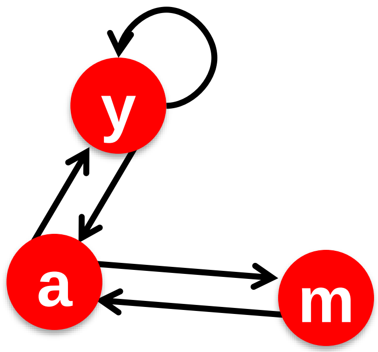
Tutorial 7: Large Graph Processing I

Li Yuan

li.yuan@u.nus.edu

National University of Singapore

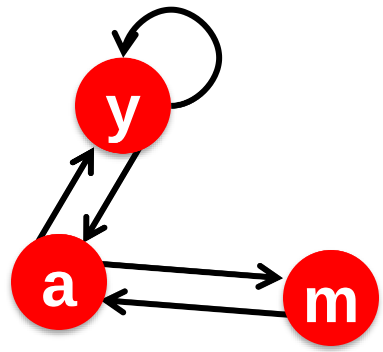
PageRank



- A page is important if it is pointed to by other important pages
- Define a rank for each page (importance score)
- Can be viewed as a random walk
  - E.g., in page a, the user has the same possibility pick page y or page m

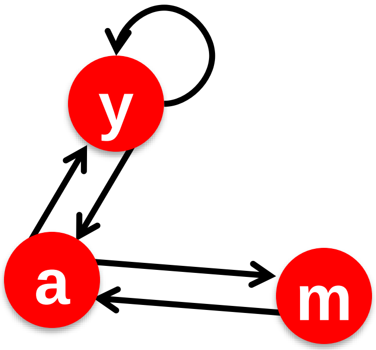
$$r_y = \frac{r_y}{2} + \frac{r_a}{2}$$
$$r_a = \frac{r_y}{2} + r_m$$
$$r_m = \frac{r_a}{2}$$

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



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

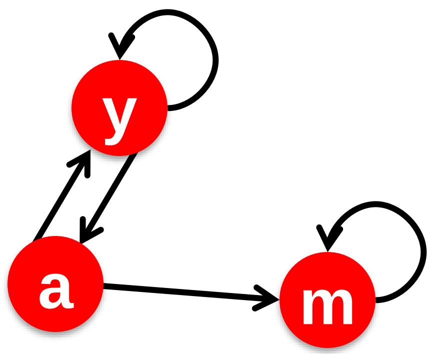
$$\begin{bmatrix} r_y \\ r_a \\ r_m \end{bmatrix} = \begin{bmatrix} \frac{1}{2} & \frac{1}{2} & 0 \\ \frac{1}{2} & 0 & 1 \\ 0 & \frac{1}{2} & 0 \end{bmatrix} \cdot \begin{bmatrix} r_y \\ r_a \\ r_m \end{bmatrix}$$



$$\begin{bmatrix} r_y \\ r_a \\ r_m \end{bmatrix} = \begin{array}{c|c} & \begin{array}{ccc} \text{y} & \text{a} & \text{m} \end{array} \\ \begin{array}{c} \text{y} \\ \text{a} \\ \text{m} \end{array} & \begin{bmatrix} \frac{1}{2} & \frac{1}{2} & 0 \\ \frac{1}{2} & 0 & 1 \\ 0 & \frac{1}{2} & 0 \end{bmatrix} \end{array} \cdot \begin{bmatrix} r_y \\ r_a \\ r_m \end{bmatrix}$$

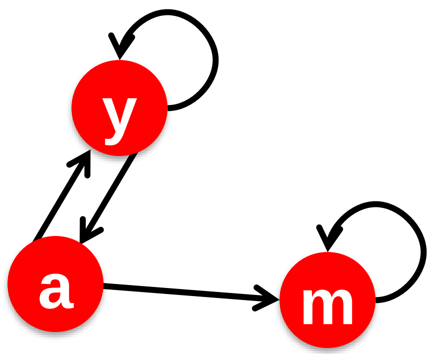
$$\boldsymbol{r} = \boldsymbol{M} \cdot \boldsymbol{r}$$

- If $i \rightarrow j$, then $\boldsymbol{M}_{ji} = \frac{1}{d_i}$ else $\boldsymbol{M}_{ji} = 0$
- $\boldsymbol{M}$ is a column stochastic matrix, columns sum to 1



- Spider traps: all out-links are within the group
  - Eventually spider traps absorb all importance

- Solution: Random teleport
  - With probability β, follow a link at random
  - With probability $1 - \beta$, jump to a random page



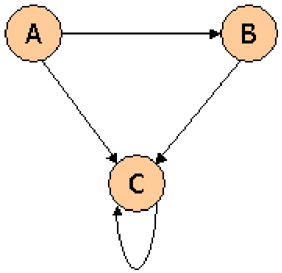
$$A = \beta M + (1-\beta)\left[\frac{1}{N}\right]_{N\times N}$$

$$\beta \times \begin{array}{c c} & \begin{array}{ccc} \mathrm{y} & \mathrm{a} & \mathrm{m} \end{array} \\ \begin{array}{c} \mathrm{y} \\ \mathrm{a} \\ \mathrm{m} \end{array} & \begin{bmatrix} \frac{1}{2} & \frac{1}{2} & 0 \\ \frac{1}{2} & 0 & 0 \\ 0 & \frac{1}{2} & 1 \end{bmatrix} \end{array} + (1-\beta) \times \begin{array}{c c} & \begin{array}{ccc} \mathrm{y} & \mathrm{a} & \mathrm{m} \end{array} \\ \begin{array}{c} \mathrm{y} \\ \mathrm{a} \\ \mathrm{m} \end{array} & \begin{bmatrix} \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \\ \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \\ \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \end{bmatrix} \end{array} = \boldsymbol{A}$$

$$\boldsymbol{r} = \boldsymbol{A} \cdot \boldsymbol{r}$$

Problem 1

Consider three Web pages with the following links:



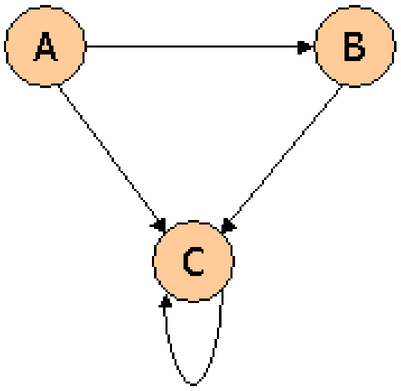
Suppose we compute PageRank with a β of 0.7, and we introduce the additional constraint that the sum of the PageRanks of the three pages must be 3, to handle the problem that otherwise any multiple of a solution will also be a solution. Compute the PageRanks a, b, and c of the three pages A, B, and C, respectively.

Solution 1

$$M = \begin{array}{c c} & \begin{array}{ccc} A & B & C \end{array} \\ \begin{array}{c} A \\ B \\ C \end{array} & \begin{bmatrix} 0 & 0 & 0 \\ \frac{1}{2} & 0 & 0 \\ \frac{1}{2} & 1 & 1 \end{bmatrix} \end{array}$$

$$N = \begin{array}{c c} & \begin{array}{ccc} A & B & C \end{array} \\ \begin{array}{c} A \\ B \\ C \end{array} & \begin{bmatrix} \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \\ \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \\ \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \end{bmatrix} \end{array}$$

$$A = \beta M + (1 - \beta) N$$
$$= \begin{bmatrix} 0.1 & 0.1 & 0.1 \\ 0.45 & 0.1 & 0.1 \\ 0.45 & 0.8 & 0.8 \end{bmatrix}$$



Solution 1

$$\begin{bmatrix} r_a \\ r_b \\ r_c \end{bmatrix} = A \cdot \begin{bmatrix} r_a \\ r_b \\ r_c \end{bmatrix}$$

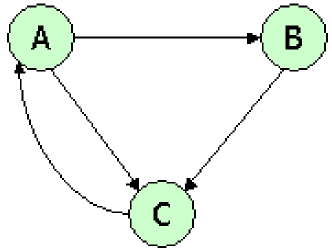
$$r_a = 0.1r_a + 0.1r_b + 0.1r_c$$
$$r_b = 0.45r_a + 0.1r_b + 0.1r_c$$
$$r_c = 0.45r_a + 0.8r_b + 0.8r_c$$

$$r_a + r_b + r_c = 3$$

$$r_a = 0.3, r_b = 0.405, r_c = 2.295$$

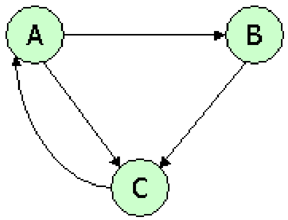
Problem 2

Consider three Web pages with the following links:



Suppose we compute PageRank with β=0.85. Write the equations for the PageRanks a, b, and c of the three pages A, B, and C, respectively.

Solution 2



$$M = \begin{bmatrix} 0 & 0 & 1 \\ \frac{1}{2} & 0 & 0 \\ \frac{1}{2} & 1 & 0 \end{bmatrix} \qquad N = \begin{bmatrix} \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \\ \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \\ \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \end{bmatrix}$$

$$A = \beta M + (1-\beta)N = \begin{bmatrix} 0.05 & 0.05 & 0.9 \\ 0.475 & 0.05 & 0.05 \\ 0.475 & 0.9 & 0.05 \end{bmatrix}$$

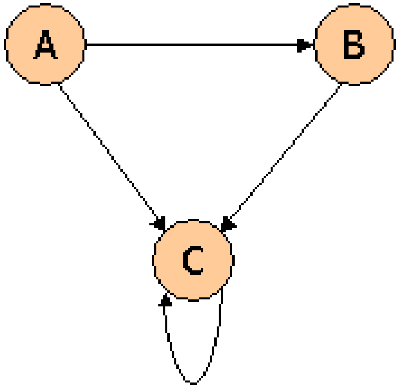
$$r_a = 0.05r_a + 0.05r_b + 0.9r_c$$
$$r_b = 0.475r_a + 0.05r_b + 0.05r_c$$
$$r_c = 0.475r_a + 0.9r_b + 0.05r_c$$

Topic-specific PageRank

- Evaluate pages not just according to their popularity, but by how close they are to a particular topic.
- Given a set S
  - S contains only pages that are relevant to the topic
  - Each page only can teleport to the page in the teleport set S

If $i \in S$, $A_{ij} = \beta M_{ij} + (1 - \beta)/|S|$
Otherwise $A_{ij} = \beta M_{ij}$

$$M = \begin{bmatrix} 0 & 0 & 0 \\ \frac{1}{2} & 0 & 0 \\ \frac{1}{2} & 1 & 1 \end{bmatrix}$$



Standard PageRank: $N = \begin{bmatrix} \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \\ \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \\ \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \end{bmatrix}$

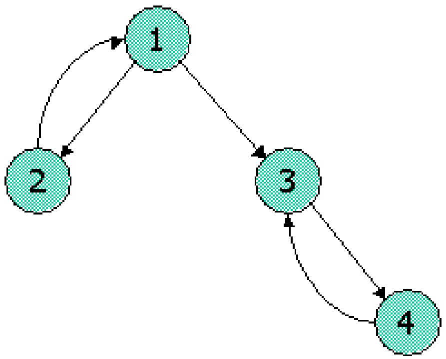
If teleport set S={A,B}

Topic-sepecific PageRank: $N = \begin{array}{c} \\ A \\ B \\ C \end{array} \begin{array}{c} \begin{array}{ccc} A & B & C \end{array} \\ \begin{bmatrix} \frac{1}{2} & \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} & \frac{1}{2} \\ 0 & 0 & 0 \end{bmatrix} \end{array}$

$$A = \beta M + (1 - \beta) N$$

Problem 3

Consider the following link topology.

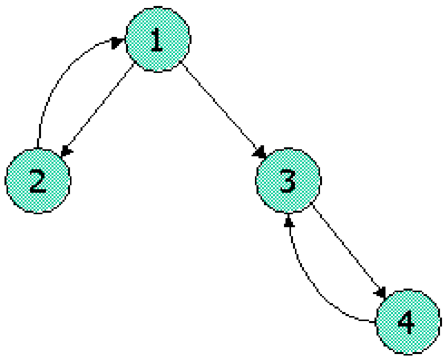


Compute the Topic-Specific PageRank for the following link topology. Assume that pages selected for the teleport set are nodes 1 and 2 and that in the teleport set, the weight assigned for node 1 is twice that of node 2. Assume further that the teleport probability, (1 - beta), is 0.3.

Solution 3

$$M = \begin{bmatrix} 0 & 1 & 0 & 0 \\ \frac{1}{2} & 0 & 0 & 0 \\ \frac{1}{2} & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{bmatrix} \qquad N_e = \begin{bmatrix} \frac{1}{2} & \frac{1}{2} & \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} & \frac{1}{2} & \frac{1}{2} \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

$$N = \begin{bmatrix} \frac{2}{3} & \frac{2}{3} & \frac{2}{3} & \frac{2}{3} \\ \frac{1}{3} & \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$



Solution 3

$$A = \beta M + (1-\beta)N = \begin{bmatrix} 0.2 & 0.9 & 0.2 & 0.2 \\ 0.45 & 0.1 & 0.1 & 0.1 \\ 0.35 & 0 & 0 & 0.7 \\ 0 & 0 & 0.7 & 0 \end{bmatrix}$$

$$\begin{bmatrix} tr_1 \\ tr_2 \\ tr_3 \\ tr_4 \end{bmatrix} = A \cdot \begin{bmatrix} tr_1 \\ tr_2 \\ tr_3 \\ tr_4 \end{bmatrix}, tr_1 + tr_2 + tr_3 + tr_4 = 1$$

$$tr_1 = 0.3576, tr_2 = 0.2252, tr_3 = 0.2454, tr_4 = 0.1718$$

Acknowledgement

Thanks to Li Qinbin for making these slides.

liqinbin@u.nus.edu