Advanced Algorithms (XIV)

Shanghai Jiao Tong University

Chihao Zhang

June 8, 2020

Mixing Time via Coupling

Mixing Time via Coupling

The state space Ω

Mixing Time via Coupling

The state space Ω

Transition matrix $P \in \mathbb{R}^{\Omega \times \Omega}$

Mixing Time via Coupling

The state space Ω

Transition matrix $P \in \mathbb{R}^{\Omega \times \Omega}$

Two chains $(X_0, X_1, \ldots)$ and $(Y_0, Y_1, \ldots)$

Mixing Time via Coupling

The state space Ω

Transition matrix $P \in \mathbb{R}^{\Omega \times \Omega}$

Two chains $(X_0, X_1, \ldots)$ and $(Y_0, Y_1, \ldots)$

A distance $d : \Omega \times \Omega \to \mathbb{R}_{\geq 0}$

Mixing Time via Coupling

The state space Ω

Transition matrix $P \in \mathbb{R}^{\Omega \times \Omega}$

Two chains $(X_0, X_1, \ldots)$ and $(Y_0, Y_1, \ldots)$

A distance $d : \Omega \times \Omega \to \mathbb{R}_{\geq 0}$

Two chains are "coupled" so that:

Mixing Time via Coupling

The state space Ω

Transition matrix $P \in \mathbb{R}^{\Omega \times \Omega}$

Two chains $(X_0, X_1, \ldots)$ and $(Y_0, Y_1, \ldots)$

A distance $d : \Omega \times \Omega \to \mathbb{R}_{\geq 0}$

Two chains are "coupled" so that:

$$\mathbf{E}[d(X_{t+1}, Y_{t+1}) \mid (X_t, Y_t)] \leq (1 - \alpha) \cdot d(X_t, Y_t)$$

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

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

Recall the mixing time

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

Recall the mixing time

The mixing time $\tau_{\text{mix}}(\varepsilon)$ is the first time t such that the total variation distance between X_t and π is at most ε, for any initial X_0

$$\tau_{\text{mix}}(\varepsilon) = \max_{\mu_0} \min_{t \geq 0} d_{\text{TV}}(\mu_0^T P^t, \pi) \leq \varepsilon$$

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

Recall the mixing time

The mixing time $\tau_{\text{mix}}(\varepsilon)$ is the first time t such that the total variation distance between X_t and π is at most ε, for any initial X_0

$$\tau_{\text{mix}}(\varepsilon) = \max_{\mu_0} \min_{t\geq 0} d_{\text{TV}}(\mu_0^T P^t, \pi) \leq \varepsilon$$

By coupling lemma

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

Recall the mixing time

The mixing time $\tau_{\text{mix}}(\varepsilon)$ is the first time t such that the total variation distance between X_t and π is at most ε, for any initial X_0

$$\tau_{\text{mix}}(\varepsilon) = \max_{\mu_0} \min_{t \geq 0} d_{\text{TV}}(\mu_0^T P^t, \pi) \leq \varepsilon$$

By coupling lemma

$$d_{\text{TV}}(X_t, Y_t) \leq \mathbf{Pr}[X_t \neq Y_t] = \mathbf{Pr}[d(X_t, Y_t) > 0]$$

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

Recall the mixing time

The mixing time $\tau_{\text{mix}}(\varepsilon)$ is the first time t such that the total variation distance between X_t and π is at most ε, for any initial X_0

$$\tau_{\text{mix}}(\varepsilon) = \max_{\mu_0} \min_{t\geq 0} d_{\text{TV}}(\mu_0^T P^t, \pi) \leq \varepsilon$$

By coupling lemma

$$d_{\text{TV}}(X_t, Y_t) \leq \mathbf{Pr}[X_t \neq Y_t] = \mathbf{Pr}[d(X_t, Y_t) > 0]$$

For finite Ω, we assume WLOG that $\min_{x,y\in\Omega: x\neq y} d(x, y) = 1$

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

Recall the mixing time

> The mixing time $\tau_{\text{mix}}(\varepsilon)$ is the first time t such that the total variation distance between X_t and π is at most ε, for any initial X_0
>
> $$\tau_{\text{mix}}(\varepsilon) = \max_{\mu_0} \min_{t\geq 0} d_{\text{TV}}(\mu_0^T P^t, \pi) \leq \varepsilon$$

By coupling lemma

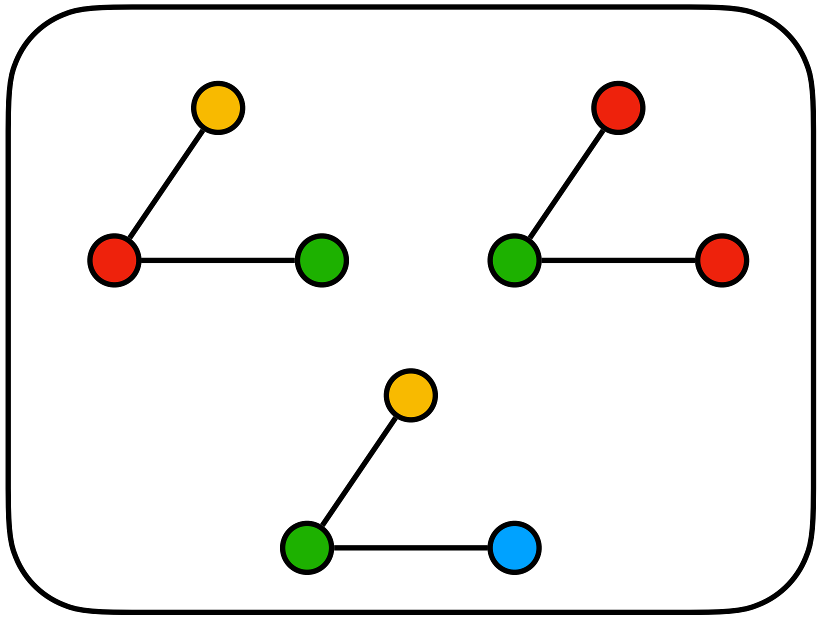
$$d_{\text{TV}}(X_t, Y_t) \leq \mathbf{Pr}[X_t \neq Y_t] = \mathbf{Pr}[d(X_t, Y_t) > 0]$$

For finite Ω, we assume WLOG that $\min_{x,y\in\Omega:x\neq y} d(x, y) = 1$

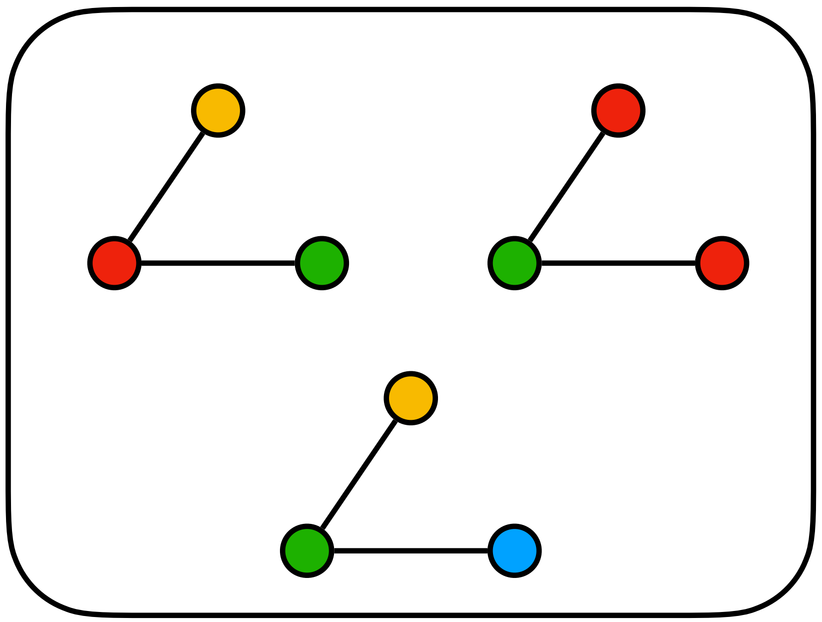
$$\begin{aligned}\mathbf{Pr}[d(X_t, Y_t) > 0] &= \mathbf{Pr}[d(X_t, Y_t) \geq 1] \\ &\leq \mathbf{E}[d(X_t, Y_t)] \leq (1-\alpha)^t \cdot d(X_0, Y_0)\end{aligned}$$

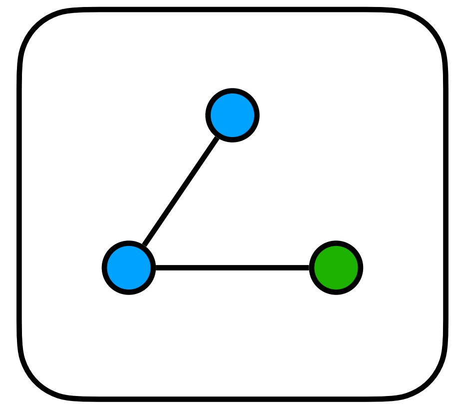
Sampling Proper Colorings

Sampling Proper Colorings

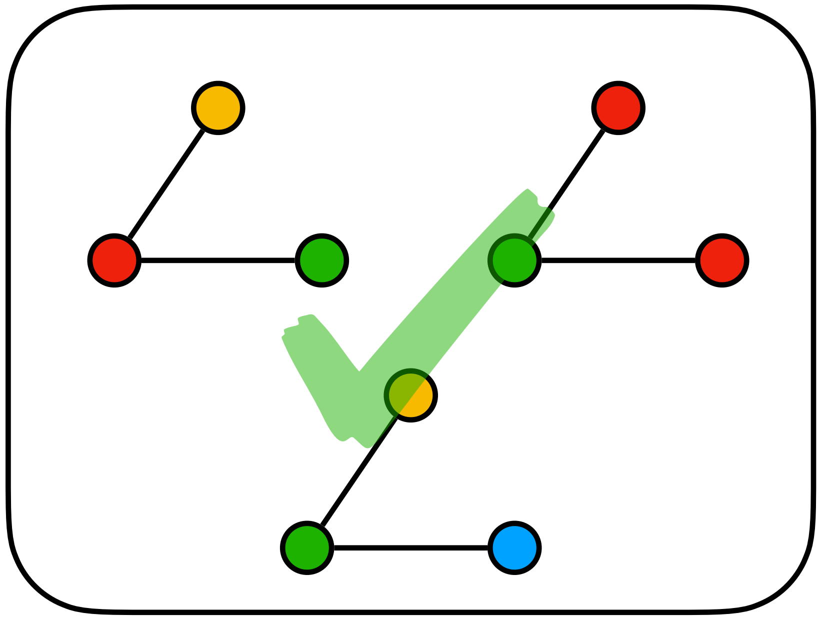


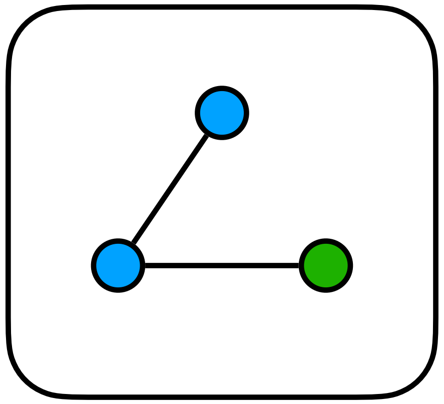
Sampling Proper Colorings



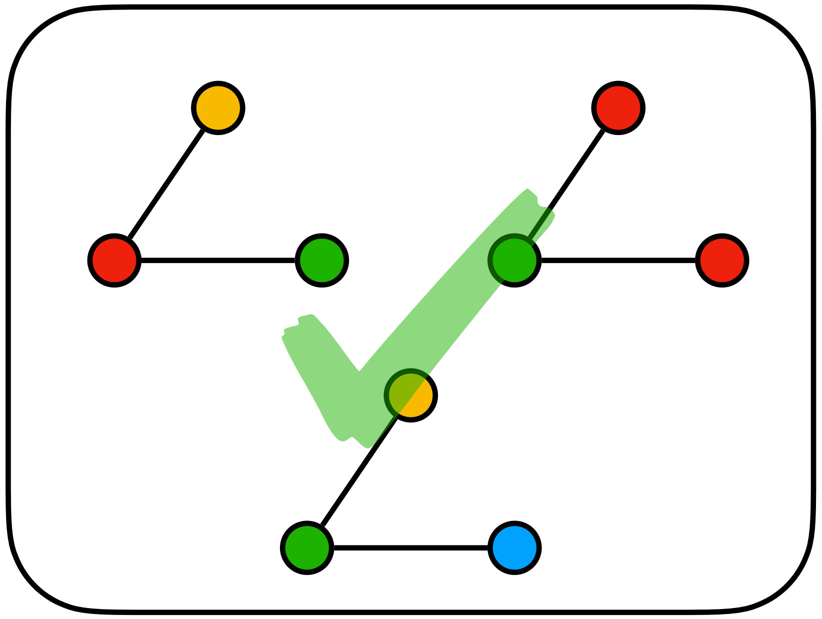


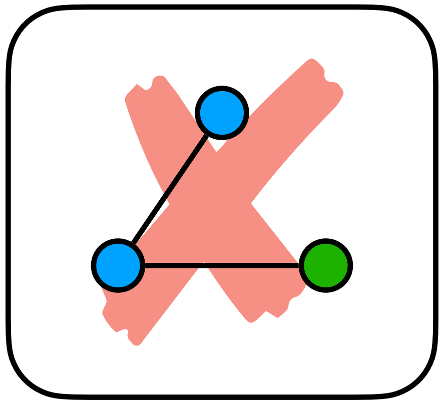
Sampling Proper Colorings



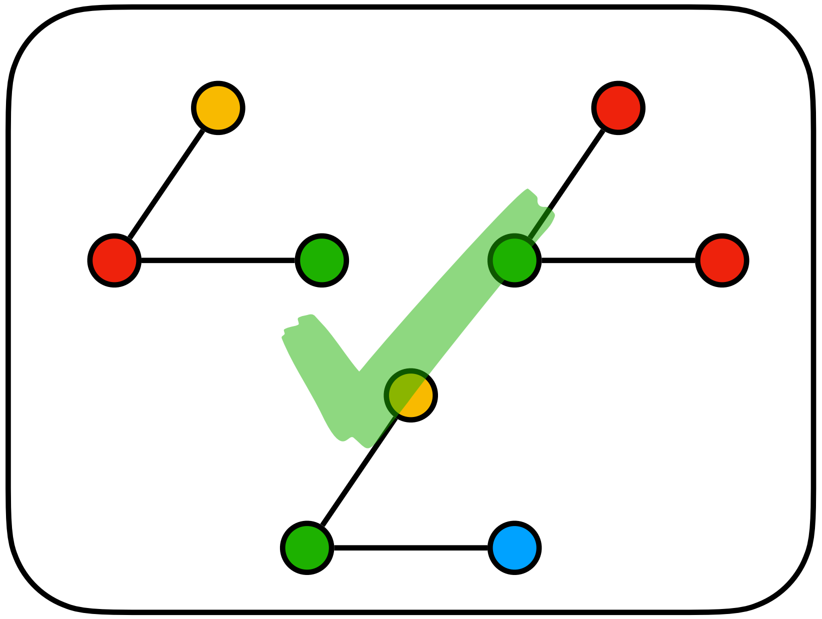


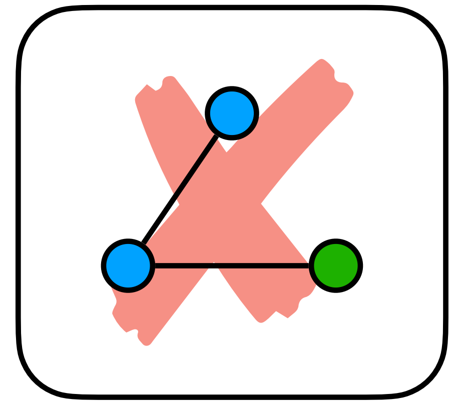
Sampling Proper Colorings





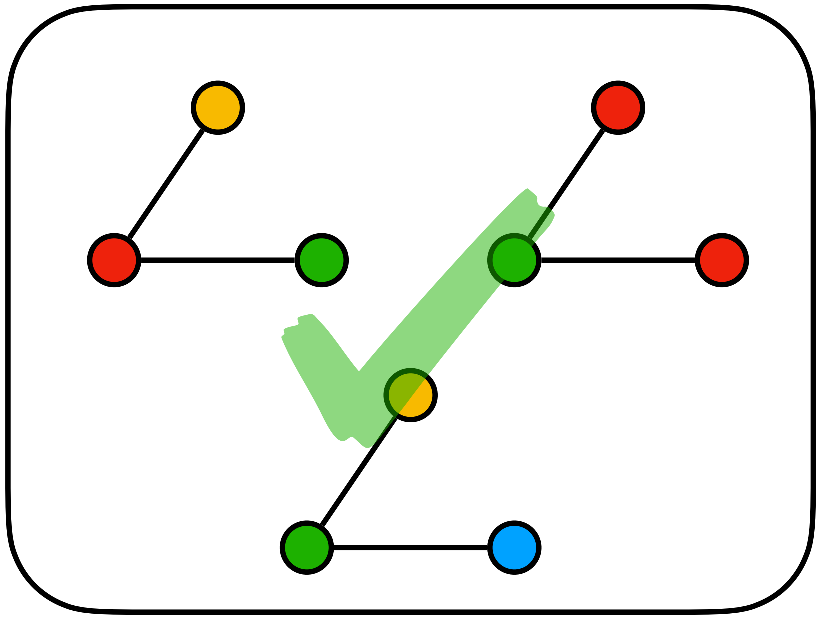
Sampling Proper Colorings

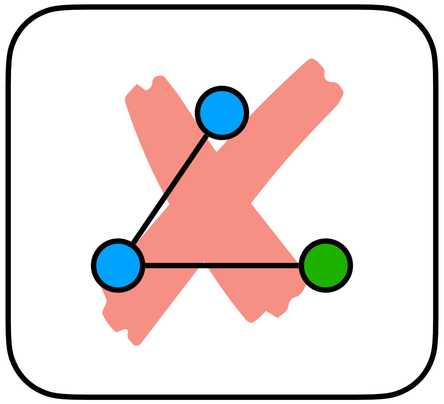




q - the number of proper colorings

Sampling Proper Colorings

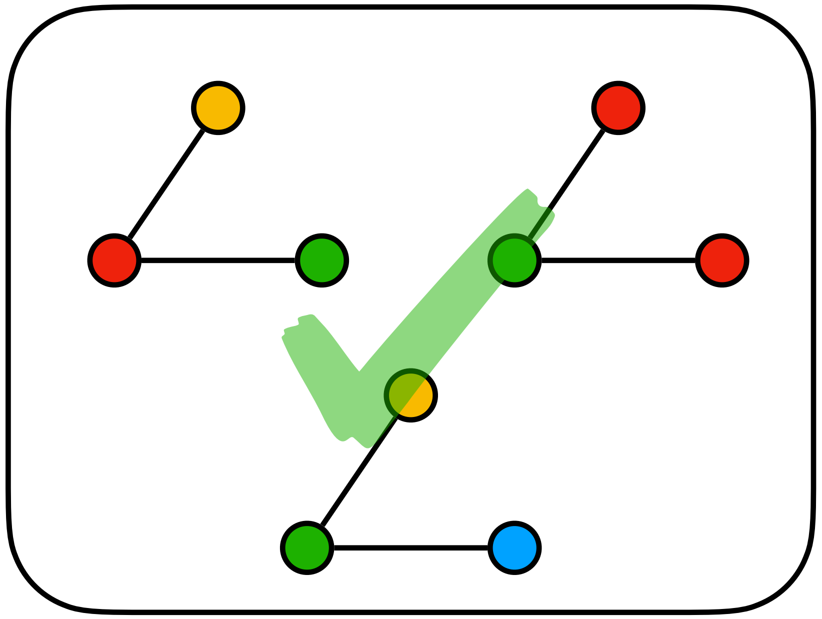


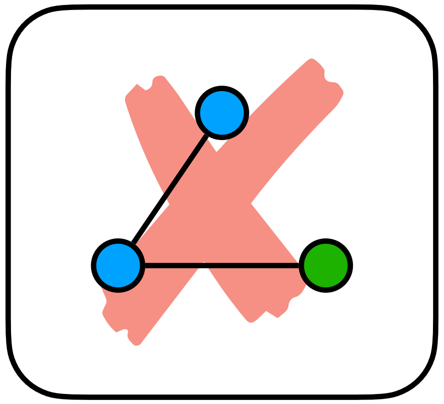


q - the number of proper colorings

G - a graph of maximum degree Δ

Sampling Proper Colorings





q - the number of proper colorings

G - a graph of maximum degree Δ

Is G colorable using q colors?

The problem is **NP**-hard in general

The problem is **NP**-hard in general

We consider the case when $q > \Delta$

The problem is **NP**-hard in general

We consider the case when $q > \Delta$

Consider the chain obtained via the "Metropolis Rule"

The problem is **NP**-hard in general

We consider the case when $q > \Delta$

Consider the chain obtained via the "Metropolis Rule"

- Pick $v \in V$ and $c \in [q]$ u.a.r.
- Recolor v with c if possible

The problem is **NP**-hard in general

We consider the case when $q > \Delta$

Consider the chain obtained via the "Metropolis Rule"

- Pick $v \in V$ and $c \in [q]$ u.a.r.
- Recolor v with c if possible

The chain is irreducible when $q \geq \Delta + 2$

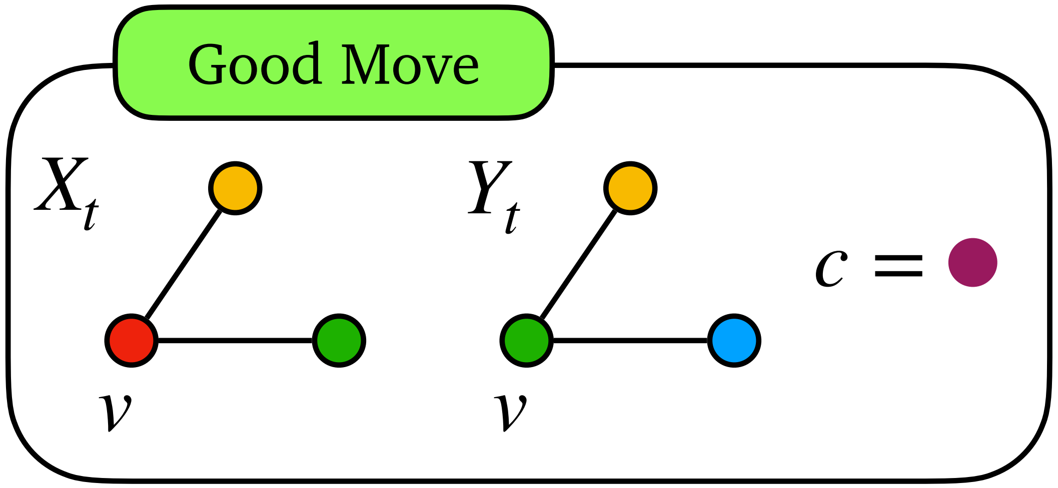
The Coupling

The Coupling

Two chains choose the same v and c

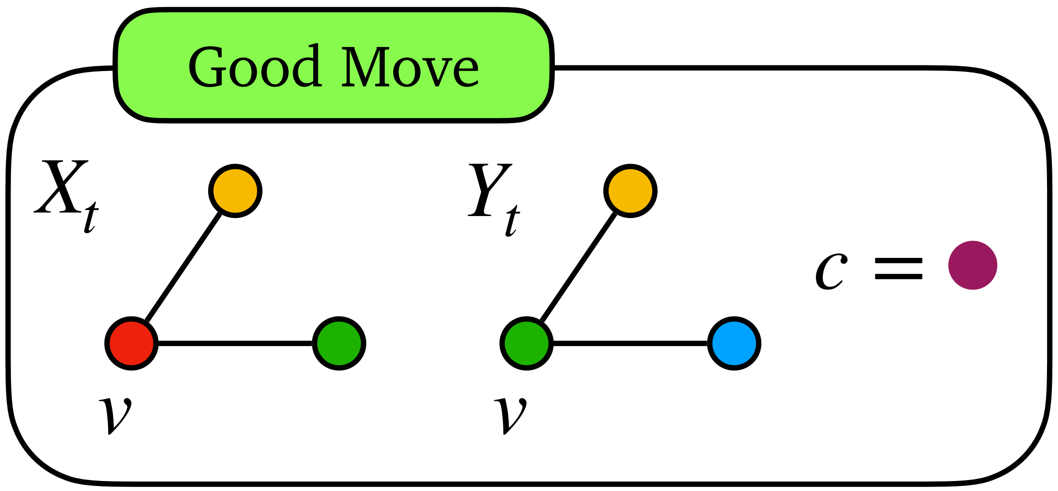
The Coupling

Two chains choose the same v and c



The Coupling

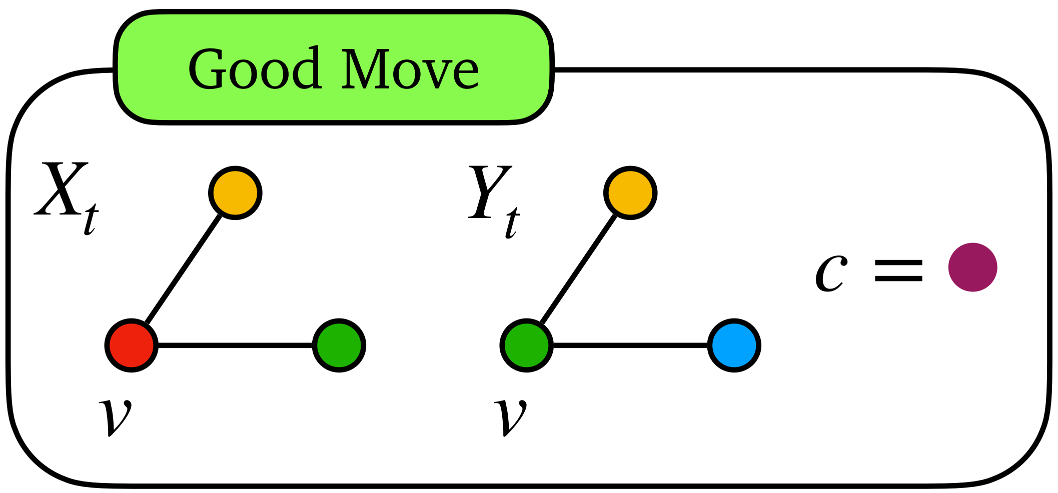
Two chains choose the same v and c



$$d(X_{t+1}, Y_{t+1}) = d(X_t, Y_t) - 1$$

The Coupling

Two chains choose the same v and c

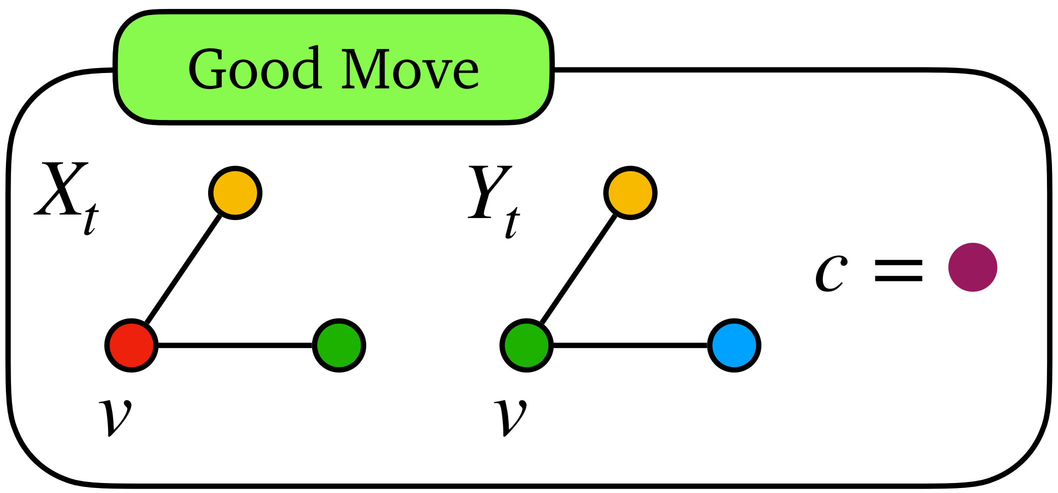


$$d(X_{t+1}, Y_{t+1}) = d(X_t, Y_t) - 1$$

$$\mathbf{Pr}[\,\cdot\,] \geq \frac{d(X_t, Y_t)}{N} \cdot \frac{q - 2(\Delta - 1)}{q}$$

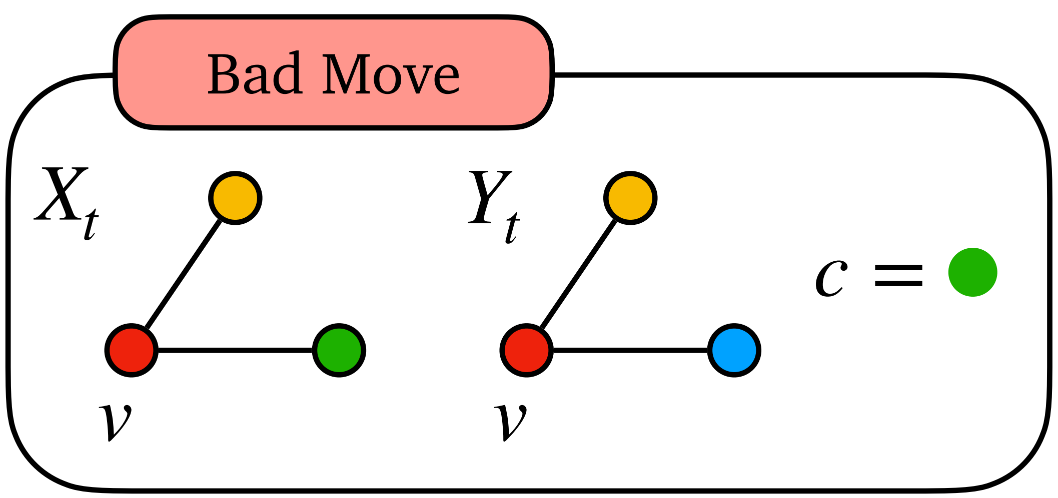
The Coupling

Two chains choose the same v and c



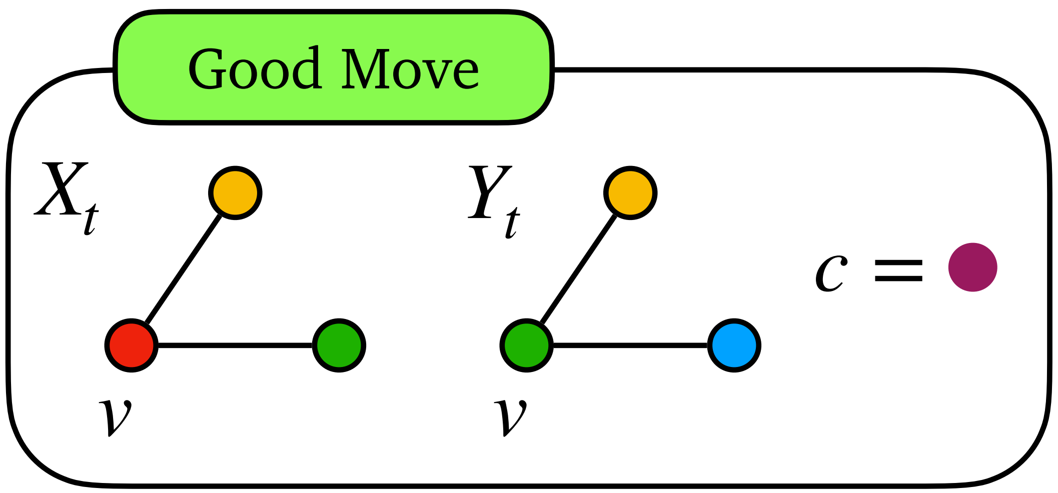
$$d(X_{t+1}, Y_{t+1}) = d(X_t, Y_t) - 1$$

$$\mathbf{Pr}[\,\cdot\,] \geq \frac{d(X_t, Y_t)}{N} \cdot \frac{q - 2(\Delta - 1)}{q}$$



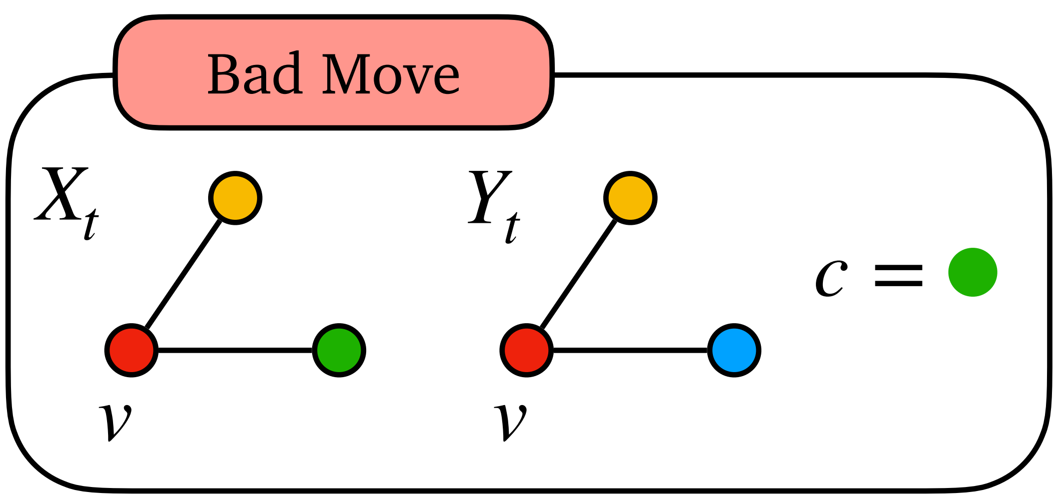
The Coupling

Two chains choose the same v and c



$$d(X_{t+1}, Y_{t+1}) = d(X_t, Y_t) - 1$$

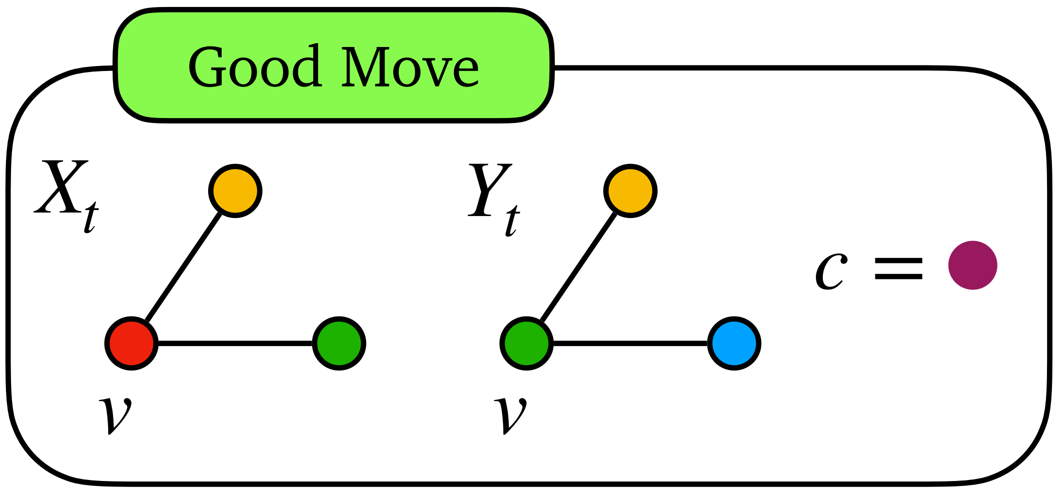
$$\mathbf{Pr}[\,\cdot\,] \geq \frac{d(X_t, Y_t)}{N} \cdot \frac{q - 2(\Delta - 1)}{q}$$



$$d(X_{t+1}, Y_{t+1}) = d(X_t, Y_t) + 1$$

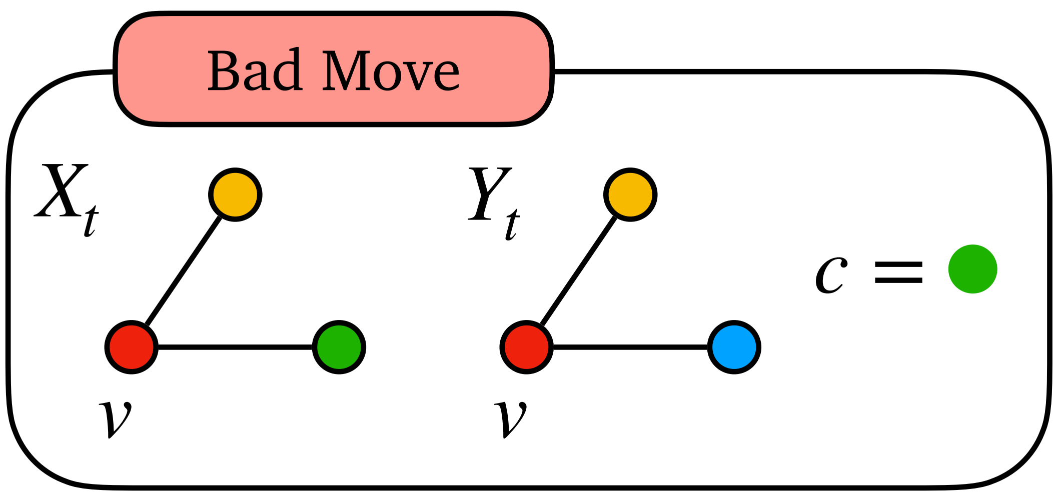
The Coupling

Two chains choose the same v and c



$$d(X_{t+1}, Y_{t+1}) = d(X_t, Y_t) - 1$$

$$\mathbf{Pr}[\,\cdot\,] \geq \frac{d(X_t, Y_t)}{N} \cdot \frac{q - 2(\Delta - 1)}{q}$$



$$d(X_{t+1}, Y_{t+1}) = d(X_t, Y_t) + 1$$

$$\mathbf{Pr}[\,\cdot\,] \leq \frac{2d(X_t, Y_t)\Delta}{Nq}$$

$$\begin{aligned}\mathbf{E}[d(X_{t+1}, Y_{t+1}) \mid (X_t, Y_t)] &\le d(X_t, Y_t) \cdot \left(1 + \frac{2\Delta - (q - 2\Delta + 2))}{qN}\right) \\ &= d(X_t, Y_t) \cdot \left(1 - \frac{q - 4\Delta + 2}{qN}\right)\end{aligned}$$

$$\mathbf{E}[d(X_{t+1}, Y_{t+1}) \mid (X_t, Y_t)] \leq d(X_t, Y_t) \cdot \left(1 + \frac{2\Delta - (q - 2\Delta + 2))}{qN}\right)$$
$$= d(X_t, Y_t) \cdot \left(1 - \frac{q - 4\Delta + 2}{qN}\right)$$

So if $q \geq 4\Delta - 1$, we have

$$\begin{aligned}\mathbf{E}[d(X_{t+1}, Y_{t+1}) \mid (X_t, Y_t)] &\le d(X_t, Y_t) \cdot \left(1 + \frac{2\Delta - (q - 2\Delta + 2))}{qN}\right) \\ &= d(X_t, Y_t) \cdot \left(1 - \frac{q - 4\Delta + 2}{qN}\right)\end{aligned}$$

So if $q \ge 4\Delta - 1$, we have

$$\mathbf{E}[d(X_{t+1}, Y_{t+1}) \mid (X_t, Y_t)] \le \left(1 - \frac{1}{qN}\right) d(X_t, Y_t)$$

$$\mathbf{E}[d(X_{t+1}, Y_{t+1}) \mid (X_t, Y_t)] \le d(X_t, Y_t) \cdot \left(1 + \frac{2\Delta - (q - 2\Delta + 2))}{qN}\right)$$

$$= d(X_t, Y_t) \cdot \left(1 - \frac{q - 4\Delta + 2}{qN}\right)$$

So if $q \ge 4\Delta - 1$, we have

$$\mathbf{E}[d(X_{t+1}, Y_{t+1}) \mid (X_t, Y_t)] \le \left(1 - \frac{1}{qN}\right) d(X_t, Y_t)$$

In other words, $\{d(X_t, Y_t)\}_{t \ge 0}$ is a super martingale

Recall the mixing time

The mixing time $\tau_{\text{mix}}(\varepsilon)$ is the first time t such that the total variation distance between X_t and π is at most ε, for any initial X_0

$$\tau_{\text{mix}}(\varepsilon) = \max_{\mu_0} \min_{t \ge 0} d_{\text{TV}}(\mu_0^T P^t, \pi) \le \varepsilon$$

By coupling lemma

$$d_{\text{TV}}(X_t, Y_t) \le \mathbf{Pr}[X_t \ne Y_t] = \mathbf{Pr}[d(X_t, Y_t) > 0]$$

For finite Ω, we assume WLOG that $\min_{x,y \in \Omega : x \ne y} d(x, y) = 1$

$$\begin{aligned}\mathbf{Pr}[d(X_t, Y_t) > 0] &= \mathbf{Pr}[d(X_t, Y_t) \ge 1] \\ &\le \mathbf{E}[d(X_t, Y_t)] \le (1 - \alpha)^t \cdot d(X_0, Y_0)\end{aligned}$$

$$\mathbf{E}[d(X_{t+1}, Y_{t+1}) \mid (X_t, Y_t)] \le d(X_t, Y_t) \cdot \left(1 + \frac{2\Delta - (q - 2\Delta + 2))}{qN}\right)$$

$$= d(X_t, Y_t) \cdot \left(1 - \frac{q - 4\Delta + 2}{qN}\right)$$

So if $q \ge 4\Delta - 1$, we have

$$\mathbf{E}[d(X_{t+1}, Y_{t+1}) \mid (X_t, Y_t)] \le \left(1 - \frac{1}{qN}\right) d(X_t, Y_t)$$

In other words, $\{d(X_t, Y_t)\}_{t\ge 0}$ is a super martingale

Recall the mixing time

The mixing time $\tau_{\mathrm{mix}}(\varepsilon)$ is the first time t such that the total variation distance between X_t and π is at most ε, for any initial X_0

$$\tau_{\mathrm{mix}}(\varepsilon) = \max_{\mu_0} \min_{t\ge 0} d_{\mathrm{TV}}(\mu_0^T P^t, \pi) \le \varepsilon$$

By coupling lemma

$$d_{\mathrm{TV}}(X_t, Y_t) \le \mathbf{Pr}[X_t \ne Y_t] = \mathbf{Pr}[d(X_t, Y_t) > 0]$$

For finite Ω, we assume WLOG that $\min_{x,y\in\Omega: x\ne y} d(x, y) = 1$

$$\mathbf{Pr}[d(X_t, Y_t) > 0] = \mathbf{Pr}[d(X_t, Y_t) \ge 1]$$
$$\le \mathbf{E}[d(X_t, Y_t)] \le (1-\alpha)^t \cdot d(X_0, Y_0)$$

$$d_{\mathrm{TV}}(X_t, Y_t) \le \left(1 - \frac{1}{qN}\right)^t \cdot N \le \varepsilon$$

$$\mathbf{E}[d(X_{t+1}, Y_{t+1}) \mid (X_t, Y_t)] \le d(X_t, Y_t) \cdot \left(1 + \frac{2\Delta - (q - 2\Delta + 2))}{qN}\right)$$

$$= d(X_t, Y_t) \cdot \left(1 - \frac{q - 4\Delta + 2}{qN}\right)$$

So if $q \ge 4\Delta - 1$, we have

$$\mathbf{E}[d(X_{t+1}, Y_{t+1}) \mid (X_t, Y_t)] \le \left(1 - \frac{1}{qN}\right) d(X_t, Y_t)$$

In other words, $\{d(X_t, Y_t)\}_{t \ge 0}$ is a super martingale

Recall the mixing time

The mixing time $\tau_{\text{mix}}(\varepsilon)$ is the first time t such that the total variation distance between X_t and π is at most ε, for any initial X_0

$$\tau_{\text{mix}}(\varepsilon) = \max_{\mu_0} \min_{t \ge 0} d_{\text{TV}}(\mu_0^T P^t, \pi) \le \varepsilon$$

By coupling lemma

$$d_{\text{TV}}(X_t, Y_t) \le \mathbf{Pr}[X_t \ne Y_t] = \mathbf{Pr}[d(X_t, Y_t) > 0]$$

For finite Ω, we assume WLOG that $\min_{x,y \in \Omega : x \ne y} d(x, y) = 1$

$$\mathbf{Pr}[d(X_t, Y_t) > 0] = \mathbf{Pr}[d(X_t, Y_t) \ge 1]$$
$$\le \mathbf{E}[d(X_t, Y_t)] \le (1 - \alpha)^t \cdot d(X_0, Y_0)$$

$$d_{\text{TV}}(X_t, Y_t) \le \left(1 - \frac{1}{qN}\right)^t \cdot N \le \varepsilon$$

$$\implies \tau_{mix}(\varepsilon) \le qN\left(\log N + \log \varepsilon^{-1}\right)$$

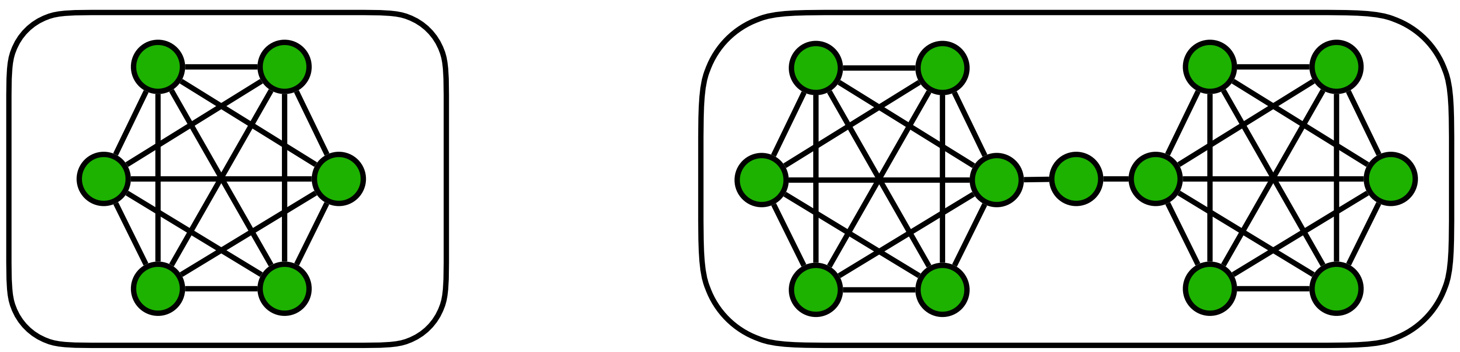
Geometric View of Mixing

Geometric View of Mixing

A Markov chain is a random walk on the state space

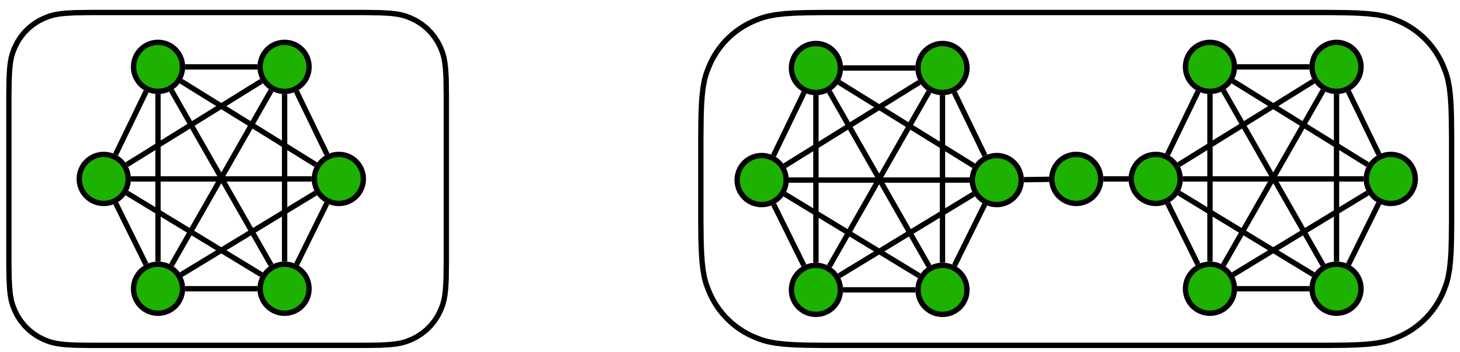
Geometric View of Mixing

A Markov chain is a random walk on the state space



Geometric View of Mixing

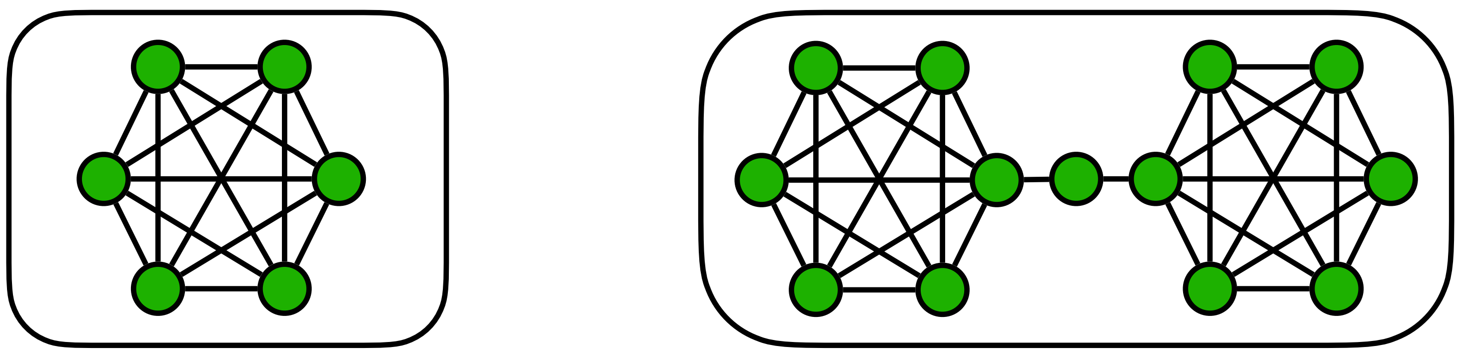
A Markov chain is a random walk on the state space



Which random walk mixes faster?

Geometric View of Mixing

A Markov chain is a random walk on the state space



Which random walk mixes faster?

We will develop tools to formalize the intuition

Back to Graph Spectrum