

Big Data Analytics Assignment 3

Due Date: 23:59 Jun 22, 2020

Every Student **MUST** include the following statement, together with his/her signature in the submitted homework.

I declare that the assignment submitted is original except for source material explicitly acknowledged, and that the same or related material has not been previously submitted for another course. I also acknowledge that I am aware of University policy and regulations on honesty in academic work, and of the disciplinary guidelines and procedures applicable to breaches of such policy and regulations.

Signed(Student _____) Date: _____

Name _____ SID _____

Submission notice:

- Submit your homework to Email xuhl@dgut.edu.cn

General homework policies:

A student may discuss the problems with others. However, the work a student turns in must be created COMPLETELY by oneself ALONE. A student may not share ANY written work or pictures, nor may one copy answers from any source other than one's own brain.

Each student **MUST LIST** on the homework paper the **name of every person he/she has discussed or worked with**. If the answer includes content from any other source, the student **MUST STATE THE SOURCE**. Failure to do so is cheating and will result in sanctions. Copying answers from someone else is cheating even if one lists their name(s) on the homework.

If there is information you need to solve a problem but the information is not stated in the problem, try to find the data somewhere. If you cannot find it, state what data you need, make a reasonable estimate of its value, and justify any assumptions you make. You will be graded not only on whether your answer is correct, but also on whether you have done an intelligent analysis.

Part A. PageRank Programming

In this exercise, you are required to implement the PageRank algorithms on the given dataset, and then output the top 100 nodes ranked by their PageRank scores.

The description of the dataset is included in the first few lines of the dataset.

You need to submit the source code and comment your source code so that the tutor can follow it. Submit your result in a plain text.

References:

[1] DataSet

<https://xuhappy.github.io/courses/BigData/homework/web-Google.txt.gz>