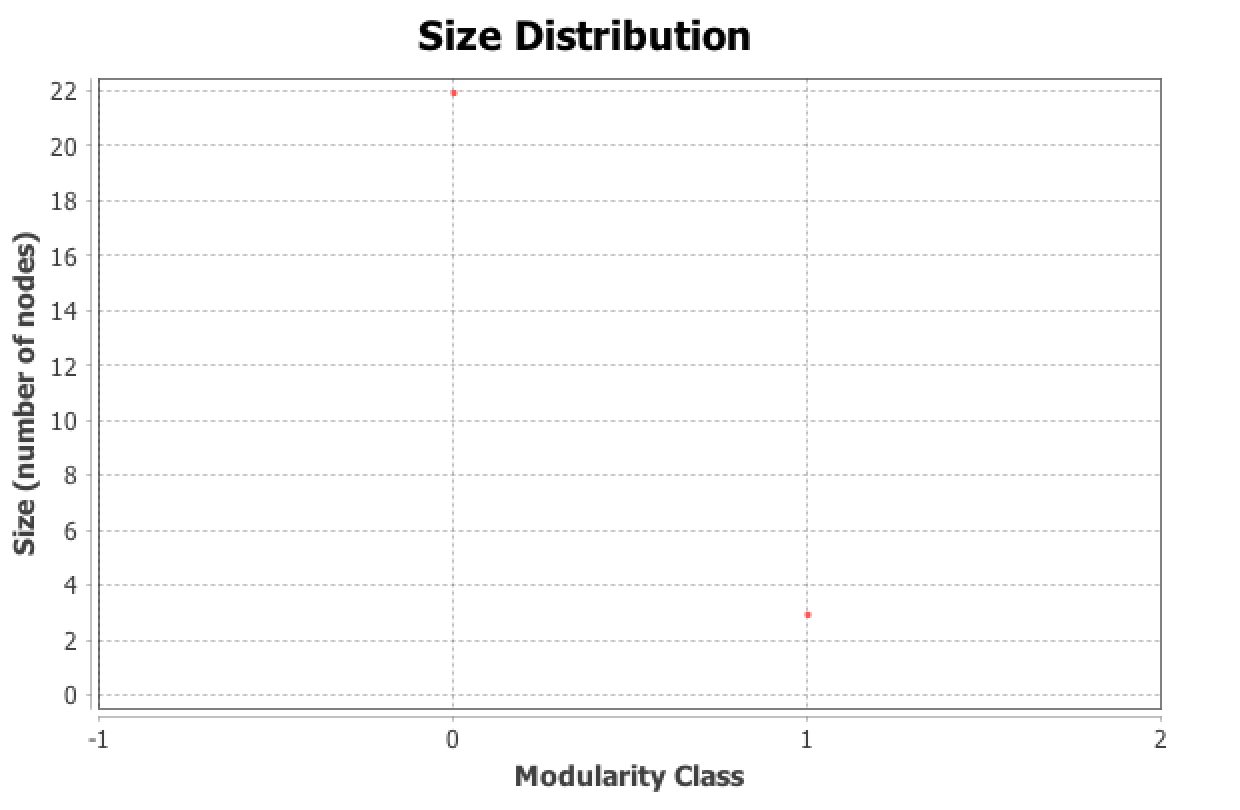
**Assignment 3**

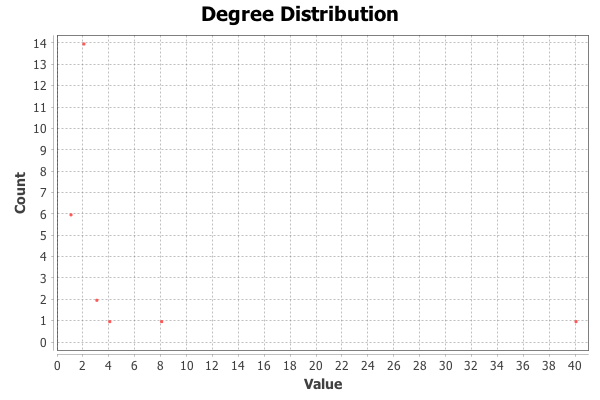
1. The dataset given to us is an email network. To import the dataset into Gephi to visualize it, I modified the headings ‘From’ and ‘To’ to ‘Source’ and ‘Target’ respectively so that the dataset can be imported as an edge list. I modified the date column to a date format acceptable by Gephi. This allowed me to use the dates as a timestamp. As we inherently don’t provide weights to these edges and Gephi doesn’t support parallel edges, multiple edges from one node to another where transformed into a weighted edge, with the weight signifying the number of emails sent from that node to a particular recipient.
2. I ran the inbuilt community detection algorithm using the randomizing and weights as parameters with a resolution of 0.5. This resolution would help me get more communities in the graph. I got the following results: -

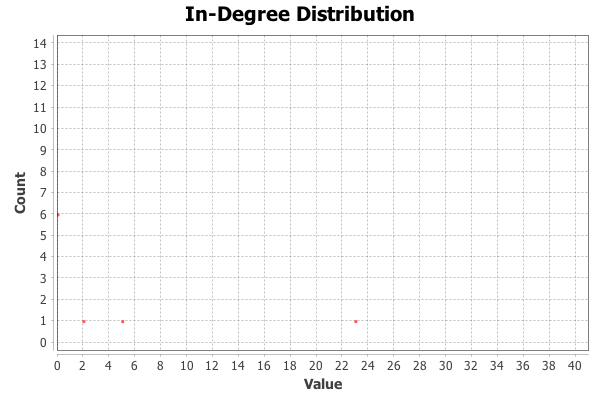
* Modularity: 0.070
* Modularity with resolution: -0.368
* Number of Communities: 2

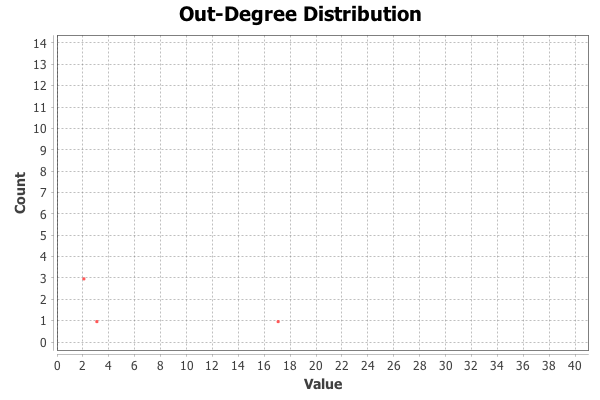


1. I got the average degree centrality as 3.680.

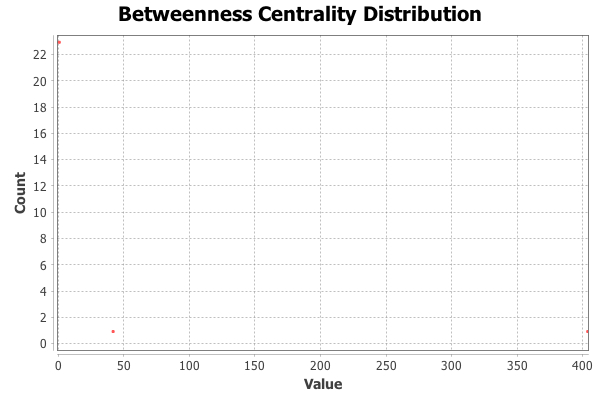
The degree centrality of an undirected network means the average number of nodes accessible by a node. In an email network like ours, which is directed, the degree centrality will be the average of the in-degree as well as the out-degrees. This means that on an average, people in this network have contacted or have been contacted by someone else 1.84 times, i.e, their average in-degree and out-degree is 1.84.







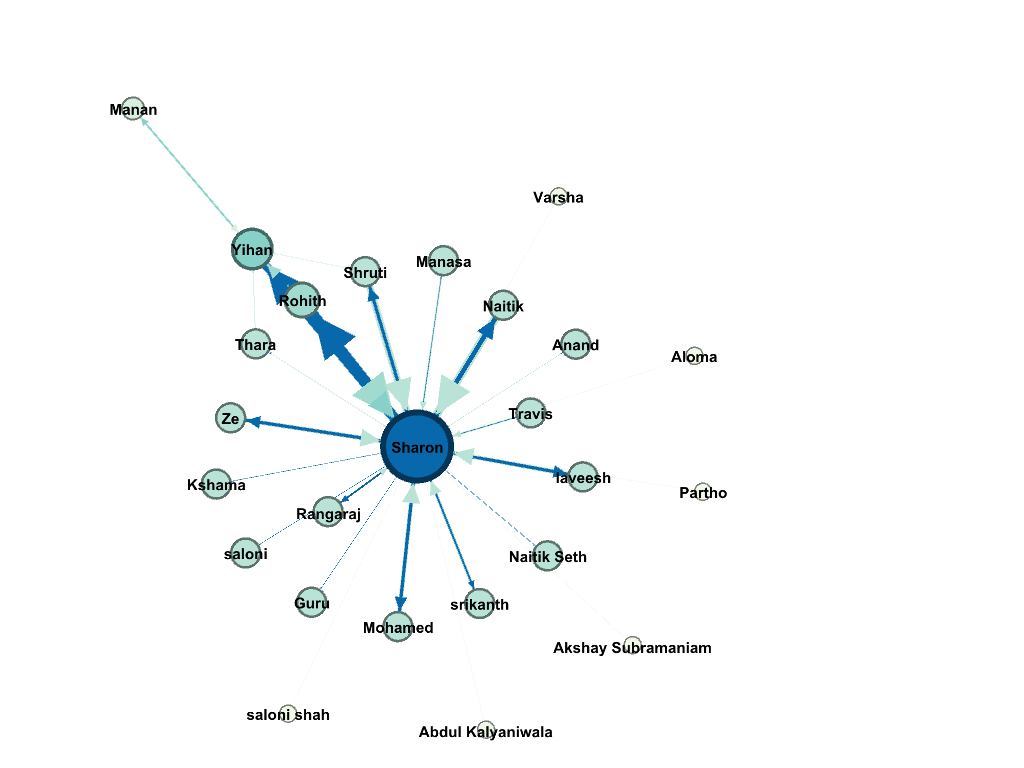
1. Betweenness centrality measures how often a node appears on the shortest paths from the nodes in the network. For this network the average betweeness centrality comes out to be **17.76**. This means that an email sender/receiver in this network on an average comes 17.76 times when trying to find the shortest path between two nodes.



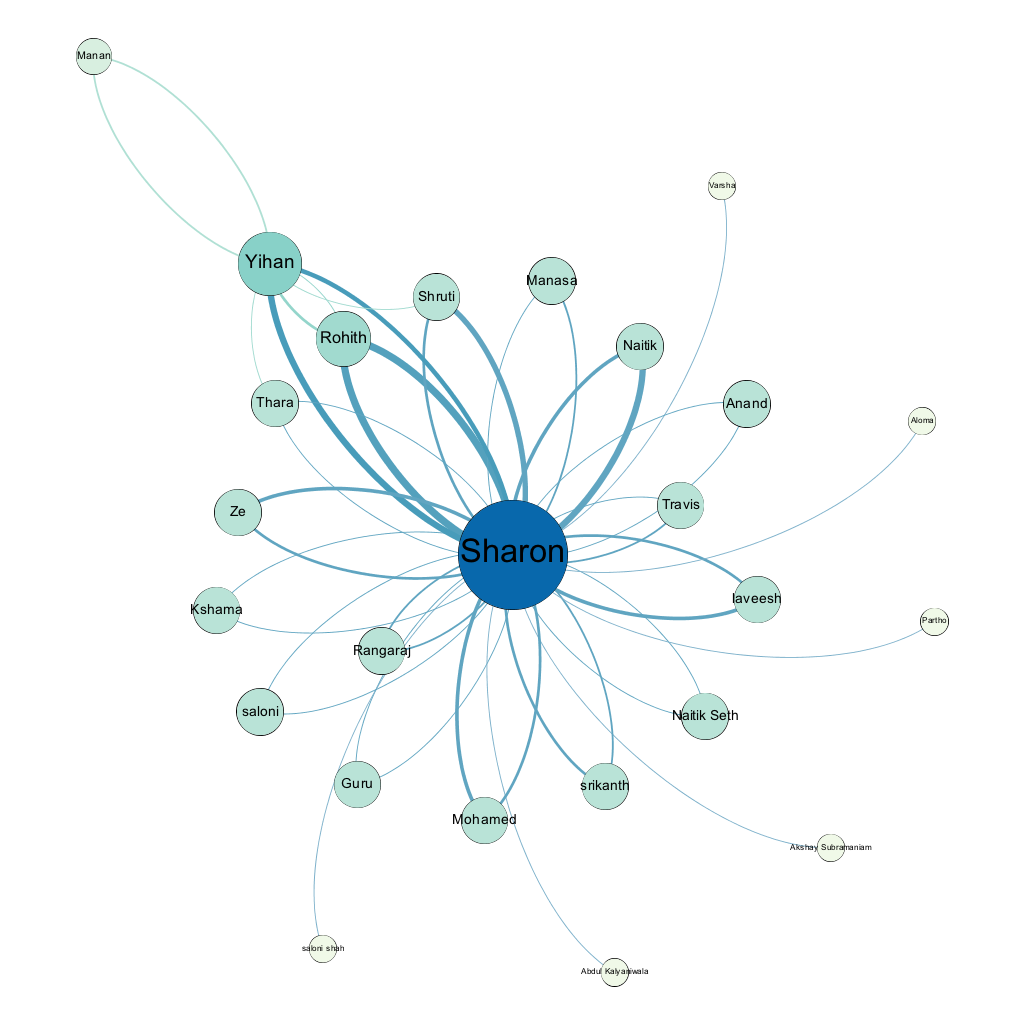
1. Eigenvector centrality is the measure of nodes importance in a network based on the nodes connections. Now, if we check each nodes eigenvector centrality, we get that ‘Sharon’ is the most important node with an eigenvector centrality of 1.0.

The average eigenvector centrality of the network will give us on an average how important is each node in the network. For this network, it comes out to be 0.2151.

1. The Reference Layout that I got is this.



The visualization generated on exporting is shown below



The nodes have been colored and sized based on the eigenvector centrality. The minimum node size has been set to 15 and the maximum node size to 60. This allows the nodes to be sufficiently big, hence visualized easily. The label names are the people in the network who have sent emails.

I decided to choose ‘Yifan Hu’. ‘Yifan Hu’ is a force directed method which gives us answers much faster than the other ones (Force Atlas, Fruchterman Reingold). It makes a solar-system type of network with the most important and the largest node being the “sun”. Other nodes (Yihan) can have moons(Manan). Our dataset had a lot of correspondences to and from “Sharon”, hence it is perfectly alright to make “Sharon” the center of our network. Since Yifan Hu focuses on the attraction and repulsion in the neighbourhood, it is perfect for our email network since the nodes with similar harmonic betweeness, have been potrayed around the same radius with the most important node being in the centre. Since the modularity is 2, we can see that there is a mini network around Yihan, so this graph depicts that, keeping in mind the overall betweeness of each node. If certain nodes have the same betweeness, their weight is taken into account, which gives us a clear idea of close are the nodes to our centre- Sharon.