Name: Zhang Jingbin
Student NO:MB954741
CISC7019 Web Mining
Due: 09/20/2019

Assignment #1

Problem 1. (5.1.2) : Compute the PageRank of each page in Fig. 1, assuming $\beta = 0.8$.

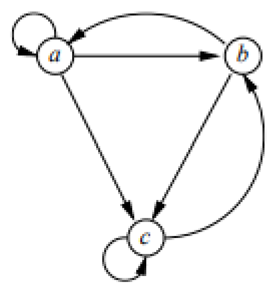

Figure 1: An example graph for exercises

Solution:

$$PageRank\ equation: \begin{cases} A_{ji} = \beta M_{ji} + \dfrac{(1-\beta)}{N} \\ r = A \cdot r \end{cases}$$

from Fig.1,the adjacency matrix $M = \begin{bmatrix} \frac{1}{3} & \frac{1}{2} & 0 \\ \frac{1}{3} & 0 & \frac{1}{2} \\ \frac{1}{3} & \frac{1}{2} & \frac{1}{2} \end{bmatrix}$,the Random Teleports rate:$\beta = 0.8$.

So the random telepot formulation as follows.

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

Using the iteration technique and initial $r^0 = \begin{bmatrix} \frac{1}{3} \\ \frac{1}{3} \\ \frac{1}{3} \end{bmatrix}$

$r^1 = 0.8 \times \begin{bmatrix} \frac{1}{3} & \frac{1}{2} & 0 \\ \frac{1}{3} & 0 & \frac{1}{2} \\ \frac{1}{3} & \frac{1}{2} & \frac{1}{2} \end{bmatrix} \cdot r^0 + 0.2 \times \begin{bmatrix} \frac{1}{3} \\ \frac{1}{3} \\ \frac{1}{3} \end{bmatrix}$

$r^1 = \begin{bmatrix} \frac{13}{45} \\ \frac{13}{45} \\ \frac{19}{45} \end{bmatrix}$

follow the random telepot formulation 1, we can get

$$r = \begin{bmatrix} \frac{1}{3} \\ \frac{1}{3} \\ \frac{1}{3} \end{bmatrix}, \begin{bmatrix} \frac{13}{45} \\ \frac{13}{45} \\ \frac{19}{45} \end{bmatrix}, \begin{bmatrix} \frac{7}{27} \\ \frac{211}{675} \\ \frac{289}{675} \end{bmatrix} \cdots \cdots, \begin{bmatrix} \frac{7}{27} \\ \frac{25}{81} \\ \frac{35}{81} \end{bmatrix}$$

According to formula 1, we can get

$$r = \left(\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} - 0.8 \times \begin{bmatrix} \frac{1}{3} & \frac{1}{2} & 0 \\ \frac{1}{3} & 0 & \frac{1}{2} \\ \frac{1}{3} & \frac{1}{2} & \frac{1}{2} \end{bmatrix} \right)^{-1} \cdot \begin{bmatrix} \frac{1}{15} \\ \frac{1}{15} \\ \frac{1}{15} \end{bmatrix}$$

$$r = \begin{bmatrix} \frac{55}{27} & \frac{10}{9} & \frac{20}{27} \\ \frac{100}{81} & \frac{55}{27} & \frac{110}{81} \\ \frac{140}{81} & \frac{50}{27} & \frac{235}{81} \end{bmatrix} \cdot \begin{bmatrix} \frac{1}{15} \\ \frac{1}{15} \\ \frac{1}{15} \end{bmatrix}$$

$$r = \begin{bmatrix} \frac{7}{27} \\ \frac{25}{81} \\ \frac{35}{81} \end{bmatrix}$$

Problem 2. (5.1.6) : Suppose we recursively eliminate dead ends from the graph, solve the remaining graph, and estimate the PageRank for the dead-end pages as described in Section 5.1.4. Suppose the graph is a chain of dead ends, headed by a node with a self-loop, as suggested in Fig. 2. What would be the PageRank assigned to each of the nodes?

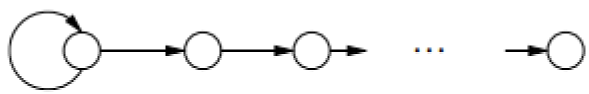

Figure 2: A chain of dead ends

Solution:

After estimate dead-end pages recursively, we can get an new graph in figure 3.

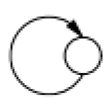

Figure 3: After estimate dead-end

Suppose the Page 1 has an initail value 1.
So we can get this equation:
$r_1 = 1$
The remaining Page PageRank values are:
$r_2 = \frac{1}{2} \times r_1 = \frac{1}{2}$
$r_i = r_{i-1} = \frac{1}{2}, where\ i \geq 3$
Becasue the sum of the PageRanks exceed 1, we need to normalizate.

$$R = \begin{cases} r_1 = \frac{1}{1+\frac{1}{2}(n-1)} = \frac{2}{n+1} \\ r_i = \frac{\frac{1}{2}}{1+\frac{1}{2}(n-1)} = \frac{1}{n+1} & , where\ i \geq 2 \end{cases}$$

Problem 3. (5.4.1) : In Section 5.4.2 we analyzed the spam farm of Fig. 4, where every supporting page links back to the target page. Repeat the analysis for a spam farm in which:

(a) Each supporting page links to itself instead of to the target page.

(b) Each supporting page links nowhere.

(c) Each supporting page links both to itself and to the target page.

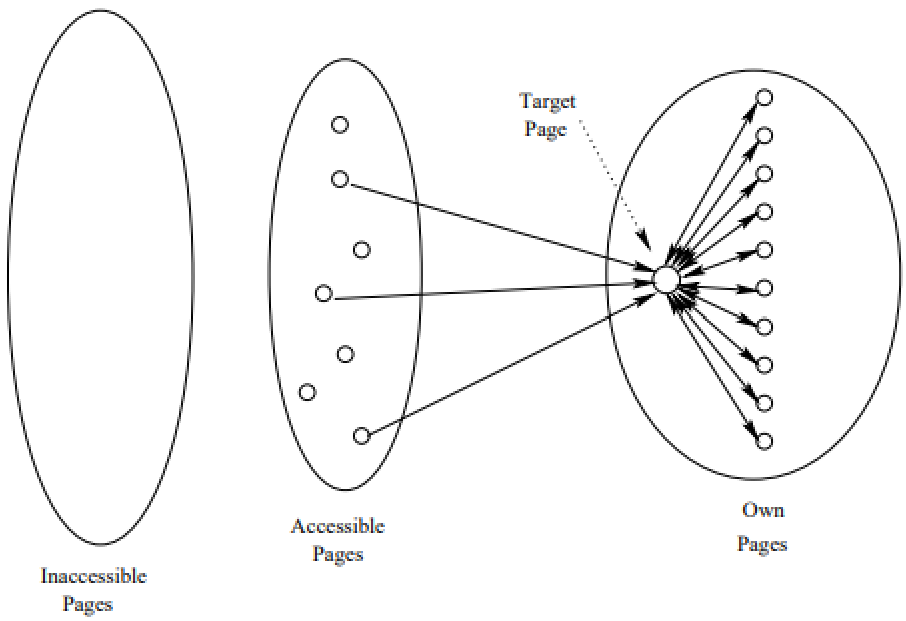

Figure 4: The Web from the point of view of the link spammer

Solution:
Suppose the number of supporting pages is M, the number of the whole pages is N.β means a random telepot rate, the rank contributed by accessible pages are X, the page rank of target page is y, and the rank of each supporting page is z.

(a) Each supporting page links to itself instead of to the target page.
$z = \beta(\frac{y}{M} + z) + \frac{(1-\beta)}{N}$
$y = X + \frac{(1-\beta)}{N}$
Because y do not have M, spam farm's function is invalid.

(b) Each supporting page links nowhere.
if each supporting page links nowhere, supporting page will become dead end page.
So
$y = X + \frac{(1-\beta)}{N}$
$z = \frac{y}{M} + \frac{(1-\beta)}{N}$
Because y do not have M, spam farm's function is invalid.

(c) Each supporting page links both to itself and to the target page.
$z = \beta(\frac{y}{M} + \frac{z}{2}) + \frac{(1-\beta)}{N}$

$z = \frac{2y\beta}{(1-\beta)M} + \frac{2-2\beta}{(2-\beta)N}$
$y = X + \beta M z + \frac{(1-\beta)}{N} = X + \beta M(\frac{2y\beta}{(1-\beta)M} + \frac{2-2\beta}{(2-\beta)N}) + \frac{(1-\beta)}{N}$
where $\frac{(1-\beta)}{N} \approx 0$
So $y \approx X + \frac{y\beta^2}{(2-\beta)} + \frac{(1-\beta)\beta M}{(2-\beta)N} = \frac{2-\beta}{(2+\beta)(1-\beta)}X + \frac{\beta M}{(2+\beta)N}$
Suppose $\beta = 0.8$, therefore $\frac{2-\beta}{(2+\beta)(1-\beta)} = 2.14$ and $\frac{\beta}{(2+\beta)} = 0.2857$, which means If M is large enough, y can be large enough too.

Problem 4. (5.5.1) : Compute the hubbiness and authority of each of the nodes in our original Web graph of Fig. 5

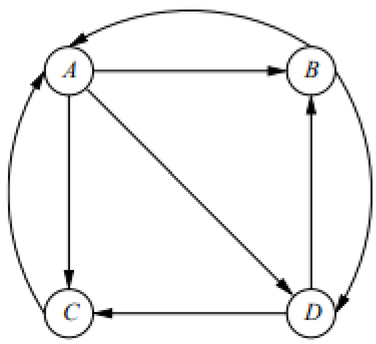

Figure 5: A hypothetical example of the Web

Solution:
Adjacency matrix

$$A = \begin{bmatrix} 0 & 1 & 1 & 1 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 \end{bmatrix}, A^T = \begin{bmatrix} 0 & 1 & 1 & 0 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 \\ 1 & 1 & 0 & 0 \end{bmatrix}, A \cdot A^T = \begin{bmatrix} 3 & 1 & 0 & 2 \\ 1 & 2 & 1 & 0 \\ 0 & 1 & 1 & 0 \\ 2 & 0 & 0 & 2 \end{bmatrix}, \ A^T \cdot A = \begin{bmatrix} 2 & 0 & 0 & 1 \\ 0 & 2 & 2 & 1 \\ 0 & 2 & 2 & 1 \\ 1 & 1 & 1 & 2 \end{bmatrix}$$

Initialize:$a_j^{(0)} = \frac{1}{\sqrt{N}} = \frac{1}{2}, h_j^{(0)} = \frac{1}{\sqrt{N}} = \frac{1}{2}$
$\because h = \lambda A \cdot a, \ a = \mu A^T \cdot h$
where $\lambda = \frac{1}{\sum h_i}, \ \mu = \frac{1}{\sum a_i}$
$\therefore a^* = \mu\lambda A^T \cdot A \cdot a^*$
$\therefore h^* = \mu\lambda A \cdot A^T \cdot h^*$
a^* is the principal eigenvector of $\mu\lambda A^T \cdot A$
h^* is the principal eigenvector of $\mu\lambda A \cdot A^T$

$$a = \begin{bmatrix} 0.5 \\ 0.5 \\ 0.5 \\ 0.5 \end{bmatrix}, \begin{bmatrix} 0.125 \\ 0.25 \\ 0.25 \\ 0.375 \end{bmatrix}, \begin{bmatrix} 0.111 \\ 0.278 \\ 0.278 \\ 0.333 \end{bmatrix}, \ldots, \begin{bmatrix} 0.079 \\ 0.298 \\ 0.298 \\ 0.324 \end{bmatrix}$$
$$h = \begin{bmatrix} 0.5 \\ 0.5 \\ 0.5 \\ 0.5 \end{bmatrix}, \begin{bmatrix} 0.375 \\ 0.25 \\ 0.125 \\ 0.25 \end{bmatrix}, \begin{bmatrix} 0.389 \\ 0.222 \\ 0.167 \\ 0.222 \end{bmatrix}, \ldots, \begin{bmatrix} 0.410 \\ 0.180 \\ 0.144 \\ 0.266 \end{bmatrix}$$