Assignment 1: Rank Your Page

2018 Fall EECS205002 Linear Algebra

Due: 2018/10/17

Google's most famous algorithm is PageRank, which sorts the web pages that contain the searched keywords. The idea behind PageRank is Markov chain, which works as follows. Let's image there is a person, such as Roger, surfing the web. Roger first starts from a random page, and follows the outgoing links in the page randomly. After going to a page, Roger starts this process repeatedly. If such web surfing goes long enough, what is the probability of each page that Roger is reading? The higher probability of the page that Roger is reading, the higher rank the page will be.

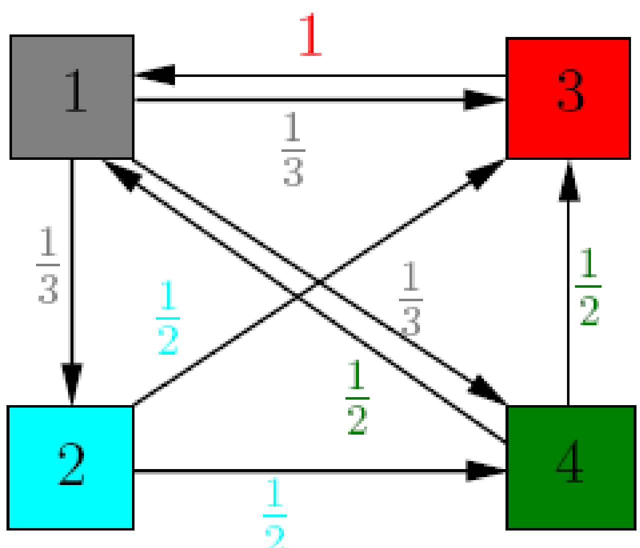

Figure 1: Example of PageRank

Figure shows an example. [1] Suppose there are four pages, and the numbers on the arrow are the probability to click the link. For example, page 1 has three outgoing links, so each link has 1/3 probability to be chosen. The transition of

[1](`http://pi.math.cornell.edu/~mec/Winter2009/RalucaRemus/Lecture3/lecture3.html`)

webpages can be described as an matrix, called transition matrix. The one of above example is

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

Assume Roger was at page 4 in the beginning. The probability of which page is reading can be expressed as a vector

$$v_0 = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 1 \end{bmatrix}.$$

After the first move, the probability becomes

$$v_1 = Av_0 = \begin{bmatrix} 0 & 0 & 1 & 1/2 \\ 1/3 & 0 & 0 & 0 \\ 1/3 & 1/2 & 0 & 1/2 \\ 1/3 & 1/2 & 0 & 0 \end{bmatrix} \begin{bmatrix} 0 \\ 0 \\ 0 \\ 1 \end{bmatrix} = \begin{bmatrix} 1/2 \\ 0 \\ 1/2 \\ 0 \end{bmatrix}.$$

After the second move, the probability becomes

$$v_2 = Av_1 = \begin{bmatrix} 0 & 0 & 1 & 1/2 \\ 1/3 & 0 & 0 & 0 \\ 1/3 & 1/2 & 0 & 1/2 \\ 1/3 & 1/2 & 0 & 0 \end{bmatrix} \begin{bmatrix} 1/2 \\ 0 \\ 1/2 \\ 0 \end{bmatrix} = \begin{bmatrix} 1/2 \\ 1/6 \\ 1/6 \\ 1/6 \end{bmatrix}.$$

If we continue this process, we may get a $v_i = v_{i-1}$ for some the ith moves. In that case, v_i is called the steady state, and the numbers in v_i is used as the rank of the pages.

Note that the summation of numbers in v_0, v_1 and v_2 is 1, and they are all between 0 and 1. (why?)

1 Assignment in Python

1. Design a simple, connected, directed, weighted graph with N nodes and M links, where $10 \leq N \leq 15$ and $20 \leq M \leq 30$. The weights must satisfy the probability condition, which means they are between $(0, 1]$ and the summation of weights from a node is 1. Write its probability matrix A.

2. Give 10 different initial vectors v of length N, whose elements must satisfy the probability condition. Compute $A^k v$ for $k = 1, 2, \ldots$ until $\|A^{i+1}v - A^i v\| \leq 10^{-5}$. Compare their final results to see if they are all the same (with difference less than 10^{-5}).

3. After each matrix-vector multiplication, check if the elements of $A^k v$ satisfy the probability condition. If not, how do you fix it?

2 Submission

1. Write a report in PDF file that includes (a) your graph, (b) your matrix A, (c) 10 different initial vectors, their plots of convergence, and a discussion about their differences, and (d) methods to fix the last problem.

2. Python code of the second problem.

3. Zip them and submit to iLMS system