Big Data Analytics

ESSEC

Olga Klopp

Home work 5 Solution: Link Analysis

1. (**Exercise 5.1.1 MMDS book)** Compute the PageRank of each page in Fig. 5.7, assuming no taxation.

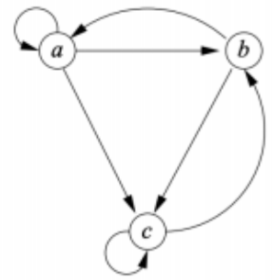

Figure 5.7: An example graph for exercises

Solution:

We have the following transition matrix:

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

Let $r = (r_1, r_2, r_3)$ be the vector of ranks. We have the following system:

$$\begin{cases} r_1 &= r_1/3 + r_2/2 \\ r_2 &= r_1/3 + r_3/2 \\ r_3 &= r_1/3 + r_2/2 + r_3/2 \end{cases}$$

and $r_1 + r_2 + r_3 = 1$. We get the following solution: $(3/13, 4/13, 6/13)$.

2. (**Exercise 5.1.2 MMDS book)** Compute the PageRank of each page in Fig. 5.7, assuming taxation with $\beta = 0.8$

Solution:

Assuming taxation we have the following equation:

$$u_{n+1} = 0.8Mu_n + 0.2e/3 \tag{1}$$

where $u_n = [u_n(1), u_n(2), u_n(3)]^T$ and $e = [1, 1, 1]^T$. This matrix equation is equivalent to the following system:

$$\begin{cases} u_{n+1}(1) &= \frac{4}{15}u_n(1) + \frac{2}{5}u_n(2) + \frac{1}{15} \\ u_{n+1}(2) &= \frac{4}{15}u_n(1) + \frac{2}{5}u_n(3) + \frac{1}{15} \\ u_{n+1}(3) &= \frac{4}{15}v + \frac{2}{5}u_n(2) + \frac{2}{5}u_n(3) + \frac{1}{15} \end{cases}$$

Starting with $u_0 = (1/3, 1/3, 1/3)$ after few iterations we get: $(13/45, 13/45, 19/45)$, $(0.259, 0.312, 0.428)$, $(0.261, 0.307, 0.432)$, $(0.259, 0.309, 0.431)$, $(0.259, 0.3085, 0.432)\ldots$.

Let $A = 0.8M$ and $b = 0.2e/3$. Applying (1) recursively we get: $u_1 = Au_0 + b$, $u_2 = A^2u_0 + Ab + b$, $u_2 = A^3u_0 + A^2b + Ab + b$, ...

$$u_n = A^n u_0 + \sum_{k=0}^{n-1} A^k b.$$

Note that $A^n = (0.8)^n M^n$ and $(0.8)^n \to 0$ when $n \to \infty$. On the other hand, $M^n u_0$ converges to the stationary state. This implies that $A^n u_0 \to 0$. We also have that $\sum_{k=0}^{\infty} A^k = (1 - A)^{-1}$. Thus our solution is given by $u_\infty = (1 - A)^{-1}\, b$. We have that

$$(1 - A)^{-1} = \begin{bmatrix} 55/27 & 10/9 & 20/27 \\ 100/81 & 55/27 & 110/81 \\ 140/81 & 50/27 & 235/81 \end{bmatrix}$$

and $u_\infty = [7/27, 25/81, 35/81] \approx [0.259, 0.3086, 0.432]$.

3. (**Exercise 5.2.1 MMDS book)** Suppose we wish to store an $n \times n$ Boolean matrix (0 and 1 elements only). We could represent it by the bits themselves, or we could represent the matrix by listing the positions of the 1's as pairs of integers, each integer requiring $\log_2(n)$ bits. The former is suitable for dense matrices; the latter is suitable for sparse matrices. How sparse must the matrix be (i.e., what fraction of the elements should be 1's) for the sparse representation to save space?

   Solution:

   An $n \times n$ boolean matrix takes n^2 bits of space. A sparse matrix is a matrix in which most of the elements are zero. So, we represent the 1's as a pair of integers, each of which takes $\log_2(n)$ bits. If our matrix has N non-zero elements we need $N \log_2(n)$ bits. The sparse representation saves space if $2N \log_2(n) < n^2$, that is, the fraction of of 1's $N/n^2 < (\log_2(n^2))^{-1}$.

4. (**Exercise 5.2.2 MMDS book)** Using the method of Ex 3, represent the transition matrices of the graph from Figure 5.7.

   Solution:

   $$\{(1, 1); (1, 2); (2, 1); (2, 3); (3, 1); (3, 2); (3, 3)\}$$