

# **Assignment 4: Community Detection**

## Environment Versions

Spark: 2.2.1

Python: 2.7

Scala: 2.11

## **Task 1 : Calculating Edge Betweenness**

Calculating edge betweenness using '*Piyush\_Umate\_Betweenness.py*'

```
spark-submit Piyush_Umate_Betweenness.py <rating_file_path>
```

Time taken: 29.3946559429 sec

Here,

**rating\_file\_path** is file path of input ratings file

Calculating edge betweenness using '*Piyush\_Umate\_hw4.jar*' in Scala

```
spark-submit --class Betweenness Piyush_Umate_hw4.jar <rating_file_path>
```

Time taken: 58sec

Here,

**rating\_file\_path** is file path of input ratings file

## **Task 2 : Community Detection**

Calculating communities using '*Piyush\_Umate\_Community.py*'

```
spark-submit Piyush_Umate_Community.py <rating_file_path>
```

Time taken: 258.077067137 sec

Here,

**rating\_file\_path** is file path of input ratings file

Calculating communities using '*Piyush\_Umate\_hw4.jar*'

```
spark-submit --class Community Piyush_Umate_hw4.jar <rating_file_path>  
Time taken: 227sec
```

Here,

**rating\_file\_path** is file path of input ratings file

## Bonus: Sparkling Graph

Executing community detection using Pscan implemented in '*Piyush\_Umate\_hw4.jar*'

```
spark-submit --class Community_Sparkling Piyush_Umate_hw4.jar <rating_file_path>  
Time taken: 12sec
```

Here,

**rating\_file\_path** is file path of input ratings file

Possible improvements

1. I have taken epsilon as 0.8305, but this value can be adjusted to get better communities.