Algorithm Foundations of Data Science and Engineering

Lecture 7: Random Walk

MING GAO

DaSE @ ECNU
(for course related communications)
mgao@dase.ecnu.edu.cn

Apr. 8, 2019

Outline

1. Motivation
2. Markov Chain and Random Walk
3. Page Rank
   - Problem Formulation
   - PageRank Algorithm
   - Improvements of PageRank Algorithm

Joint Probability

The r.v.s $X_1, \cdots, X_n$ are a sequence of discrete r.v.s, then joint pmf $P(X_1 = x_1, \cdots, X_n = x_n)$ can be computed as

- If $X_1, \cdots, X_n$ are mutually independent r.v.s, $\prod_{i=1}^{n} P(X_i = x_i)$;
- $P(X_1 = x_1) \prod_{i=2}^{n} P(X_i = x_i | X_{i-1} = x_{i-1}, \cdots, X_1 = x_1)$.

Joint Probability

The r.v.s $X_1, \cdots, X_n$ are a sequence of discrete r.v.s, then joint pmf $P(X_1 = x_1, \cdots, X_n = x_n)$ can be computed as

- If $X_1, \cdots, X_n$ are mutually independent r.v.s, $\prod_{i=1}^{n} P(X_i = x_i)$;
- $P(X_1 = x_1) \prod_{i=2}^{n} P(X_i = x_i | X_{i-1} = x_{i-1}, \cdots, X_1 = x_1)$.

Remarks:

- When r.v.s are dependent, the generative model is hard to scalable to large scale applications;

Joint Probability

The r.v.s $X_1, \cdots, X_n$ are a sequence of discrete r.v.s, then joint pmf $P(X_1 = x_1, \cdots, X_n = x_n)$ can be computed as

- If $X_1, \cdots, X_n$ are mutually independent r.v.s, $\prod_{i=1}^n P(X_i = x_i)$;
- $P(X_1 = x_1) \prod_{i=2}^n P(X_i = x_i | X_{i-1} = x_{i-1}, \cdots, X_1 = x_1)$.

Remarks:

- When r.v.s are dependent, the generative model is hard to scalable to large scale applications;
- We need to simplify the model in practise.

Joint Probability

The r.v.s $X_1, \cdots, X_n$ are a sequence of discrete r.v.s, then joint pmf $P(X_1 = x_1, \cdots, X_n = x_n)$ can be computed as

- If $X_1, \cdots, X_n$ are mutually independent r.v.s, $\prod_{i=1}^{n} P(X_i = x_i)$;
- $P(X_1 = x_1) \prod_{i=2}^{n} P(X_i = x_i | X_{i-1} = x_{i-1}, \cdots, X_1 = x_1)$.

Remarks:

- When r.v.s are dependent, the generative model is hard to scalable to large scale applications;
- We need to simplify the model in practise.
  - The first-order correlation in language model can be defined as:

$$P(X_i = x_i | X_{i-1} = x_{i-1}, \cdots, X_1 = x_1) = P(X_i = x_i | X_{i-1} = x_{i-1}).$$

Joint Probability

The r.v.s $X_1, \cdots, X_n$ are a sequence of discrete r.v.s, then joint pmf $P(X_1 = x_1, \cdots, X_n = x_n)$ can be computed as

- If $X_1, \cdots, X_n$ are mutually independent r.v.s, $\prod_{i=1}^{n} P(X_i = x_i)$;
- $P(X_1 = x_1) \prod_{i=2}^{n} P(X_i = x_i | X_{i-1} = x_{i-1}, \cdots, X_1 = x_1)$.

Remarks:

- When r.v.s are dependent, the generative model is hard to scalable to large scale applications;
- We need to simplify the model in practise.
  - The first-order correlation in language model can be defined as:

    $$P(X_i = x_i | X_{i-1} = x_{i-1}, \cdots, X_1 = x_1) = P(X_i = x_i | X_{i-1} = x_{i-1}).$$

  - For financial applications, the stock price can be modeled by $t-$order correlation $P(X_i = x_i | X_{i-1} = x_{i-1}, \cdots, X_1 = x_1) = P(X_i = x_i | X_{i-1} = x_{i-1}, \cdots, X_{i-t} = x_{i-t})$.

Gambling

Suppose that you start with \$10, and you wager \$1 on an unending, fair, coin toss indefinitely, or until you lose all of your money.

Gambling

Suppose that you start with \$10, and you wager \$1 on an unending, fair, coin toss indefinitely, or until you lose all of your money.
If X_n represents the number of dollars you have after n tosses, with $X_0 = 10$, then the sequence $\{X_n : n \in \mathbb{N}\}$ is a r.v. sequence.

Gambling

Suppose that you start with \$10, and you wager \$1 on an unending, fair, coin toss indefinitely, or until you lose all of your money.
If X_n represents the number of dollars you have after n tosses, with $X_0 = 10$, then the sequence $\{X_n : n \in \mathbb{N}\}$ is a r.v. sequence.

- If I know that you have \$12 now, then it would be expected that with even odds, you will either have \$11 or \$13 after the next toss.

Gambling

Suppose that you start with \$10, and you wager \$1 on an unending, fair, coin toss indefinitely, or until you lose all of your money.

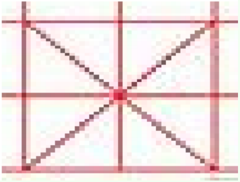
If X_n represents the number of dollars you have after n tosses, with $X_0 = 10$, then the sequence $\{X_n : n \in \mathbb{N}\}$ is a r.v. sequence.

- If I know that you have \$12 now, then it would be expected that with even odds, you will either have \$11 or \$13 after the next toss.
- This guess is not improved by the added knowledge that you started with \$10, then went up to \$11, down to \$10, up to \$11, and then to \$12.

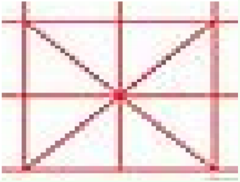
In this example, your money satisfies the first-order correlation, i.e.,

$$P(X_n = x_n | X_{n-1} = x_{n-1}, \cdots, X_1 = x_1) = P(X_n = x_n | X_{n-1} = x_{n-1}).$$

Chess

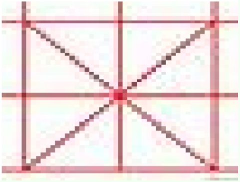


Chess



If X_n represents the location of the king in the $n-$th step, then the sequence $\{X_n : n \in \mathbb{N}\}$ is a r.v. sequence.

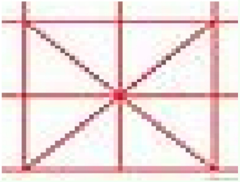
Chess



If X_n represents the location of the king in the $n-$th step, then the sequence $\{X_n : n \in \mathbb{N}\}$ is a r.v. sequence.

- A walk of king in the graph is a sequence of vertices $v_1, v_2, \cdots, v_n, \cdots$, not necessarily distinct, such that (v_i, v_{i+1}) is an edge in the graph.

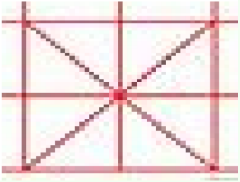
Chess



If X_n represents the location of the king in the $n-$th step, then the sequence $\{X_n : n \in \mathbb{N}\}$ is a r.v. sequence.

- A walk of king in the graph is a sequence of vertices $v_1, v_2, \cdots, v_n, \cdots$, not necessarily distinct, such that (v_i, v_{i+1}) is an edge in the graph.
- The length of a walk is number of edges in the path, equivalently it is equal to n.

Chess



If X_n represents the location of the king in the $n-$th step, then the sequence $\{X_n : n \in \mathbb{N}\}$ is a r.v. sequence.

- A walk of king in the graph is a sequence of vertices $v_1, v_2, \cdots, v_n, \cdots$, not necessarily distinct, such that (v_i, v_{i+1}) is an edge in the graph.
- The length of a walk is number of edges in the path, equivalently it is equal to n.

In this example, the location of king also satisfies the first-order correlation, i.e.,

$$P(X_n = v_n | X_{n-1} = v_{n-1}, \cdots, X_1 = v_1) = P(X_n = v_n | X_{n-1} = v_{n-1}).$$

Random walk

If X_n represents the $n-$th Web page your visited, then the sequence $\{X_n : n \in \mathbb{N}\}$ is a r.v. sequence.

Random walk

If X_n represents the $n-$th Web page your visited, then the sequence $\{X_n : n \in \mathbb{N}\}$ is a r.v. sequence.

- Currently, you locate in the page w_i. Next step, you will visit the page w_{i+1}. A walk of you in the Web graph is a sequence of pages $w_1, w_2, \cdots, w_n, \cdots$, such that (w_i, w_{i+1}) is an edge in the graph.

Random walk

If X_n represents the $n-$th Web page your visited, then the sequence $\{X_n : n \in \mathbb{N}\}$ is a r.v. sequence.

- Currently, you locate in the page w_i. Next step, you will visit the page w_{i+1}. A walk of you in the Web graph is a sequence of pages $w_1, w_2, \cdots, w_n, \cdots$, such that (w_i, w_{i+1}) is an edge in the graph.
- The length of a walk is number of edges in the path, equivalently it is equal to n.

Random walk

If X_n represents the $n-$th Web page your visited, then the sequence $\{X_n : n \in \mathbb{N}\}$ is a r.v. sequence.

- Currently, you locate in the page w_i. Next step, you will visit the page w_{i+1}. A walk of you in the Web graph is a sequence of pages $w_1, w_2, \cdots, w_n, \cdots$, such that (w_i, w_{i+1}) is an edge in the graph.
- The length of a walk is number of edges in the path, equivalently it is equal to n.

In this example, the visited pages also satisfies the first-order correlation, i.e.,

$$P(X_n = w_n | X_{n-1} = w_{n-1}, \cdots, X_1 = w_1) = P(X_n = w_n | X_{n-1} = w_{n-1}).$$

Random process

A random process is a collection of r.v.s indexed by some set I, taking values in a set S.

Random process

A random process is a collection of r.v.s indexed by some set I, taking values in a set S.

- I is the index set, usually time, e.g., $\mathcal{Z}^+$, $\mathcal{R}$, and $\mathcal{R}^+$, etc.

Random process

A random process is a collection of r.v.s indexed by some set I, taking values in a set S.

- I is the index set, usually time, e.g., $\mathcal{Z}^+$, $\mathcal{R}$, and $\mathcal{R}^+$, etc.
- S is the state space, e.g., $\mathcal{Z}^+$, $\mathcal{R}^n$, and $\{a, b, c\}$, etc.

Random process

A random process is a collection of r.v.s indexed by some set I, taking values in a set S.

- I is the index set, usually time, e.g., $\mathcal{Z}^+$, $\mathcal{R}$, and $\mathcal{R}^+$, etc.
- S is the state space, e.g., $\mathcal{Z}^+$, $\mathcal{R}^n$, and $\{a, b, c\}$, etc.

Random process

A random process is a collection of r.v.s indexed by some set I, taking values in a set S.

- I is the index set, usually time, e.g., $\mathcal{Z}^+$, $\mathcal{R}$, and $\mathcal{R}^+$, etc.
- S is the state space, e.g., $\mathcal{Z}^+$, $\mathcal{R}^n$, and $\{a, b, c\}$, etc.

We classify random processes according to

- Index set can be discrete or continuous;

Random process

A random process is a collection of r.v.s indexed by some set I, taking values in a set S.

- I is the index set, usually time, e.g., $\mathcal{Z}^+$, $\mathcal{R}$, and $\mathcal{R}^+$, etc.
- S is the state space, e.g., $\mathcal{Z}^+$, $\mathcal{R}^n$, and $\{a, b, c\}$, etc.

We classify random processes according to

- Index set can be discrete or continuous;
- State space can be finite, countable or uncountable (continuous).

Random process

A random process is a collection of r.v.s indexed by some set I, taking values in a set S.

- I is the index set, usually time, e.g., $\mathcal{Z}^+$, $\mathcal{R}$, and $\mathcal{R}^+$, etc.
- S is the state space, e.g., $\mathcal{Z}^+$, $\mathcal{R}^n$, and $\{a, b, c\}$, etc.

We classify random processes according to

- Index set can be discrete or continuous;
- State space can be finite, countable or uncountable (continuous).

Examples

- Gambling;

Random process

A random process is a collection of r.v.s indexed by some set I, taking values in a set S.

- I is the index set, usually time, e.g., $\mathcal{Z}^+$, $\mathcal{R}$, and $\mathcal{R}^+$, etc.
- S is the state space, e.g., $\mathcal{Z}^+$, $\mathcal{R}^n$, and $\{a, b, c\}$, etc.

We classify random processes according to

- Index set can be discrete or continuous;
- State space can be finite, countable or uncountable (continuous).

Examples

- Gambling;
- Chess;

Random process

A random process is a collection of r.v.s indexed by some set I, taking values in a set S.

- I is the index set, usually time, e.g., $\mathcal{Z}^+$, $\mathcal{R}$, and $\mathcal{R}^+$, etc.
- S is the state space, e.g., $\mathcal{Z}^+$, $\mathcal{R}^n$, and $\{a, b, c\}$, etc.

We classify random processes according to

- Index set can be discrete or continuous;
- State space can be finite, countable or uncountable (continuous).

Examples

- Gambling;
- Chess;
- Random walk;

Random process

A random process is a collection of r.v.s indexed by some set I, taking values in a set S.

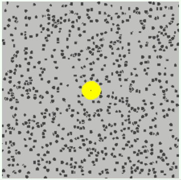
- I is the index set, usually time, e.g., $\mathcal{Z}^+$, $\mathcal{R}$, and $\mathcal{R}^+$, etc.
- S is the state space, e.g., $\mathcal{Z}^+$, $\mathcal{R}^n$, and $\{a, b, c\}$, etc.

We classify random processes according to

- Index set can be discrete or continuous;
- State space can be finite, countable or uncountable (continuous).

Examples

- Gambling;
- Chess;
- Random walk;
- Brown motion.



Markov property

More formally, $X(t)$ is Markovian if the following property holds:

Markov property

More formally, $X(t)$ is Markovian if the following property holds:

$$\begin{aligned} P(X(t_{n+1}) = j_{n+1} | X(t_n) = j_n, \cdots, X(t_0) = j_0) \\ = P(X(t_{n+1}) = j_{n+1} | X(t_n) = j_n) \end{aligned}$$

for all finite sequence of times $t_0 < \cdots < t_n < t_{n+1} \in I$ and of states $j_0, \cdots, j_n < j_{n+1} \in S$.

Markov property

More formally, $X(t)$ is Markovian if the following property holds:

$$\begin{aligned} P(X(t_{n+1}) = j_{n+1} | X(t_n) = j_n, \cdots, X(t_0) = j_0) \\ = P(X(t_{n+1}) = j_{n+1} | X(t_n) = j_n) \end{aligned}$$

for all finite sequence of times $t_0 < \cdots < t_n < t_{n+1} \in I$ and of states $j_0, \cdots, j_n < j_{n+1} \in S$.

- A random process is called a **Markov Process** if conditional on the current state, its future is independent of its past.

Markov property

More formally, $X(t)$ is Markovian if the following property holds:

$$\begin{aligned} P(X(t_{n+1}) &= j_{n+1} | X(t_n) = j_n, \cdots, X(t_0) = j_0) \\ &= P(X(t_{n+1}) = j_{n+1} | X(t_n) = j_n) \end{aligned}$$

for all finite sequence of times $t_0 < \cdots < t_n < t_{n+1} \in I$ and of states $j_0, \cdots, j_n < j_{n+1} \in S$.

- A random process is called a **Markov Process** if conditional on the current state, its future is independent of its past.
- Process $X_0, X_1, \cdots$ satisfies the Markov property if

$$P(X_{n+1} = x_{n+1} | X_n = x_n, \cdots, X_0 = x_0) = P(X_{n+1} = x_{n+1} | X_n = x_n)$$

for all n and all $x_i \in \Omega$.

Markov property

More formally, $X(t)$ is Markovian if the following property holds:

$$\begin{aligned} P(X(t_{n+1}) = j_{n+1} | X(t_n) = j_n, \cdots, X(t_0) = j_0) \\ = P(X(t_{n+1}) = j_{n+1} | X(t_n) = j_n) \end{aligned}$$

for all finite sequence of times $t_0 < \cdots < t_n < t_{n+1} \in I$ and of states $j_0, \cdots, j_n < j_{n+1} \in S$.

- A random process is called a **Markov Process** if conditional on the current state, its future is independent of its past.
- Process $X_0, X_1, \cdots$ satisfies the Markov property if

$$P(X_{n+1} = x_{n+1} | X_n = x_n, \cdots, X_0 = x_0) = P(X_{n+1} = x_{n+1} | X_n = x_n)$$

  for all n and all $x_i \in \Omega$.
- The term Markov property refers to the memoryless property of a stochastic process.

Discrete time Markov Chain

A stochastic process $X_0, X_1, \cdots$ of discrete time and discrete space is a **Markov chain** if it satisfies the Markov property, i.e.,

Discrete time Markov Chain

A stochastic process $X_0, X_1, \cdots$ of discrete time and discrete space is a **Markov chain** if it satisfies the Markov property, i.e.,

$$P(X_{n+1} = x_{n+1} | X_n = x_n, \cdots, X_0 = x_0) = P(X_{n+1} = x_{n+1} | X_n = x_n)$$

for all n and all $x_i \in \Omega$.

Discrete time Markov Chain

A stochastic process $X_0, X_1, \cdots$ of discrete time and discrete space is a **Markov chain** if it satisfies the Markov property, i.e.,

$$P(X_{n+1} = x_{n+1} | X_n = x_n, \cdots, X_0 = x_0) = P(X_{n+1} = x_{n+1} | X_n = x_n)$$

for all n and all $x_i \in \Omega$.

- A Markov chain $\{X_t\}$ is said to be time homogeneous if $P(X_{s+t} = j | X_s = i)$ is independent of s. When this holds, putting $s = 0$ gives

$$P(X_{s+t} = j | X_s = i) = P(X_t = j | X_0 = i).$$

Discrete time Markov Chain

A stochastic process $X_0, X_1, \cdots$ of discrete time and discrete space is a **Markov chain** if it satisfies the Markov property, i.e.,

$$P(X_{n+1} = x_{n+1} | X_n = x_n, \cdots, X_0 = x_0) = P(X_{n+1} = x_{n+1} | X_n = x_n)$$

for all n and all $x_i \in \Omega$.

- A Markov chain $\{X_t\}$ is said to be time homogeneous if $P(X_{s+t} = j | X_s = i)$ is independent of s. When this holds, putting $s = 0$ gives

$$P(X_{s+t} = j | X_s = i) = P(X_t = j | X_0 = i).$$

- If moreover $P(X_{n+1} = j | X_n = i) = P_{ij}$ is independent of n, then X is said to be time homogeneous Markov chain.

Examples

$\Omega = \{A, B\}$;

Examples

$\Omega = \{A, B\}$;

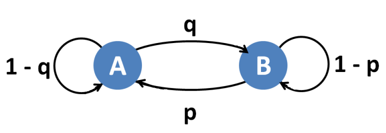
$$\mu(A,B) = q, \mu(A,A) = 1-q,$$
$$\mu(B,A) = p, \mu(B,B) = 1-p$$

Examples

$\Omega = \{A, B\}$;

$$\mu(A, B) = q, \mu(A, A) = 1 - q,$$
$$\mu(B, A) = p, \mu(B, B) = 1 - p$$

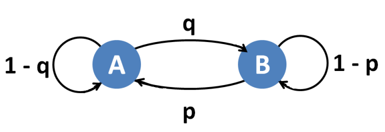
and $\pi(A) = 1, \pi(B) = 0$, so that at time 0 we always start at A.



Examples

$\Omega = \{A, B\}$;

$$\mu(A, B) = q, \mu(A, A) = 1 - q,$$
$$\mu(B, A) = p, \mu(B, B) = 1 - p$$



and $\pi(A) = 1, \pi(B) = 0$, so that at time 0 we always start at A.

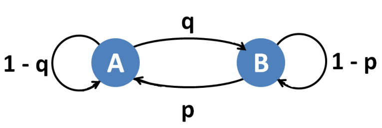
$\Omega = \mathbb{Z}$;

Examples

$\Omega = \{A, B\}$;

$$\mu(A,B) = q, \mu(A,A) = 1-q,$$
$$\mu(B,A) = p, \mu(B,B) = 1-p$$

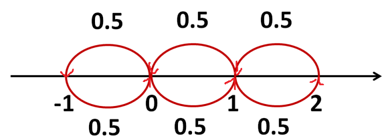
and $\pi(A) = 1, \pi(B) = 0$, so that at time 0 we always start at A.



$\Omega = \mathbb{Z}$;

$$\mu(a, a-1) = \frac{1}{2}, \mu(a, a+1) = \frac{1}{2},$$
$$\mu(a,b) = 0, \text{ if } b \neq a \pm 1, \forall a \in \mathbb{Z}$$

and $\pi(10) = 1$, $\pi(a) = 0$ if $a \neq 10$, so at time 0 we always start at 10.



Example of random walk

A graph $G = (V, E)$ consists of a vertex set V and an edge set E, where the elements of E are unordered pairs of vertices:
$E \subset \{(x, y) : x, y \in V, x \neq y\}$.

Example of random walk

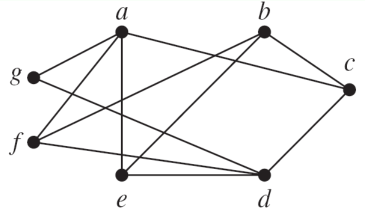
A graph $G = (V, E)$ consists of a vertex set V and an edge set E, where the elements of E are unordered pairs of vertices:
$E \subset \{(x, y) : x, y \in V, x \neq y\}$.
The degree $deg(x)$ of a vertex x is the number of neighbours of x.

Example of random walk

A graph $G = (V, E)$ consists of a vertex set V and an edge set E, where the elements of E are unordered pairs of vertices:
$E \subset \{(x, y) : x, y \in V, x \neq y\}$.
The degree $deg(x)$ of a vertex x is the number of neighbours of x.
In this case, $\Omega = V$ and is finite, $\mu(x, y) = \frac{1}{deg(x)}$ if y is a neighbour of x and $\mu(x, y) = 0$ otherwise.

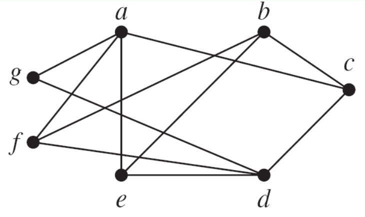
Example of random walk

A graph $G = (V, E)$ consists of a vertex set V and an edge set E, where the elements of E are unordered pairs of vertices:
$E \subset \{(x, y) : x, y \in V, x \neq y\}$.
The degree $deg(x)$ of a vertex x is the number of neighbours of x.
In this case, $\Omega = V$ and is finite,
$\mu(x, y) = \frac{1}{deg(x)}$ if y is a neighbour of x and $\mu(x, y) = 0$ otherwise.
$\pi(a) = 1$, and $\pi(x) = 0$ for every $x \in V \setminus \{a\}$, so at time 0 we start from vertex a.



Example of random walk

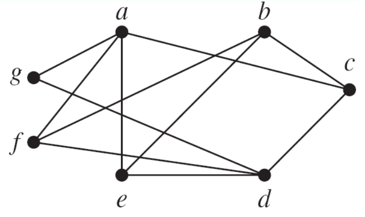
A graph $G = (V, E)$ consists of a vertex set V and an edge set E, where the elements of E are unordered pairs of vertices:
$E \subset \{(x, y) : x, y \in V, x \neq y\}$.
The degree $deg(x)$ of a vertex x is the number of neighbours of x.
In this case, $\Omega = V$ and is finite, $\mu(x, y) = \frac{1}{deg(x)}$ if y is a neighbour of x and $\mu(x, y) = 0$ otherwise.
$\pi(a) = 1$, and $\pi(x) = 0$ for every $x \in V \backslash \{a\}$, so at time 0 we start from vertex a.



$deg(a) = 4, deg(b) = 3$
$deg(c) = 3, deg(d) = 4$
$deg(e) = 3, deg(f) = 3$
$deg(g) = 2$

Example of random walk

A graph $G = (V, E)$ consists of a vertex set V and an edge set E, where the elements of E are unordered pairs of vertices:
$E \subset \{(x, y) : x, y \in V, x \neq y\}$.
The degree $deg(x)$ of a vertex x is the number of neighbours of x.
In this case, $\Omega = V$ and is finite, $\mu(x, y) = \frac{1}{deg(x)}$ if y is a neighbour of x and $\mu(x, y) = 0$ otherwise.
$\pi(a) = 1$, and $\pi(x) = 0$ for every $x \in V \backslash \{a\}$, so at time 0 we start from vertex a.



$deg(a) = 4, deg(b) = 3$
$deg(c) = 3, deg(d) = 4$
$deg(e) = 3, deg(f) = 3$
$deg(g) = 2$

Thus, a graph can be considered as a random walk, i.e., a Markov chain.

Probability transition matrix

Definition

Let a Markov chain have $P_{x,y}^{(t+1)} = P[X_{t+1} = y | X_t = x]$, and the finite state space be $\Omega = [n]$. This gives us a **probability transition matrix** $P^{(t+1)}$ at time t.

Probability transition matrix

Definition

Let a Markov chain have $P_{x,y}^{(t+1)} = P[X_{t+1} = y | X_t = x]$, and the finite state space be $\Omega = [n]$. This gives us a **probability transition matrix** $P^{(t+1)}$ at time t.The probability transition matrix is an $n \times n$ matrix of nonnegative entries such that the sum over each row of $P^{(t)}$ is 1, since $\forall n$ and $\forall x_i \in \Omega$

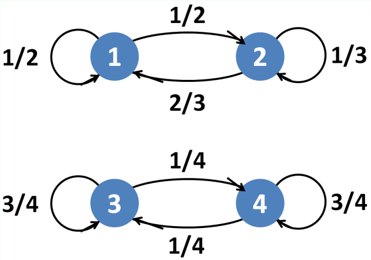
$$\sum_y P_{x,y}^{(t+1)} = \sum_y P[X_{t+1} = y | X_t = x] = \frac{\sum_y P[X_{t+1} = y, X_t = x]}{P[X_t = x]} = 1$$

Probability transition matrix

Definition

Let a Markov chain have $P^{(t+1)}_{x,y} = P[X_{t+1} = y | X_t = x]$, and the finite state space be $\Omega = [n]$. This gives us a **probability transition matrix** $P^{(t+1)}$ at time t.The probability transition matrix is an $n \times n$ matrix of nonnegative entries such that the sum over each row of $P^{(t)}$ is 1, since $\forall n$ and $\forall x_i \in \Omega$

$$\sum_y P^{(t+1)}_{x,y} = \sum_y P[X_{t+1} = y | X_t = x] = \frac{\sum_y P[X_{t+1} = y, X_t = x]}{P[X_t = x]} = 1$$

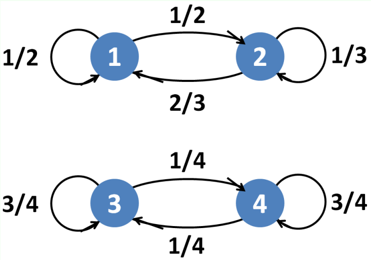


Probability transition matrix

Definition

Let a Markov chain have $P_{x,y}^{(t+1)} = P[X_{t+1} = y | X_t = x]$, and the finite state space be $\Omega = [n]$. This gives us a **probability transition matrix** $P^{(t+1)}$ at time t.The probability transition matrix is an $n \times n$ matrix of nonnegative entries such that the sum over each row of $P^{(t)}$ is 1, since $\forall n$ and $\forall x_i \in \Omega$

$$\sum_y P_{x,y}^{(t+1)} = \sum_y P[X_{t+1} = y | X_t = x] = \frac{\sum_y P[X_{t+1} = y, X_t = x]}{P[X_t = x]} = 1$$



- For the time homogeneous Markov chain,

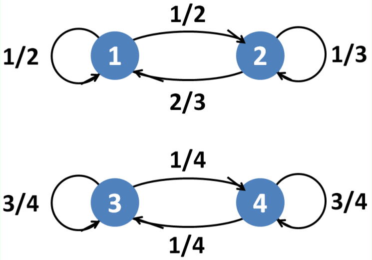
$$P = \begin{pmatrix} \frac{1}{2} & \frac{1}{2} & 0 & 0 \\ \frac{1}{3} & \frac{2}{3} & 0 & 0 \\ 0 & 0 & \frac{3}{4} & \frac{1}{4} \\ 0 & 0 & \frac{1}{4} & \frac{3}{4} \end{pmatrix}$$

Probability transition matrix

Definition

Let a Markov chain have $P_{x,y}^{(t+1)} = P[X_{t+1} = y | X_t = x]$, and the finite state space be $\Omega = [n]$. This gives us a **probability transition matrix** $P^{(t+1)}$ at time t.The probability transition matrix is an $n \times n$ matrix of nonnegative entries such that the sum over each row of $P^{(t)}$ is 1, since $\forall n$ and $\forall x_i \in \Omega$

$$\sum_y P_{x,y}^{(t+1)} = \sum_y P[X_{t+1} = y | X_t = x] = \frac{\sum_y P[X_{t+1} = y, X_t = x]}{P[X_t = x]} = 1$$



- For the time homogeneous Markov chain,

$$P = \begin{pmatrix} \frac{1}{2} & \frac{1}{2} & 0 & 0 \\ \frac{1}{3} & \frac{2}{3} & 0 & 0 \\ 0 & 0 & \frac{3}{4} & \frac{1}{4} \\ 0 & 0 & \frac{1}{4} & \frac{3}{4} \end{pmatrix}$$

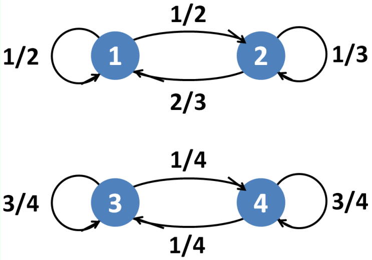
- $P_{1,2} = P[X_{t+1} = 2 | X_t = 1] = \frac{1}{2}$ and $P_{1,3} = P[X_{t+1} = 3 | X_t = 1] = 0$

Probability transition matrix

Definition

Let a Markov chain have $P_{x,y}^{(t+1)} = P[X_{t+1} = y | X_t = x]$, and the finite state space be $\Omega = [n]$. This gives us a **probability transition matrix** $P^{(t+1)}$ at time t.The probability transition matrix is an $n \times n$ matrix of nonnegative entries such that the sum over each row of $P^{(t)}$ is 1, since $\forall n$ and $\forall x_i \in \Omega$

$$\sum_y P_{x,y}^{(t+1)} = \sum_y P[X_{t+1} = y | X_t = x] = \frac{\sum_y P[X_{t+1} = y, X_t = x]}{P[X_t = x]} = 1$$



- For the time homogeneous Markov chain,

$$P = \begin{pmatrix} \frac{1}{2} & \frac{1}{2} & 0 & 0 \\ \frac{1}{3} & \frac{2}{3} & 0 & 0 \\ 0 & 0 & \frac{3}{4} & \frac{1}{4} \\ 0 & 0 & \frac{1}{4} & \frac{3}{4} \end{pmatrix}$$

- $P_{1,2} = P[X_{t+1} = 2 | X_t = 1] = \frac{1}{2}$ and $P_{1,3} = P[X_{t+1} = 3 | X_t = 1] = 0$
- $\sum_{i=1}^4 P_{1,i} = 1$, while $\sum_{i=1}^4 P_{i,1} \neq 1$.

State distribution

Definition

Let $\pi^{(t)}$ be the **state distribution** of the chain at time t, that $\pi_x^{(t)} = P[X_t = x]$.

State distribution

Definition

Let $\pi^{(t)}$ be the **state distribution** of the chain at time t, that $\pi_x^{(t)} = P[X_t = x]$.
For a finite chain, $\pi^{(t)}$ is a vector of n nonnegative entries such that $\sum_x \pi_x^{(t)} = 1$. Then, it holds that $\pi^{(t+1)} = \pi^{(t)} P^{(t+1)}$. We apply the law of total probability.

State distribution

Definition

Let $\pi^{(t)}$ be the **state distribution** of the chain at time t, that $\pi_x^{(t)} = P[X_t = x]$.
For a finite chain, $\pi^{(t)}$ is a vector of n nonnegative entries such that $\sum_x \pi_x^{(t)} = 1$. Then, it holds that $\pi^{(t+1)} = \pi^{(t)} P^{(t+1)}$. We apply the law of total probability.

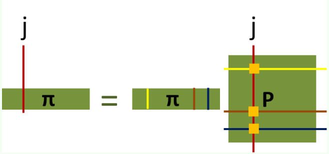
$$
\begin{aligned}
&\pi_y^{(t+1)} \\
&= \sum_x P[X_{t+1} = y | X_t = x] P[X_t = x] \\
&= \sum_x \pi_x^{(t)} P_{x,y}^{(t+1)} \\
&= (\pi^{(t)} P^{(t+1)})_y
\end{aligned}
$$

State distribution

Definition

Let $\pi^{(t)}$ be the **state distribution** of the chain at time t, that $\pi_x^{(t)} = P[X_t = x]$.
For a finite chain, $\pi^{(t)}$ is a vector of n nonnegative entries such that $\sum_x \pi_x^{(t)} = 1$. Then, it holds that $\pi^{(t+1)} = \pi^{(t)} P^{(t+1)}$. We apply the law of total probability.

$$\begin{aligned}
&\pi_y^{(t+1)} \\
&= \sum_x P[X_{t+1} = y | X_t = x] P[X_t = x] \\
&= \sum_x \pi_x^{(t)} P_{x,y}^{(t+1)} \\
&= (\pi^{(t)} P^{(t+1)})_y
\end{aligned}$$



Example of State distribution

Example

$$(0.4, 0.6, 0, 0) \cdot \begin{pmatrix} \frac{1}{2} & \frac{1}{2} & 0 & 0 \\ \frac{1}{3} & \frac{2}{3} & 0 & 0 \\ 0 & 0 & \frac{3}{4} & \frac{1}{4} \\ 0 & 0 & \frac{1}{4} & \frac{3}{4} \end{pmatrix} = (0.4, 0.6, 0, 0).$$

Example of State distribution

Example

$$(0.4, 0.6, 0, 0) \cdot \begin{pmatrix} \frac{1}{2} & \frac{1}{2} & 0 & 0 \\ \frac{1}{3} & \frac{2}{3} & 0 & 0 \\ 0 & 0 & \frac{3}{4} & \frac{1}{4} \\ 0 & 0 & \frac{1}{4} & \frac{3}{4} \end{pmatrix} = (0.4, 0.6, 0, 0).$$

$$(0, 0, 0.5, 0.5) \cdot \begin{pmatrix} \frac{1}{2} & \frac{1}{2} & 0 & 0 \\ \frac{1}{3} & \frac{2}{3} & 0 & 0 \\ 0 & 0 & \frac{3}{4} & \frac{1}{4} \\ 0 & 0 & \frac{1}{4} & \frac{3}{4} \end{pmatrix} = (0, 0, 0.5, 0.5).$$

Example of State distribution

Example

$$(0.4, 0.6, 0, 0) \cdot \begin{pmatrix} \frac{1}{2} & \frac{1}{2} & 0 & 0 \\ \frac{1}{3} & \frac{2}{3} & 0 & 0 \\ 0 & 0 & \frac{3}{4} & \frac{1}{4} \\ 0 & 0 & \frac{1}{4} & \frac{3}{4} \end{pmatrix} = (0.4, 0.6, 0, 0).$$

$$(0, 0, 0.5, 0.5) \cdot \begin{pmatrix} \frac{1}{2} & \frac{1}{2} & 0 & 0 \\ \frac{1}{3} & \frac{2}{3} & 0 & 0 \\ 0 & 0 & \frac{3}{4} & \frac{1}{4} \\ 0 & 0 & \frac{1}{4} & \frac{3}{4} \end{pmatrix} = (0, 0, 0.5, 0.5).$$

$$(0.1, 0.9, 0, 0) \cdot \begin{pmatrix} \frac{1}{2} & \frac{1}{2} & 0 & 0 \\ \frac{1}{3} & \frac{2}{3} & 0 & 0 \\ 0 & 0 & \frac{3}{4} & \frac{1}{4} \\ 0 & 0 & \frac{1}{4} & \frac{3}{4} \end{pmatrix} = (0.35, 0.65, 0, 0).$$

Stationary distributions

Definition

A **stationary distribution** of a finite Markov chain with transition matrix P is a probability distribution π such that $\pi P = \pi$.

Stationary distributions

Definition

A **stationary distribution** of a finite Markov chain with transition matrix P is a probability distribution π such that $\pi P = \pi$.

- For some Markov chains, no matter what the initial distribution is, after running the chain for a while, the distribution of the chain approaches the stationary distribution.

Stationary distributions

Definition

A **stationary distribution** of a finite Markov chain with transition matrix P is a probability distribution π such that $\pi P = \pi$.

- For some Markov chains, no matter what the initial distribution is, after running the chain for a while, the distribution of the chain approaches the stationary distribution.
- E.g., $P^{20} = \begin{pmatrix} 0.4 & 0.6 & 0 & 0 \\ 0.4 & 0.6 & 0 & 0 \\ 0 & 0 & 0.5 & 0.5 \\ 0 & 0 & 0.5 & 0.5 \end{pmatrix}$. The chain could converge to any distribution which is a linear combination of $(0.4, 0.6, 0, 0)$ and $(0, 0, 0.5, 0.5)$.

Stationary distributions

Definition

A **stationary distribution** of a finite Markov chain with transition matrix P is a probability distribution π such that $\pi P = \pi$.

- For some Markov chains, no matter what the initial distribution is, after running the chain for a while, the distribution of the chain approaches the stationary distribution.
- E.g., $P^{20} = \begin{pmatrix} 0.4 & 0.6 & 0 & 0 \\ 0.4 & 0.6 & 0 & 0 \\ 0 & 0 & 0.5 & 0.5 \\ 0 & 0 & 0.5 & 0.5 \end{pmatrix}$. The chain could converge to any distribution which is a linear combination of $(0.4, 0.6, 0, 0)$ and $(0, 0, 0.5, 0.5)$.We observe that the original chain P can be broken into two disjoint Markov chains, which have their own stationary distributions.

Stationary distributions

Definition

A **stationary distribution** of a finite Markov chain with transition matrix P is a probability distribution π such that $\pi P = \pi$.

- For some Markov chains, no matter what the initial distribution is, after running the chain for a while, the distribution of the chain approaches the stationary distribution.
- E.g., $P^{20} = \begin{pmatrix} 0.4 & 0.6 & 0 & 0 \\ 0.4 & 0.6 & 0 & 0 \\ 0 & 0 & 0.5 & 0.5 \\ 0 & 0 & 0.5 & 0.5 \end{pmatrix}$. The chain could converge to any distribution which is a linear combination of $(0.4, 0.6, 0, 0)$ and $(0, 0, 0.5, 0.5)$.We observe that the original chain P can be broken into two disjoint Markov chains, which have their own stationary distributions. We say that the chain is **reducible**.

Irreducibility

Definition

State y is accessible from state x if it is possible for the chain to visit state y if the chain starts in state x, in other words, $P^n(x, y) > 0, \forall n$. State x **communicates** with state y if y is accessible from x and x is accessible from y.

Irreducibility

Definition

State y is accessible from state x if it is possible for the chain to visit state y if the chain starts in state x, in other words, $P^n(x, y) > 0, \forall n$. State x **communicates** with state y if y is accessible from x and x is accessible from y. We say that the Markov chain is **irreducible** if all pairs of states communicates.

Irreducibility

Definition

State y is accessible from state x if it is possible for the chain to visit state y if the chain starts in state x, in other words, $P^n(x, y) > 0, \forall n$. State x **communicates** with state y if y is accessible from x and x is accessible from y. We say that the Markov chain is **irreducible** if all pairs of states communicates.

- y is accessible from x means that y is connected from x in the transition graph, i.e., there is a directed path from x to y

Irreducibility

Definition

State y is accessible from state x if it is possible for the chain to visit state y if the chain starts in state x, in other words, $P^n(x, y) > 0, \forall n$. State x **communicates** with state y if y is accessible from x and x is accessible from y. We say that the Markov chain is **irreducible** if all pairs of states communicates.

- y is accessible from x means that y is connected from x in the transition graph, i.e., there is a directed path from x to y
- x communicates with y means that x and y are strongly connected in the transition graph

Irreducibility

Definition

State y is accessible from state x if it is possible for the chain to visit state y if the chain starts in state x, in other words, $P^n(x, y) > 0, \forall n$. State x **communicates** with state y if y is accessible from x and x is accessible from y. We say that the Markov chain is **irreducible** if all pairs of states communicates.

- y is accessible from x means that y is connected from x in the transition graph, i.e., there is a directed path from x to y
- x communicates with y means that x and y are strongly connected in the transition graph
- A finite Markov chain is irreducible if and only if its transition graph is strongly connected

Irreducibility

Definition

State y is accessible from state x if it is possible for the chain to visit state y if the chain starts in state x, in other words, $P^n(x, y) > 0, \forall n$. State x **communicates** with state y if y is accessible from x and x is accessible from y. We say that the Markov chain is **irreducible** if all pairs of states communicates.

- y is accessible from x means that y is connected from x in the transition graph, i.e., there is a directed path from x to y
- x communicates with y means that x and y are strongly connected in the transition graph
- A finite Markov chain is irreducible if and only if its transition graph is strongly connected
- The Markov chain with transition matrix P is not irreducible

Irreducibility

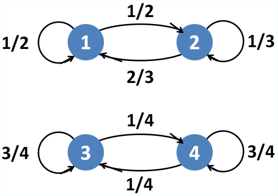
Definition

State y is accessible from state x if it is possible for the chain to visit state y if the chain starts in state x, in other words, $P^n(x, y) > 0, \forall n$. State x **communicates** with state y if y is accessible from x and x is accessible from y. We say that the Markov chain is **irreducible** if all pairs of states communicates.

- y is accessible from x means that y is connected from x in the transition graph, i.e., there is a directed path from x to y
- x communicates with y means that x and y are strongly connected in the transition graph
- A finite Markov chain is irreducible if and only if its transition graph is strongly connected
- The Markov chain with transition matrix P is not irreducible
- The communication relation satisfies reflexive, symmetric, and transitive.

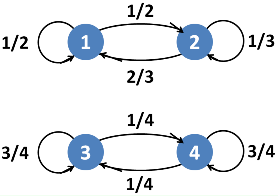
Examples of irreducibility

Examples of irreducibility



Whether is the random walk irreducibility or not?

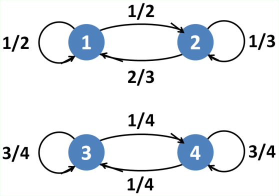
Examples of irreducibility



Whether is the random walk irreducibility or not?

Not! Why?

Examples of irreducibility

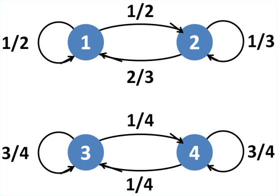


Whether is the random walk irreducibility or not?

Not! Why?

There are two connected components in this graph. Thus, not path exists for the pair of vertices v_1 and v_3.

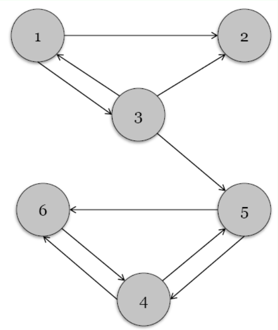
Examples of irreducibility



Whether is the random walk irreducibility or not?

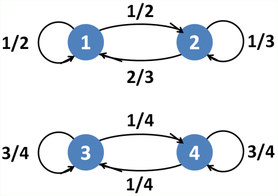
Not! Why?

There are two connected components in this graph. Thus, not path exists for the pair of vertices v_1 and v_3.



Whether is the random walk irreducibility or not?

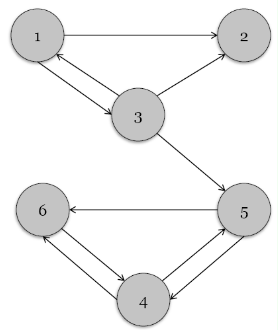
Examples of irreducibility



Whether is the random walk irreducibility or not?

Not! Why?

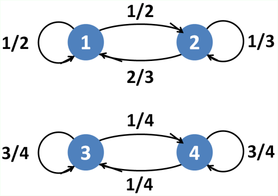
There are two connected components in this graph. Thus, not path exists for the pair of vertices v_1 and v_3.



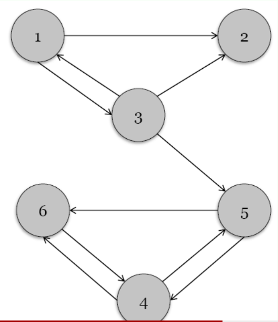
Whether is the random walk irreducibility or not?

Not! Why?

Examples of irreducibility



Whether is the random walk irreducibility or not?
Not! Why?
There are two connected components in this graph. Thus, not path exists for the pair of vertices v_1 and v_3.



Whether is the random walk irreducibility or not?
Not! Why?
For v_2, it only has the in-edges, i.e., there is not path from v_2 to another vertices, e.g., v_1.

Period

Definition

The **period** of a state x is the greatest common divisor (gcd), such that $d_x = gcd\{n|(P^n)_{x,x} > 0\}$.

Period

Definition

The **period** of a state x is the greatest common divisor (gcd), such that $d_x = gcd\{n|(P^n)_{x,x} > 0\}$.
If $\forall n \geq 1$, $(P^n)_{x,x} = 0$, then $d_x = \infty$.

Period

Definition

The **period** of a state x is the greatest common divisor (gcd), such that $d_x = gcd\{n|(P^n)_{x,x} > 0\}$.
If $\forall n \geq 1$, $(P^n)_{x,x} = 0$, then $d_x = \infty$.

- For example, suppose that the period of state x is $d_x = 3$. Then, starting from state x, chain $x, \bigcirc, \bigcirc, \square, \bigcirc, \bigcirc, \square, \cdots,$ only the squares are possible to be x.

Period

Definition

The **period** of a state x is the greatest common divisor (gcd), such that $d_x = gcd\{n|(P^n)_{x,x} > 0\}$.
If $\forall n \geq 1$, $(P^n)_{x,x} = 0$, then $d_x = \infty$.

- For example, suppose that the period of state x is $d_x = 3$. Then, starting from state x, chain $x, \bigcirc, \bigcirc, \square, \bigcirc, \bigcirc, \square, \cdots,$ only the squares are possible to be x.
- In the transition graph of a finite Markov chain, $(P^n)_{x,x} > 0$ is equivalent to that x is on a cycle of length n.

Period

Definition

The **period** of a state x is the greatest common divisor (gcd), such that $d_x = gcd\{n|(P^n)_{x,x} > 0\}$.
If $\forall n \geq 1$, $(P^n)_{x,x} = 0$, then $d_x = \infty$.

- For example, suppose that the period of state x is $d_x = 3$. Then, starting from state x, chain $x, \bigcirc, \bigcirc, \square, \bigcirc, \bigcirc, \square, \cdots,$ only the squares are possible to be x.
- In the transition graph of a finite Markov chain, $(P^n)_{x,x} > 0$ is equivalent to that x is on a cycle of length n.
- Period of a state x is the greatest common devisor of the lengths of cycles passing x.

Period

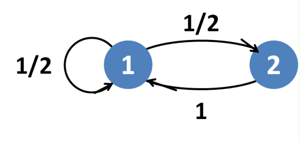
Definition

The **period** of a state x is the greatest common divisor (gcd), such that $d_x = gcd\{n|(P^n)_{x,x} > 0\}$.
If $\forall n \geq 1$, $(P^n)_{x,x} = 0$, then $d_x = \infty$.

- For example, suppose that the period of state x is $d_x = 3$. Then, starting from state x, chain $x, \bigcirc, \bigcirc, \square, \bigcirc, \bigcirc, \square, \cdots,$ only the squares are possible to be x.
- In the transition graph of a finite Markov chain, $(P^n)_{x,x} > 0$ is equivalent to that x is on a cycle of length n.
- Period of a state x is the greatest common devisor of the lengths of cycles passing x.

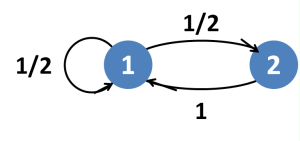
A state is **aperiodic** if its period is 1. A Markov chain is **aperiodic** if all its states are aperiodic.

Example period



What is the period of state 1?

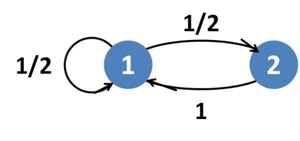
Example period



What is the period of state 1?

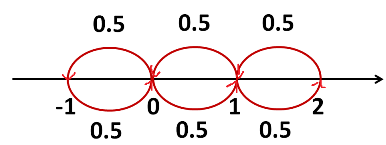
Period of a state x will be 1, if it has a self-loop.

Example period



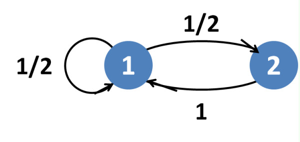
What is the period of state 1?

Period of a state x will be 1, if it has a self-loop.

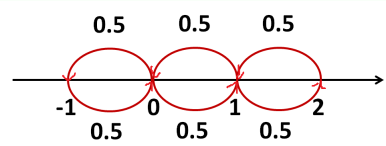


What is the period of state 0?

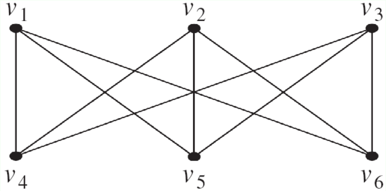
Example period



What is the period of state 1?
Period of a state x will be 1, if it has a self-loop.



What is the period of state 0?



What is the period of vertex v_1?

Properties of period

Theorem

1. If the states x and y communicate, then $d_x = d_y$.

Properties of period

Theorem

1. If the states x and y communicate, then $d_x = d_y$.Consequently, if the Markov chain is irreducible, then all states have the same period.

Properties of period

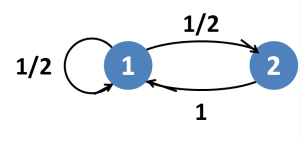
Theorem

1. If the states x and y communicate, then $d_x = d_y$.Consequently, if the Markov chain is irreducible, then all states have the same period.
2. We have $(P^n)_{x,x} = 0$ if $n(\mod d_x) \neq 0$.

Properties of period

Theorem

1. If the states x and y communicate, then $d_x = d_y$.Consequently, if the Markov chain is irreducible, then all states have the same period.
2. We have $(P^n)_{x,x} = 0$ if $n \pmod{d_x} \neq 0$.

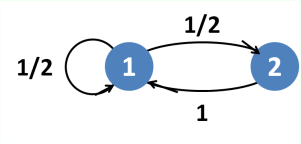


What are the periods of all states in the graph?

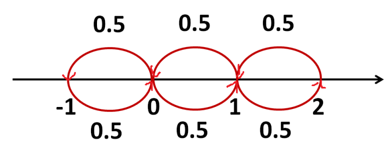
Properties of period

Theorem

1. If the states x and y communicate, then $d_x = d_y$. Consequently, if the Markov chain is irreducible, then all states have the same period.
2. We have $(P^n)_{x,x} = 0$ if $n \pmod{d_x} \neq 0$.



What are the periods of all states in the graph?



What are the periods of all states in the graph?

Existence of the stationary distribution

Theorem

Let $X_0, X_1, \cdots$, be an **irreducible** and **aperiodic** Markov chain with transition matrix P. Then, $\lim_{n\to\infty}(P^n)_{ij}$ exists and independent of i, denoted as $\lim_{n\to\infty}(P^n)_{ij} = \pi(j)$. We also have

Existence of the stationary distribution

Theorem

Let $X_0, X_1, \cdots$, be an **irreducible** and **aperiodic** Markov chain with transition matrix P. Then, $\lim_{n\to\infty}(P^n)_{ij}$ exists and independent of i, denoted as $\lim_{n\to\infty}(P^n)_{ij} = \pi(j)$. We also have

- $$\lim_{n\to\infty} P^n = \begin{pmatrix} \pi(1) & \pi(2) & \cdots & \pi(j) & \cdots \\ \pi(1) & \pi(2) & \cdots & \pi(j) & \cdots \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ \pi(1) & \pi(2) & \cdots & \pi(j) & \cdots \\ \cdots & \cdots & \cdots & \cdots & \cdots \end{pmatrix} \tag{1}$$

Existence of the stationary distribution

Theorem

Let $X_0, X_1, \cdots$, be an **irreducible** and **aperiodic** Markov chain with transition matrix P. Then, $\lim_{n\to\infty}(P^n)_{ij}$ exists and independent of i, denoted as $\lim_{n\to\infty}(P^n)_{ij} = \pi(j)$. We also have

- $$\lim_{n\to\infty} P^n = \begin{pmatrix} \pi(1) & \pi(2) & \cdots & \pi(j) & \cdots \\ \pi(1) & \pi(2) & \cdots & \pi(j) & \cdots \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ \pi(1) & \pi(2) & \cdots & \pi(j) & \cdots \\ \cdots & \cdots & \cdots & \cdots & \cdots \end{pmatrix} \tag{1}$$

- $\pi(j) = \sum_{i=1}^{\infty} \pi(i) P_{ij}$, and $\sum_{i=1}^{\infty} \pi(i) = 1$.

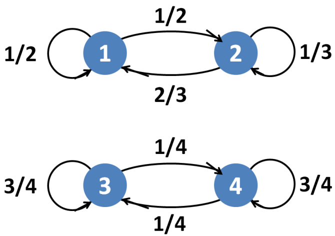
Existence of the stationary distribution

Theorem

Let $X_0, X_1, \cdots$, be an **irreducible** and **aperiodic** Markov chain with transition matrix P. Then, $\lim_{n\to\infty}(P^n)_{ij}$ exists and independent of i, denoted as $\lim_{n\to\infty}(P^n)_{ij} = \pi(j)$. We also have

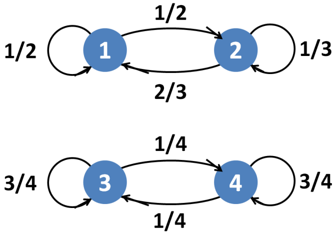
- $$\lim_{n\to\infty} P^n = \begin{pmatrix} \pi(1) & \pi(2) & \cdots & \pi(j) & \cdots \\ \pi(1) & \pi(2) & \cdots & \pi(j) & \cdots \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ \pi(1) & \pi(2) & \cdots & \pi(j) & \cdots \\ \cdots & \cdots & \cdots & \cdots & \cdots \end{pmatrix} \tag{1}$$
- $\pi(j) = \sum_{i=1}^{\infty} \pi(i)P_{ij}$, and $\sum_{i=1}^{\infty} \pi(i) = 1$.
- π is the unique and non-negative solution for equation $\pi P = \pi$.

Counter examples

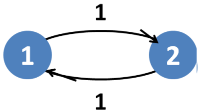


Both $(0.4, 0.6, 0, 0)$ and $(0, 0, 0.5, 0.5)$ are stationary distributions of this Markov chain.

Counter examples



Both $(0.4, 0.6, 0, 0)$ and $(0, 0, 0.5, 0.5)$ are stationary distributions of this Markov chain.



The transition matrix is $P = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$. For $n \in N$, we have

$$P^{2n} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \text{ and } P^{2n+1} = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$$

Counter example Cont'd

The transition matrix for a Markov chain is $\begin{pmatrix} 0 & 1 & 0 & 0 \\ \frac{1}{4} & 0 & \frac{3}{4} & 0 \\ 0 & \frac{1}{2} & 0 & \frac{1}{2} \\ 0 & 0 & 1 & 0 \end{pmatrix}$

Counter example Cont'd

The transition matrix for a Markov chain is $\begin{pmatrix} 0 & 1 & 0 & 0 \\ \frac{1}{4} & 0 & \frac{3}{4} & 0 \\ 0 & \frac{1}{2} & 0 & \frac{1}{2} \\ 0 & 0 & 1 & 0 \end{pmatrix}$

a. Whether is this chain irreducible? Why?

Counter example Cont'd

The transition matrix for a Markov chain is $\begin{pmatrix} 0 & 1 & 0 & 0 \\ \frac{1}{4} & 0 & \frac{3}{4} & 0 \\ 0 & \frac{1}{2} & 0 & \frac{1}{2} \\ 0 & 0 & 1 & 0 \end{pmatrix}$

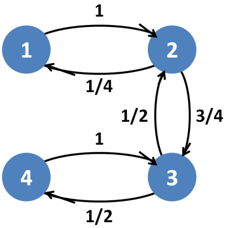
a. Whether is this chain irreducible? Why?

b. Is this chain aperiodic? If it is, show your reason; if not, what is the period?

Counter example Cont'd

The transition matrix for a Markov chain is $\begin{pmatrix} 0 & 1 & 0 & 0 \\ \frac{1}{4} & 0 & \frac{3}{4} & 0 \\ 0 & \frac{1}{2} & 0 & \frac{1}{2} \\ 0 & 0 & 1 & 0 \end{pmatrix}$

a. Whether is this chain irreducible? Why?
b. Is this chain aperiodic? If it is, show your reason; if not, what is the period?

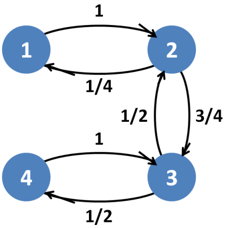


Counter example Cont'd

The transition matrix for a Markov chain is $\begin{pmatrix} 0 & 1 & 0 & 0 \\ \frac{1}{4} & 0 & \frac{3}{4} & 0 \\ 0 & \frac{1}{2} & 0 & \frac{1}{2} \\ 0 & 0 & 1 & 0 \end{pmatrix}$

a. Whether is this chain irreducible? Why?

b. Is this chain aperiodic? If it is, show your reason; if not, what is the period?



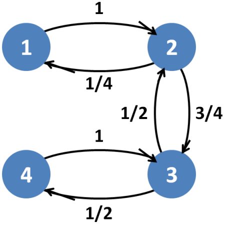
a. It is irreducible. Every pair of vertices communicates each other.

Counter example Cont'd

The transition matrix for a Markov chain is $\begin{pmatrix} 0 & 1 & 0 & 0 \\ \frac{1}{4} & 0 & \frac{3}{4} & 0 \\ 0 & \frac{1}{2} & 0 & \frac{1}{2} \\ 0 & 0 & 1 & 0 \end{pmatrix}$

a. Whether is this chain irreducible? Why?

b. Is this chain aperiodic? If it is, show your reason; if not, what is the period?



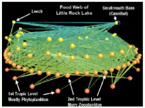
a. It is irreducible. Every pair of vertices communicates each other.

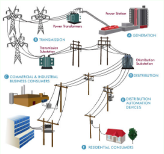
b. Period of every vertex is 2.

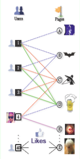
Outline

Graphs - why should we care?

Motivation





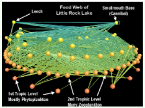


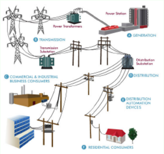
Graphs in real world

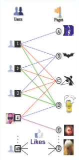
- "YahooWeb graph": 1B vertices(Web sites), 6B edges (links)

Graphs - why should we care?

Motivation



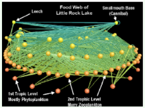



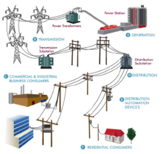
Graphs in real world

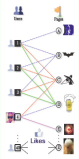
- "YahooWeb graph": 1B vertices(Web sites), 6B edges (links)
- Facebook, Twitter, etc: more than 1B users

Graphs - why should we care?

Motivation







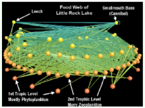
Graphs in real world

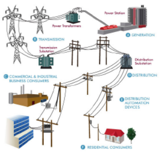
- "YahooWeb graph": 1B vertices(Web sites), 6B edges (links)
- Facebook, Twitter, etc: more than 1B users
- Food Web: all biologies, food chain

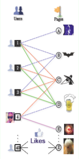
Graphs - why should we care?

Motivation







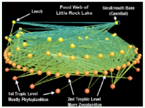



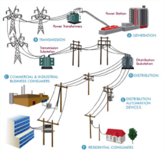
Graphs in real world

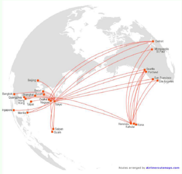
- “YahooWeb graph”: 1B vertices(Web sites), 6B edges (links)
- Facebook, Twitter, etc: more than 1B users
- Food Web: all biologies, food chain
- Power-grid: vertices (plants or consumers), edges (power lines)

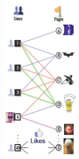
Graphs - why should we care?

Motivation









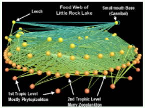
Graphs in real world

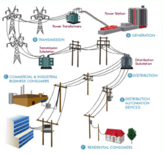
- “YahooWeb graph”: 1B vertices(Web sites), 6B edges (links)
- Facebook, Twitter, etc: more than 1B users
- Food Web: all biologies, food chain
- Power-grid: vertices (plants or consumers), edges (power lines)
- Airline route: vertices (airports), edges (flights)

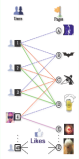
Graphs - why should we care?

Motivation









Graphs in real world

- "YahooWeb graph": 1B vertices(Web sites), 6B edges (links)
- Facebook, Twitter, etc: more than 1B users
- Food Web: all biologies, food chain
- Power-grid: vertices (plants or consumers), edges (power lines)
- Airline route: vertices (airports), edges (flights)
- Adoption: users purchase products, adopt services, etc.

Graphs - why should we care? Cont'd

Graphs model the relations between entities, e.g., person VS. person, Web page VS. Web pages, airport VS. airport, paper VS. paper, person VS. item, and entity VS. entity, etc.

Graphs - why should we care? Cont'd

Graphs model the relations between entities, e.g., person VS. person, Web page VS. Web pages, airport VS. airport, paper VS. paper, person VS. item, and entity VS. entity, etc.

- Many graphs are huge, such as social networks, citation networks, and adoption networks, etc, full of untrusted information, random things, and web spams, etc.

Graphs - why should we care? Cont'd

Graphs model the relations between entities, e.g., person VS. person, Web page VS. Web pages, airport VS. airport, paper VS. paper, person VS. item, and entity VS. entity, etc.

- Many graphs are huge, such as social networks, citation networks, and adoption networks, etc, full of untrusted information, random things, and web spams, etc.
- A natural question is how to search in the Web. It is a research field, called **information retrieval**.

Graphs - why should we care? Cont'd

Graphs model the relations between entities, e.g., person VS. person, Web page VS. Web pages, airport VS. airport, paper VS. paper, person VS. item, and entity VS. entity, etc.

- Many graphs are huge, such as social networks, citation networks, and adoption networks, etc, full of untrusted information, random things, and web spams, etc.
- A natural question is how to search in the Web. It is a research field, called **information retrieval**.
- How to find the relevant objects in a small and trusted set?

Graphs - why should we care? Cont'd

Graphs model the relations between entities, e.g., person VS. person, Web page VS. Web pages, airport VS. airport, paper VS. paper, person VS. item, and entity VS. entity, etc.

- Many graphs are huge, such as social networks, citation networks, and adoption networks, etc, full of untrusted information, random things, and web spams, etc.
- A natural question is how to search in the Web. It is a research field, called **information retrieval**.
- How to find the relevant objects in a small and trusted set?
  - Web contains many sources of information who to "trust"? (Trick: Trustworthy pages may point to each other!)

Graphs - why should we care? Cont'd

Graphs model the relations between entities, e.g., person VS. person, Web page VS. Web pages, airport VS. airport, paper VS. paper, person VS. item, and entity VS. entity, etc.

- Many graphs are huge, such as social networks, citation networks, and adoption networks, etc, full of untrusted information, random things, and web spams, etc.
- A natural question is how to search in the Web. It is a research field, called **information retrieval**.
- How to find the relevant objects in a small and trusted set?
  - Web contains many sources of information who to "trust"? (Trick: Trustworthy pages may point to each other!)
  - What is the "best" answer to query "DaSE"? (Trick: Ranking the related documents)

Ranking vertices on the graph

Ranking vertices on the graph

- All web pages are not equally "important".

Ranking vertices on the graph

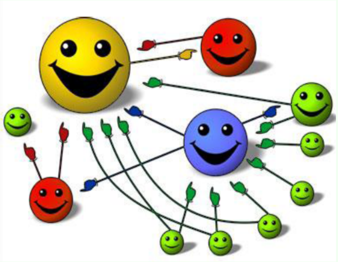
- All web pages are not equally "important".
- There is a large diversity in the web-graph node connectivity. Let's rank the pages by the link structure!

Ranking vertices on the graph

- All web pages are not equally "important".
- There is a large diversity in the web-graph node connectivity. Let's rank the pages by the link structure!
  - Page is more important if it has more links.

Ranking vertices on the graph

- All web pages are not equally “important”.
- There is a large diversity in the web-graph node connectivity. Let’s rank the pages by the link structure!
    - Page is more important if it has more links.
    - Are all in-links equal?



Outline

Intuition

- Web pages are important if people visit them a lot.

Intuition

- Web pages are important if people visit them a lot.
- But we cannot watch everybody using the Web.

Intuition

- Web pages are important if people visit them a lot.
- But we cannot watch everybody using the Web.
- A good surrogate for visiting pages is to assume people follow links randomly.

Intuition

- Web pages are important if people visit them a lot.
- But we cannot watch everybody using the Web.
- A good surrogate for visiting pages is to assume people follow links randomly.
- Leads to random surfer model:
  - Start at a random page and follow random outlinks repeatedly, from whatever page you are at (state transition).

Intuition

- Web pages are important if people visit them a lot.
- But we cannot watch everybody using the Web.
- A good surrogate for visiting pages is to assume people follow links randomly.
- Leads to random surfer model:
  - Start at a random page and follow random outlinks repeatedly, from whatever page you are at (state transition).
  - PageRank = limiting probability of being at a page (stationary distribution).

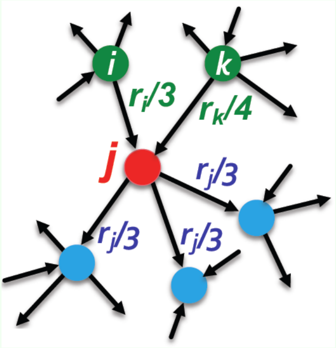
Intuition

- Web pages are important if people visit them a lot.
- But we cannot watch everybody using the Web.
- A good surrogate for visiting pages is to assume people follow links randomly.
- Leads to random surfer model:
  - Start at a random page and follow random outlinks repeatedly, from whatever page you are at (state transition).
  - PageRank = limiting probability of being at a page (stationary distribution).
  - Solve the recursive equation: "importance of a page = its share of the importance of each of its predecessor pages" (Equivalent to the random-surfer definition of PageRank)

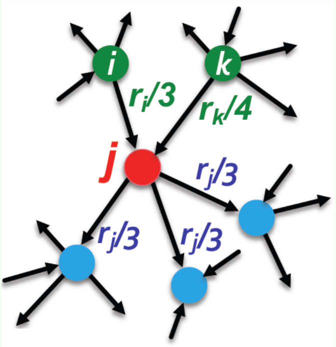
Intuition

- Web pages are important if people visit them a lot.
- But we cannot watch everybody using the Web.
- A good surrogate for visiting pages is to assume people follow links randomly.
- Leads to random surfer model:
  - Start at a random page and follow random outlinks repeatedly, from whatever page you are at (state transition).
  - PageRank = limiting probability of being at a page (stationary distribution).
  - Solve the recursive equation: "importance of a page = its share of the importance of each of its predecessor pages" (Equivalent to the random-surfer definition of PageRank)
- The problem is formulated to compute the stationary distribution of the Web graph.

Recursive formulation

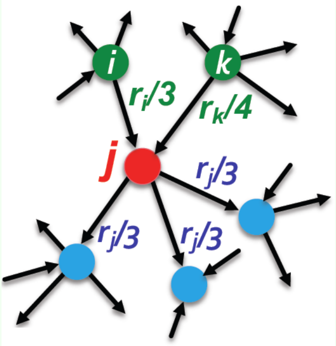


Recursive formulation



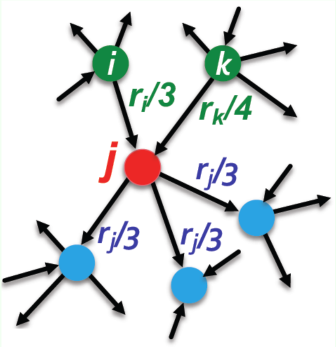
- Each link's vote is proportional to the importance of its source page.

Recursive formulation



- Each link's vote is proportional to the importance of its source page.
- If page v_j with importance r_j has n_j out-links, each link gets $\frac{r_j}{n_j}$ votes, i.e., $\frac{1}{3}r_j$.

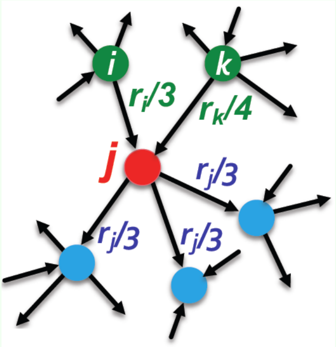
Recursive formulation



- Each link's vote is proportional to the importance of its source page.
- If page v_j with importance r_j has n_j out-links, each link gets $\frac{r_j}{n_j}$ votes, i.e., $\frac{1}{3}r_j$.
- Page v_j's own importance is the sum of the votes on its in-links.

$$r_j = \frac{1}{3}r_i + \frac{1}{4}r_k.$$

Recursive formulation

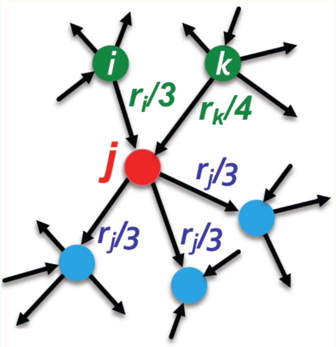


- Each link's vote is proportional to the importance of its source page.
- If page v_j with importance r_j has n_j out-links, each link gets $\frac{r_j}{n_j}$ votes, i.e., $\frac{1}{3}r_j$.
- Page v_j's own importance is the sum of the votes on its in-links.

$$r_j = \frac{1}{3}r_i + \frac{1}{4}r_k.$$

A "vote" from an important page is worth more.

Recursive formulation



- Each link's vote is proportional to the importance of its source page.
- If page v_j with importance r_j has n_j out-links, each link gets $\frac{r_j}{n_j}$ votes, i.e., $\frac{1}{3}r_j$.
- Page v_j's own importance is the sum of the votes on its in-links.

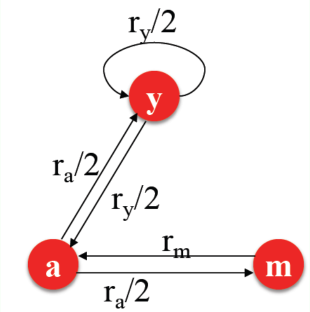
$$r_j = \frac{1}{3}r_i + \frac{1}{4}r_k.$$

A "vote" from an important page is worth more.Furthermore, a page is important if it is pointed to by other important pages, i.e.,

$$r_j = \sum_{i \to j} \frac{r_i}{n_i}.$$

Example

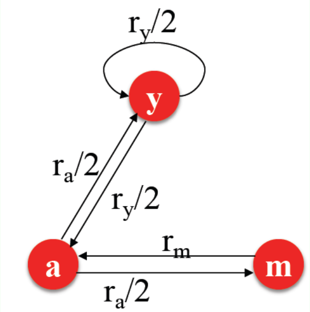
"Flow" equations



$$r_y = \frac{1}{2} r_y + \frac{1}{2} r_a$$
$$r_a = \frac{1}{2} r_y + r_m$$
$$r_m = \frac{1}{2} r_a$$

Example

"Flow" equations

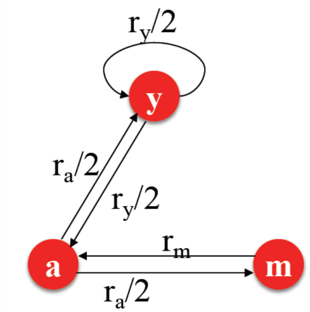


$$r_y = \frac{1}{2} r_y + \frac{1}{2} r_a$$
$$r_a = \frac{1}{2} r_y + r_m$$
$$r_m = \frac{1}{2} r_a$$

- 3 equations, 3 unknowns, no constants. No unique solution

Example

"Flow" equations

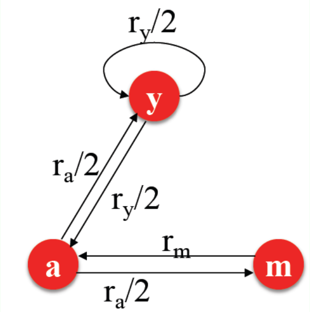


$$r_y = \frac{1}{2} r_y + \frac{1}{2} r_a$$
$$r_a = \frac{1}{2} r_y + r_m$$
$$r_m = \frac{1}{2} r_a$$

- 3 equations, 3 unknowns, no constants. No unique solution
- Additional constraint forces uniqueness $r_a + r_y + r_a = 1$. We have the solution $r_y = \frac{2}{5}, r_a = \frac{2}{5}, r_m = \frac{1}{5}$.

Example

"Flow" equations



$$r_y = \frac{1}{2} r_y + \frac{1}{2} r_a$$
$$r_a = \frac{1}{2} r_y + r_m$$
$$r_m = \frac{1}{2} r_a$$

- 3 equations, 3 unknowns, no constants. No unique solution
- Additional constraint forces uniqueness $r_a + r_y + r_a = 1$. We have the solution $r_y = \frac{2}{5}, r_a = \frac{2}{5}, r_m = \frac{1}{5}$.

Gaussian elimination method works for small examples, but we need a better method for large web-size graphs.

Matrix formulation

The Web can be modeled as a random walk.

Matrix formulation

The Web can be modeled as a random walk.

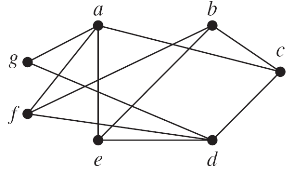
- Let vertex v_i has n_i out-links;

Matrix formulation

The Web can be modeled as a random walk.

- Let vertex v_i has n_i out-links;
- We define the probability transition matrix $M = (M_{ij}) \in R^{n \times n}$, where

$$M_{ij} = \begin{cases} \frac{1}{n_i}, & \text{if } v_i \to v_j \\ 0, & \text{otherwise.} \end{cases}$$

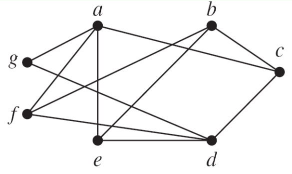


Matrix formulation

The Web can be modeled as a random walk.

- Let vertex v_i has n_i out-links;
- We define the probability transition matrix $M = (M_{ij}) \in R^{n \times n}$, where

$$M_{ij} = \begin{cases} \frac{1}{n_i}, & \text{if } v_i \to v_j \\ 0, & \text{otherwise.} \end{cases}$$



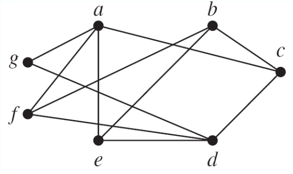
- Ranking vector r: vector with an entry per page:

Matrix formulation

The Web can be modeled as a random walk.

- Let vertex v_i has n_i out-links;
- We define the probability transition matrix $M = (M_{ij}) \in R^{n \times n}$, where

$$M_{ij} = \begin{cases} \frac{1}{n_i}, & \text{if } v_i \to v_j \\ 0, & \text{otherwise.} \end{cases}$$

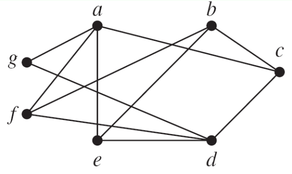


- Ranking vector r: vector with an entry per page:
  - r_i is the importance score of page v_i;
  - $\sum_{i=1}^{n} r_i = 1$.

Matrix formulation

The Web can be modeled as a random walk.

- Let vertex v_i has n_i out-links;
- We define the probability transition matrix $M = (M_{ij}) \in R^{n \times n}$, where



$$M_{ij} = \begin{cases} \frac{1}{n_i}, & \text{if } v_i \to v_j \\ 0, & \text{otherwise.} \end{cases}$$

- Ranking vector r: vector with an entry per page:
  - r_i is the importance score of page v_i;
  - $\sum_{i=1}^{n} r_i = 1$.
- The flow equations can be written as

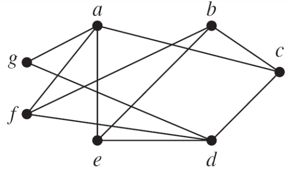
$$r^T = r^T \cdot M \text{ or } r = M^T \cdot r.$$

Matrix formulation

The Web can be modeled as a random walk.

- Let vertex v_i has n_i out-links;
- We define the probability transition matrix $M = (M_{ij}) \in R^{n\times n}$, where

$$M_{ij} = \begin{cases} \frac{1}{n_i}, & \text{if } v_i \rightarrow v_j \\ 0, & \text{otherwise.} \end{cases}$$



- Ranking vector r: vector with an entry per page:
  - r_i is the importance score of page v_i;
  - $\sum_{i=1}^{n} r_i = 1$.
- The flow equations can be written as

$$r^T = r^T \cdot M \text{ or } r = M^T \cdot r.$$

That is, r is the stationary distribution of the random walk.

Eigenvector formulation

So the ranking vector r is an eigenvector of the probability transition matrix M with the largest eigenvalue 1.

Eigenvector formulation

So the ranking vector r is an eigenvector of the probability transition matrix M with the largest eigenvalue 1.

- According to the power method

  1: Pick a starting vector $\mathbf{r}^{(0)} = (\frac{1}{n}, \cdots, \frac{1}{n})$;

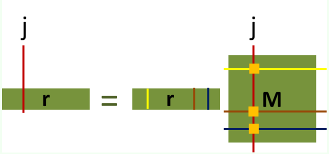
  2: For $k = 1, 2, \cdots$;

  3: Let $\mathbf{r}^{(t)} = \mathbf{r}^{(t-1)} \cdot M$;

Eigenvector formulation

So the ranking vector r is an eigenvector of the probability transition matrix M with the largest eigenvalue 1.

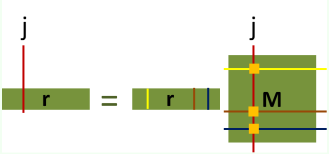
- According to the power method
  1: Pick a starting vector $\mathbf{r}^{(0)} = (\frac{1}{n}, \cdots, \frac{1}{n})$;
  2: For $k = 1, 2, \cdots$;
  3: Let $\mathbf{r}^{(t)} = \mathbf{r}^{(t-1)} \cdot M$;
- Simple and efficient algorithm
  1: set $\mathbf{r}^{(0)} = (\frac{1}{n}, \cdots, \frac{1}{n})$;
  2: For $k = 1, 2, \cdots$;
  3: For $j = 1, 2, \cdots, n$;
  4: Let $r_j^{(t)} = \sum_{i \to j} \frac{r_i}{n_i}$;



Eigenvector formulation

So the ranking vector r is an eigenvector of the probability transition matrix M with the largest eigenvalue 1.

- According to the power method
  1: Pick a starting vector $\mathbf{r}^{(0)} = (\frac{1}{n}, \cdots, \frac{1}{n})$;
  2: For $k = 1, 2, \cdots$;
  3: Let $\mathbf{r}^{(t)} = \mathbf{r}^{(t-1)} \cdot M$;
- Simple and efficient algorithm
  1: set $\mathbf{r}^{(0)} = (\frac{1}{n}, \cdots, \frac{1}{n})$;
  2: For $k = 1, 2, \cdots$;
  3: For $j = 1, 2, \cdots, n$;
  4: Let $r_j^{(t)} = \sum_{i \to j} \frac{r_i}{n_i}$;



The improved algorithm has less time and space consumption.

Outline

1. Motivation

2. Markov Chain and Random Walk

3. Page Rank
   - Problem Formulation
   - PageRank Algorithm
   - Improvements of PageRank Algorithm

Does this converge?

Does it converge to what we want?

Does this converge?

Does it converge to what we want? Are results reasonable?

Does this converge?

Does it converge to what we want? Are results reasonable?

- The answers will be true if the random walk has unique stationary distribution, i.e., is the random walk on the Web graph irreducible and aperiodic?

Does this converge?

Does it converge to what we want? Are results reasonable?

- The answers will be true if the random walk has unique stationary distribution, i.e., is the random walk on the Web graph irreducible and aperiodic?
- The Web graph may have multiple connected components.

Does this converge?

Does it converge to what we want? Are results reasonable?

- The answers will be true if the random walk has unique stationary distribution, i.e., is the random walk on the Web graph irreducible and aperiodic?
- The Web graph may have multiple connected components.
- How to solve the problem?

Does this converge?

Does it converge to what we want? Are results reasonable?

- The answers will be true if the random walk has unique stationary distribution, i.e., is the random walk on the Web graph irreducible and aperiodic?
- The Web graph may have multiple connected components.
- How to solve the problem?

We need to revise the random walk.

Does this converge?

Does it converge to what we want? Are results reasonable?

- The answers will be true if the random walk has unique stationary distribution, i.e., is the random walk on the Web graph irreducible and aperiodic?
- The Web graph may have multiple connected components.
- How to solve the problem?

We need to revise the random walk.

- The transition matrix is revised as

$$\widetilde{M} = \beta M + (1-\beta)\left[\frac{1}{n}\right]_{n\times n}$$

Does this converge?

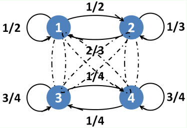
Does it converge to what we want? Are results reasonable?

- The answers will be true if the random walk has unique stationary distribution, i.e., is the random walk on the Web graph irreducible and aperiodic?
- The Web graph may have multiple connected components.
- How to solve the problem?

We need to revise the random walk.

- The transition matrix is revised as

$$\widetilde{M} = \beta M + (1-\beta)\left[\frac{1}{n}\right]_{n\times n}$$



Does this converge?

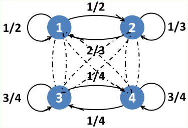
Does it converge to what we want? Are results reasonable?

- The answers will be true if the random walk has unique stationary distribution, i.e., is the random walk on the Web graph irreducible and aperiodic?
- The Web graph may have multiple connected components.
- How to solve the problem?

We need to revise the random walk.

- The transition matrix is revised as

$$\widetilde{M} = \beta M + (1-\beta)\left[\frac{1}{n}\right]_{n\times n}$$



- At each step, the random surfer has two options

Does this converge?

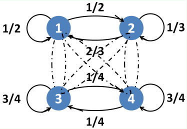
Does it converge to what we want? Are results reasonable?

- The answers will be true if the random walk has unique stationary distribution, i.e., is the random walk on the Web graph irreducible and aperiodic?
- The Web graph may have multiple connected components.
- How to solve the problem?

We need to revise the random walk.

- The transition matrix is revised as

$$\widetilde{M} = \beta M + (1-\beta)\left[\frac{1}{n}\right]_{n\times n}$$



- At each step, the random surfer has two options
  - With probability β, it follows original graph structure;

Does this converge?

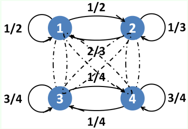
Does it converge to what we want? Are results reasonable?

- The answers will be true if the random walk has unique stationary distribution, i.e., is the random walk on the Web graph irreducible and aperiodic?
- The Web graph may have multiple connected components.
- How to solve the problem?

We need to revise the random walk.

- The transition matrix is revised as

$$\widetilde{M} = \beta M + (1-\beta)\left[\frac{1}{n}\right]_{n\times n}$$



- At each step, the random surfer has two options
  - With probability β, it follows original graph structure;
  - With probability $1-\beta$, it jumps to some random pages.

Does this converge?

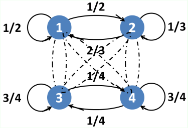
Does it converge to what we want? Are results reasonable?

- The answers will be true if the random walk has unique stationary distribution, i.e., is the random walk on the Web graph irreducible and aperiodic?
- The Web graph may have multiple connected components.
- How to solve the problem?

We need to revise the random walk.

- The transition matrix is revised as

$$\widetilde{M} = \beta M + (1-\beta)\left[\frac{1}{n}\right]_{n\times n}$$



- At each step, the random surfer has two options
  - With probability β, it follows original graph structure;
  - With probability $1-\beta$, it jumps to some random pages.
  - In practice, β will be 0.85 or 0.9.

Improvement

The original transition matrix M is sparse, while $\widetilde{M}$ is very dense.

Improvement

The original transition matrix M is sparse, while $\widetilde{M}$ is very dense.

- Given $n = 1$ billion pages, if we need 4 bytes for each entry, 2 billion entries for vectors (approx. 8GB)

Improvement

The original transition matrix M is sparse, while $\widetilde{M}$ is very dense.

- Given $n = 1$ billion pages, if we need 4 bytes for each entry, 2 billion entries for vectors (approx. 8GB)
- Thus, matrix $\widetilde{M}$ has n^2 entries, then 10^{18} is a large number!

Improvement

The original transition matrix M is sparse, while $\widetilde{M}$ is very dense.

- Given $n = 1$ billion pages, if we need 4 bytes for each entry, 2 billion entries for vectors (approx. 8GB)
- Thus, matrix $\widetilde{M}$ has n^2 entries, then 10^{18} is a large number!

We rearranging the equation $\widetilde{M} = \beta M + (1-\beta)\left[\frac{1}{n}\right]_{n\times n}$ as

$$
\begin{aligned}
r_j &= \sum_{i=1}^{n} \widetilde{M}_{ij} r_i = \sum_{i=1}^{n} (\beta M_{ij} + \frac{1-\beta}{n}) r_i \\
&= \sum_{i=1}^{n} \beta M_{ij} r_i + \sum_{i=1}^{n} \frac{1-\beta}{n} r_i = \sum_{i=1}^{n} \beta M_{ij} r_i + \frac{1-\beta}{n}
\end{aligned}
$$

Thus, we get a new matrix formulation

$$
r^T = \beta r^T \cdot M + [\frac{1-\beta}{n}]_n
$$

PageRank: the complete algorithm

- Input: Graph $G = (V, E)$ and parameter β;

PageRank: the complete algorithm

- Input: Graph $G = (V, E)$ and parameter β;
- Output: PageRank vector $\mathbf{r}^{new}$.

1: set $\mathbf{r}^{old} = (\frac{1}{n}, \cdots, \frac{1}{n})$;
2: repeat until convergence: $\sum_j |r_j^{new} - r_j^{old}| < \epsilon$;
3: $\forall j : r_j'^{new} = \sum_{i \to j} \frac{r_i^{old}}{n_i}$;
$r_j'^{new} = 0$ if in-degree of j is 0;
Now revise the random walk:
4: $\forall j : r_j^{new} = \beta r_j'^{new} + \frac{1-s}{n}$, where $s = \sum_j r_j'^{new}$;
5: $\mathbf{r}^{old} = \mathbf{r}^{new}$;

Take-home messages

- Motivation
- Markov Chain and Random Walk
- PageRank
  - Problem Formulation
  - PageRank Algorithm
  - Improvements of PageRank Algorithm