Jacob John 16BCE2205
7th August 2017

Experiment #02 - GOOGLE'S MECHANISM FOR RANKING WEB-PAGES

Aim: Pages are ranked using pagerank algorithm. Understand the mathematics behind the most successful search engine, i.e. Google by using a simplified version of the *Random Surfer Algorithm.* This algorithm is devised by Larry Page and Sergey Brin.

Problem statement: To used eigenvalues and eigenvectors to represent and calculate pageranks for webpages similar to the Random Surfer Algorithm.

Mathematical background: *Pagerank* is a way of measuring the importance of website pages according to Google. Pagerank is a function that assigns a real number to each page in the web.

PR(E) denotes pagerank of "E", i.e., the numerical weight assigned to the page "E". *PR(.)* denotes the PageRank of and *L(.)* denotes number of outbound links.

Outbound links are meant to take you elsewhere and are going to direct you to another specific webpage or website.

Damping factor is the probability at any step that the person will continue in the same page is a damping factor D. Generally, the damping factor is considered to be around 0.85 despite based on results by various studies.

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

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

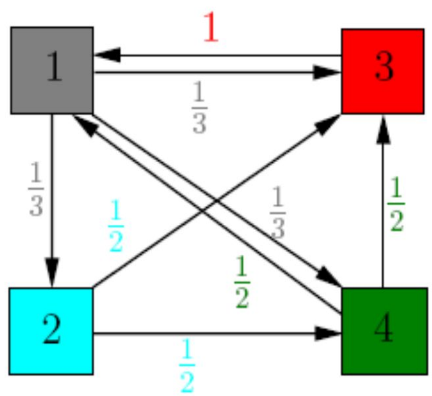
Where, N is the number of documents in the collection.

Transition matrices or *stochastic matrices* (randomly determined; having a random probability distribution or pattern that may be analyzed statistically but may not be predicted precisely) and are matrices used to display the transition of a Markov chain. Each of its entries is a non-negative real number representing a probability. For a transition matrix, the maximum eigenvalue will be 1.

Right stochastic matrix is a real square matrix with each row summing to 1. While, *left stochastic matrix* is a real square matrix with each column summing to 1. Furthermore, *doubly square matrix* is a square matrix of nonnegative real numbers with each row and column summing to one.

Questions:

Q: Consider a tiny web consisting of four pages only as depicted in the below diagram which shows the "links" from one web page to another web page.



A real web may consist of millions of web pages.

A: Now by the formula, the equations we have using the diagram are,

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

The above system can be written as $X = AX$. X is the eigenvector corresponding to the eigenvalue 1. Since, the largest absolute value of the eigenvalues of the transition matrix is 1. A will be given as:

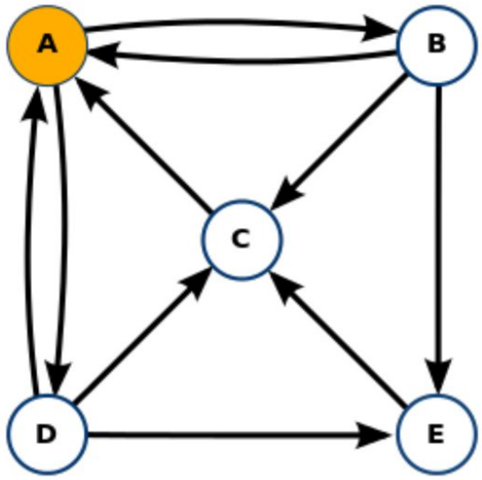
```
>> A = [0 0 1 1/2;1/3 0 0 0;1/3 1/2 0 1/2;1/3 1/2 0 0]

A =

         0         0         1    0.5000
    0.3333         0         0         0
    0.3333    0.5000         0    0.5000
    0.3333    0.5000         0         0
```

The eigenvector corresponding the eigenvalue 1, provides a non-trivial solution to the system and is called *importance vector* or *pagerank vector.*

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

A =

         0         0    1.0000    0.5000
    0.3333         0         0         0
    0.3333    0.5000         0    0.5000
    0.3333    0.5000         0         0

>> [V,D] = eig(A)

V =

   0.7210 + 0.0000i   0.7552 + 0.0000i   0.7552 + 0.0000i   0.5065 + 0.0000i
   0.2403 + 0.0000i  -0.3037 - 0.3461i  -0.3037 + 0.3461i  -0.6057 + 0.0000i
   0.5408 + 0.0000i  -0.0932 + 0.2747i  -0.0932 - 0.2747i  -0.3815 + 0.0000i
   0.3605 + 0.0000i  -0.3584 + 0.0714i  -0.3584 - 0.0714i   0.4807 + 0.0000i


D =

   1.0000 + 0.0000i   0.0000 + 0.0000i   0.0000 + 0.0000i   0.0000 + 0.0000i
   0.0000 + 0.0000i  -0.3606 + 0.4110i   0.0000 + 0.0000i   0.0000 + 0.0000i
   0.0000 + 0.0000i   0.0000 + 0.0000i  -0.3606 - 0.4110i   0.0000 + 0.0000i
   0.0000 + 0.0000i   0.0000 + 0.0000i   0.0000 + 0.0000i  -0.2788 + 0.0000i

>> u = V(:,1)

u =

    0.7210
    0.2403
    0.5408
    0.3605

>> x = u/sum(u)

x =

    0.3871
    0.1290
    0.2903
    0.1935
```

Which suggests that the Page #1 is the most important page and Page #2 is the least important page in the given web.

Q: Write the transition matrix for the webs shown below and arrange the pages in the order of their importance.



A: Now by the formula, the equations we have using the diagrams are:

$$x_1 = 0x_1 + \tfrac{1}{3}x_2 + 1x_3 + \tfrac{1}{3}x_4 + 0x_5 ...(1)$$
$$x_2 = \tfrac{1}{2}x_1 + 0x_2 + 0x_3 + 0x_4 + 0x_5 ...(2)$$
$$x_3 = 0x_1 + \tfrac{1}{3}x_2 + 0x_3 + \tfrac{1}{3}x_4 + \tfrac{1}{3}x_5 ...(3)$$
$$x_4 = \tfrac{1}{2}x_1 + 0x_2 + 0x_3 + 0x_4 + 0x_5 ...(2)$$
$$x_5 = 0x_1 + \tfrac{1}{3}x_2 + 0x_3 + \tfrac{1}{3}x_4 + 0x_5 ...(5)$$

The above system can be written as $X = AX$. X is the eigenvector corresponding to the eigenvalue 1. Since, the largest absolute value of the eigenvalues of the transition matrix is 1. A will be given as:

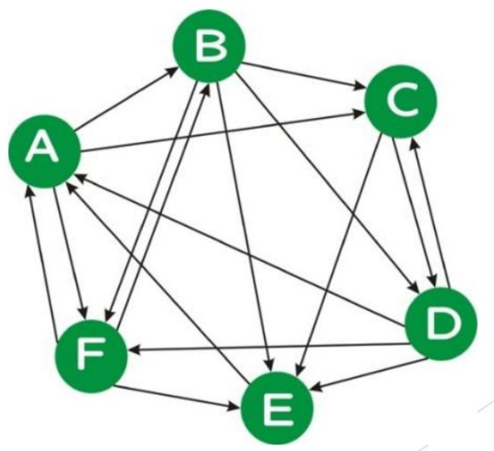
```
>> A = [0 1/3 1 1/3 0;1/2 0 0 0 0;0 1/3 0 1/3 1/3;1/2 0 0 0 0;0 1/3 0 1/3 0];
>> A

A =

         0    0.3333    1.0000    0.3333         0
    0.5000         0         0         0         0
         0    0.3333         0    0.3333    0.3333
    0.5000         0         0         0         0
         0    0.3333         0    0.3333         0
```

The eigenvector corresponding the eigenvalue 1, provides a non-trivial solution to the system and is called *importance vector* or *pagerank vector.*

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

A =

         0    0.3333    1.0000    0.3333         0
    0.5000         0         0         0         0
         0    0.3333         0    0.3333    0.3333
    0.5000         0         0         0         0
         0    0.3333         0    0.3333         0


>> [V,D] = eig(A)

V =

  Columns 1 through 4

   0.6967 + 0.0000i  -0.2575 - 0.3619i  -0.2575 + 0.3619i  -0.3599 + 0.0000i
   0.3818 + 0.0000i  -0.1883 + 0.3650i  -0.1883 - 0.3650i   0.4320 + 0.0000i
   0.3810 + 0.0000i   0.3633 - 0.2774i   0.3633 + 0.2774i  -0.1381 + 0.0000i
   0.3818 + 0.0000i  -0.1883 + 0.3650i  -0.1883 - 0.3650i   0.4320 + 0.0000i
   0.2790 + 0.0000i   0.5064 + 0.0000i   0.5064 + 0.0000i  -0.6915 + 0.0000i

  Column 5

  -0.0000 + 0.0000i
   0.7071 + 0.0000i
  -0.0000 + 0.0000i
  -0.7071 + 0.0000i
   0.0000 + 0.0000i


D =

  Columns 1 through 4

   0.9123 + 0.0000i   0.0000 + 0.0000i   0.0000 + 0.0000i   0.0000 + 0.0000i
   0.0000 + 0.0000i  -0.2479 + 0.4806i   0.0000 + 0.0000i   0.0000 + 0.0000i
   0.0000 + 0.0000i   0.0000 + 0.0000i  -0.2479 - 0.4806i   0.0000 + 0.0000i
   0.0000 + 0.0000i   0.0000 + 0.0000i   0.0000 + 0.0000i  -0.4165 + 0.0000i
   0.0000 + 0.0000i   0.0000 + 0.0000i   0.0000 + 0.0000i   0.0000 + 0.0000i

  Column 5

   0.0000 + 0.0000i
   0.0000 + 0.0000i
   0.0000 + 0.0000i
   0.0000 + 0.0000i
   0.0000 + 0.0000i

>> u = V(:,1)

u =

    0.6967
    0.3818
    0.3810
```

```
    0.3818
    0.2790

>> x = u/sum(u)

x =

    0.3286
    0.1801
    0.1797
    0.1801
    0.1316
```

Q: Write the transition matrix for the webs shown below and arrange the pages in the order of their importance.



A: Now by the formula, the equations we have using the diagrams are:

$x_1 = 0x_1 + \frac{1}{3}x_2 + 1x_3 + \frac{1}{3}x_4 + 0x_5 ...(1)$
$x_2 = \frac{1}{2}x_1 + 0x_2 + 0x_3 + 0x_4 + 0x_5 ...(2)$
$x_3 = 0x_1 + \frac{1}{3}x_2 + 0x_3 + \frac{1}{3}x_4 + \frac{1}{3}x_5 ...(3)$
$x_4 = \frac{1}{2}x_1 + 0x_2 + 0x_3 + 0x_4 + 0x_5 ...(2)$
$x_5 = 0x_1 + \frac{1}{3}x_2 + 0x_3 + \frac{1}{3}x_4 + 0x_5 ...(5)$

The above system can be written as $X = AX$. X is the eigenvector corresponding to the eigenvalue 1. Since, the largest absolute value of the eigenvalues of the transition matrix is 1. A will be given as:

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

```
A =

         0    0.3333    1.0000    0.3333         0
    0.5000         0         0         0         0
         0    0.3333         0    0.3333    0.3333
    0.5000         0         0         0         0
         0    0.3333         0    0.3333         0
```

The eigenvector corresponding the eigenvalue 1, provides a non-trivial solution to the system and is called *importance vector* or *pagerank vector*.

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

A =

         0    0.3333    1.0000    0.3333         0
    0.5000         0         0         0         0
         0    0.3333         0    0.3333    0.3333
    0.5000         0         0         0         0
         0    0.3333         0    0.3333         0

>> [V,D] = eig(A)

V =

   0.6967 + 0.0000i  -0.2575 - 0.3619i  -0.2575 + 0.3619i  -0.3599 + 0.0000i   0.0000 +
0.0000i
   0.3818 + 0.0000i  -0.1883 + 0.3650i  -0.1883 - 0.3650i   0.4320 + 0.0000i   0.7071 +
0.0000i
   0.3810 + 0.0000i   0.3633 - 0.2774i   0.3633 + 0.2774i  -0.1381 + 0.0000i  -0.0000 +
0.0000i
   0.3818 + 0.0000i  -0.1883 + 0.3650i  -0.1883 - 0.3650i   0.4320 + 0.0000i  -0.7071 +
0.0000i
   0.2790 + 0.0000i   0.5064 + 0.0000i   0.5064 + 0.0000i  -0.6915 + 0.0000i  -0.0000 +
0.0000i


D =

   0.9123 + 0.0000i   0.0000 + 0.0000i   0.0000 + 0.0000i   0.0000 + 0.0000i   0.0000 +
0.0000i
   0.0000 + 0.0000i  -0.2479 + 0.4806i   0.0000 + 0.0000i   0.0000 + 0.0000i   0.0000 +
0.0000i
   0.0000 + 0.0000i   0.0000 + 0.0000i  -0.2479 - 0.4806i   0.0000 + 0.0000i   0.0000 +
0.0000i
   0.0000 + 0.0000i   0.0000 + 0.0000i   0.0000 + 0.0000i  -0.4165 + 0.0000i   0.0000 +
0.0000i
   0.0000 + 0.0000i   0.0000 + 0.0000i   0.0000 + 0.0000i   0.0000 + 0.0000i   0.0000 +
0.0000i

>> u = V(:,1)
```

```
u =

    0.6967
    0.3818
    0.3810
    0.3818
    0.2790
```

Attestation:

DATE __/__/__

EXPERIMENT - 1 : Stress distribution in a tower bridge

$$A = \begin{bmatrix} 2 & 0 & 4 \\ 0 & 6 & 0 \\ 4 & 0 & 2 \end{bmatrix} \quad \text{eig } A = \begin{bmatrix} -2 \\ 6 \\ 6 \end{bmatrix}$$

EXPERIMENT - 2 : Google's page ranking algorithm

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

$$x = \begin{matrix} 0.3286 \\ 0.1801 \\ 0.1797 \\ 0.1316 \end{matrix}$$

11/9/17