CS 429 Information Retrieval

Assignment 4: Page Rank Algorithm

Points: 50

Submission deadline: Tuesday, 04/17/18, 11:59 PM

Late submission deadline (10% penalty): Saturday, 04/21/18, 11:59 PM

Note: This is individual work.

In this assignment, you will implement the page rank algorithm as discussed in class (see relevant lecture notes)

- Your program should read the directed graph of the webpages from a text file with the following format. You can also implement code that will automatically generate input files for various

---

(# of pages)

(# of links)

(src page 1) (dest page2) // an outgoing link from webpage page1 to page2

(src page1) (dest page1)

Figure 1 Format of the input file

numbers of pages, so as to test your code effectively. A few example input files are provided with the assignment. Note the pageids range from 0 to N-1 where N is the total no. of pages.

- You will implement the iterative algorithm that uses the power method to generate the page rank. The teleportation rate α=0.15.

- Your program should print the pageranks for the top 10 pages (both page id and pagerank value). It should also store the pageranks for all the pages (both page id and pagerank value) in the output file "out.txt".

Other instructions:

- Implement in Python 3.0
- Comment your code appropriately
- You may reuse the code from earlier assignment

Attachments:

- Skeleton code – implement the functions in the code. Use additional functions as needed.

- Sample input files.

Submission:

Submit the following files on blackboard as a .zip file.

1. pagerank.py

2. Output.txt: containing the output generated by your code for an input file. Also include the input file.