Seoul National University

M1522.001400-001 Introduction to Data Mining

Homework 3: Link Analysis (Chapter 5)

2017-18538 Hwang SunYoung

1.

(a)

Let page i has d_i out-links. If i->j, then M_{ji} = $1/d_i$ else M_{ji} = 0.

Let matrix M be the column stochastic adjacency matrix of the graph in figure 1.

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

(b)

The google matrix A is

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

$$A = 0.8 \times \begin{bmatrix} 0 & 0 & \frac{1}{3} & 0 & 0 \\ \frac{1}{2} & 0 & 0 & 1 & \frac{1}{2} \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & \frac{1}{3} & 0 & 0 \\ \frac{1}{2} & 0 & \frac{1}{3} & 0 & \frac{1}{2} \end{bmatrix} + (1-0.8) \times \begin{bmatrix} \frac{1}{5} & \frac{1}{5} & \frac{1}{5} & \frac{1}{5} & \frac{1}{5} \\ \frac{1}{5} & \frac{1}{5} & \frac{1}{5} & \frac{1}{5} & \frac{1}{5} \\ \frac{1}{5} & \frac{1}{5} & \frac{1}{5} & \frac{1}{5} & \frac{1}{5} \\ \frac{1}{5} & \frac{1}{5} & \frac{1}{5} & \frac{1}{5} & \frac{1}{5} \\ \frac{1}{5} & \frac{1}{5} & \frac{1}{5} & \frac{1}{5} & \frac{1}{5} \end{bmatrix}$$

$$= \begin{bmatrix} \frac{1}{25} & \frac{1}{25} & \frac{23}{75} & \frac{1}{25} & \frac{1}{25} \\ \frac{11}{25} & \frac{1}{25} & \frac{1}{25} & \frac{21}{25} & \frac{11}{25} \\ \frac{1}{25} & \frac{21}{25} & \frac{1}{25} & \frac{1}{25} & \frac{1}{25} \\ \frac{1}{25} & \frac{1}{25} & \frac{23}{75} & \frac{1}{25} & \frac{1}{25} \\ \frac{11}{25} & \frac{1}{25} & \frac{23}{75} & \frac{1}{25} & \frac{11}{25} \end{bmatrix}$$

Rank vector r after the third iteration is

$$r = \begin{bmatrix} \frac{239}{1875} \\ \frac{153}{625} \\ \frac{91}{375} \\ \frac{239}{1875} \\ \frac{161}{625} \end{bmatrix}$$

(c) convergence value of each entry in the rank vector r is

$$\begin{bmatrix} 0.11 \\ 0.27 \\ 0.26 \\ 0.11 \\ 0.25 \end{bmatrix}$$

2.

$n \times n$ Boolean matrix to store value. (0 and 1 elements only)

Let f that a fraction. (0≤f≤1)

f is a fraction of 1's in the matrix.

① we need n^2 space to represent the matrix by the bits themselves.

② we need $2fn^2\lceil log_2 n \rceil$ space to represent the matrix by listing the positions of the 1's as pair of integers.

When we use ② to save the space, it must be $n^2 \geq 2fn^2\lceil log_2 n \rceil$.

$$f \leq \frac{1}{2\lceil log_2 n \rceil}$$

The matrix must be sparse below f.

3.

Let rank vector r of graph consists of a clique of n nodes and a single additional node that is the successor of each of the n nodes in the clique. At dead-end node, we can "always teleport" other node(include itself) in same probability. Length is n+1.

$$r = \begin{bmatrix} p \\ p \\ \vdots \\ p \\ p \\ q \end{bmatrix}$$

Sum of element np+q=1

Let matrix M be the n+1 × n+1 column stochastic adjacency matrix of graph consists of a clique of n nodes and a single additional node that is the successor of each of the n nodes in the clique.

$$\mathrm{M} = \begin{bmatrix} 0 & \frac{1}{n} & \cdots & \frac{1}{n} & \frac{1}{n+1} \\ \frac{1}{n} & 0 & \cdots & \frac{1}{n} & \frac{1}{n+1} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \frac{1}{n} & \frac{1}{n} & \cdots & 0 & \frac{1}{n+1} \\ \frac{1}{n} & \frac{1}{n} & \frac{1}{n} & \frac{1}{n} & \frac{1}{n+1} \end{bmatrix}$$

Pagerank equation is $\mathrm{r} = \beta \mathrm{Mr} + (1-\beta)\left[\frac{1}{N}\right]_N$

$$\begin{bmatrix} p \\ p \\ \vdots \\ p \\ p \\ q \end{bmatrix} = \beta \begin{bmatrix} 0 & \frac{1}{n} & \cdots & \frac{1}{n} & \frac{1}{n+1} \\ \frac{1}{n} & 0 & \cdots & \frac{1}{n} & \frac{1}{n+1} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \frac{1}{n} & \frac{1}{n} & \cdots & 0 & \frac{1}{n+1} \\ \frac{1}{n} & \frac{1}{n} & \frac{1}{n} & \frac{1}{n} & \frac{1}{n+1} \end{bmatrix} \begin{bmatrix} p \\ p \\ \vdots \\ p \\ p \\ q \end{bmatrix} + (1-\beta)\left[\frac{1}{n+1}\right]_{n+1} = \begin{bmatrix} \beta\left(\frac{p(n-1)}{n} + \frac{q}{n+1}\right) + (1-\beta)\frac{1}{n+1} \\ \vdots \\ \vdots \\ \vdots \\ \beta\left(\frac{p(n-1)}{n} + \frac{q}{n+1}\right) + (1-\beta)\frac{1}{n+1} \\ \beta\left(p + \frac{q}{n+1}\right) + (1-\beta)\frac{1}{n+1} \end{bmatrix}$$

In this equation we can get q= $\beta\left(p + \frac{q}{n+1}\right) + (1-\beta)\frac{1}{n+1}$

Put np+q=1 into $\beta\left(p + \frac{q}{n+1}\right) + (1-\beta)\frac{1}{n+1}$, then $q = \frac{\beta + \mathrm{n}}{n^2 + n + \beta}$

Put $q = \frac{\beta + \mathrm{n}}{n^2 + n + \beta}$ into np+q=1, then $p = \frac{\mathrm{n}}{n^2 + n + \beta}$

Pagerank vector r is

$$r = \begin{bmatrix} \frac{\mathrm{n}}{n^2+n+\beta} \\ \frac{\mathrm{n}}{n^2+n+\beta} \\ \vdots \\ \frac{\mathrm{n}}{n^2+n+\beta} \\ \frac{\mathrm{n}}{n^2+n+\beta} \\ \frac{\beta+\mathrm{n}}{n^2+n+\beta} \end{bmatrix}$$

4.

(a)

Node	Hub score
A	9
B	7
C	2
D	4

(b)

Node	Updated authority score
A	2
B	9
C	11
D	16

(c)

A is link matrix for the graph of Figure 3 and A^t is its transpose.

$$A = \begin{bmatrix} 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \qquad A^t = \begin{bmatrix} 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 \\ 1 & 1 & 0 & 0 \end{bmatrix}$$

Start iteration according to iterative update rule,

1st : $a_1 = A^t h_0$, $h_1 = Aa_1 = AA^t h_0$

2nd: $a_2 = A^t h_1 = A^t A A^t h_0$, $h_2 = A a_2 = A A^t A A^t h_0 = (AA^t)^2 h_0$

3rd: $a_3 = A^t h_2 = A^t (AA^t)^2 h_0$, $h_3 = A a_3 = A A^t (AA^t)^2 h_0 = (AA^t)^3 h_0$

kth: $a_k = A^t (AA^t)^{k-1} h_0$, $h_k = (AA^t)^k h_0$

general update equation for the k-step hub-authority computation is

$a_k = A^t (AA^t)^{k-1} h_0$, $h_k = (AA^t)^k h_0$