Marvel Dataset Report

Network and Algorithms

01.01.2021

PREPARED BY

Javid Guliyev

Sadig Gojayev

Majid Gurbanli

Chapter 1

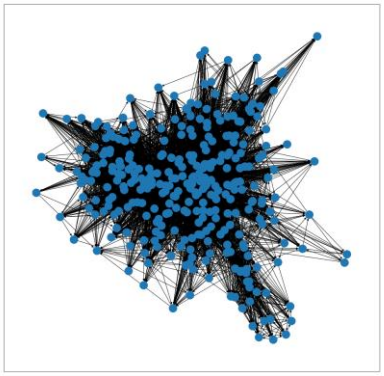
Introduction

1.1 Introduction to Project

The main point of the project is to measure the different measures and metrics such as degree centrality, betweenness centrality and graph representation on this metrics. For this purpose we use marvel dataset.

1.2 Dataset

As a dataset we used Marvel dataset which is about relationship between Marvel characters. Above-mentioned dataset contains 327 unimodal nodes and 9891 weighted edges. The relationships between characters were calculated based on how many times two characters appeared together in the same comic book.



The visual(graph) representation of dataset

Chapter 2

2.1 Measures and Metrics

- Degree centrality
- Closeness Centrality
- Betweenness Centrality
- Network Diameter
- Network Density
- Network Average Shortest Path Length

Above-mentioned measures and metrics were used in our project.

Chapter 3

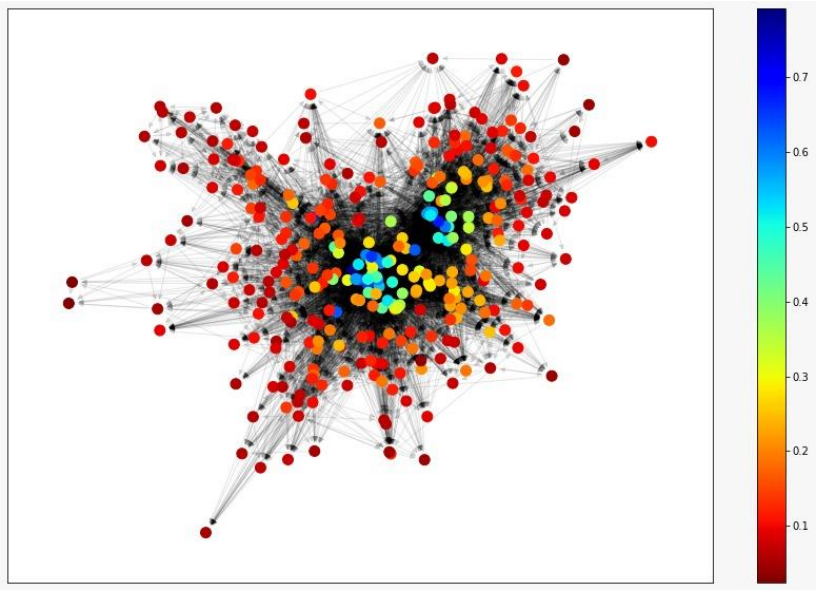
3.1 Degree Centrality

The degree centrality of a node is simply its degree—the number of edges it has. For finding degree centrality of a node we divide maximum possible score of node to n – 1 where n is a number of nodes. For finding degree centrality of network we can sum all centrality values of nodes. The formula of degree centrality is:

$$C_D(v) = \deg(v)$$

Picture 2

The visual(graph) representaion of degree centrality is :



Picture 3

The numerical value of degree centrality is 60.68.

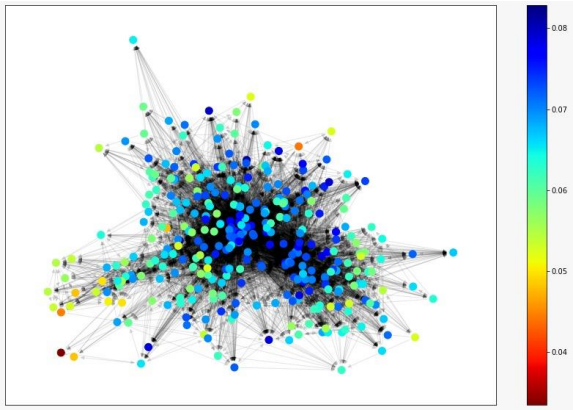
3.2 Closeness Centrality

Closeness centrality of a node indicates how close a node is to all other nodes in the network. It is calculated as the average of the shortest path length from the node to every other node in the network. Closeness centrality of a network is a sum of shortest path of closeness centrality of a node to all other nodes. The formula of Closeness Centrality is

$$C(x) = \frac{1}{\sum_y d(y,x)}.$$

Picture 5

The visual representation of closeness centrality is:



Picture 4

The numerical value of closeness centrality is 21.66.

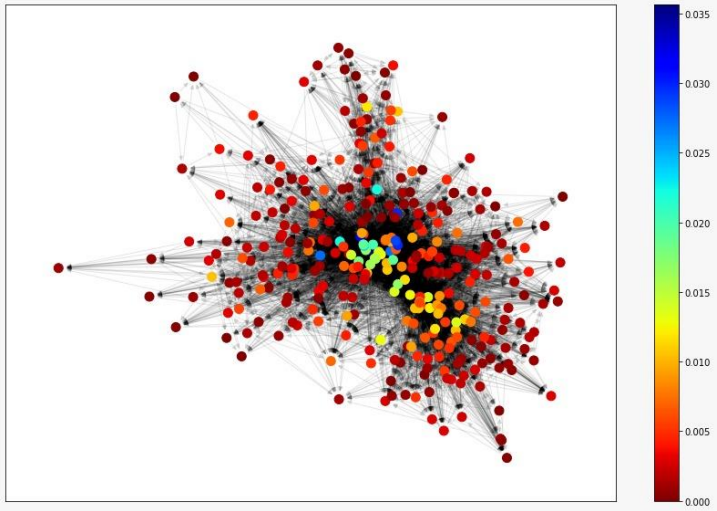
3.3 Betweenness Centrality

The centrality captures how much a given node is in-between others. This metric is measured with the number of shortest paths (between any couple of nodes in the graphs) that passes through the target node *u* (denoted $\sigma\sigma_{v,w}(u)$). The formula of betweenness centrality is :

$$g(v) = \sum_{s \neq v \neq t} \frac{\sigma_{st}(v)}{\sigma_{st}}$$

Picture 5

Betweenness Centrality of network is a cumulative sum of all betweenness centralities of each node. The visual representation of betweenness Centrality is:



Picture 6

The numerical value of betweenness centrality is 1.556.

3.3 Network Density

A network's density is the number of connections divided by the number of possible connections. A completely linked network has a density of 1. Formula of network density of directed graph is :

$$d = m/(n)(n-1)$$

Picture 7

Where m is a number of edges and n is a number of nodes. The numerical value of network density is 0.09.

3.4 Network Diameter

It is the shortest distance between the two most distant nodes in the network. In other words, once the shortest path length from every node to all other nodes is calculated, the diameter is the longest of all the calculated path lengths. The numerical value of our marvel dataset is 4.

3.5 Network Average Shortest Path Length

Average path length is a concept in network topology that is defined as the average number of steps along the shortest paths for all possible pairs of network nodes. The formula of Network Average Shortest Path Length is :

$$l_G = \frac{1}{n \cdot (n-1)} \cdot \sum_{i \neq j} d(v_i, v_j),$$

The numerical value for NASPL is 2.123

Chapter 4

4.1 Relationships between metrics.

For this purpose we create a function which remove the certain nodes that affect graph less. Moreover with such manner we decrease the complexity of graph representation. For each metric we remove part of nodes that contains metric values which are less than average value of given metric. After this evaluation we do not evaluate metric value for new group of nodes.

4.1.1 Betweenness centrality

For the betweenness centrality 219 nodes which were below the average were removed.

```
Network diameter                 : 3
Network density                  : 0.11838006230529595
Network avr_shortest_p length    : 2.1953963309103495
Betweenness Centrality           : 0.5415159273111136
Degree Centrality                : 22.717791411042946
Closeness Centrality             : 7.152810901406879
```

The previous value(without replacement) of betweenness centrality was 1.556. The initial value of it is 0.541. The value decreased nearly 66%.

4.1.2 Closeness Centrality

For the closeness centrality 150 nodes which were below the average were removed.

```
Network diameter                 : 3
Network density                  : 0.09530688238315357
Network avr_shortest_p length    : 2.199280945043657
Betweenness Centrality           : 0.789090544870158
Degree Centrality                : 33.245398773006144
Closeness Centrality             : 11.71150289772625
```

The previous value(without replacement) of closeness centrality was 21.66. The initial value of it is 11.711. The value decreased nearly 46%.

4.1.3 Degree Centrality

For the degree centrality 222 nodes which were below the average were removed.

```
Network diameter                 : 3
Network density                  : 0.11913919413919413
Network avr_shortest_p length    : 2.193681318681319
Betweenness Centrality           : 0.520149443131982
Degree Centrality                : 22.15950920245399
Closeness Centrality             : 6.937349230076724
```

The previous value(without replacement) of degree centrality was 60.68. The initial value of it is 22.15. The value decreased nearly 62%.

4.2 Comparison table

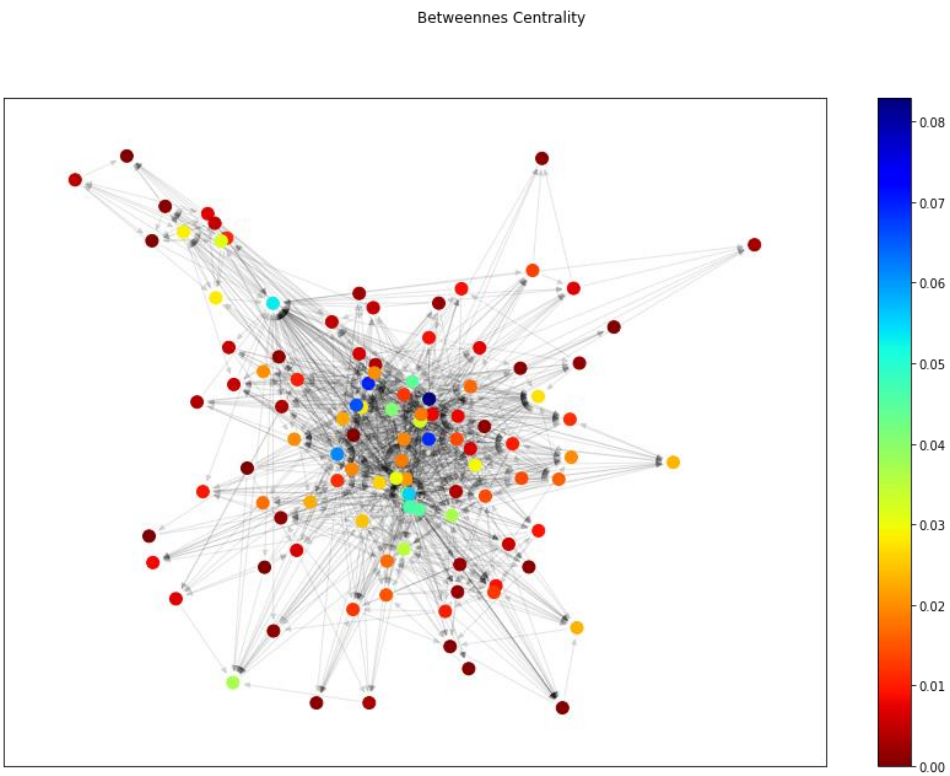
Measures	Normal	Betweenness	Closeness	Degree
Diameter	4	3	3	3
Density	0.0928	0.1184	0.0953	0.1191
NASPL	2.1232	2.1954	2.1993	2.1937
Betweenness	1.5553	0.5415	0.7891	0.5201
Degree	60.6810	22.7178	33.2454	22.1595
Closeness	21.6639	7.1528	11.7115	6.9373

4.3 Conclusion

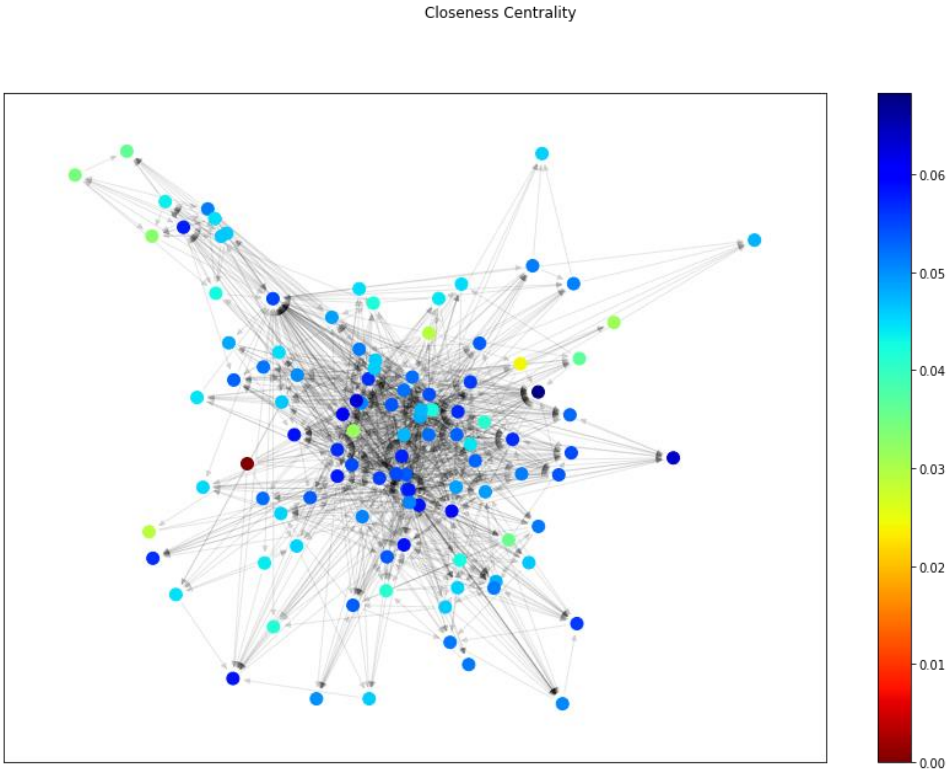
The main goal of the project was done. Certain measurements were calculated and Graph was represented. For each metric the average value was measured to decrease the complexity of the representation O(n). The Diameter changed to 3 from 4, because we dropped each time more than half of the nodes. The degree centrality composes more higher values than its average that the 62% of data was dropped for re-representation. As the closeness centrality related with degree centrality, we can see that when we extract higher values of closeness centrality, the degree centrality gets his highest cumulative sum. Approximately half of closeness centrality values are above the average of its values (46%). The last but not least, with betweenness centrality we dropped 66% nodes that contains the values less than average of its attribute. It means that 34% nodes contain high proportion of shortest paths. Final the Relation table was created.

4.4 Visual results of measures

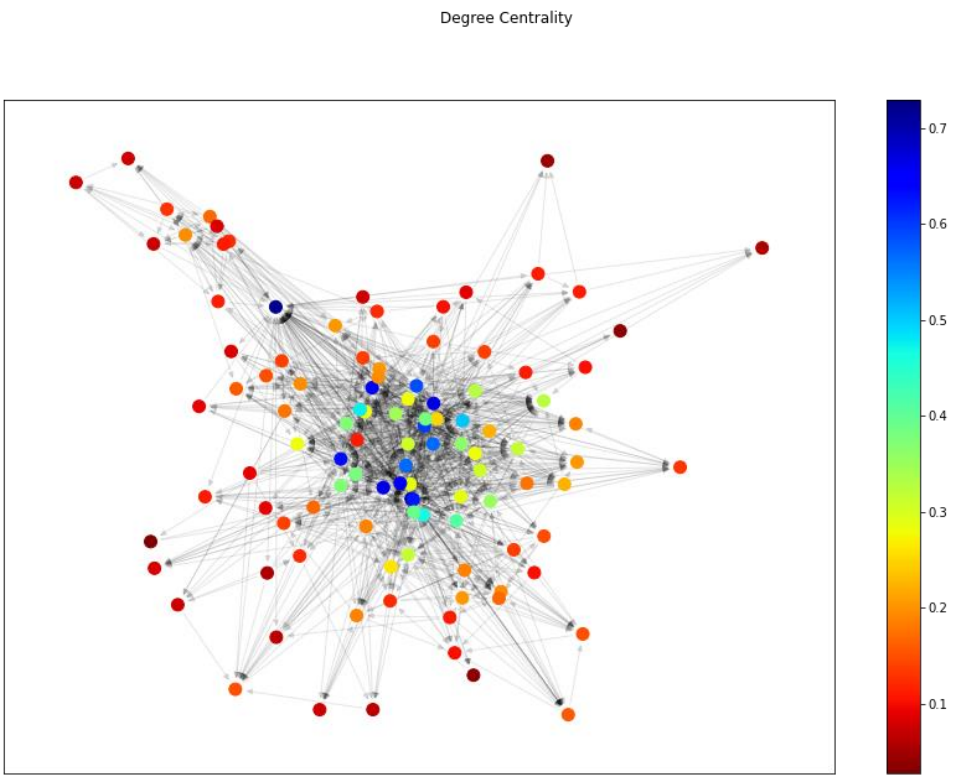
4.4.1 Betweenness Centrality



Picture 8. Betweenness Centrality

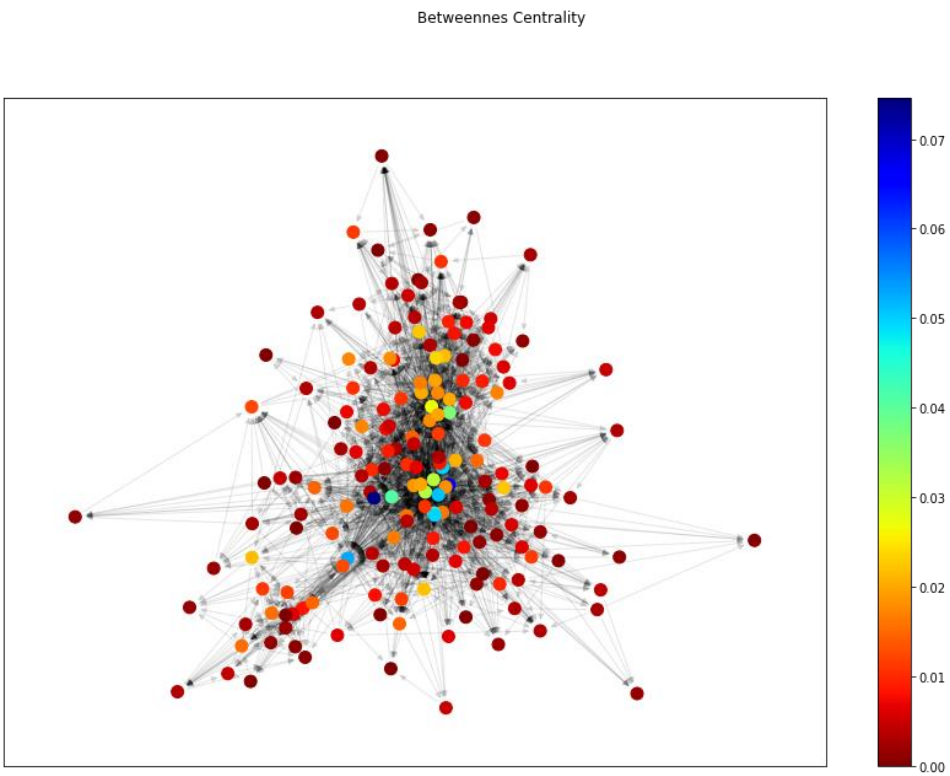


Picture 9. Closeness Centrality

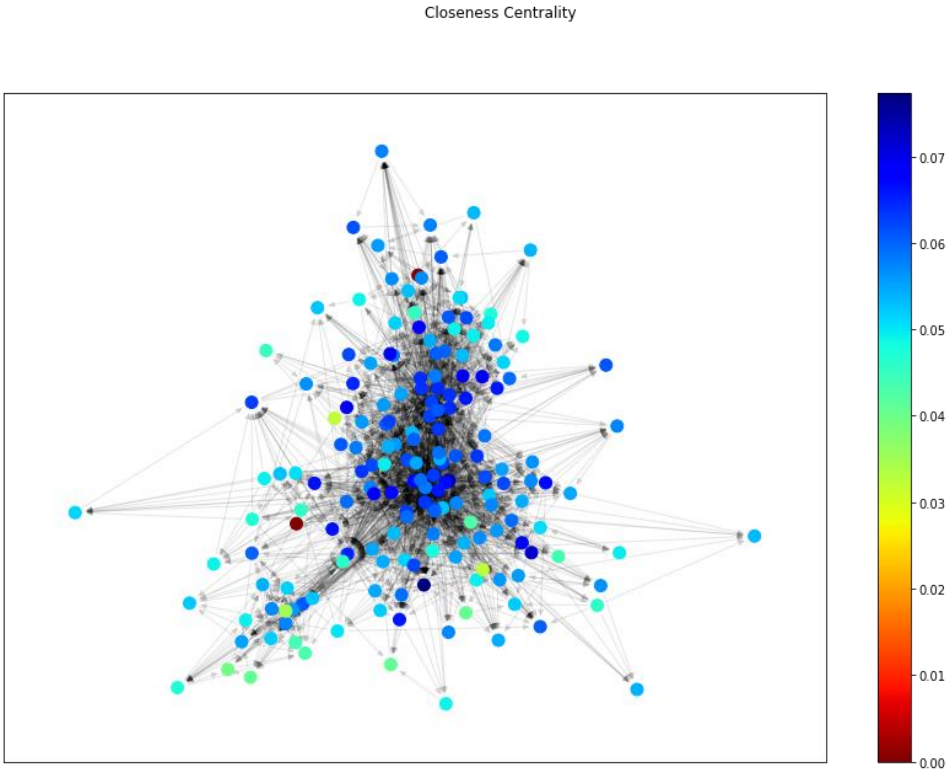


Picture 10. Degree Centrality

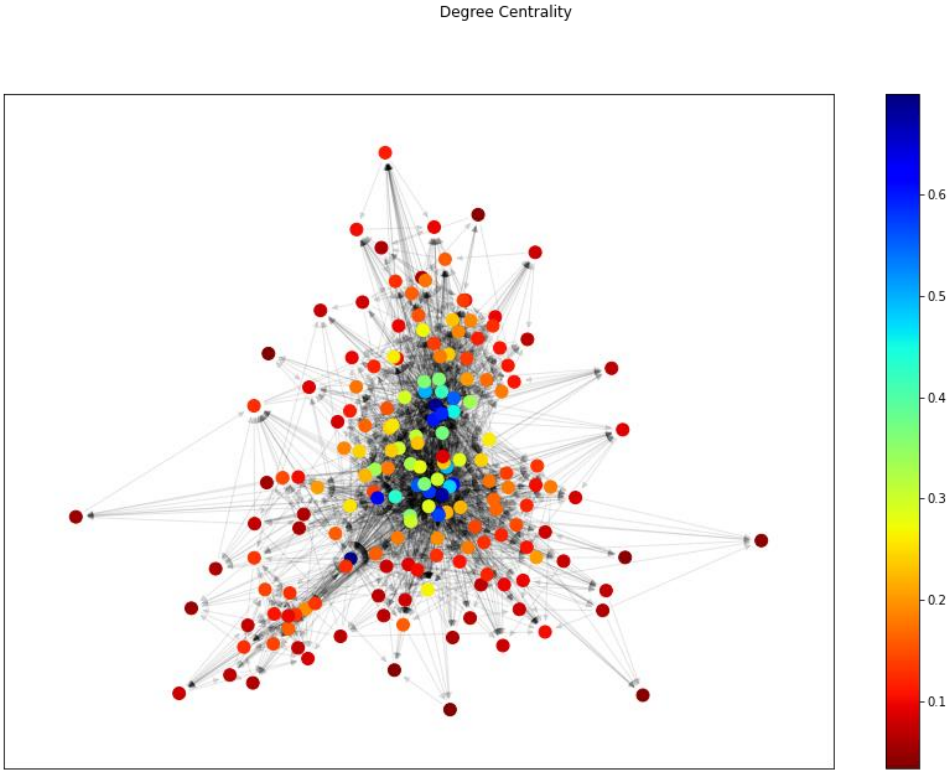
4.4.2 Closeness Centrality



Picture 11. Betweenness Centrality

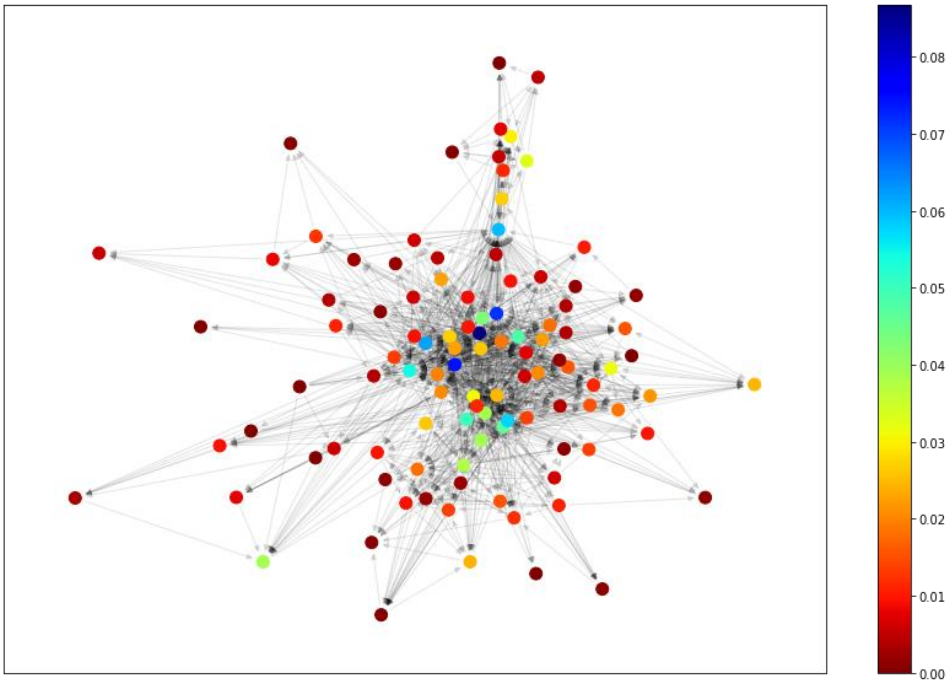


Picture 12. Closeness Centrality

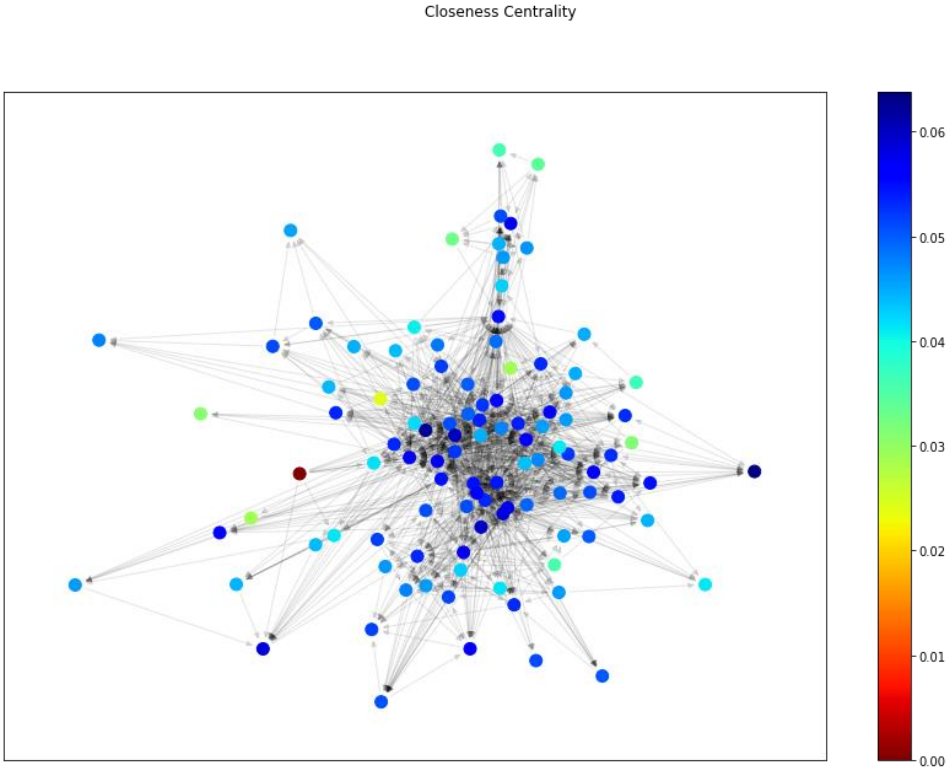


Picture 13. Degree Centrality

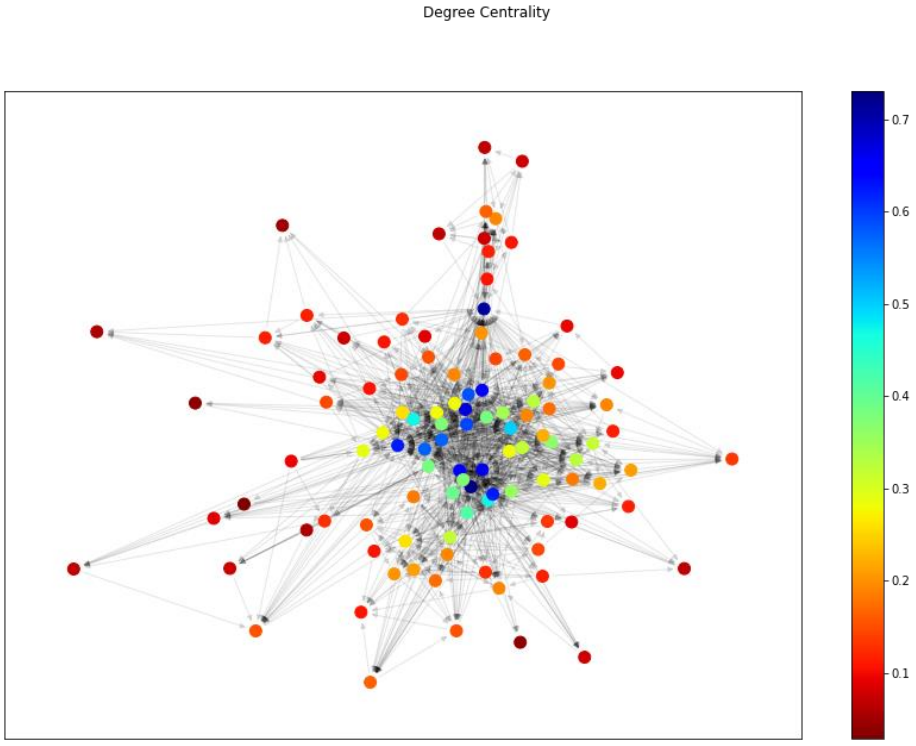
4.4.2 Degree Centrality



Picture 14. Betweenness Centrality



Picture 15. Closeness Centrality



Picture 16. Degree Centrality