In this project, you will use spark (pyspark or scala) to analyze social network data.

**Q1**

**Write a spark script to find total number of common friends for any possible friend pairs.** The key idea is that if two people are friend then they have a lot of mutual/common friends.

For example,

Alice’s friends are Bob, Sam, Sara, Nancy Bob’s friends are Alice, Sam, Clara, Nancy Sara’s friends are Alice, Sam, Clara, Nancy

As Alice and Bob are friend and so, their mutual friend list is [Sam, Nancy]

As Sara and Bob are not friend and so, their mutual friend list is empty. (**In this case you may exclude them from your output**).

Input files:

1. soc-LiveJournal1Adj.txt

**The input contains the adjacency list and has multiple lines in the following format:**

<User><TAB><Friends>

Here, <User> is a unique integer ID corresponding to a unique user and <Friends> is a comma-separated list of unique IDs (<User> ID) corresponding to the friends of the user. Note that the friendships are mutual (i.e., edges are undirected): if A is friend with B then B is also friend with A. The data provided is consistent with that rule as there is an explicit entry for each side of each edge. So when you make the pair, always consider (A, B) or (B, A) for user A and B but not both.

**Output: The output should contain one line per user in the following format:**

<User\_A>, <User\_B><TAB><Mutual/Common Friend Number>

where <User\_A> & <User\_B> are unique IDs corresponding to a user A and B (A and B are friend). < Mutual/Common Friend Number > is total number of common friends between user A and user B.

**Q2.**

Please answer this question by using data sets below.

1. soc-LiveJournal1Adj.txt

2. userdata.txt

The userdata.txt consists of column1 : userid

column2 : firstname column3 : lastname column4 : address column5: city column6 :state column7 : zipcode column8 :country column9 :username

column10 : date of birth.

Find top-10 friend pairs by their total number of common friends. For each top-10 friend pair print detail information in decreasing order of total number of common friends. More specifically the output format can be:

<Total number of Common Friends><TAB><First Name of User A><TAB><Last Name of User A> <TAB><address of User A><TAB><First Name of User B><TAB><Last Name of User B><TAB>

<address of User B>

…