

Assignment #1: COMP4434 Big Data Analytics

Due Date: 23:59pm, Monday, 8 Feb. 2021

Question 1

[10 marks]

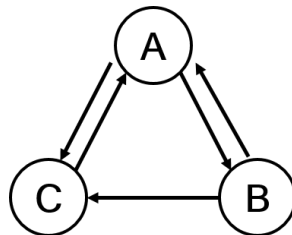
“Social networks have developed rapidly in recent years. According to eMarketer's forecast, the total number of social Internet users in China will increase by 4.8% in 2020, reaching 859.1 million. The peak volume of Sina Weibo posts reached a new high. At 0:00:00 during the Chinese New Year in 2020, a total of 32,312 Weibo posts were posted simultaneously.”

Please specify which Vs in 4V are reflected in the text above and explain the reason in detail. [10 marks]

Question 2

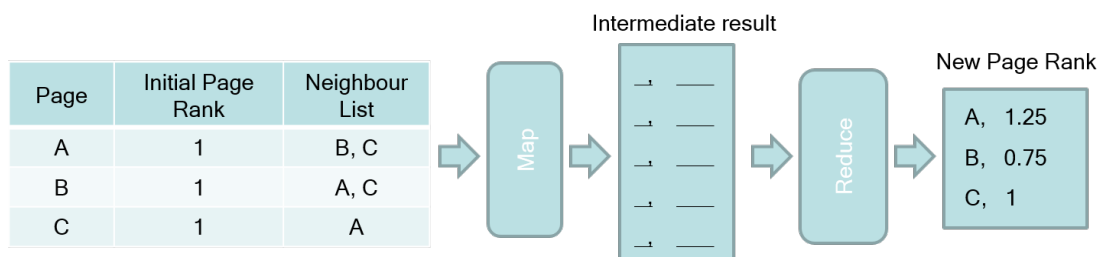
[15 marks]

Consider an imaginary web of 3 web pages, as shown in the figure below:



Assume that the initial page rank of each web page is 1 and the damping factor is 0.5.

- Calculate the page rank values of A, B, C for the first three iterations. Approximate the results to 3 decimal places. [5 marks]
- If the approximate page rank values stay unchanged in iterations, we consider that the page rank values reach convergence. Write the number of iterations required for page rank values to converge and give the final page rank values for A, B, and C. (Programming is encouraged) [5 marks]
- The following graph illustrates the process of PageRank algorithm in MapReduce framework. Calculate the intermediate result with calculation process. [5 marks]



Question 3**[25 marks]**

Extracting part of the census data, we can get the following child-parent relationship table:

Child	Parent
Tom	Lucy
Tom	Jack
Jone	Lucy
Jone	Jack
Lucy	Mary
Lucy	Ben
Jack	Alice
Jack	Jesse
Terry	Alice
Terry	Jesse
Philip	Terry
Philip	Alma
Mark	Terry
Mark	Alma

We need to use MapReduce to find the grandchild-grandparent relationship (example: Tom-Mary) from this table.

- Explain how you implement the map and reduce functions (including the key-value pair definition) in pseudo code and show the intermediate results by each mapper and the output by each reducer. (Using 2 mappers and 2 reducers, consider the rank and shuffle module is predefined.) [15 marks]
- Implement the map and reduce function using python language. And upload the source code file. [10 marks]