TEAM 9 – DJJM
SOCIAL COLLABORATION OF NETWORK SCIENTISTS

CNT5805 – Final Project Report

Section One: Project Summary

In order to achieve a general idea or to understand data, visualization of data is very important. When the data is visualized through networks, structures or graphical representation, it provides a better understanding of the information available. This research project provides a broader perspective of social collaboration between network scientists. The dataset utilized in the project to create a social collaborative network between scientists consists of information regarding scientists working on network theory and performing experiments related to the field of network science.

The key motivation for selecting social collaboration of network scientists is to understand the concept of social network analysis in real world scenarios and how an individual or community is connected with the other based on a common field. Through social networks, connections are made, analyzing a social network provides a broader perspective of how vast the network is and how much it elongates. Visualizing a social network provides comprehensive information on how a network works and how individual entities (nodes) form a network through connections (edges) with multiple individuals present in the network. The dataset contains names of the scientists who are authors of the research papers or a common experiment performed in the field of network science as nodes and the edges in the network represent if the scientists collaborated on the same research paper or experiment. The edges are connections through which the scientists (nodes) are connected with each other in the network if they make joint contributions in the field of network science. The source of the mentioned social collaboration dataset is as follows.

Data Source - http://vlado.fmf.uni-lj.si/pub/networks/data/collab/netscience.htm

The primary purpose to analyze the social collaboration dataset is to educate the reader about the structure of a social network and get insights on how the group of scientists collaborate with each other, perform different experiments and provide research in areas related to network science. The subject of network science is a vast area which comprises a lot of multiple fields. Numerous scientists and scholars work in this field and provide information to get a clearer picture of networks. How they collaborate with other network scientists as well to research on topics related to network, its integrations, formations, and varied areas within the field and what insights can be gained from this dataset, how network scientists interact, which countries do scientists belong to, their gender, their knowledge based on how much they have contributed to the network field, communities formed by specific scientists, number of scientists contributing maximum in this field etc. can be known through visualization of this network. Every network showcases a lot of insights through which various research questions can be answered. The following research questions are answered within the research project.

Research Questions

1. **Are there preferred scientists as well as outlier fewer collaborative scientists?**

   The number of scientists collaborating with maximum 1 scientist or scientists contributing individually for research in network science. As the network is a social network, the scientists who don't collaborate with other scientists or have minimal connections in the field are preferred to be outliers in the network.For example:- 40% of scientists from the whole network have minimal connections or contribute solely in the field.

2. **Are there any distinct social circles or communities that exist within the network? For instance, Are there any communities that are closely grouped and have maximum output for the networking field?**

   For instance there is a specific community of scientists which has collaborated with maximum scientists, or there is a community with maximum contribution in the network field. Communities are formed in the network when a specific scientist has connections with other respective scientists which in turn creates a social network, formed for sole purpose for making contributions in the field of network science.

3. **Are there any social hierarchies within the scientists' collaboration? Are there any Systems of social organization in which some individuals enjoy a higher social status than others? Can we find if there's a repeating theme with one individual within some groups, then that person is higher on the social hierarchy?**

   An individual scientist who possesses a higher social status is the one with maximum connections and the one who is responsible for forming major communities within the network. That scientist is higher in social hierarchy and is considered to be the leader of the organization, the one who has collaborated with maximum other respective scientists and provided research and innovations in the field of network science.

4. **Which country of origin for the scientist provides maximum insight into the field of network?**

   The maximum number of scientists belonging to a specific country. The countries who provide maximum contribution in the field of network science. This can help to discern which country has maximum resources and are willing to contribute in plenty to the network field. For example - 30% of the scientists in the network belong to the United States of America.

5. **Which is the predominant gender of the scientists that contributes the most to the networking field?**

   The gender of scientists, male or female, which one of them is predominant and provides more contribution in the field of network science. For example - around 70% of scientists contributing to the network field are females or males.

6. **What are the hubs of scientists who have contributed the most in the field of network science?**

   Scientists forming larger hubs in the network, the one with common connections. The scientists form enclosed communities within the network, having fewer connections with other respective scientists outside of their community in the network. For instance there is a specific group of scientists or a scientist who collaborates with scientists of his/her own community or only with the scientists of the same country. This results in the formation of a hub within the network.

7. **Are there any scientists who don't collaborate with anyone and provide their own research in the field of network?**

   The scientist who does not collaborate with other respective scientists. They contribute solely in the field of network science. For example - 10% of the scientists in the network contribute individually in the field. Maximum scientists who contribute individually are either male or female.

8. **Who are the most influential scientists in the network?**

   The scientists having the most connections, higher weighted degrees within the network are considered to be the most influential scientists. The scientists who have contributed to more than 10 research papers or experimental studies by collaborating with other respective scientists are considered to be influential.

9. **Do scientists from cross background co-author a paper together?**

   Scientists belonging to different countries and collaborating to provide their research and innovations in the field of network science. The cross background characteristic can be discovered within larger communities where there are maximum scientists

Section Two: Initial Graph

The graph of the network contains 1589 nodes and 2743 edges and it is an undirected network. The network possesses weighted edges.

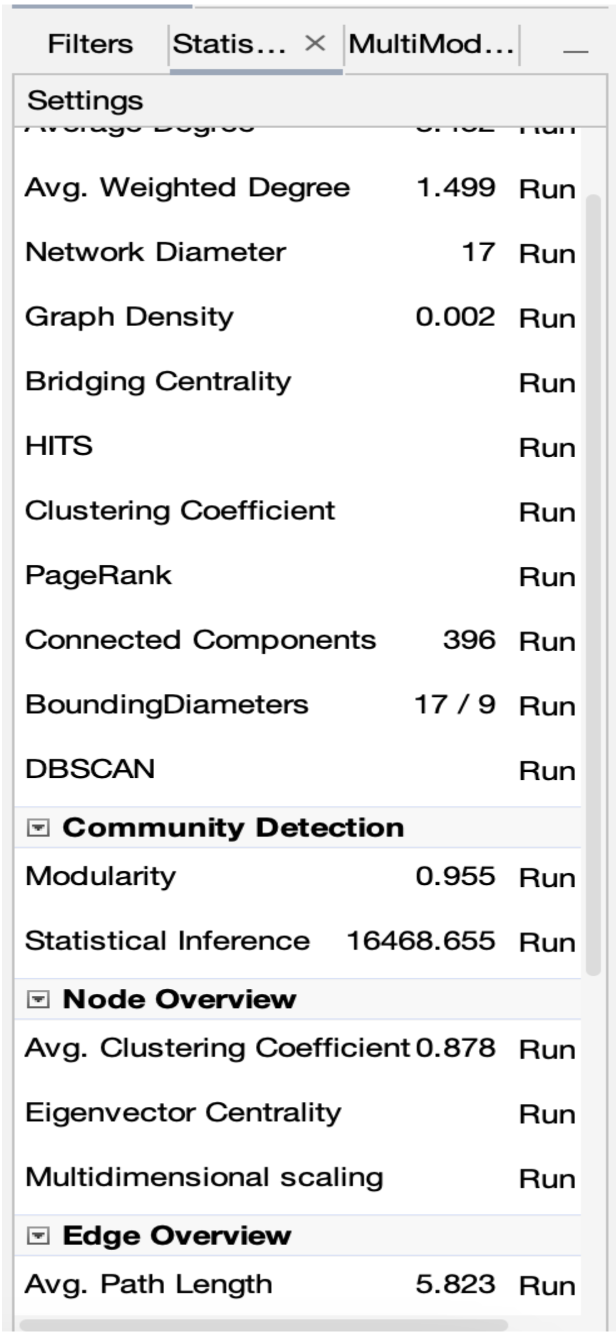

Figure 2.1

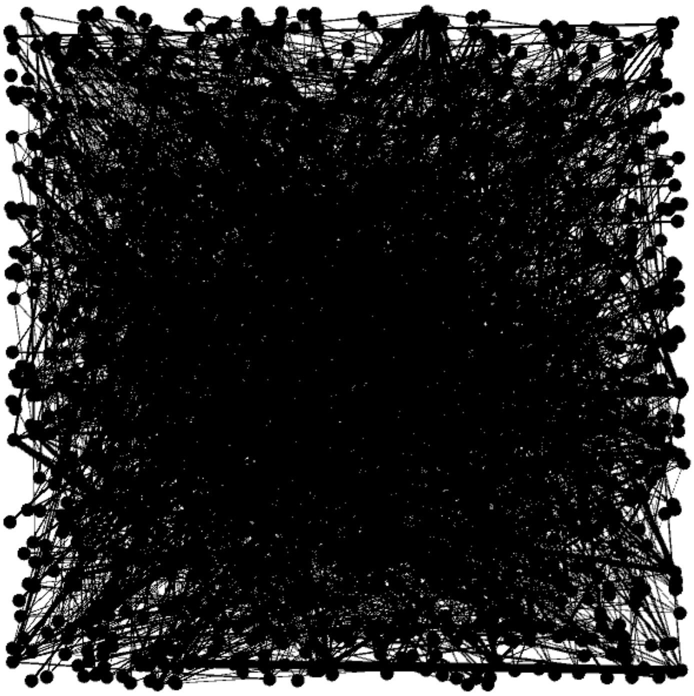

Figure 2.2

The graphical network obtained on importing the .gml file of social collaboration network scientists is an undirected and a scale-free network. The network consists of 1589 nodes and 2743 edges. The nodes of the network represent the names of the scientists who are authors of research papers and who perform experiments in the field of network science, and the edge represents the names of scientists who contribute together or co-author on the same research paper. Initially, the graph appears juxtaposed as it contains names of numerous scientists and doesn't display any specific hub or formation of communities.

To discover the hubs and communities in the network, certain filters including texts, layout algorithms and color coding can be employed in order to recognize them correctly. The network being vast would display formation of communities as it is a social network. Network density is important as it reveals how connected the network is compared to how connected the network might be and it ranges from 0 to 1. The density of the social collaboration of network scientists network is low (0.002) as the number of nodes and edges are relatively high which would result in maximum connections.

Network density can be calculated using the total number of edges divided by the total number of possible edges. This is equivalent to dividing the cardinality by the network size. This is shown by the equation **m/(n(n-1)/2)** where **m** is equal to the number of edges and **n** is equal to the number of nodes. Therefore the Network Density is equal to 2743 / (1589(1589-1)/2) = 2743 / 1261666 = 0.002

Section Three: Layout Algorithms

Layout One: Dual Circle

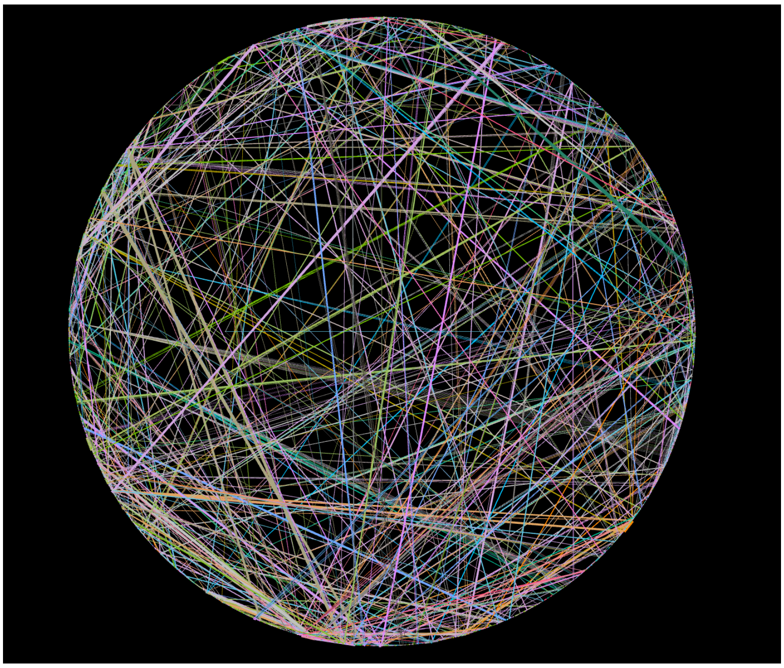

Figure 3.1 Dual Circle Layout Algorithm

When applying the **Dual Circle Layout Algorithm**, the nodes of the respective social collaboration network are arranged in a circular order based on their ID. The network can also be modified by altering the layout algorithms and sizing the nodes based on metrics like degree, betweenness centrality etc.

Layout Two: Fruchterman Reingold

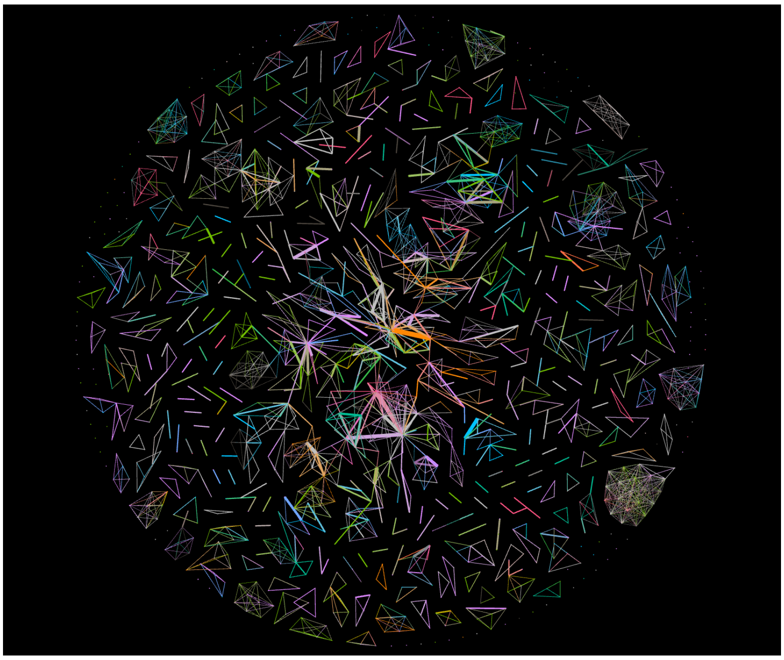

Figure 3.2 Fruchterman Reingold Layout Algorithm

When applying the **Fruchterman Reingold Layout Algorithm**, a circular spiral network is formed considering the tension or force between two nodes in the network. The network is stabilized by altering the nodes and modifying the force between them. By modifying area, gravity or speed in “Gephi” the graph can be stabilized.

In the network of social collaboration of network scientists, the nodes (scientists) with maximum connections with other respective scientists have maximum tension or force between them when they work on the same paper or study. It depicts that these nodes have maximum members under their community and are in control of flow of information within their community. The algorithm shows formation of various clusters and also multiple disjointed nodes revealing that many scientists do not possess connections (edges) with other scientists.

Layout Three: Force Atlas Layout Algorithm

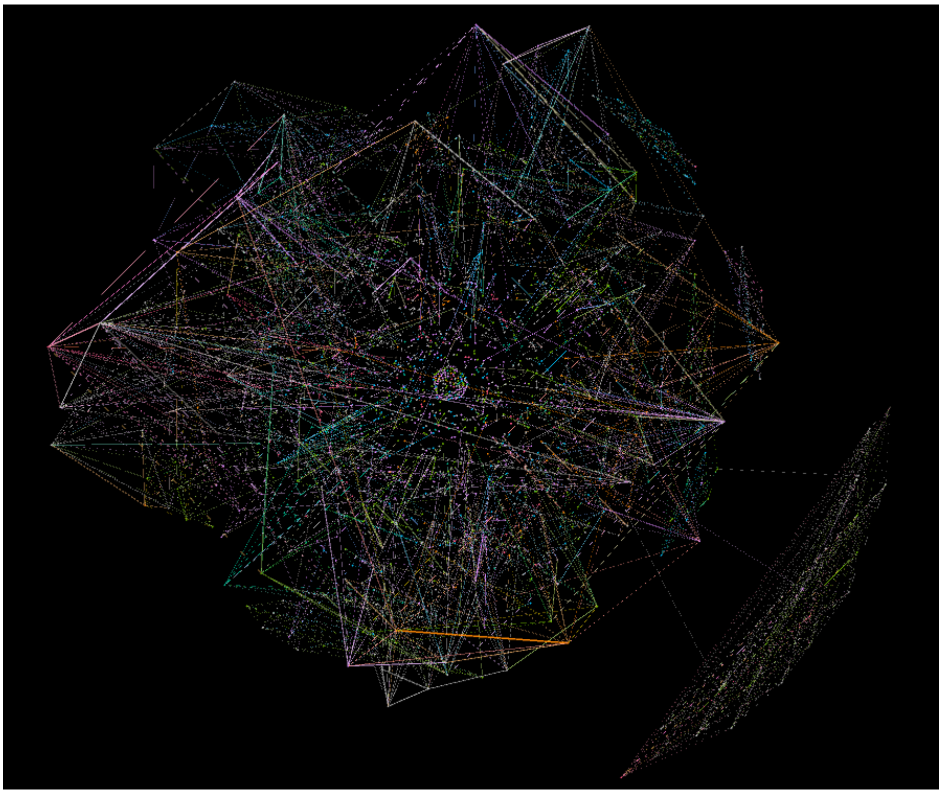

Figure 3.3 Force Atlas Layout Algorithm

Force Atlas Layout Algorithm, on its application, attracts nodes with more connections with each other and the nodes with less connections are repulsed. In the network of social collaboration of network scientists, the scientists with maximum connections and who contribute most in the networking field appear together and scientists with fewer connections and almost no connections appear at the center of the network, juxtaposed and totally isolated from the rest of the communities of the network.

Layout Four: Isometric Layout Algorithm

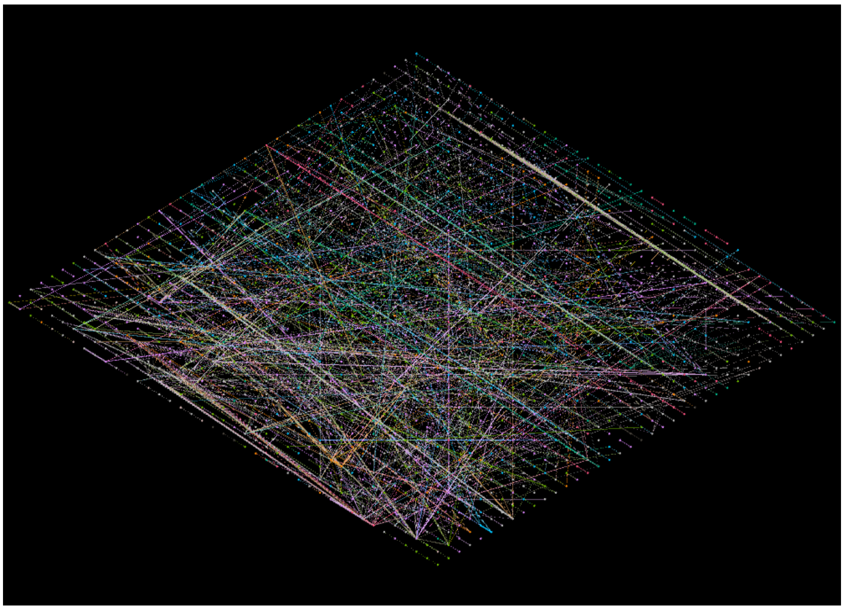

Figure 3.4 Isometric Layout Algorithm

When applying **Isometric Layout Algorithm**, 3D coordinates of the network nodes are generated. Visual representation of three-dimensional nodes in two dimensions is formed in the network. The network is divided into different Z layers which are used for partition and ranking in the network. The value of “Z” can contain any characteristic from a network like node Node ID, degree, betweenness centrality etc. On modifying the Z-maximum level and Z-distance stability of the network can be achieved. The scientists (nodes) with fewer connections are organized on the border of the network and as the network’s depth increases, the scientists (nodes) with maximum connections are visible and a layered model of the network is formed.

Layout Five: Circle Pack Layout Algorithm

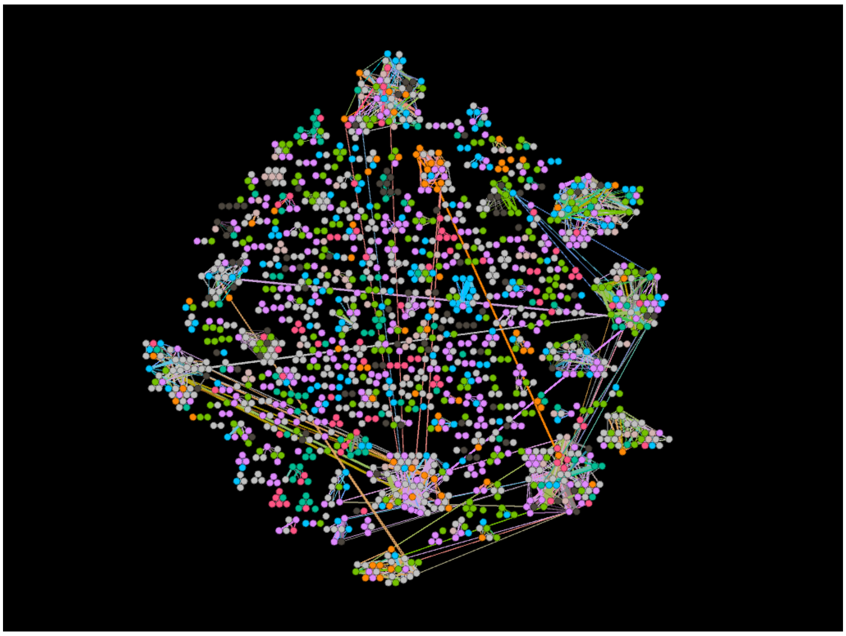

Figure 3.5 Circle Pack Layout Algorithm

Circle Pack Layout Algorithm, on its application groups the nodes based on its attributes in a circular pattern. On applying this algorithm, clear communities in the network are visible which makes the network more manageable and navigable. A varied social network of network scientists can be discerned based on the formation of communities and altering the network based on different hierarchical characteristics would form a more profound network.

At this point, the **Circle Pack Layout Algorithm** seems more convincing in visualizing the vast social network of scientists. As the algorithm arranges nodes based on its attributes and hierarchical characteristics. Multiple characteristics can be considered and the network can be modified to make it stable and readable to the user.

Circle Pack Layout Algorithm projects out the communities in any vast social network in a much more simple way as it allows modification of the network based on statistical and hierarchical characteristics.

Section Four: Graph Visualization

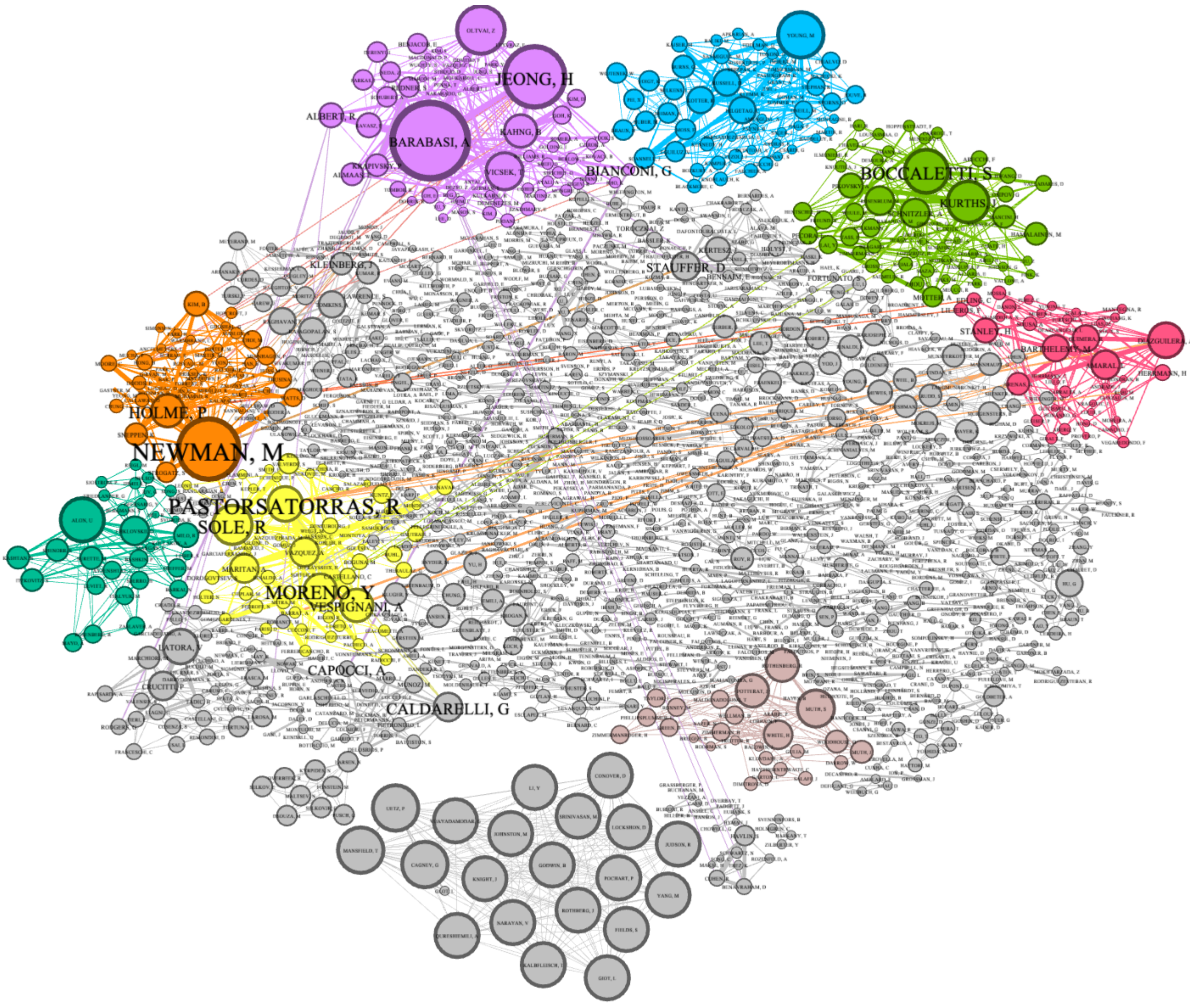

Figure 4.1 Modified Network

The latest network obtained after altering the characteristics of the network shows formation of communities. The scientists (nodes) with maximum collaborations are visible based on sizing of nodes and connections with distant scientists (nodes). The network obtained is not a single connected network and has many cliques and triads as well.

The nodes are sized as shown in the figure Figure 5.2 based on the degree of the node ranging from 5 to 130. Some of the influential scientists with maximum collaboration include Barabasi A, Jeong H, Newman M, Pastorsatorras R, Bocalletti S etc. in the network.

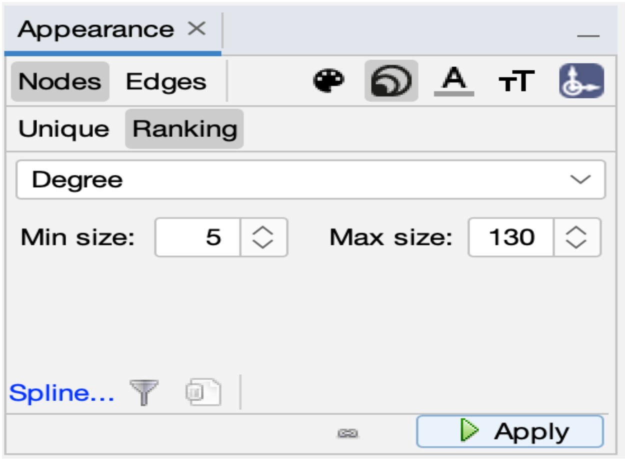

Figure 4.2 Size Modification based on degree

The network is color coded and partitioned based on the modularity class of the network. The network contains 9 different communities when partitioned with a modularity class. The modification based on modularity class is as shown in the figure 5.3.

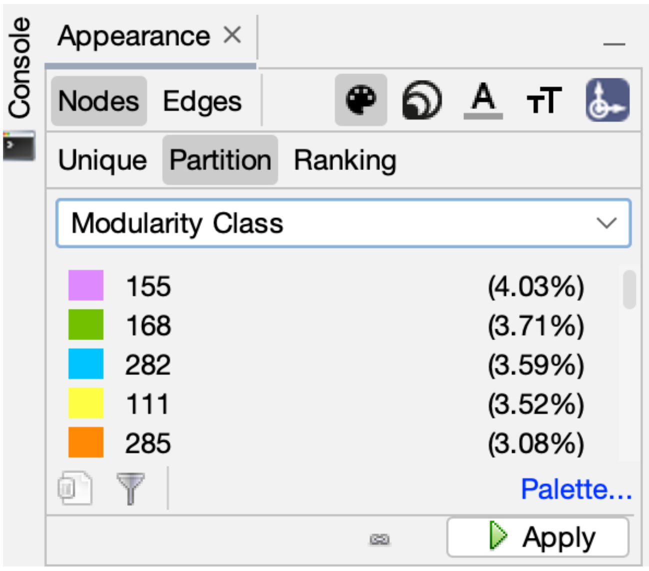

Figure 4.3 Partitioning nodes based on modularity class

The social collaboration of network scientists network is a vast social network with high modularity indicating dense connections with the scientists in the respective communities but sparse connections when it comes to collaborating with the scientists belonging to different communities.

Section 5: Statistics

Average Degree:

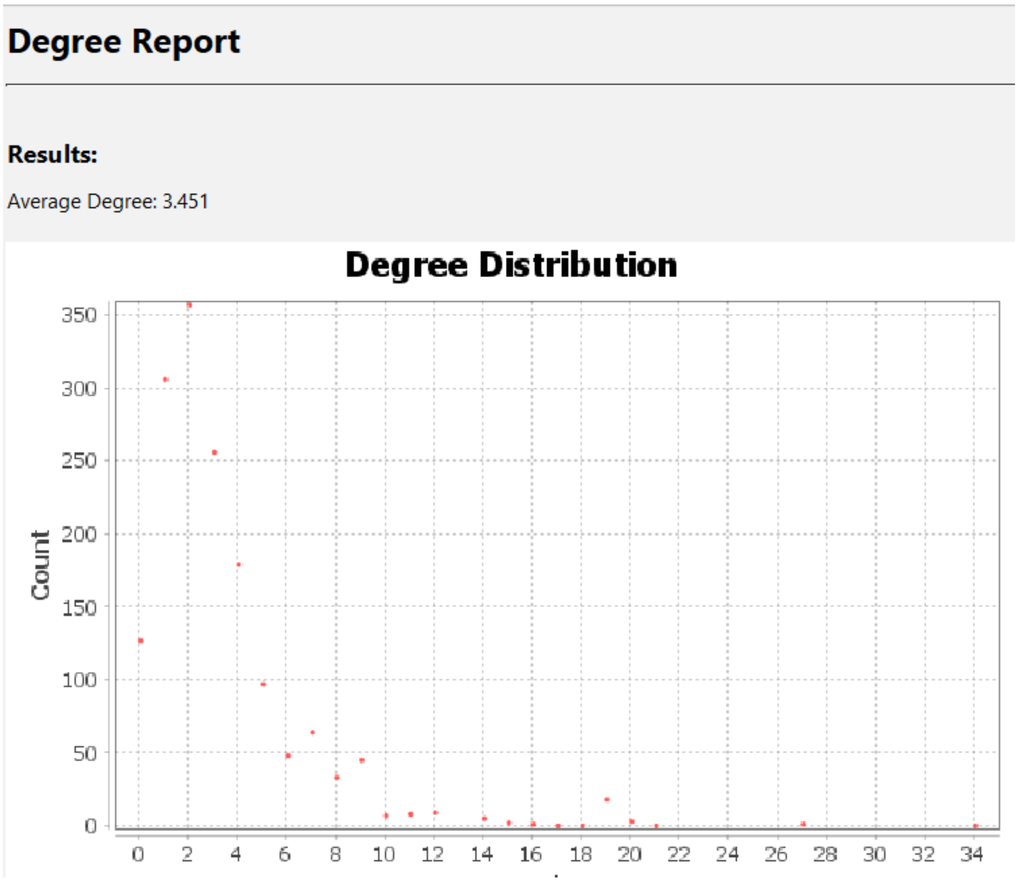

Figure 5.1 Degree Distribution Report

The average degree report shows that each node has on average a degree of 3.451. This means on average each scientist is connected to at least 3 scientists. The degree range is very useful as it can show the amount of collaboration within the network. For example- one could filter the degree range to 0 and see which scientists did not collaborate and had independent research.

Graph Density:

HTML Report

Graph Density Report

Parameters:

Network Interpretation: undirected

Results:

Density: 0.002

Figure 5.2 Graph Density Report

The density of the graph is how close the network is to completion. The calculated density is 0.002. This means that the network is sparse.

Clustering:

Results:

Average Clustering Coefficient: 0.878
Total triangles: 3764
The Average Clustering Coefficient is the mean value of individual coefficients.

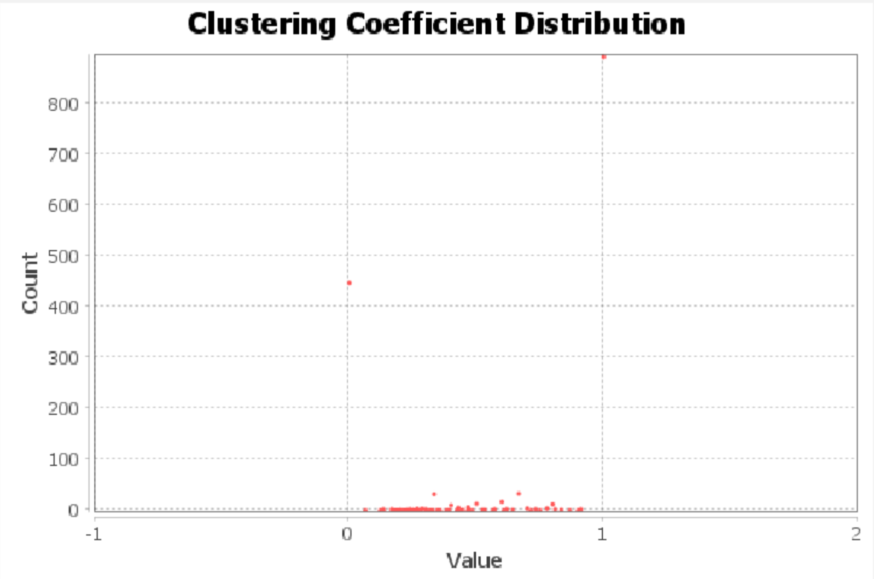

Figure 5.3 Clustering Coefficient Report

The cluster coefficient measures which nodes will likely form clusters.The cluster coefficient ranges between 0 and 1. According to figure 5.3 ,the average cluster coefficient for each node is 0.878. This means within this network of scientists, there will likely be the formation of communities within the network.

Betweenness Centrality:

Id	Label	Interval	gender	country	giantcomponent	Eccentricity	Closeness Centrality	Harmonic Closeness Centrality	Betweenness Centrality	Degree	Eigenvector Centrality	Clustering Coefficient	Number of triangles
78	NEWMAN, M		male	USA	10	9.0	0.256619	0.331917	28300.564474	27	0.073739	0.062678	22
150	PASTORSATO...		male	ITA	10	11.0	0.247059	0.310298	24592.767656	15	0.059112	0.171429	18
281	SOLE, R		male	USA	10	10.0	0.249012	0.304821	19249.904317	17	0.050428	0.169118	23
216	BOCCALETTI, S		male	ITA	10	13.0	0.187686	0.25601	18200.0	19	0.061566	0.181287	31
34	JEONG, H		male	JPN	10	10.0	0.229648	0.314303	17858.002513	27	0.16835	0.182336	64
756	HOLME, P		male	NED	10	9.0	0.243087	0.30442	16506.036277	14	0.063261	0.21978	20
301	CALDARELLI, ...		male	USA	10	10.0	0.232902	0.277717	15786.012583	12	0.034587	0.227273	15
203	CAPOCCI, A		male	ITA	10	9.0	0.216619	0.250832	12446.912583	6	0.017801	0.4	6
151	VESPIGNANI, ...		male	ITA	10	11.0	0.230628	0.288913	11143.696397	14	0.050616	0.197802	18
33	BARABASI, A		male	HUN	10	10.0	0.213318	0.303135	10834.47342	34	0.182854	0.13369	75
46	STAUFFER, D		male	BEL	10	10.0	0.216371	0.271602	10575.906104	12	0.028681	0.121212	8
219	KURTHS, J		male	IRE	10	14.0	0.162791	0.228921	8911.833333	18	0.071707	0.267974	41
30	ALBERT, R		male	USA	10	10.0	0.211765	0.272847	6516.850269	10	0.07286	0.266667	12
307	KLEINBERG, J		male	GER	10	10.0	0.20884	0.25724	6203.0	9	0.039327	0.333333	12
72	STANLEY, H		male	USA	10	10.0	0.212479	0.250698	5154.06655	10	0.025674	0.266667	12
654	KAHNG, B		male	KOR	10	11.0	0.192759	0.252845	4987.970474	14	0.05335	0.230769	21
327	LATORA, V		male	NED	10	11.0	0.190237	0.248744	4601.692492	15	0.046138	0.257143	27
53	VICSEK, T		male	SRB	10	11.0	0.194344	0.262879	3950.075974	16	0.111044	0.333333	40
184	BENNAIM, E		male	USA	10	13.0	0.145497	0.181	3321.0	4	0.010114	0.333333	2
697	SCHNITZLER, ...		male	GER	10	15.0	0.141467	0.192462	3321.0	12	0.049387	0.469697	31
596	FORTUNATO,...		male	ITA	10	10.0	0.195956	0.233578	3249.606104	3	0.014198	0.333333	1
55	KRAPIVSKY, P		male	RUS	10	12.0	0.16875	0.222446	3051.676079	9	0.025948	0.222222	8
51	ALMAAS, E		male	CYP	10	11.0	0.181556	0.23923	3000.763153	9	0.057863	0.25	9
473	MOTTER, A		male	ENG	10	15.0	0.14094	0.185187	2966.166667	8	0.021179	0.285714	8
757	EDLING, C		male	NED	10	9.0	0.219767	0.255296	2858.424883	5	0.019997	0.7	7
758	LILJEROS, F		male	DEN	10	9.0	0.219767	0.255296	2858.424883	5	0.019997	0.7	7
121	SNEPPEN, K		male	ENG	10	10.0	0.20667	0.251068	2267.214286	9	0.024704	0.194444	7
136	PECORA, L		male	MEX	10	14.0	0.158824	0.200217	1870.0	7	0.015255	0.238095	5
303	MUNOZ, M		male	ESP	10	10.0	0.202464	0.232652	1869.0	7	0.014433	0.285714	6

Figure 5.4 Network Data Table

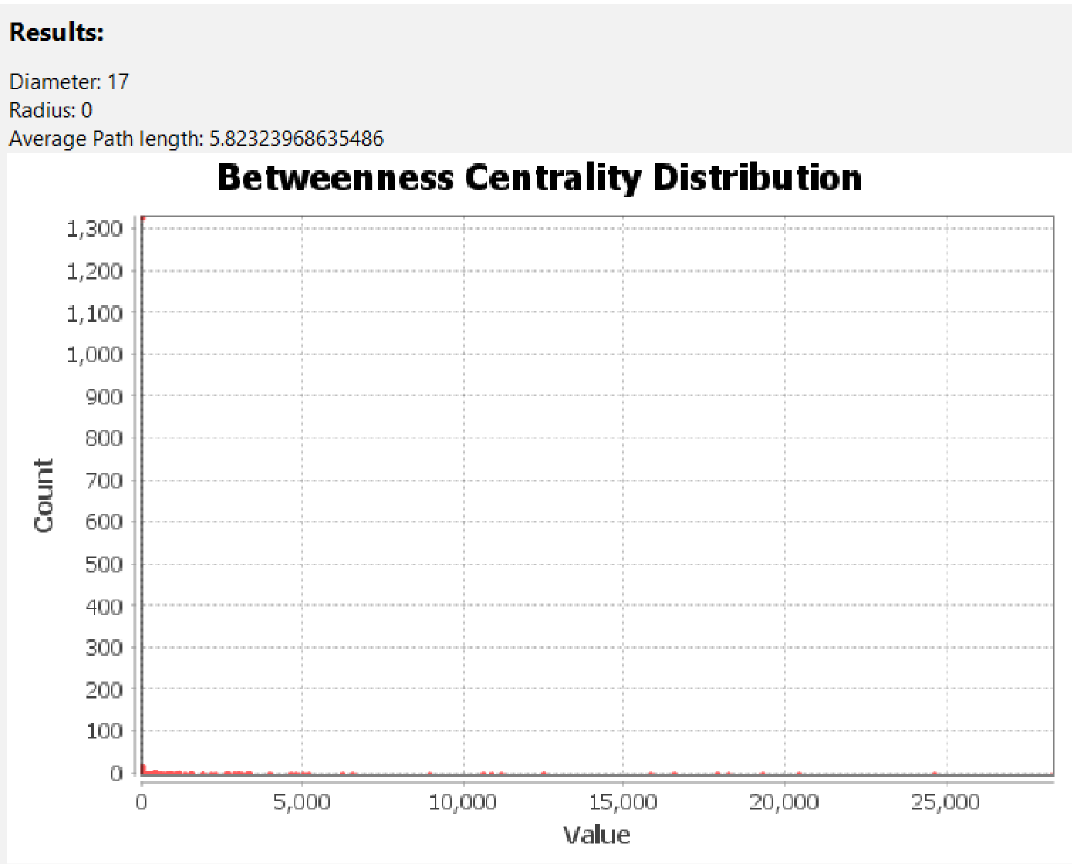

Figure 5.5 Betweenness Centrality Distribution Report

Betweenness centrality shows the centrality of a node in the network. It calculates the number of times a node lies on the shortest path to other nodes. A node with a high betweenness centrality is a node that is very influential within the network. According to the graph in Figure 5.5, the Betweenness Centrality was filtered in Descending order to show scientists with the most influence within the network. Scientist Newman has the highest betweenness centrality which shows he is the most influential scientist within the network.

Closeness Centrality Distribution:

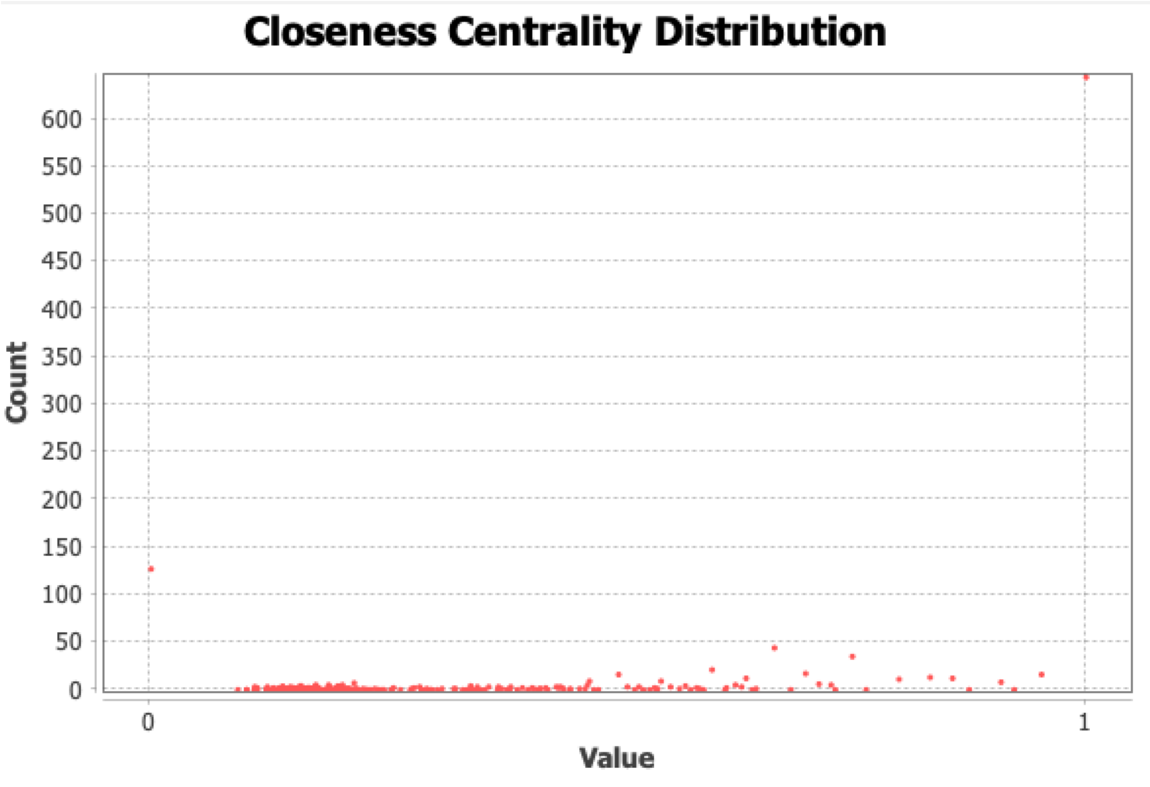

Figure 5.6 Closeness Centrality Distribution Report

Average shortest distance from one node to another is obtained by closeness centrality distribution and its value range is from 0 to 1. The distribution plot of closeness centrality shows the scientists who have maximum collaborations and their count.

Harmonic Closeness Centrality Distribution:

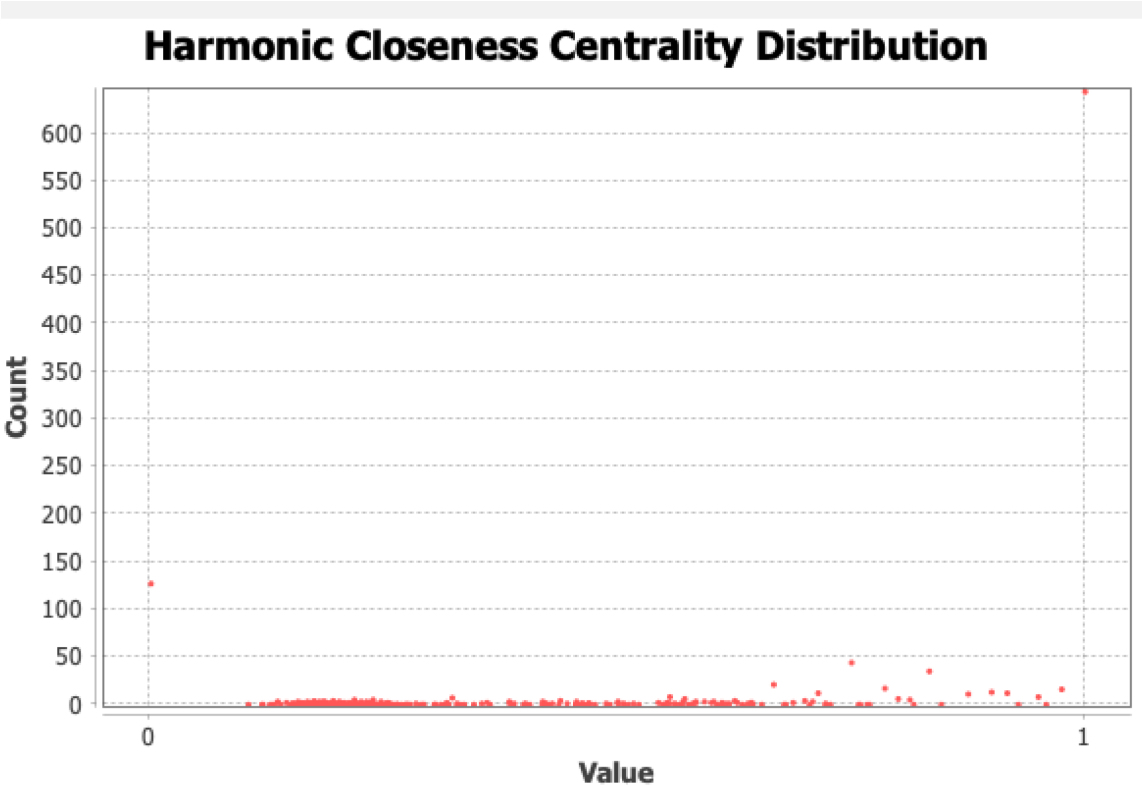

Figure 5.7 Harmonic Closeness Centrality Distribution Report

From harmonic closeness centrality, size of the node can be altered. To analyze the graph correctly so that all nodes are visible, harmonic closeness centrality is used. The values of nodes with maximum connections/edges are increased and based on that the community structure of the network can be visualized. In the social collaboration network, as visualized using "Gephi" there are many nodes (scientists) with maximum collaborations whose harmonic closeness centrality can be increased to make the network more visually understandable.

Eccentricity Distribution:

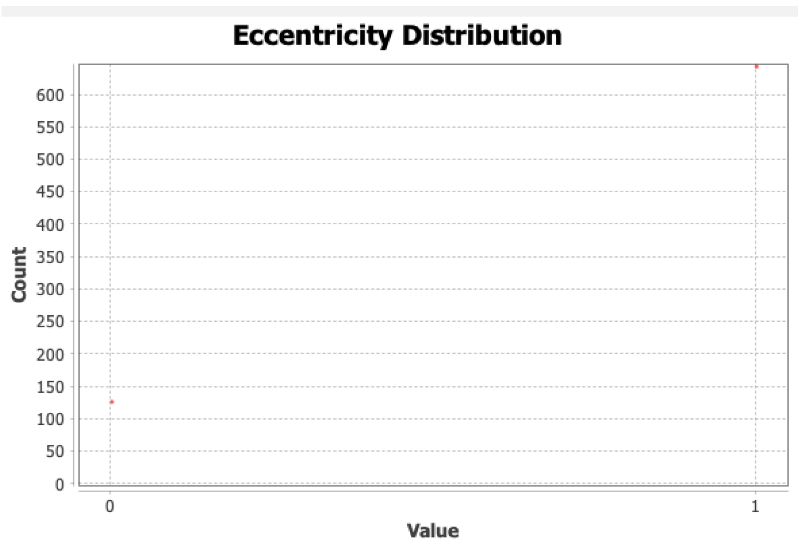

Figure 5.8 Eccentricity Distribution Report

The Eccentricity Distribution shows the maximum distance from the node having maximum connection to the node having least connections. In the social collaboration network, the maximum distance between the most influential scientist to the least influential scientists ranges from approximately 125 to 700 (more than 700) as shown in Figure 5.8.

Modularity:

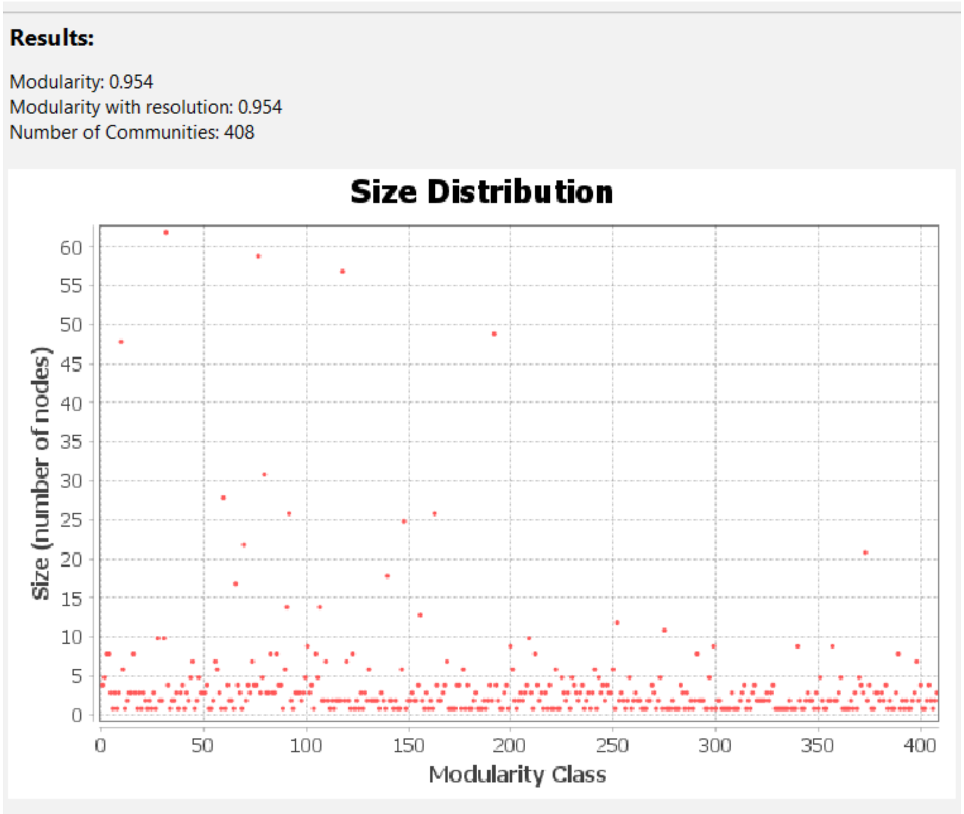

Figure 5.9 Modularity Class Report

Modularity of the network can be achieved based on the factor of total communities created within the network. Modularity size distribution segregates the nodes into clusters based on their importance in the network through percentage substitution. According to Figure 5.6, there are about 408 communities formed within the network.

Section Six: Filtered Network

Giant Component Filter:

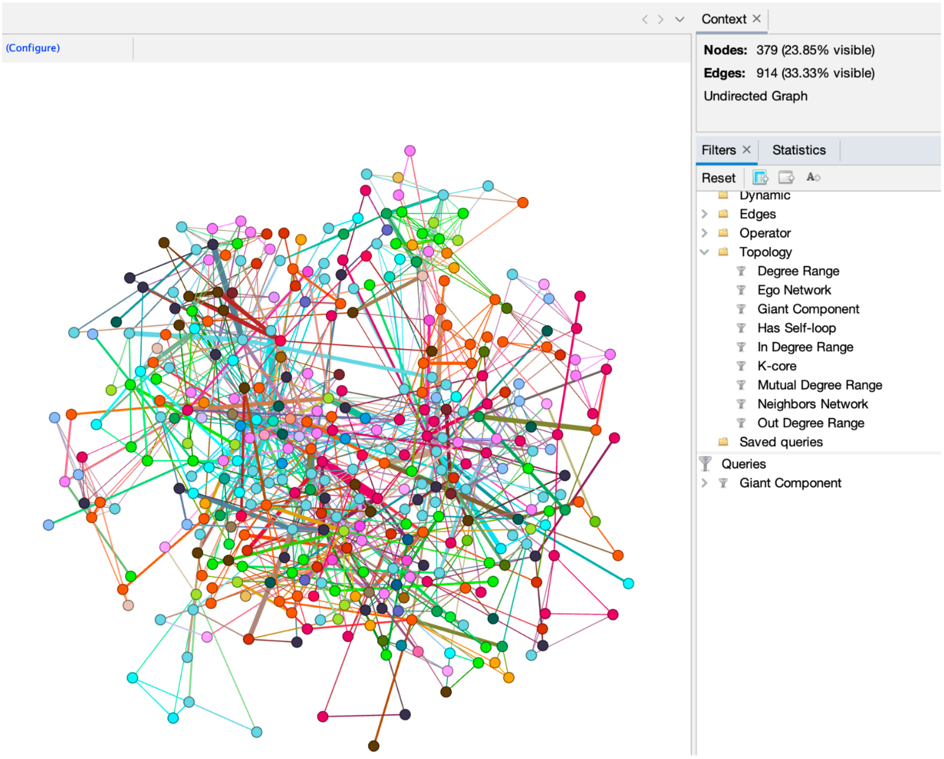

Figure 6.1 Giant Component Filter

Applying the giant component filter, the nodes that contain a large proportion of the edges in the network are visible. This filter removes close to three quarters of the nodes, and about two thirds of the edges. With these statistics, it can be inferred that not many network scientists collaborate very closely.

Ego Network Filter:

Ego Network filter reveals total connections of any respective node based on its ID. Total connections of some of the influential scientists are taken into consideration as shown in the figures below:

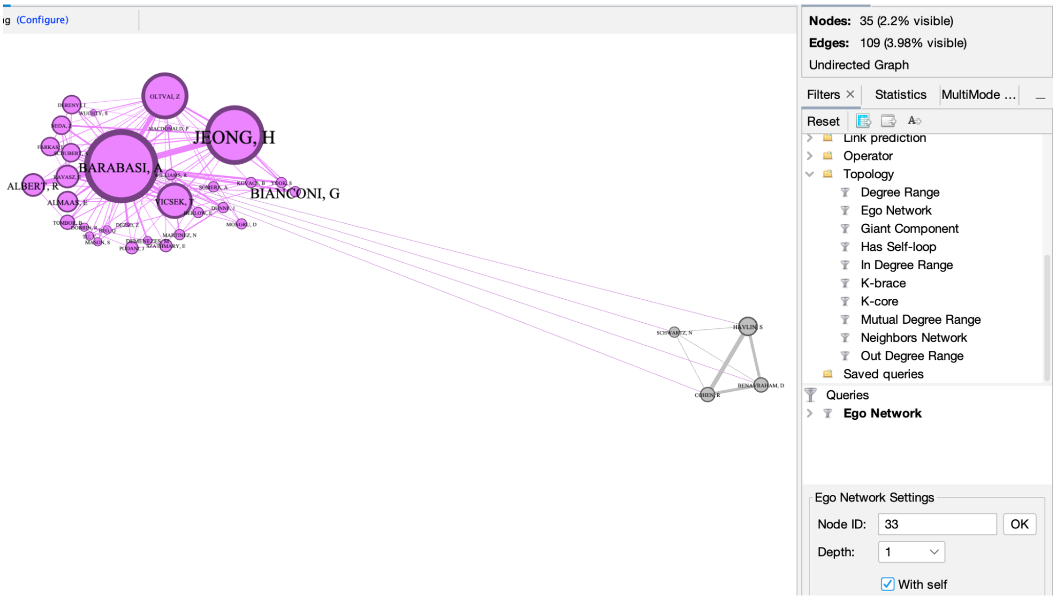

Figure 6.2.1 Scientist Barabasi, A collaborations

Scientist Barabasi, A possesses 34 (35-1 = 34 i.e. considering scientist Barabasi, A himself as a node) connections with the respective scientists, stating that he has collaborated with 34 different scientists and provided research in the field of network science.

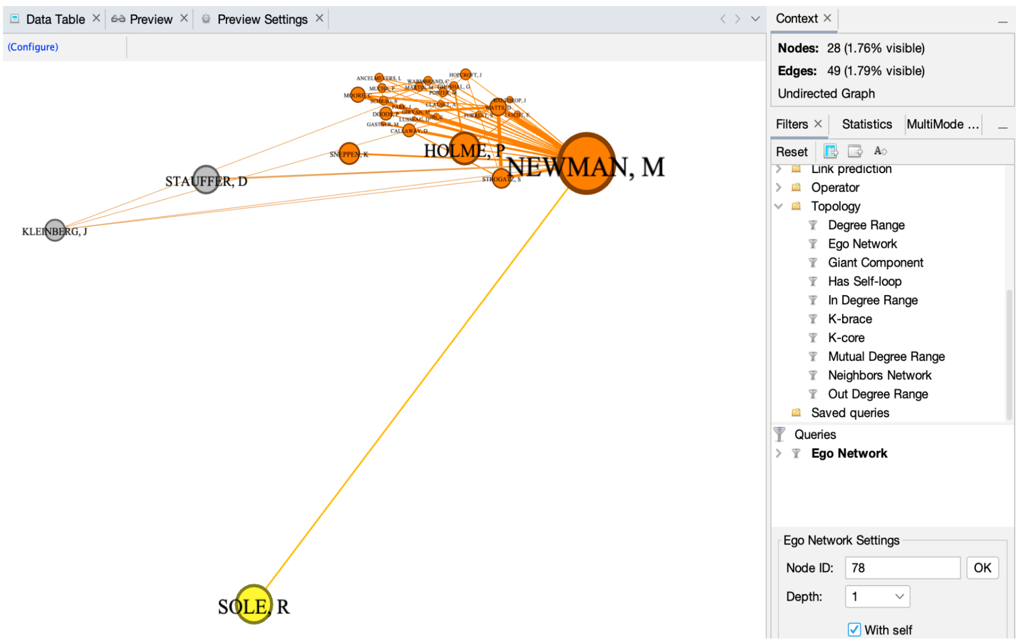

Figure 6.2.2 Scientist Newman, M collaborations

Scientist Newman, M possesses 27 (28-1 = 27 i.e. considering scientist Newman, M himself as a node) connections with the respective scientists, stating that he has collaborated with 27 different scientists and provided research in the field of network science.

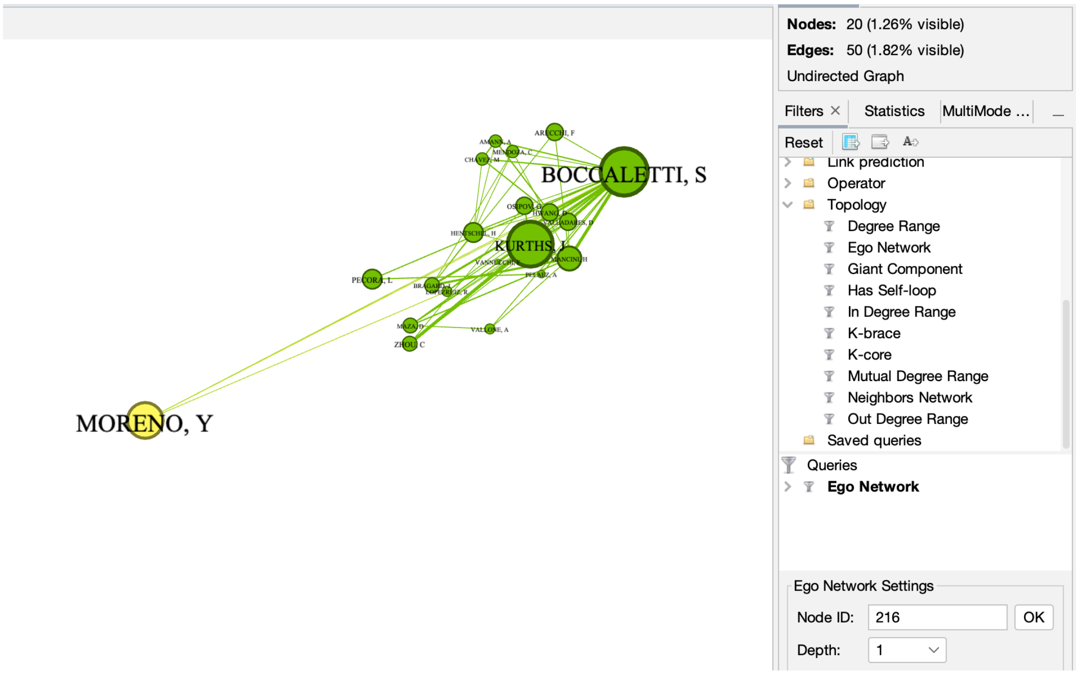

Figure 6.2.3 Scientist Boccaletti, S collaborations

Scientist Boccaletti, S possesses 19 (20-1 = 19 i.e. considering scientist Boccaletti, M himself as a node) connections with the respective scientists, stating that he has collaborated with 19 different scientists and provided research in the field of network science.

The depth of the node ID can also be increased revealing the collaborations that scientist made through reference of the scientists he/she has collaborated individually with. Increasing the depth in this network of social collaboration doesn't prove fruitful as when contributing towards a research study direct collaborations need to be fulfilled in order to make reliable connections.

K-Brace Filter:

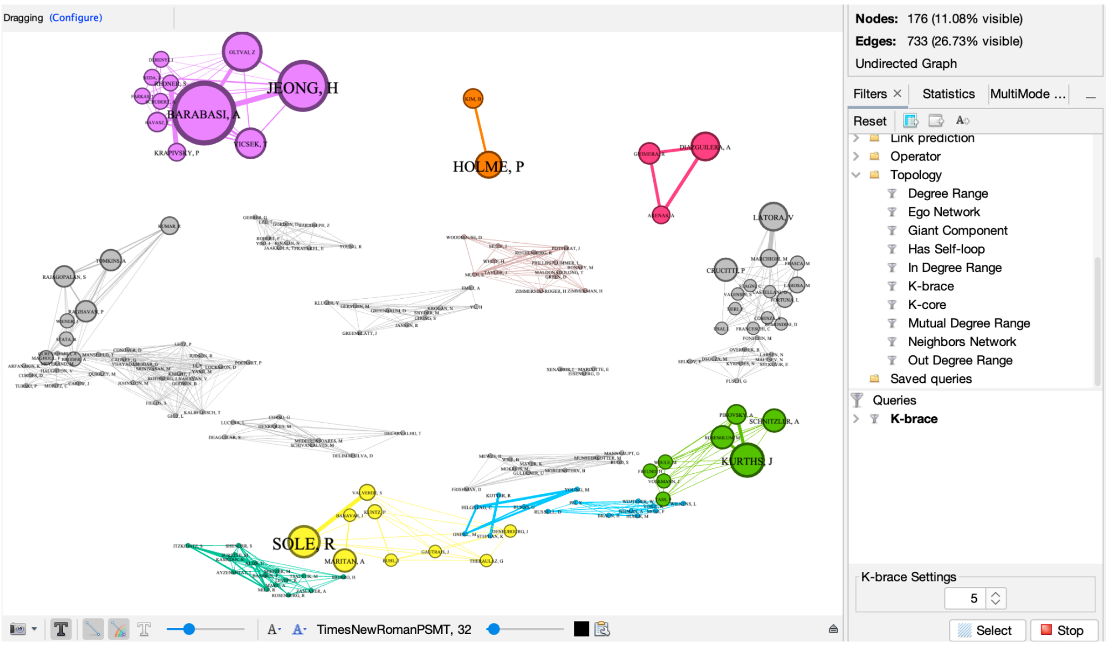

Figure 6.3 Network with K-Brace filter, when k=5

K-brace filter shows the edge's fixity. The number of nodes possessing common neighbors and based on that network is filtered. In Figure 6.3, the K-Brace value is set to 5, so the nodes having 5 common neighbors with another node are removed.

Similarly, as the value of K-Brace is increased the main node with diverse connections with different nodes are revealed. K-Brace filter on its application, deletes all edges of embeddedness less than the value of k and deletes nodes having single connection. This helps the main node (scientist) for allowing proper flow of information in the network based on trusted nodes (other scientists the main scientist has collaborated with).

The usage of different filters is utilized to answer research questions mentioned in Section 1. Filters are primary tools for performing analysis and finding various insights within the network which can be inferred by the usage of a few filters used above. Interesting insights are revealed with the usage of a few filters like Giant component revealing the most influential and main scientists with maximum connections, Ego network filter which is used to reveal the connection of any respective scientists and discern with how many scientists the specific scientist collaborates with considering his/her own community and different communities, and K-Brace filter revealing common connections of the scientists and setting different values to find referential collaborations.

Section Seven: Network Features

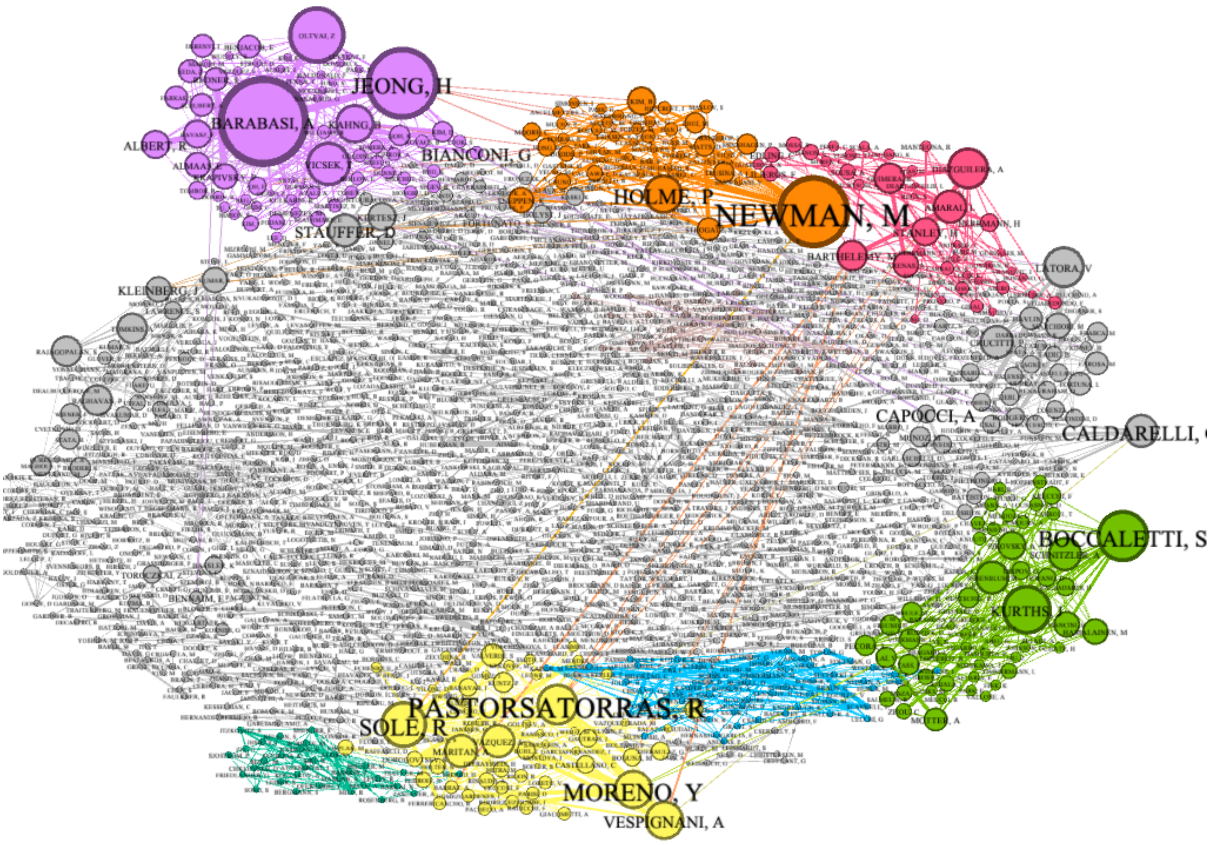

Figure 7.1 Final Network Communities

One of the network features found was communities. From figure 5.3 it is shown that the clustering coefficient highly suggests the formation of clusters. This is relevant to the study as to determine if there are any distinct groupings amongst the scientists. Figure 7.1 supports this fact as it is shown that several clusters have formed. This is indicative of the communities of scientists as nine different communities of the scientists are formed.

Homophily is identified within the network. There are some of the scientists that have hardly any connections outside of their own community. Robustness of the network is based on the centrality distribution of the network. There are numerous nodes in the network which are disconnected portraying the scientists who do not collaborate with any of the respective scientists and provide their own contribution in the field of network science. The network becomes disconnected in some cases losing its robustness due to disconnectedness. The network contains data about a large number of scientists revealing a good amount of spread and formation of social networks based on the factor of network science.

Section 8: Final Results and Contemplation

The following insights have been procured based on the research questions mentioned in section one and the inferences are made as follows:-

1. **Are there preferred scientists as well as outlier fewer collaborative scientists?**

The number of scientists collaborating with maximum 1 scientist or scientists contributing individually for research in network science. As the network is a social network, the scientists who don't collaborate with other scientists or have minimal connections in the field are preferred to be outliers in the network.For example:- 40% of scientists from the whole network have minimal connections or contribute solely in the field.

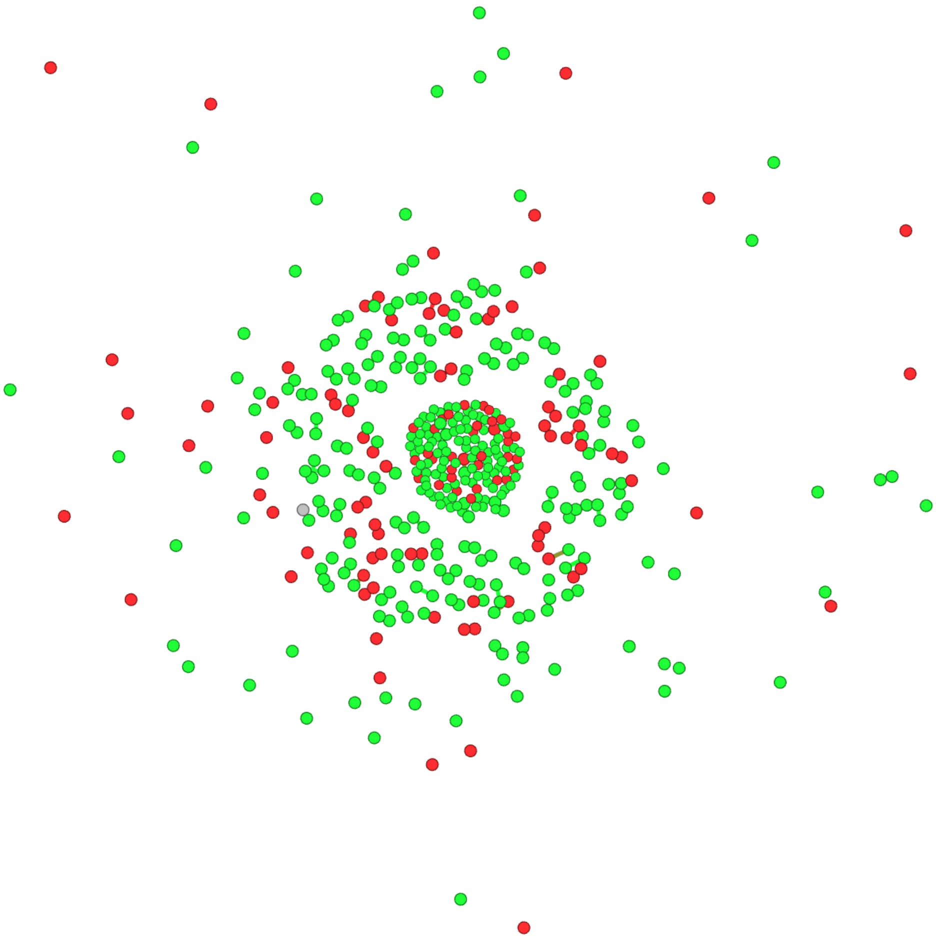

Figure 8.1 Research Question 1

There are 431 scientists who collaborate very rarely and if they do, it is maximum with 1 scientist in the field of network science and provide their contribution in the field. The above network is partitioned based on gender, males (red) females (green).

2. **Are there any distinct social circles or communities that exist within the network? For instance, Are there any communities that are closely grouped and have maximum output for the networking field.?**

For instance there is a specific community of scientists which has collaborated with maximum scientists, or there is a community with maximum contribution in the network field. Communities are formed in the network when a specific scientist has connections with other respective scientists which in turn creates a social network, formed for sole purpose for making contributions in the field of network science.

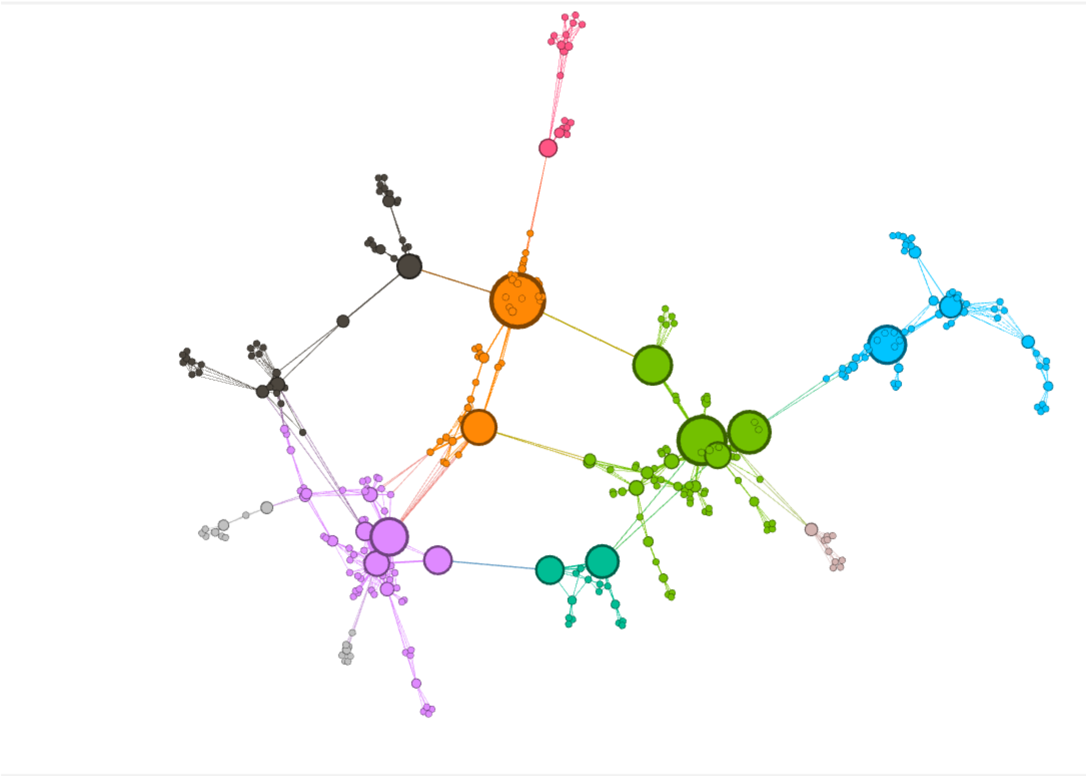

Figure 8.2 Research Question 2

To extract and visualize a more manageable network, the network was filtered by the Giant component as well as filtering out all nodes with degrees less than 2. This resulted in 10 communities being formed. Leaders of the top 5 communities are shown in the table below.

Orange Community	Newman, M
Green Community	Pastorsatorras, R
Blue Community	Bocalleti, S
Dark Green Community	Caldarelli, G
Purple Community	Jeong H

3. **Are there any social hierarchies within the scientists' collaboration? Are there any Systems of social organization in which some individuals enjoy a higher social status than others? Can we find if there's a repeating theme with one individual within some groups, then that person is higher on the social hierarchy?**

An individual scientist who possesses a higher social status is the one with maximum connections and the one who is responsible for forming major communities within the network. That scientist is higher in social hierarchy and is considered to be the leader of the organization, the one who has collaborated with maximum other respective scientists and provided research and innovations in the field of network science.

For the mentioned researched question Top 5 communities providing maximum contribution have been taken into consideration and based on that the scientist having higher social status i.e. the scientists with maximum collaborations in the field have been deduced.

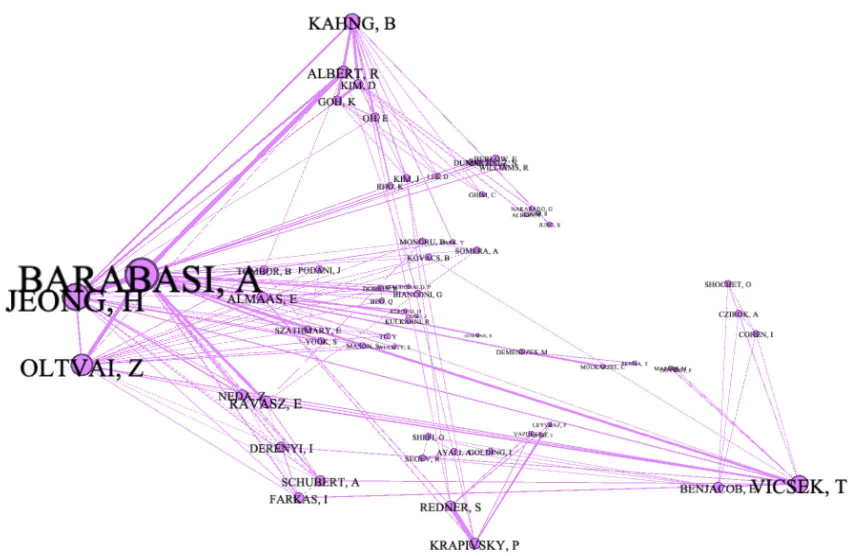

Figure 8.3.1 Research Question 3:Community with the highest contribution

In the community with the highest contribution in the field of network science, the scientists **Barabasi, A, Jeong, H and Oltvai, Z** are the scientists having higher social status, meaning they have the maximum collaborations with the scientists of their own respective community.

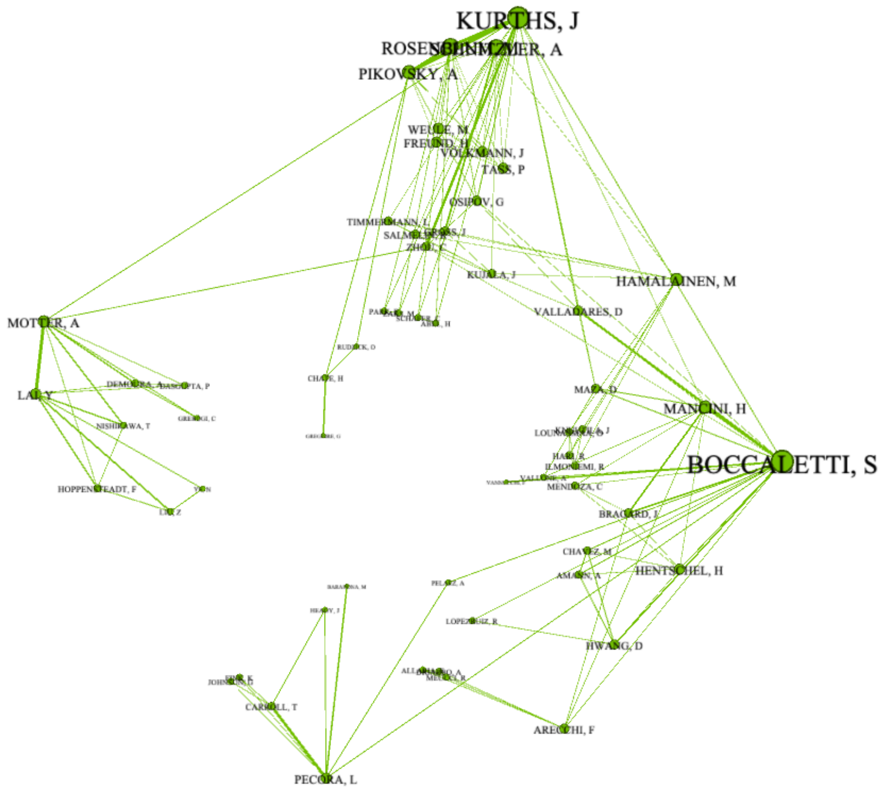

Figure 8.3.2 Research Question 3:Community with the second highest contribution

In the community with the second highest contribution in the field of network science, the scientists **Boccaletti, S,** and **Kurthis, J** are the scientists having higher social status, meaning they have the maximum collaborations with the scientists of their own respective community.

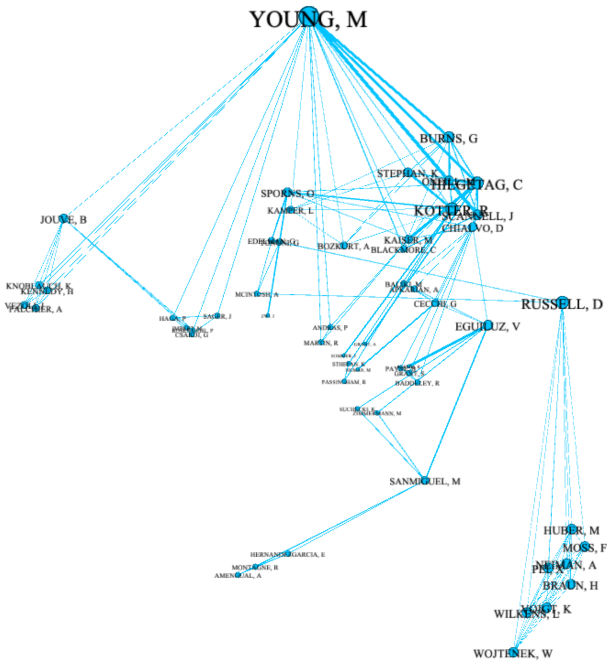

Figure 8.3.3 Research Question 3:Community with the third highest contribution

In the community with the third highest contribution in the field of network science, the scientist **Young, M** is the only scientist having higher social status, meaning the only scientist having maximum collaborations with the scientists of his own respective community.

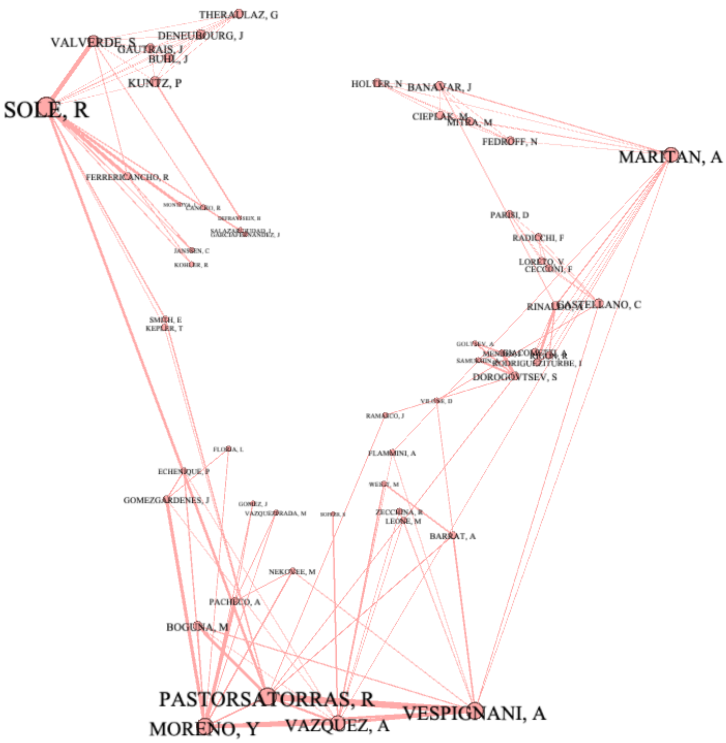

Figure 8.3.4 Research Question 3:Community with the fourth highest contribution

In the community with the fourth highest contribution in the field of network science, the scientists **Sole, R, Maritan, A, Pastorsatorras, R, Moreno, Y, Vasquez, A** and **Vespignani, A** are the scientists having higher social status, meaning they have the maximum collaborations with the scientists of their own respective community.

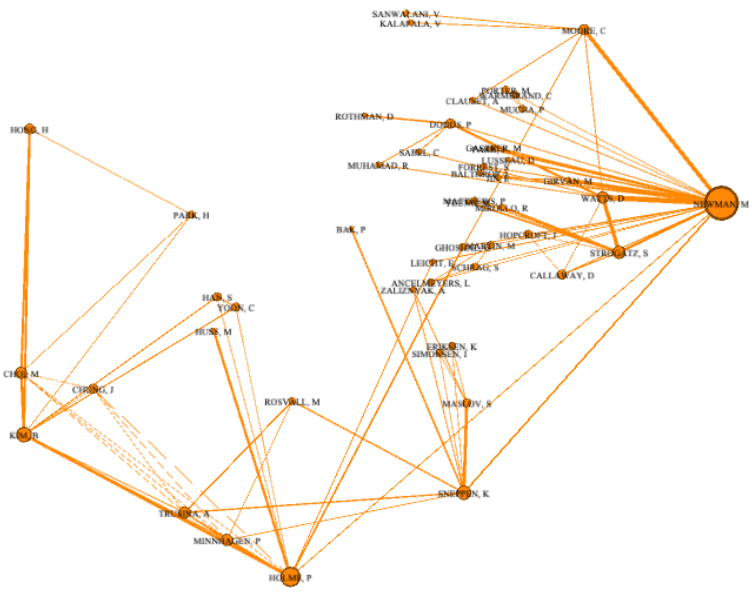

Figure 8.3.5 Research Question 3:Community with the fifth highest contribution

In the community with the fifth highest contribution in the field of network science, the scientist **Newman, M** is the only scientist having higher social status, meaning the only scientist having maximum collaborations with the scientists of his own respective community.

The scientists in the respective communities as well as in the whole network possess higher social status by forming maximum collaborations with other respective scientists present irrespective of the background or gender.

4. **Which country of origin for the scientist provides maximum insight into the field of network?**

The maximum number of scientists belonging to a specific country. The countries who provide maximum contribution in the field of network science. This can help to discern which country has maximum resources and are willing to contribute in plenty to the network field. For example - 30% of the scientists in the network belong to the United States of America.

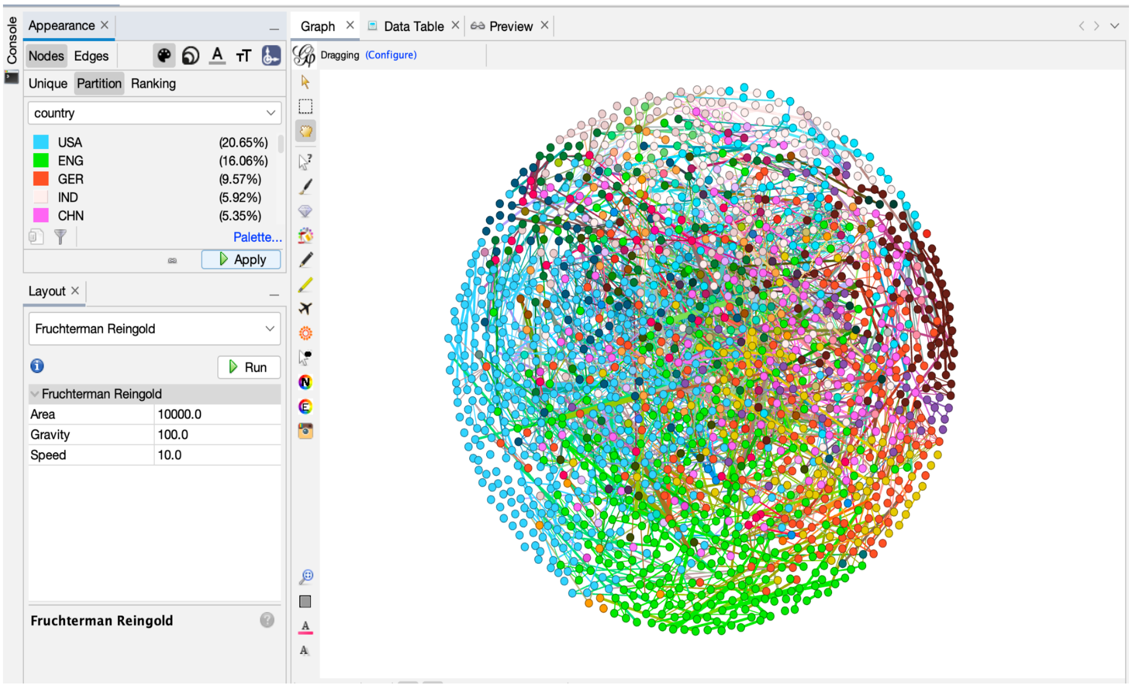

Figure 8.4 Research Question 4

The top 5 countries who contributed the most in the field of network science were **The United States of America (USA), England, Germany, India** and **China**. Out of which the U.S. had the largest number of scientists who contribute to this field i.e. around **20.66%** followed by England with **16.06%**, Germany with **9.57%**, India with **5.92%** , and lastly China with **5.35%**. The network contains numerous other countries as well who provide their contribution in the field of network science.

5. **What is the predominant gender of the scientists that contributed the most to the networking field?**

The gender of scientists, male or female, which one of them is predominant and provides more contribution in the field of network science. For example - around 70% of scientists contributing to the network field are females or males.

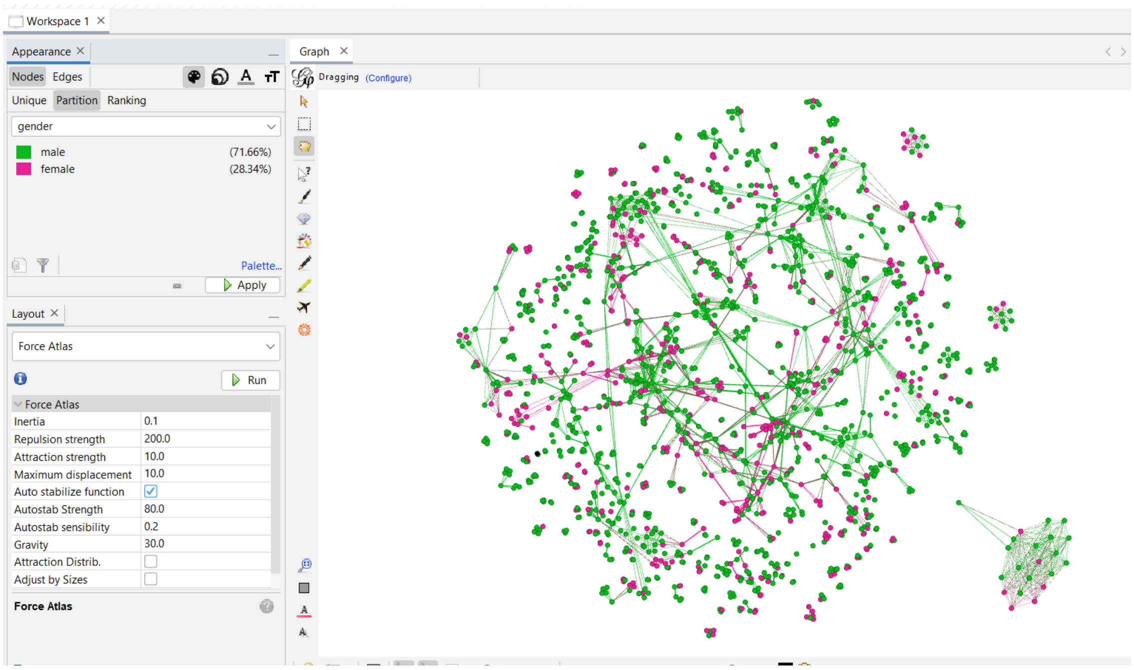

Figure 8.5 Research Question 5

In the network, male gender(green) contributes more to the network compared to females(pink). Out of the 1589 network scientists, 71.66% of the scientists are male and 28.34% of the scientists are female.

6. **What are the hubs of scientists who have contributed the most in the field of network science?**

Scientists forming larger hubs in the network are the ones with common connections. The scientists who form enclosed communities within the network, having fewer connections with other respective scientists outside of their community in the network. For instance there is a specific group of scientists or a scientist who collaborates with scientists of his/her own community or only with the scientists of the same country. This results in the formation of a hub within the network.

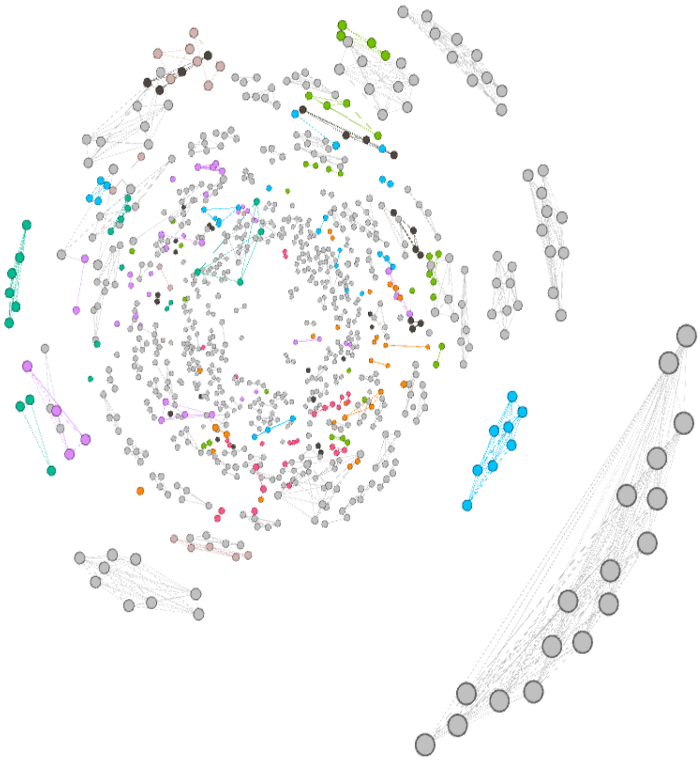

Figure 8.6 Research Question 6

Specific Hubs/Communities of scientists belonging to different backgrounds are formed within the network who collaborate within their specific community only or have minimal connections with scientists belonging to a different community.The network shows scientists having connections enclosed within their specific community and not having any distant connections with other scientists.

7. **Are there any scientists who don't collaborate with anyone and provide their own research in the field of network?**

There are scientists who do not collaborate with other respective scientists. They contribute solely in the field of network science. For example - 10% of the scientists in the network contribute individually in the field. Maximum scientists who contribute individually are either male or female.

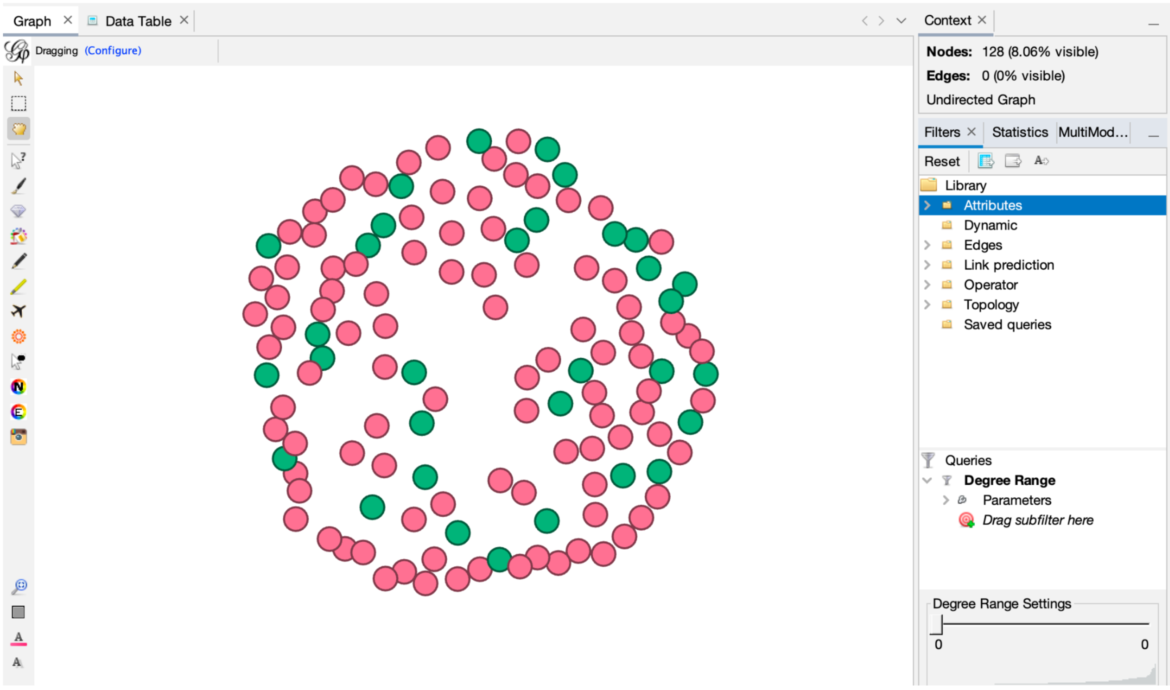

Figure 8.7 Research Question 7

There are 128 scientists who don't collaborate with any other network scientists. They contribute exclusively by themselves in the field of network science. There are 30 female scientists and 98 male scientists belonging to various countries who provide their contribution individually in this field. Approximately 8% of the scientists provide contribution individually without collaborating with any of the respective scientists in the network field.

8. Who are the most influential scientists in the network?

The scientists having the most connections, as well as a higher weighted degree within the network are considered to be the most influential scientists. The scientists who have contributed to more than 10 research papers or experimental studies by collaborating with other respective scientists are considered to be influential.

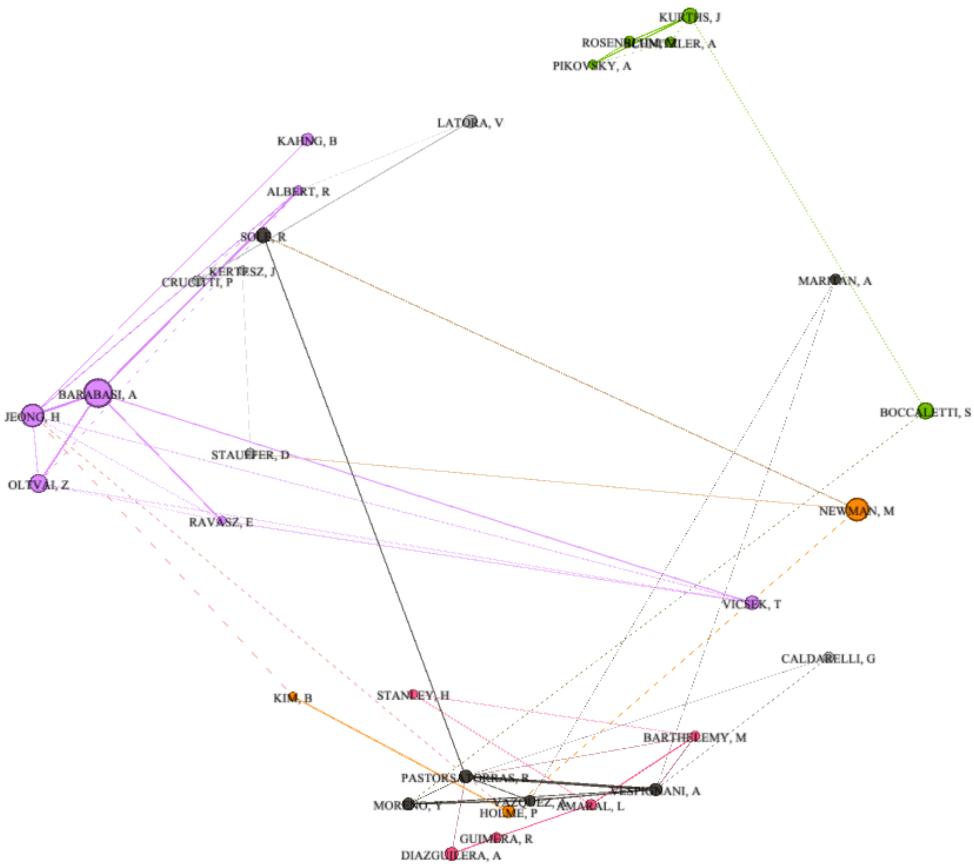

Figure 8.8 Research Question 8

There are 31 scientists in the network who are the most influential. They are most influential based on the fact that they have contributed to more than 10 research papers or experiments in the field of network science by collaborating with other scientists.

9. Do scientists from cross background co-author a paper together?

Scientists belonging to different countries and collaborating to provide their research and innovations in the field of network science. The cross background characteristic can be discovered within larger communities where there are maximum scientists

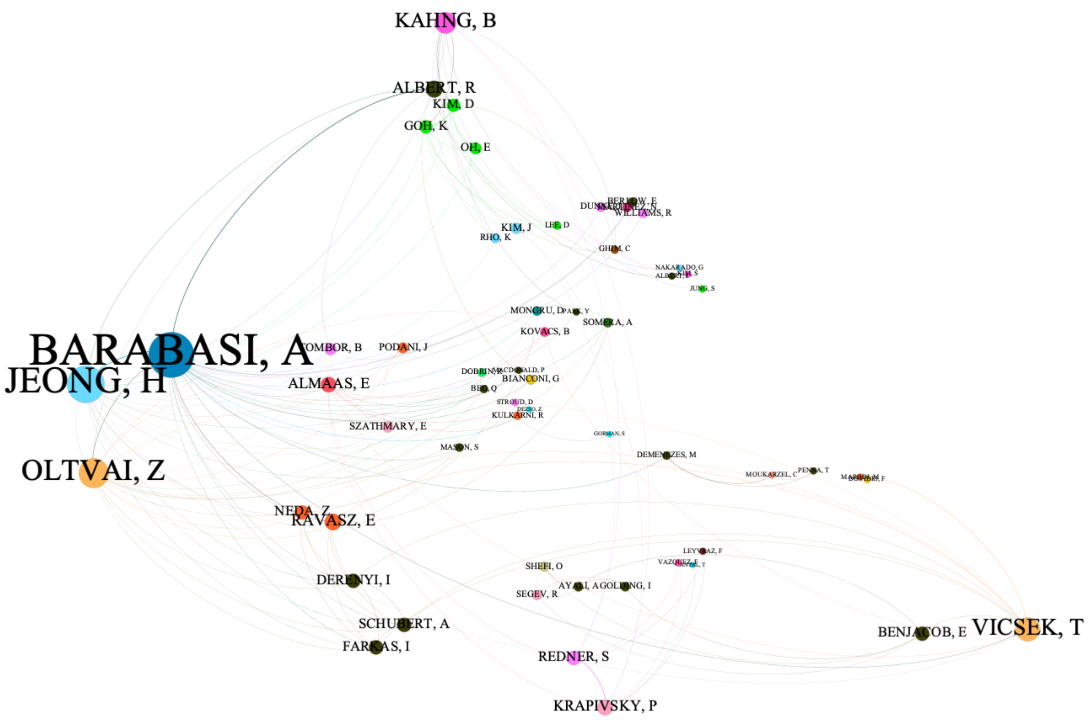

Figure 8.9 Research Question 9

The above network shows the community (containing 64 scientists) with maximum contributions in the field of network science . It showcases many scientists from different backgrounds co-authoring many research papers or taking part in experimental studies related to network science. The backgrounds are based on the fact that the scientists belong to different countries.

The research analysis of above mentioned questions reveals interesting insights within the social collaboration of network scientists. The insights obtained reveal there are a considerable amount of scientists contributing to the field of network science portraying how they contribute, how much they contribute, the impact they have in this field, metadata about scientists revealing their origin and gender etc. all this provides a considerable amount of understanding of the network.

Final Contemplation

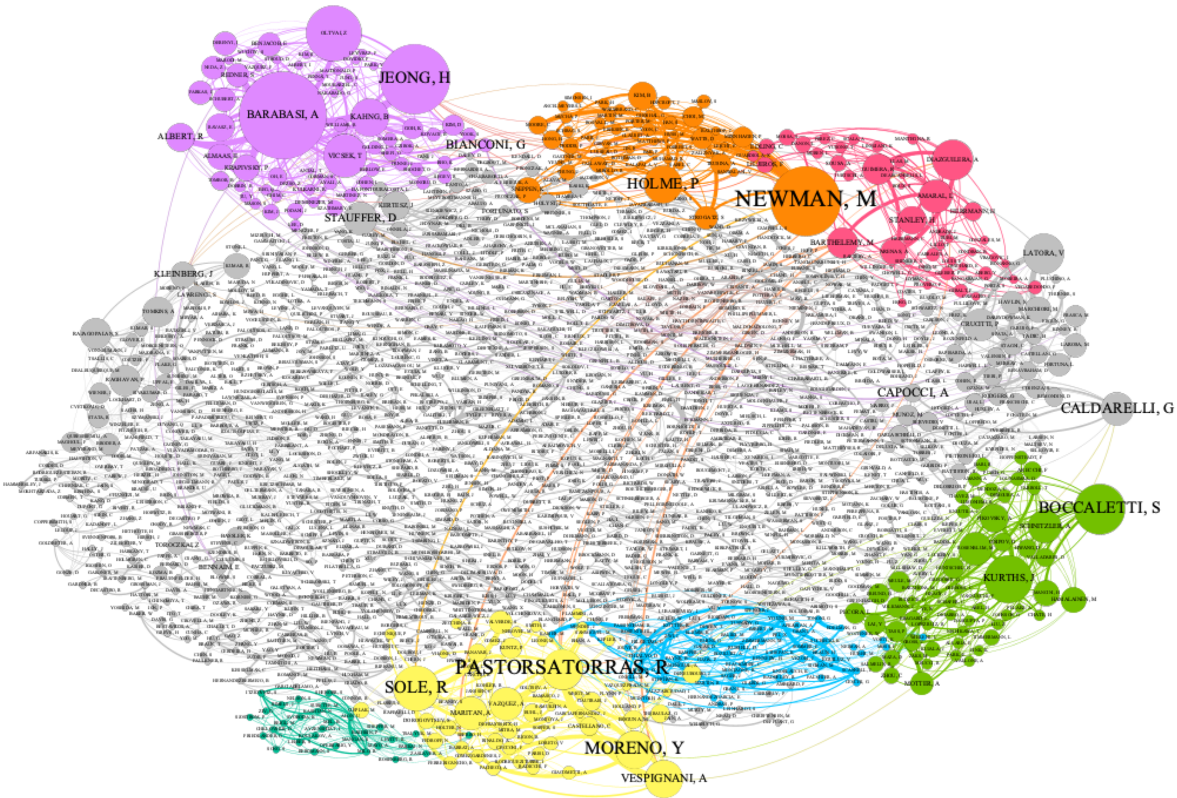

Figure 8.10 Final Network

What could have made the study more interesting, would be the **ages of the scientists**. This metric could be used to gain further insight on how this factor played into collaboration between network scientists, to ascertain whether or not preferences based on age exist. Another metric that could have made the study more relevant would be the metadata about **educational institutions the scientists or professors belong to**. If the specific institution where the scientist belongs to is revealed, a good amount of hubs and communities in the network can be created and can also help discern the world renowned institutes who provide their contribution in the field of network science.

The dataset of Social Collaboration of Network Scientists contained the node IDs, labels, specific edge weights and degrees of the respective scientists within the dataset. The dataset also contained metadata about the scientists' gender and country of origin. The dataset contained metadata about only a few scientists, the metadata about the remaining scientists was obtained online through a data source mentioned as under.

Data Source - https://en.wikipedia.org/wiki/List_of_network_scientists

The project workload was divided into segments as follows:

Final Project Tasks	Assigned To
Network Visualization with Gephi using different algorithms	Jainam, Dylan
Statistical Analysis	Jalique, Matthew
Research Questions (2 each to every member)	Jainam, Dylan, Jalique, Matthew
Final Project Report	Jainam, Dylan, Jalique, Matthew
Final Project Presentation	Jainam, Dylan, Jalique, Matthew

In terms of dividing the workload between our group members, we were all very agreeable as to how we divided the workload, and we all supported each other in our shared contribution. We did not have any problems or gripes within the group during our contributions to the final project.

Something unexpected was the workload the size of our dataset brought. Labeling the nodes was more difficult as the large number of nodes prompted a lot of label overlap. There are a lot of edges as well, so working with the data proved interesting. In class, we had not yet worked with a data set this large, so it provided a great learning opportunity to become more familiar with working on a data set this large.

Social Collaboration of Network Scientists reveals the most elementary way in which a social network is formed when it is focused on a common goal. The network dataset is large, containing data about various scientists providing a general idea of social structure that can be observed in the real world and how collaborations with people working in the same field can influence the network's structure. Innumerable number of scientists working in the network field and providing their contribution by collaborating with other respective scientists evolves the field. Analyzing a social network through visualization can help one understand the fluidity of the network and find interesting insights.

The collaboration of various network scientists in providing research and performing experimental studies in the field of network science is visualized using "Gephi", the collaboration between the scientists takes place whenever they co-author or contribute in a group to provide research or perform an experimental study in the field of network science. This social network of network scientists also contains metadata about scientists' gender and their country of origin. The network is visualized through the nodes and edges in "Gephi". Interesting insights are found when the network is analyzed through statistics, plotting distribution plots, applying various layout algorithms, and filtering the data. The network of social collaboration of network scientists divided into nine distinct communities when partitioned using modularity class. The most influential scientists having maximum collaborations in the network and the number of connections they have in their own community and also the outside community is calculated through centrality distribution. Using various techniques of statistical analysis like finding average weighted degree, network diameters, eccentricity distribution etc. provides an elementary idea about the social network and their connections. Using filters of topology, edges and attributes the network navigation and flow of information is discerned.

Employing the mentioned features of "Gephi" network becomes more fluid and understandable.The formation of social network and how it operates based on factor of contributing to the common field i.e. network science, the formation of specific communities, the scientists collaborating with other scientists and the amount of their diverse connections with other scientists, considering background and various characteristics of the scientists themselves, considering all these factors various inferences are made. Using "Gephi" to form the social network of network scientists and visualizing it provides a conception that any network irrespective of its size and vastness can be made understandable when visualized in a proper manner.