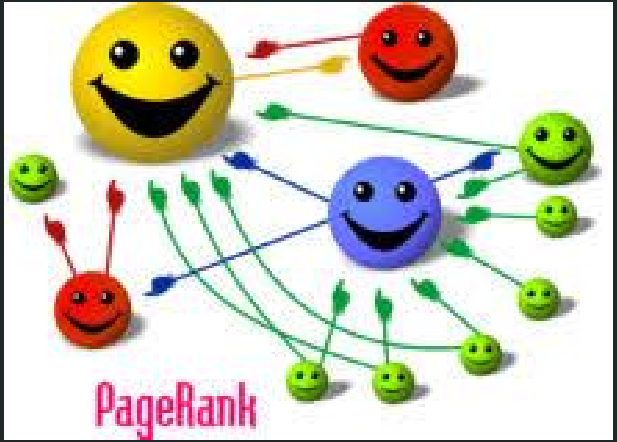
Google PageRank



Kenny Leung

October 5, 2017

PageRank

- Link analysis algorithm used by Google Search

- Rank webpages in their search engine results based on relative importance

- First algorithm used by Google to order search engine results

“PageRank relies on the uniquely democratic nature of the web by using its vast link structure as an indicator of an individual page’s value.”

Markov Chain - Simple Random Walk

- each node {A to E} represents a webpage

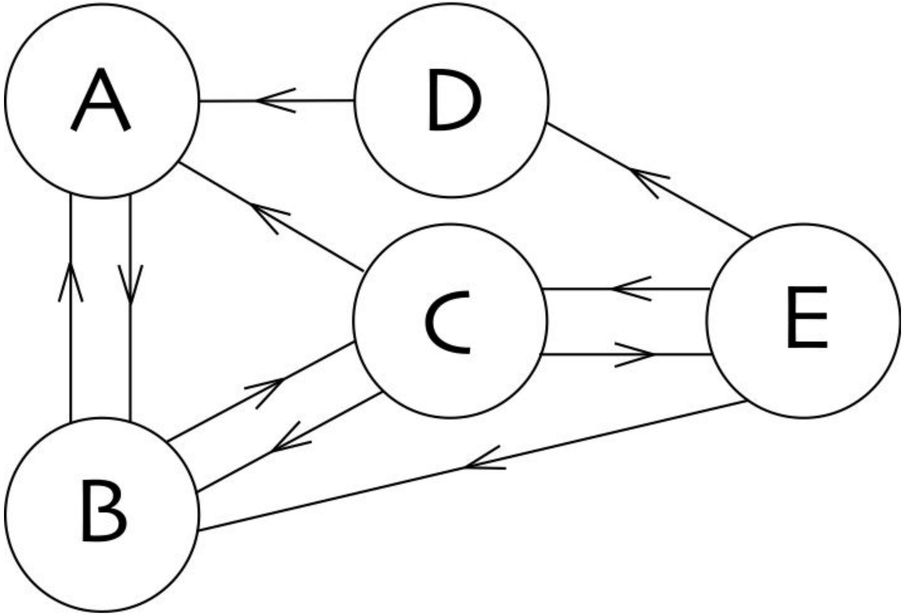
- each arrow represents a hyperlink

- each state is which page you are on

<u>Simple Random Walk</u>

- start in at a page and pick a random link

to click with equal probability



Markov Chain - Simple Random Walk

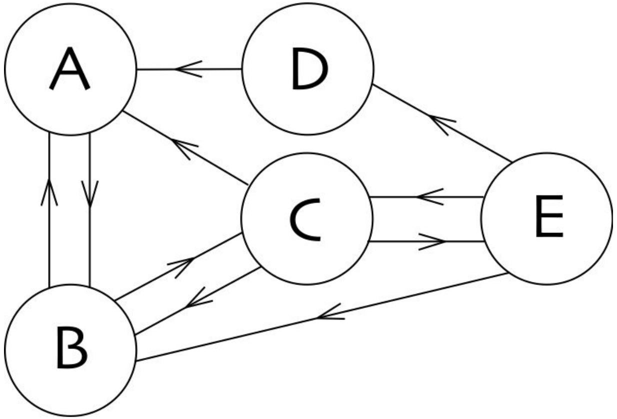
Matrix P - Markov Transition Matrix

1. sum of entries of any column of P is equal to 1

2. all entries are greater or equal to zero

$$P = \begin{array}{cc} & \begin{array}{ccccc} A & B & C & D & E \end{array} \\ & \left(\begin{array}{ccccc} 0 & \frac{1}{2} & \frac{1}{3} & 1 & 0 \\ 1 & 0 & \frac{1}{3} & 0 & \frac{1}{3} \\ 0 & \frac{1}{2} & 0 & 0 & \frac{1}{3} \\ 0 & 0 & 0 & 0 & \frac{1}{3} \\ 0 & 0 & \frac{1}{3} & 0 & 0 \end{array} \right) \begin{array}{c} A \\ B \\ C \\ D \\ E \end{array} \end{array}$$

Markov Chain Process



Markov Chain - Simple Random Walk

Transition Matrix

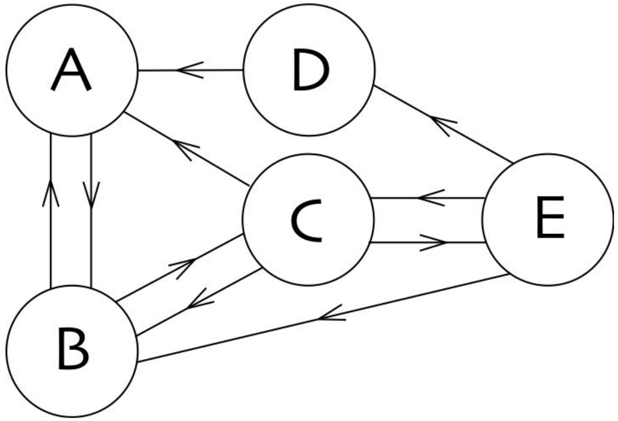
Next step (n+1):

$$P = \begin{array}{c} \begin{matrix} A & B & C & D & E \end{matrix} \\ \begin{pmatrix} 0 & \frac{1}{2} & \frac{1}{3} & 1 & 0 \\ 1 & 0 & \frac{1}{3} & 0 & \frac{1}{3} \\ 0 & \frac{1}{2} & 0 & 0 & \frac{1}{3} \\ 0 & 0 & 0 & 0 & \frac{1}{3} \\ 0 & 0 & \frac{1}{3} & 0 & 0 \end{pmatrix} \end{array} \begin{array}{c} A \\ B \\ C \\ D \\ E \end{array}$$

After two steps (n+2):

$$P^2 = \begin{array}{c} \begin{matrix} A & B & C & D & E \end{matrix} \\ \begin{pmatrix} \frac{1}{2} & \frac{1}{6} & \frac{1}{6} & 0 & \frac{11}{18} \\ 0 & \frac{2}{3} & \frac{4}{9} & 1 & \frac{1}{9} \\ \frac{1}{2} & 0 & \frac{5}{18} & 0 & \frac{1}{6} \\ 0 & 0 & \frac{1}{9} & 0 & 0 \\ 0 & \frac{1}{6} & 0 & 0 & \frac{1}{9} \end{pmatrix} \end{array} \begin{array}{c} A \\ B \\ C \\ D \\ E \end{array}$$

Markov Chain Process



Markov Chain - Simple Random Walk

After 32 steps (n+32):

$$P^{32} = \begin{array}{c} \begin{array}{ccccc} A & B & C & D & E \end{array} \\ \left(\begin{array}{ccccc} 0.293 & 0.293 & 0.293 & 0.293 & 0.293 \\ 0.390 & 0.390 & 0.390 & 0.390 & 0.390 \\ 0.220 & 0.220 & 0.220 & 0.220 & 0.220 \\ 0.024 & 0.024 & 0.024 & 0.024 & 0.024 \\ 0.073 & 0.073 & 0.073 & 0.073 & 0.073 \end{array} \right) \end{array} \begin{array}{c} \\ A \\ B \\ C \\ D \\ E \end{array}$$

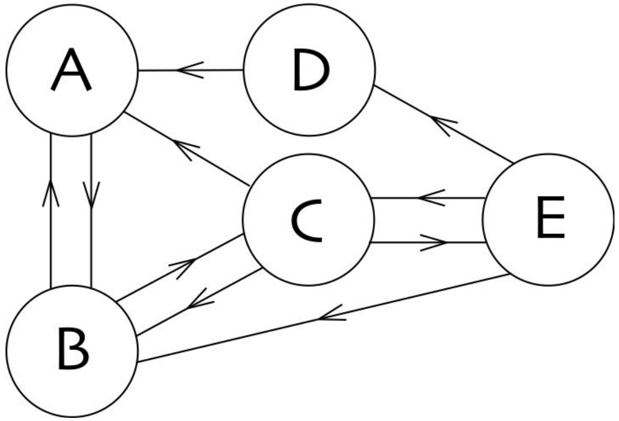
- all columns the same

- independent of starting webpage

- stationary matrix / vector

$$\pi^t = (0.293, 0.390, 0.220, 0.024, 0.073)$$

Markov Chain Process



Google Algorithm

Example is an over-simplification of a complex algorithm!

Continual Development

- reflects the taste of the surfer on the web

- robust to abuse from those who try to improve the ranking of their page

Source: http://blog.kleinproject.org/?p=280

Other Uses

- Entirely general and applies to any graph or network in any domain

Used in:

    - social and information network analysis

    - link prediction and recommendation