COMP 6730 Advanced Database Systems Homework 3

1 Problem 1:Exercise 5.1.2

Exercise 5.1.2: Compute the PageRank of each page in Fig. 5.7, assuming $\beta = 0.8$.

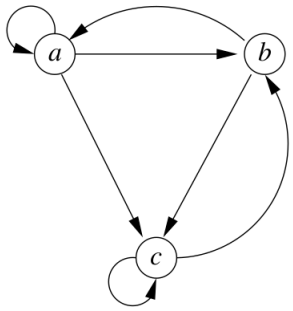

Figure 5.7: An example graph for exercises

Answer:
According to the PageRank equation:

$$r_j = \sum_{i \to j} \beta \frac{r_i}{d_i} + (1 - \beta)\frac{1}{N}$$

First, try to build the stochastic adjacency matrix M using the following relationship about their importance r_i

$$r_a = r_b/2 + r_a/3$$

$$r_b = r_a/3 + r_c/2$$

$$r_c = r_a/3 + r_b/2 + r_c/2$$

The we can get:

$$M = \begin{pmatrix} 1/3 & 1/2 & 0 \\ 1/3 & 0 & 1/2 \\ 1/3 & 1/2 & 1/2 \end{pmatrix} \tag{1}$$

Add their result and get the final matrix as follows:

$$r_j = \sum_{i \to j} \begin{pmatrix} 1/3 & 7/15 & 1/15 \\ 1/3 & 1/15 & 7/15 \\ 1/3 & 7/15 & 7/15 \end{pmatrix}$$

Using the equation $r = M * r$, do the iteration multiple times, and then we can get the following result:
1st iteration: r = (0.33 0.33 0.33)
2nd iteration: r = (0.29 0.29 0.42)
3rd iteration: r = (0.26 0.31 0.43)
...
100th iteration: r = (0.2592 0.3086 0.4320)

After that it's quite stable, so the PageRank of a is **0.26**, b is **0.31**, c is **0.43**

2 Problem 2: Exercise 5.3.1

Exercise 5.3.1: Compute the topic-sensitive PageRank for the graph of Fig.5.15, assuming the teleport set is:
(a) A only.
(b) A and C.

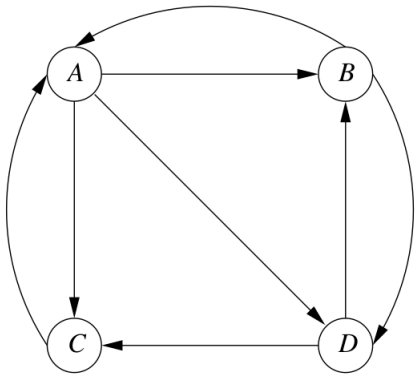

Figure 5.15: Repeat of example Web graph

Answer:
Here for the topic-sensitive, we should consider the topic when using the random jump, which means that only jumping to the node that related with the topic, the general form is

$$v' = \beta M v + (1 - \beta) e_s / |S|$$

As the first problem shows, first try to build the matrix M, which is:

$$M = \begin{pmatrix} 0 & 1/2 & 1 & 0 \\ 1/3 & 0 & 0 & 1/2 \\ 1/3 & 0 & 0 & 1/2 \\ 1/3 & 1/2 & 0 & 0 \end{pmatrix}$$

(a) For A only, the vector should be (1 0 0 0), as A is the only case, use the equation above with $\beta = 0.8$, try to start as the vector $v = 1 \quad 0 \quad 0 \quad 0$, since A relates to the topic. NOTE that the initial distribution has no effect on the limit or the final result, doing multiple iterations and the final value reaches:

$$\text{v} = (0.4286\ 0.1905\ 0.1905\ 0.1905)$$

so the topic-sensitive PageRank for a, b, c, d is **0.43**, **0.19**, **0.19**, **0.19** respectively.

(b) For A and C, the biased vector should be (1/2 0 1/2 0), other functions are the same, using the same method we can get:

$$\text{v} = (0.3857\ 0.1714\ 0.2714\ 0.1714)$$

so the topic-sensitive PageRank for a, b, c, d is **0.39**, **0.17**, **0.27**, **0.17** respectively.

For some details, please refer to the code below.

Appendix: Code

ex2.m

```
M = [0 1/2 1 0; 1/3 0 0 1/2; 1/3 0 0 1/2; 1/3 1/2 0 0];
v = [1 0 0 0]'; % initial value no effects for final result
%cons = [1 0 0 0]'; % for case a), has different bias
cons = [0.5 0 0.5 0]'; % for case b),

for i = 1:100
    v = 0.8 * M * v + 0.2 * cons;
end
% ends, v is the final vector for (a,b,c,d)
```