Advanced Algorithms (XII)

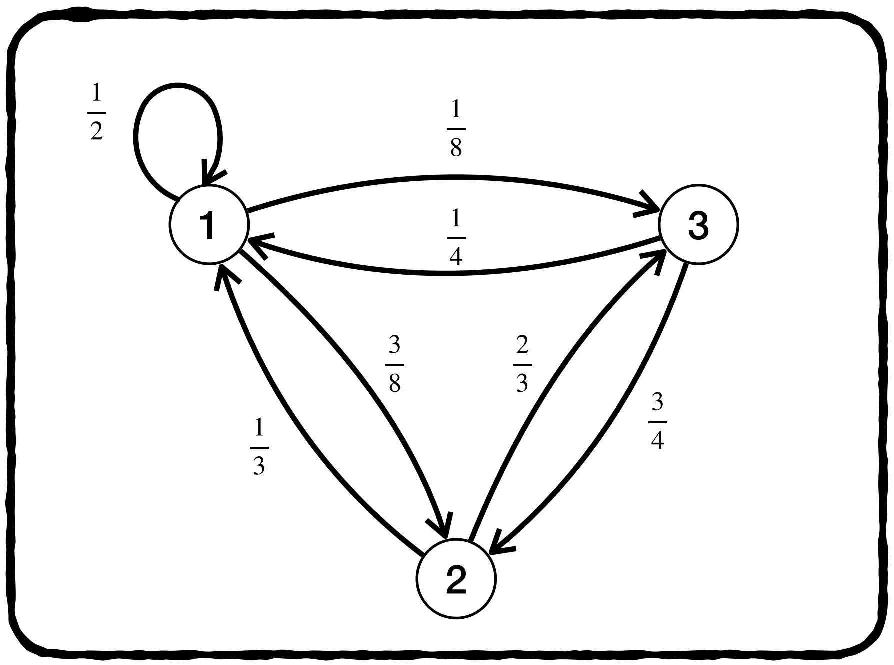
Shanghai Jiao Tong University

Chihao Zhang

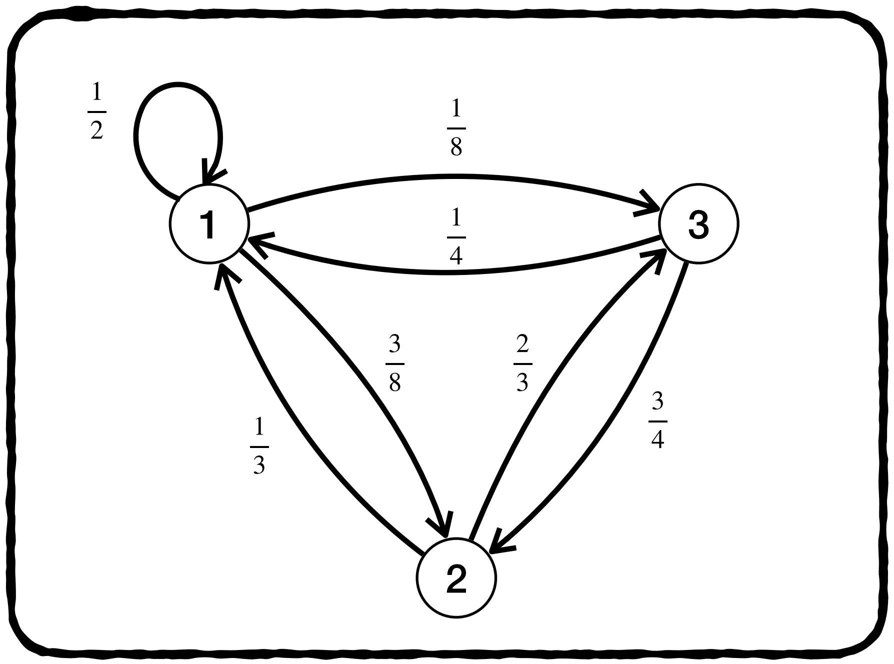
May 25, 2020

Random Walk on a Graph

Random Walk on a Graph

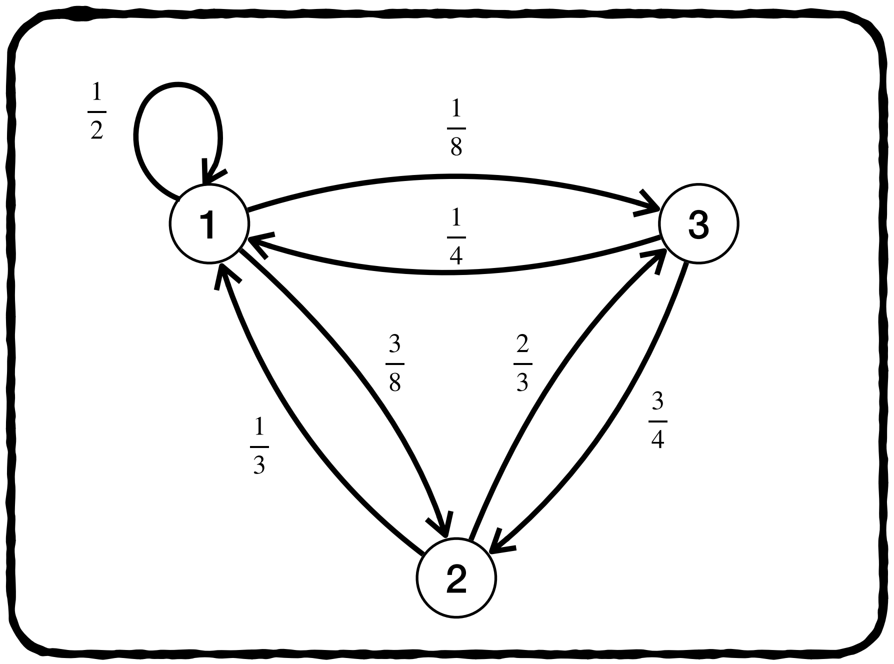


Random Walk on a Graph



$$P = [p_{ij}]_{1 \leq i,j \leq n} = \begin{bmatrix} \frac{1}{2} & \frac{3}{8} & \frac{1}{8} \\ \frac{1}{3} & 0 & \frac{2}{3} \\ \frac{1}{3} & \frac{3}{4} & 0 \end{bmatrix}$$

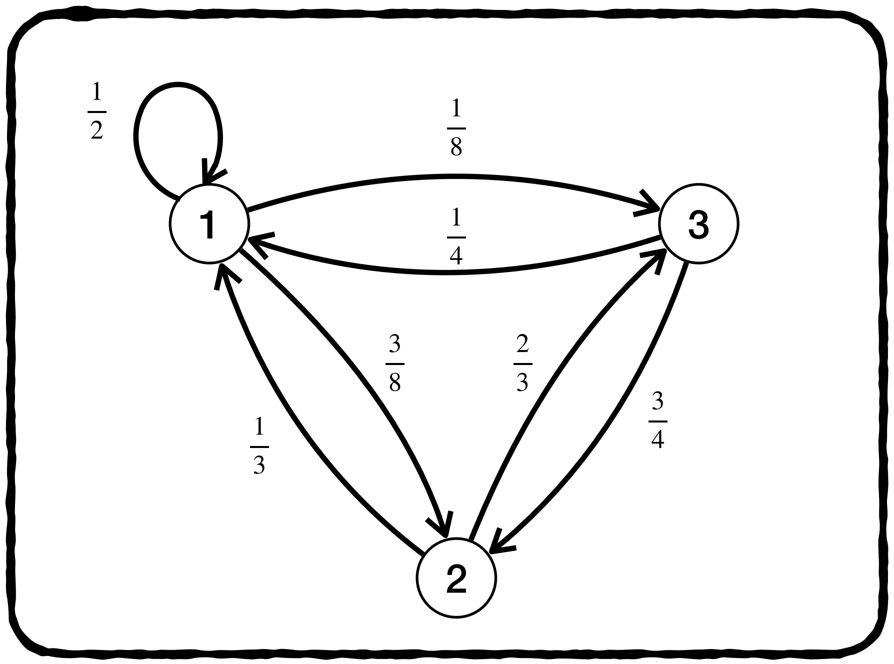
Random Walk on a Graph



$$P = [p_{ij}]_{1 \le i,j \le n} = \begin{bmatrix} \frac{1}{2} & \frac{3}{8} & \frac{1}{8} \\ \frac{1}{3} & 0 & \frac{2}{3} \\ \frac{1}{3} & \frac{3}{4} & 0 \end{bmatrix}$$

$$p_{ij} = \mathbf{Pr}[X_{t+1} = j \mid X_t = i]$$

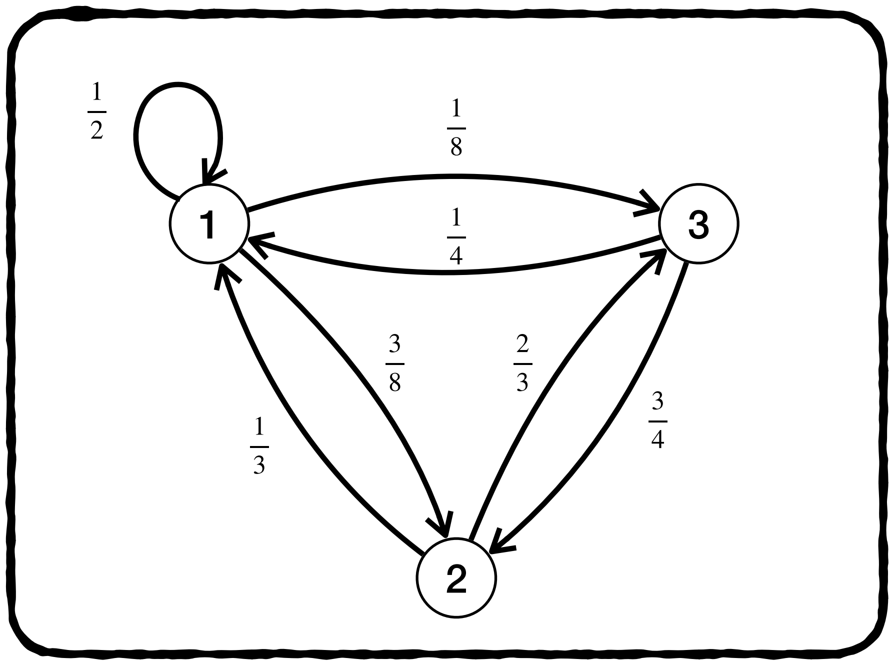
Random Walk on a Graph



$$P = [p_{ij}]_{1 \le i,j \le n} = \begin{bmatrix} \frac{1}{2} & \frac{3}{8} & \frac{1}{8} \\ \frac{1}{3} & 0 & \frac{2}{3} \\ \frac{1}{3} & \frac{3}{4} & 0 \end{bmatrix}$$

$$p_{ij} = \mathbf{Pr}[X_{t+1} = j \mid X_t = i] \qquad \forall t \ge 0, \mu_t^T = \mu_0^T P^t$$

Random Walk on a Graph



$$P = [p_{ij}]_{1 \le i,j \le n} = \begin{bmatrix} \frac{1}{2} & \frac{3}{8} & \frac{1}{8} \\ \frac{1}{3} & 0 & \frac{2}{3} \\ \frac{1}{3} & \frac{3}{4} & 0 \end{bmatrix}$$

$$p_{ij} = \mathbf{Pr}[X_{t+1} = j \mid X_t = i] \qquad \forall t \ge 0,\ \mu_t^T = \mu_0^T P^t$$

Stationary distribution π: $\pi^T P = \pi^T$

Fundamental Theorem of Markov Chains

Fundamental Theorem of Markov Chains

We study a few basic questions regarding a chain:

Fundamental Theorem of Markov Chains

We study a few basic questions regarding a chain:

- Does a stationary distribution always exist?

Fundamental Theorem of Markov Chains

We study a few basic questions regarding a chain:

- Does a stationary distribution always exist?
- If so, is the stationary distribution unique?

Fundamental Theorem of Markov Chains

We study a few basic questions regarding a chain:

- Does a stationary distribution always exist?
- If so, is the stationary distribution unique?
- If so, does any initial distribution converge to it?

Existence of Stationary Distribution

Existence of Stationary Distribution

Yes, any Markov chain has a stationary distribution

Existence of Stationary Distribution

Yes, any Markov chain has a stationary distribution

Perron-Frobenius

Any positive matrix $n \times n$ matrix A has a positive real eigenvalue λ with $\rho(A) = \lambda$. Moreover, its eigenvector is positive.

Existence of Stationary Distribution

Yes, any Markov chain has a stationary distribution

Perron-Frobenius

Any positive matrix matrix $n \times n$ matrix A has a positive real eigenvalue λ with $\rho(A) = \lambda$. Moreover, its eigenvector is positive.

$$\lambda(P^T) = \lambda(P) = 1$$

Existence of Stationary Distribution

Yes, any Markov chain has a stationary distribution

Perron-Frobenius

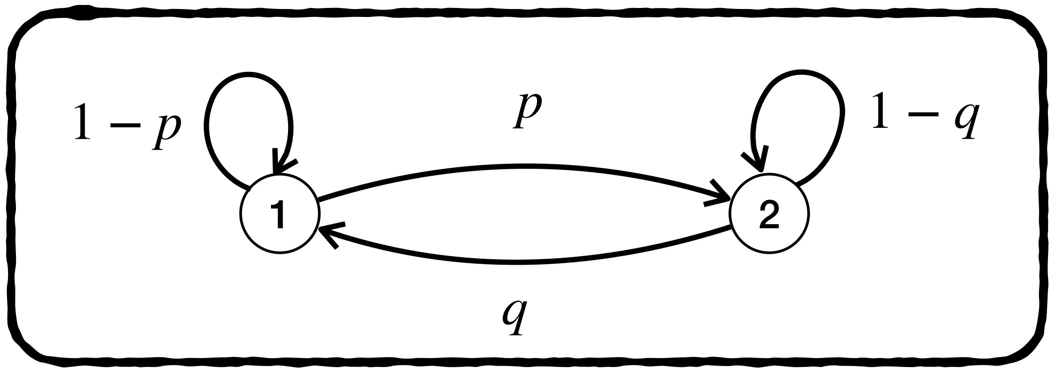
Any positive matrix $n \times n$ matrix A has a positive real eigenvalue λ with $\rho(A) = \lambda$. Moreover, its eigenvector is positive.

$$\lambda(P^T) = \lambda(P) = 1$$

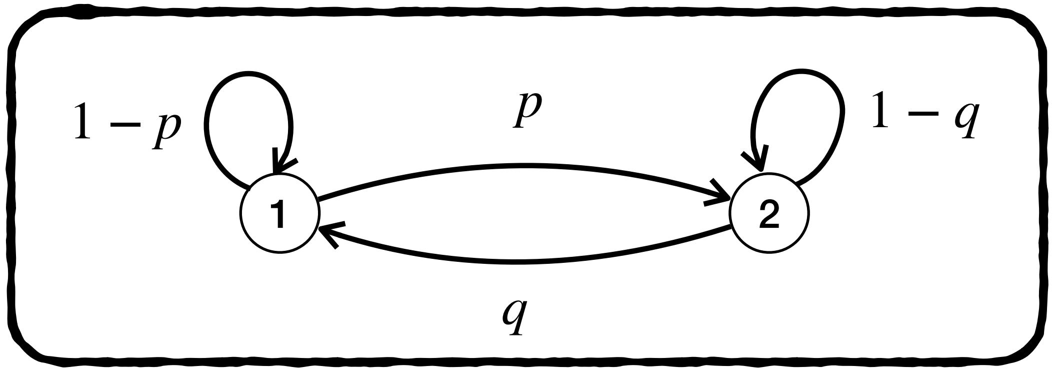
The positive eigenvector is π

Uniqueness and Convergence

Uniqueness and Convergence

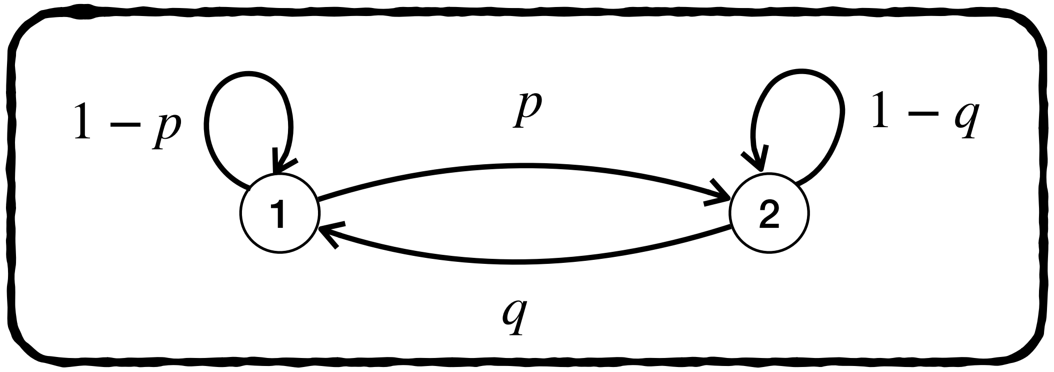


Uniqueness and Convergence



$$P = \begin{bmatrix} 1-p & p \\ q & 1-q \end{bmatrix}$$

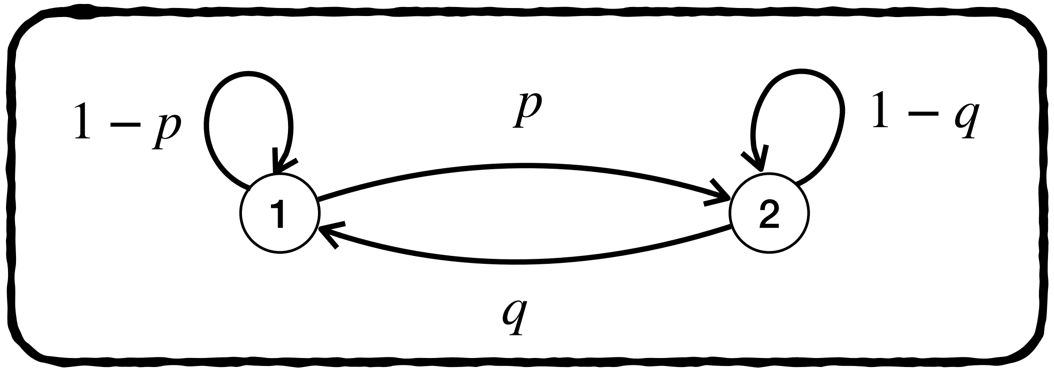
Uniqueness and Convergence



$$P = \begin{bmatrix} 1-p & p \\ q & 1-q \end{bmatrix}$$

$$\pi = \left(\frac{q}{p+q}, \frac{p}{p+q} \right)^T \text{ is a stationary dist. of } P$$

Uniqueness and Convergence

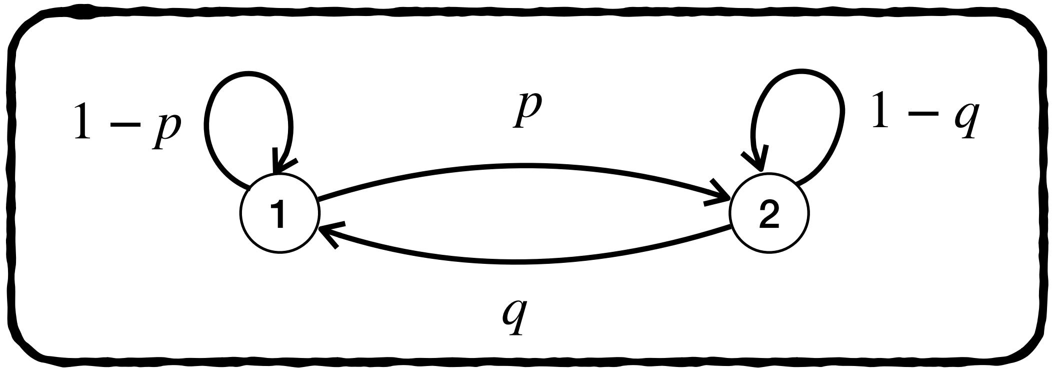


$$P = \begin{bmatrix} 1-p & p \\ q & 1-q \end{bmatrix}$$

$\pi = \left(\dfrac{q}{p+q}, \dfrac{p}{p+q} \right)^T$ is a stationary dist. of P

Start from an arbitrary $\mu_0 = \big(\mu(1), \mu(2)\big)^T$

Uniqueness and Convergence



$$P = \begin{bmatrix} 1-p & p \\ q & 1-q \end{bmatrix}$$

$$\pi = \left(\frac{q}{p+q}, \frac{p}{p+q} \right)^T \text{ is a stationary dist. of } P$$

Start from an arbitrary $\mu_0 = \big(\mu(1), \mu(2)\big)^T$

Compute $\|\mu_0^T P^t - \pi^T\|$

$$\Delta_t = |\mu_t(1) - \pi(1)|$$

$$\Delta_t = |\mu_t(1) - \pi(1)|$$

$$\begin{aligned}
\Delta_{t+1} &= \left| \mu_{t+1}(1) - \frac{q}{p+q} \right| \\
&= \left| \mu_t(1-p) + (1-\mu_t(1))q - \frac{q}{p+q} \right| \\
&= (1-p-q)\left| \mu_t(1) - \frac{q}{p+q} \right| = (1-p-q) \cdot \Delta_t
\end{aligned}$$

$$\Delta_t = |\mu_t(1) - \pi(1)|$$

$$\Delta_{t+1} = \left| \mu_{t+1}(1) - \frac{q}{p+q} \right|$$

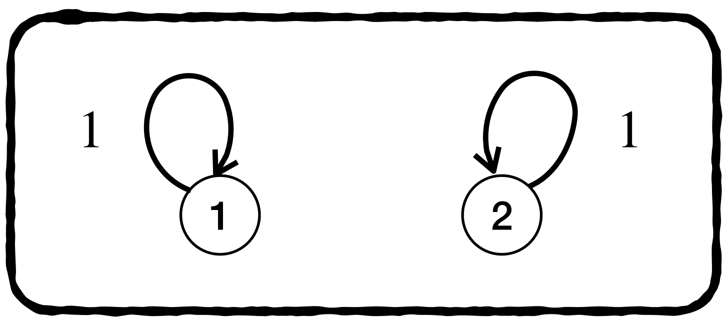
$$= \left| \mu_t(1-p) + (1-\mu_t(1))q - \frac{q}{p+q} \right|$$

$$= (1-p-q)\left| \mu_t(1) - \frac{q}{p+q} \right| = (1-p-q)\cdot\Delta_t$$

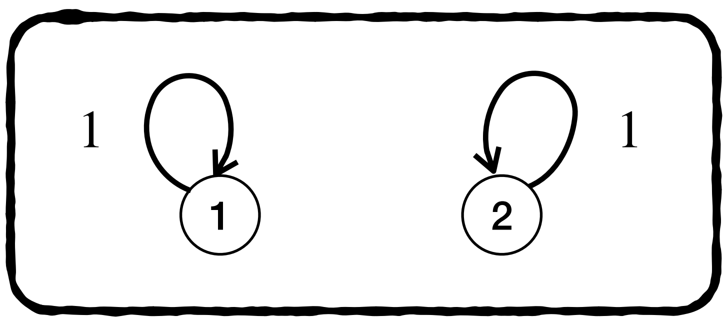
Since $p, q \in [0,1]$, there are two ways to prohibit $\Delta_t \to 0$: $p = q = 1$ or $p = q = 0$

$$p = q = 0$$

$$p = q = 0$$

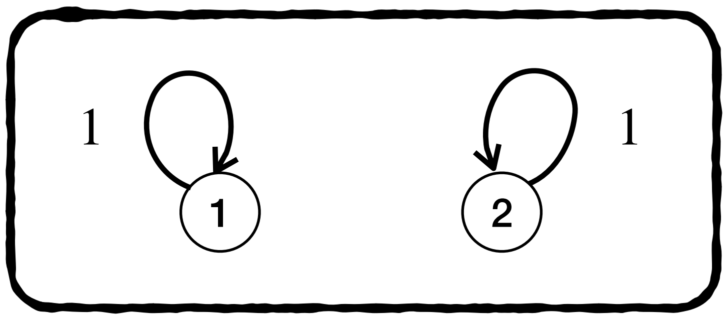


$$p = q = 0$$



$$\forall t,\ \Delta_t = \Delta_0$$

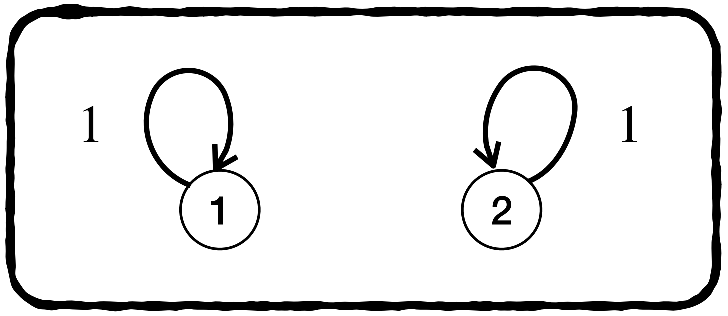
$$p = q = 0$$



$$\forall t,\ \Delta_t = \Delta_0$$

The graph is disconnected

$$p = q = 0$$

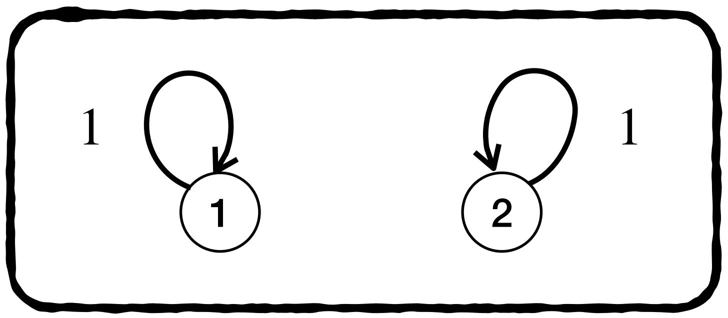


$$\forall t,\ \Delta_t = \Delta_0$$

The graph is disconnected

The chain is called reducible

$$p = q = 0$$



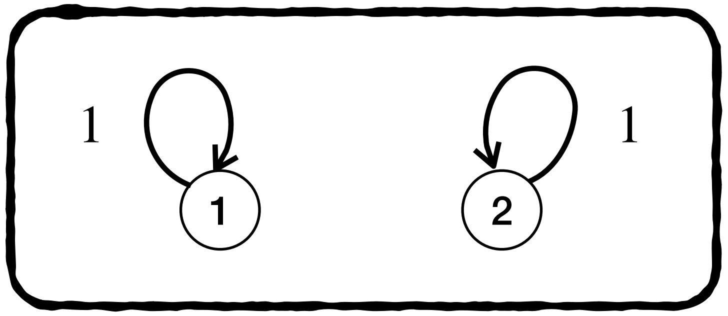
$$\forall t,\ \Delta_t = \Delta_0$$

The graph is disconnected

The chain is called reducible

In this case, the stationary distribution is not unique

$$p = q = 0$$



$$\forall t,\ \Delta_t = \Delta_0$$

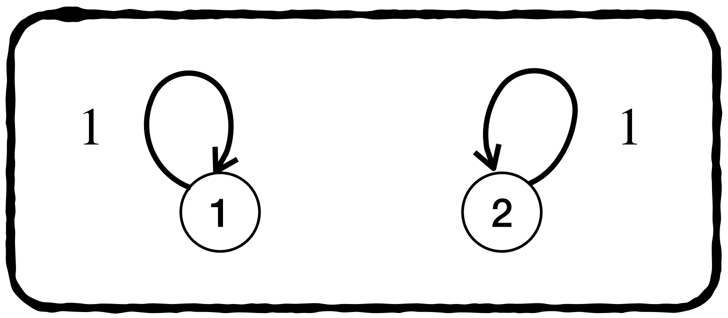
The graph is disconnected

The chain is called reducible

In this case, the stationary distribution is not unique

Chain = convex combination of small chains

$$p = q = 0$$



$\forall t,\ \Delta_t = \Delta_0$

The graph is disconnected

The chain is called reducible

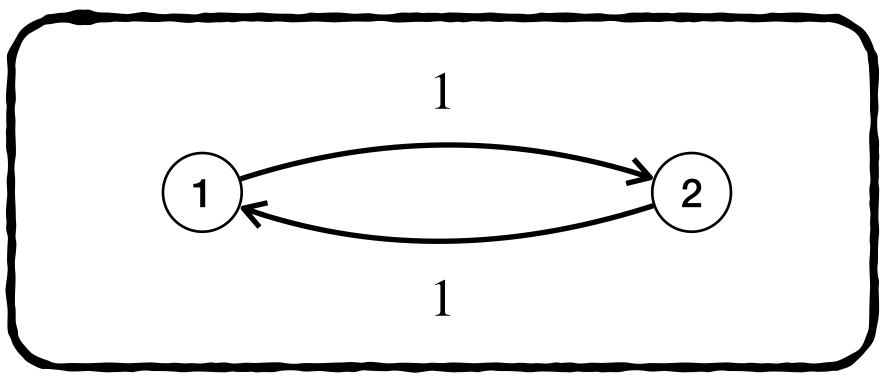
In this case, the stationary distribution is not unique

Chain = convex combination of small chains

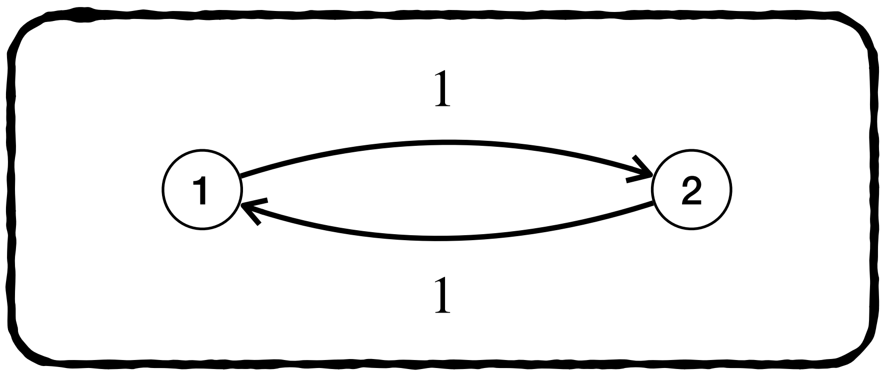
Stationary distribution=convex combination of “small” distributions

$$p = q = 1$$

$$p = q = 1$$

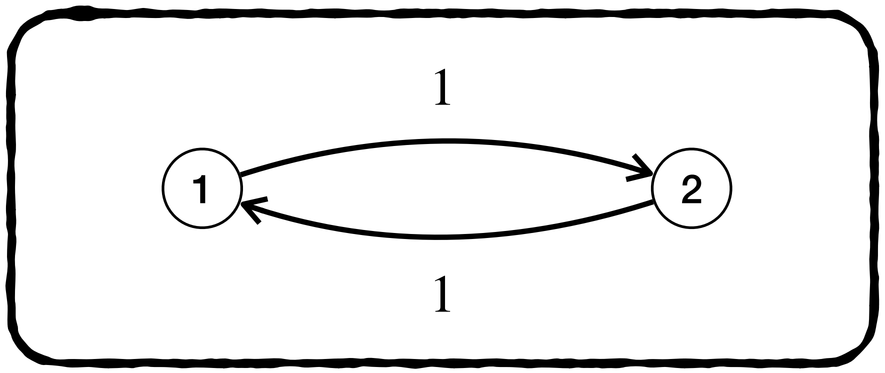


$$p = q = 1$$



$$\forall t,\ \Delta_t = -\,\Delta_{t-1}$$

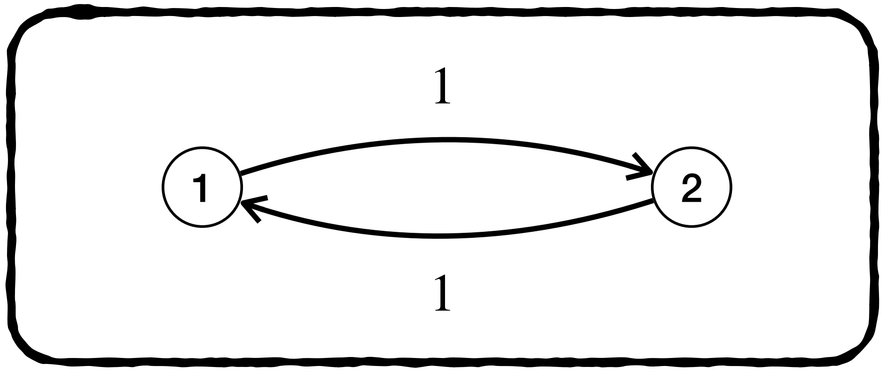
$$p = q = 1$$



$$\forall t,\ \Delta_t = -\Delta_{t-1}$$

The graph is bipartite

$$p = q = 1$$

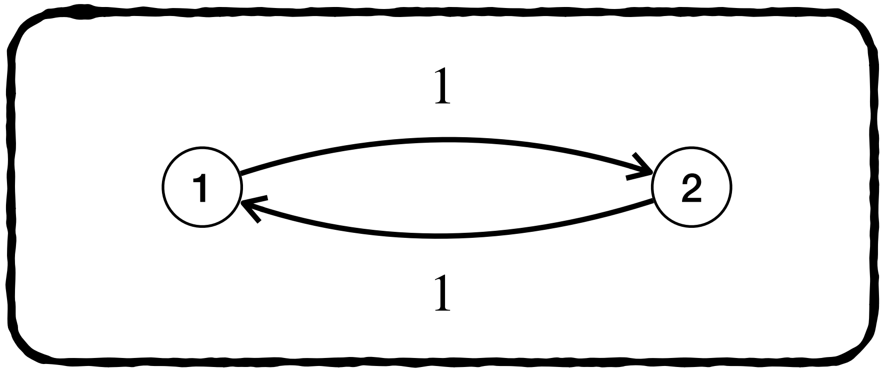


$$\forall t, \; \Delta_t = -\Delta_{t-1}$$

The graph is bipartite

The chain is called periodic

$$p = q = 1$$



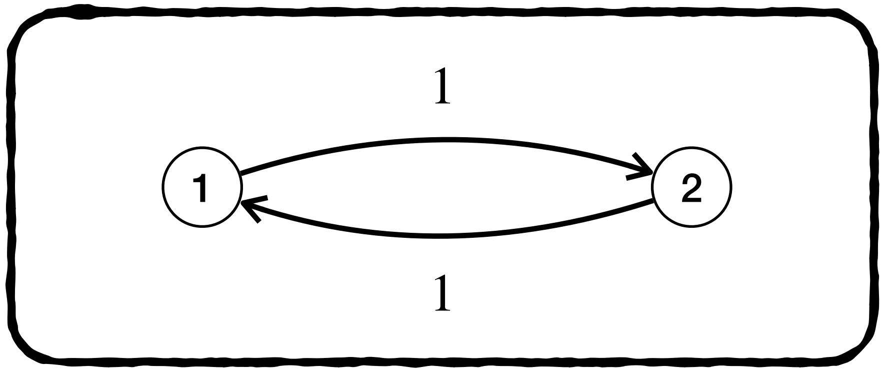
$\forall t,\ \Delta_t = -\Delta_{t-1}$

The graph is bipartite

The chain is called periodic

Formally, $\exists v,\ \gcd_{C \in C_v} |C| > 1$

$$p = q = 1$$



$\forall t,\ \Delta_t = -\Delta_{t-1}$

The graph is bipartite

The chain is called periodic

Formally, $\exists v,\ \mathrm{gcd}_{C \in C_v} |C| > 1$

In this case, not all initial distribution converges to the stationary distribution

Fundamental Theorem of Markov Chains

Fundamental Theorem of Markov Chains

If a finite chain P is irreducible and aperiodic, then it has a unique stationary distribution π. Moreover, for any initial distribution μ, it holds that

$$\lim_{t \to \infty} \mu^T P^t = \pi^T$$

Fundamental Theorem of Markov Chains

If a finite chain P is irreducible and aperiodic, then it has a unique stationary distribution π. Moreover, for any initial distribution μ, it holds that

$$\lim_{t \to \infty} \mu^T P^t = \pi^T$$

(Show on board, see the note for details)

Reversible Chains

Reversible Chains

We study a special family of Markov chains called reversible chains

Reversible Chains

We study a special family of Markov chains called reversible chains

Their transition graphs are undirected

$$x \to y \iff y \to x$$

Reversible Chains

We study a special family of Markov chains called reversible chains

Their transition graphs are undirected

$$x \to y \iff y \to x$$

A chain P and a distribution π satisfies *detailed balance condition*:

Reversible Chains

We study a special family of Markov chains called reversible chains

Their transition graphs are undirected

$$x \to y \iff y \to x$$

A chain P and a distribution π satisfies *detailed balance condition*:

$$\forall x, y \in V, \pi(x) \cdot P(x, y) = \pi(y) \cdot P(y, x)$$

Reversible Chains

We study a special family of Markov chains called reversible chains

Their transition graphs are undirected

$$x \to y \iff y \to x$$

A chain P and a distribution π satisfies *detailed balance condition*:

$$\forall x, y \in V,\ \pi(x) \cdot P(x, y) = \pi(y) \cdot P(y, x)$$

Then π is a stationary distribution of P

We study reversible chains because

We study reversible chains because

- They are quite general. For any π, one can define an reversible P whose stationary distribution is π

We study reversible chains because

- They are quite general. For any π, one can define an reversible P whose stationary distribution is π

Helpful for Sampling

We study reversible chains because

- They are quite general. For any π, one can define an reversible P whose stationary distribution is π

  Helpful for Sampling

- We have powerful tools (spectral method) to analyze reversible chains

Spectral Decomposition Theorem

Spectral Decomposition Theorem

An $n \times n$ symmetric matrix A has n real eigenvalues $\lambda_1, \ldots, \lambda_n$ with corresponding eigenvectors $\mathbf{v}_1, \ldots, \mathbf{v}_n$ which are orthogonal. Moreover, it holds that

$$A = V \Lambda V^T$$

Spectral Decomposition Theorem

An $n \times n$ symmetric matrix A has n real eigenvalues $\lambda_1, \ldots, \lambda_n$ with corresponding eigenvectors $\mathbf{v}_1, \ldots, \mathbf{v}_n$ which are orthogonal. Moreover, it holds that

$$A = V \Lambda V^T$$

where $V = [\mathbf{v}_1, \ldots, \mathbf{v}_n]$ and $\Lambda = \operatorname{diag}(\lambda_1, \ldots, \lambda_n)$

Spectral Decomposition Theorem

An $n \times n$ symmetric matrix A has n real eigenvalues $\lambda_1, \ldots, \lambda_n$ with corresponding eigenvectors $\mathbf{v}_1, \ldots, \mathbf{v}_n$ which are orthogonal. Moreover, it holds that

$$A = V \Lambda V^T$$

where $V = [\mathbf{v}_1, \ldots, \mathbf{v}_n]$ and $\Lambda = \mathrm{diag}(\lambda_1, \ldots, \lambda_n)$

Equivalently, $A = \sum_{i=1}^{n} \lambda_i \mathbf{v}_i \mathbf{v}_i^T$

Spectral Decomposition Theorem for Reversible Chains

Spectral Decomposition Theorem for Reversible Chains

π is a stationary distribution of a reversible chain P

Spectral Decomposition Theorem for Reversible Chains

π is a stationary distribution of a reversible chain P

Define an inner product $\langle \cdot , \cdot \rangle_\pi$ on $\mathbb{R}^n$:

Spectral Decomposition Theorem for Reversible Chains

π is a stationary distribution of a reversible chain P

Define an inner product $\langle \cdot , \cdot \rangle_\pi$ on $\mathbb{R}^n$:

$$\langle \mathbf{x}, \mathbf{y} \rangle_\pi = \sum_{i=1}^{n} \pi(i) \cdot \mathbf{x}(i) \cdot \mathbf{y}(i) = \mathbf{x}^T D_\pi \mathbf{y},$$

Spectral Decomposition Theorem for Reversible Chains

π is a stationary distribution of a reversible chain P

Define an inner product $\langle \cdot , \cdot \rangle_\pi$ on $\mathbb{R}^n$:

$$\langle \mathbf{x}, \mathbf{y} \rangle_\pi = \sum_{i=1}^{n} \pi(i) \cdot \mathbf{x}(i) \cdot \mathbf{y}(i) = \mathbf{x}^T D_\pi \mathbf{y},$$

where $D_\pi = \mathrm{diag}(\pi_1, \ldots, \pi_n)$

Spectral Decomposition Theorem for Reversible Chains

π is a stationary distribution of a reversible chain P

Define an inner product $\langle \cdot , \cdot \rangle_\pi$ on $\mathbb{R}^n$:

$$\langle \mathbf{x}, \mathbf{y} \rangle_\pi = \sum_{i=1}^{n} \pi(i) \cdot \mathbf{x}(i) \cdot \mathbf{y}(i) = \mathbf{x}^T D_\pi \mathbf{y},$$

where $D_\pi = \mathrm{diag}(\pi_1, \ldots, \pi_n)$

Consider the Hilbert space $\mathbb{R}^n$ endowed with $\langle \cdot , \cdot \rangle_\pi$

Let $P \in \mathbb{R}^{n\times n}$ be reversible with respect to π. It has n real eigenvalues $\lambda_1, \ldots, \lambda_n$ with corresponding eigenvectors $\mathbf{v}_1, \ldots, \mathbf{v}_n$ which are orthogonal in $(\mathbb{R}^n, \langle \cdot\,, \rangle_\pi)$. Moreover

$$P = \sum_{i=1}^{n} \lambda_i \mathbf{v}_i \mathbf{v}_i^T D_\pi$$

Let $P \in \mathbb{R}^{n\times n}$ be reversible with respect to π. It has n real eigenvalues $\lambda_1, \ldots, \lambda_n$ with corresponding eigenvectors $\mathbf{v}_1, \ldots, \mathbf{v}_n$ which are orthogonal in $(\mathbb{R}^n, \langle \cdot , \rangle_\pi)$. Moreover

$$P = \sum_{i=1}^{n} \lambda_i \mathbf{v}_i \mathbf{v}_i^T D_\pi$$

Spectral Decomposition Theorem

An $n \times n$ symmetric matrix A has n real eigenvalues $\lambda_1, \ldots, \lambda_n$ with corresponding eigenvectors $\mathbf{v}_1, \ldots, \mathbf{v}_n$ which are orthogonal. Moreover, it holds that

$$A = V\Lambda V^T$$

where $V = [\mathbf{v}_1, \ldots, \mathbf{v}_n]$ and $\Lambda = \mathrm{diag}(\lambda_1, \ldots, \lambda_n)$

Equivalently, $A = \sum_{i=1}^{n} \lambda_i \mathbf{v}_i \mathbf{v}_i^T$

Let $P \in \mathbb{R}^{n \times n}$ be reversible with respect to π. It has n real eigenvalues $\lambda_1, \ldots, \lambda_n$ with corresponding eigenvectors $\mathbf{v}_1, \ldots, \mathbf{v}_n$ which are orthogonal in $(\mathbb{R}^n, \langle \cdot , \rangle_\pi)$. Moreover

$$P = \sum_{i=1}^{n} \lambda_i \mathbf{v}_i \mathbf{v}_i^T D_\pi$$

Proof. Reduce to the symmetric case.

Spectral Decomposition Theorem

An $n \times n$ symmetric matrix A has n real eigenvalues $\lambda_1, \ldots, \lambda_n$ with corresponding eigenvectors $\mathbf{v}_1, \ldots, \mathbf{v}_n$ which are orthogonal. Moreover, it holds that

$$A = V \Lambda V^T$$

where $V = [\mathbf{v}_1, \ldots, \mathbf{v}_n]$ and $\Lambda = \mathrm{diag}(\lambda_1, \ldots, \lambda_n)$

Equivalently, $A = \sum_{i=1}^{n} \lambda_i \mathbf{v}_i \mathbf{v}_i^T$

Properties of Eigenvalues

Properties of Eigenvalues

π is a stationary distribution of a reversible chain P

Properties of Eigenvalues

π is a stationary distribution of a reversible chain P

The eigenvalues of P are $\lambda_1 \leq \lambda_2 \ldots \leq \lambda_n$

Properties of Eigenvalues

π is a stationary distribution of a reversible chain P

The eigenvalues of P are $\lambda_1 \leq \lambda_2 \ldots \leq \lambda_n$

- $\lambda_n = 1$

Properties of Eigenvalues

π is a stationary distribution of a reversible chain P

The eigenvalues of P are $\lambda_1 \leq \lambda_2 \ldots \leq \lambda_n$

- $\lambda_n = 1$
- $\lambda_1 \geq -1$ and $\lambda_1 = -1$ if and only if P is bipartite

Properties of Eigenvalues

π is a stationary distribution of a reversible chain P

The eigenvalues of P are $\lambda_1 \leq \lambda_2 \ldots \leq \lambda_n$

- $\lambda_n = 1$
- $\lambda_1 \geq -1$ and $\lambda_1 = -1$ if and only if P is bipartite
- $\lambda_{n-1} = 1$ if and only if P is reducible

Properties of Eigenvalues

π is a stationary distribution of a reversible chain P

The eigenvalues of P are $\lambda_1 \leq \lambda_2 \ldots \leq \lambda_n$

Proof next week!

- $\lambda_n = 1$
- $\lambda_1 \geq -1$ and $\lambda_1 = -1$ if and only if P is bipartite
- $\lambda_{n-1} = 1$ if and only if P is reducible

$$\boxed{P = \sum_{i=1}^{n} \lambda_i \mathbf{v}_i \mathbf{v}_i^T D_\pi}$$

$$\boxed{P = \sum_{i=1}^{n} \lambda_i \mathbf{v}_i \mathbf{v}_i^T D_\pi}$$

$$P^t = \sum_{i=1}^{n} \lambda_i^t \mathbf{v}_i \mathbf{v}_i^T D_\pi$$

$$P = \sum_{i=1}^{n} \lambda_i \mathbf{v}_i \mathbf{v}_i^T D_\pi$$

$$P^t = \sum_{i=1}^{n} \lambda_i^t \mathbf{v}_i \mathbf{v}_i^T D_\pi$$

- $\lambda_n = 1$
- $\lambda_1 \geq -1$ and $\lambda_1 = -1$ if and only if P is bipartite
- $\lambda_{n-1} = 1$ if and only if P is reducible

$$P = \sum_{i=1}^{n} \lambda_i \mathbf{v}_i \mathbf{v}_i^T D_\pi$$

$$P^t = \sum_{i=1}^{n} \lambda_i^t \mathbf{v}_i \mathbf{v}_i^T D_\pi$$

- $\lambda_n = 1$
- $\lambda_1 \geq -1$ and $\lambda_1 = -1$ if and only if P is bipartite
- $\lambda_{n-1} = 1$ if and only if P is reducible

If P is irreducible ($\lambda_{n-1} < 1$) and aperiodic ($\lambda_1 > -1$)

$$P = \sum_{i=1}^{n} \lambda_i \mathbf{v}_i \mathbf{v}_i^T D_\pi$$

- $\lambda_n = 1$
- $\lambda_1 \geq -1$ and $\lambda_1 = -1$ if and only if P is bipartite
- $\lambda_{n-1} = 1$ if and only if P is reducible

$$P^t = \sum_{i=1}^{n} \lambda_i^t \mathbf{v}_i \mathbf{v}_i^T D_\pi$$

If P is irreducible ($\lambda_{n-1} < 1$) and aperiodic ($\lambda_1 > -1$)

$$\lim_{t\to\infty} P^t = \mathbf{1}\mathbf{1}^T D_\pi = \begin{bmatrix} \pi^T \\ \pi^T \\ \vdots \\ \pi^T \end{bmatrix}$$