

CS550: Massive Data Mining

Homework 1

Due 11:59pm Monday, February 23, 2026
Please see the homework file for late policy

Submission Instructions

Honor Code Students may have discussions about the homework with peers. However, each student must write down their solutions independently to show they understand the solution well enough in order to reconstruct it by themselves. Students should clearly mention the names of all the other students with whom they have discussions about the homework. Directly using the code or solutions obtained from the web or from others is considered an honor code violation. We check all the submissions for plagiarism and take the honor code seriously, and we hope students to do the same.

Discussions (People with whom you discussed ideas used in your answers):

On-line or hardcopy documents used as part of your answers:

I acknowledge and accept the Honor Code.

(Signed)_____

If you are not printing this document out, please type your initials above.

Answer to Question 1

Algorithm Description

The problem is solved with a single MapReduce job.

Map phase: For each user U with friend list $F = \{f_1, f_2, \dots, f_n\}$, the mapper performs two kinds of emissions:

1. For every pair (f_i, f_j) in F (i.e., two friends of U), emit $(f_i, (f_j, 1))$ and $(f_j, (f_i, 1))$. This records that f_i and f_j share U as a mutual friend, so they are potential recommendations for each other.
2. For every friend $f_i \in F$, emit $(U, (f_i, -1))$ to mark that U and f_i are already friends.

Reduce phase: For each user U , the reducer aggregates all emitted values. It sums up the mutual-friend counts for each candidate and filters out candidates that are already direct friends (marked with -1). The remaining candidates are sorted in decreasing order of mutual-friend count, with ties broken by ascending user ID. The top 10 are output as recommendations.

Recommendations for Specified Users

User	Recommendations
924	439, 2409, 6995, 11860, 15416, 43748, 45881
8941	8943, 8944, 8940
8942	8939, 8940, 8943, 8944
9019	9022, 317, 9023
9020	9021, 9016, 9017, 9022, 317, 9023
9021	9020, 9016, 9017, 9022, 317, 9023
9022	9019, 9020, 9021, 317, 9016, 9017, 9023
9990	13134, 13478, 13877, 34299, 34485, 34642, 37941
9992	9987, 9989, 35667, 9991
9993	9991, 13134, 13478, 13877, 34299, 34485, 34642, 37941

Answer to Question 2(a)

Answer to Question 2(b)

Answer to Question 2(c)

Answer to Question 2(d)

Answer to Question 2(e)

Answer to Question 3(a)

Answer to Question 3(b)

Answer to Question 3(c)