

Homework 4

Technology : Python: 2.7 Scala: 2.11 Spark: 2.2.1

Directory:

Surbhi_Batra_hw2

- Surbhi_Batra
 - Surbhi_Batra_Description
- OutputFiles
 - Surbhi_Batra_Betweenness.txt
 - Surbhi_Batra_Community.txt
 - Surbhi_Batra_Task3.txt
- Solution
 - Surbhi_Batra_Betweenness.py
 - Surbhi_Batra_Community.py
 - Surbhi_Batra_Task3.py

Algorithm:

>>>Execute these Commands after : cd/to/Surbhi_Batra_hw4/Solution

Python:

TASK : 1

Betweenness

Command :

\$SPARK_HOME/bin/spark-submit Surbhi_Batra_Betweenness.py ratings.csv

Time taken : 60.0701379776 seconds

TASK : 2

Community

\$SPARK_HOME/bin/spark-submit Surbhi_Batra_Community.py ratings.csv

Optimal No of Communities : 224

Modularity : 0.0164918517156

Time Taken : 496.938520908 seconds

Note: I have used Networkx so do pip install networkx if it is not present already on the system.

TASK : 3 : Bonus Task

Command : \$SPARK_HOME/bin/spark-submit Surbhi_Batra_Task3.py ratings.csv

Modularity = 0.121062978854

Before running the file, I installed the following libraries using the following commands on my linux.

```
sudo python -m pip install community
```

```
sudo python -m pip install python-louvain
```