

Assignment 4 – ReadMe

I have run the code on **Mac** terminal. Following is the information about the same:

Environment requirements-

I have used Python 2.7 and Spark 2.2.1 to complete this assignment

Note: Paths can be relative to current directory or absolute.

Note: There should NOT be any spaces in the file path or file name

Task 1

Name of python file: Nupur_Shukla_Betweenness.py

Command to run: \$SPARK_HOME/bin/spark-submit Nupur_Shukla_Betweenness.py <Path of ratings.csv file>

Example: \$SPARK_HOME/bin/spark-submit Nupur_Shukla_Betweenness.py
"/Users/nupur/Desktop/Assignment_04/data/ratings.csv"

Time: 84 seconds

Task 2

Name of python file: Nupur_Shukla_Community.py

Command to run: \$SPARK_HOME/bin/spark-submit Nupur_Shukla_Community.py <Path of ratings.csv file>

Example: \$SPARK_HOME/bin/spark-submit Nupur_Shukla_Community.py
"/Users/nupur/Desktop/Assignment_04/data/ratings.csv"

Time: 320 seconds

Bonus

Name of python file: Nupur_Shukla_Bonus.py

Command to run: \$SPARK_HOME/bin/spark-submit Nupur_Shukla_Bonus.py <Path of ratings.csv file>

Example: \$SPARK_HOME/bin/spark-submit Nupur_Shukla_Bonus.py
"/Users/nupur/Desktop/Assignment_04/data/ratings.csv"

Library used: I have used **Python networkx** library's girvan_newman method to detect communities for this task. I have generated 4 communities using this library.

Time: 417 seconds