Network Analysis

By Abhinav Chadha, Sanjeet Walia

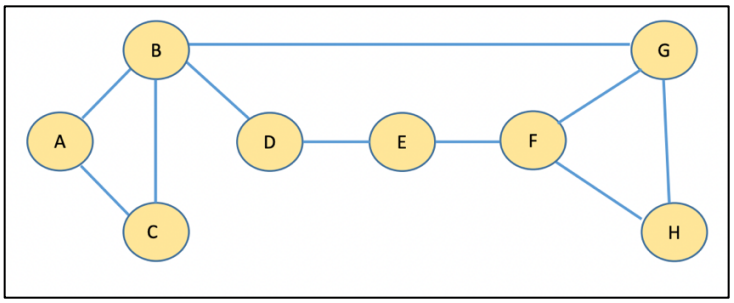
Network analysis is a set of integrated techniques to depict relations among actors (nodes) and to analyze the social structures that emerge from the recurrence of these relations. The basic assumption is that better explanations of social phenomena are yielded by analysis of the relations among entities.

To understand the importance of a node in a system, we measure its Network using three centralities:

1. ***Degree Centrality***
   Refers to the total number of nodes connected to the actor in discussion. To calculate Degree Centrality, we take numerical sum of the edges connected to the node.

$$DC(x) = \sum nodes\ originating\ from\ x$$

Example,



$DC(A) = 2$
$DC(B) = 4$
And so on.

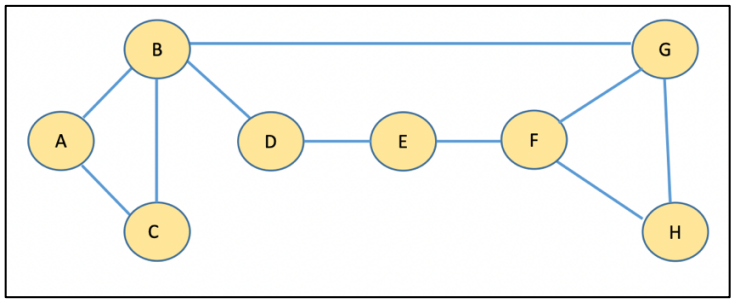
2. ***Closeness Centrality***
   Closeness centrality indicates how close a node is to all other nodes in the network. It is calculated as the average of the shortest path length from the node to every other node in the network.

The closeness centrality of a node x is given by,

$$CC(x) = \frac{\mathrm{N} - 1}{\sum_{y \neq x} d(y,x)}$$

$d(y,x) \rightarrow length\ of\ shortest\ path\ between\ nodes\ y\ and\ x$

$$N \rightarrow total\ number\ of\ nodes$$



Example, the closeness centrality of node F is calculated as,
$N = 8$
$d(A,F) = A \rightarrow B \rightarrow G \rightarrow F = 3$
$d(B,F) = B \rightarrow G \rightarrow F = 2$
$d(C,F) = C \rightarrow B \rightarrow G \rightarrow F = 3$
$d(D,F) = D \rightarrow E \rightarrow F = 2$
$d(E,F) = E \rightarrow F = 1$
$d(G,F) = G \rightarrow F = 1$
$d(H,F) = H \rightarrow F = 1$

Putting the above values in formula,

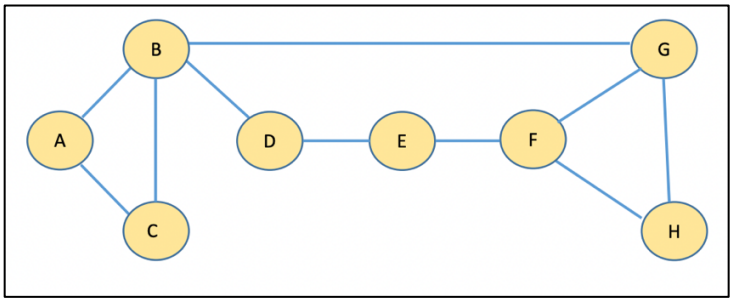
$$CC(F) = \frac{8-1}{3+2+3+2+1+1+1} = \frac{7}{13} = 0.538$$

3. *Betweenness Centrality*

As an indicator of how many paths a node is on, its betweenness centrality demonstrates how many paths it is a part of, which represent that node's ability to make connections to other groups in the graph.

The betweenness centrality of a node v is given by,

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



Example, the betweenness centrality of Node D is given by,

$$\frac{\sigma_{AB}(D)}{\sigma_{AB}} = \frac{0}{1} = 0 \qquad \frac{\sigma_{BE}(D)}{\sigma_{BE}} = \frac{1}{1} = 1 \qquad \frac{\sigma_{CH}(D)}{\sigma_{CH}} = \frac{0}{1} = 0$$

$$\frac{\sigma_{AC}(D)}{\sigma_{AC}} = \frac{0}{1} = 0 \qquad \frac{\sigma_{BF}(D)}{\sigma_{BF}} = \frac{0}{1} = 0 \qquad \frac{\sigma_{EF}(D)}{\sigma_{EF}} = \frac{0}{1} = 0$$

$$\frac{\sigma_{AE}(D)}{\sigma_{AE}} = \frac{1}{1} = 1 \qquad \frac{\sigma_{BG}(D)}{\sigma_{BG}} = \frac{0}{1} = 0 \qquad \frac{\sigma_{EG}(D)}{\sigma_{EG}} = \frac{0}{1} = 0$$

$$\frac{\sigma_{AF}(D)}{\sigma_{AF}} = \frac{0}{1} = 0 \qquad \frac{\sigma_{BH}(D)}{\sigma_{BH}} = \frac{0}{1} = 0 \qquad \frac{\sigma_{EH}(D)}{\sigma_{EH}} = \frac{0}{1} = 0$$

$$\frac{\sigma_{AG}(D)}{\sigma_{AG}} = \frac{0}{1} = 0 \qquad \frac{\sigma_{CE}(D)}{\sigma_{CE}} = \frac{1}{1} = 1 \qquad \frac{\sigma_{FG}(D)}{\sigma_{FG}} = \frac{0}{1} = 0$$

$$\frac{\sigma_{AH}(D)}{\sigma_{AH}} = \frac{0}{1} = 0 \qquad \frac{\sigma_{CF}(D)}{\sigma_{CF}} = \frac{0}{1} = 0 \qquad \frac{\sigma_{FH}(D)}{\sigma_{FH}} = \frac{0}{1} = 0$$

$$\frac{\sigma_{BC}(D)}{\sigma_{BC}} = \frac{0}{1} = 0 \qquad \frac{\sigma_{CG}(D)}{\sigma_{CG}} = \frac{0}{1} = 0 \qquad \frac{\sigma_{GH}(D)}{\sigma_{GH}} = \frac{0}{1} = 0$$

Putting the above values in formula,

$$BC(D) = 1 + 1 + 1 = 3$$

For small network, it is easy to perform these calculations, but for a network with multiple actors we need a software like GEPHI which solves these numerical values for us.

We use two different networks to analyze the Centralities and other important factors of our network.

A. Employee's network

Using the modularity tool (under Statistics), we can calculate the different number of departments which exist in the company.

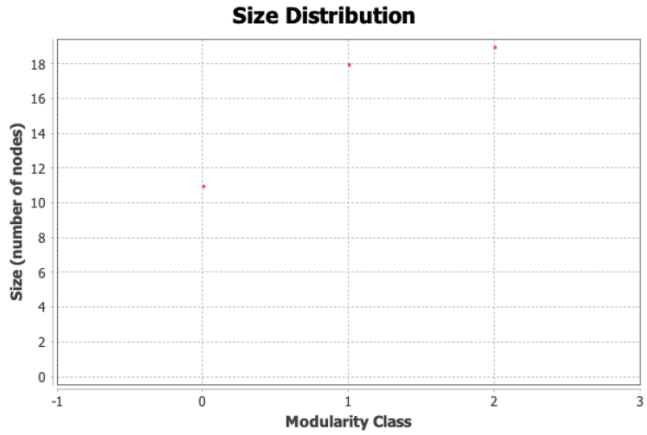
Modularity Report

Parameters:

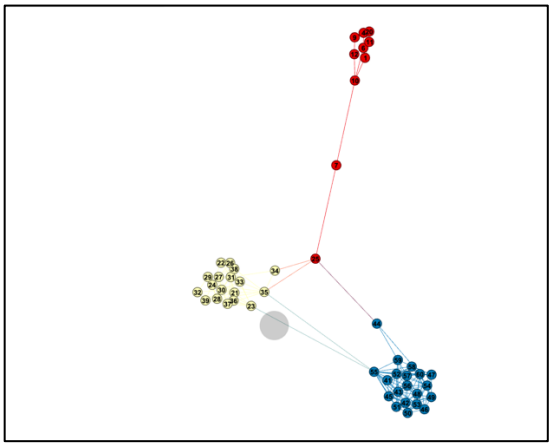
Randomize: On
Use edge weights: On
Resolution: 1.0

Results:

Modularity: 0.533
Modularity with resolution: 0.533
Number of Communities: 3

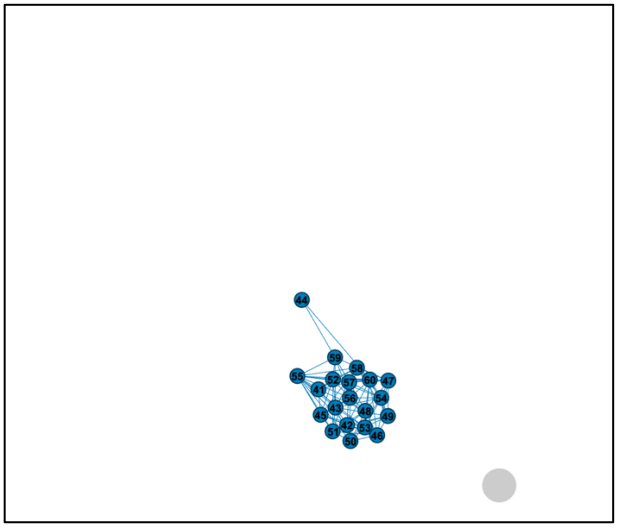


As evident from this Modularity report, we see a total of 3 departments in this network. It can also be seen from the graph image of the network, as below,



To identify an employee that is in the same department as a particular employee (say employee 46, as an example) we can look at the data table in Gephi. When we check the data table, we see ID 46 is modularity class 2. We can also now filter for modularity class 2 and can look at the employees in the same department, as depicted by the images below.

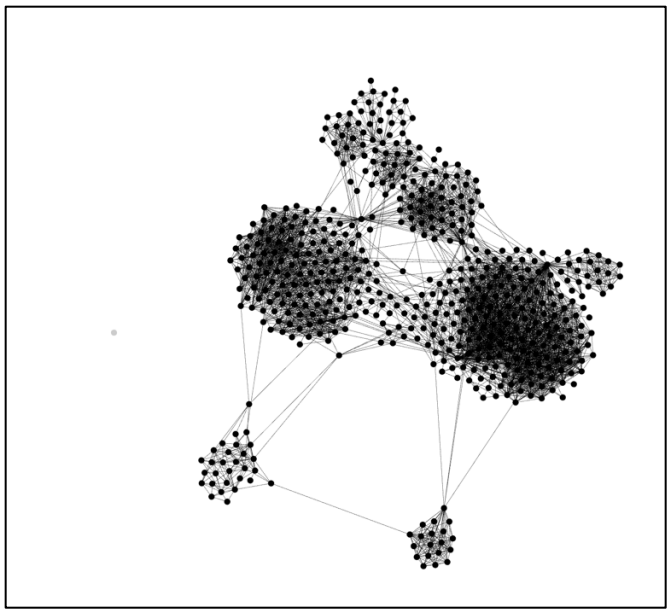
Id	Label	Interval	Modularity Class
41	41		2
42	42		2
43	43		2
44	44		2
45	45		2
46	46		2
47	47		2
48	48		2
49	49		2
50	50		2
51	51		2
52	52		2
53	53		2
54	54		2
55	55		2
56	56		2
57	57		2
58	58		2
59	59		2
60	60		2



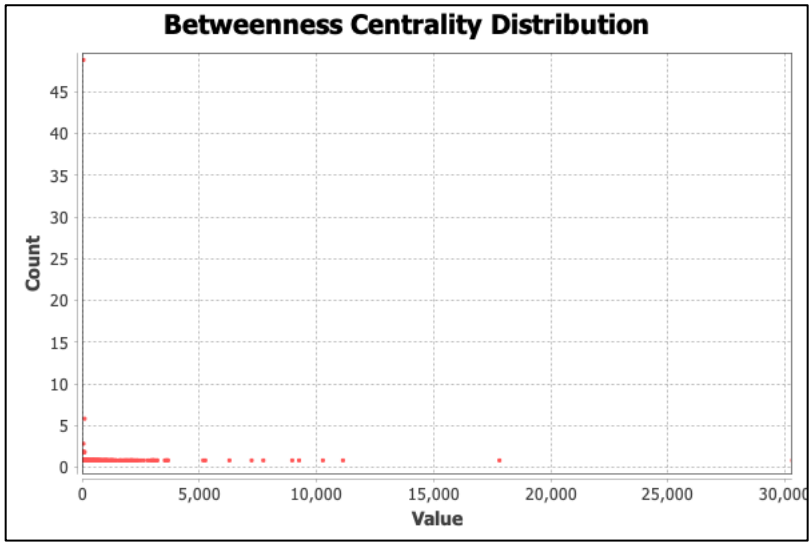
B. FB Network

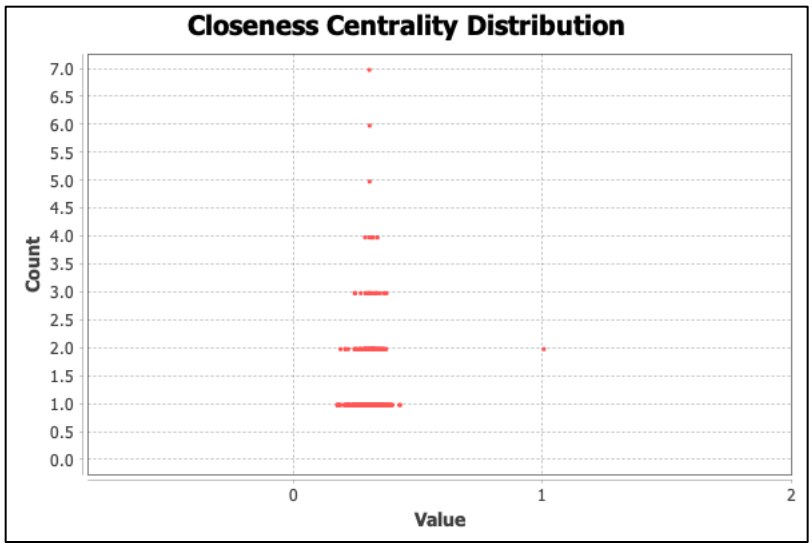
In this network, We see 534 nodes and 4813 edges. This suggests, we are working on a database of 534 Facebook friends in this graph.

We can apply various layouts, here we are using “Force Atlas 2” and then “Noverlap” to generate the following network graph.

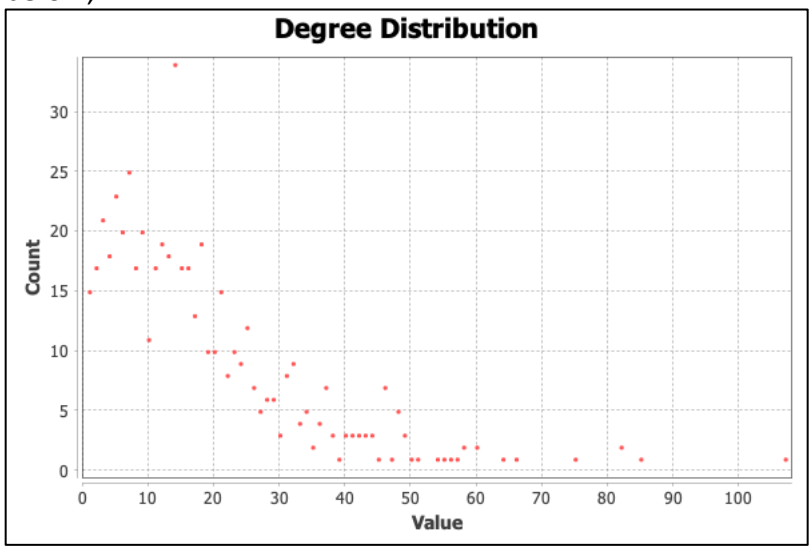


Using the network diameter function we can see the Betweenness Centrality Distribution and Closeness Centrality Distribution for this graph.

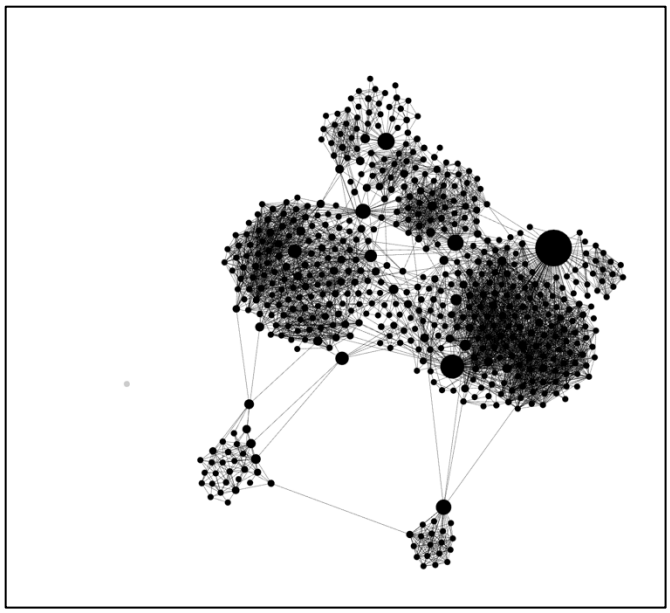




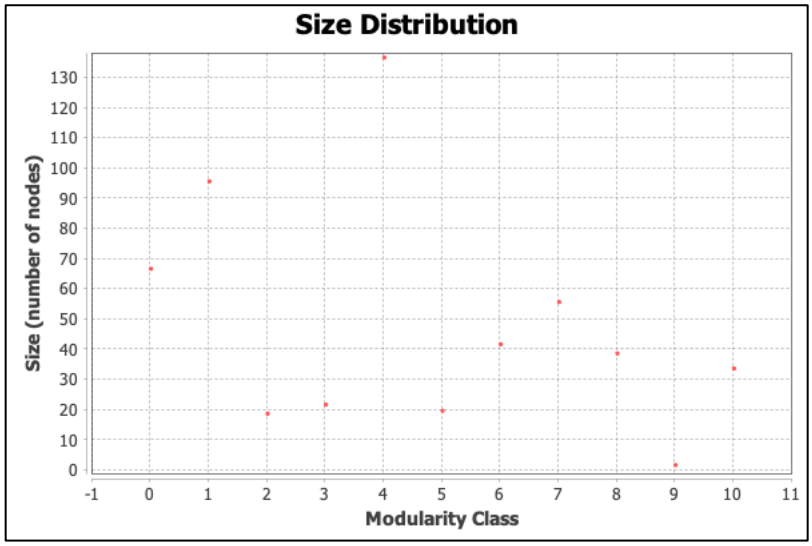
The average degree for this distribution is 18.026. The distribution can be seen below,



We can also rescale the nodes based on their centralities. As an example, see below the graph made by rescaling the nodes by their betweenness centrality.



Similar to above, to determine the different clusters in our network we can look at the modularity of the graph. We see 11 different communities in our graph.



Modularity: 0.682
Modularity with resolution: 0.682
Number of Communities: 11

To show it on the graph, we have colored the graph based on their community.