# Proposal for

# Social Network Analysis (SNA) with Python

Name Here

Contents

[Proposal for 1](#_Toc100968471)

[Social Network Analysis (SNA) with Python 1](#_Toc100968472)

[Introduction: 2](#_Toc100968473)

[Objective: 2](#_Toc100968474)

[Significance: 2](#_Toc100968475)

[Challenges: 2](#_Toc100968476)

[Methodology: 3](#_Toc100968477)

[Analysis of Results/Findings: 4](#_Toc100968478)

[Importing Libraries: 4](#_Toc100968479)

[Social Network Basics: 5](#_Toc100968480)

[Symmetric Networks (undirected): 5](#_Toc100968481)

[Asymmetric Networks (coordinated): 6](#_Toc100968482)

[Weighted Networks: 6](#_Toc100968483)

[Clustering coefficient: 7](#_Toc100968484)

[Degree: 8](#_Toc100968485)

[Distance: 8](#_Toc100968486)

[Breadth-first hunt: 8](#_Toc100968487)

[Eccentricity: 9](#_Toc100968488)

[Facebook Case Study: 9](#_Toc100968489)

[Conclusion: 12](#_Toc100968490)

[References: 12](#_Toc100968491)

## Introduction:

Social Network Analysis is the process where we try to investigate the social structures by using the networks and graph theories. In the social network, there are Nodes and Edges, based on this we make a network and build a graph through graph theory. The Node is the person whose network we are building and the Edge is the connection with the Node. In simple words, the edge represents the relationship between the Node and the network. These social media networks (SNA) are to analyze the patterns of relationships between the people in a specific group.

# Objective:

The main and basic objective and goal of this project are understanding the Social Media Networks (SNAs), how these networks work, and how they can help us know the connection between people in a group. We will be doing all the analysis in Python by using the Networkx module, NetworkX is a Python language software package for the creation, manipulation, and study of the structure, dynamics, and function of complex networks. It is used to study large complex networks represented in form of graphs with nodes and edges. Using networkx we can load and store complex networks.

At the start of our project, we will understand the basic idea of social networking and at the end of our project, we will perform a case study on Facebook data and plot the network graph using Python’s NetwrokX module.

In the project, we will go through the different types of social networks, we will be learning about the network coefficients, degrees, differences, etc.

# Significance:

Social Network Analysis is a reasonable technique that can dependably screen the communications in a web-based PBL climate. Utilizing SNA could uncover significant data about the course, the gathering, and individual understudies. The bits of knowledge created by SNA might be valuable with regards to learning examination to assist with observing understudies' movement. This way we can see the whole network of people and their connections.

# Challenges:

Social Network Analysis is a reasonable technique that can dependably screen the communications in a web-based PBL climate. Utilizing SNA could uncover significant data about the course, the gathering, and individual understudies. The bits of knowledge created by SNA might be valuable with regards to learning examination to assist with observing understudies' movement. The data collection is the main challenge, some social media platforms don’t allow anyone to easily access and get the data of their users and their circle of connection, in such a situation we have to perform advanced data mining techniques to scrape the data.

# Methodology:

* What were the possible approaches to solve the problem?

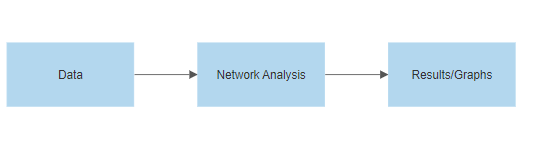
The goal of this project is to analyze the social network data often called Social Network Analysis, the purpose of doing so is to see the deep-down network we make on social media, as a user may have hundreds of circle connections with other people. Similarly, that connection will have its connections and so on. To investigate the structure of social network analysis, we used a Python module called NetworkX. NetworkX is a Python language programming bundle for the creation, control, and investigation of the design, elements, and capacity of intricate networks. It is utilized to concentrate on enormous complex organizations addressed in the type of charts with hubs and edges. Utilizing NetworkX we can load and store complex organizations.

* Why did you choose the specific one?

The reason we choose this project is that Social Network Analysis is because Social Network Analysis (SNA) can give an understanding of social impacts inside groups, and distinguish social issues. The exploration directed for SNA is keen on people, yet the actual examination centers around availability: how people work together. Social Network Analysis is a down-to-earth strategy that can dependably screen the collaborations in an internet-based PBL climate. Utilizing SNA can uncover significant data about the course, for example, the overall action and the dynamic groups. Individuals work together.

* Briefly describe your approach with technical details? E.g., Flowchart of the process, etc.

We started by understanding the basics of social network analysis. These networks are of two types. Symmetric and Asymmetric networks, to any further, we have to understand these two types. We tried to understand these types in detail. There are also weighted networks etc. The approach to the clustering of networks, degree, distance, BFS, etc. in detail.



Flow chart of the whole process

* Describe your deliverables? E.g., data description, application description with screenshots, etc.

We downloaded the data of Facebook Network from ‘stanford.edu’, the data is authentic and contain the network data of users of Facebook. This dataset comprises 'circles' (or 'companions records') from Facebook. Facebook data was gathered from overview members utilizing this Facebook application. The dataset incorporates hub highlights (profiles), circles, and self-image networks.

Facebook data has been anonymized by trading the Facebook-inward ids for every client with another worth. Additionally, while highlight vectors from this dataset have been given, the understanding of those elements has been darkened. For example, where the first dataset may have contained a component "political=Democratic Party", the new information would just hold back "political=anonymized highlight 1". In this way, utilizing the anonymized information it is feasible to decide if two clients have similar political affiliations, yet not what their individual political affiliations address.



The shape of the data

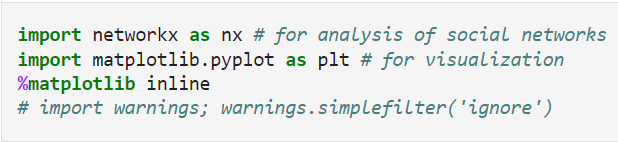
# Analysis of Results/Findings:

We utilize the module NetworkX in this instructional exercise. It is a Python bundle for the creation, control, and investigation of the design, elements, and elements of mind-boggling networks. If you work with Anaconda, you can introduce the bundle as follows:

* conda install -c anaconda network

After installing the package, we will move further by importing the libraries.

## Importing Libraries:



After importing all the required libraries, let’s understand the Networks types.

### Social Network Basics:

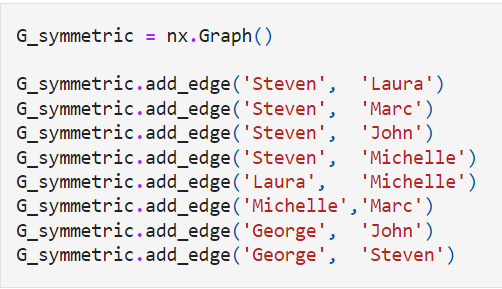
Each network comprises of:

Hubs: The people whose network we are building.

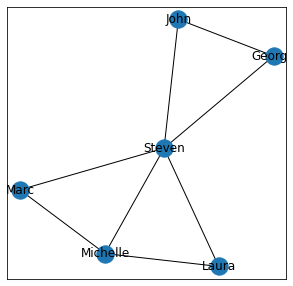
Edges: The association between the hubs. It addresses a connection between the hubs of the organization.

### Symmetric Networks (undirected):

The primary organization that we make is a gathering who cooperates. This is known as an asymmetric organization because the relationship "cooperating" is symmetric: If An is connected with B, B is additionally connected with A.



We are making a manual Symmetric Network, we are adding edges to the network, the following plot explains everything.



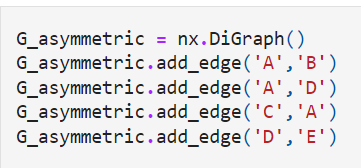
The number of nodes is 6 and the edges are 8, we can see from the graph the relationship of network users which is undirected, Jhon worked is connected with Georg and Steven, and Steven is connected to three other users, and so on.

### Asymmetric Networks (coordinated):

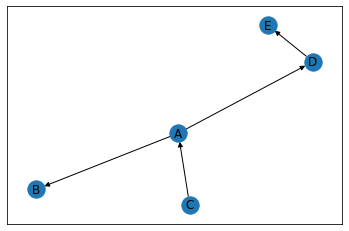
Considering the possibility that the connection between hubs is 'offspring of', then the relationship is as of now not symmetric. This is the situation assuming that somebody follows another person on Twitter. Or on the other hand on account of hyperlinks.

If an is the offspring of B, B isn't an offspring of A. Such an organization where the relationship is unbalanced (An is connected with B, doesn't imply that B is related with A) is called an Asymmetric organization.

We can assemble the awry organization in NetworkX utilizing the DiGraph strategy, which is shy of Directional Graph.



let's visualize the Asymmetric Network.

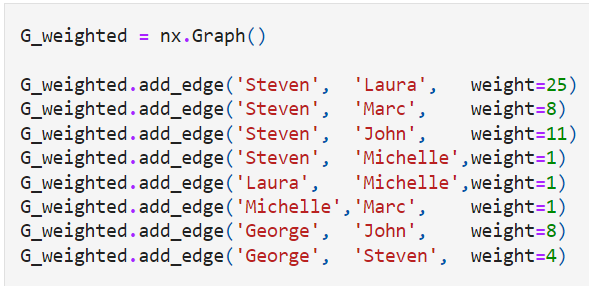


The above Network is Asymmetric Network which is a directed network.

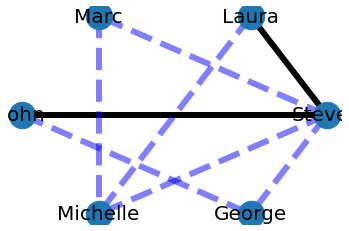
### Weighted Networks:

Till now we had networks without loads, however, potential organizations are made with loads, for instance, if in our underlying organization we consider the number of ventures done altogether, we will get a weighted Network.

Allow us to make one again of the representatives, however, this time we add weight to the organization, each edge has a weight connoting the number of ventures they have done together.



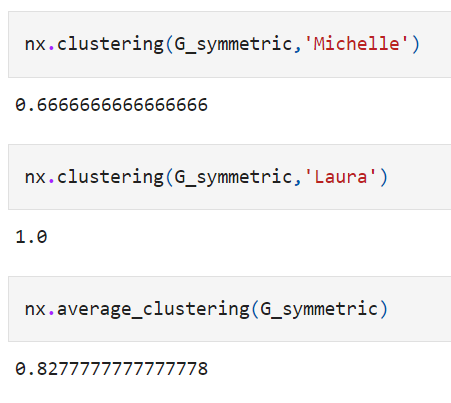
As we are now adding the weights to our network, it will look like shown below.



### Clustering coefficient:

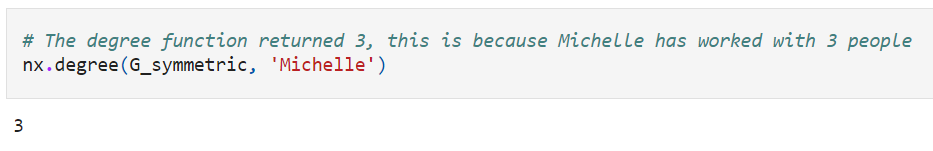
It is seen that individuals who share associations in an informal organization will more often than not structure affiliations. As such, there is an inclination in an informal organization to frame groups. We can decide the groups of a hub, neighborhood bunching coefficient, which is the negligible portion of sets of the hub's companions (that is associations) that are associated with one another. To decide the nearby bunching coefficient, we utilize nx.clustering(Graph, Node) work.

In the symmetric worker organization, you will observe that Michelle has a neighborhood bunching coefficient of 0.67 and Laura has a nearby grouping coefficient of 1. The normal bunching coefficient (amount of all the neighborhood grouping coefficients isolated by the number of hubs) for the symmetric representative organization is 0.867.



### Degree:

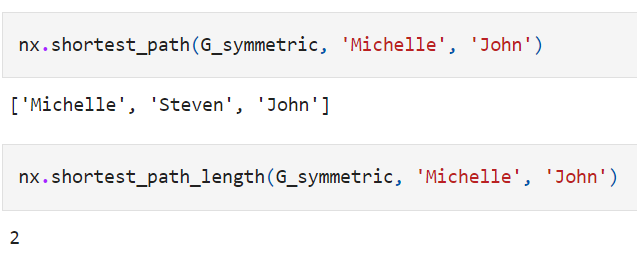
The degree of a node defines the number of connections a node has. NetworkX has the function degree which we can use to determine the degree of a node in the network.



This will return a value of 3, as Michelle has worked with three employees in the network.

### Distance:

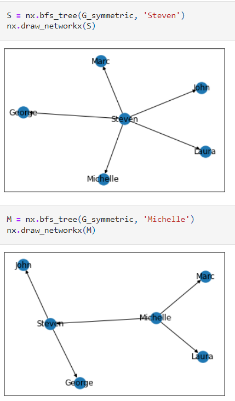
We can likewise decide the briefest way between two hubs and their length in NetworkX utilizing nx.shortest\_path(Graph, Node1, Node2) and nx.shortest\_path\_length(Graph, Node1, Node2) works individually.



### Breadth-first hunt:

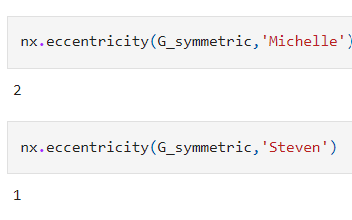
We can observe the distance of a hub from every hub in the organization utilizing expansiveness first inquiry calculation, beginning from that hub. networkX gives the capacity bfs\_tree to get it done.

Thus on the off chance that you utilize M = nx.bfs\_tree(G\_symmetric, 'Michelle') and presently draw this tree, we will get an organization structure telling how we can arrive at different hubs of the organization beginning from Michelle .de2) works individually.



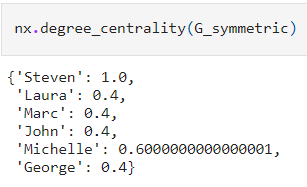
### Eccentricity:

The unusualness of a hub is characterized as the biggest distance among An and any remaining hubs. It tends to be tracked down utilizing nx. eccentricity() work. In the symmetric worker organization, Michelle has a whimsy of 2, and Steven has a flightiness of 1 (he is associated with every other hub).



1.4.1 Degree Centrality

Individuals most well-known or more preferred generally are the ones who have more companions. Degree centrality is a proportion of the number of associations a specific hub has in the organization. It depends on the way that significant hubs have numerous associations. NetworkX has the capacity degree\_centrality() to work out the degree centrality of the relative multitude of hubs of an organization.

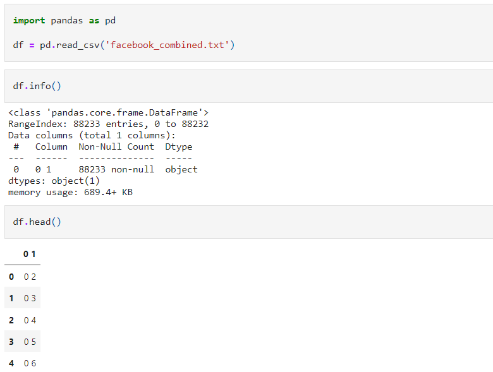


# Facebook Case Study:

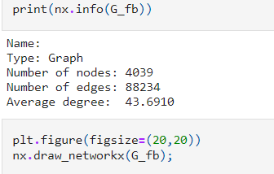
This dataset comprises 'circles' (or 'companions records') from Facebook. Facebook information was gathered from overview members utilizing this Facebook application. The dataset incorporates hub highlights (profiles), circles, and self-image organizations.

Facebook data has been anonymized by trading the Facebook-inward ids for every client with another worth. Additionally, while highlight vectors from this dataset have been given, the translation of those elements has been clouded. For example, where the first dataset may have contained a component "political=Democratic Party", the new information would just hold back "political=anonymized include 1". In this way, utilizing the anonymized information it is feasible to decide if two clients have similar political affiliations, but not what their individual political affiliations address.

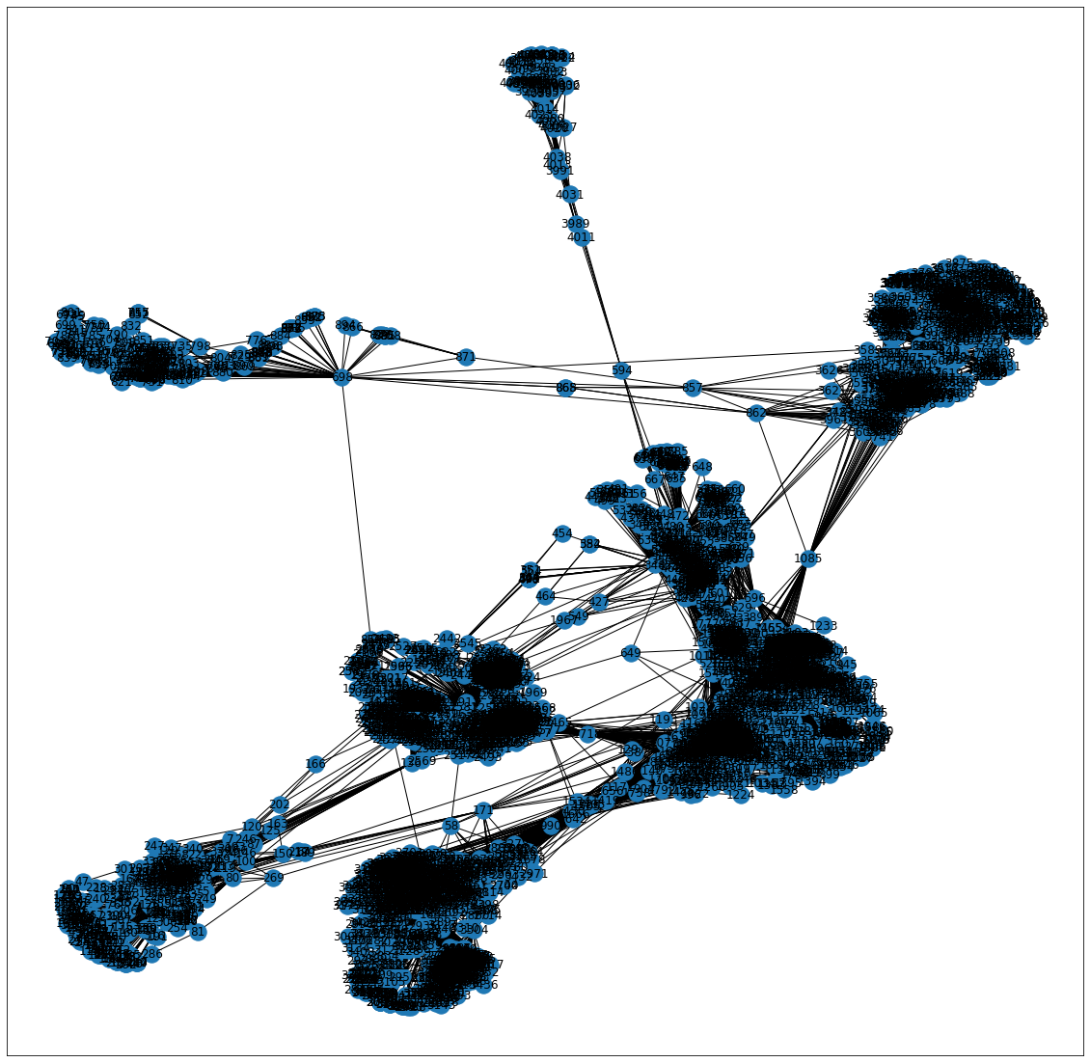
Allow us to begin with the Facebook information, for our investigation here we will utilize Facebook joined inner self organizations dataset, which contains the collective organization often people's Facebook companions list. You can download the required facebook\_combined.txt document from the Stanford University website. We read the record and develop the Graph:



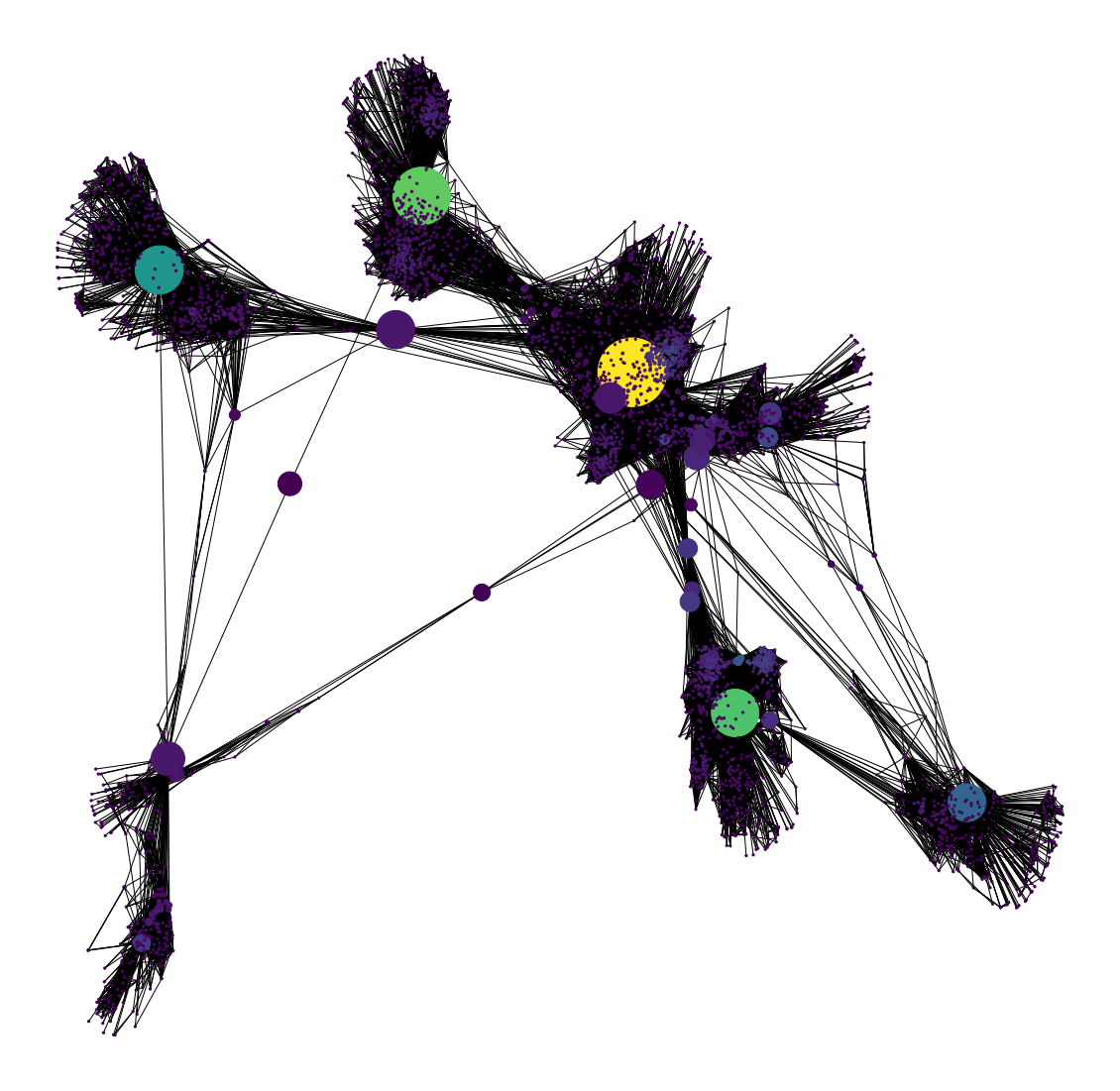
We will read the data using pandas and print the data. Now we will plot the Facebook Network data to see the whole network in the form of a graph.



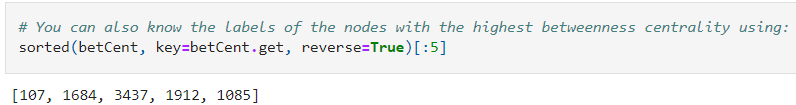
By running the above script, we will get what we want.



The above graph shows the whole network we have in our dataset. We can also visualize the network such that the node color varies with Degree and node size with Betweenness Centrality. The code to do this is:



You can also know the labels of the nodes with the highest betweenness centrality using:



We can see that a few hubs are normal between Degree Centrality, which is a proportion of degree, and Betweenness Centrality which controls the data stream. It is regular that more associated hubs likewise lie in the briefest ways between different hubs. The hub 1912 is a significant hub as it is essential as indicated by each of the three centrality estimates that we had thought of.

# Conclusion:

The characterizing component of social network analysis is its emphasis on the construction of connections, going from easygoing associates to close bonds Footnote Social organization examination expects that connections are significant. It guides and measures formal and casual connections to get what works with or obstructs the information streams that tight spot collaborating units, viz., who can say for sure whom, and who imparts what data and information to whom by what correspondence media (e.g., information and data, voice, or video communications). In this case study, we tried to understand the basics of social media networks by analyzing different types of networks like Asymmetric and Symmetric networks, etc. In the end, we have drawn the SNA of Facebook by plotting some graphs, these graphs help us understand the network of each user with other people and so on.

# References:

Dataset: <https://snap.stanford.edu/data/egonets-Facebook.html>

Paper: <https://link.springer.com/chapter/10.1007/978-981-10-0983-9_9>