CS-430 HW4 Fall 2022 14 points

Submission instructions

- ***Due date: Friday, Dec. 2nd, 11:59 pm Central Time (i.e. local time in Chicago)***

- Late submissions and submissions violating these instructions will NOT be accepted.

*- **Absolutely no handwritten submissions. No credit will be given for such submissions.***

- Teamwork is allowed (max. 4 students/team). Individual submissions are also OK.

- Upload the following files to Blackboard
*- (1) your HW report (**pdf format only; the reports in formats other than pdf will be disregarded);***
- (2) the source codes of your programs.
The Beacon students: upload your submissions to LMS.

*- **One submission per team only.** Write down names, A#, and section numbers (i.e. live, online, Beacon) of all the team members on the front page. Do **not** submit multiple copies of your assignment (e.g. by each team member). It is very confusing and will be penalized. **Clearly indicate how each team members contributed to your teamwork.***

- If you use any additional materials to solve the HW problems (e.g. textbooks, research papers, websites, etc.), reference them.

*- **Hao Ding (hding9@hawk.iit.edu)** is responsible for grading this assignment. Feel free to ask questions if you have any doubts but don't send him or me:*
- Your partial solutions with inquiries "Is that what you expect?"
- Questions the answers to, may give explicit hints on how to solve the problems.

The objective of HW4 is to investigate the PageRank Algorithm proposed by Sergey Brin and Larry Page, the founders of Google. Before you start working on this assignment **watch this video:** *http://www.amsi.org.au/ESA_Senior_Years/media/Google_PageRank.html*
and read the report [1]. Terminology and notation used in the problem statement below (e.g. hyperlink graph, row-normalized hyperlink matrix, dangling node, Google matrix, damping factor, PageRank row vector) **have been introduced in [1].**

1. Consider a 10-page mini web represented by a directed hyperlink graph whose topology is specified by the following sequence of numbers (determined by the prime powers):
   13, 16, 17, 19, 23, 25, 27, 29, 31, 32, 37, 41, 43, 47, 49, 53, 59, 61, 64, 67, 71, 73, 79, 81, 83, 89, 97, 101, 103, 107, 109
   that define starting and ending vertices of all directed edges, e.g. 13 means that there is a hyperlink from page P_1 to page P_3, 47 - a hyperlink from P_4 to P_7, 107 - a hyperlink from P_{10} to P_7, etc. (for a 3-digit number the first 2 digits represent a starting vertex).

   (a) (1 point) Find the row-normalized hyperlink matrix H of this graph. Are there any dangling nodes in this hyperlink graph?

(b) (2 points) Find the Google matrix G of this graph. Assume the damping factor $d=0.85$.

(c) (2 points) You are **NOT** allowed to use the formula for the number of iterations (given in [1], page 15, item 2b) needed to obtain numerical values of the PageRank row vector $\boldsymbol{v}$. Propose an alternative termination criterion for this iterative algorithm.

(d) (4 points) Find iteratively numerical values of all elements of the PageRank row vector $\boldsymbol{v}=[r(P_1), r(P_2),..., r(P_{10})]$. **Use your termination criterion proposed in item (c).** How many iterations do you need to obtain your numerical results? Assume the damping factor $d=0.85$ and $\boldsymbol{v}^{(0)}=[0.1, 0.1, 0.1,..., 0.1]$. **Implement this iterative algorithm by yourself (absolutely no Excel or program libraries).**

(e) (2 points) Plot the evolution of numerical values of $r(P_5)$ and $r(P_7)$ for subsequent iteration steps of the algorithm for:

(e.1) $d=0.55$ and $\boldsymbol{v}^{(0)}=[0.1, 0.1, 0.1,..., 0.1]$

(e.2) $d=0.85$ and $\boldsymbol{v}^{(0)}=[0.1, 0.1, 0.1,..., 0.1]$

(f) (2 points) Assume $d=0.85$ and run your algorithm for three different initial PageRank row vectors $\boldsymbol{v}_1^{(0)}$, $\boldsymbol{v}_2^{(0)}$ and $\boldsymbol{v}_3^{(0)}$. What are the final numerical values of $\boldsymbol{v}_1$, $\boldsymbol{v}_2$ and $\boldsymbol{v}_3$? Are they identical or different?

(g) (1 point) Compare the numerical values of the PageRank row vector $\boldsymbol{v}$ for two different values of the damping factor $d=0.85$ and $d=1$. What is the basic difference between them? Explain it.

Submit source codes of all programs you have used to produce your results.

References

[1] Davey B. A., *Google PageRank,* Australian Mathematical Sciences Institute, University of Melbourne, 2013, https://www.amsi.org.au/teacher_modules/pdfs/Maths_delivers/Pagerank5.pdf