



Assignment 4 -Programming (9%)
CSI2110 : Data Structure and Algorithms
Due by Tuesday December 3, 2019 at 23:55

The objective of this programming assignment is to apply the PageRank algorithm on a very large graph representing a social network. This algorithm will allow you to identify the most influential person in this network.

The PageRank algorithm has been designed in 1998 by Larry Page, Sergey Brin, Rajeev Motwani and Terry Winograd. This algorithm is used to compute the level of importance of a web page; it is based on the idea that the more a page is referenced by important pages, the more this one is important. This is the algorithm that leads to the creation of Google and it is still at the core of its research engine.

The PageRank Algorithm

The algorithm computes the PageRank (PR) of a node A in a directed graph using a recursive definition:

$$PR(A) = (1-d) + d (PR(T_1)/C(T_1) + \dots + PR(T_n)/C(T_n))$$

Here d is a damping factor that we will set to 0.85. Nodes T_1, T_2, \dots, T_n are the nodes that connect to A (i.e. having a link going from T_i to A). The term $C(T_i)$ is the number of links outgoing from T_i . The idea is that a node receives its PR from the node connected to it and it distributes its PR to the node it connects to.

Since the computation of the PR of a node requires the knowledge of the PR of its neighboring nodes, the full computation of all PRs has to be done in an iterative way. You first set the PR of all nodes to 1.0. You then sequentially update the PR of each node using the current PR value of its neighbors. You have to repeat this process several times before the PRs stabilize to their true value. The higher the PR of a node is, the more influence (or importance) this node has in the network.

Instructions:

You are given a file containing the description of the network to be analyzed. You also have the source code reading this file and extracting the information about the nodes and their adjacency lists.

You have to identify the 10 most influential nodes of this network (based on their PR value).

You must submit a short report (suggested max 2 pages) in file `report.pdf` that contains:

- A brief description of your classes and methods.
- A description of how you obtain the solution (number of iterations, compute time, etc.)
- Optional: any additional information to be seen by the marker.

We ask you to submit a zip file named `student_number_a4.zip` containing:

- your code: all commented `.java` files required to run your program. Make sure to write your name and student number at the beginning of each file.
- a file `report.pdf` containing the report
- `Graph.java` exactly as provided
- `CSI2510.java` exactly as provided
- `PageRank.java` with no changes made to the provided template.
 - implement the incomplete method(s)
 - can add new private methods/variables
 - no new public methods/variables
 - no change to the public interface of the class.
- `email-dnc.edges`
- `TestPageRank.java` exactly as provided
-