Lecture 11: Markov Chains, Random Walks, Mixing Time, Page Rank

Rafael Oliveira

University of Waterloo
Cheriton School of Computer Science

rafael.oliveira.teaching@gmail.com

June 15, 2021

Overview

- Introduction
  - Why Random Walks & Markov Chains?
  - Basics on Theory of (finite) Markov Chains
- Main Topics
  - Fundamental Theorem of Markov Chains
  - Page Rank
- Conclusion
- Acknowledgements

What is a Random Walk?

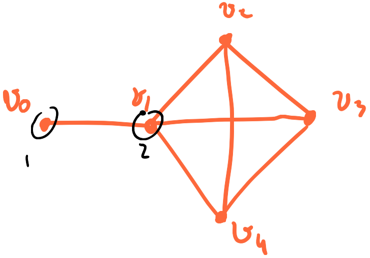
Given a graph $G(V, E)$

1. random walk starts from a vertex v_0
2. at each time step it moves to a *uniformly random neighbor* of the current vertex in the graph

$$v_{t+1} \leftarrow_R N_G(v_t)$$

$v_0 \xrightarrow{1} v_1 \xrightarrow{1/4} v_2$

$v_1 \xrightarrow{1/4} v_0$



$v_0 \xrightarrow{1} v_1 \xrightarrow{1/4} v_0 \xrightarrow{1} v_1 \xrightarrow{1/4} v_2 \xrightarrow{1/3} v_3$

What is a Random Walk?

Given a graph $G(V, E)$

1. random walk starts from a vertex v_0
2. at each time step it moves to a *uniformly random neighbor* of the current vertex in the graph

$$v_{t+1} \leftarrow_R N_G(v_t)$$

$p^{(0)} = (1, 0, \ldots, 0)$

$v_0 = 1$

Basic questions involving random walks:

graph G $\quad V = \{1, \ldots, n\}$

at time t $\quad p^{(t)} = (p_1^{(t)}, p_2^{(t)}, \ldots, p_n^{(t)})$

$p_i^{(t)} \leftarrow$ probability of being at i at time t

What is a Random Walk?

Given a graph $G(V, E)$

1. random walk starts from a vertex v_0
2. at each time step it moves to a *uniformly random neighbor* of the current vertex in the graph

$$v_{t+1} \leftarrow_R N_G(v_t)$$

Basic questions involving random walks:

- *Stationary distribution:* does the random walk converge to a "stable" distribution? If it does, what is this distribution?

$p^{(t)} = p^{(t+1)}$ "stable"

stationary distribution

What is a Random Walk?

Given a graph $G(V, E)$

1. random walk starts from a vertex v_0
2. at each time step it moves to a *uniformly random neighbor* of the current vertex in the graph

$$v_{t+1} \leftarrow_R N_G(v_t)$$

Basic questions involving random walks:

- *Stationary distribution:* does the random walk converge to a "stable" distribution? If it does, what is this distribution?
- *Mixing time:* how long does it take for the walk to converge to the stationary distribution?

today's lecture

What is a Random Walk?

Given a graph $G(V, E)$

1. random walk starts from a vertex v_0
2. at each time step it moves to a *uniformly random neighbor* of the current vertex in the graph

$$v_{t+1} \leftarrow_R N_G(v_t)$$

Basic questions involving random walks:

- *Stationary distribution:* does the random walk converge to a "stable" distribution? If it does, what is this distribution?
- *Mixing time:* how long does it take for the walk to converge to the stationary distribution?
- *Hitting time:* starting from a vertex v_0, what is expected number of steps until it reaches a vertex v_f?

What is a Random Walk?

Given a graph $G(V, E)$

1. random walk starts from a vertex v_0
2. at each time step it moves to a *uniformly random neighbor* of the current vertex in the graph

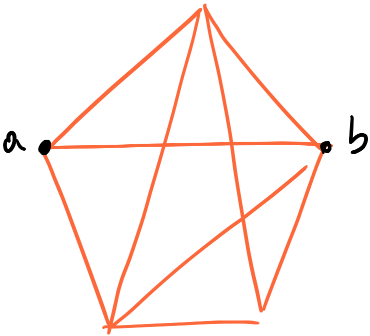
$$v_{t+1} \leftarrow_R N_G(v_t)$$

Basic questions involving random walks:

- *Stationary distribution:* does the random walk converge to a "stable" distribution? If it does, what is this distribution?
- *Mixing time:* how long does it take for the walk to converge to the stationary distribution?
- *Hitting time:* starting from a vertex v_0, what is expected number of steps until it reaches a vertex v_f?
- *Cover time:* how long does it take to reach every vertex of the graph at least once?

Random Walk: Example

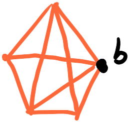
- Suppose $G(V, E) = K_n$, the complete graph, $a, b \in V$ two vertices



Random Walk: Example

- Suppose $G(V, E) = K_n$, the complete graph, $a, b \in V$ two vertices
  1. What is expected number of steps to reach b in simple random walk starting at a? (i.e., hitting time)

$$x_t = \Pr[\text{reach } b \text{ first time at time } t \text{ starting at } a]$$



$$x_t = \left(\frac{n-2}{n-1}\right)^{t-1} \cdot \underbrace{\frac{1}{n-1}}_{\text{reaching } b \text{ at time } t}$$

$$\mathbb{E}[\text{steps to reach } b] = \sum_{t \geq 1} t \cdot x_t = \sum_{t \geq 1} \frac{t}{n-1} \cdot \left(\frac{n-2}{n-1}\right)^{t-1}$$

$$= \frac{1}{n-1} \sum_{t \geq 1} t \left(\frac{n-2}{n-1}\right)^{t-1} = \frac{1}{n-1}\left(\frac{d}{dx} \underbrace{\sum_{t \geq 1} x^t}_{\frac{x}{1-x}}\right)\Bigg|_{x = \frac{n-2}{n-1}} = \frac{1}{n-1} \cdot \frac{1}{(1-x)^2}\Bigg|_{x = \frac{n-2}{n-1}} = \boxed{n-1}$$

$$\frac{1 \cdot (1-x) - x(-1)}{(1-x)^2} = \frac{1}{(1-x)^2}$$

Random Walk: Example

- Suppose $G(V, E) = K_n$, the complete graph, $a, b \in V$ two vertices
  1. What is expected number of steps to reach b in simple random walk starting at a? (i.e., hitting time)
  2. Starting from a, what is expected number of steps to visit all vertices? (i.e, cover time)

Random Walk: Example

- Suppose $G(V, E) = K_n$, the complete graph, $a, b \in V$ two vertices
  1. What is expected number of steps to reach b in simple random walk starting at a? (i.e., hitting time)
  2. Starting from a, what is expected number of steps to visit all vertices? (i.e, cover time)
  3. Stationary Distribution?

$$\pi = \left(\frac{1}{n}, \frac{1}{n}, \frac{1}{n}, \ldots, \frac{1}{n}\right)$$

$$\Pr[a \text{ in next step}] = \sum_{i=1}^{n} \Pr[\text{currently at } i] \cdot \Pr[i \to a]$$

$$= \sum_{i=1}^{n} \underbrace{\frac{1}{n}}_{\text{because of } \pi} \cdot \frac{1}{n-1}\left(1 - \delta_{ia}\right) = \frac{(n-1)}{n(n-1)} = \frac{1}{n}$$

$$\delta_{ia} = \begin{cases} 1 & \text{if } i = a \\ 0 & \text{otherwise} \end{cases}$$

Random Walk: Example

- Suppose $G(V, E) = K_n$, the complete graph, $a, b \in V$ two vertices
  1. What is expected number of steps to reach b in simple random walk starting at a? (i.e., hitting time)
  2. Starting from a, what is expected number of steps to visit all vertices? (i.e, cover time)
  3. Stationary Distribution?
  4. Mixing time? (we'll do it later)

Random Walk: Example

- Suppose $G(V, E) = K_n$, the complete graph, $a, b \in V$ two vertices
  1. What is expected number of steps to reach b in simple random walk starting at a? (i.e., hitting time)
  2. Starting from a, what is expected number of steps to visit all vertices? (i.e, cover time)
  3. Stationary Distribution?
  4. Mixing time? (we'll do it later)
- **Practice question:** Compare question 2 to coupon collector problem!

$$(n-1)\underbrace{H_{n-1}}_{\text{harmonic number}}$$

What is a Markov Chain?

Random walk is a special kind of *stochastic process*:

$$\Pr[X_t = v_t \mid \overbrace{X_0 = v_0, \ldots, X_{t-1} = v_{t-1}}^{\text{history of random walk}}] = \Pr[X_t = v_t \mid \underbrace{X_{t-1} = v_{t-1}]}_{\text{only depend on information at time } t-1}$$

"forgetful stochastic process"

X_t random variable taking values over V

where will the random walk be at time t

What is a Markov Chain?

Random walk is a special kind of *stochastic process*:

$$\Pr[X_t = v_t \mid X_0 = v_0, \ldots, X_{t-1} = v_{t-1}] = \Pr[X_t = v_t \mid X_{t-1} = v_{t-1}]$$

Probability that we are at vertex v_t at time t only depends on the state of our process at time $t-1$.

What is a Markov Chain?

Random walk is a special kind of *stochastic process*:

$$\boxed{\Pr[X_t = v_t \mid X_0 = v_0, \ldots, X_{t-1} = v_{t-1}] = \Pr[X_t = v_t \mid X_{t-1} = v_{t-1}]}$$

Probability that we are at vertex v_t at time t only depends on the state of our process at time $t - 1$.

Process is "*forgetful*"

Markov chain is characterized by this property.

Why study Markov Chains and Random Walks?

Markov Chains and Random Walks are ubiquitous in randomized algorithms.

Why study Markov Chains and Random Walks?

Markov Chains and Random Walks are ubiquitous in randomized algorithms.

- Page Rank algorithm (today's lecture)

Why study Markov Chains and Random Walks?

Markov Chains and Random Walks are ubiquitous in randomized algorithms.

- Page Rank algorithm (today's lecture)
- Approximation algorithms for counting problems [Karp, Luby & Madras]
    - Permanent of non-negative matrices [Jerrum, Vigoda & Sinclair] (*great final project topic!*)

Why study Markov Chains and Random Walks?

Markov Chains and Random Walks are ubiquitous in randomized algorithms.

- Page Rank algorithm (today's lecture)
- Approximation algorithms for counting problems [Karp, Luby & Madras]
  - Permanent of non-negative matrices [Jerrum, Vigoda & Sinclair] (*great final project topic!*)
- Sampling Problems
  - Gibbs sampling in statistical physics (*great final project topic!*)
  - many more places

Why study Markov Chains and Random Walks?

Markov Chains and Random Walks are ubiquitous in randomized algorithms.

- Page Rank algorithm (today's lecture)
- Approximation algorithms for counting problems [Karp, Luby & Madras]
  - Permanent of non-negative matrices [Jerrum, Vigoda & Sinclair] (*great final project topic!*)
- Sampling Problems
  - Gibbs sampling in statistical physics (*great final project topic!*)
  - many more places
- Probability amplification without too much randomness (efficient)
  - random walks on expander graphs (*great final project topic!*)

Why study Markov Chains and Random Walks?

Markov Chains and Random Walks are ubiquitous in randomized algorithms.

- Page Rank algorithm (today's lecture)
- Approximation algorithms for counting problems [Karp, Luby & Madras]
  - Permanent of non-negative matrices [Jerrum, Vigoda & Sinclair] (*great final project topic!*)
- Sampling Problems
  - Gibbs sampling in statistical physics (*great final project topic!*)
  - many more places
- Probability amplification without too much randomness (efficient)
  - random walks on expander graphs (*great final project topic!*)
- many more

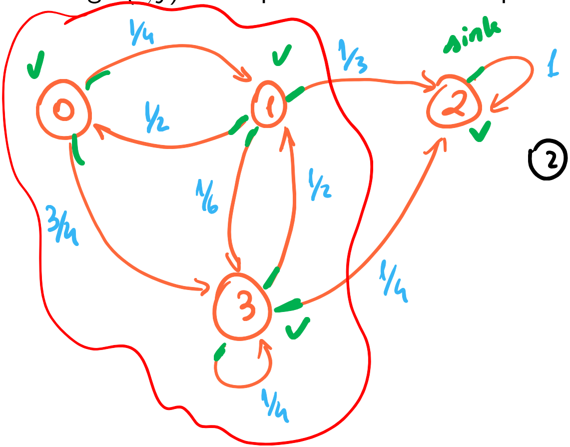
Representing Finite Markov Chains

Markov chain can be seen as weighted directed graph.

Representing Finite Markov Chains

Markov chain can be seen as weighted directed graph.

- Vertex is a state of Markov chain
- edge (i, j) corresponds to transition probability from i to j



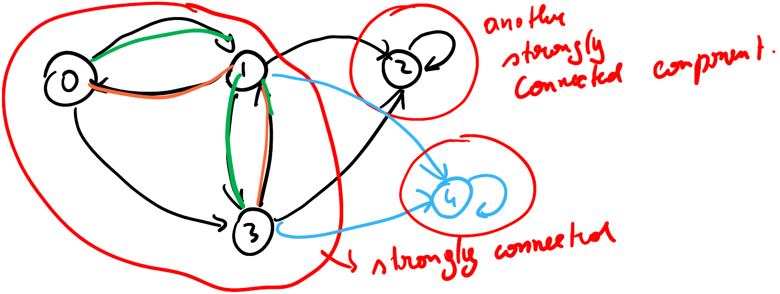
① weights on edges are ≥ 0

② sum of weights coming out of each node $= 1$ (probability distribution)

Representing Finite Markov Chains

Markov chain can be seen as weighted directed graph.

- Vertex is a state of Markov chain
- edge (i, j) corresponds to transition probability from i to j



- Markov Chain *irreducible* if underlying directed graph is *strongly connected* (i.e. there is directed path from i to j for any pair $i, j \in V$)

Representing Finite Markov Chains

Markov chain can be seen in weighted adjacency matrix format.

Representing Finite Markov Chains

Markov chain can be seen in weighted adjacency matrix format.

- $P \in \mathbb{R}^{n \times n}$ transition matrix

Representing Finite Markov Chains

Markov chain can be seen in weighted adjacency matrix format.

- $P \in \mathbb{R}^{n \times n}$ transition matrix
- entry $P_{i,j}$ corresponds to transition probability from i to j

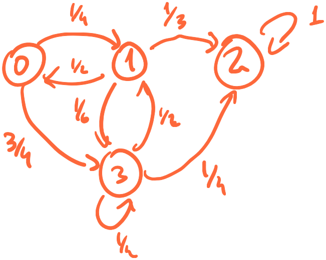
Representing Finite Markov Chains

Markov chain can be seen in weighted adjacency matrix format.

- $P \in \mathbb{R}^{n \times n}$ transition matrix
- entry $P_{i,j}$ corresponds to transition probability from i to j
- $p_t \in \mathbb{R}^n$ probability vector: $p_t(i) := \Pr[\text{being at } \textit{state } i \text{ at } \textit{time } t]$

Representing Finite Markov Chains

Markov chain can be seen in weighted adjacency matrix format.



$$P = \begin{array}{c|cccc} & 0 & 1 & 2 & 3 \\ \hline 0 & 0 & 1/4 & 0 & 3/4 \\ 1 & 1/2 & 0 & 1/3 & 1/6 \\ 2 & 0 & 0 & 1 & 0 \\ 3 & 0 & 1/2 & 1/4 & 1/4 \end{array}$$

$p_0 = (1, 0, 0, 0)$

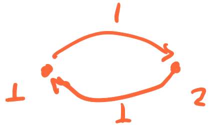
$p_1 = p_0 \cdot P = (\underset{0}{0}, \underset{1}{1/4}, \underset{2}{0}, \underset{3}{3/4})$

- $P \in \mathbb{R}^{n \times n}$ transition matrix
- entry $P_{i,j}$ corresponds to transition probability from i to j
- $\boxed{p_t \in \mathbb{R}^n}$ probability vector: $p_t(i) :=$ Pr[being at *state i* at *time t*]
- Transition given by

row vector

$$\boxed{p_{t+1} = p_t \cdot P}$$

Properties of Markov Chains



- *Period* of a state i is:

$$\gcd\{t \in \mathbb{N} \mid P^t_{i,i} > 0\}$$

That is, gcd of all times t such that the probability of starting at state i and being back at i at time t is positive

$P^t_{i,i}$ = Pr [back at state i after t steps if started at i]

$P^1_{i,i}$ = Pr. stay at state i

$P^2_{i,i}$ = Pr $[\, i \to v \to i \,]$

Properties of Markov Chains

- *Period* of a state i is:

$$\gcd\{t \in \mathbb{N} \mid P^t_{i,i} > 0\}$$

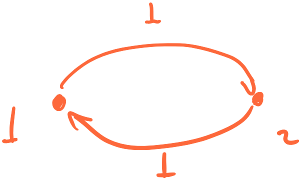
  That is, gcd of all times t such that the probability of starting at state i and being back at i at time t is positive
- State i is *aperiodic* if its period is 1.

Properties of Markov Chains

- *Period* of a state i is:

$$\gcd\{t \in \mathbb{N} \mid P^t_{i,i} > 0\}$$

  That is, gcd of all times t such that the probability of starting at state i and being back at i at time t is positive
- State i is *aperiodic* if its period is 1.
- Markov Chain aperiodic if *all states* are aperiodic (otherwise periodic)



in periodic Markov chain we may never reach stationary distribution!

$\pi = (\frac{1}{2}, \frac{1}{2})$

$p_0 = (1, 0)$

Properties of Markov Chains

- *Period* of a state i is:

$$\gcd\{t \in \mathbb{N} \mid P^t_{i,i} > 0\}$$

  That is, gcd of all times t such that the probability of starting at state i and being back at i at time t is positive
- State i is *aperiodic* if its period is 1.
- Markov Chain aperiodic if *all states* are aperiodic (otherwise periodic)
  - Bipartite graphs yield periodic Markov Chains

Properties of Markov Chains

- *Period* of a state i is:

$$\gcd\{t \in \mathbb{N} \mid P^t_{i,i} > 0\}$$

  That is, gcd of all times t such that the probability of starting at state i and being back at i at time t is positive
- State i is *aperiodic* if its period is 1.
- Markov Chain aperiodic if *all states* are aperiodic (otherwise periodic)
  - Bipartite graphs yield periodic Markov Chains

Lemma

For any finite, irreducible and aperiodic Markov Chain, there exists $T < \infty$ such that

$$P^t_{i,j} > 0 \quad \textit{for any } i, j \in V \textit{ and } t \geq T.$$

at some point we will reach every vertex with > 0 probability if we take enough steps

Stationary Distributions

Definition (Stationary Distribution)

A stationary distribution of a Markov Chain is a probability distribution $\pi \in \mathbb{R}^n$ such that

$$\pi P = \pi.$$

transition matrix

Stationary Distributions

Definition (Stationary Distribution)

A stationary distribution of a Markov Chain is a probability distribution $\pi \in \mathbb{R}^n$ such that

$$\pi P = \pi.$$

- Informally, π is an "equilibrium/fixed point" state, as we have $\pi = \pi P^t$ for any $t \geq 0$.

$$\pi P^2 = (\pi P) \cdot P = \pi P = \pi$$

Stationary Distributions

Definition (Stationary Distribution)

A stationary distribution of a Markov Chain is a probability distribution $\pi \in \mathbb{R}^n$ such that

$$\pi P = \pi.$$

- Informally, π is an "equilibrium/fixed point" state, as we have $\pi = \pi P^t$ for any $t \geq 0$.
- **Intuition:** If we run finite, irreducible and aperiodic Markov Chain long enough, we will converge to a stationary distribution.

Stationary Distributions

Definition (Stationary Distribution)

A stationary distribution of a Markov Chain is a probability distribution $\pi \in \mathbb{R}^n$ such that

$$\pi P = \pi.$$

- Informally, π is an "equilibrium/fixed point" state, as we have $\pi = \pi P^t$ for any $t \geq 0$.
- **Intuition:** If we run finite, irreducible and aperiodic Markov Chain long enough, we will converge to a stationary distribution.
  - what do you mean by converge here?

Stationary Distributions

Definition (Stationary Distribution)

A stationary distribution of a Markov Chain is a probability distribution $\pi \in \mathbb{R}^n$ such that

$$\pi P = \pi.$$

- Informally, π is an "equilibrium/fixed point" state, as we have $\pi = \pi P^t$ for any $t \geq 0$.
- **Intuition:** If we run finite, irreducible and aperiodic Markov Chain long enough, we will converge to a stationary distribution.
  - what do you mean by converge here?
- Given two distributions $p, q \in \mathbb{R}^n$, their *total variational distance* is

$$\Delta_{TV}(p,q) = \frac{1}{2}\sum_{i=1}^{n} |p_i - q_i| = \frac{1}{2} \cdot \cancel{\|p - q\|_1}$$

$\|p-q\|_1$

Stationary Distributions

Definition (Stationary Distribution)

A stationary distribution of a Markov Chain is a probability distribution $\pi \in \mathbb{R}^n$ such that

$$\pi P = \pi.$$

- Informally, π is an "equilibrium/fixed point" state, as we have $\pi = \pi P^t$ for any $t \geq 0$.
- **Intuition:** If we run finite, irreducible and aperiodic Markov Chain long enough, we will converge to a stationary distribution.
  - what do you mean by converge here?
- Given two distributions $p, q \in \mathbb{R}^n$, their *total variational distance* is

$$\Delta_{TV}(p, q) = \frac{1}{2} \sum_{i=1}^{n} |p_i - q_i| = \frac{1}{2} \cdot \|p, q\|_1$$

- p_t *converges* to q iff $\lim_{t \to \infty} \Delta_{TV}(p_t, q) = 0$

Mixing Time of Markov Chains

Definition (Mixing Time)

The ε-mixing time of a Markov Chain is the smallest t such that

$$\Delta_{TV}(p_t, \pi) \leq \varepsilon$$

regardless of the initial starting distribution p_0.

Practice problem: For complete graph K_n compute its ε-mixing time.

Hint: look at eigenvalues and eigenvectors of transition matrix $P = \begin{pmatrix} 0 & & 1 \\ & \ddots & \\ 1 & & 0 \end{pmatrix} \cdot \frac{1}{n-1}$

Mixing Time of Markov Chains

Definition (Mixing Time)

The ε-mixing time of a Markov Chain is the smallest t such that

$$\Delta_{TV}(p_t, \pi) \leq \varepsilon$$

regardless of the initial starting distribution p_0.

- Eigenvalues $\lambda_1 = 1, \lambda_2 = \cdots = \lambda_n = -1/(n-1)$, corresponding eigenvectors $v_1, \ldots, v_n$ (orthonormal)

K_n complete graph

all one's matrix

$$P = \begin{pmatrix} 0 & & \frac{1}{n-1} \\ & \ddots & \\ \frac{1}{n-1} & & 0 \end{pmatrix} = -\frac{I}{n-1} + \frac{J}{n-1}$$

$$P = \frac{1}{n-1}\begin{pmatrix} 0 & & 1 \\ & \ddots & \\ 1 & & 0 \end{pmatrix} = \underbrace{\frac{1}{n-1} J}_{\text{rank } 1} - \frac{1}{n-1} I$$

$$J = \vec{1} \cdot \vec{1}^T$$

$$\begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix} (1 \ 1 \cdots 1)$$

eigenvectors

$$v_1 = \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix} \cdot \frac{1}{\sqrt{n}}$$

$$v_2$$

$$\vdots$$

$$\underbrace{v_n}_{\text{orthonormal}}$$

orthonormal (l_2-norm $=1$)

eigenvalues

$$\lambda_1 = 1$$

$$\lambda_2 = -1/n-1$$

$$\lambda_3 = -1/n-1$$

$$\vdots$$

$$\lambda_n = -1/n-1$$

eigenvectors of P:

$i > 1$

$$J \cdot v_i = 0$$

$$v_i \perp \vec{1}$$

$$P v_i = -\frac{1}{n-1} I \cdot v_i$$

$$= -\frac{1}{n-1} \cdot v_i$$

$$P = \lambda_1 v_1 v_1^T + \sum_{i=2}^{n} \lambda_i v_i v_i^T$$

$$\lambda_1 = 1 \qquad \lambda_2 = \dots = \lambda_n = -\frac{1}{n-1}$$

$$P^t = \lambda_1^t v_1 v_1^T + \sum_{i=2}^{n} \lambda_i^t v_i v_i^T$$

$$= v_1 v_1^T + \underbrace{\sum_{i=2}^{n} \lambda_i^t v_i v_i^T}_{\text{if } t \text{ large then}}$$

$$\lambda_i \longmapsto 0$$

$$\lambda_i^t = \left(-\frac{1}{n-1}\right)^t \qquad \text{if } t = \log_{n-1}(1/\epsilon)$$

$$\lambda_i^t = \left(-\frac{1}{n-1}\right)^t$$

$$\boxed{t = \log_{n-1}\left(\frac{n}{\epsilon}\right)}$$ then

$$|\lambda_i^t| = \frac{\epsilon}{n}$$

$$P^t = \underbrace{v_1 v_1^T}_{\text{stationary distribution}} + \frac{\epsilon}{n} \cdot (-1)^t \underbrace{\sum_{i=2}^{n} v_i v_i^T}_{\text{everything } \perp \text{ to stationary}}$$

$\vec{q} \in \mathbb{R}^n$ initial probability distribution

$$\vec{q} = \alpha_1 v_1 + \alpha_2 v_2 + \dots + \alpha_n v_n$$

$$\boxed{P^t \vec{q} = \lambda_1^t \alpha_1 v_1 + \lambda_2^t \alpha_2 v_2 + \dots + \lambda_n^t \alpha_n v_n}$$

$$\|P^{(t)} q - v_1\|_1 = \underbrace{\left|\left(\lambda_1^t \alpha_1 - 1\right)\right|}_{\leq \epsilon} + \underbrace{\sum_{j=2}^{n} \underbrace{|\lambda_j^t|}_{\leq \frac{\epsilon}{n}}}_{\leq \epsilon}$$

$$\sum_{i=1}^{n} \lambda_i^t \alpha_i = 1$$

Template:

- find stationary distribution
- get eigenvalues of the transition matrix of Markov chain

  $\lambda_1 = 1 \quad \lambda_2 = \dots = \lambda_n = -1/n-1$
- need to compute t such that

  $$|\lambda_j^t| < \frac{\epsilon}{n} \text{ for all } j > 1 .$$
- t is your mixing time

- Introduction
  - Why Random Walks & Markov Chains?
  - Basics on Theory of (finite) Markov Chains
- Main Topics
  - Fundamental Theorem of Markov Chains
  - Page Rank
- Conclusion
- Acknowledgements

Fundamental Theorem of Markov Chains

- The *return time* from state i to itself is defined as

$$H_{i,i} := \underbrace{\min\{t \geq 1 \mid \underbrace{X_t = i}_{\text{at } i \text{ in time } t}, \underbrace{X_0 = i}_{\text{started at } i}\}}_{\text{first time}}$$

Fundamental Theorem of Markov Chains

- The *return time* from state i to itself is defined as

$$H_{i,i} := \min\{t \geq 1 \mid X_t = i, X_0 = i\}$$

- *Expected return time:* defined as $h_{i,i} := \mathbb{E}[H_{i,i}]$.

Fundamental Theorem of Markov Chains

- The *return time* from state i to itself is defined as

$$H_{i,i} := \min\{t \geq 1 \mid X_t = i, X_0 = i\}$$

- *Expected return time:* defined as $h_{i,i} := \mathbb{E}[H_{i,i}]$.

Theorem (Fundamental Theorem of Markov Chains)

Any finite, irreducible and aperiodic Markov Chain has the following properties:

1. *There exists a unique stationary distribution* π*, where* $\boxed{\pi_i > 0}$ *for all* $i \in [n]$
2. *The sequence of distributions* $\{p_t\}_{t \geq 0}$ *will converge to* π*, no matter what the initial distribution is*
3. $$\boxed{\pi_i = \lim_{t \to \infty} P^t_{i,i} = \frac{1}{h_{i,i}}}$$

Fundamental Theorem of Markov Chains

Theorem (Fundamental Theorem of Markov Chains)

Any finite, irreducible and aperiodic Markov Chain has the following properties:

1. *There is unique stationary distribution* π*, where* $\pi_i > 0$ *for all* $i \in [n]$
2. *For every distribution* $p_0 \in \mathbb{R}^n_{\geq 0}$, $\lim_{t \to \infty} \underbrace{p_0 \cdot P^t}_{p_t} = \pi$
3. $$\pi_i = \lim_{t \to \infty} P^t_{i,i} = \frac{1}{h_{i,i}}$$

Intuition for proof of this theorem:

Fundamental Theorem of Markov Chains

Theorem (Fundamental Theorem of Markov Chains)

Any finite, irreducible and aperiodic Markov Chain has the following properties:

1. *There is unique stationary distribution π, where $\pi_i > 0$ for all $i \in [n]$*
2. *For every distribution $p_0 \in \mathbb{R}^n_{\geq 0}$,* $\lim_{t \to \infty} p_0 \cdot P^t = \pi$
3. $$\pi_i = \lim_{t \to \infty} P^t_{i,i} = \frac{1}{h_{i,i}}$$

Intuition for proof of this theorem:

- two random walks are "indistinguishable" after they "meet" at the *same vertex v* at a particular *time t*

Fundamental Theorem of Markov Chains

Theorem (Fundamental Theorem of Markov Chains)

Any finite, irreducible and aperiodic Markov Chain has the following properties:

1. *There is unique stationary distribution* π, *where* $\pi_i > 0$ *for all* $i \in [n]$
2. *For every distribution* $p_0 \in \mathbb{R}^n_{\geq 0}$, $\lim_{t \to \infty} p_0 \cdot P^t = \pi$
3. $$\pi_i = \lim_{t \to \infty} P^t_{i,i} = \frac{1}{h_{i,i}}$$

Intuition for proof of this theorem:

- two random walks are "indistinguishable" after they "meet" at the *same vertex v* at a particular *time t*
- By finiteness, irreducibility and aperiodicity, two walks will meet with positive probability (and thus by Markov property) become *same distribution*

Fundamental Theorem of Markov Chains

Theorem (Fundamental Theorem of Markov Chains)

Any finite, irreducible and aperiodic Markov Chain has the following properties:

1. *There is unique stationary distribution* π, *where* $\pi_i > 0$ *for all* $i \in [n]$
2. *For every distribution* $p_0 \in \mathbb{R}^n_{\geq 0}$, $\lim_{t \to \infty} p_0 \cdot P^t = \pi$
3. $$\pi_i = \lim_{t \to \infty} P^t_{i,i} = \frac{1}{h_{i,i}}$$

If our underlying graph is undirected:

Fundamental Theorem of Markov Chains

Theorem (Fundamental Theorem of Markov Chains)

Any finite, irreducible and aperiodic Markov Chain has the following properties:

1. *There is unique stationary distribution π, where $\pi_i > 0$ for all $i \in [n]$*
2. *For every distribution $p_0 \in \mathbb{R}^n_{\geq 0}$,* $\lim_{t\to\infty} p_0 \cdot P^t = \pi$
3. $$\pi_i = \lim_{t\to\infty} P^t_{i,i} = \frac{1}{h_{i,i}}$$

If our underlying graph is undirected:

- If A_G adjacency matrix of $G(V,E)$ and $D = diag(d_1, d_2, \ldots, d_n)$, transition matrix:

$$\boxed{P = D^{-1} \cdot A_G}$$

$$D = \begin{pmatrix} d_1 & & & \\ & d_2 & & \\ & & \ddots & \\ & & & d_n \end{pmatrix}$$

Fundamental Theorem of Markov Chains

Theorem (Fundamental Theorem of Markov Chains)

Any finite, irreducible and aperiodic Markov Chain has the following properties:

1. *There is unique stationary distribution* π, *where* $\pi_i > 0$ *for all* $i \in [n]$
2. *For every distribution* $p_0 \in \mathbb{R}^n_{\geq 0}$, $\lim_{t \to \infty} p_0 \cdot P^t = \pi$
3. $$\pi_i = \lim_{t \to \infty} P^t_{i,i} = \frac{1}{h_{i,i}}$$

If our underlying graph is undirected:

- If A_G adjacency matrix of $G(V, E)$ and $D = diag(d_1, d_2, \ldots, d_n)$, transition matrix:
$$P = D^{-1} \cdot A_G$$
- Note that in this case, easy to guess stationary distribution: $\boxed{\pi \cdot P = \pi}$

Practice problem: prove this!

$$\pi_i = \frac{d_i}{2m}, \quad m = |E|$$

Fundamental Theorem of Markov Chains

If our underlying graph is undirected:

Fundamental Theorem of Markov Chains

If our underlying graph is undirected:

- In this case, easy to guess stationary distribution:

$$\pi_i = \frac{d_i}{2m}, \quad m = |E|$$

Fundamental Theorem of Markov Chains

If our underlying graph is undirected:

- In this case, easy to guess stationary distribution:

$$\pi_i = \frac{d_i}{2m}, \quad m = |E|$$

- If A_G adjacency matrix of $G(V, E)$ and $D = diag(d_1, d_2, \ldots, d_n)$, transition matrix:

$$P = D^{-1} \cdot A_G$$

Fundamental Theorem of Markov Chains

If our underlying graph is undirected:

- In this case, easy to guess stationary distribution:

$$\pi_i = \frac{d_i}{2m}, \quad m = |E|$$

- If A_G adjacency matrix of $G(V, E)$ and $D = diag(d_1, d_2, \ldots, d_n)$, transition matrix:

$$P = D^{-1} \cdot A_G$$

  - P not symmetric, but *similar* to a symmetric matrix:

$$D^{1/2} P D^{-1/2} = D^{1/2} D^{-1} A_G D^{-1/2} = \underbrace{D^{-1/2} A_G D^{-1/2}}_{\text{symmetric}} = P'$$

Fundamental Theorem of Markov Chains

If our underlying graph is undirected:

- If A_G adjacency matrix of $G(V, E)$ and $D = diag(d_1, d_2, \ldots, d_n)$, transition matrix:

$$P = D^{-1} \cdot A_G$$

  - P not symmetric, but *similar* to a symmetric matrix:

$$D^{1/2} P D^{-1/2} = D^{1/2} D^{-1} A_G D^{-1/2} = D^{-1/2} A_G D^{-1/2} = P'$$

  - P and P' has $\underbrace{\textit{same eigenvalues}}_{\text{similarity (conjugation)}}$! And P' has only $\underbrace{\textit{real eigenvalues}}_{\text{symmetric matrices} \Rightarrow \text{real eigenvalues (spectral thm)}}$!

Fundamental Theorem of Markov Chains

If our underlying graph is undirected:

- If A_G adjacency matrix of $G(V, E)$ and $D = diag(d_1, d_2, \dots, d_n)$, transition matrix:

$$P = D^{-1} \cdot A_G$$

  - P not symmetric, but *similar* to a symmetric matrix:

$$D^{1/2} P D^{-1/2} = D^{1/2} D^{-1} A_G D^{-1/2} = D^{-1/2} A_G D^{-1/2} = P'$$

  - P and P' has *same eigenvalues*! And P' has only *real eigenvalues*!
  - Eigenvectors of P are $D^{-1/2} v_i$ where v_i are eigenvectors of P'. And v_i can be taken to form *orthonormal basis*.

Fundamental Theorem of Markov Chains

- Stationary distribution: $\pi_i = \dfrac{d_i}{2m}, \quad m = |E|$
- Transition matrix: $P = D^{-1} \cdot A_G$
  - P *similar* to a symmetric matrix: $P' = D^{-1/2} A_G D^{-1/2}$
  - P and P' has *same eigenvalues*! And P' has only *real eigenvalues*!
  - Eigenvectors of P are $D^{-1/2} v_i$ where v_i are eigenvectors of P'. And v_i can be taken to form *orthonormal basis*.

Fundamental Theorem of Markov Chains

- Stationary distribution: $\pi_i = \dfrac{d_i}{2m}, \quad m = |E|$
- Transition matrix: $P = D^{-1} \cdot A_G$
  - P *similar* to a symmetric matrix: $P' = D^{-1/2} A_G D^{-1/2}$
  - P and P' has *same eigenvalues*! And P' has only *real eigenvalues*!
  - Eigenvectors of P are $D^{-1/2} v_i$ where v_i are eigenvectors of P'. And v_i can be taken to form *orthonormal basis*.
  - Graph strongly connected $\Rightarrow$ *Perron-Frobenius* for irreducible non-negative matrices

Fundamental Theorem of Markov Chains

- Stationary distribution: $\pi_i = \dfrac{d_i}{2m}, \quad m = |E|$
- Transition matrix: $P = D^{-1} \cdot A_G$
  - P *similar* to a symmetric matrix: $P' = D^{-1/2} A_G D^{-1/2}$
  - P and P' has *same eigenvalues*! And P' has only *real eigenvalues*!
  - Eigenvectors of P are $D^{-1/2} v_i$ where v_i are eigenvectors of P'. And v_i can be taken to form *orthonormal basis*.
  - Graph strongly connected $\Rightarrow$ *Perron-Frobenius* for irreducible non-negative matrices
    - *unique* eigenvector whose eigenvalue has *max absolute value*
    - eigenvector has *all positive coordinates*
    - eigenvalue is *positive*

$\lambda_1, \ldots, \lambda_n$

$|\lambda_1| > |\lambda_j|$

$\lambda_1 > 0$

Fundamental Theorem of Markov Chains

- Stationary distribution: $\boxed{\pi_i = \frac{d_i}{2m},}$ $m = |E|$
- Transition matrix: $P = D^{-1} \cdot A_G$
  - P *similar* to a symmetric matrix: $P' = D^{-1/2} A_G D^{-1/2}$
  - P and P' has *same eigenvalues*! And P' has only *real eigenvalues*!
  - Eigenvectors of P are $D^{-1/2} v_i$ where v_i are eigenvectors of P'. And v_i can be taken to form *orthonormal basis*.
  - Graph strongly connected $\Rightarrow$ *Perron-Frobenius* for irreducible non-negative matrices
    - *unique* eigenvector whose eigenvalue has *max absolute value*
    - eigenvector has *all positive coordinates*
    - eigenvalue is *positive*
  - This eigenvector is π!

π has all positive coordinates!

Fundamental Theorem of Markov Chains

- Stationary distribution: $\pi_i = \dfrac{d_i}{2m}, \quad m = |E|$
- Transition matrix: $P = D^{-1} \cdot A_G$
  - P *similar* to a symmetric matrix: $P' = D^{-1/2} A_G D^{-1/2}$
  - P and P' has *same eigenvalues*! And P' has only *real eigenvalues*!
  - Eigenvectors of P are $D^{-1/2} v_i$ where v_i are eigenvectors of P'. And v_i can be taken to form *orthonormal basis*.
  - Graph strongly connected $\Rightarrow$ *Perron-Frobenius* for irreducible non-negative matrices
    - *unique* eigenvector whose eigenvalue has *max absolute value*
    - eigenvector has *all positive coordinates*
    - eigenvalue is *positive*
  - This eigenvector is π!
  - All random walks converge to π, as we wanted to show.

$\pi P = \pi$ max eigenvalue = all other eigenvalues are <1

$$q P^t = q \cdot \left(\sum_{i=1}^{n} \lambda_i^t \cdot v_i v_i^T \right) \xrightarrow{t \to \infty} \pi \quad \xrightarrow{t\to\infty} 0.$$

Page Rank

- **Setting:** we have a directed graph describing relationships between set of webpages.

Page Rank

- **Setting:** we have a directed graph describing relationships between set of webpages.

  There is a directed edge (i, j) if there is a link from page i to page j.

Page Rank

- **Setting:** we have a directed graph describing relationships between set of webpages.

  There is a directed edge (i, j) if there is a link from page i to page j.
- **Goal:** want algorithm to “rank” how important a page is.

Page Rank

- **Setting:** we have a directed graph describing relationships between set of webpages.

  There is a directed edge (i, j) if there is a link from page i to page j.
- **Goal:** want algorithm to "rank" how important a page is.
- *Intuition:* if many other pages link to a particular page, then the linked page must be important!

Page Rank

- **Setting:** we have a directed graph describing relationships between set of webpages.

  There is a directed edge (i, j) if there is a link from page i to page j.
- **Goal:** want algorithm to "rank" how important a page is.
- *Intuition:* if many other pages link to a particular page, then the linked page must be important!

Algorithm (Page Rank Algorithm)

1. *Initially, each page has pagerank value* $\frac{1}{n}$

Page Rank

- **Setting:** we have a directed graph describing relationships between set of webpages.

  There is a directed edge (i, j) if there is a link from page i to page j.
- **Goal:** want algorithm to "rank" how important a page is.
- *Intuition:* if many other pages link to a particular page, then the linked page must be important!

Algorithm (Page Rank Algorithm)

1. *Initially, each page has pagerank value* $\frac{1}{n}$
2. *In each step, each page:*

Page Rank

- **Setting:** we have a directed graph describing relationships between set of webpages.

  There is a directed edge (i, j) if there is a link from page i to page j.
- **Goal:** want algorithm to "rank" how important a page is.
- *Intuition:* if many other pages link to a particular page, then the linked page must be important!

Algorithm (Page Rank Algorithm)

1. *Initially, each page has pagerank value* $\frac{1}{n}$
2. *In each step, each page:*
   1. *divides its pagerank value equally to its outgoing link,*

Page Rank

- **Setting:** we have a directed graph describing relationships between set of webpages.

  There is a directed edge (i, j) if there is a link from page i to page j.
- **Goal:** want algorithm to "rank" how important a page is.
- *Intuition:* if many other pages link to a particular page, then the linked page must be important!

Algorithm (Page Rank Algorithm)

1. *Initially, each page has pagerank value* $\frac{1}{n}$
2. *In each step, each page:*
   1. *divides its pagerank value equally to its outgoing link,*
   2. *sends these equal shares to the pages it points to,*

Page Rank

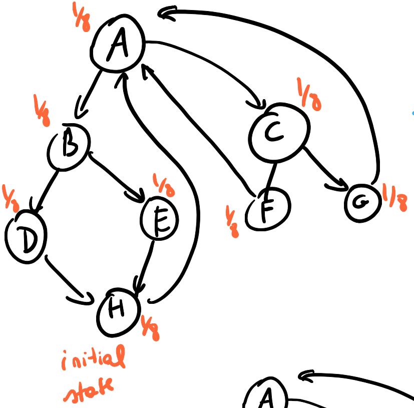
- **Setting:** we have a directed graph describing relationships between set of webpages.

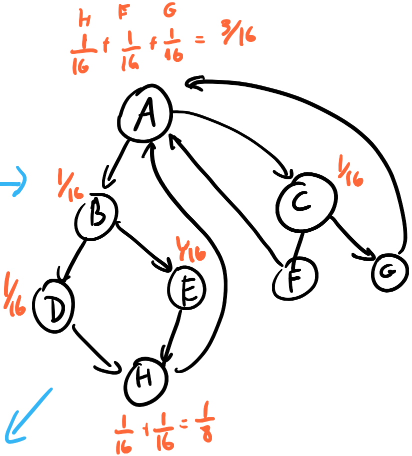
  There is a directed edge (i, j) if there is a link from page i to page j.
- **Goal:** want algorithm to "rank" how important a page is.
- *Intuition:* if many other pages link to a particular page, then the linked page must be important!

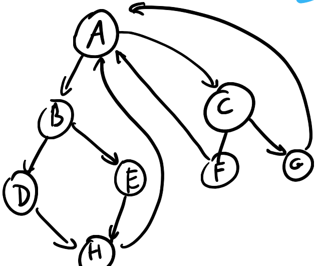
Algorithm (Page Rank Algorithm)

1. *Initially, each page has pagerank value* $\frac{1}{n}$
2. *In each step, each page:*
   1. *divides its pagerank value equally to its outgoing link,*
   2. *sends these equal shares to the pages it points to,*
   3. *updates its new pagerank value to be the sum of shares it receives.*

Page Rank - Example







Page Rank

Algorithm (Page Rank Algorithm)

1. *Initially, each page has pagerank value* $\frac{1}{n}$
2. *In each step, each page:*
   1. *divides its pagerank value equally to its outgoing link,*
   2. *sends these equal shares to the pages it points to,*
   3. *updates its new pagerank value to be the sum of shares it receives.*

Page Rank

Algorithm (Page Rank Algorithm)

1. *Initially, each page has pagerank value $\frac{1}{n}$*
2. *In each step, each page:*
   1. *divides its pagerank value equally to its outgoing link,*
   2. *sends these equal shares to the pages it points to,*
   3. *updates its new pagerank value to be the sum of shares it receives.*

- *Equilibrium* of pagerank values equal to probabilities of *stationary distribution* of random walk

$$P \in \mathbb{R}^{n \times n}, \quad P_{i,j} = \frac{1}{\delta^{out}(i)}$$

Page Rank

Algorithm (Page Rank Algorithm)

1. *Initially, each page has pagerank value* $\frac{1}{n}$
2. *In each step, each page:*
   1. *divides its pagerank value equally to its outgoing link,*
   2. *sends these equal shares to the pages it points to,*
   3. *updates its new pagerank value to be the sum of shares it receives.*

- *Equilibrium* of pagerank values equal to probabilities of *stationary distribution* of random walk

$$P \in \mathbb{R}^{n \times n}, \quad P_{i,j} = \frac{1}{\delta^{out}(i)}$$

- Pagerank values and transition probabilities satisfy same equations:

Practice problem! check this!

$$p_{t+1}(j) = \sum_{\underbrace{i:(i,j)\in E}_{\text{every vertex } i \text{ s.t. there is edge } (i,j)}} \frac{p_t(i)}{\delta^{out}(i)} \quad \Rightarrow \quad p_{t+1} = p_t \cdot P$$

Page Rank

Algorithm (Page Rank Algorithm)

1. *Initially, each page has pagerank value* $\frac{1}{n}$
2. *In each step, each page:*
   1. *divides its pagerank value equally to its outgoing link,*
   2. *sends these equal shares to the pages it points to,*
   3. *updates its new pagerank value to be the sum of shares it receives.*

- *Equilibrium* of pagerank values equal to probabilities of *stationary distribution* of random walk

$$P \in \mathbb{R}^{n \times n}, \quad P_{i,j} = \frac{1}{\delta^{out}(i)}$$

- Pagerank values and transition probabilities satisfy same equations:

$$p_{t+1}(j) = \sum_{i:(i,j)\in E} \frac{p_t(i)}{\delta^{out}(i)} \quad \Rightarrow \quad p_{t+1} = p_t \cdot P$$

- If graph finite, irreducible and aperiodic, fundamental theorem guarantees stationary distribnution.

Page Rank

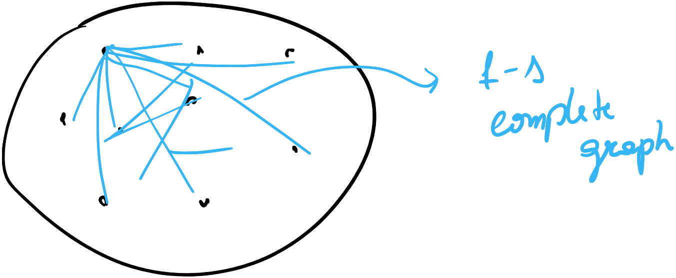
- In practice, directed graph may not satisfy fundamental theorem's conditions

Page Rank

- In practice, directed graph may not satisfy fundamental theorem's conditions
- Modify original graph as follows:

Page Rank

- In practice, directed graph may not satisfy fundamental theorem's conditions
- Modify original graph as follows:
  - Fix number $0 < s < 1$
  - Divide s fraction of its pagerank value to its neighbors,
  - $1 - s$ fraction of its pagerank value to all nodes evenly



Page Rank

- In practice, directed graph may not satisfy fundamental theorem's conditions
- Modify original graph as follows:
  - Fix number $0 < s < 1$
  - Divide s fraction of its pagerank value to its neighbors,
  - $1 - s$ fraction of its pagerank value to all nodes evenly
- Equivalent to the random walk:

Page Rank

- In practice, directed graph may not satisfy fundamental theorem's conditions
- Modify original graph as follows:
  - Fix number $0 < s < 1$
  - Divide s fraction of its pagerank value to its neighbors,
  - $1 - s$ fraction of its pagerank value to all nodes evenly
- Equivalent to the random walk:
  - With probability s go to one of its neighbors (uniformly at random),
  - With probability $1 - s$ go to random page (uniformly at random)

complete graph

Page Rank

- In practice, directed graph may not satisfy fundamental theorem's conditions
- Modify original graph as follows:
  - Fix number $0 < s < 1$
  - Divide s fraction of its pagerank value to its neighbors,
  - $1 - s$ fraction of its pagerank value to all nodes evenly
- Equivalent to the random walk:
  - With probability s go to one of its neighbors (uniformly at random),
  - With probability $1 - s$ go to random page (uniformly at random)
- Now resulting graph is *strongly connected* and *aperiodic* $\Rightarrow$ unique stationary distribution

Page Rank

- In practice, directed graph may not satisfy fundamental theorem's conditions
- Modify original graph as follows:
  - Fix number $0 < s < 1$
  - Divide s fraction of its pagerank value to its neighbors,
  - $1 - s$ fraction of its pagerank value to all nodes evenly
- Equivalent to the random walk:
  - With probability s go to one of its neighbors (uniformly at random),
  - With probability $1 - s$ go to random page (uniformly at random)
- Now resulting graph is *strongly connected* and *aperiodic* $\Rightarrow$ unique stationary distribution
- This modification does not change "relative importance" of vertices

Conclusion

Markov Chains and Random Walks are ubiquitous in randomized algorithms.

- Page Rank algorithm (today's lecture)
- Approximation algorithms for counting problems [Karp, Luby & Madras]
  - Permanent of non-negative matrices [Jerrum, Vigoda & Sinclair]
- Sampling Problems
  - Gibbs sampling in statistical physics
  - many more places
- Probability amplification without too much randomness (efficient)
  - Random walks on expander graphs
- many more

Acknowledgement

- Lecture based largely on:
  - Lap Chi's notes
  - [Motwani & Raghavan 2007, Chapter 6]
- See Lap Chi's notes at `https://cs.uwaterloo.ca/~lapchi/cs466/notes/L11.pdf`
- Also see Lap Chi's notes `https://cs.uwaterloo.ca/~lapchi/cs466/notes/L14.pdf` for a proof of fundamental theorem of Markov chains for undirected graphs.

References I

Motwani, Rajeev and Raghavan, Prabhakar (2007)
Randomized Algorithms

Karp, R.M. and Luby, M. and Madras, N. (1989)
Monte-Carlo approximation algorithms for enumeration problems.
Journal of algorithms, 10(3), pp.429-448.

Jerrum, M. and Sinclair, A. and Vigoda, E. (2004)
A polynomial-time approximation algorithm for the permanent of a matrix with nonnegative entries.
Journal of the ACM (JACM), 51(4), pp.671-697.