- We deal with discrete-time Markov chains on a finite state space Ω.

- We deal with discrete-time Markov chains on a finite state space Ω.
- A sequence $\{X_t \in \Omega\}_{t=0}^{\infty}$ of random variables is a Markov chain (MC), with state space Ω, if

$$\Pr[X_{t+1} = y \mid X_t = x_t, ..., X_0 = x_0] = \Pr[X_{t+1} = y \mid X_t = x_t]$$

for all $t \in \mathbb{N}$ and all $x_0, ..., x_t \in \Omega$.

- We deal with discrete-time Markov chains on a finite state space Ω.
- A sequence $\{X_t \in \Omega\}_{t=0}^{\infty}$ of random variables is a Markov chain (MC), with state space Ω, if

$$\Pr[X_{t+1} = y \mid X_t = x_t, ..., X_0 = x_0] = \Pr[X_{t+1} = y \mid X_t = x_t]$$

  for all $t \in \mathbb{N}$ and all $x_0, ..., x_t \in \Omega$.
- This is called the Markovian property.

- We deal with discrete-time Markov chains on a finite state space Ω.
- A sequence $\{X_t \in \Omega\}_{t=0}^{\infty}$ of random variables is a Markov chain (MC), with state space Ω, if

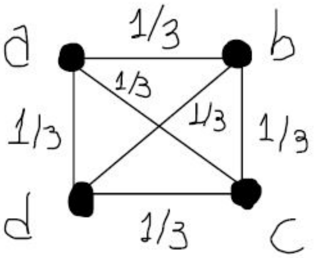
$$\Pr[X_{t+1} = y \mid X_t = x_t, ..., X_0 = x_0] = \Pr[X_{t+1} = y \mid X_t = x_t]$$

  for all $t \in \mathbb{N}$ and all $x_0, ..., x_t \in \Omega$.
- This is called the Markovian property.
- Time-homogeneous MCs are the ones for which the probability $\Pr[X_{t+1} = y \mid X_t = x]$ does not depend on t. In this case we write

$$P(x, y) = \Pr[X_{t+1} = y \mid X_t = x]$$

  where P is the transition matrix of the MC.

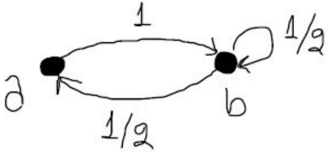
Example 1



$$P = \begin{array}{c} \\ a \\ b \\ c \\ d \end{array} \begin{array}{c} \begin{array}{cccc} a & b & c & d \end{array} \\ \begin{pmatrix} 0 & 1/3 & 1/3 & 1/3 \\ 1/3 & 0 & 1/3 & 1/3 \\ 1/3 & 1/3 & 0 & 1/3 \\ 1/3 & 1/3 & 1/3 & 0 \end{pmatrix} \end{array}$$

$$X_0 = a, X_1 = b, X_2 = d, X_3 = b, \ldots$$

Example 2



$$P = \begin{array}{c} \\ a \\ b \end{array} \begin{array}{c} \begin{array}{cc} a & b \end{array} \\ \begin{pmatrix} 0 & 1 \\ 1/2 & 1/2 \end{pmatrix} \end{array}$$

$$X_0 = a, X_1 = b, X_2 = b, X_3 = b, X_4 = a, X_5 = b, \ldots$$

Transition matrix

- Each row of the transition matrix P is a distribution.

Transition matrix

- Each row of the transition matrix P is a distribution.
- P describes single-step transition probabilities.

Transition matrix

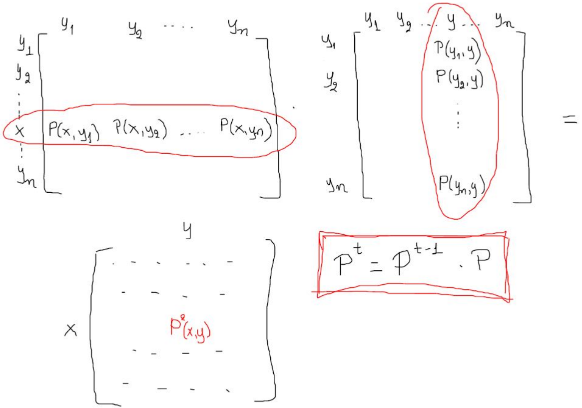
- Each row of the transition matrix P is a distribution.
- P describes single-step transition probabilities.
- The t-step transition probabilities are given inductively by

$$P^t(x, y) := \begin{cases} I(x, y), & \text{if } t = 0 \\ \sum_{y' \in \Omega} P^{t-1}(x, y')P(y', y), & \text{if } t > 0 \end{cases}$$

Transition matrix

- Each row of the transition matrix P is a distribution.
- P describes single-step transition probabilities.
- The t-step transition probabilities are given inductively by

$$P^t(x,y) := \begin{cases} I(x,y), & \text{if } t = 0 \\ \sum_{y' \in \Omega} P^{t-1}(x,y')P(y',y), & \text{if } t > 0 \end{cases}$$



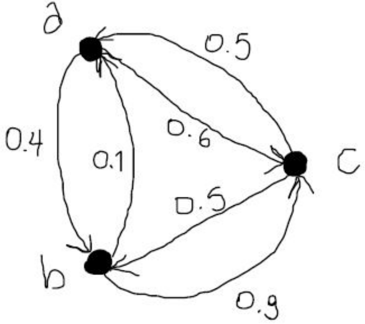
Transition matrix

- Each row of the transition matrix P is a distribution.
- P describes single-step transition probabilities.
- The t-step transition probabilities are given inductively by

$$P^t(x,y) := \begin{cases} I(x,y), & \text{if } t = 0 \\ \sum_{y' \in \Omega} P^{t-1}(x,y')P(y',y), & \text{if } t > 0 \end{cases}$$

- So P^t describes t-step transition probabilities.

Example 3



$$P = \begin{array}{c} \\ a \\ b \\ c \end{array} \begin{array}{c} \begin{array}{ccc} a & b & c \end{array} \\ \begin{bmatrix} 0 & 0.4 & 0.6 \\ 0.1 & 0 & 0.9 \\ 0.5 & 0.5 & 0 \end{bmatrix} \end{array}$$

Start distribution: $\sigma_0 = (1, 0, 0)$ (start in a)
After one step: $\sigma_1 = \sigma_0 P = (0, 0.4, 0.6)$
After two steps: $\sigma_2 = \sigma_1 P = \sigma_0 P^2 = (0.34, 0.3, 0.36)$
After t steps: $\sigma_t = \sigma_{t-1} P = \sigma_0 P^t$

Stationary distribution

$$\begin{bmatrix} \pi(a) & \pi(b) & \pi(c) & \pi(d) \end{bmatrix} \begin{bmatrix} P(a,a) & P(a,b) & P(a,c) & P(a,d) \\ P(b,a) & P(b,b) & P(b,c) & P(b,d) \\ P(c,a) & P(c,b) & P(c,c) & P(c,d) \\ P(d,a) & P(d,b) & P(d,c) & P(d,d) \end{bmatrix} =$$

$$\begin{bmatrix} \pi(a) & \pi(b) & \pi(c) & \pi(d) \end{bmatrix}$$

$$\pi(a) = \pi(a)P(a,a) + \pi(b)P(b,a) + \pi(c)P(c,a) + \pi(d)P(d,a)$$

Stationary distribution

- A stationary distribution of an MC with transition matrix P is a distribution $\pi : \Omega \to [0,1]$ such that

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

- In other words, $\pi \cdot P = \pi$.

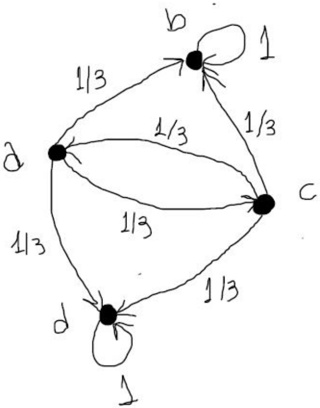
Definition of irreducibility

Definition

An MC is irreducible if for all $x, y \in \Omega$, there exists a $t > 0$, such that $P^t(x, y) > 0$ (there exists a path in the transition graph from every state to every other state).

Example 4

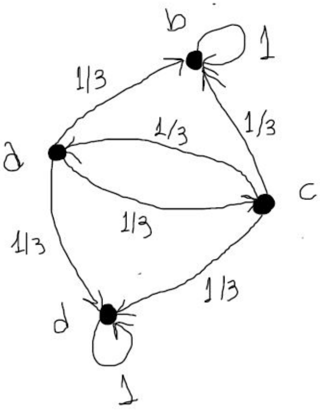
Not irreducible



$$P = \begin{bmatrix} 0 & 1/3 & 1/3 & 1/3 \\ 0 & 1 & 0 & 0 \\ 1/3 & 1/3 & 0 & 1/3 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

Example 4

Not irreducible



$$P = \begin{bmatrix} 0 & 1/3 & 1/3 & 1/3 \\ 0 & 1 & 0 & 0 \\ 1/3 & 1/3 & 0 & 1/3 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

Stationary distributions: $\pi_1 = (0, 1, 0, 0)$, $\pi_2 = (0, 0, 0, 1)$, $\pi_3 = (0, 0.5, 0, 0.5)$, ...

Definition of aperiodicity

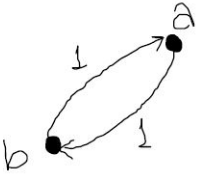
Definition

An MC is aperiodic if $\gcd\{t \mid P^t(x,x) > 0\} = 1$ for all $x \in \Omega$ (for each state x, the gcd of all walk lengths from x to x is 1).

In the case of an irreducible MC, it is sufficient to verify the condition $\gcd\{t \mid P^t(x,x) > 0\} = 1$ for just one state $x \in \Omega$.

Example 5

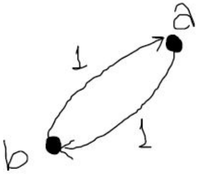
Not aperiodic



$$P = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$$

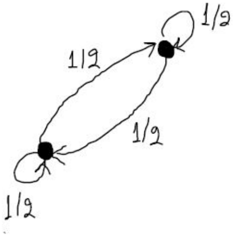
Example 5

Not aperiodic



$$P = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$$

Lazy MC (a self-loop at every state)



$$P = \begin{bmatrix} 1/2 & 1/2 \\ 1/2 & 1/2 \end{bmatrix}$$

A (finite sate) MC is ergodic if and only if it is irreducible and aperiodic.

Theorem

An ergodic MC has a unique stationary distribution π.
Moreover, the MC tends to π in the sense that $P^t(x, y) \to \pi(y)$, as $t \to \infty$, for all $x \in \Omega$.

Note that the MC eventually forgets its starting state.

Detailed balance condition

Theorem

Suppose P is the transition matrix of an MC. If the function $\pi' : \Omega \to [0,1]$ satisfies

$$\pi'(x)P(x,y) = \pi'(y)P(y,x), \text{ for all } x, y \in \Omega$$

and,

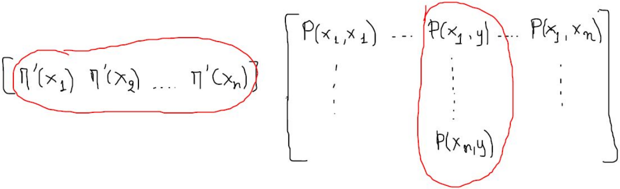
$$\sum_{x \in \Omega} \pi'(x) = 1,$$

then π' is the stationary distribution of the MC.
If, in addition, the MC is ergodic, then π' is the unique stationary distribution.

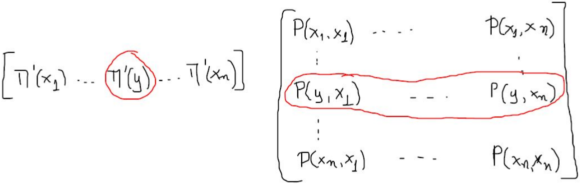
An MC for which the detailed balance condition holds is called time-reversible.

Proof. Suppose X_t is distributed as π'. Then,

$$\Pr[X_{t+1} = y] = \sum_{x \in \Omega} \pi'(x) P(x, y) =$$



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



□

Example 6

Random walk on the *n*-dimensional hypercube

- $\Omega = \{0, 1\}^n$.
- Start with some $x \in \Omega$.

Example 6

Random walk on the *n*-dimensional hypercube

- $\Omega = \{0, 1\}^n$.
- Start with some $x \in \Omega$.
- Description of a transition: Choose a bit u.a.r. and flip it with probability $1/2$.

Example 6

Random walk on the *n*-dimensional hypercube

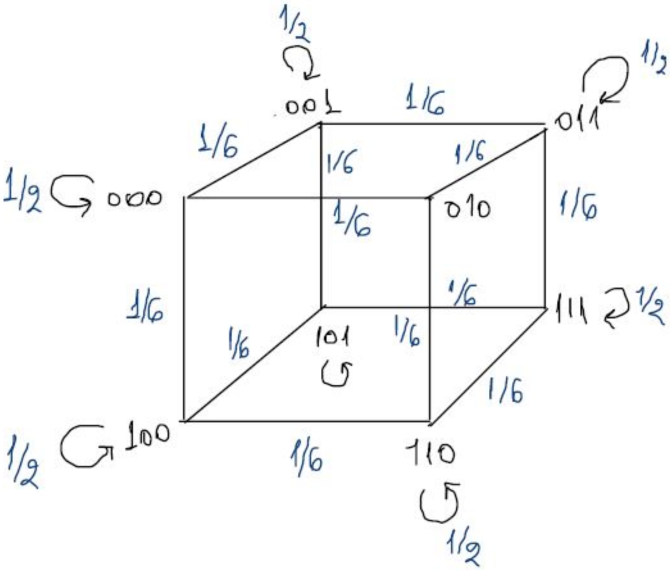
- $\Omega = \{0, 1\}^n$.
- Start with some $x \in \Omega$.
- Description of a transition: Choose a bit u.a.r. and flip it with probability $1/2$.
- This MC is irreducible and aperiodic. Why?

Example 6

Random walk on the *n*-dimensional hypercube

- $\Omega = \{0, 1\}^n$.
- Start with some $x \in \Omega$.
- Description of a transition: Choose a bit u.a.r. and flip it with probability $1/2$.
- This MC is irreducible and aperiodic. Why?
- Its stationary distribution is π with $\pi(x) = \frac{1}{2^n}$, for all $x \in \Omega$. Why?

Example 6



1/2
001
1/6
011
1/2
1/6
1/6
1/6
1/2
000
010
1/6
1/6
1/6
1/6
111
1/2
1/6
101
1/6
1/6
1/2
100
1/6
110
1/2

Example 7

Generating uniformly random matchings

- $\Omega = \mathcal{M}(G) =$ the set of all matchings in a graph G.
- Start with a matching, i.e. $X_0 = M_0 \in \mathcal{M}(G)$.

Example 7

Generating uniformly random matchings

- $\Omega = \mathcal{M}(G) =$ the set of all matchings in a graph G.
- Start with a matching, i.e. $X_0 = M_0 \in \mathcal{M}(G)$.
- Suppose $X_t = M$. The next state is the result of the following trial.
  1. With probability $\frac{1}{2}$ set $X_{t+1} \leftarrow M$ and halt.
  2. Otherwise, choose $e \in E(G)$ and set $M' \leftarrow M \oplus \{e\}$.
  3. If $M' \in \mathcal{M}(G)$ then $X_{t+1} \leftarrow M'$, else $X_{t+1} \leftarrow M$.

Example 7

Generating uniformly random matchings

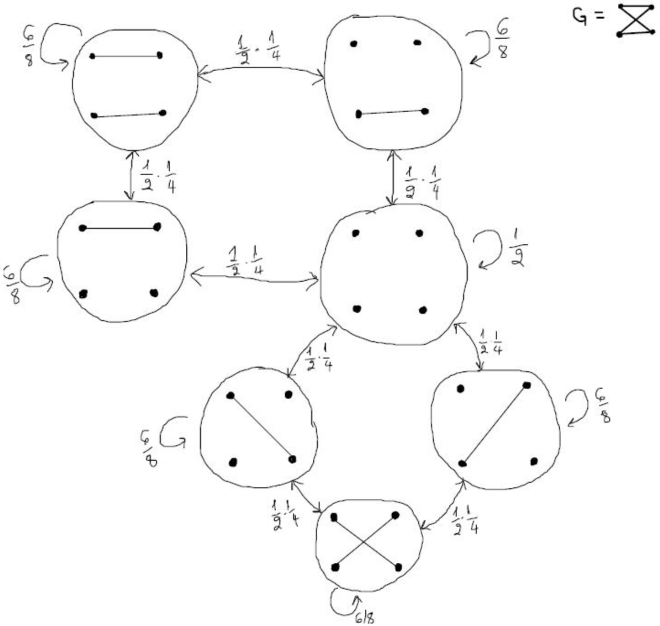
- $\Omega = \mathcal{M}(G) =$ the set of all matchings in a graph G.
- Start with a matching, i.e. $X_0 = M_0 \in \mathcal{M}(G)$.
- Suppose $X_t = M$. The next state is the result of the following trial.
  1. With probability $\frac{1}{2}$ set $X_{t+1} \leftarrow M$ and halt.
  2. Otherwise, choose $e \in E(G)$ and set $M' \leftarrow M \oplus \{e\}$.
  3. If $M' \in \mathcal{M}(G)$ then $X_{t+1} \leftarrow M'$, else $X_{t+1} \leftarrow M$.
- This MC is irreducible and aperiodic. Why?

Example 7

Generating uniformly random matchings

- $\Omega = \mathcal{M}(G) =$ the set of all matchings in a graph G.
- Start with a matching, i.e. $X_0 = M_0 \in \mathcal{M}(G)$.
- Suppose $X_t = M$. The next state is the result of the following trial.
  1. With probability $\frac{1}{2}$ set $X_{t+1} \leftarrow M$ and halt.
  2. Otherwise, choose $e \in E(G)$ and set $M' \leftarrow M \oplus \{e\}$.
  3. If $M' \in \mathcal{M}(G)$ then $X_{t+1} \leftarrow M'$, else $X_{t+1} \leftarrow M$.
- This MC is irreducible and aperiodic. Why?
- Its stationary distribution is the uniform distribution over all matchings (Exercise).

Example 7



Definition of the mixing time

- Suppose (X_t) is an ergodic MC on countable state space Ω with transition matrix P and initial state $X_0 = x \in \Omega$.

Definition of the mixing time

- Suppose (X_t) is an ergodic MC on countable state space Ω with transition matrix P and initial state $X_0 = x \in \Omega$.
- For $t \in \mathbb{N}$, the distribution of X_t is naturally denoted $P^t(x, \cdot)$.

Definition of the mixing time

- Suppose (X_t) is an ergodic MC on countable state space Ω with transition matrix P and initial state $X_0 = x \in \Omega$.
- For $t \in \mathbb{N}$, the distribution of X_t is naturally denoted $P^t(x, \cdot)$.
- Let π denote the sationary distribution of the MC, i.e. the limit of $P^t(x, \cdot)$ as $t \to \infty$.

Definition of the mixing time

- Suppose (X_t) is an ergodic MC on countable state space Ω with transition matrix P and initial state $X_0 = x \in \Omega$.
- For $t \in \mathbb{N}$, the distribution of X_t is naturally denoted $P^t(x, \cdot)$.
- Let π denote the sationary distribution of the MC, i.e. the limit of $P^t(x, \cdot)$ as $t \to \infty$.
- The rate of convergence to stationarity of (X_t) is measured by its mixing time from initial state x:

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

Lemma

The total variation distance $||P^t(x,\cdot) - \pi||_{TV}$ of the t-step distribution from sationarity is non-increasing with respect to t.

Lemma

The total variation distance $||P^t(x,\cdot) - \pi||_{TV}$ of the t-step distribution from sationarity is non-increasing with respect to t.

- The definition of the mixing time is equivalent to

$$\tau_x(\varepsilon) := \min\{t \mid ||P^s(x,\cdot) - \pi||_{TV} \leq \varepsilon, \text{ for all } s \geq t\}.$$

Lemma

The total variation distance $||P^t(x,\cdot) - \pi||_{TV}$ of the t-step distribution from sationarity is non-increasing with respect to t.

- The definition of the mixing time is equivalent to

$$\tau_x(\varepsilon) := \min\{t \mid ||P^s(x,\cdot) - \pi||_{TV} \leq \varepsilon, \text{ for all } s \geq t\}.$$

- The definition of mixing time that is independent of the initial state is $\tau(\varepsilon) = \max_{x\in\Omega} \tau_x(\varepsilon)$.

Lemma

The total variation distance $||P^t(x,\cdot) - \pi||_{TV}$ of the t-step distribution from sationarity is non-increasing with respect to t.

- The definition of the mixing time is equivalent to

$$\tau_x(\varepsilon) := \min\{t \mid ||P^s(x,\cdot) - \pi||_{TV} \leq \varepsilon, \text{ for all } s \geq t\}.$$

- The definition of mixing time that is independent of the initial state is $\tau(\varepsilon) = \max_{x\in\Omega} \tau_x(\varepsilon)$.
- We define $\tau_{mix} := \tau(1/4)$ (i.e. the first time t at which $||P^t(x,\cdot) - \pi||_{TV} \leq \frac{1}{4}$).
  Then, $||P^{k\tau_{mix}}(x,\cdot) - \pi||_{TV} \leq 2^{-k}$ for every $k \in \mathbb{N}$.
  Equivalently,

$$\tau(\varepsilon) \leq \lceil \log \varepsilon^{-1} \rceil \tau_{mix}.$$

Overview

- Let $G = (V, E)$ be a graph and $Q = \{1, 2, ..., q\}$ be a set of colors.
- We wish to sample a random q-coloring uniformly from all possible proper q-colorings on the vertices of G.
- If this sampler is an fpaus, then there is an fpras for counting the proper q-colorings of G.

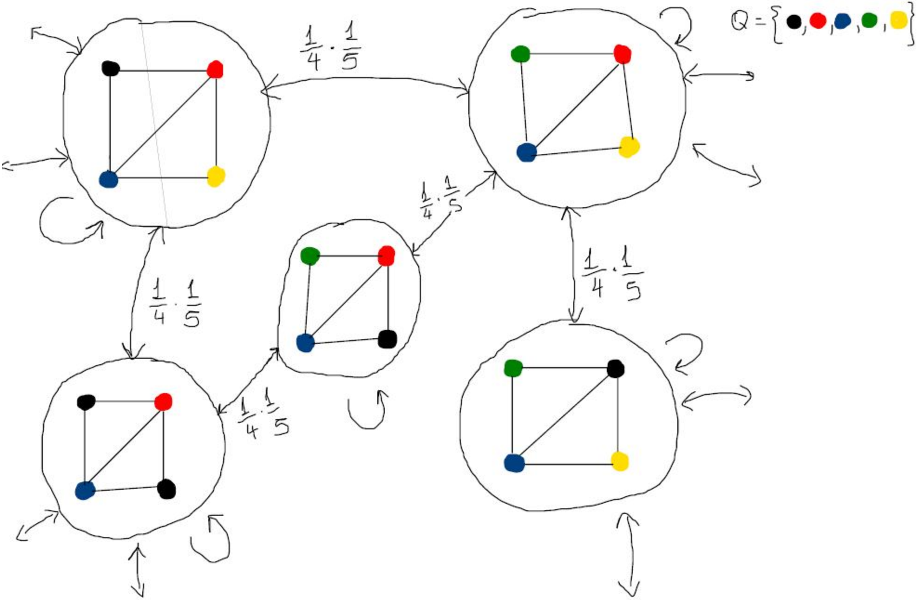
- Let Ω be the set of proper q-colorings of G.
- We will describe a time-homogeneous MC on Ω.
- We are considering a q-coloring as a function $C : V \rightarrow Q$, so we denote by $X_t(u)$ the color of vertex u in the state that the MC is in, after t steps.

Trial defining an MC on q-colorings

Suppose $X_t = C$. The next state is the result of the following trial.

1. Choose a vertex $v \in V$ and a color $c \in Q$ u.a.r.
2. Change the color of v to c, i.e. $X_{t+1}(v) \leftarrow c$, if the result would be a legal coloring. Otherwise, keep the colors of all vertices the same.

Example 8


$Q = \{ \bullet, \bullet, \bullet, \bullet, \bullet \}$
$\frac{1}{4} \cdot \frac{1}{5}$
$\frac{1}{4} \cdot \frac{1}{5}$
$\frac{1}{4} \cdot \frac{1}{5}$
$\frac{1}{4} \cdot \frac{1}{5}$
$\frac{1}{4} \cdot \frac{1}{5}$

- Does this MC yield an fpaus?

- Can the problem of counting q-colorings have an fpras?

- Which counting problems can have an fpras?

Which counting problems can have an fpras?

- If $\#\textsc{Sat}$ has an fpras then RP $=$ NP.
- More generally, if a problem in $\#$P with an NP-complete decision version admits an fpras, then RP $=$ NP.

Which counting problems can have an fpras?

- If $\#\textsc{Sat}$ has an fpras then $\mathsf{RP} = \mathsf{NP}$.
- More generally, if a problem in $\#\mathsf{P}$ with an NP-complete decision version admits an fpras, then $\mathsf{RP} = \mathsf{NP}$.
- We focus on counting problems that have an easy decision version, i.e. in BPP.
- Most counting problems with an fpras, have a decision version in P.

Can the problem of counting q-colorings have an fpras?

Decision version: Is there a q-coloring in a graph G of max degree Δ?

- $q \geq \Delta + 1$: There is a coloring of G. Simply choose a color for each vertex in some fixed order. Then, at every step, there is at least one color not among the neighbors of v.

Can the problem of counting q-colorings have an fpras?

Decision version: Is there a q-coloring in a graph G of max degree Δ?

- $q \geq \Delta + 1$: There is a coloring of G. Simply choose a color for each vertex in some fixed order. Then, at every step, there is at least one color not among the neighbors of v.

- $q < \Delta$: In general, it is NP-hard to decide if there is a coloring of G. If $\Delta \geq 4$, it is NP-hard to determine if G is $(\Delta - 1)$-colorable.

Can the problem of counting q-colorings have an fpras?

Decision version: Is there a q-coloring in a graph G of max degree Δ?

- $q \geq \Delta + 1$: There is a coloring of G. Simply choose a color for each vertex in some fixed order. Then, at every step, there is at least one color not among the neighbors of v.

- $q < \Delta$: In general, it is NP-hard to decide if there is a coloring of G. If $\Delta \geq 4$, it is NP-hard to determine if G is $(\Delta - 1)$-colorable.

- $q = \Delta$: Brook's theorem states that there exists a coloring of G iff
  - $\Delta > 2$ and there is no $(\Delta + 1)$-clique in G, or
  - $\Delta = 2$ and there is no odd cycle in G.

  Brook's theorem yields a polynomial-time algorithm for constructing a q-coloring.

Theorems

- Jerrum (1995): The above Markov chain converges to the uniform distribution over Ω and it has $\tau_{mix} = \mathcal{O}(n \log n)$ for $q \geq 2\Delta + 1$.

  We are going to prove it in two steps:

  1. for $q \geq 4\Delta + 1$.
  2. for $q \geq 2\Delta + 1$.

- Vigoda (1999): $\mathcal{O}(n^2 \log n)$ mixing for $q \geq \frac{11}{6}\Delta$.

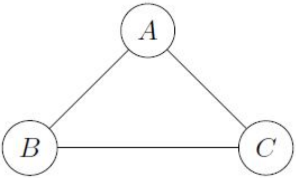
- Hayes & Sinclair (2005): $\Omega(n \log n)$ lower bound.

Conjectures

- If $q \geq \Delta + 2$, the above Markov chain has $\tau_{mix} = \mathcal{O}(n \log n)$.
  Martinelli, Sinclair & Weitz (2006): This conjecture is known to hold for Δ-regular trees.
- If $q = \Delta + 1$, there exists a Markov chain with polynomial mixing time.

The MC is ergodic

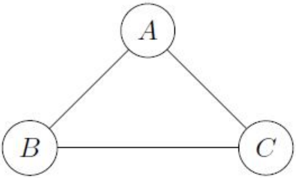
- If $q = \Delta + 1$, then the above MC is not irreducible. For example, we cannot move out of the following state.



where $Q = \{A, B, C\}$ and $\Delta = 2$.

The MC is ergodic

- If $q = \Delta + 1$, then the above MC is not irreducible. For example, we cannot move out of the following state.

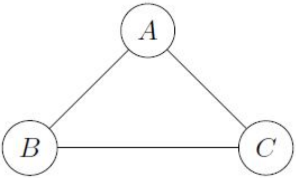


where $Q = \{A, B, C\}$ and $\Delta = 2$.

- If $q \geq \Delta + 2$, then the above MC is irreducible.

The MC is ergodic

- If $q = \Delta + 1$, then the above MC is not irreducible. For example, we cannot move out of the following state.



where $Q = \{A, B, C\}$ and $\Delta = 2$.

- If $q \geq \Delta + 2$, then the above MC is irreducible.

- The above MC is aperiodic, since when choosing the color c, it may happen to be $X_t(v)$. So there exists a self-loop at every state.

The stationary distribution of the MC

The stationary distribution of the above MC is uniform over Ω.

- Let $C_i, C_j \in \Omega$ be two different proper colorings.

The stationary distribution of the MC

The stationary distribution of the above MC is uniform over Ω.

- Let $C_i, C_j \in \Omega$ be two different proper colorings.
- Then, $P(C_i, C_j) = P(C_j, C_i) = \frac{1}{nq}$, if C_j results from C_i by recoloring some vertex $v \in V(G)$
(and $P(C_i, C_j) = P(C_j, C_i) = 0$, otherwise).

The stationary distribution of the MC

The stationary distribution of the above MC is uniform over Ω.

- Let $C_i, C_j \in \Omega$ be two different proper colorings.
- Then, $P(C_i, C_j) = P(C_j, C_i) = \frac{1}{nq}$, if C_j results from C_i by recoloring some vertex $v \in V(G)$
(and $P(C_i, C_j) = P(C_j, C_i) = 0$, otherwise).
- So, $\pi(C_i)P(C_i, C_j) = \pi(C_j)P(C_j, C_i)$, where $\pi(C_i) = \pi(C_j) = \frac{1}{|\Omega|}$.

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

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

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

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,

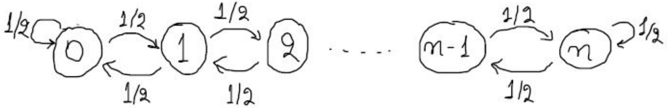
$$\sum_{y' \in \Omega} \widehat{P}((x, y), (x', y')) = P(x, x'),$$

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

Example 9

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

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



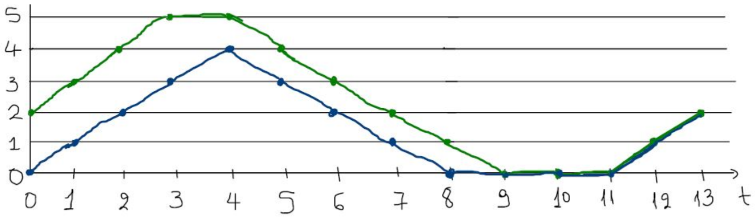
- Add either $+1$ or -1, each with probability $1/2$, to the current state if possible.

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

Example 9

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

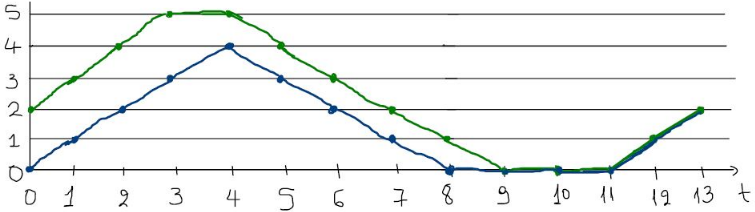
- $X_0 = x,\ Y_0 = y$.
- At the (t+1)-th step, choose $b_{t+1} \in \{-1, 1\}$ u.a.r.
- Attempt to add b_{t+1} to both X_t and Y_t.



Example 9

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 $t = t(\varepsilon)$,

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

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 $t = t(\varepsilon)$,

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

By the definition of the total variation distance, $||P^t(x, \cdot) - \pi||_{TV} \leq \varepsilon$. □

Bounding the mixing time of the MC

Theorem

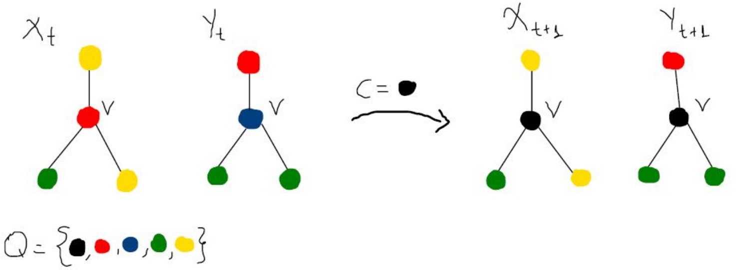
The mixing time of the above MC is $\tau_{mix} = \mathcal{O}(n \log n)$ for $q \geq 4\Delta + 1$.

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.

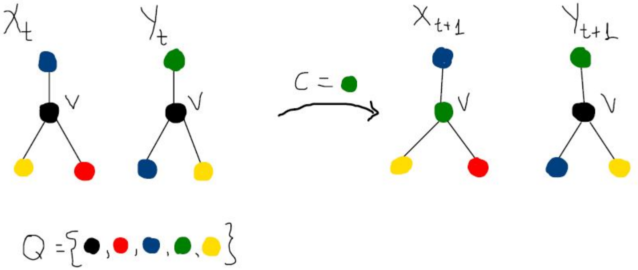
1. Good moves ($D_{t+1} = D_t - 1$): v has different colors in X_t and Y_t, and c does not appear in the neighborhood of v in either X_t or Y_t.



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

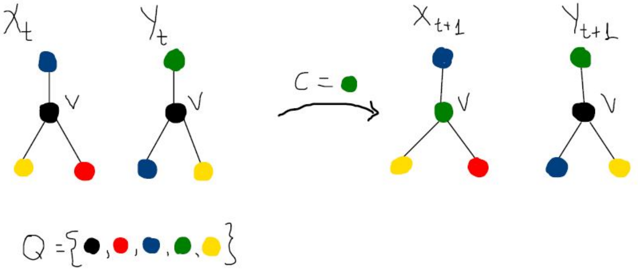
Proof cont.

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

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

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

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

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

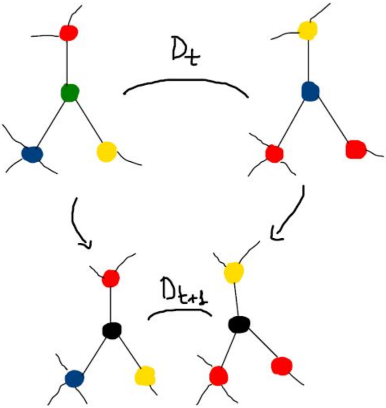
for $q \geq 4\Delta + 1$.

□

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

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

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