CSE 6240 - Spring 2015

Web Search & Text Mining

Homework 2

01/19/2015

Due: 01/25/2015 23:59

Page Rank

Description

Implement the Page Rank algorithm on a directed graph. Here we assume alpha = 0.85, and use the equation:

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

Where d means alpha, N is the number of pages and L number of outbound links.

You may use any programming language you like, but C++, Java or Python is strongly recommended. Please make sure your program uses standard IO stream and uses input and output format below (sample files are provided).

Input Format

In the first line, 2 integers *n* and *m*, separated by space, denote the number of nodes and edges.

In the following *m* lines, each line has 2 integers *x, y*, which shows an edge *x -> y*. Node ids (unique) range from 1 to *n*. Duplicate edges may occur and should be counted multiple times.

Output Format

N lines, line *i* has the node I's PR value.

Sample Input

4 6

2 1

2 3

3 1

4 1

4 2

4 3

Sample Output

0.12687

0.04812

0.06857

0.0375

Deliverable

The deliverable should contain three files, please put all the files in to a directory named "HW2-{YOUR FIRST NAME}-{YOUR LAST NAME}":

1. Code. For example, PR.py, PR.cpp

2. Documentation explaining your code/methods.

3. README.txt showing how to run your code.

Please archive the folder and name it as "HW2-{YOUR FIRST NAME}-{YOUR LAST NAME}.zip". and upload it to T-square.

Scoring metrics:

(1) Your code (50%)

(2) The documentation explaining your code/methods (20%) and README (10%) showing how to run your code.

(3) Results of the test cases that will be used to evaluate your code. (20%)