

**Azka Ali**

**19I-0597**

**Findings:**

* To find the most popular user, we have to know about the user with maximum number of friends (that is, the biggest circle)
* For this purpose, we’re using .circle files of each user in order to identify the biggest circle
* To achieve this, each user’s circles are sent to the mapper.py which basically sums each circle up in order to find the largest sum
* Then the result is sorted and shuffled automatically while reducer.py picks out the largest sum; identifying user with the biggest circle in the given dataset
* The same output is written on Hadoop cluster as well, which is in file 107.circles, circle name is circle6
* For the second part, we send the biggest circle along with its nodes to the hadoop cluster as an output of part1’s reducer, this way both parts have the same reducer
* Later we start comparing all the other circles point by point with the biggest circle
* This way we can calculate the similarity between both and get a result
* The reducer can thus find the maximum of similarity and result in the user with maximum similarity with the user with biggest circle

**Hadoop Cluster Output:**

This prints the name of the file containing user with biggest circle, along with an agregation of users in the circle, thereby showing the user with biggest friends circle.

