

## Assignment: PageRank

PageRank is the basis of Google's ranking of web pages in search results. Given a directed graph where pages are nodes and the links between pages are edges, the algorithm calculates the likelihood, or rank, that a page will be visited. The rank of a page is determined recursively by the ranks of the pages that link to it. In this definition, pages that have a lot of incoming links or links from highly ranked pages are likely to be highly ranked. This idea is quite simple and yet powerful enough to produce search results that correspond well to human expectations of which pages are important. For this assignment you need to do the following tasks:

1. [5 points] Implement the PageRank algorithm in Python with  $\alpha = 0.85$ .
2. [5 points] Using the PageRank algorithm, find the top 15 popular airports using this data <https://www.dropbox.com/s/f0xu05l38hayk36/routes.csv?dl=0>. The file contains directions between airports. The first column is a code of the first airport, and the second column is the code of the second (destination) airport.
3. [5 points] Using the PageRank algorithm, find the top 10 cited publications in the Cora dataset <https://relational.fit.cvut.cz/dataset/CORA>
4. [5 points] The value of  $\alpha$  changes the PageRank vector. You will vary  $\alpha$  between .75 and .95 to see how the PageRank vector changes. The way in which you measure or describe this change is up to you. Some examples might be:
  - a. How different are the top airports/publications for each  $\alpha$ ? How different are the PageRanks of the top airports/publications?
  - b. What is the largest change in any airport/publication's PageRank as  $\alpha$  changes?
  - c. How does the PageRank vector change as a whole?

You don't have to do any particular one of the above. If you think of a different way to measure the change, you are entirely free to use that method. For whatever method you choose for measuring the change in the PageRank vector, explain why you chose that metric and to what extent the PageRank vector changes using that metric. Write any interesting observations you find.

Submit your code (1 file) and results (1 PDF file) via Blackboard.