Social Network Analysis

The Social Network of Astronauts

Lapakis Gerasimos 3170089

Introduction

As I was browsing Kaggle to brainstorm ideas, I stumbled upon many datasets, but the one I will be analyzing was the one that attracted me the most. I always had an interest for space and astronauts, mainly due to my visits to the National Air and Space Museum in USA and Memorial Museum of Cosmonautics in Moscow. The concept of this project is an undirected graph, where nodes are astronauts with their information, such as name, occupation, sex, number of missions and many more! The edges are created between them, if they have been on a mission together. Each edge has information like the nodes (year and name of mission and more).

Data Preparation

To generate the data, I created two csv files, one for edges and one for nodes, astoedges and astronodes respectively, using Pandas.

```
import pandas as pd

import numpy as np

pd.set_option('display.max_columns', None)
```

```
astro = pd.read_csv('astronauts.csv')
astronodes = astro.drop_duplicates(subset=['number','name']).copy()
astronodes = astronodes[['number','name','sex','nationality','occupation','year_of_birth','military_civilian','year_of_selection','total_number_of_missions']]
astronodes.rename(columns={'number':'Id'}, inplace=True)
astronodes.rename(columns={'name':'Label'}, inplace=True)
astronodes.to_csv('astronodes.csv', encoding='utf-8')
```

```
astroedges = astro.groupby('mission_title')['number'].apply(list)
astroedges = astroedges.to_frame().reset_index()
astroedges.rename(columns={'number':'Target'}, inplace=True)
```

```
astroedges = astroedges.explode('Target')
```

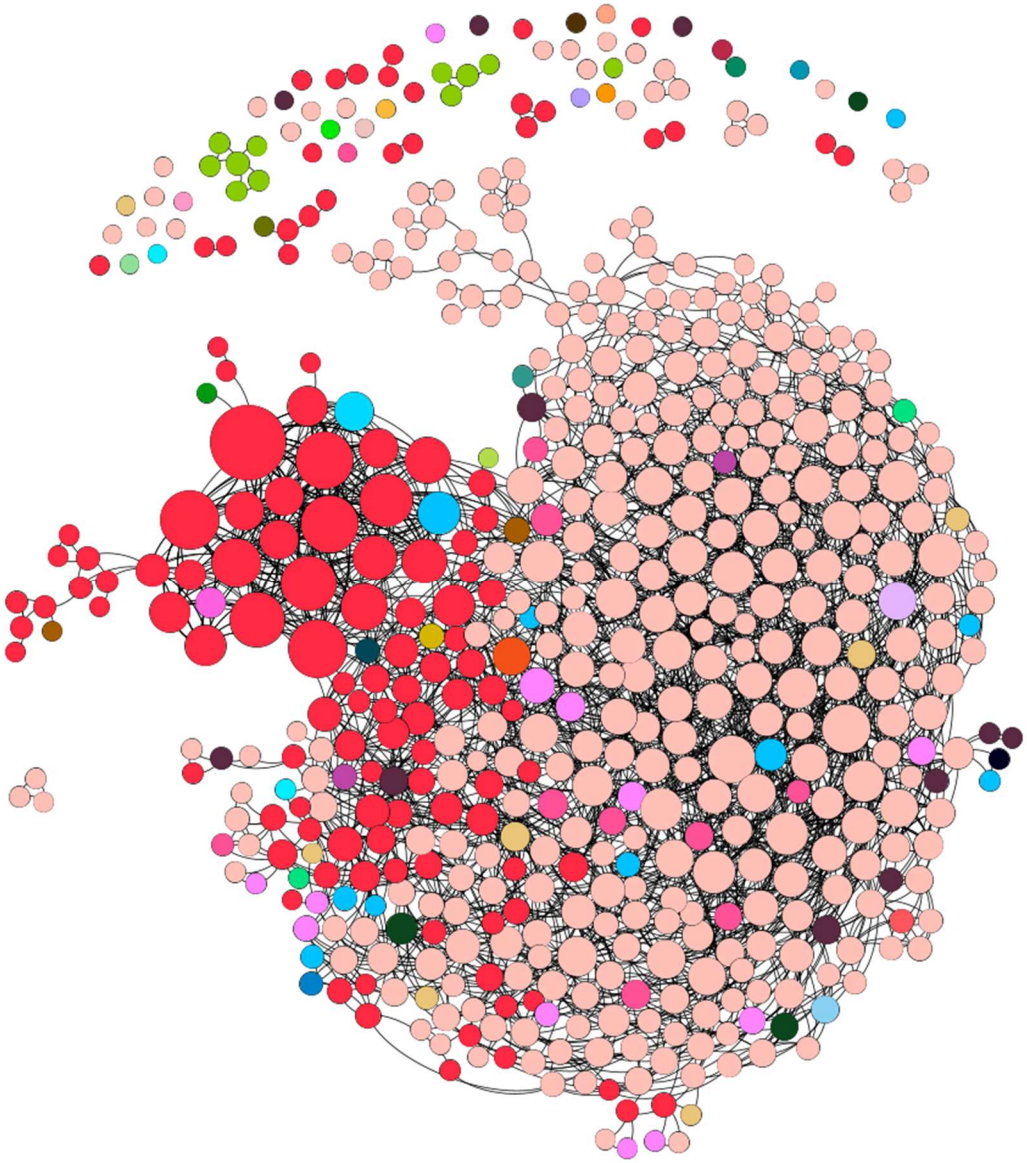
```
astroedges = pd.merge(astro,
              astroedges,
              on='mission_title',
              how='inner')
```

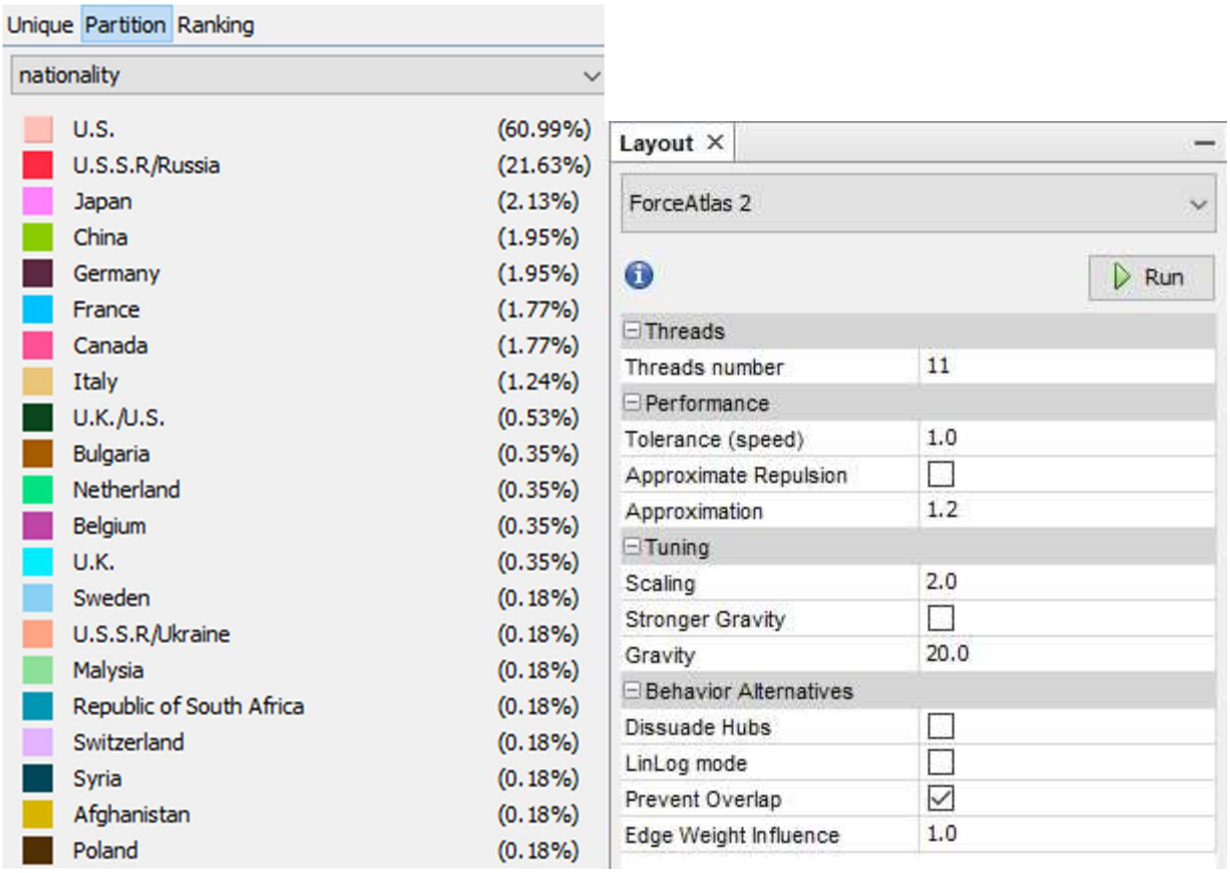
```
astroedges.rename(columns={'number':'Source'}, inplace=True)
astroedges['Type'] = 'Undirected'
astroedges['Weight'] = '1'
astroedges = astroedges.loc[astroedges['Source']!=astroedges['Target']]
astroedges = astroedges[['Source','Target','year_of_mission','mission_title','ascend_shuttle','in_orbit','descend_shuttle']]
```

```
astroedges.to_csv('astroedges.csv', encoding='utf-8')
```

Graphical Representation of Network

For the representation of the graph I used Gephi and some of its plug-ins. Firstly, I partitioned the nodes by nationality colors, as it is a better way to get an understanding of my network. Then, I used the Forced Atlas 2 Layout with gravity set to 20 and “Prevent Overlap” checked, to get a cleaner representation.





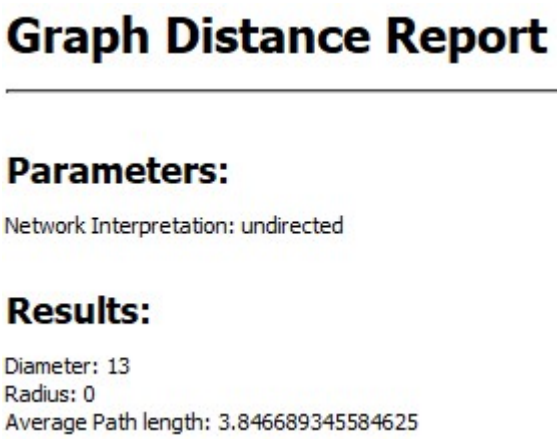
Topology

(Edges, Vectors)

Each graph is defined by two sets, V (nodes) and E (edges). In our case each node is an astronaut that has been to space, and has various attributes. According to this dataset there are 564 Nodes - astronauts, so V = 564. Our nodes are connected by edges, which show if an astronaut has been in a mission with another astronaut. Total number of edges is E = 2675.

Average Path Length

Now average path length is 3.846, which is quite interesting, as this means that one astronaut is on average less than 4 hops away from every other astronaut. On the other hand diameter is 13, which means that the 2 most distant astronauts need 13 hops to connect (the shortest path between 2 nodes). Diameter is more than 3 times the value of average path length, as in most networks.



Graph Distance Report

Parameters:

Network Interpretation: undirected

Results:

Diameter: 13
Radius: 0
Average Path length: 3.846689345584625

Graph Density

Graph density measures how close the network is to be complete. In this case, graph density came out to be 0.017, which means that this graph is quite sparse.

Results:

Density: 0.017

Even if we exclude the weakly connected components (the ones with id other than 0), we are left with 478 nodes and a graph density of 0.023. These astronauts are the ones that have only been to space alone one time or with a small group consisting of many astronauts, than never participated in another mission.

Results:

Density: 0.023

Our graph's topology is even comparable to the one of US air transportation as it has 546 nodes, 2781 edges, 0.02 density and 10.2 average degree, as we have 564, 2675, 0.017 and 9.486.

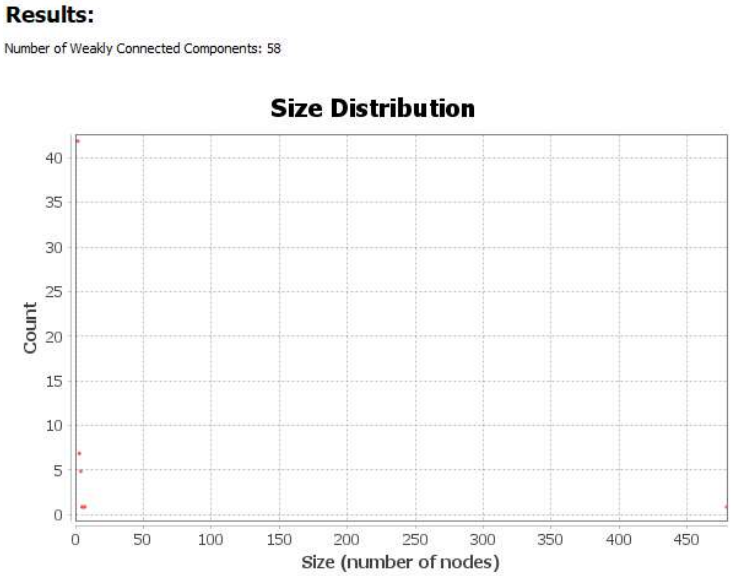
Network	Type	Nodes (N)	Links (L)	Density (d)	Average degree ($\langle k \rangle$)
Facebook Northwestern Univ.		10,567	488,337	0.009	92.4
IMDB movies and stars		563,443	921,160	0.000006	3.3
IMDB co-stars	W	252,999	1,015,187	0.00003	8.0
Twitter US politics	DW	18,470	48,365	0.0001	2.6
Enron Email	DW	87,273	321,918	0.00004	3.7
Wikipedia math	D	15,220	194,103	0.0008	12.8
Internet routers		190,914	607,610	0.00003	6.4
US air transportation		546	2,781	0.02	10.2
World air transportation		3,179	18,617	0.004	11.7
Yeast protein interactions		1,870	2,277	0.001	2.4
C. elegans brain	DW	297	2,345	0.03	7.9
Everglades ecological food web	DW	69	916	0.2	13.3

Source: Slides, Lecture 3 of course Social Network Analysis.

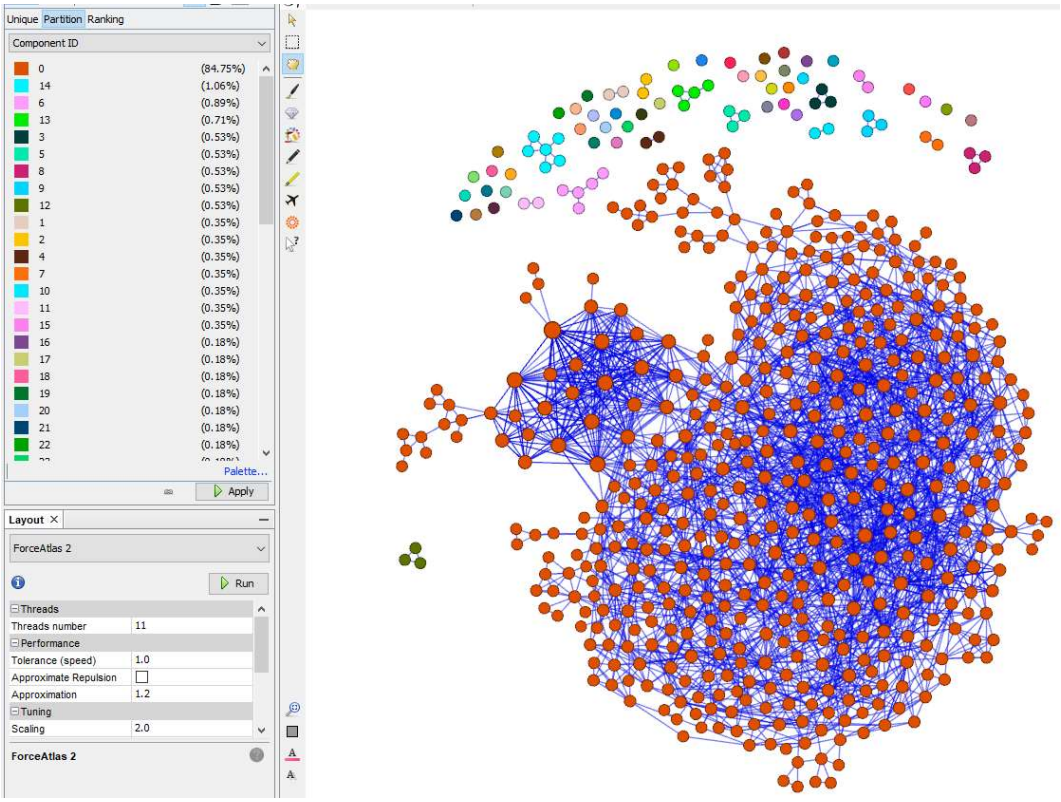
Connected Components

We can see from the Gephi Graph that there are 58 weakly connected components, which is totally normal for our graph, as there are many astronauts, especially in the early stages of space exploration, that were travelling to space alone or in small groups! So all the nodes that are not connected to our main component are added to this category, 86/564 nodes to be exact. Note that there are nodes that have Degree greater than zero that are in this category. These are the ones that have only been to missions together and with no one else.

What is interesting is that there are some quite famous names there, like Yuri Gagarin (first human in outer space) and Mark Richard Shuttleworth (first African from an independent country to travel to space). Finally, the weakly components that are consisted of multiple nodes have not mixed nationalities except one, Soyuz 36/35(1980) with Bertalan(Hungary) and Kubasov(USSR).



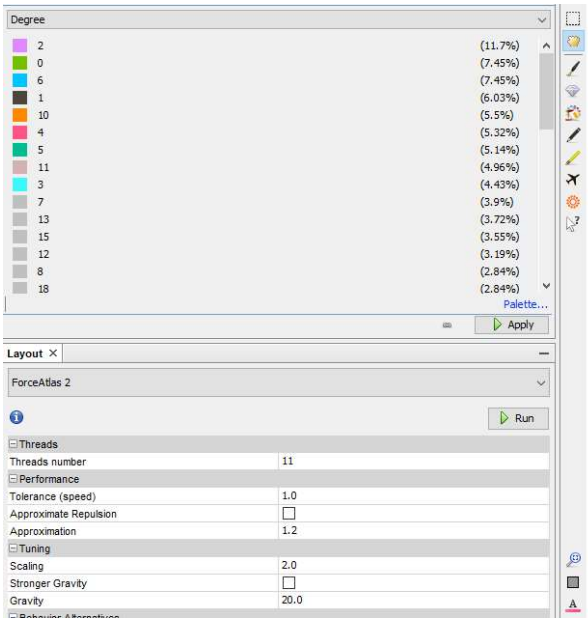
84.75% of all nodes are in the giant component(ID=0), while the 2nd biggest component(ID=14) is 1.06% of all nodes. In the following graph, nodes are colored based on component ID.

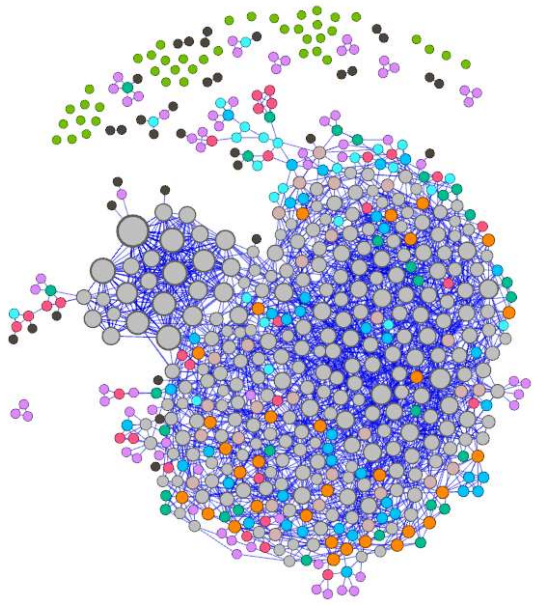


Degree

Now, it is time to talk about the degree of this graph, or rather average degree. Degree shows how many edges a node has and average degree is calculated by 2E/V, in our case, 5350/564=9.4858 ~ 9.486, which is what Gephi calculated. This graph is undirected so there is no in degree and out degree.

The 5 most common Degree values are 2(66 nodes with 30 edges), 0(42 nodes with, well, 0 edges), 6 (42 nodes with 16 edges), 1(34 nodes with 7 edges) and 10 (31 nodes with 19 edges). As expected the nodes that are part of the weakly connected components have a degree of 0 to 3.





Id	Label	Degree	Component ID	total_number_of_miss...	sex	nationality	occupation	military_civilian	year_of_selection	year_of_birth
86	Dzhanibekov, Vladimir	65	0	5	male	U.S.S.R/Russia	commander	military	1970	1942
93	Popov, Leonid	46	0	3	male	U.S.S.R/Russia	commander	military	1970	1945
99	Strekalov, Gennadi	45	0	5	male	U.S.S.R/Russia	MSP	civilian	1973	1940
95	Malyshev, Yuri	44	0	2	male	U.S.S.R/Russia	commander	military	1967	1941
111	Savitskaya, Svetlana	44	0	2	female	U.S.S.R/Russia	MSP	civilian	1980	1948
91	Lyakhov, Vladimir	42	0	3	male	U.S.S.R/Russia	commander	military	1967	1941
98	Kizim, Leonid	41	0	3	male	U.S.S.R/Russia	commander	military	1965	1941
100	Savinykh, Viktor	39	0	3	male	U.S.S.R/Russia	flight engineer	civilian	1978	1940
88	Ivanchenkov, Aleksandr	35	0	2	male	U.S.S.R/Russia	flight engineer	military	1973	1940
192	Ross, Jerry L.	34	0	7	male	U.S.	MSP	military	1980	1948
282	Brown, Curtis L., Jr.	33	0	6	male	U.S.	pilot	military	1987	1956
110	Serebrov, Aleksandr	32	0	4	male	U.S.S.R/Russia	flight engineer	civilian	1978	1944
123	Aleksandrov, Aleksandr	31	0	2	male	U.S.S.R/Russia	flight engineer	civilian	1978	1943
183	Volkov, Alexander	30	0	3	male	U.S.S.R/Russia	MSP	military	1976	1948
136	Solovyov, Vladimir	29	0	2	male	U.S.S.R/Russia	flight engineer	civilian	1978	1946
197	Chang-Diaz, Franklin R.	29	0	7	male	U.S.	MSP	civilian	1980	1950
288	Helms, Susan J.	28	0	5	female	U.S.	MSP	military	1990	1958
274	Bowersox, Kenneth D.	28	0	5	male	U.S.	pilot	military	1987	1956
108	Chr��tien, Jean-	27	0	3	male	France	MSP	military	1980	1938
271	Foale, C. Michael	27	0	6	male	U.S.	MSP	civilian	1987	1957

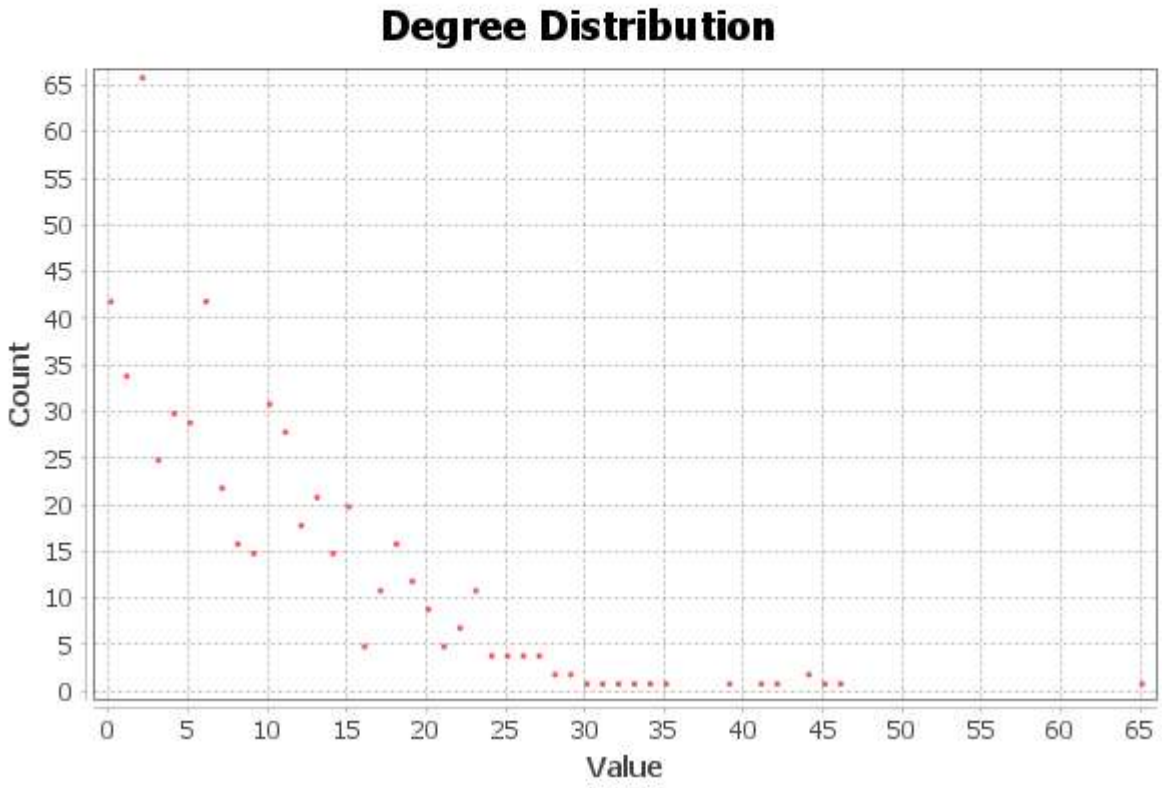
The above top 20, shows that only 2 astronauts with high degree are females. There was one astronaut form France, 13 from USSR and 6 from USA. The military/civilian and occupation attributes do not play a role for an astronaut to be connected with others, 8 were civilian and 12 were from military. Number of missions varies, but Vladimir

Dzhanibekov has a degree of 65 which means that within 5 missions, he has travelled with 65 different astronauts!

Degree Report

Results:

Average Degree: 9.486



Centrality

Moving on to centralities measures, we will change the nodes' size and color, according to which centrality we will be examining. Each node displays astronaut's name and number of missions.

Betweenness Centrality

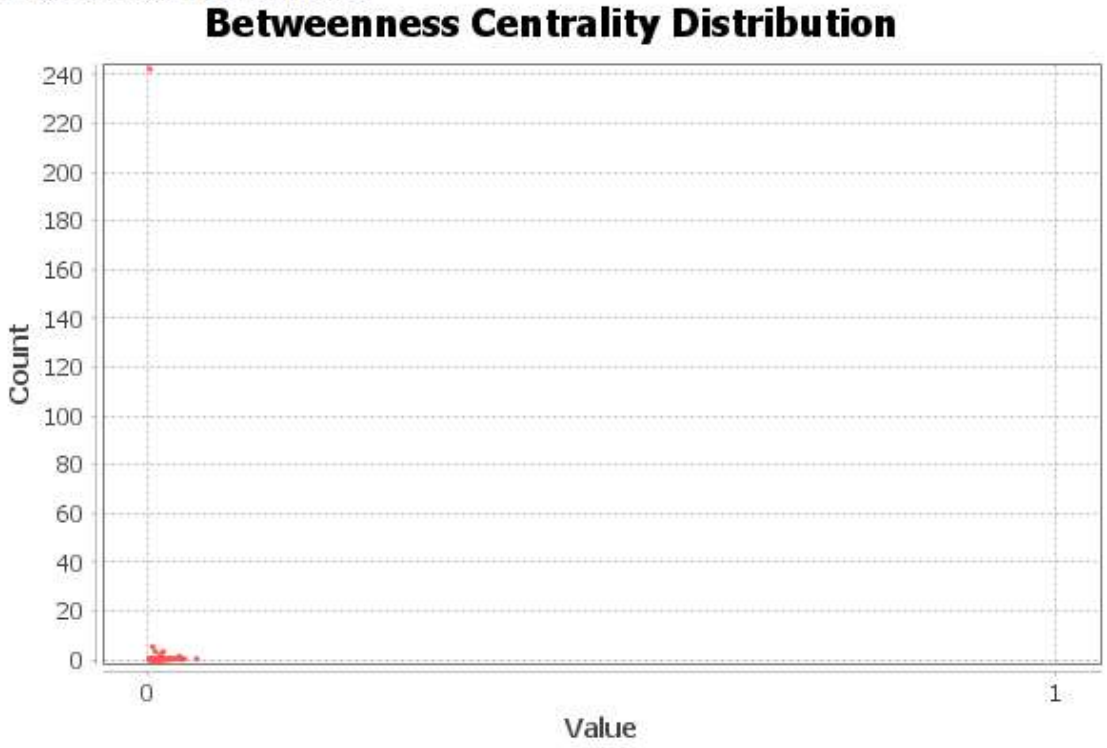
Another important measure of centrality is betweenness, as it is based on shortest paths. It counts the number of times a node acts as a bridge along the shortest path between two other nodes. Contrary to closeness centrality, it doesn't seem to be affected by total number of missions. Michael Collins has the third biggest betweenness centrality, although he has participated in only 2 missions, but he is connected with Young John, who has the highest value, due to them being together on Gemini 10. In addition, Ross Jerry, has been to 7 missions and has almost half the betweenness centrality of Young John, while Franklin Chang-Diaz who has 7 missions has a mere 0.01 betweenness centrality. All these astronauts have considerable influence within the network as they control the information passing between others.

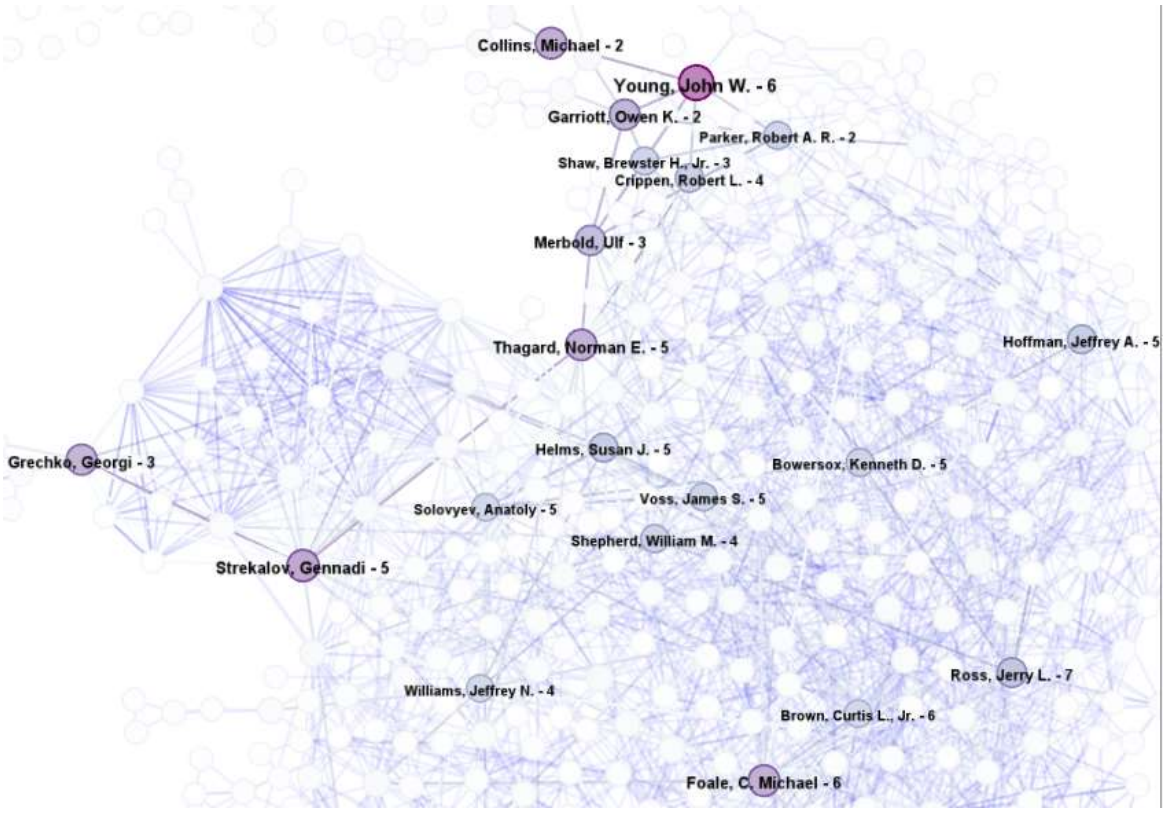
That means despite them having the most trips to space, they are "not important" to the network of information flow of astronauts' network. Again, we see that in our top 20 there are 16 astronauts from US, 3 from USSR and only one from Germany. Finally, Helms Susan is the only female in our list. There are 11 militaries and 9 civilians. Number of missions doesn't seem important as it varies between astronauts.

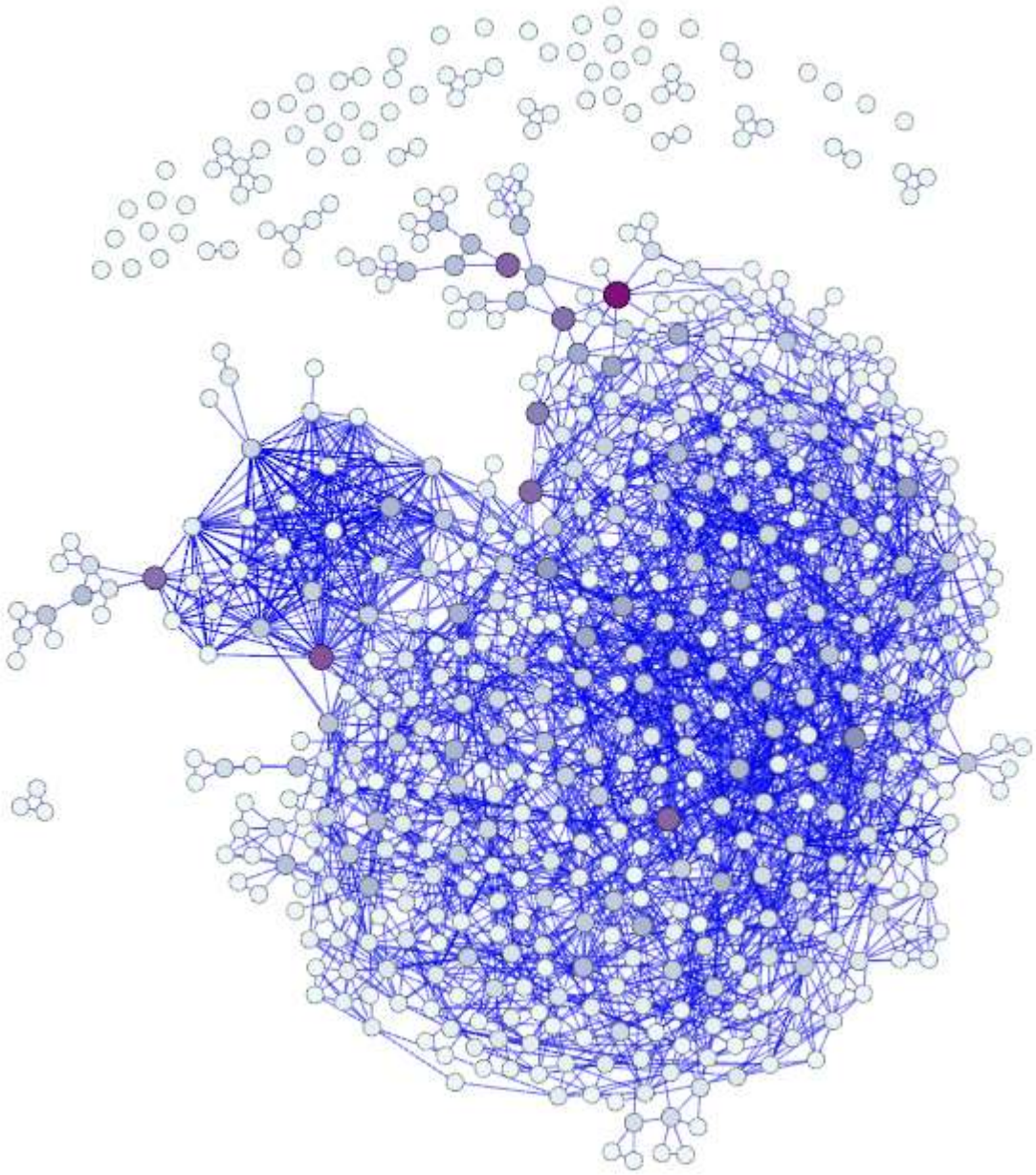
Id	Label	Betweenness Cent...	Component ID	total_number_of_miss...	sex	nationality	occupation	military_civilian	year_of_selection	year_of_birth
17	Young, John W.	0.051451	0	6	male	U.S.	pilot	military	1962	1930
99	Strekalov, Gennadi	0.038217	0	5	male	U.S.S.R/Russia	MSP	civilian	1973	1940
27	Collins, Michael	0.035271	0	2	male	U.S.	pilot	civilian	1963	1930
271	Foale, C. Michael	0.035217	0	6	male	U.S.	MSP	civilian	1987	1957
122	Thagard, Norman E.	0.03444	0	5	male	U.S.	MSP	civilian	1978	1943
75	Grechko, Georgi	0.032401	0	3	male	U.S.S.R/Russia	flight engineer	civilian	1968	1931
63	Garriott, Owen K.	0.032401	0	2	male	U.S.	MSP	civilian	1965	1930
131	Merbold, Ulf	0.029433	0	3	male	Germany	PSP	civilian	1978	1941
192	Ross, Jerry L.	0.026793	0	7	male	U.S.	MSP	military	1980	1948
288	Helms, Susan J.	0.023619	0	5	female	U.S.	MSP	military	1990	1958
163	Hoffman, Jeffrey A.	0.023046	0	5	male	U.S.	MSP	civilian	1978	1944
102	Crippen, Robert L.	0.021736	0	4	male	U.S.	pilot	military	1969	1937
214	Shepherd, William M.	0.020607	0	4	male	U.S.	MSP	military	1984	1949
274	Bowersox, Kenneth D.	0.020566	0	5	male	U.S.	pilot	military	1987	1956
262	Voss, James S.	0.020149	0	5	male	U.S.	MSP	military	1987	1949
128	Shaw, Brewster H., Jr.	0.019995	0	3	male	U.S.	pilot	military	1978	1945
129	Parker, Robert A. R.	0.019613	0	2	male	U.S.	MSP	civilian	1967	1936
208	Solovyev, Anatoly	0.018349	0	5	male	U.S.S.R/Russia	commander	military	1976	1948
396	Williams, Jeffrey N.	0.01758	0	4	male	U.S.	MSP	military	1996	1958
398	Burbank, Daniel C.	0.017575	0	3	male	U.S.	MSP	military	1996	1961

Results:

Diameter: 13
Radius: 0
Average Path length: 3.846689345584625





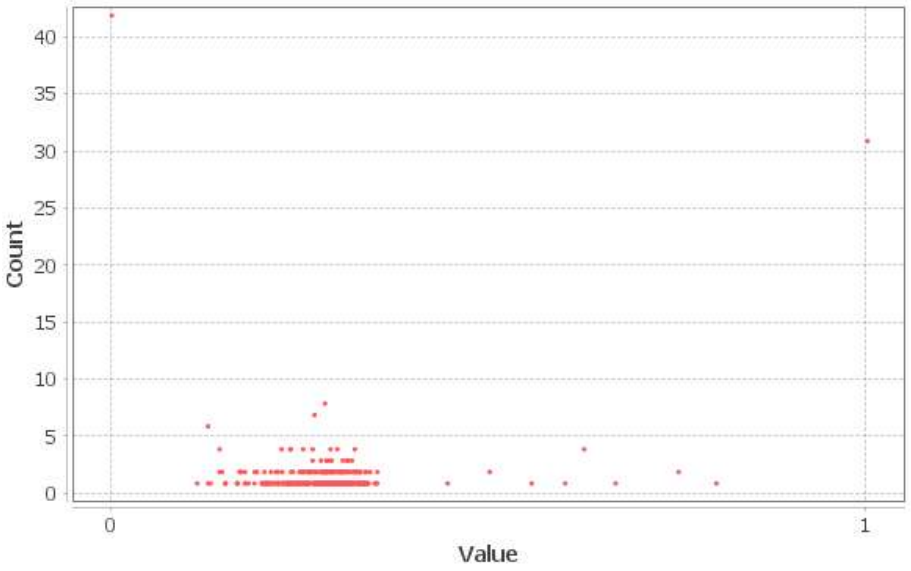


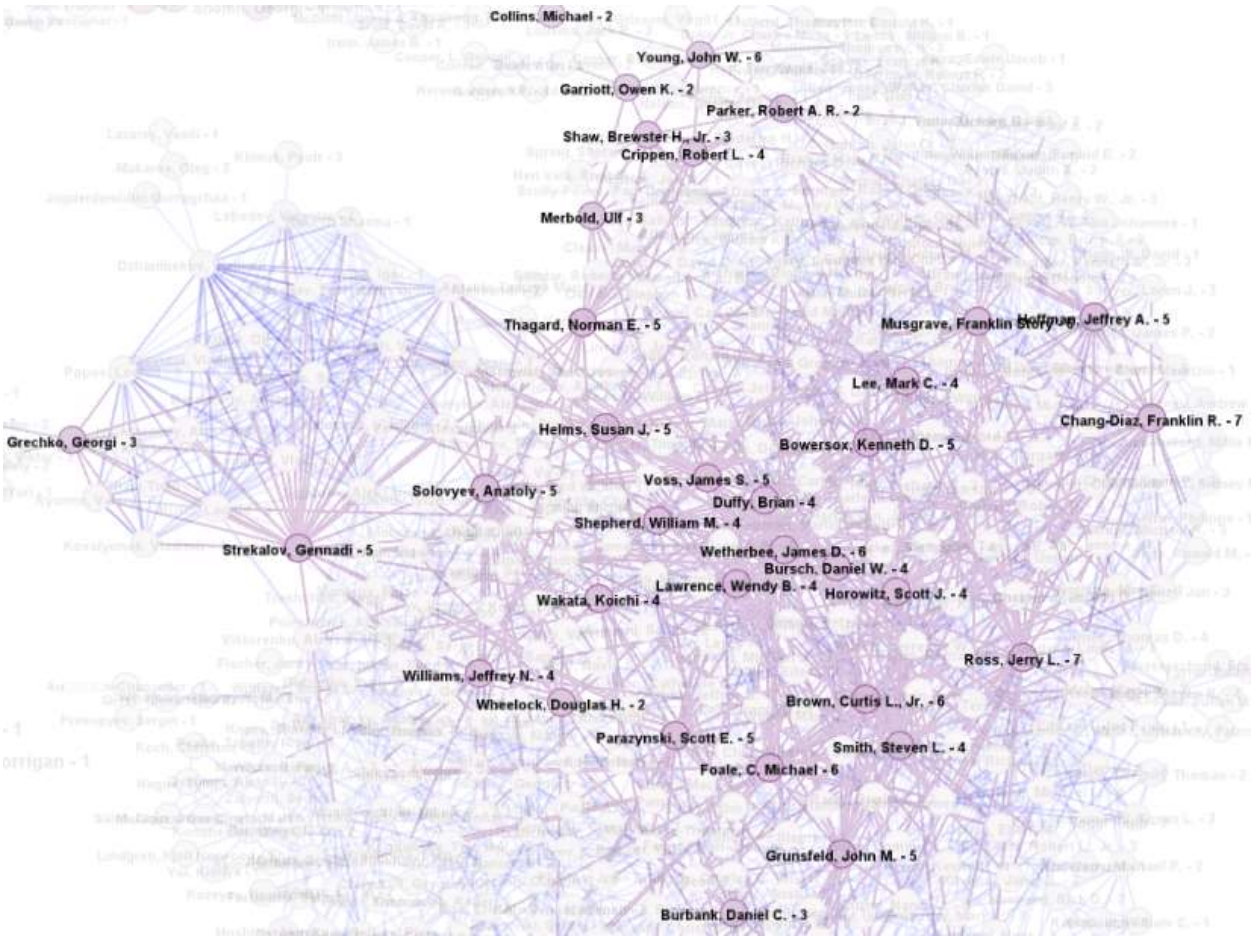
Closeness Centrality

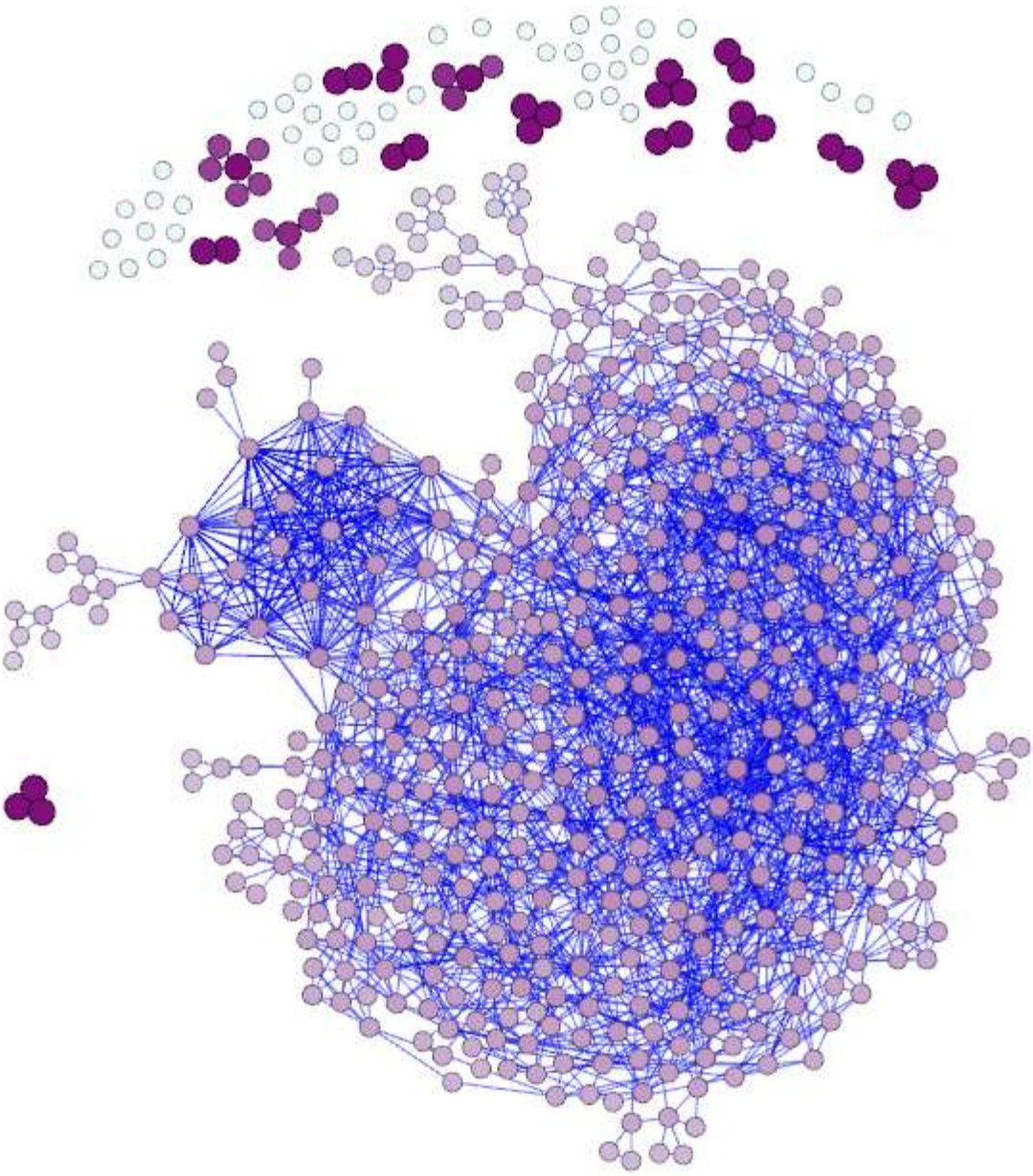
Firstly, we can see that the weakly connected components have 0 closeness centrality, if they have degree 0 and in every other case a pretty high value. This happens as it is a measure of centrality based on the sum of the length of the shortest paths between the node and all other nodes in the graph and each component is considered a mini-graph. Now, let's talk about our main component with Component ID = 0. Here are the top 20 closeness centrality astronauts. Only 2 are non USA citizens, one from USSR and one from Japan and only 2 are females. Here, we observe that all of them have been from 4 to 7 missions, which is what we expected, as closeness centrality shows how easily a node can reach other nodes and as our edges are created through astronauts being together in a mission.

Id	Label	Closeness Centra...	Component ID	total_number_of_miss...	sex	nationality	occupation	military_civilian	year_of_selection	year_of_birth
274	Bowersox, Kenneth D.	0.351252	0	5	male	U.S.	pilot	military	1987	1956
282	Brown, Curtis L., Jr.	0.351252	0	6	male	U.S.	pilot	military	1987	1956
271	Foale, C. Michael	0.350993	0	6	male	U.S.	MSP	civilian	1987	1957
192	Ross, Jerry L.	0.349963	0	7	male	U.S.	MSP	military	1980	1948
288	Helms, Susan J.	0.347668	0	5	female	U.S.	MSP	military	1990	1958
214	Shepherd, William M.	0.341935	0	4	male	U.S.	MSP	military	1984	1949
326	Grunsfeld, John M.	0.341935	0	5	male	U.S.	MSP	civilian	1992	1958
226	Wetherbee, James D.	0.339744	0	6	male	U.S.	pilot	military	1984	1952
302	Bursch, Daniel W.	0.339019	0	4	male	U.S.	MSP	military	1990	1957
346	Horowitz, Scott J.	0.338058	0	4	male	U.S.	pilot	military	1992	1957
218	Lee, Mark C.	0.337819	0	4	male	U.S.	MSP	military	1984	1952
262	Voss, James S.	0.336864	0	5	male	U.S.	MSP	military	1987	1949
327	Lawrence, Wendy B.	0.336864	0	4	female	U.S.	MSP	military	1992	1959
208	Solovyev, Anatoly	0.336627	0	5	male	U.S.S.R/Russia	commander	military	1976	1948
343	Wakata, Koichi	0.335207	0	4	male	Japan	MSP	civilian	1992	1963
319	Smith, Steven L.	0.335207	0	4	male	U.S.	MSP	civilian	1992	1958
122	Thagard, Norman E.	0.334502	0	5	male	U.S.	MSP	civilian	1978	1943
323	Parazynski, Scott E.	0.334268	0	5	male	U.S.	MSP	civilian	1992	1961
117	Musgrave, Franklin Story	0.333566	0	6	male	U.S.	MSP	civilian	1967	1935
270	Duffy, Brian	0.332868	0	4	male	U.S.	pilot	military	1985	1953

Closeness Centrality Distribution



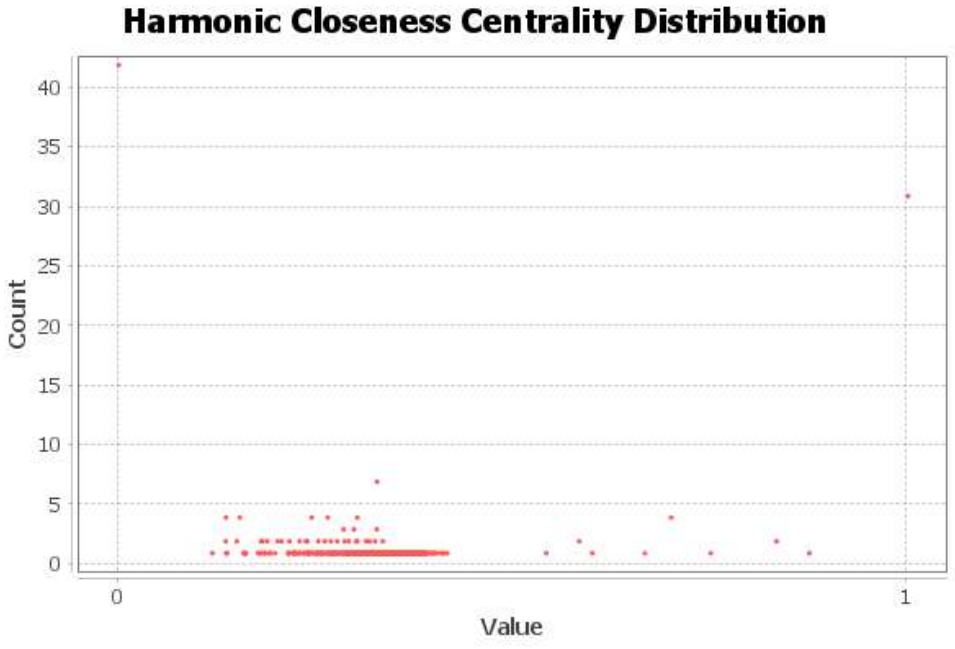


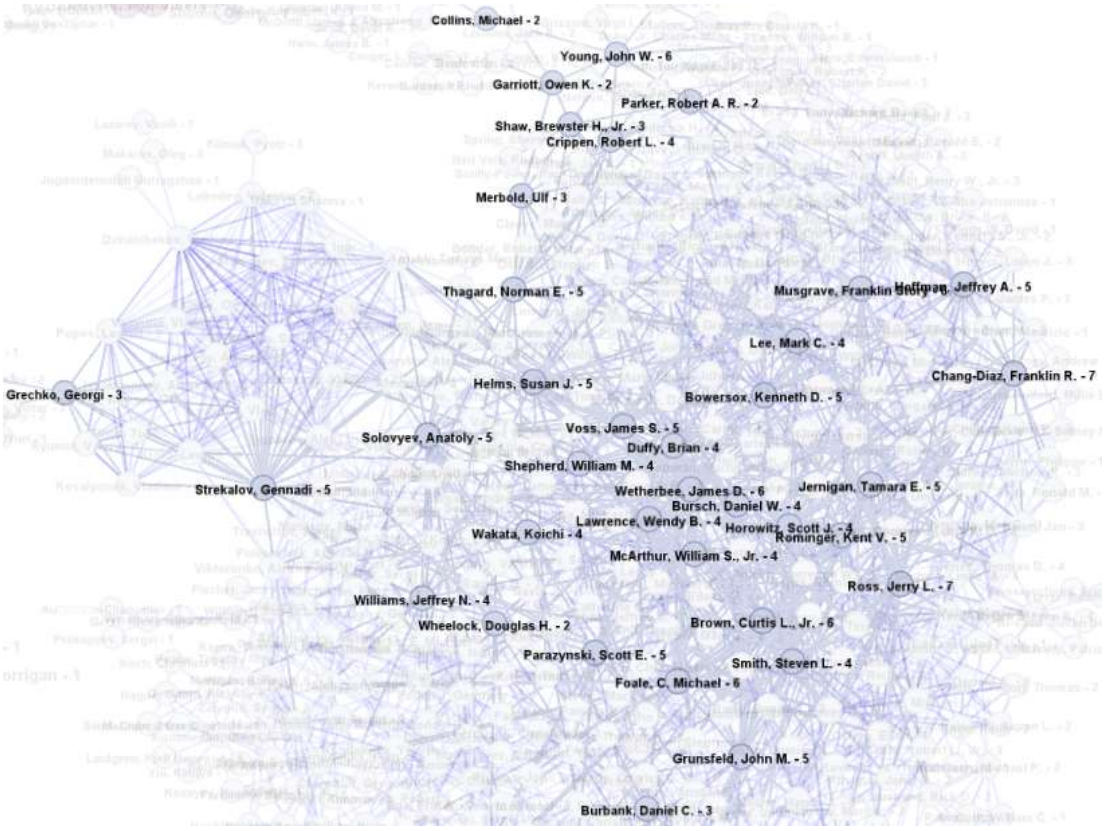


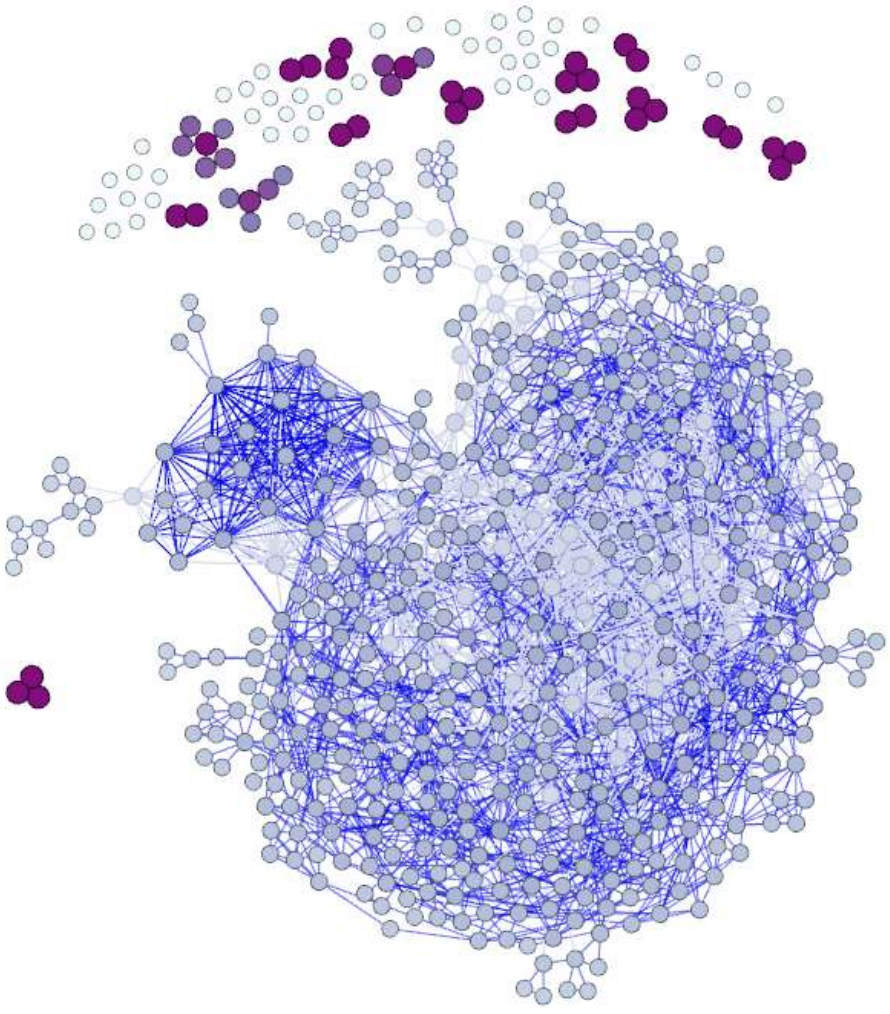
Harmonic Closeness Centrality

Harmonic Closeness was invented to help with dealing with unconnected graphs. So in our case, it is more suitable as we have many weakly connected components. The results are quite similar with the ones of closeness centrality's top 20, but most astronauts have scored a higher centrality value. There are 18 astronauts from USA and only 2 from USSR and all of them have been from 4 to 7 missions. There are 17 male and 3 female astronauts in this list. Finally, there are 6 civilians and 14 militaries

Id	Label	Harmonic Closeness C...	Component ID	total_number_of_miss...	sex	nationality	occupation	military_civilian	year_of_selection	year_of_birth
282	Brown, Curtis L., Jr.	0.416357	0	6	male	U.S.	pilot	military	1987	1956
192	Ross, Jerry L.	0.412965	0	7	male	U.S.	MSP	military	1980	1948
274	Bowersox, Kenneth D.	0.409541	0	5	male	U.S.	pilot	military	1987	1956
271	Foale, C. Michael	0.409367	0	6	male	U.S.	MSP	civilian	1987	1957
288	Helms, Susan J.	0.404245	0	5	female	U.S.	MSP	military	1990	1958
226	Wetherbee, James D.	0.401164	0	6	male	U.S.	pilot	military	1984	1952
326	Grunsfeld, John M.	0.400025	0	5	male	U.S.	MSP	civilian	1992	1958
319	Smith, Steven L.	0.396679	0	4	male	U.S.	MSP	civilian	1992	1958
208	Solovyev, Anatoly	0.396462	0	5	male	U.S.S.R/Russia	commander	military	1976	1948
323	Parazynski, Scott E.	0.395713	0	5	male	U.S.	MSP	civilian	1992	1961
262	Voss, James S.	0.395553	0	5	male	U.S.	MSP	military	1987	1949
214	Shepherd, William M.	0.391108	0	4	male	U.S.	MSP	military	1984	1949
218	Lee, Mark C.	0.390743	0	4	male	U.S.	MSP	military	1984	1952
335	Rominger, Kent V.	0.389935	0	5	male	U.S.	pilot	military	1992	1956
346	Horowitz, Scott J.	0.389381	0	4	male	U.S.	pilot	military	1992	1957
327	Lawrence, Wendy B.	0.389206	0	4	female	U.S.	MSP	military	1992	1959
305	McArthur, William S., Jr.	0.387921	0	4	male	U.S.	MSP	military	1990	1951
302	Bursch, Daniel W.	0.387564	0	4	male	U.S.	MSP	military	1990	1957
254	Jernigan, Tamara E.	0.38697	0	5	female	U.S.	MSP	civilian	1985	1959
99	Strekalov, Gennadi	0.38674	0	5	male	U.S.S.R/Russia	MSP	civilian	1973	1940




Collins, Michael - 2
Young, John W. - 6
Garriott, Owen K. - 2
Parker, Robert A. R. - 2
Shaw, Brewster H., Jr. - 3
Crippen, Robert L. - 4
Merbold, Ulf - 3
Thagard, Norman E. - 5
Musgrave, Franklin Story - 6
Hoffman, Jeffrey A. - 5
Lee, Mark C. - 4
Chang-Diaz, Franklin R. - 7
Helms, Susan J. - 5
Grechko, Georgi - 3
Bowersox, Kenneth D. - 5
Voss, James S. - 5
Solovyev, Anatoly - 5
Duffy, Brian - 4
Shepherd, William M. - 4
Strekalov, Gennadi - 5
Wetherbee, James D. - 6
Jernigan, Tamara E. - 5
Bursch, Daniel W. - 4
Lawrence, Wendy B. - 4
Horowitz, Scott J. - 4
Wakata, Koichi - 4
Romingér, Kent V. - 5
McArthur, William S., Jr. - 4
Ross, Jerry L. - 7
Williams, Jeffrey N. - 4
Wheelock, Douglas H. - 2
Brown, Curtis L., Jr. - 6
Parazynski, Scott E. - 5
Smith, Steven L. - 4
Foale, C. Michael - 6
Grunsfeld, John M. - 5
Burbank, Daniel C. - 3



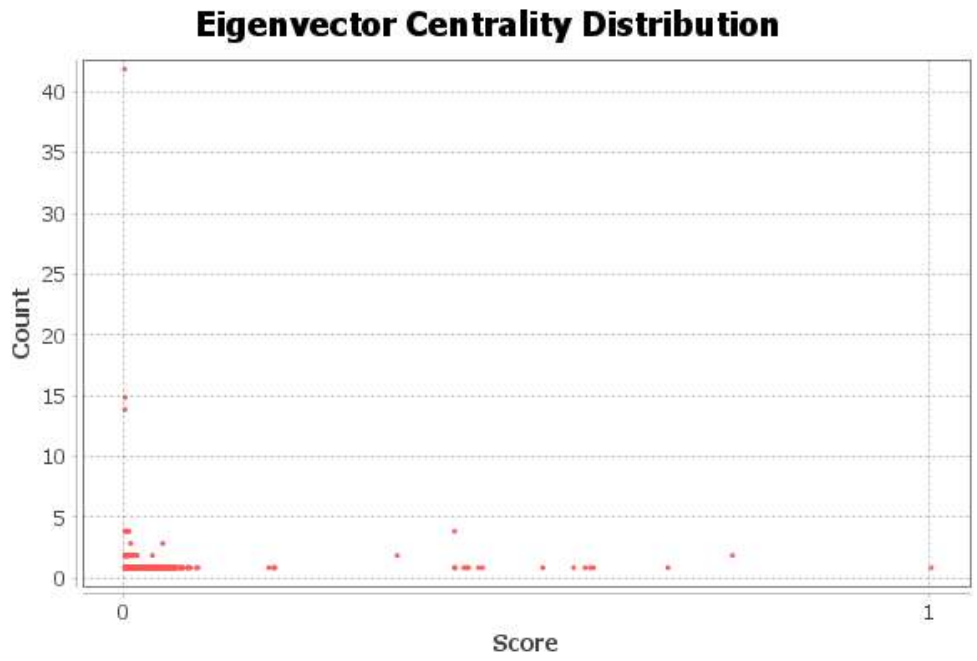
Eigenvector Centrality

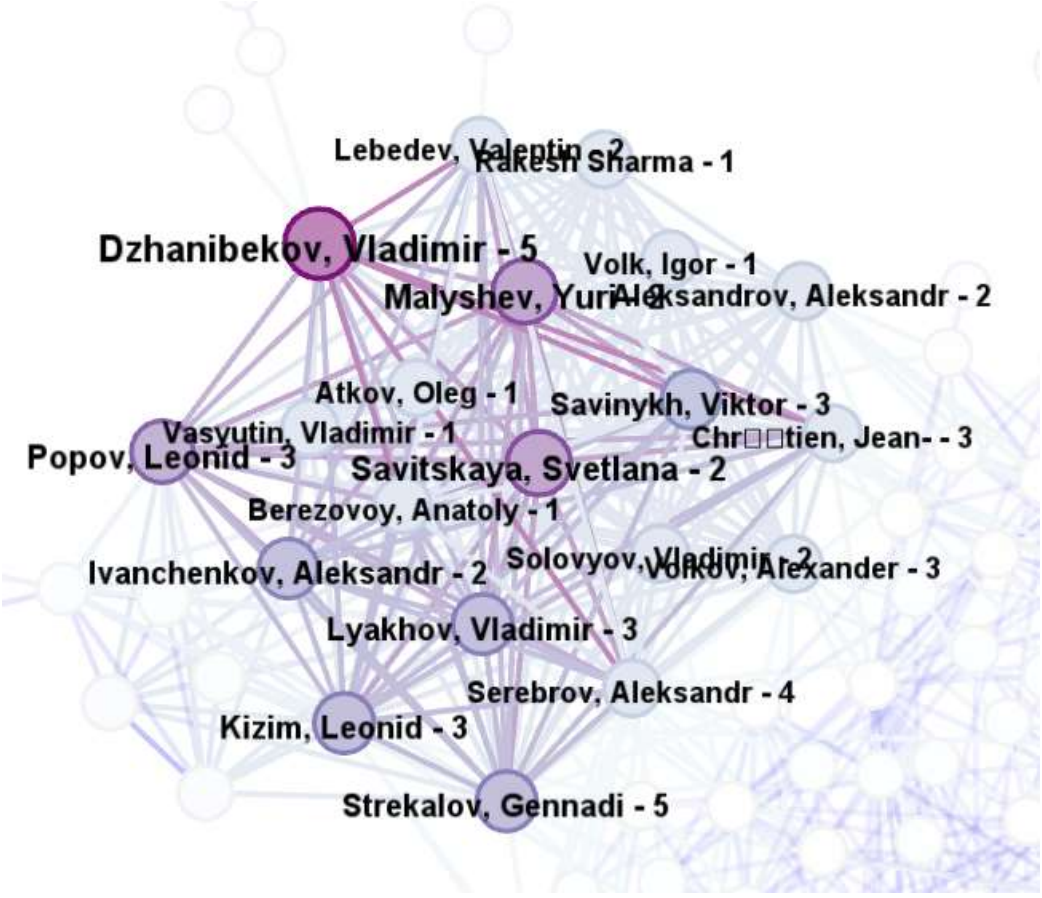
Concluding the centrality chapter (for now), we will discuss the eigenvector centrality, a measure of the influence of a node in a network, based on connected nodes. This is why we can see many similarities between the following table and the one of Degree table. Something weird is that opposed to previous centralities, this one's top 20 doesn't include any US astronauts, but 18 from USSR, 1 from France and 1 from India. The 3rd astronaut with the highest eigenvector centrality value is Svetlana Savitskaya, the 2nd woman who has been to space. In the end, we can see that the top 20 nodes are concentrated only on a part of the graph, due to the definition of eigenvector centrality. Number of missions doesn't matter here, as we see Svetlana Savitskaya 2nd on the list with only 2 missions. 8 civilians appear on our list and 12 militaries.

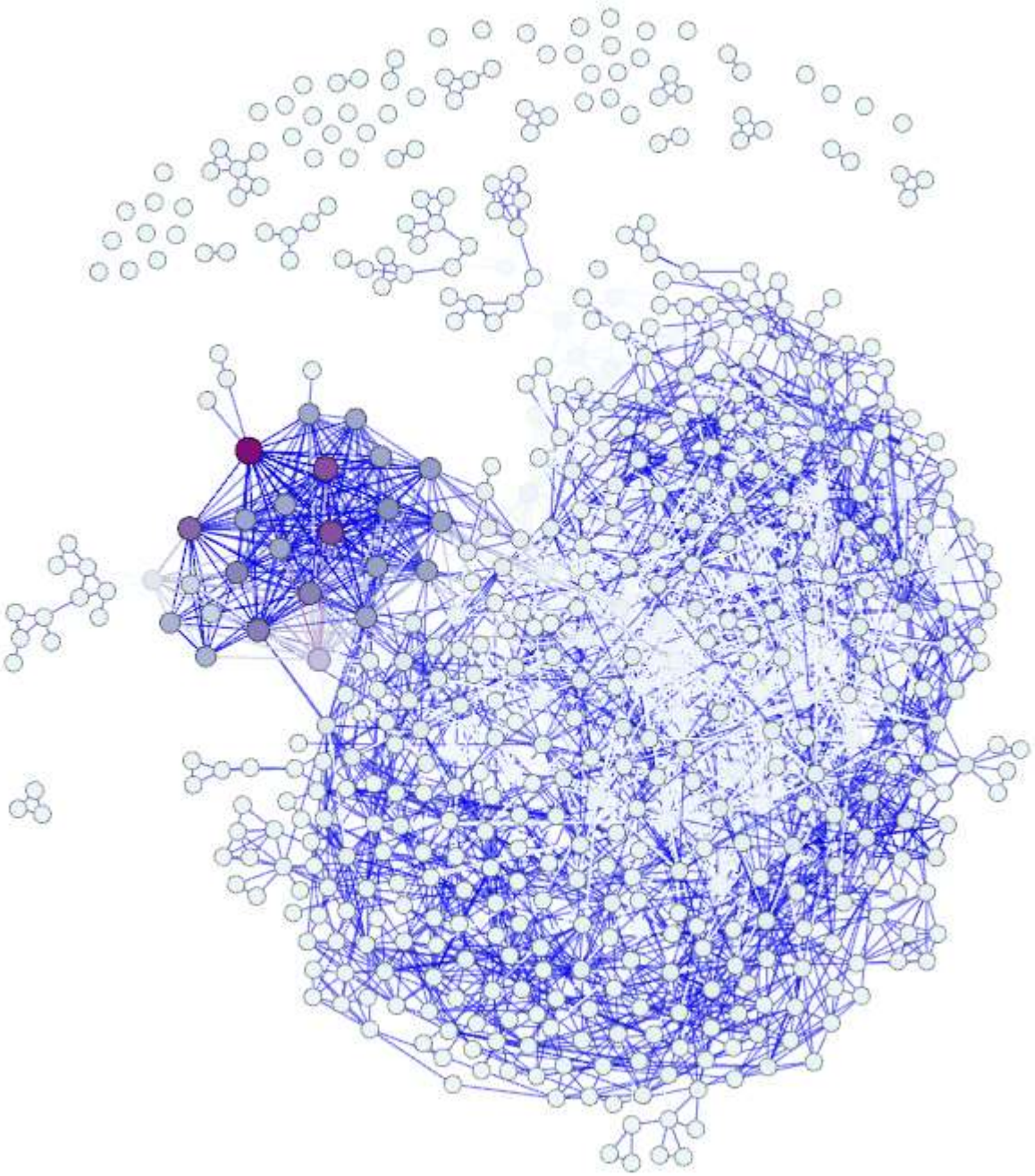
Id	Label	Eigenvector Centrality	Component ID	total_number_of_mis...	sex	nationality	occupation	military_civilian	year_of_selection	year_of_birth
86	Dzhanibekov, Vladimir	1.0	0	5	male	U.S.S.R/Russia	commander	military	1970	1942
111	Savitskaya, Svetlana	0.75364	0	2	female	U.S.S.R/Russia	MSP	civilian	1980	1948
95	Malyshev, Yuri	0.75364	0	2	male	U.S.S.R/Russia	commander	military	1967	1941
93	Popov, Leonid	0.673603	0	3	male	U.S.S.R/Russia	commander	military	1970	1945
98	Kizim, Leonid	0.581292	0	3	male	U.S.S.R/Russia	commander	military	1965	1941
99	Strekalov, Gennadi	0.577273	0	5	male	U.S.S.R/Russia	MSP	civilian	1973	1940
91	Lyakhov, Vladimir	0.570952	0	3	male	U.S.S.R/Russia	commander	military	1967	1941
88	Ivanchenkov, Aleksandr	0.556805	0	2	male	U.S.S.R/Russia	flight engineer	military	1973	1940
100	Savinykh, Viktor	0.518415	0	3	male	U.S.S.R/Russia	flight engineer	civilian	1978	1940
123	Aleksandrov, Aleksandr	0.443628	0	2	male	U.S.S.R/Russia	flight engineer	civilian	1978	1943
136	Solovyov, Vladimir	0.43869	0	2	male	U.S.S.R/Russia	flight engineer	civilian	1978	1946
110	Serebrov, Aleksandr	0.426376	0	4	male	U.S.S.R/Russia	flight engineer	civilian	1978	1944
183	Volkov, Alexander	0.424187	0	3	male	U.S.S.R/Russia	MSP	military	1976	1948
108	Chr��tien, Jean-	0.420494	0	3	male	France	MSP	military	1980	1938
70	Lebedev, Valentin	0.409728	0	2	male	U.S.S.R/Russia	flight engineer	civilian	1972	1942
107	Berezovoy, Anatoly	0.409118	0	1	male	U.S.S.R/Russia	commander	military	1970	1942
137	Atkov, Oleg	0.409118	0	1	male	U.S.S.R/Russia	MSP	civilian	1983	1949
138	Rakesh Sharma	0.409118	0	1	male	India	pilot	military	1982	1949
143	Volk, Igor	0.409118	0	1	male	U.S.S.R/Russia	MSP	military	1977	1937
182	Vasyutin, Vladimir	0.409118	0	1	male	U.S.S.R/Russia	commander	military	1976	1952

Network Interpretation: undirected
Number of iterations: 100
Sum change: 0.043803589954842674

Results:

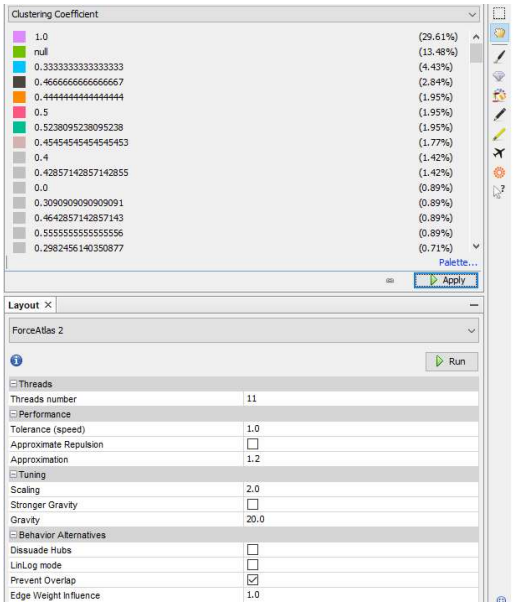


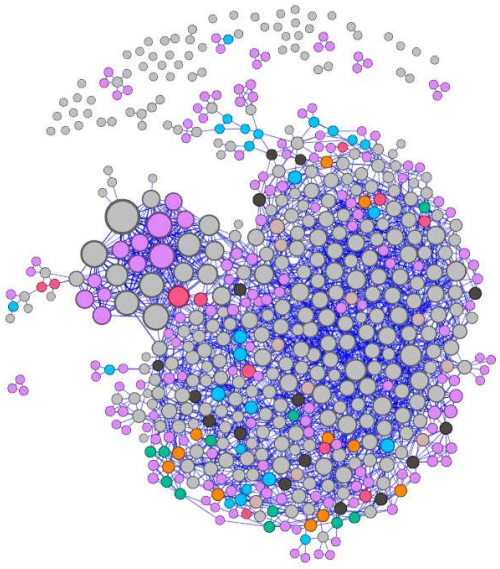


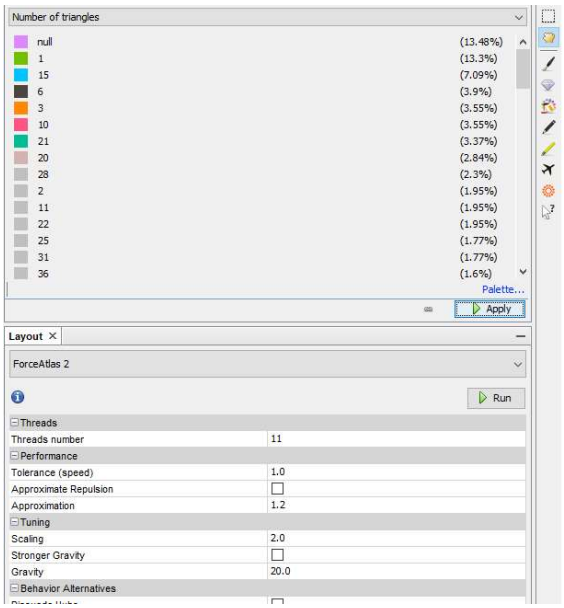


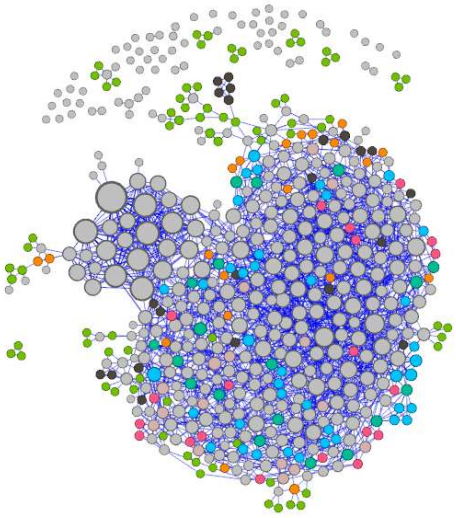
Clustering Coefficient

According to Gephi's Github: The neighborhood of a node, u, is the set of nodes that are connected to u. If every node in the neighborhood of u is connected to every other node in the neighborhood of u, then the neighborhood of u is complete and will have a clustering coefficient of 1. If no nodes in the neighborhood of u are connected, then the clustering coefficient will be 0. When running the default node overview statistic "Avg. Clustering Coefficient", we get that 81 nodes have clustering coefficient of 0, while 167 have 1, but the average is 0.592.





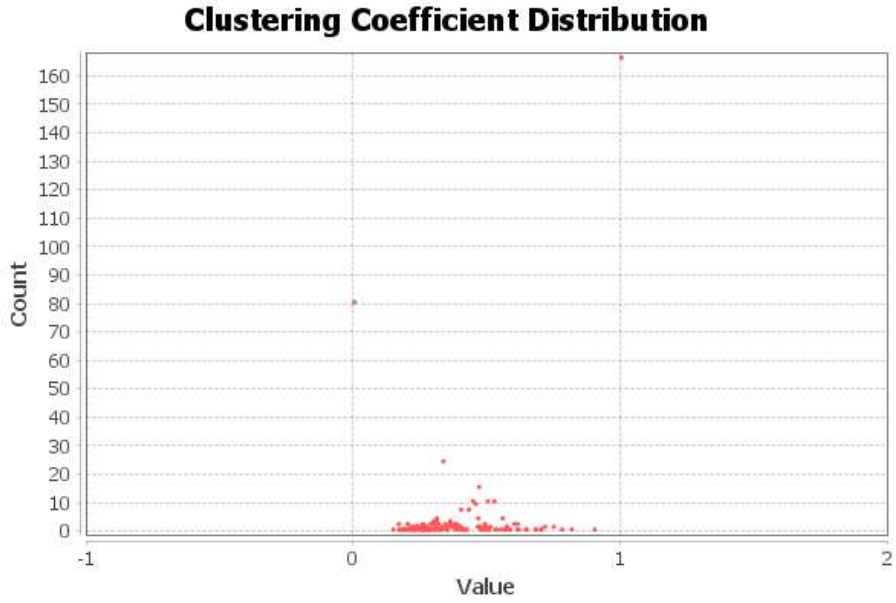




Now the definition of a triangle is: if nodes B and C have a friend A in common, then the formation of an edge between B and C produces a situation in which all three nodes A,B, and C have edges connecting each other. According to gephi there are 4335 triangles in this graph.

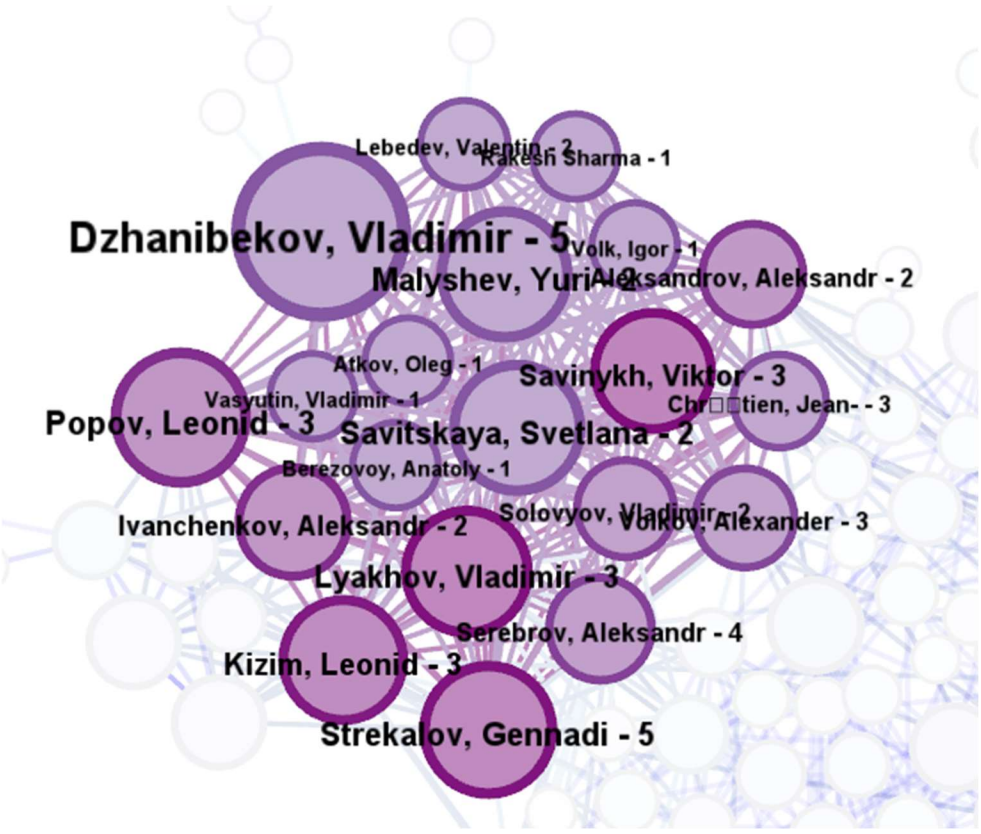
Results:

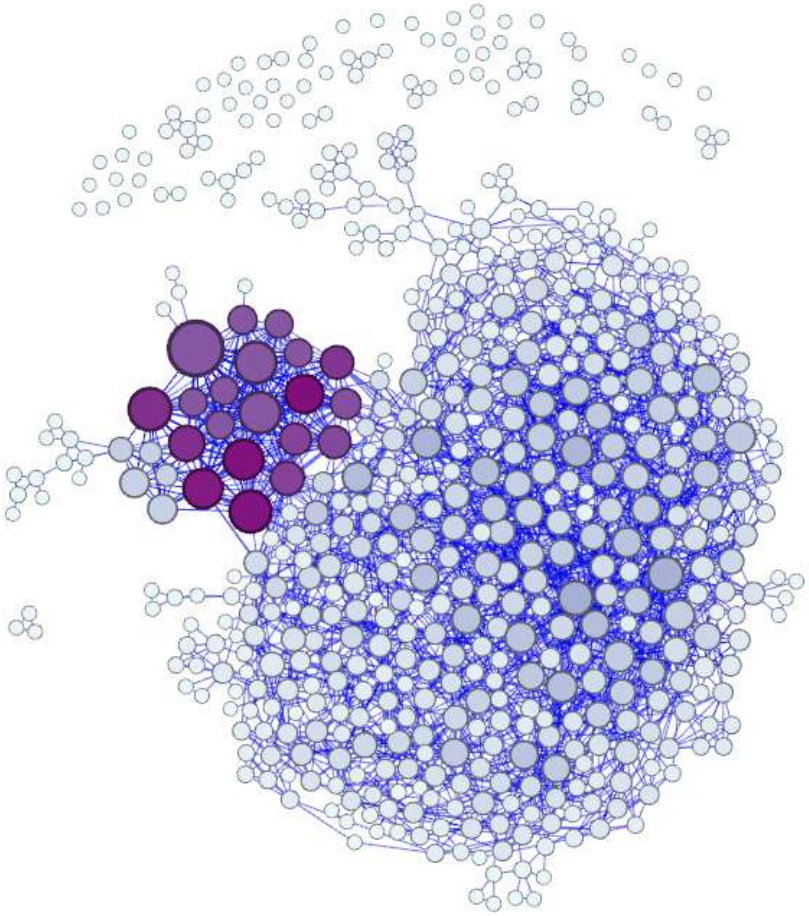
Average Clustering Coefficient: 0.592
Total triangles: 4335
The Average Clustering Coefficient is the mean value of individual coefficients.



Id	Label	Number of triangles	total_number_of_mis...	sex	nationality	occupation	military_civilian	year_of_selec...	year_of_birth
100	Savinykh, Viktor	234	3	male	U.S.S.R/Russia	flight engineer	civilian	1978	1940
91	Lyakhov, Vladimir	231	3	male	U.S.S.R/Russia	commander	military	1967	1941
99	Strekalov, Gennadi	228	5	male	U.S.S.R/Russia	MSP	civilian	1973	1940
98	Kizim, Leonid	223	3	male	U.S.S.R/Russia	commander	military	1965	1941
88	Ivanchenkov, Aleksandr	206	2	male	U.S.S.R/Russia	flight engineer	military	1973	1940
93	Popov, Leonid	206	3	male	U.S.S.R/Russia	commander	military	1970	1945
123	Aleksandrov, Aleksandr	199	2	male	U.S.S.R/Russia	flight engineer	civilian	1978	1943
110	Serebrov, Aleksandr	189	4	male	U.S.S.R/Russia	flight engineer	civilian	1978	1944
136	Solovyov, Vladimir	188	2	male	U.S.S.R/Russia	flight engineer	civilian	1978	1946
183	Volkov, Alexander	183	3	male	U.S.S.R/Russia	MSP	military	1976	1948
108	Chr��tien, Jean-	177	3	male	France	MSP	military	1980	1938
107	Berezovoy, Anatoly	171	1	male	U.S.S.R/Russia	commander	military	1970	1942
70	Lebedev, Valentin	171	2	male	U.S.S.R/Russia	flight engineer	civilian	1972	1942
86	Dzhanibekov, Vladimir	171	5	male	U.S.S.R/Russia	commander	military	1970	1942
95	Malyshev, Yuri	171	2	male	U.S.S.R/Russia	commander	military	1967	1941
111	Savitskaya, Svetlana	171	2	female	U.S.S.R/Russia	MSP	civilian	1980	1948
137	Atkov, Oleg	171	1	male	U.S.S.R/Russia	MSP	civilian	1983	1949
138	Rakesh Sharma	171	1	male	India	pilot	military	1982	1949
143	Volk, Igor	171	1	male	U.S.S.R/Russia	MSP	military	1977	1937
182	Vasyutin, Vladimir	171	1	male	U.S.S.R/Russia	commander	military	1976	1952

Sorting by number of triangles we can see that there are astronauts with many triangles. This list is quite similar with the one of eigenvector centrality. Even the concentration of nodes is almost identical to the one of eigenvector centrality!

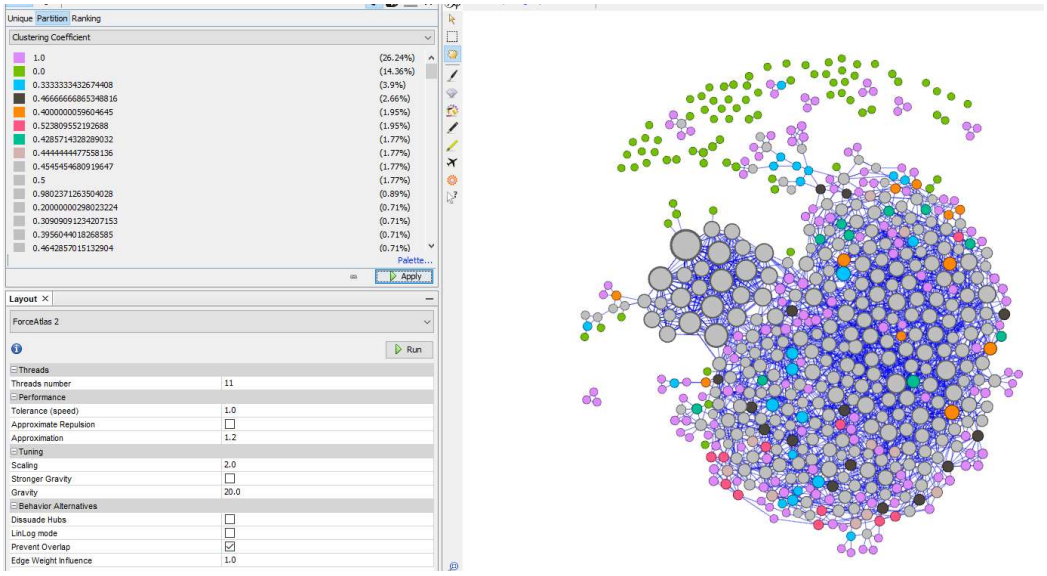




There is a Gephi plugin named Clustering coefficient which gives different numbers, to the default one. It has 2 methods, one basic method and one triangle method.

The Basic Method gives an average coefficient of ~0.516, a value slightly lower than the default method.

General C = 0.5163995368102341

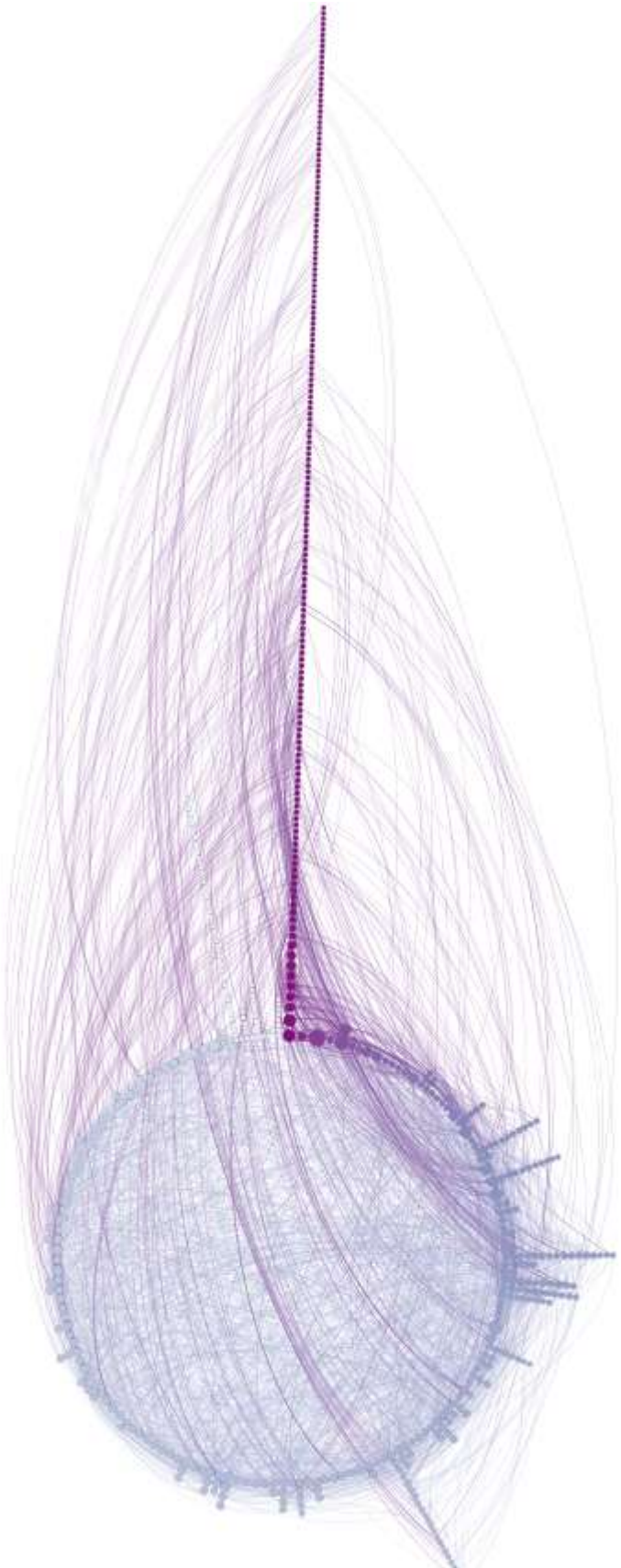


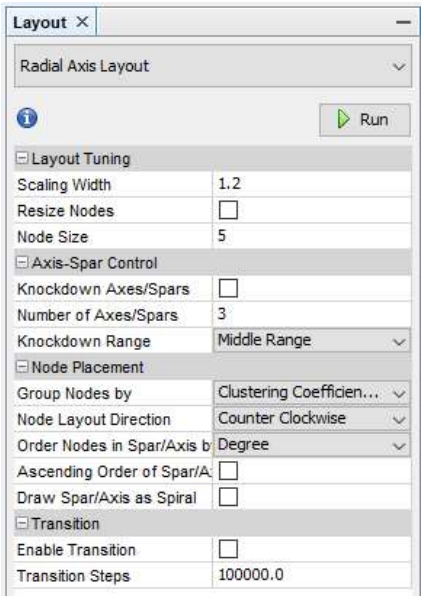
The Triangle Method has average coefficient 0.4810505509376526 and 7015 triangles.

Number of triangles: 7015
Number of paths (Length 2): 43748
Value of Clustering Coefficient: 0.4810505509376526

We can see that despite the average coefficient being lower, the number of triangles is greatly increased and the astronauts that had null clustering coefficient are eliminated!

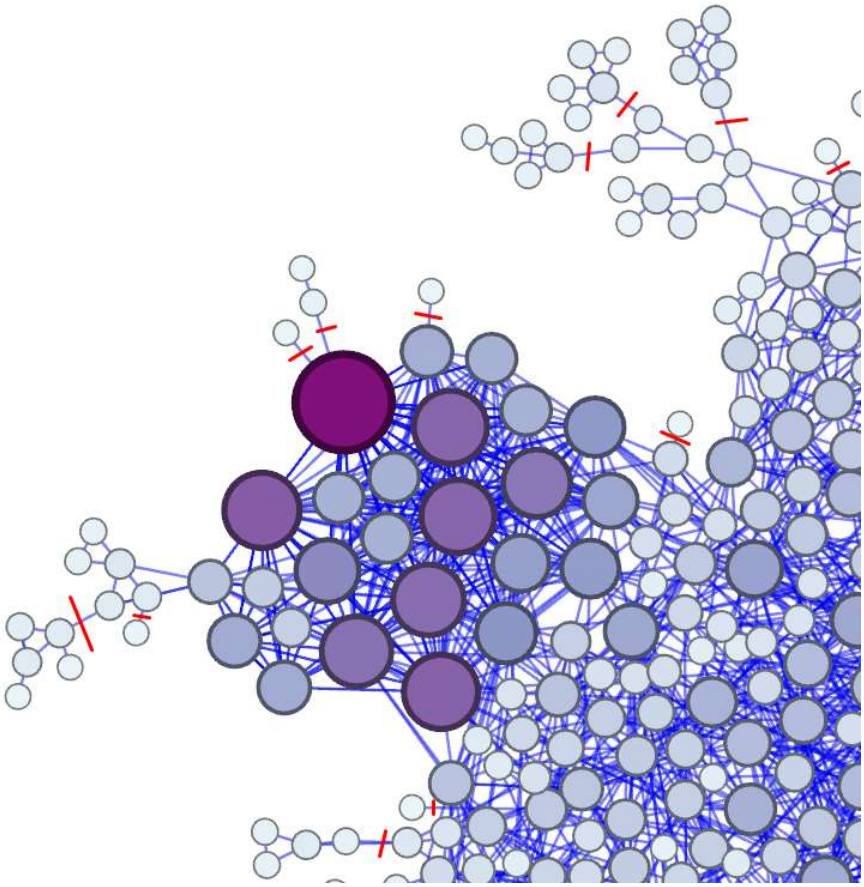
Here, we are switching to Radial Axis Layout (included in Circular Layout Plugin), to better demonstrate the clustering coefficient, grouping by it.





Bridges

By definition, a bridge is an edge of a graph whose deletion increases the graph's number of connected components. Unfortunately, I was not able to find any tools or plugins to locate bridges with Gephi. As I was searching the plugin tab I found Bridging Centrality, which is a centrality index based on information flow and topological locality in networks. The following image is from our graph, ranked by degree. If we remove the edges crossed by a red line, the definition of bridge is fulfilled, so these edges are bridges!

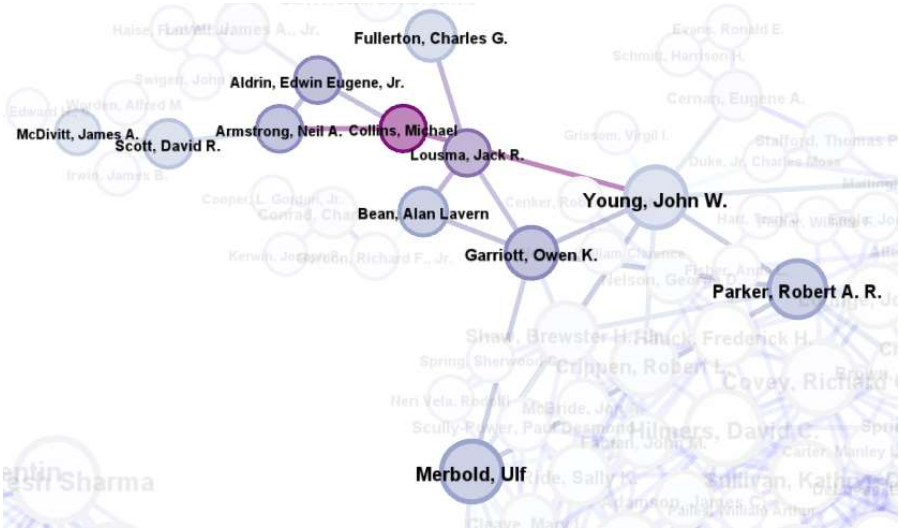


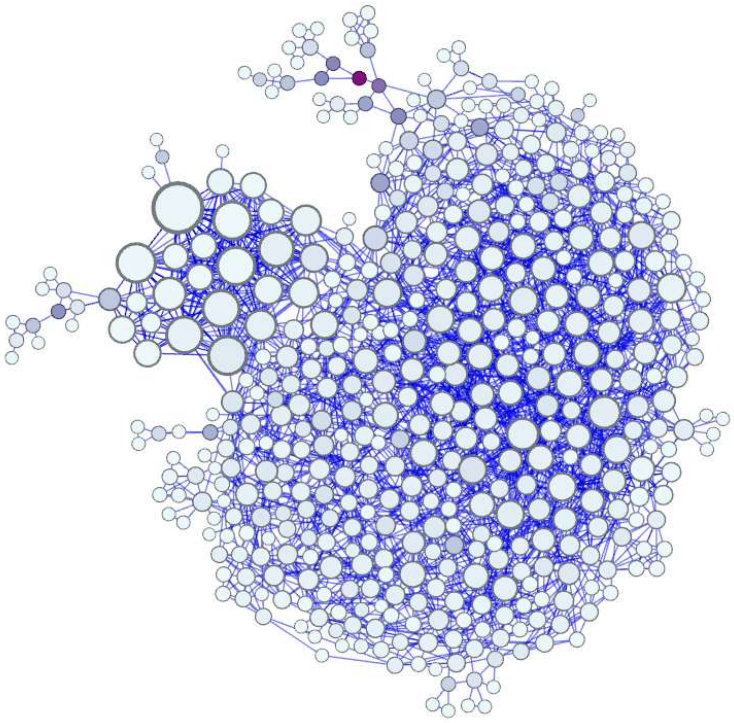
Bridging Centrality

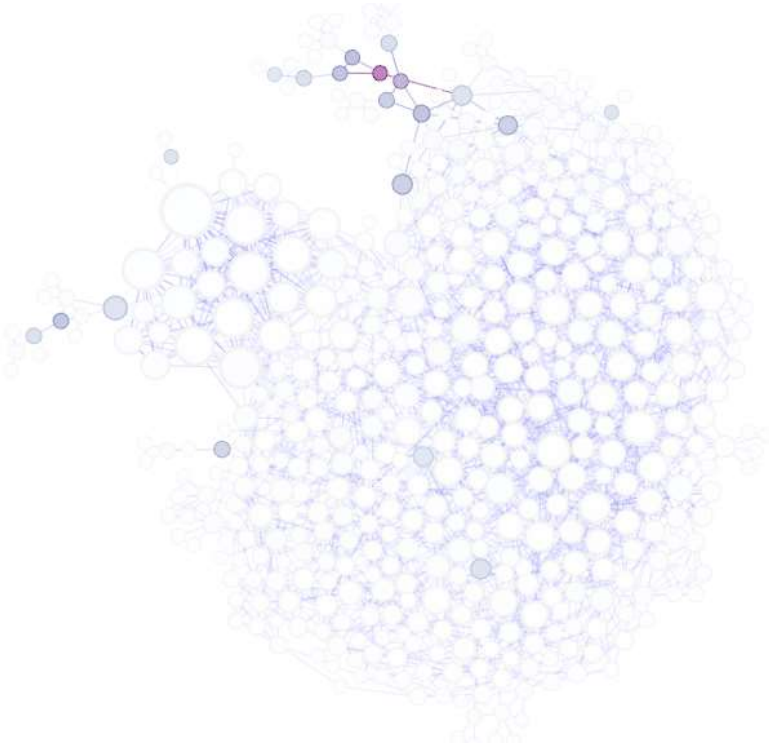
The bridging centrality that we discussed above, appears to be quite important as it emphasizes both on information flow and the probability of finding a local bridge. We had to remove the weakly connected components again, as an exception was thrown. This top 20's values are quite low as the top one is Michael Colins with ~0.02 bridging centrality. Number of missions doesn't matter and all astronauts are males. In addition 5 out of 20 are from USSR, 1 is from Germany and the res 14 are form USA. 11 are militaries and 9 are civilians.

Id	Label	Bridging Centrality	total_number_of_missions	sex	nationality	occupation	military_civilian	year_of_selec...	year_of_birth
27	Collins, Michael	0.021626763913...	2	male	U.S.	pilot	civilian	1963	1930
62	Lousma, Jack R.	0.011236859183...	2	male	U.S.	pilot	military	1966	1936
29	Aldrin, Edwin Eugene, Jr.	0.008315275795...	2	male	U.S.	pilot	military	1963	1930
63	Garriott, Owen K.	0.007780874801...	2	male	U.S.	MSP	civilian	1965	1930
24	Armstrong, Neil A.	0.007559341632...	2	male	U.S.	commander	civilian	1962	1930
41	Filipchenko, Anatoly	0.006848822118...	2	male	U.S.S.R/Russia	commander	military	1963	1928
44	Bean, Alan Lavern	0.004032044954...	2	male	U.S.	pilot	military	1963	1932
131	Merbold, Ulf	0.003955854729...	3	male	Germany	PSP	civilian	1978	1941
129	Parker, Robert A. R.	0.003575958435...	2	male	U.S.	MSP	civilian	1967	1936
506	Surayev, Maksim	0.003257870244...	2	male	U.S.S.R/Russia	commander	military	1997	1972
106	Fullerton, Charles G.	0.002499867871...	2	male	U.S.	pilot	military	1969	1936
25	Scott, David R.	0.002277411822...	3	male	U.S.	pilot	military	1963	1932
51	Rukavishnikov, Nikolai	0.002276210655...	3	male	U.S.S.R/Russia	flight engineer	military	1968	1932
75	Grechko, Georgi	0.002128418484...	3	male	U.S.S.R/Russia	flight engineer	civilian	1968	1931
398	Burbank, Daniel C.	0.002120745284...	3	male	U.S.	MSP	military	1996	1961
17	Young, John W.	0.002100459619...	6	male	U.S.	pilot	military	1962	1930
65	Makarov, Oleg	0.002064671875...	3	male	U.S.S.R/Russia	flight engineer	civilian	1968	1933
148	Walker, Charles David	0.001886792452...	3	male	U.S.	PSP	civilian	1983	1948
362	Lu, Edward T.	0.001693560064...	3	male	U.S.	MSP	civilian	1995	1963
18	McDivitt, James A.	0.001677148846...	2	male	U.S.	commander	military	1962	1929

We can see at the graphs below, that at the top right part of our graph there are many nodes from our top 20 list. These nodes and edges connect many mini-graphs and if we remove them these parts become weakly connected components, making some of these edges bridges.







Modularity

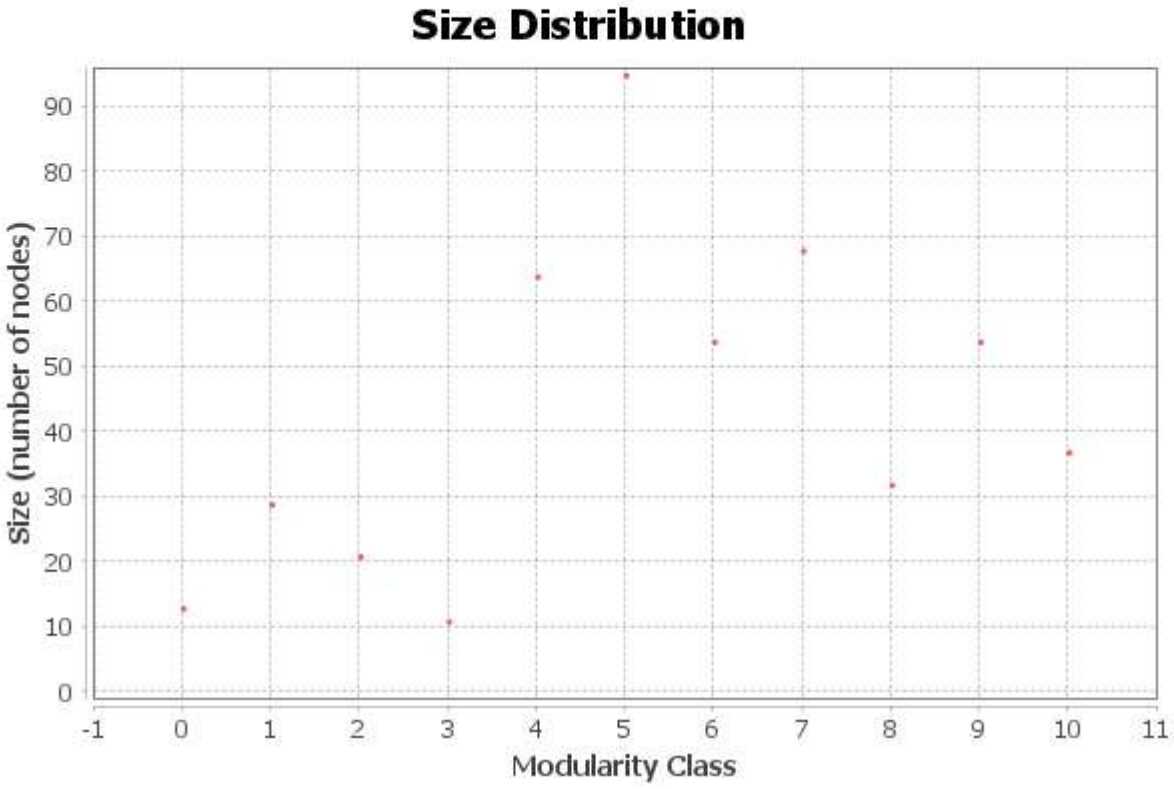
To see the modularity in this graph we will exclude the weakly connected components as they are not helping with determining communities and will create some of their own. We want to see the communities formed in the strongly connected nodes. Modularity measures the strength of division of a network into communities (modules). This is why filter Giant Component was important. We use resolution 1 and color the nodes based on their modularity class. As we can see they tend to create communities, 11 to be exact, with a modularity of 0.574.

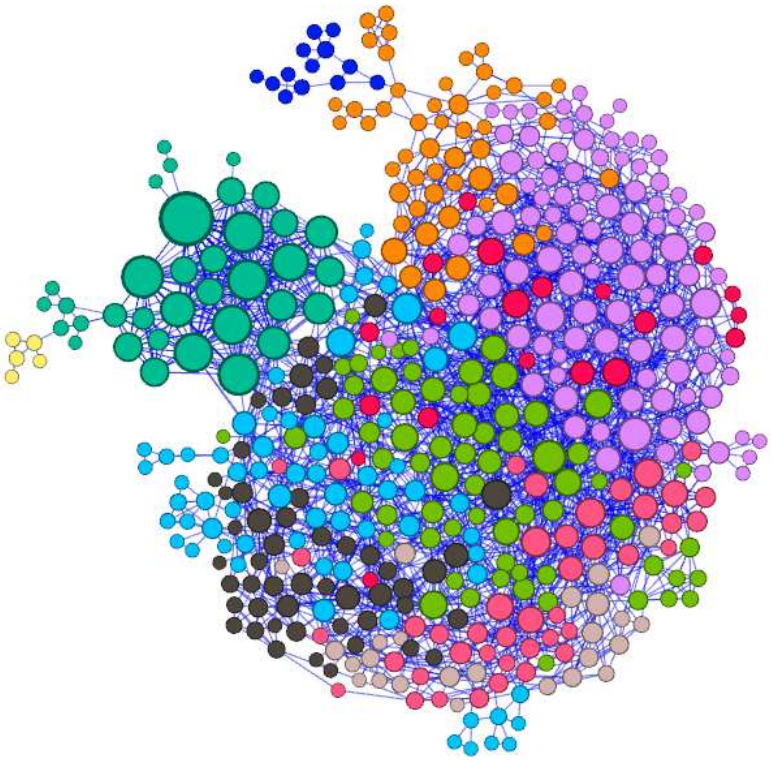
Parameters:

Randomize: On
Use edge weights: Off
Resolution: 1.0

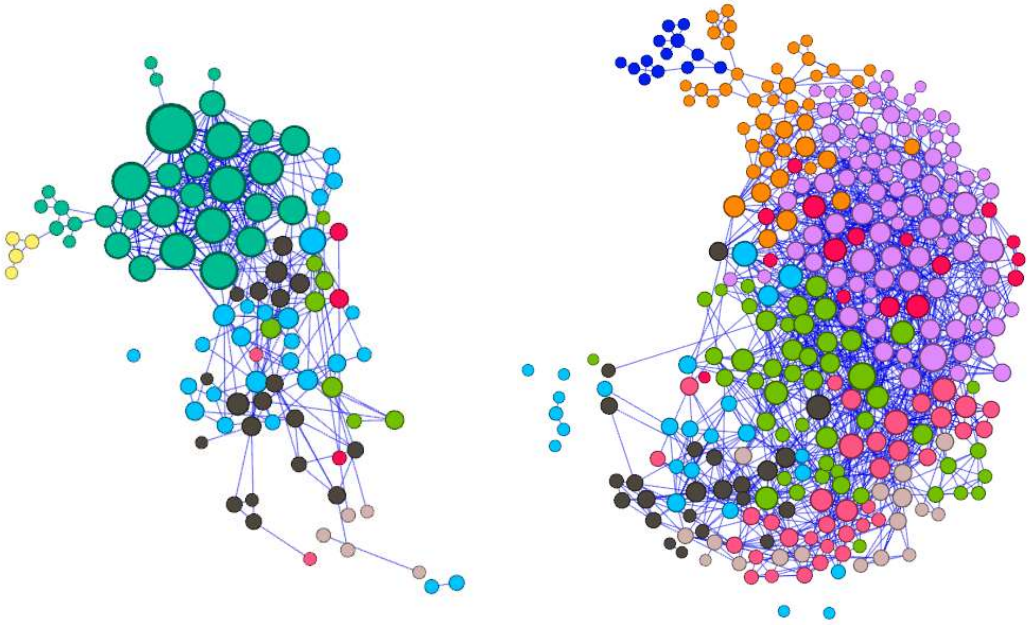
Results:

Modularity: 0.574
Modularity with resolution: 0.574
Number of Communities: 11

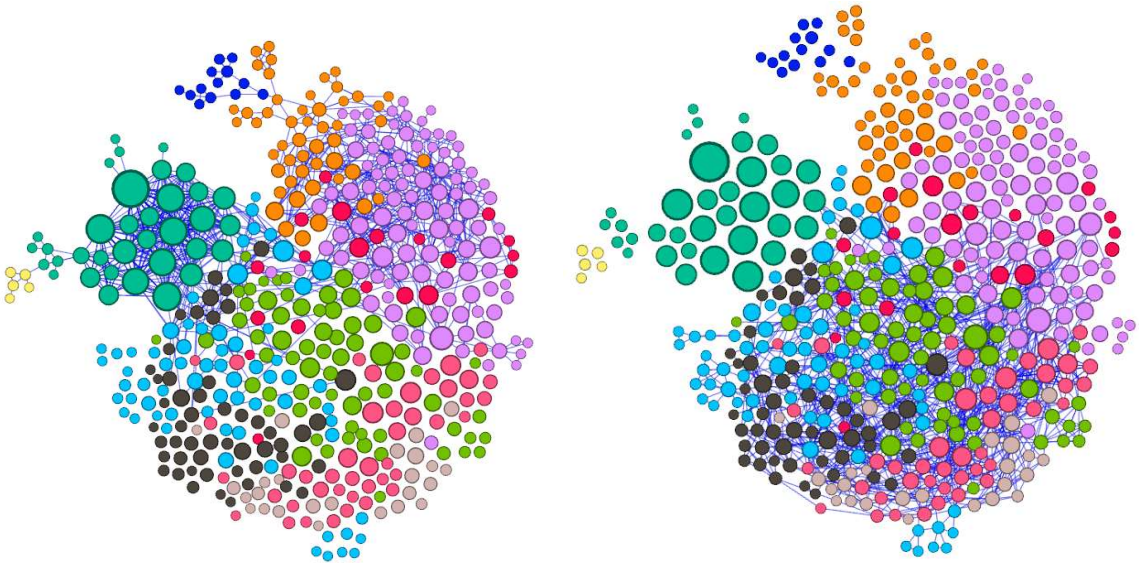




Filtering by country, we see USSR on the left and USA on the right.



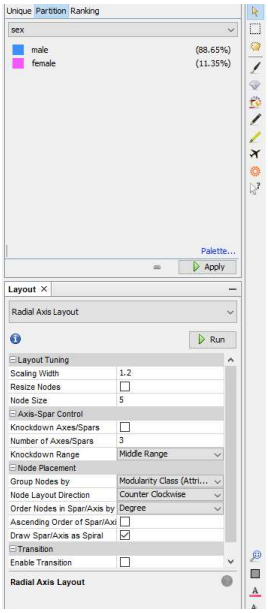
Finally observing the year of the mission we can see that communities are separated in 2 main categories based on if they were on cold war era (1947-1991) or after(1991-now). On the left there are the cold war mission links and on the right after 1991 links. While filtering sex and occupation, I couldn't see any major community being formed, based on modularity.

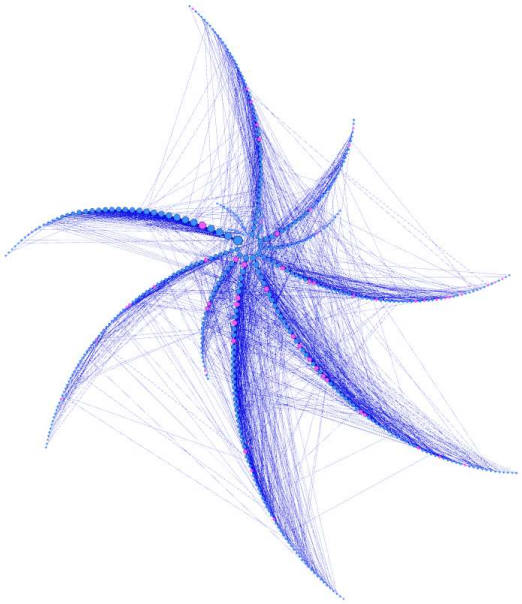


Homophily

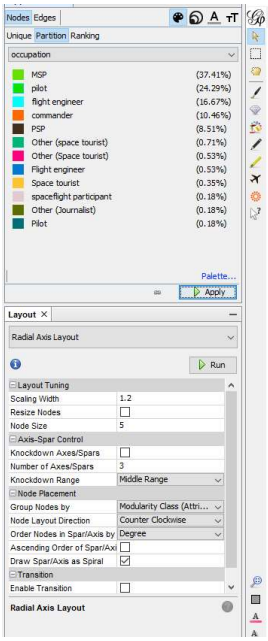
Homophily measures the tendency of individuals to associate and bond with similar others. We will examine astronauts based on some of the node characteristic using modularity class. We will once again keep the nodes of the big component. Following Gephi's tutorial we are using Radial Axis layout, grouping by modularity class and with Draw spar/axis as spiral checked, to better show links inside communities.

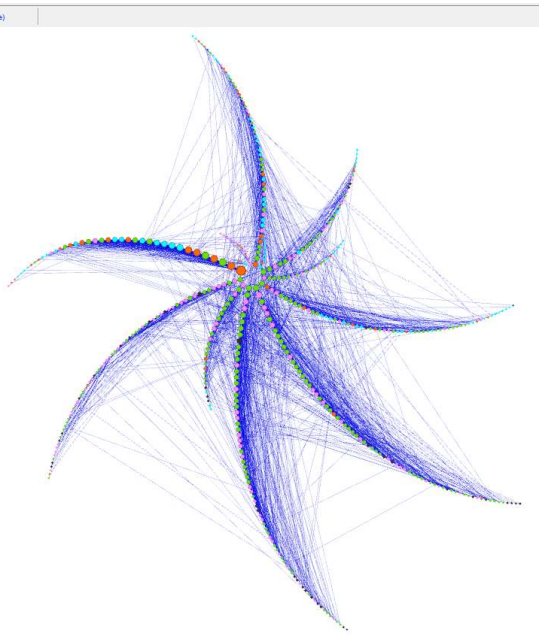
Firstly we see if gender plays any role in the formation of communities. It seems like it does, but male astronauts are 8 times more than females.



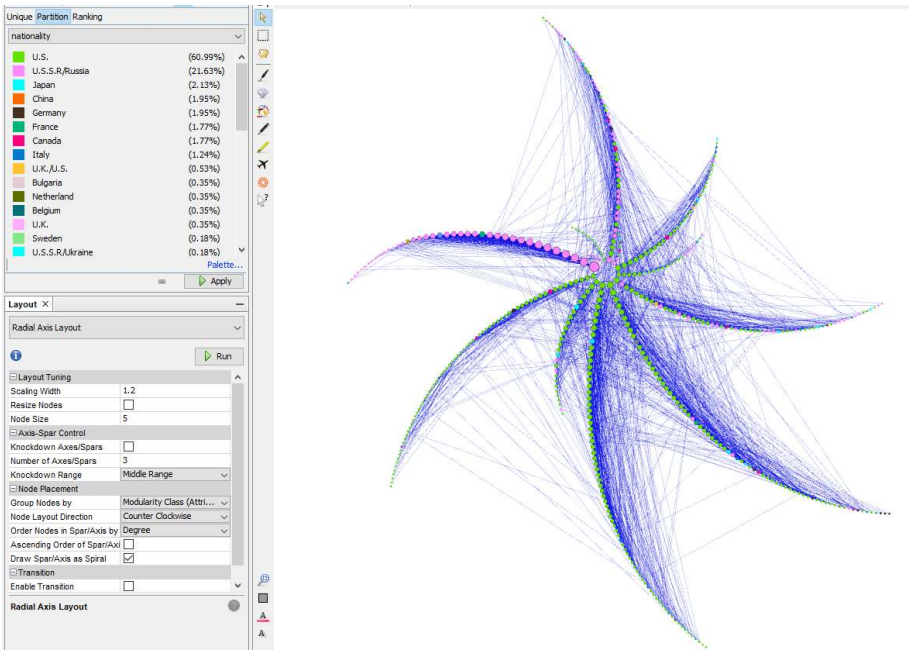


Secondly, we will examine occupation of astronauts. It doesn't seem to matter when forming communities, which is normal as most missions consist of astronauts with different roles.

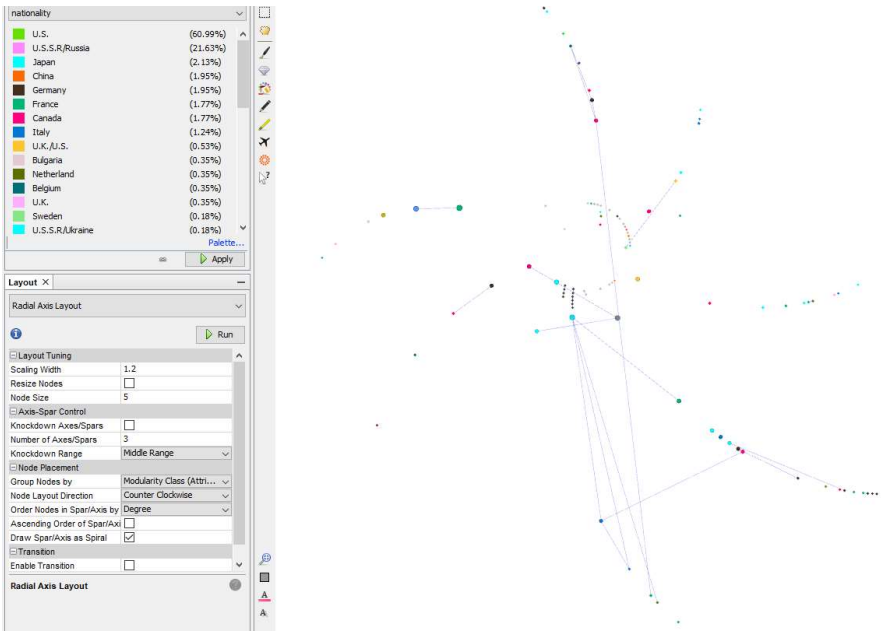




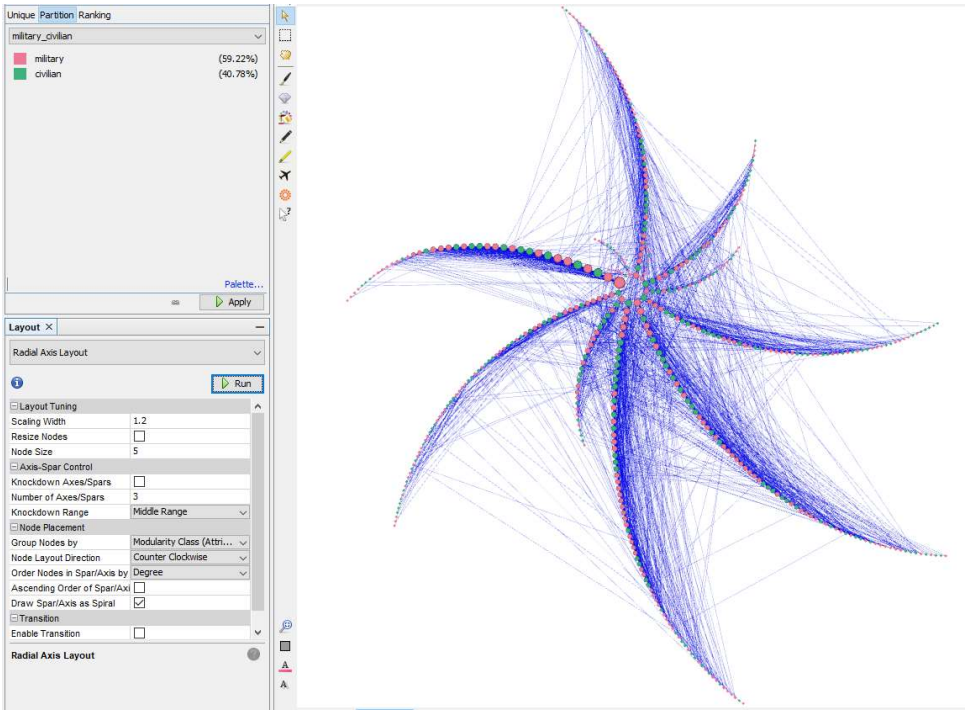
Thirdly, we will filter by nationality. Although it seems like communities are formed more than 80% are US and USSR, so we can't be sure.



Even if we exclude USA and USSR, we will see almost no communities formed, as there are only 98 astronauts by other countries other than USSR and USA.



Finally, we can see that when comparing civilians with militaries, communities don’t tend to be formed.



Conclusion

Astronauts’ “universe” is quite complex. Many famous astronauts don’t play a major role in this network, while many unknown ones are pretty important. Some communities are formed, although, due to the nature of this network we can’t be so sure. We can see that after the cold war, countries started working together more often and that there are not many prejudices. Despite the lack of female astronauts, they do play a major role in all top 20 tables. Being military or civilian doesn’t play a major role as they are quite equally distributed. As Valentina Tereshkova said, “It doesn't matter what country or what political system you are from. Space brings you together.” and that is what this report demonstrated!

Sources

https://gephi.org/tutorials/gephi-tutorial-layouts.pdf

https://www.kaggle.com/jessemostipak/astronaut-database

https://seinecle.github.io/gephi-tutorials/generated-html/simple-project-from-a-to-z-en.html

https://seinecle.github.io/gephi-tutorials/generated-html/using-filters-en.html

https://www.aueb.gr/el/content/λογότυπα