

***Facebook Influencer and User
Connection Analysis using SNA***

Introduction

Social networks are the networks that depict the relations between people in the form of a graph for different kinds of analysis. The graph to store the relationships of people is known as Sociogram. All the graph points and lines are stored in the matrix data structure called Sociomatrix.

Social network analysis, a prominent way towards understanding the human behavior by mapping and measuring the relationships and interactions between people, groups, organizations, or any information/knowledge processing entities. Following the definition by the authors of “*Social Network Analysis: Methods and Application*” research, Wasserman and Faust, social network data can be viewed as a social relational system characterized by a set of actors and their social ties. Additional information in the form of actor attribute variables or multiple relations can be part of the social relational system.

In a social network each people, group or entity are represented as a node or vertex and the relationships and interactions create links or edges between the nodes. This creates the social structure. The substructure in terms of grouping or cliques is one of the most interesting things of the social structure. The behavior of the network can be estimated from the number, size, and connections among the sub-groupings in a network.

To create visual of social network analysis, actors and relationships are the two key components. A very common and simple example for the use of social network analysis graph is webpages. On the internet, webpages can contain links to other webpages -either on their own website or another website. These links can be considered as the relationships between the actors, which are the webpages in this case. This is the backbone of search engine architecture.

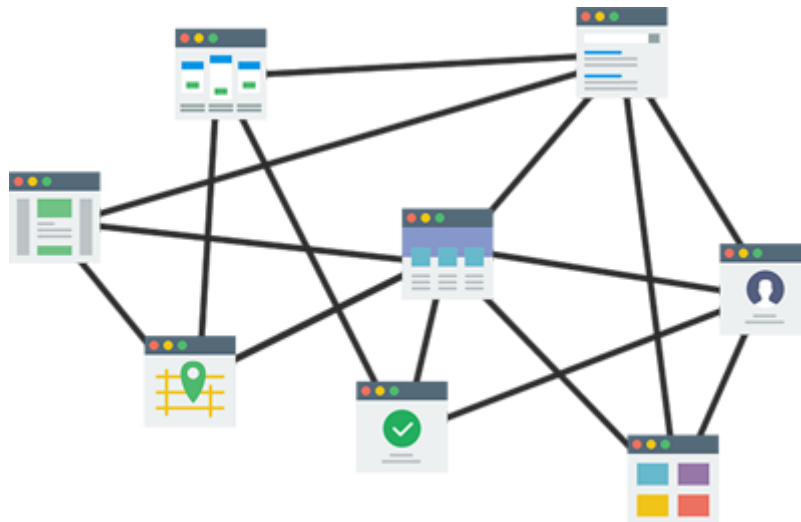


Figure 01: Relationship between webpages

The visual graph of a social network is similar to the mind mapping. The actors or edges are represented with points or dots or circles and the links are represented with lines or arrowheads. An exaple of social media graph is as following:

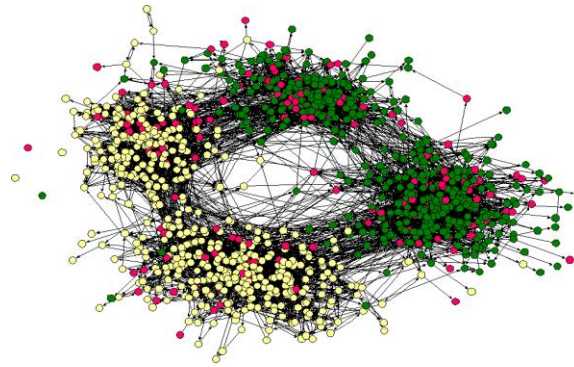


Figure 02: Social network analysis graph (Friendship network)

Background

The growth of social network analysis happened from three traditions[mcglain2010]. First is cognitive and psychologists of the 1930s, researched the groups and information flow structure among the members. Second is anthropologists of Harvard who studied and refined the premises of anthropologist A.R. Radcliff-Brown by focusing on interpersonal relations and sub-groups within social network. Thirdly, Manchester University researchers examined tribal societies and used the studies to further refine social theory and the study of relationship between communities.

Influencers on social media are very vital in terms of user engagement within the network and other user of the network from a person to a brand or company. With the help of their “influence” over others in the network various marketing promotions of products and brands are boosted to the target audience. For example, the beauty and health industry are one such market where influencers use their platform to promote contracted brands as a part of digital marketing of that brand or product.

Thus, the analysis of social media to extract influencers who have high density of connection and degree of influence over their connection can prove to be extremely vital for a company or organization. The importance of each actor within the network in the social media or the target network is the influence of that actor, and which can be analyzed using Social Network Analysis. To find such influential actors within a Facebook dataset, we used mathematical analysis such as Centrality using Social Network Analysis (SNA).

Methodology and Result analysis

Social Network Analysis is the process of evaluating the relationships shared by individual entities in a network. For this mini project we have used Facebook dataset “facebook_edgeds.csv”. Our main goal is to find the connection between users from this dataset and finding the influencer from all the networks. Facebook is constructed on the basis of SNA where the entities (people, organizations, and computers) are represented by nodes in the network. The links or edges represent the relationships between various nodes. The position of a node in the network determines whether they are the connectors, leaders, or bridges. SNA is applied to social networking platforms to determine the real influential power of a person in the network.

Table 1: Dataset Sample

User 1	User 2
0	18427
1	22208
1	15785
1	18304
1	16260

Degree is the measure of the total number of edges connected to a particular vertex. For directed networks, there are two measures of degree. In-degree is the number of connections that point inward at a vertex. There are four well-known centrality measures: degree, betweenness, closeness and eigenvector - each with its own strengths and weaknesses. The mean, median and mode are known as measures of centrality: an aim to identify the midpoint in a data set through statistical means.

Table 2: Summary of the Dataset

	User 1	User 2
count	85501.000000	85501.000000
mean	6851.224898	14672.127379
std	5336.571365	5465.247707
min	0.000000	14.000000
25%	2242.000000	10694.000000
50%	5778.000000	15606.000000
75%	10510.000000	19368.000000
max	22405.000000	22469.000000

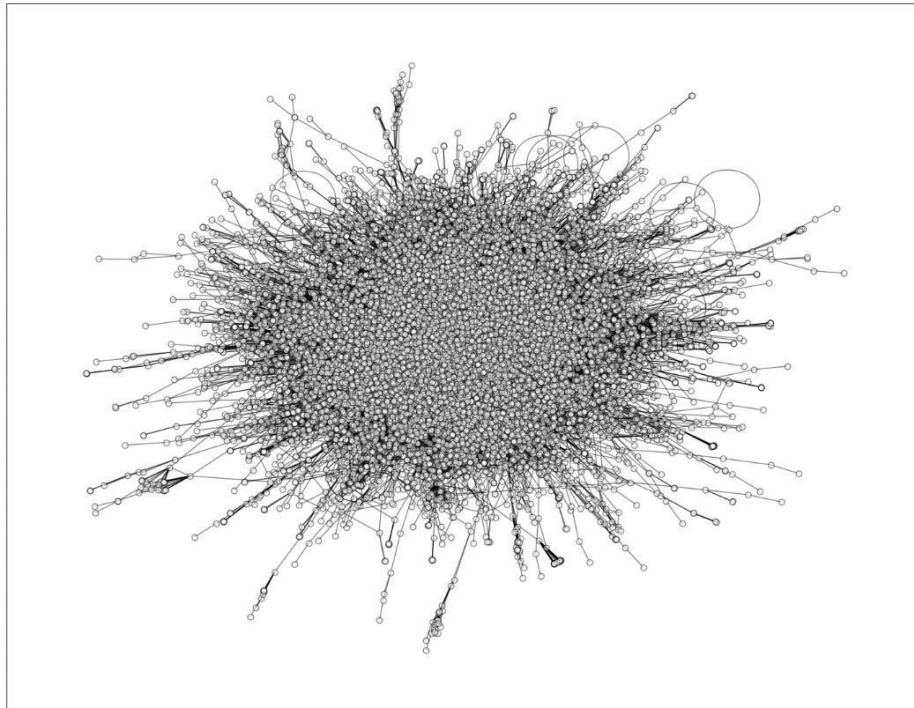


Figure 3: The network graph of the dataset

In the dataset, each row is a connection of *User 1* with *User 2*. In the Figure 3, the users are depicted as nodes (small circles) and the connection between each user is depicted as edges (lines connecting circles). Total number of nodes (unique user) is 22, 470 and total number of edges (connection) is 85, 501.

Community_Id	Community_Size	Avg_Deg_Cent	Avg_Clo_Cent	Avg_Bet_Cent	Radius	Diameter	Number_Bridges	Avg_shortest_path	
0	0	22470	0.000339	0.169954	0.000224	10	19	1	6.041997

Figure 4: Community in the dataset

Only 1 community is present in the network. The community is summarized in Figure 4. As it is the only community, the size of the community is equal to the number of nodes or users. Average Degree Centrality is 0.000339, Average Closeness Centrality is 0.169954, Average Betweenness Centrality is 0.000224. These attributes are explained as following:

Degree Centrality: The degree centrality of a node in a graph is simply a count of the number of edges that connect to it. Degree Centrality assign an importance score based simply on the number of links held by each node basically it's a simple count of the total number of connections linked to a vertex. Degree is a highly effective measure of the influence or importance of a node. The average degree centrality can be a good measurement to find the influencers in the network. Those with most connections (close or higher than average) are identified as influencers.

$$\text{Average Degree Centrality, } DC_{avg} = \frac{\text{Total Edges, } E}{\text{Total Nodes, } N}$$

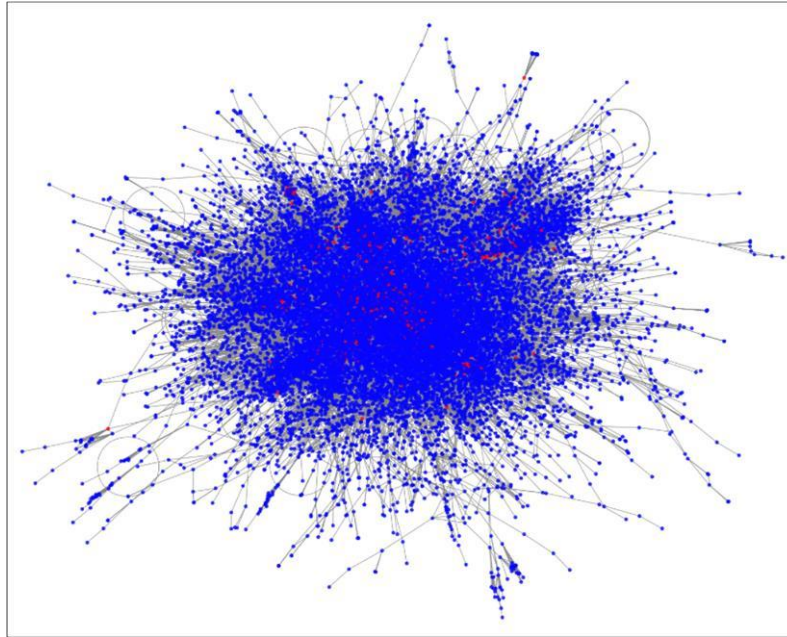


Figure 5: Average Degree Centrality of the network

In the network, the influencers found by applying degree centrality are depicted with red circles. Observing the graph indicates concentration of influencers towards the center of the graph with a few close to center and a very few at the outer edges of the network. Though the average degree centrality is a good attribute to identify influencers, it is not always best as higher connections does not always mean higher level of influence. The average degree centrality is also very low, because of which a lot of nodes are marked as influencer but very far from the center of the network.

Closeness Centrality: Closeness centrality scores each node based on their 'closeness' to all other nodes in the network. What it tells us: This measure calculates the shortest paths between all nodes, then assigns each node a score based on its sum of shortest paths. The average length of the shortest path possible from a specific node to all the other nodes in the network. The more central a node, the closer all the other nodes. This measure is sometimes used to reflect how quickly information might spread among nodes in a network.

$$\text{Closeness Centrality, } CC_i = \frac{N - 1}{\sum_j d(i, j)}$$

Where, D_{ij} = length of the shortest path between nodes i and j in the network, N is the number of nodes.

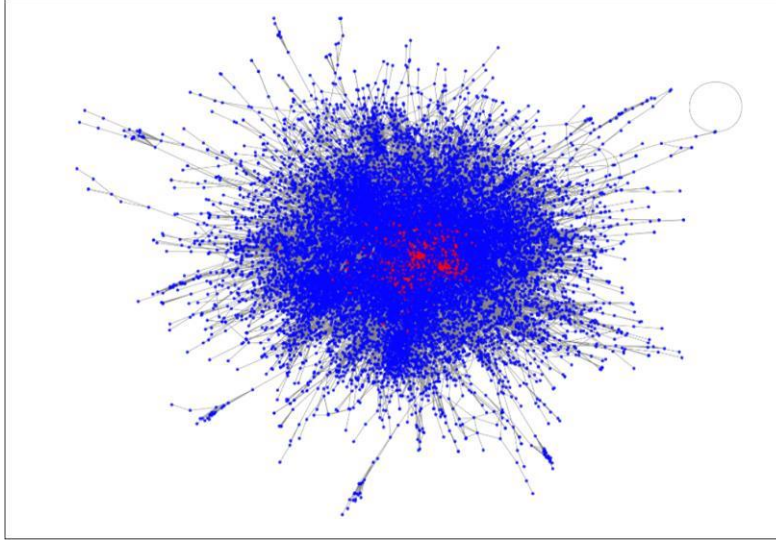


Figure 6: Closeness Centrality of the network

The maximum value for closeness centrality of a node can be 1 and the 0 is the minimum. Higher closeness centrality indicates strong bond (high rate of information flow) thus more likeliness of an influencer. The analysis of Figure 6 shows that almost all the influencers are at the very center of the network and none outside of the central region, as they should be. The average closeness centrality of the network performed very well to identify influencers than the others.

Betweenness Centrality: Betweenness centrality is a way of detecting the amount of influence a node has over the flow of information in a graph. It is often used to find nodes that serve as a bridge from one part of a graph to another. The number of times a node serves as a bridge on the shortest path between other nodes. A node that is often a bridge can control the spread of information, allowing or limiting its flow. The betweenness centrality is calculated as following:

$$\text{Betweenness Centrality, } C_B = \sum_{j < k} g_{jk}(n_i) / g_{jk}$$

Where, g_{jk} = the number of shortest paths connecting jk , and $g_{jk}(n_i)$ = the number that node i is on.

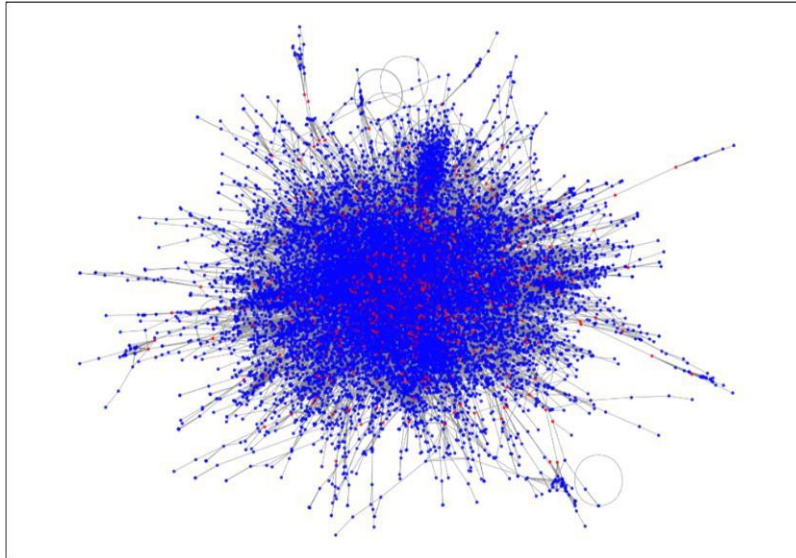


Figure 7: Betweenness Centrality of the network

The betweenness centrality is also a great measure of influence like closeness centrality. Both measures depend on the shortest paths to the node but they are different from each other as closeness centrality finds the shortest distance from all graph nodes to find the influencer in information flow and betweenness centrality quantifies the importance of the node in information flow.

Highly Influential Nodes: Among the influential nodes measured by each centrality, there are nodes which are common among all of the measures. These nodes can be identified as highly influential nodes. These nodes not only have high number of connections but also influence the information flow of the network. Thus, the density of these nodes can be seen high towards the center of the network. These nodes or users can be classified as the most influential users in the community.

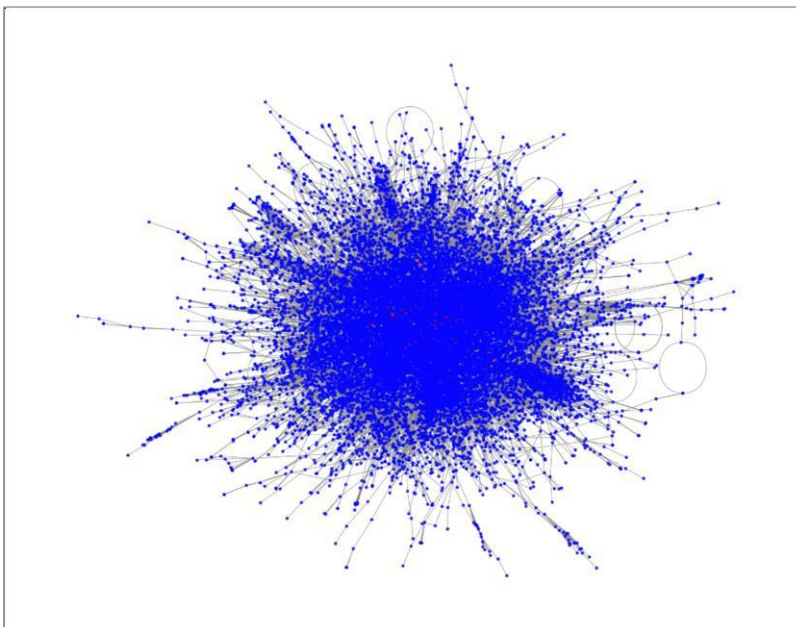


Figure 8: Highly influential nodes in the network

Bridges: A bridge is a type of social tie that connects two different groups in a social network. While bridges connect two nodes they are not necessarily influencers.

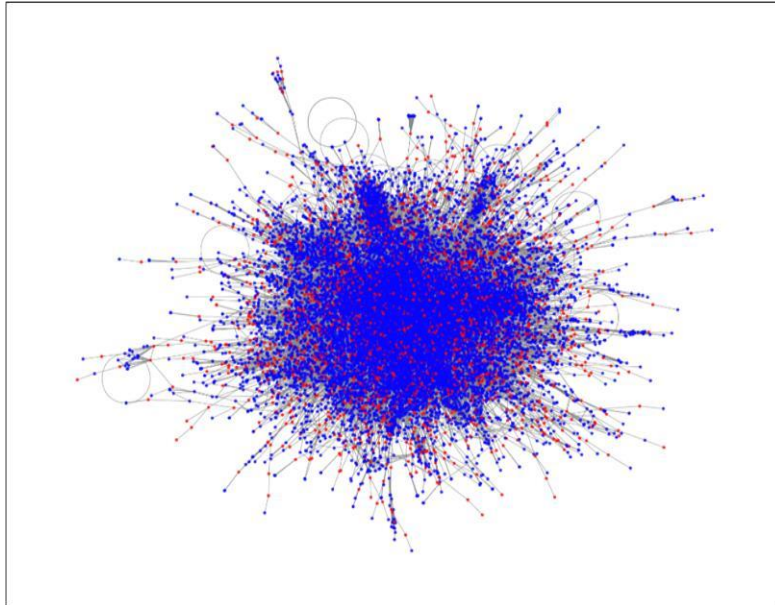


Figure 9: Bridges in the network

Bridges and influencers Among all the connection the highest node in the community which will be considered as influencer. The following network graph shows all the bridges but the influencers with the highest number of bridges. Only a very few can be seen near the center.

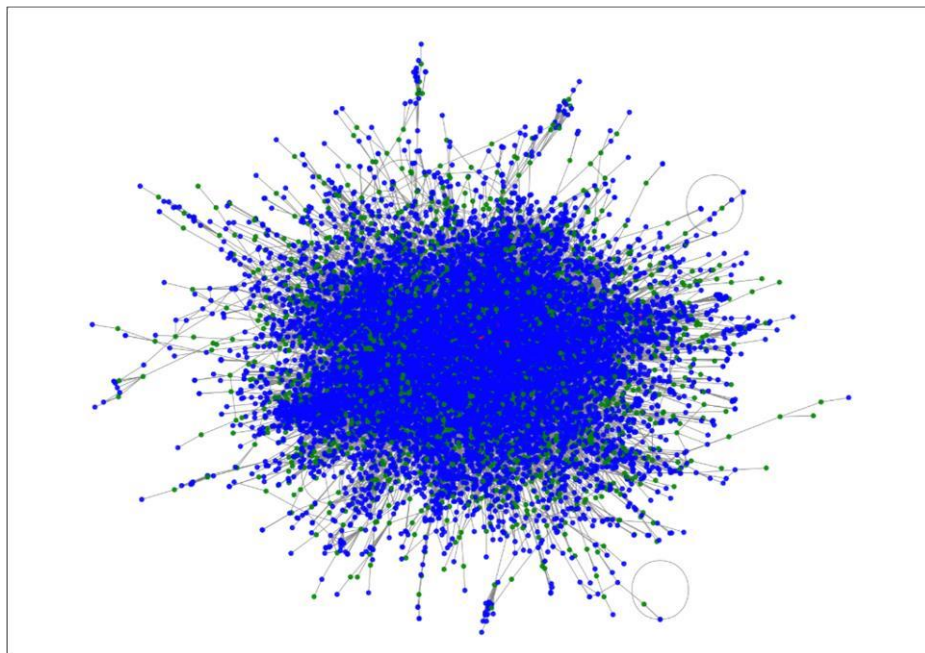


Figure 10: Bridges and influencers in the network

Importance of SNA Influence

The benefits of social networks include their ability to help people connect and stay in touch with family, friends, and new contacts; the opportunity they offer businesses to market their brands; their ability to spread useful, even vital, information instantly to individuals and institutions.

Social networks are important because they allow people to develop relationships that might not be possible due to distances of place and time. They also help boost business productivity when used for public relations, marketing, and advertising purposes. Any social media platform like Facebook, YouTube, WhatsApp, Instagram, WeChat, TikTok, Facebook Messenger, Douyin, QQ, and Sina Weibo uses SNA Influence technique.

Social networking connects individuals and businesses by allowing them to share information, ideas, and messages. Companies also use social networks to create and strengthen brand recognition, promote products and services, and answer customer queries and concerns.

Social networking involves the use of online social media platforms to connect with new and existing friends, family, colleagues, and businesses. Individuals can use social networking to announce and discuss their interests and concerns with others who may support and otherwise interact with them.

Businesses can use social networking to build a brand, sell products, grow a customer base, and strengthen customer relationships and service. Social networking may have certain disadvantages, but its benefits—including giving individuals and companies a method to spread messages instantly—make it a technology that will continue to be used and developed for years to come.

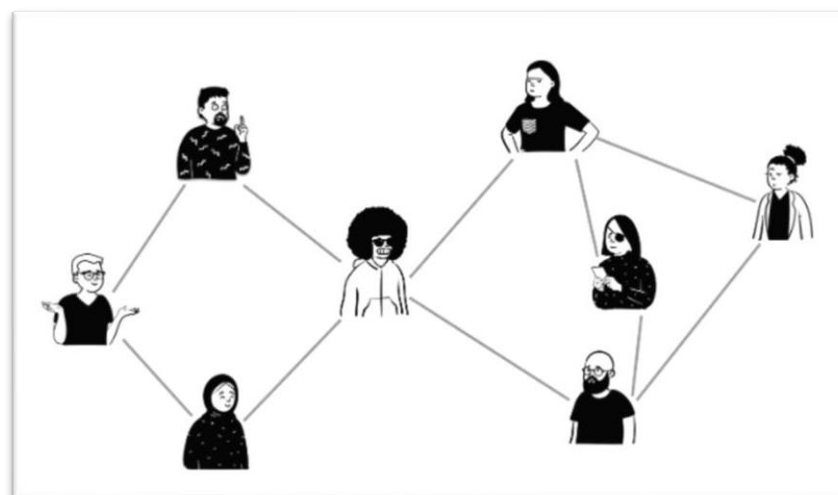


Figure 11: Visualization of influencers in the network

Future Scopes

The aim of social network analysis is to understand a community by mapping the relationships that connect them as a network, and then trying to draw out key individuals, groups within the network ('components'), and associations between the individuals. A network is simply a number of points (or 'nodes') that are connected by links. Generally, in social network analysis, the nodes are people, and the links are any social connection between them – for example, friendship, marital/family ties, or financial ties.



Figure 12: Facebook Application with Social Network Analysis

Big Data is a research field involving a large number of collaborating disciplines. Based on bibliometric data downloaded from the Web of Science, applying various SNA and visualization tools to examine the structure and patterns of interdisciplinary collaborations, most important contributors to Big Data research as well as the recently evolving overall pattern.

With the advent of digitization, the internet today acts as a rolling ball in the workplace as well as in everyday life, with people gravitating toward social media. Ever since the pandemic, internet has invaded the earth where rain is still hesitant. According to a report, half of India's population uses social media. The report further states that social network penetration is projected to hit 67% of the country's population by 2025.

Social media has become an integral part of our lives in the current digital era, assisting individuals in developing stronger bonds with others, and networking applications are now proving their worth. That is why people spend so much time online browsing social media apps. We are now surrounded by numerous social networking apps that assist us in connecting with one another. However, as technology advances, social networking apps are thinking outside the box and developing innovative functions such as connecting people in cafés for business and personal meetings. Hence, with the launch of social marketing apps the world is now gaining influence and benefiting the future in the times ahead.

Conclusion

Social Network Analysis (SNA) is the process of exploring or examining the social structure by using graph theory. It is used for measuring and analyzing the structural properties of the network. It helps to measure relationships and flows between groups, organizations, and other connected entities. We need specialized tools to study and analyze social networks. Social networks are important because they allow people to develop relationships that might not be possible due to distances of place and time. While the influencers play an active role in the flow of information and activity of the network. The

influencers activity can increase the information flow of a network as well as connection with in the network. They also help boost business productivity when used for public relations, marketing, and advertising purposes.

There are a decent number of measurements to identify the influencers in a social network, among them centrality methodologies are very popular. The centrality measurements can be said as mathematical or statistical analysis of a social network. The closeness centrality and the betweenness performed very well to find the influential nodes in the network. While the union of the degree centrality, closeness centrality and betweenness centrality helped to filter out the most influential nodes over the other nodes present in the network.