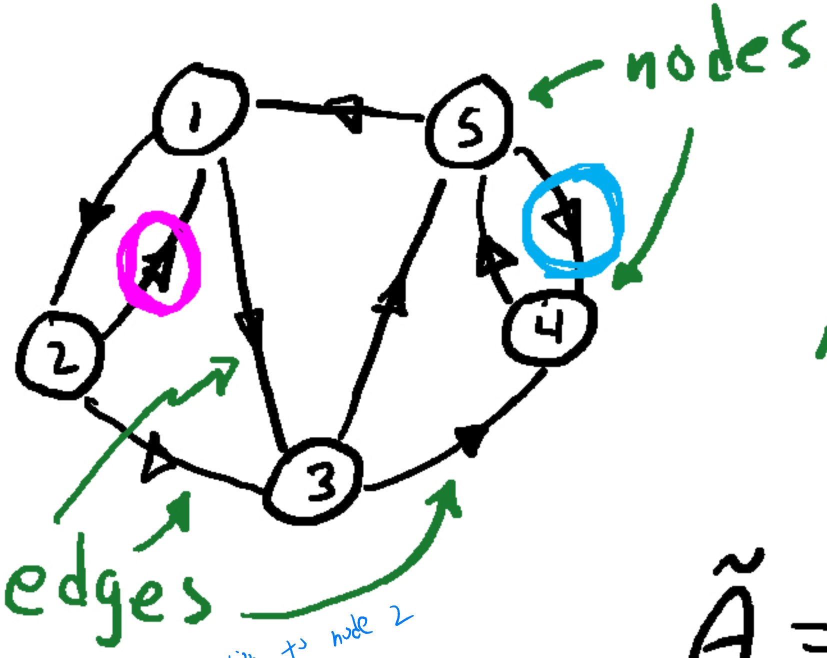
Network Graphs and the PageRank Algorithm

Objectives

- Introduce matrix representations for network graphs
- Define transition probability matrix and paths on graph
- Illustrate PageRank algorithm concepts

Matrices represent network graphs



Examples: webpages / links, cities / roads, routers / wires

(nodes above "webpages"; edges above "links")

Adjacency matrix: connection topology

$$\tilde{A} = \begin{array}{c} \text{from} \\ \begin{bmatrix} 0 & 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 1 & 0 \end{bmatrix} \end{array} \ \text{to}$$

edge from node 2 (column 2) to node 1 (row 1)

edge from node 5 (column 5) to node 4 (row 4)

25% of time, it will transition to node 2

$$A = \begin{bmatrix} 0 & 1/2 & 0 & 0 & 1/2 \\ 1/4 & 0 & 0 & 0 & 0 \\ 3/4 & 1/2 & 0 & 0 & 0 \\ 0 & 0 & 7/8 & 0 & 1/2 \\ 0 & 0 & 1/8 & 1 & 0 \end{bmatrix}$$

transition probability matrix - columns sum to 1

Transition probability matrix predicts "paths"

located at node 3

$$\underline{A} = \begin{bmatrix} 0 & 1/2 & 0 & 0 & 1/2 \\ 1/4 & 0 & 0 & 0 & 0 \\ 3/4 & 1/2 & 0 & 0 & 0 \\ 0 & 0 & 7/8 & 0 & 1/2 \\ 0 & 0 & 1/8 & 1 & 0 \end{bmatrix}$$

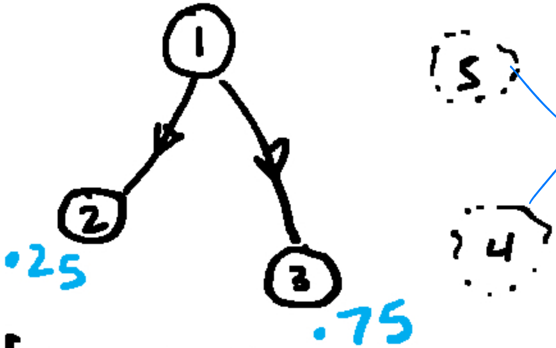
Start at node 1 : $\underline{p}_0 = [1\ 0\ 0\ 0\ 0]^T$

After 1 move: $\underline{p}_1 = \underline{A}\,\underline{p}_0 = \begin{bmatrix} 0 \\ 0.25 \\ 0.75 \\ 0 \\ 0 \end{bmatrix}$

Transition probability matrix

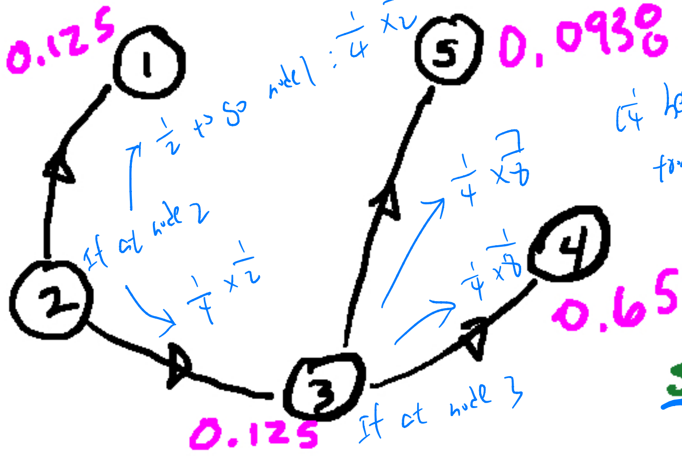
25% #2

75% #3



(No way getting to these nodes in a single step)

After 2 moves: $\underline{p}_2 = \underline{A}\,\underline{p}_1 = \begin{bmatrix} 0.125 \\ 0 \\ 0.125 \\ 0.6562 \\ 0.0938 \end{bmatrix}$

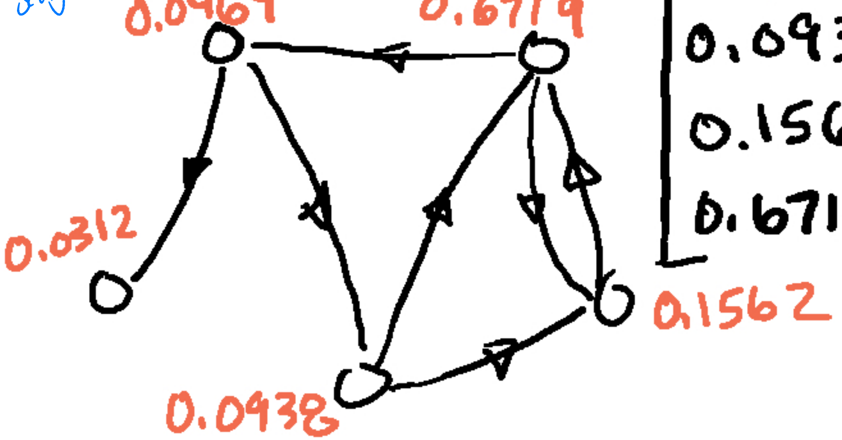


($\frac{1}{4}$ here is the probability from last step)

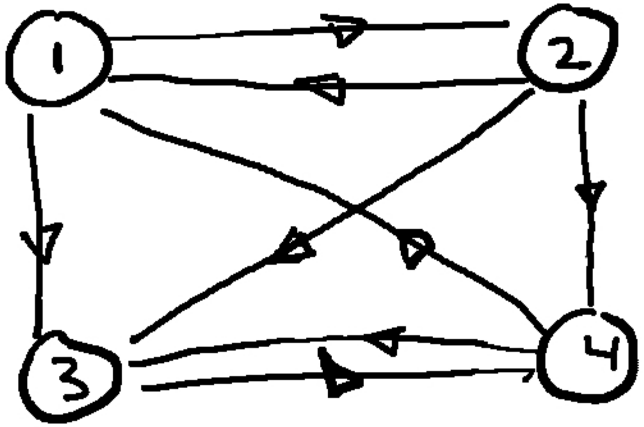
☆ Markov chain: next state depends only on current state

After 3 moves: $\underline{p}_3 = \underline{A}\,\underline{p}_2 = \begin{bmatrix} 0.0469 \\ 0.0312 \\ 0.0938 \\ 0.1562 \\ 0.6719 \end{bmatrix}$

(The probability of getting to each node from node 1 at first)



PageRank algorithm ranks web pages



where will I visit most?

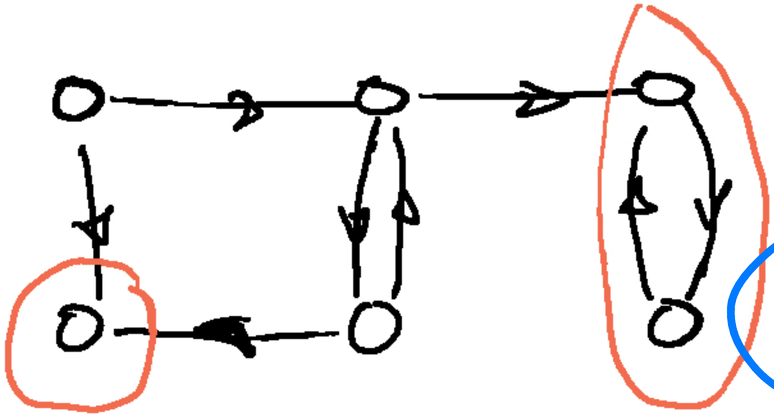
1) Adjacency matrix

$$\underline{\tilde{A}} = \begin{bmatrix} 0 & 1 & 0 & 1 \\ 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 1 \\ 0 & 1 & 1 & 0 \end{bmatrix}$$

2) Normalize columns (equal prob. outlinks)

$$\underline{A} = \begin{bmatrix} 0 & 1/3 & 0 & 1/2 \\ 1/2 & 0 & 0 & 0 \\ 1/2 & 1/3 & 0 & 1/2 \\ 0 & 1/3 & 1 & 0 \end{bmatrix}$$

3) Eliminate traps



introduce small probability to go from any node to any other node

Final transition matrix

probability of following the link

probability of random jumps

★ Transition matrix: $\underline{Q} = (1-\alpha)\underline{A} + \frac{\alpha}{N} \begin{bmatrix} 1 & 1 & \cdots & 1 \\ 1 & 1 & \cdots & 1 \\ \vdots & \vdots & & \vdots \\ 1 & 1 & \cdots & 1 \end{bmatrix}$

number of nodes

$\frac{\alpha}{N}$: random jump probability

Eigenvector of $\underline{Q}$ ranks pages

$\underline{Q}$ is irreducible (no traps) column stochastic (cols sum to 1), with nonnegative entries $\Rightarrow$

(Perron-Frobenius) Largest eval is 1, evect $\underline{p} = [p_1 \cdots p_N]^T$ satisfies $p_i > 0$, $\sum_i p_i = 1$

eigenvalue

eigenvector

$N \times 1$

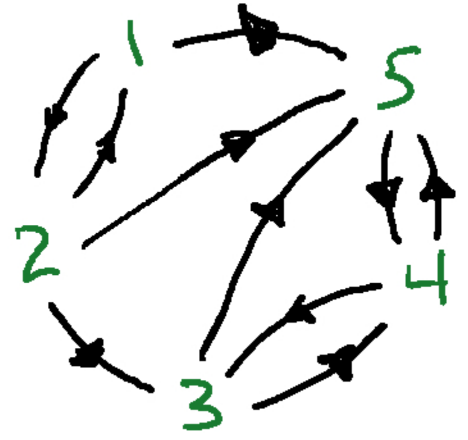
$$\underline{u} = \frac{1}{N}[1 \; 1 \cdots 1]^T \qquad \lim_{k \to \infty} \underline{Q}^k \underline{u} = \underline{p}$$

The same as the previous power iteration algorithm

Steady-State Distribution $\quad \underline{Q}\,\underline{p} = 1 \cdot \underline{p}$

$\underline{p}$ ranks importance of pages

Example:



$$\tilde{\underline{A}} = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 1 \\ 1 & 1 & 1 & 1 & 0 \end{bmatrix}$$

Normalize and $\alpha = 0.01$ give us Q (transition prob matrix)

$$\underline{Q} = \begin{bmatrix} .002 & .332 & .002 & .002 & .002 \\ .497 & .002 & .002 & .002 & .002 \\ .002 & .332 & .002 & .497 & .002 \\ .002 & .002 & .497 & .002 & .992 \\ .497 & .332 & .497 & .497 & .002 \end{bmatrix} \quad (\alpha = 0.01)$$

$$\underline{u} = \begin{bmatrix} 0.2 \\ 0.2 \\ 0.2 \\ 0.2 \\ 0.2 \end{bmatrix}, \quad \underline{Q}\underline{u} = \begin{bmatrix} .07 \\ .10 \\ .17 \\ .30 \\ .36 \end{bmatrix}, \quad \underline{Q}^2\underline{u} = \begin{bmatrix} .003 \\ .004 \\ .18 \\ .45 \\ .30 \end{bmatrix}, \cdots \underline{Q}^{10}\underline{u} = \begin{bmatrix} .003 \\ .004 \\ .22 \\ .44 \\ .33 \end{bmatrix}$$

Initial guess, assume the probability are the same

converged

main eigenvector

$$\underline{P} = \begin{array}{c} \begin{array}{ccccc} 1 & 2 & 3 & 4 & 5 \end{array} \\ \begin{bmatrix} 0.0032 & 0.0036 & 0.2211 & 0.4401 & 0.3320 \end{bmatrix} \end{array}$$