



Northeastern University, Khoury College of Computer Science

## CS 6220 Data Mining | Assignment 3

Due: February 8, 2024

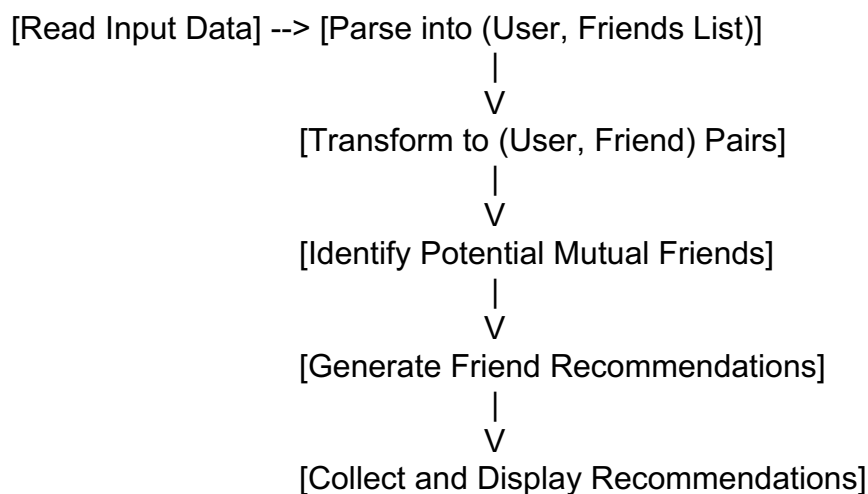
Yichen Sun

<https://github.com/LAnselet/cs6220-datamining>

### Pipeline Sketch

For any users unique integer ID, they have a comma separated list of unique IDs corresponding to the direct friends of the user with the unique ID. Furthermore, the user cannot be their recommended friends in any direct friends where they are mapped to each friend. Therefore, they shouldn't be included among the recommendation.

When a pair occurs more than once, we can add them together with `reduceByKey` and then sort them in descending order to make recommendation in limit of 10.



### Recommendations for the users

8942: 8939,8940,8943,8944

924: 439,2409,6995,11860,15416,43748,45881

9993: 9991,13134,13478,13877,34299,34485,34642,37941  
9020: 9021,9016,9017,9022,317,9023  
9019: 9022,317,9023  
8941: 8943,8944,8940  
9992: 9987,9989,35667,9991  
9022: 9019,9020,9021,317,9016,9017,9023  
9021: 9020,9016,9017,9022,317,9023  
9990: 13134,13478,13877,34299,34485,34642,37941