

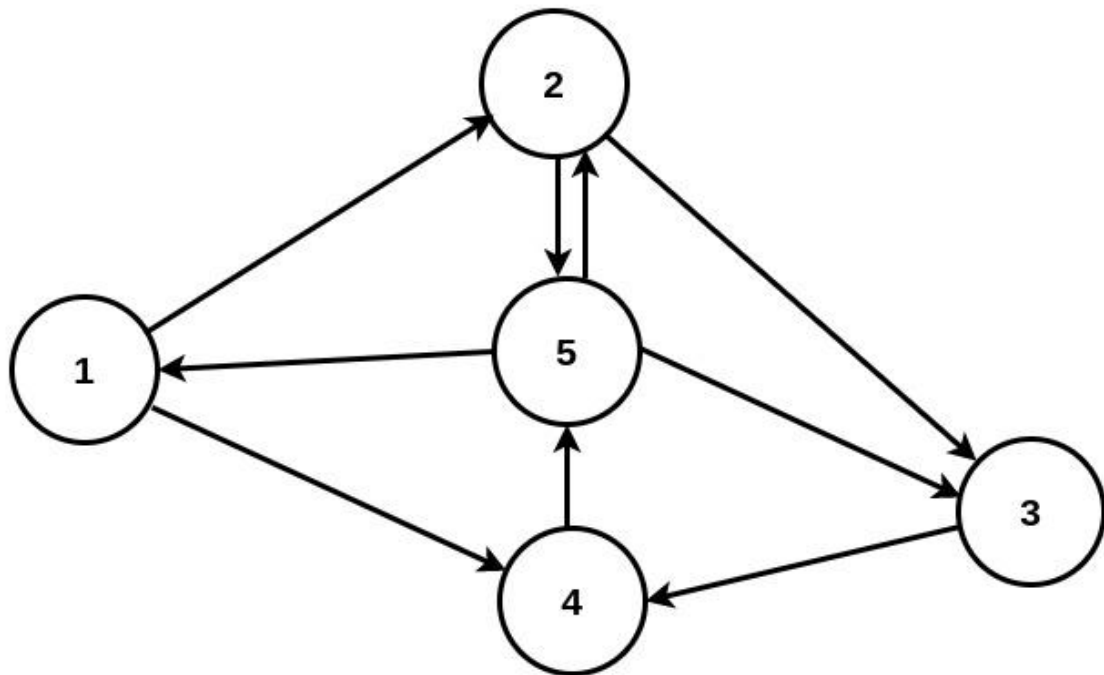
Big Data Computing

Lab - II

August 10, 2021

I. PageRank

Consider the sample directed unweighted graph represented as follows:



The information regarding the edges among the nodes are provided as part of a file labeled **input.txt**. Using the provided input file, you are provided with source code(s) and instruction manual that helps to calculate the importance of each node within the above graph through the PageRank algorithm via a map-reduce implementation. (**Note:-** Follow the steps in the instruction manual and execute the program.)

Input: A text file containing information about directed unweighted edges between different nodes in a graph.

Output: Importance of each node within the graph.

Input: *input.txt*

1 , 2 , 4
2 , 5 , 3
3 , 4
4 , 5
5 , 1 , 2 , 3

Output:

1 0.33
2 0.83
3 0.83
4 1.50
5 1.50

II. Uber pickup data analysis

Background:

In New York City, the Taxi & Limousine Commission – TLC – licenses drivers and vehicles to carry fare-paying passengers. Having a TLC driver's license (now known as the universal license), grants drivers the authority to transport passengers using a TLC licensed vehicle. You can spot a TLC licensed vehicle by the special TLC plates they wear. The plates let passengers know that the vehicle they are about to travel in is legally licensed by the TLC and has passed mechanical inspection. The plates give valuable reassurance to the potential passenger. If one aims to drive for one of these companies they have to make sure that they purchase a vehicle that they will accept on their platform.

One of the prime steps involved is to affiliate your vehicle with a TLC approved dispatch base. A dispatch base is the company you effectively drive for and who connects you to passengers seeking a ride. This can be a traditional for-hire business, or one of the newer App companies such as Uber, Lyft and Gett. These large organisations typically have their own TLC registered car bases. Contact the base of your choosing to get what is known as a base letter. The base letter is a certified document that states which base you intend to affiliate with, (which means the one you'll drive for).

Input: *uberdata.txt* (contains aggregated daily Uber trip statistics in January and February 2015)

Features:

<u>Dispatching Base Number</u>	<u>Date</u>	<u>Active Vehicles</u>	<u>Trips</u>
B02512	1/1/2015	190	1132
B02765	1/1/2015	225	1765
B02764	1/1/2015	3427	29421
B02682	1/1/2015	945	7679
B02617	1/1/2015	1228	9537

You are provided with the source code of one (Map-Reduce program) which performs the following task:

1. **Basement_trips.java** : Finds the days on which each basement has more trips.

The instructions for implementing the task are provided along with the assignment.