YL5090

HW1b

Read text into RDD:

* Key is each user, e.g. U1
* Value is a list of tuples, each tuple being the user name and number of common friends this user has with U1.

First transformation: Cartesian with itself. This will create tuples, each have two elements. Each element is again a (user, (friends)) tuple.

Second transformation: filter the above RDD to remove the tuples that 1) have the same user twice and 2) the first user is in the second user’s friend list. After this we will have all tuples being two (user, (friends)), while the two users are not the same or are friends with each other.

Map function: calculate the number of common friends in the two tuples. This is done by a intersection of tuples. After that the RDD key will be userID, the value will be a tuple : (another user, number of common friends)

Combine by Key: this step will get all the users that have common friends with a user, and group them into a list. After this step the RDD key will be userID, the value will be a list of tuples. Each tuple being (another userID, number of common friends).

Map: this step is to sort the list for each user by the number of common friends another user has with them.

Map: this step is to get the top 10 tuples for each user. The result will be key = userID, value = a list of ten tuples. Each tuple being (another userID, number of common friends).

FlatMap: break down the list into individual key-value pairs. Now key is userID, value is a tuple of (another user ID, number of common friends).