



Northeastern University, Khoury College of Computer Science

CS 6220 Data Mining | Assignment 3

Due: October 10, 2023(100 points)

Chun Wei Tseng

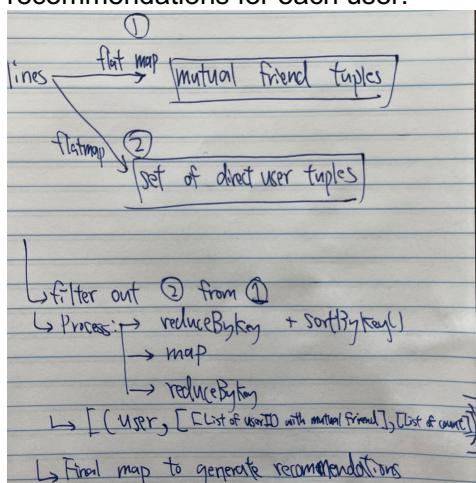
<https://github.com/Chun-Wei-Tseng/CS6220-Assignment-3>

Map Reduce in Spark

Write a Spark program that implements a simple “People You Might Know” social network friendship recommendation algorithm. The key idea is that if two people have a lot of mutual friends, then the system should recommend that they connect with each other.

Pipeline Sketch

→ Ans: For any two friends in the friend list of a user, they must have a mutual friend, which is the user. I first use flat Map with customized function to get a list of tuples of users with mutual friends and a count 1 (user1, user2, 1). Then I use flat Map with another customized function to generate a set of tuples where users a tuple are direct friends. We don't want to recommend people who are already friends with them to a user, so I use filter function to filter out those direct friends from the first list of tuples of mutual friends. After using some other reduceByKey and map functions to process the filtered data, I use customized map function, generateRecommendation, to generate recommendations for each user.



Recommendation for Users:

- **924:** ('2409', '6995', '11860', '15416', '43748', '45881')
- **8941:** ('8943', '8944')
- **8942:** ('8943', '8944')
- **9019:** ('9022', '9023')
- **9020:** ('9021', '9022', '9023')
- **9021:** ('9022', '9023')
- **9022:** ('9023')
- **9990:** ('13134', '13478', '13877', '34299', '34485', '34642', '37941')
- **9992:** ('35667', '9989', '9991')
- **9993:** ('13134', '13478', '13877', '34299', '34485', '34642', '37941')