CS 302: Design and Analysis of Algorithms
Spring 2020
Semester Project
Deadline: 15 May 2020
Submissions are accepted till 1st June 2020

DBLP listed more than 3.66 million journal articles, conference papers, and other publications on computer science in July 2016, up from about 14,000 in 1995. All important journals on computer science are tracked. Proceedings papers of many conferences are also tracked. It is mirrored at three sites across the Internet.

This semester your Design and Analysis of Algorithm project is to use a DBLP dataset from the following link.
https://drive.google.com/file/d/1_-n6eQl-7O9l9F9tHteQAH_CRsqeeUfC/view?usp=sharing

DBLP dataset has information of collaboration among multiple researchers. In this project you have to find the most important collaboration among the whole dataset. In order to do so you need to create a graph for the given data set. In your graph,   Node represents researchers and the link between them indicates collaboration among multiple researchers while weight on that link shows the number of times those researchers have worked together.
 Your tasks are following
1. Create a graph representation from a file that has 3 columns, first two columns represent author id and third represents the weight of the link. Edge weight is proportional to the amount of collaboration (bigger means collaborating quite frequently)
2. Create a **Maximum** spanning tree from the graph created in (1) . This represents the top collaborators of each researcher.
3. By using the maximum spanning tree generated in (2) enumerate the top 5 distinct collaborator pairs. Your algorithm must perform better than brute force. Credit for this question mainly depends on the algorithm efficiency along with the accuracy.

Instructions:
- You need to submit your code in Google classroom.
- This is a group project. Each group should not exceed more than 2.
- Every group should notify TA's about its members.
- No late submissions are acceptable.
- Code should be submitted with results. (Any language is allowed)

- Assignment deadline is 15 May but in order to give enough time department have decided to give extension to project till 1 june and accept submissions without any deduction.
- **Considering COVID-19 situation in Pakistan IF there are no final exams in Pakistan then this project would have considerably high weightage.**

Examples of graphs :

Following are the examples of author collaboration.