling only for adjacent states in $\tilde{\Omega}$.
- Let u be the vertex that has different colors in X_t and Y_t.
- If **$\boldsymbol{v}$ is not in the neighborhood of $\boldsymbol{u}$**, follow the previous idea.

Definition of the coupling

- We need to define the coupling only for adjacent states in $\tilde{\Omega}$.
- Let u be the vertex that has different colors in X_t and Y_t.
- If ***v* is not in the neighborhood of *u***, follow the previous idea.
- Then, we have the following good and neutral moves.

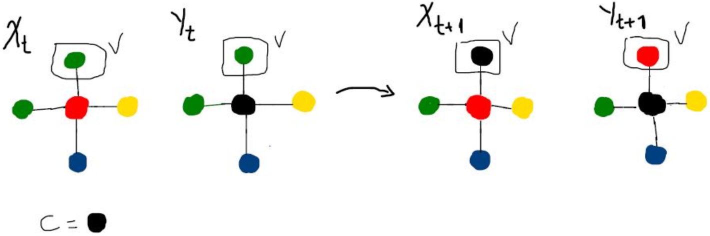


- The probability that a good move is made is $\geq \frac{1}{n} \cdot \frac{q-\Delta}{q}$.

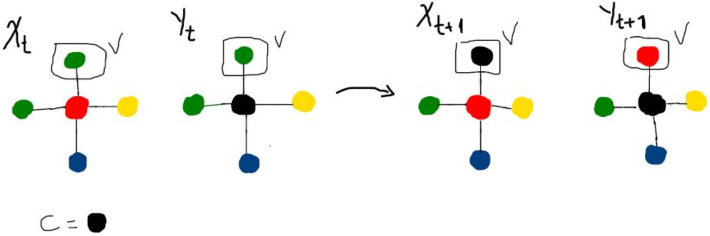
- If v is a neighbor of u, we modify the coupling.

- If v is a neighbor of u, we modify the coupling.
- If ***v* is in the neighborhood of *u***, then
  - If $c \notin \{X_t(u), Y_t(u)\}$ attempt to update v with c.
  - Otherwise, attempt to update v with c in X_t, and with the color $\{X_t(u), Y_t(u)\} \setminus \{c\}$ in Y_t.

- If v is a neighbor of u, we modify the coupling.
- If **v is in the neighborhood of u**, then
  - If $c \notin \{X_t(u), Y_t(u)\}$ attempt to update v with c.
  - Otherwise, attempt to update v with c in X_t, and with the color $\{X_t(u), Y_t(u)\} \setminus \{c\}$ in Y_t.
- Then, we have the following bad moves.



- If v is a neighbor of u, we modify the coupling.
- If **v is in the neighborhood of u**, then
  - If $c \notin \{X_t(u), Y_t(u)\}$ attempt to update v with c.
  - Otherwise, attempt to update v with c in X_t, and with the color $\{X_t(u), Y_t(u)\} \setminus \{c\}$ in Y_t.
- Then, we have the following bad moves.



- The probability that a bad move is made is $\leq \frac{\Delta}{n} \cdot \frac{1}{q}$.

Mixing time

- $\mathbb{E}[\rho(X_{t+1}, Y_{t+1})] \leq \rho(X_t, Y_t) - \frac{q-\Delta}{qn} + \frac{\Delta}{qn} = 1 - \frac{q-2\Delta}{qn} \leq 1 - \frac{1}{qn}$, since $q > 2\Delta$.

- $\tau_{mix} = \mathcal{O}(qn \log n)$.

$\square$