EXPERIMENT – II

Anirban Mukherjee

16BIT0200

GOOGLE'S MECHANISM FOR RANKING WEB PAGES

Aim:

We aim at understanding the mathematics behind the most successful search engine i.e. Google by using a simplified version of the Random Surfer Algorithm.

Problem Statements:

Finding the most relevant web pages using the PageRank Algorithm.

Matlab Commands:

Eigs – Largest eigen values and eigen vectors of a matrix.

Sum – Sum of array elements.

Description of Physical Experiment:

- The PageRank theory holds that an imaginary surfer who is randomly clicking on links will eventually stop clicking. The

PageRank value of a page reflects the chance that the random surfer will land on that page by clicking on a link. The probability, at any step, that the person will continue is a damping factor d.

- For a web of pages A, B, C, D,… the PageRank of A is given by:

  $PR(A) = \frac{1-d}{N} + d\left(\frac{PR(B)}{L(B)} + \frac{PR(C)}{L(C)} + \frac{PR(C)}{L(C)} \dots\right)$

  $PR(.)$ denotes PageRank and $L(.)$ denotes number of outbound links.

- Generally the damping factor will be set around 0.85.

- Links from a page to itself, or multiple outbound links from one single page to another single page, are ignored.

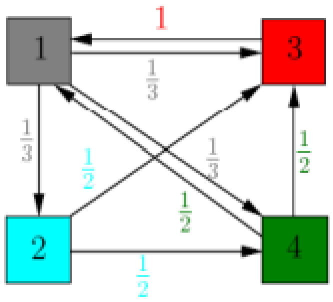
Connection to Mathematics:

Let us consider a web consisting of n-pages and let us denote the importance (PageRank) of i-th page by xi where $1 \leq i \leq n$. In the simplified random surfer model, with dampening factor $(d) = 1$, the importance of i^{th} page is calculated by the equation:

$$x_i = \sum_j \frac{x_j}{n_j}$$

when j-th page contains n_j outbound links, one of which links to the i^{th} page.

Question:



Arrange the pages in order of their importance.

<u>Solution:</u>

In equation form:

$$x_1 = 0x_1 + 0x_2 + 1x_3 + \frac{1}{2}x_4$$

$$x_2 = \frac{1}{3}x_1 + 0x_2 + 0x_3 + 0x_4$$

$$x_3 = \frac{1}{3}x_1 + \frac{1}{4}x_2 + 0x_3 + \frac{1}{2}x_4$$

$$x_4 = \frac{1}{3}x_1 + \frac{1}{2}x_2 + 0x_3 + 0x_4$$

Converting this in matrix form:

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

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

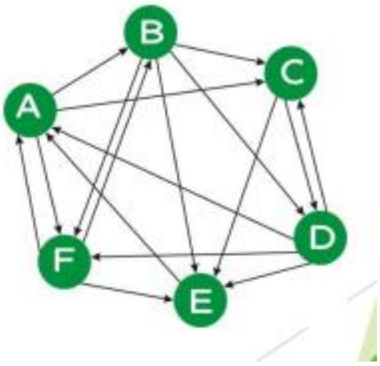
Code:

```
>> A=[0 0 1 1/2; 1/3 0 0 0; 1/3 1/2 0 1/2; 1/3 1/2 0 0];

>> [V,D]=eigs(A);

>>  u=V(:,1);

>> x=u/sum(u)


x =


    0.3871

    0.1290

    0.2903

    0.1935
```

Output:

From this, we can clearly say Page 1>Page3>Page4>Page2 is the order of significance.

Question:



Arrange the pages in order of their importance.

Solution:

In equation form:

$$x_1 = 0x_1 + 0x_2 + 0x_3 + \frac{1}{4}x_4 + 1x_5 + \frac{1}{3}x_6$$

$$x_2 = \frac{1}{3}x_1 + 0x_2 + 0x_3 + 0x_4 + 0x_5 + \frac{1}{3}x_6$$

$$x_3 = \frac{1}{3}x_1 + \frac{1}{4}x_2 + 0x_3 + \frac{1}{4}x_4 + 0x_5 + 0x_6$$

$$x_4 = 0x_1 + \frac{1}{4}x_2 + \frac{1}{2}x_3 + 0x_4 + 0x_5 + 0x_6$$

$$x_5 = 0x_1 + \frac{1}{4}x_2 + \frac{1}{2}x_3 + \frac{1}{4}x_4 + 0x_5 + \frac{1}{3}x_6$$

$$x_6 = \frac{1}{3}x_1 + \frac{1}{4}x_2 + 0x_3 + \frac{1}{4}x_4 + 0x_5 + 0x_6$$

In matrix form:

$$A = \begin{bmatrix} 0 & 0 & 0 & \frac{1}{4} & 1 & \frac{1}{3} \\ \frac{1}{3} & 0 & 0 & 0 & 0 & \frac{1}{3} \\ \frac{1}{3} & \frac{1}{4} & 0 & \frac{1}{4} & 0 & 0 \\ 0 & \frac{1}{4} & \frac{1}{2} & 0 & 0 & 0 \\ 0 & \frac{1}{4} & \frac{1}{2} & \frac{1}{4} & 0 & \frac{1}{3} \\ \frac{1}{3} & \frac{1}{4} & 0 & \frac{1}{4} & 0 & 0 \end{bmatrix}$$

```
>> A=[0 0 0 1/4 1 1/3; 1/3 0 0 0 0 1/3; 1/3 1/4 0 1/4 0 0; 0 1/4 1/2 0 0 0;
0 1/4 1/2 1/4 0 1/3; 1/3 1/4 0 1/4 0 0];

>> [V,D]=eigs(A);

>>  u=V(:,1);

>> x=u/sum(u)


x =


    0.2696

    0.1383

    0.1502

    0.1097

    0.1871

    0.1502
```

Output:

We can clearly see Page1>Page5>Page3=Page6>Page2>Page4 is the order of significance.

My Work:

The page-ranking algorithm can be used in other IT sectors as well. We as IT students, let's say we make a new social networking site or any other web application!

We can easily use this algorithm to develop a feed which will track what the user sees, reads, and watches from various sources. We can then rank them accordingly and display that to the user's feed according to relevance.

Let's consider a user visits a newspaper application we develop, we can track what content he likes to read amongst National, Business, Sports, Entertainment and various other topics. We can also track which place he's accessing it from, like Tamil Nadu. We then rank his interests and present it to the user in a format he likes to read and finds more relevant.