**NETWORK ANALYSIS ON FACEBOOK DATA**

**Objective**

In this project we are going to analyse the connections a person have through facebook. Certain measurements and functions are used to understand the connections and visualize it.

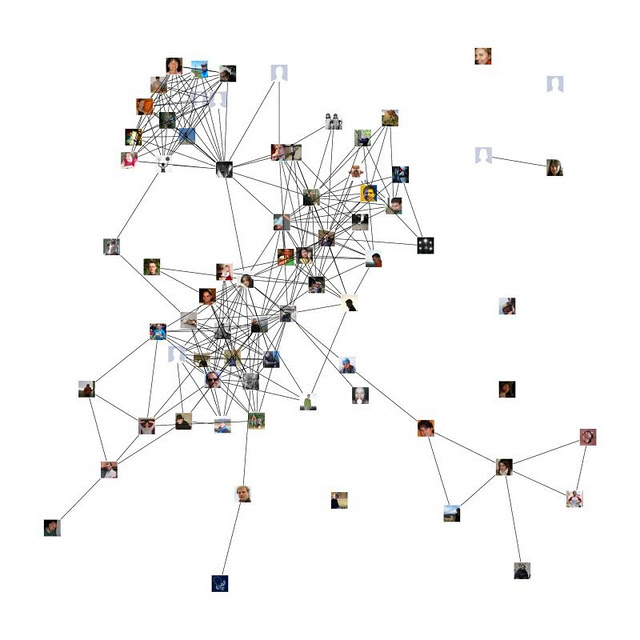
**The most important measurements include:**

* Betweenness Centrality
* Degree Centrality
* Closeness Centrality
* EigenVector Centrality

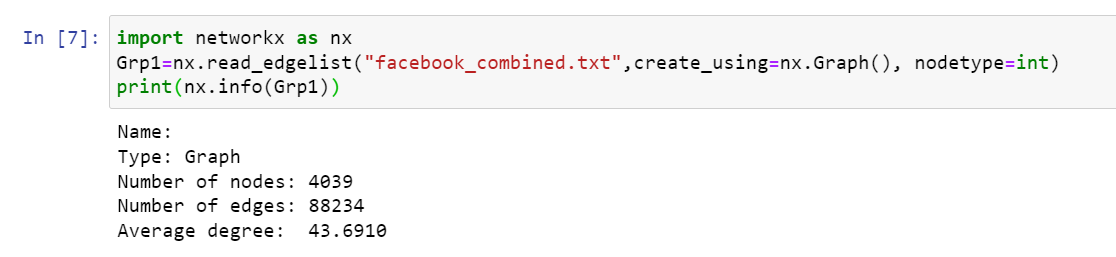
**Dataset Information**

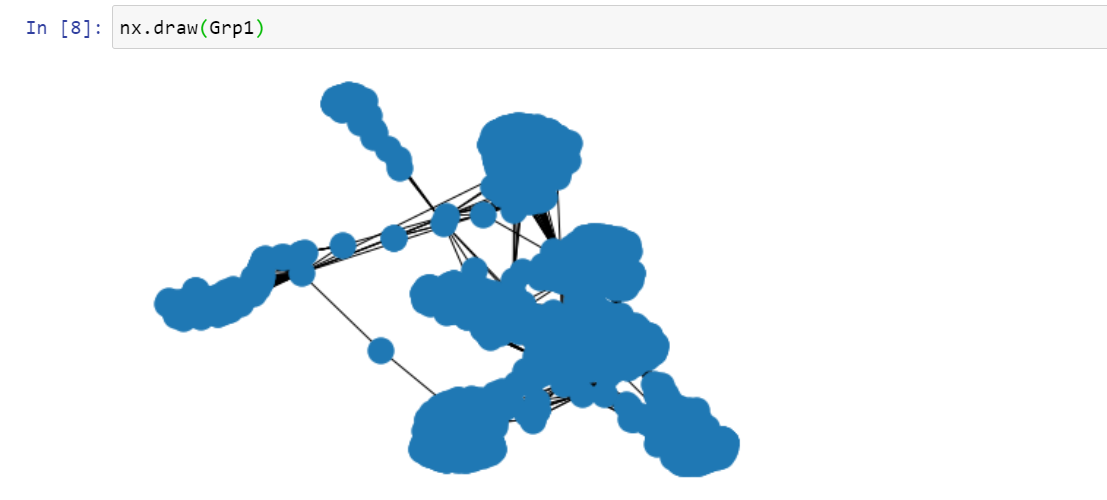
This dataset consists of ‘circles’ (or “friends lists”) from Facebook. Facebook data was collected from survey participants using this Facebook app. The dataset includes node features (profiles), circles, and ego networks.

Facebook data has been anonymized by replacing the Facebook-internal ids for each user with a new value. Also, while feature vectors from this daaset have been provided, the interpretation of those features has been obscured. For instance, where the original dataset may have contained a feature “political=Democratic Party”, the new data would simply contain “political=anonymized feature 1”. Tis, using the anonymized data it is possible to determine whether two users have the same political affiliation, but not what their individual political affiliations represent.



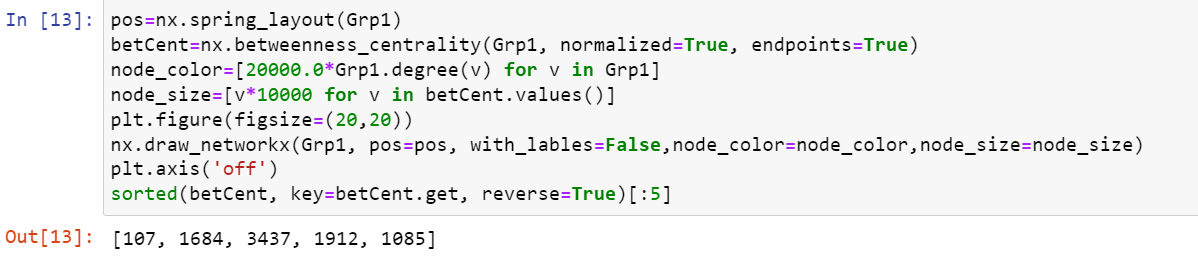
**Network Connection**

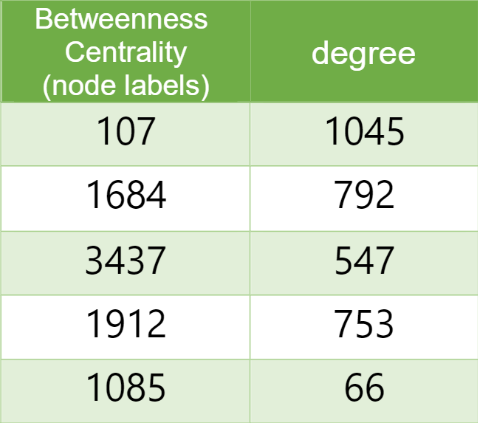


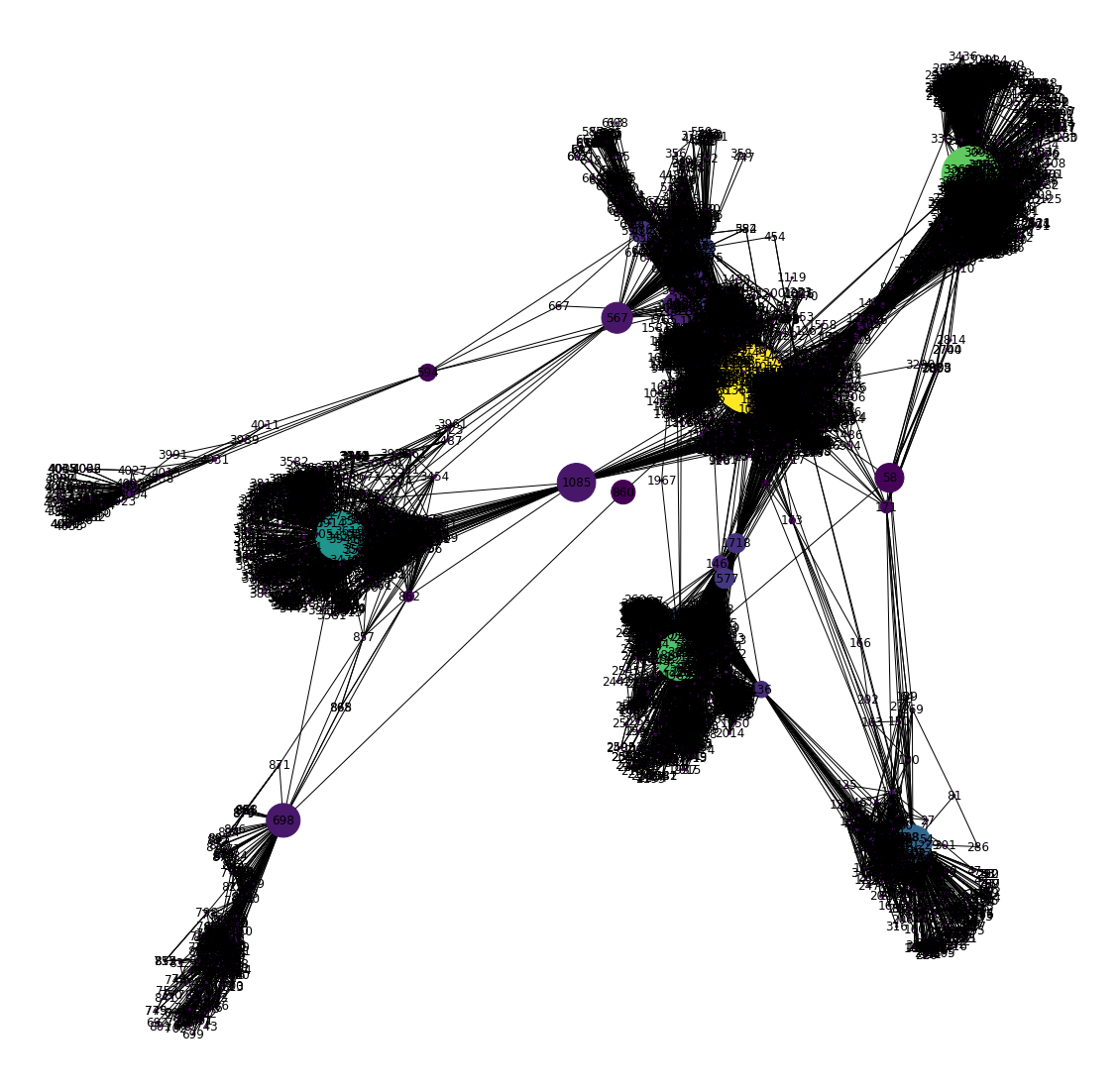


**Betweenness Centrality**

The Betweenness Centrality is the centrality of control. It represents the frequency at which a point occurs on the geodesic (shortest paths) that connected pair of points. It quantifies how many times a particular node comes in the shortest chosen path between two other nodes. The nodes with high betweenness centrality play a significant role in the communication/information flow within the network. The nodes with high betweenness centrality can have a strategic control and influence on others.

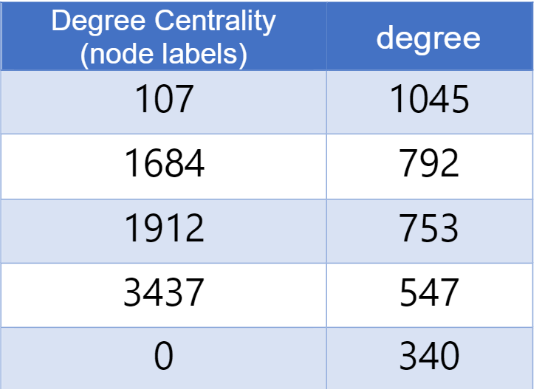


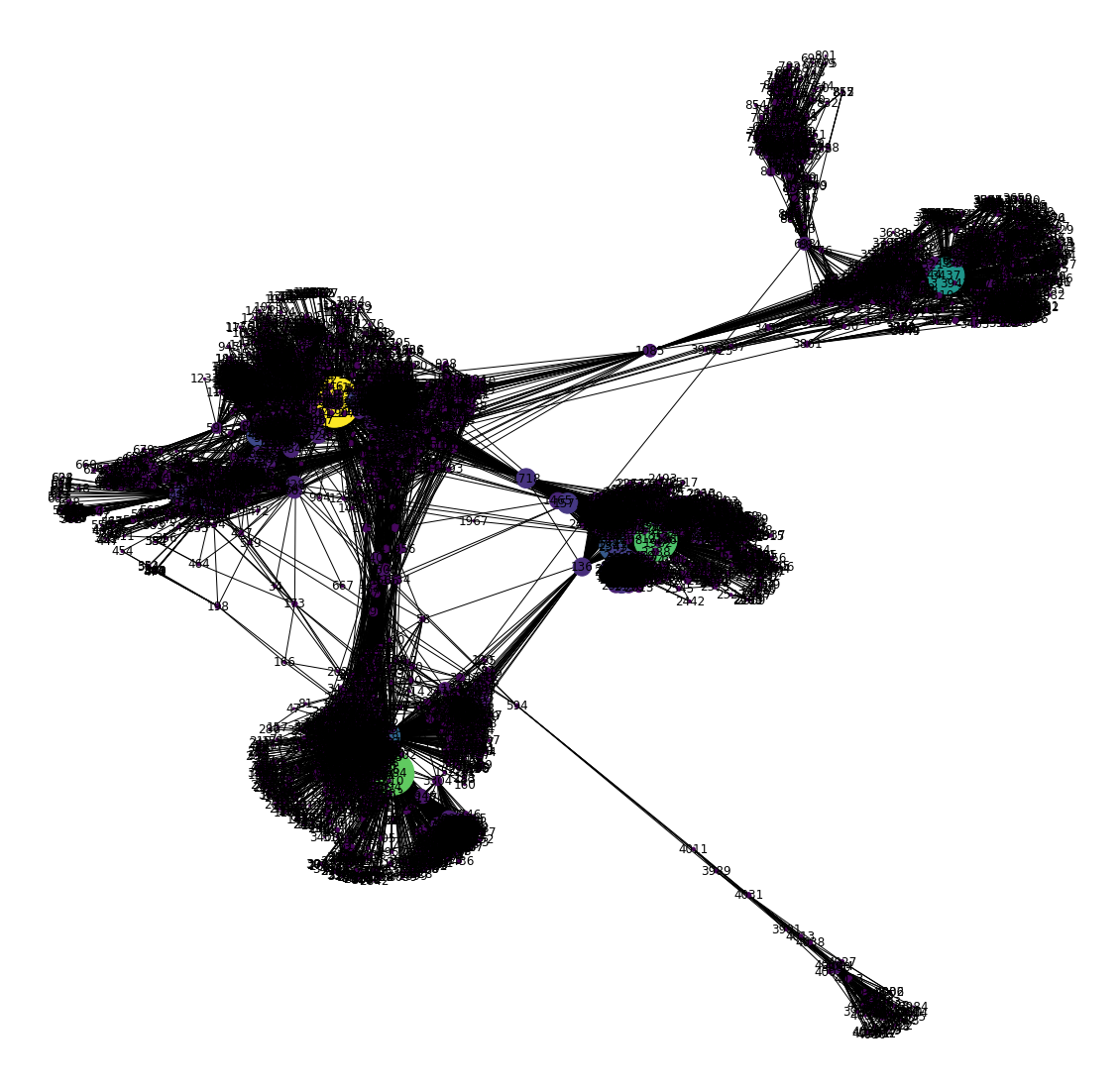




**Degree Centrality**

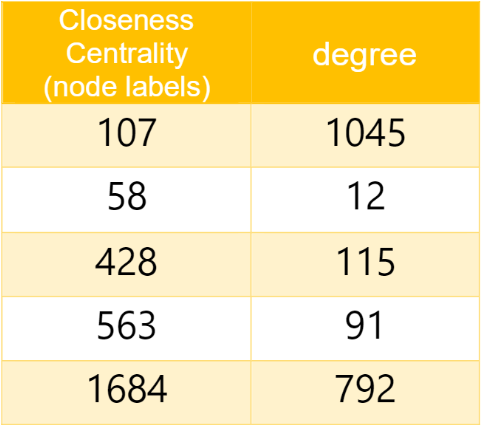
The people most popular or more liked usually are the ones who have more friends. Degree centrality is a measure of the number of connections a particular node has in the network. It is based on the fact that important nodes have many connections.





**Closeness Centrality**

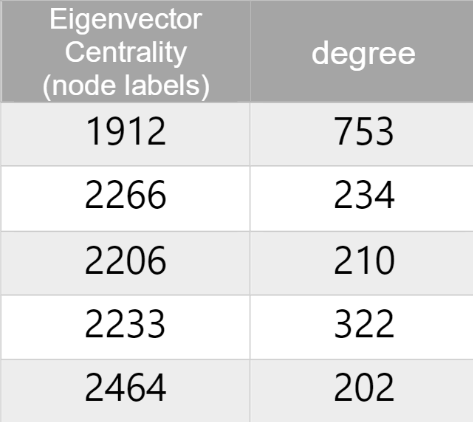
Closeness centrality we again make use of the shortest paths between nodes. We measure the distance between two nodes as the length of the shortest path between them.indicates how close a node is to all other nodes in the network. It is calculated as the average of the shortest path length from the node to every other node in the network.





### **EigenVector Centrality**

Eigenvector centrality is a measure of the influence of a node in a network. Eigenvector centrality is a more sophisticated view of centrality: A person with few connections could have a very high eigenvector centrality if those few connections are highly valued. Eigenvector centrality is a measure of exactly this. It decides that a node is important if it is connected to other important nodes. We can use the eigenvector\_centrality() function of NetworkX to calculate eigenvector centrality of all the nodes in a network.





**Conclusion**

* From the above analysis we found the nodes having betweenness, degree, closeness and EigenVector Centrality.
* The highest degree node ( Node having maximum connection) in the data is node:4038
* Therefore, Graphs can be used to figure out the most influentialpeople in a Social Network. Advertisers and Marketers can estimate the biggest bang for the marketing buck by routing their message through the most influential people in a Social Network