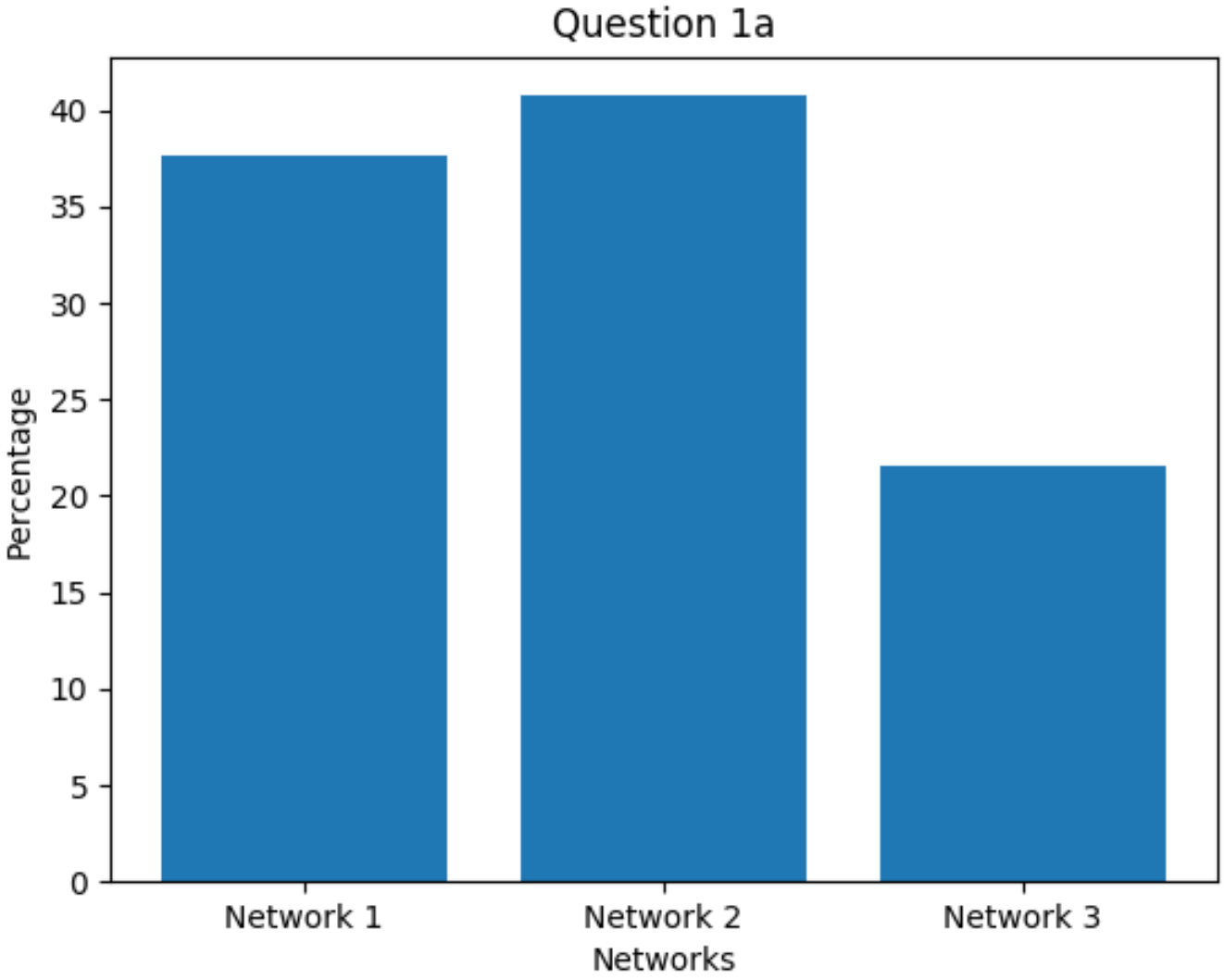
PSOSM - Quiz 3

Harkishan Singh (2017233)

Question 1

1)



2)

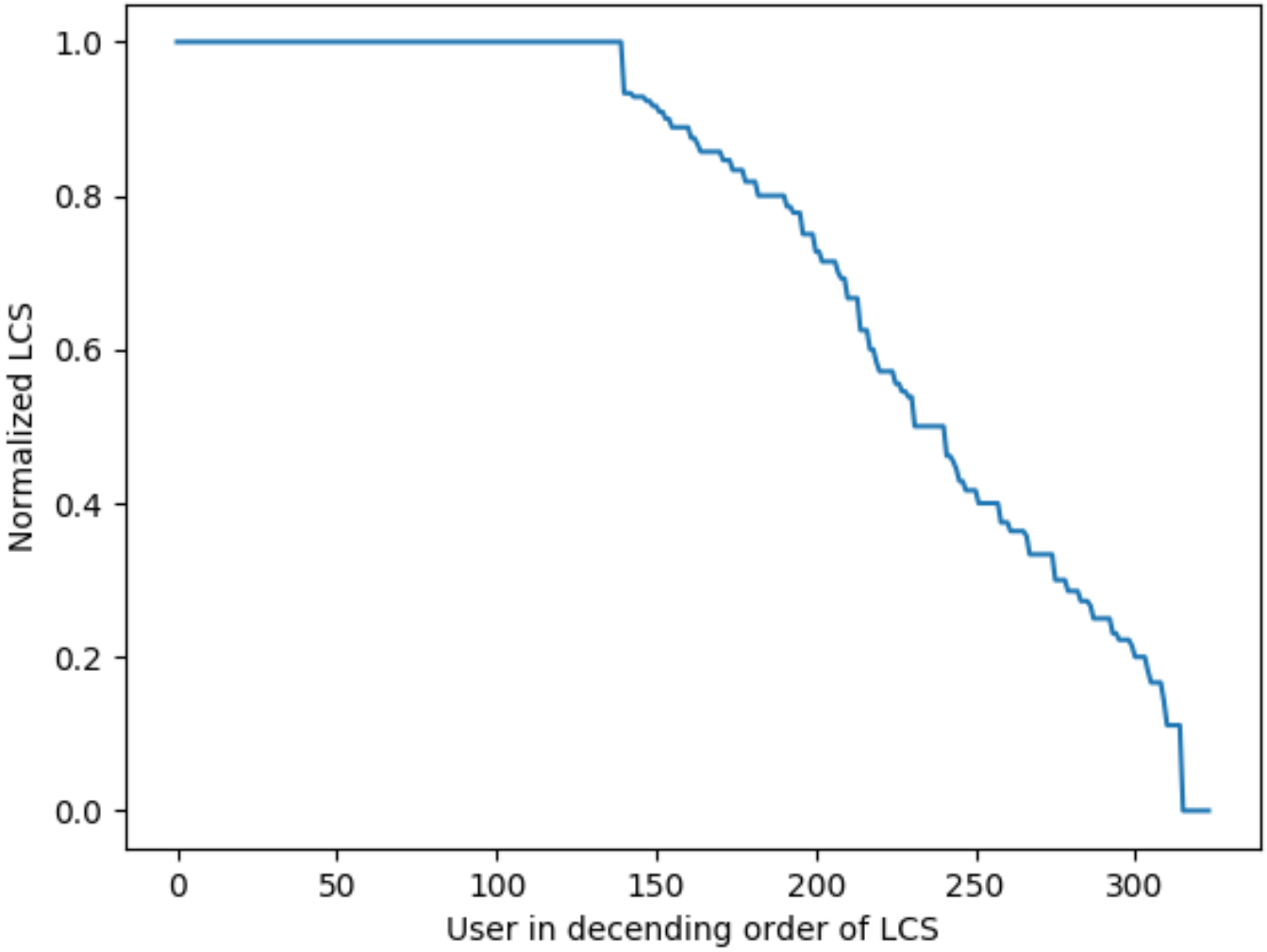
A)

```
['Mattan Griffel', 1.0]
['HerArt SheLoves', 1.0]
['Adn Montalvo Estrada', 1.0]
['Heidi Schnenberg-Hausdorf', 1.0]
['Tim Brauhn', 1.0]
```

Above screen shot have name of the user and his/her normalised LCS score.

I don't think that it is the best way to find similarity because, in this case it might happen that 2 user names are completely different but they appear at the top position due to subsequence property. I think the best way to measure similarity would be **substring** because it does not have continuous/ alternative alphabets count.

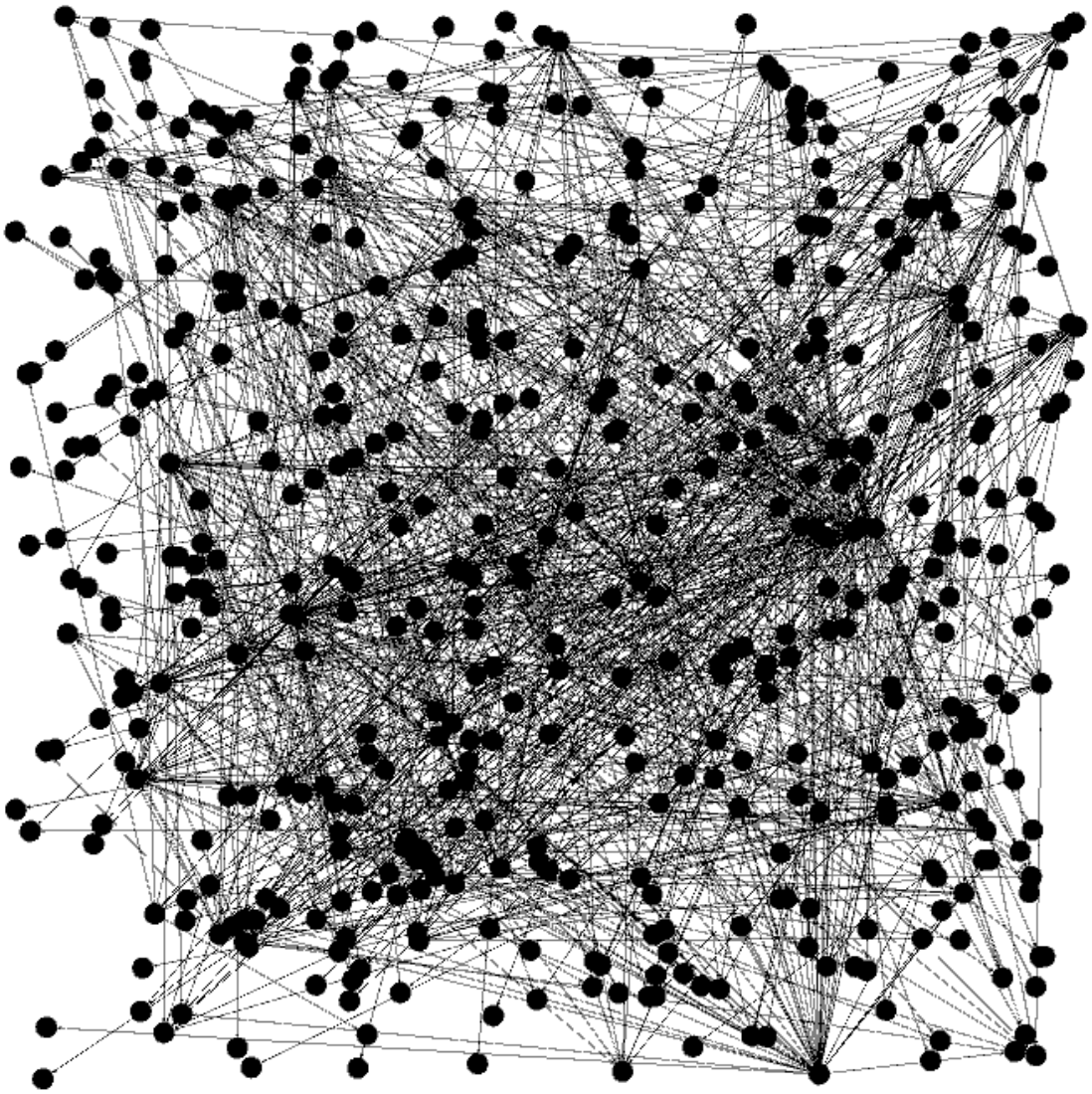
B)



X -axis have the users and Y axis have normalised LCS. Almost 150 users have LCS of 1.

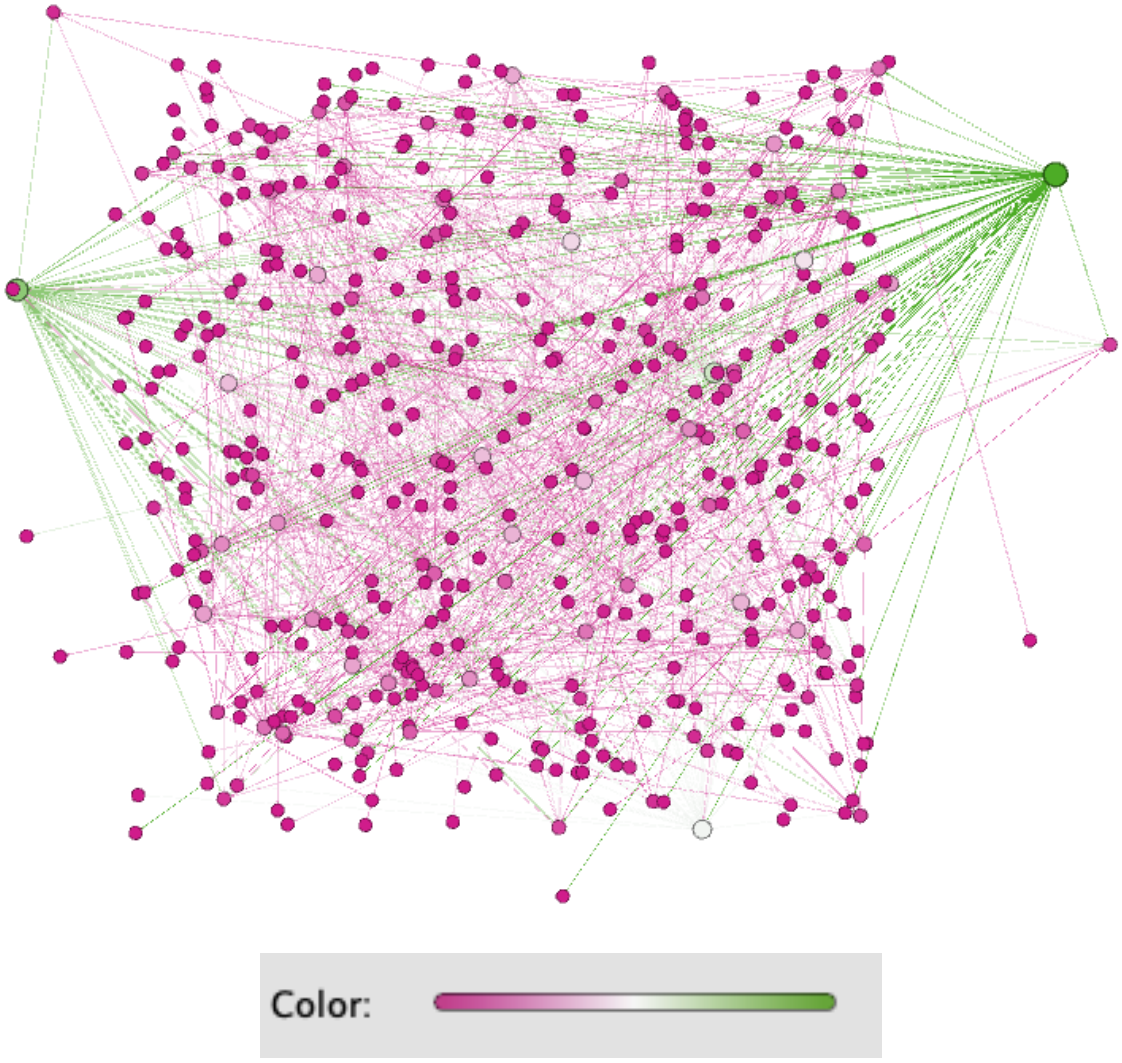
Question 2 :

Plotting the graph of the data set built from gephi :



B)

According to the degree, I have plotted the graph below : (on next page)



Here the green colour represents very high degree and purple colour shows a very low degree of the node. A light colour shows moderate degree of a node.
So, basically there are only 2 nodes with very high degree compared to those of low degree.
Here is more sorted graph :

C) Following are some stats of **clustering coefficient** of each node :

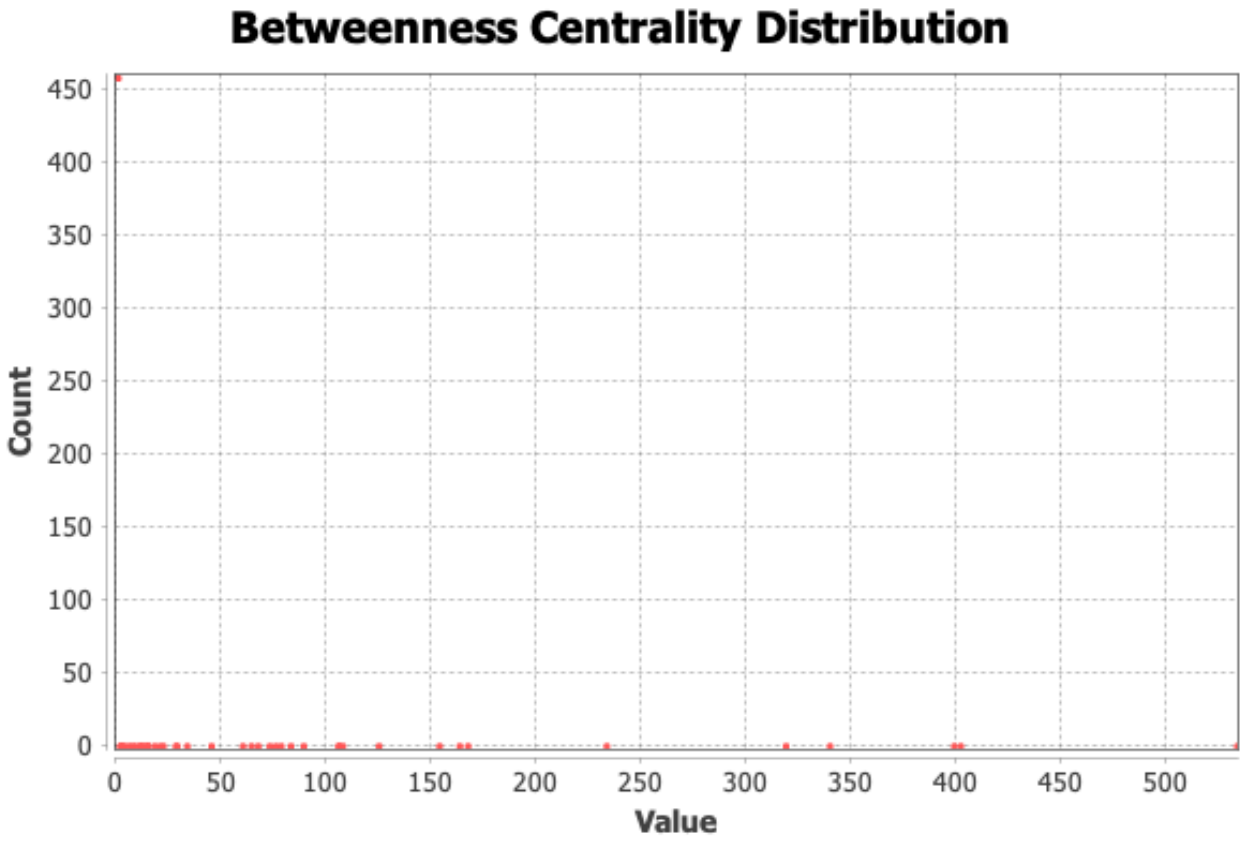
Results:

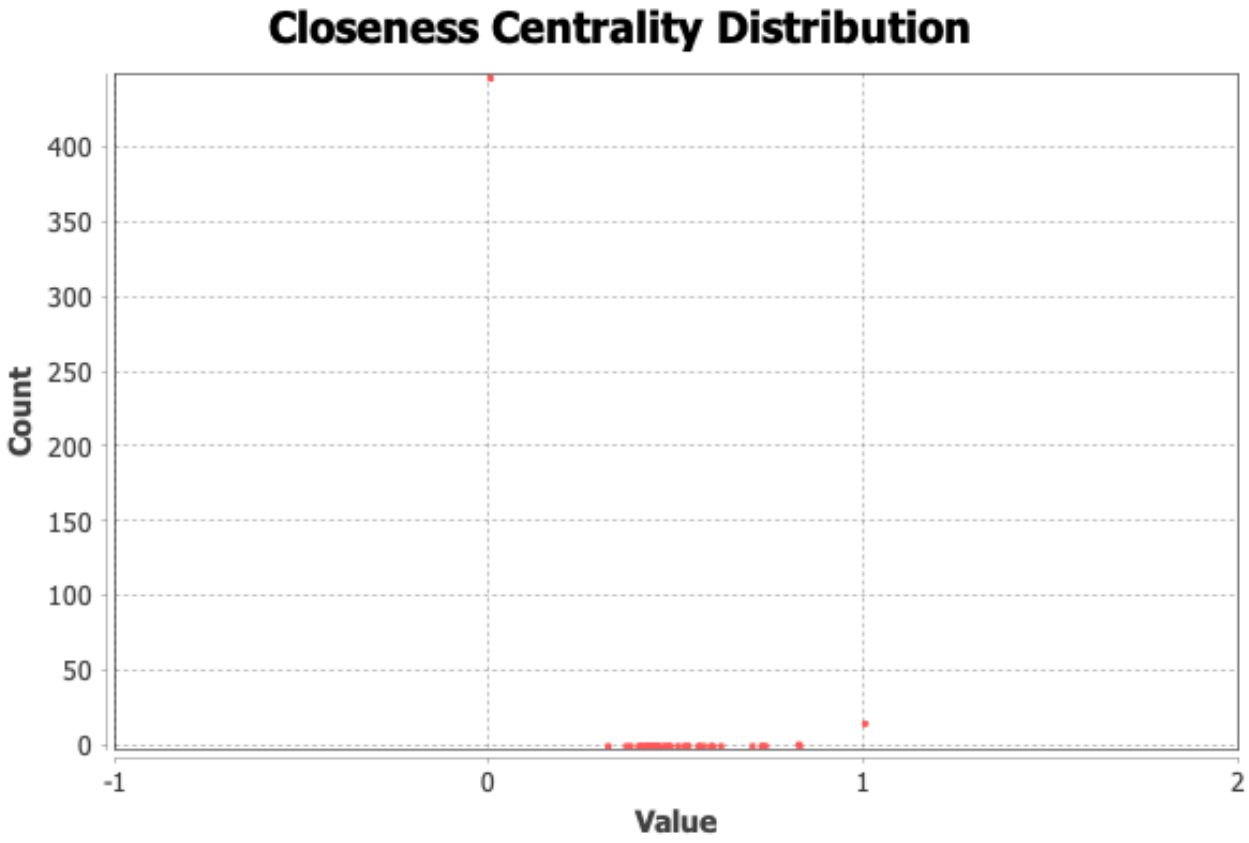
Average Clustering Coefficient: 0.340
Total triangles: 451
The Average Clustering Coefficient is the mean value of individual coefficients.

(You can look up for this in “Clustering coefficient” file)

Next page =>>

D)Here are the graphs of **Betweenness and Closeness centrality** of each node **:**





E) Diameter of the graph:

```
Diameter: 6
Radius: 0
Average Path length: 2.1354784081287046
```

You can lookup for “Network Diameter” for all the above stats provided.

F)

Looking at the above graph, it can be said that there are many paper which are published which has not cited any of the provided paper (which are represented by the leaf nodes in the network). It means that the particular paper is cited in any other paper but it has not cited any other paper.

The above statement holds true because of the fact that the average clustering coefficient is very low (=0.34). Papers are not very likely linked with each other.
Looking at the betweenness centrality graph, there are two nodes with value of over 400 which means that these paper have cited many other paper and these paper is been cited by many other paper. From this statement we can conclude that these 2 papers are very popular. (These are the only 2 nodes in 2nd graph I have provided).
Looking at closeness centrality, we can see that all nodes have almost less that 1 value which again means that graph is weakly linked.

Diameter of the graph is 6: which means that any paper in the network is linked with other paper in at most 6 paper in-between (including the 2 papers).