CS 5683: Big Data Analytics

Assignment-4: Friend Recommendation Engine

Total Points: 10 (3% toward final)

Due date: Oct 11, 2020 at 11:59pm

In this assignment, we create a simple social network friend recommendation engine. The key idea is that if two people have a lot of mutual friends, then the engine should recommend that they connect with each other. *Spark is not required. Please use helper Python program.*

Dataset: The given data file contains the adjacency list and has multiple line in the format: <USER> TAB <FRIENDS>, where <USER> is a unique integer ID corresponding to a unique user and <FRIENDS> is a comma separated list of unique IDs corresponding to the friends of the user with the unique ID <USER>. Note that the friendships are mutual (i.e., edges are undirected): if A is a friend with B, then B is also a friend with A

Algorithm: Let us use a simple algorithm such that, for each user U, the algorithm recommends N = 10 users who are not already friends with U, but have the most number of mutual friends in common with U

Output:

1. The output should contain one line per user in the following format:

   <USER>: <RECOMMENDATIONS>, where <USER> is a unique ID corresponding to a user and <RECOMMENDATIONS> is a comma separated list of unique IDs corresponding to the algorithm's recommendation of people that <USER> might know, ordered in decreasing number of mutual friends

2. Even if a user has less than 10 recommendations, output all of them in decreasing order of the number of mutual friends. If a user has no recommendations, give an empty list. If there are recommendations with the same number of mutual friends, then output those user IDs in numerically ascending order
3. Output recommendations of user IDs: 924, 1153, 6712, 8213, 12835, 23478

What to submit? Complete the helper program and submit the jupyter notebook in Canvas

Grading rubric: Check the helper program