WWW.IJRPR.COM

International Journal of Research Publication and Reviews

Journal homepage: www.ijrpr.com ISSN 2582-7421

Extracting and Analysing Web Social Network Using Network

[*1] Siddhesh S, [*2]Sri Suhas P, [*3]Dr. M. Sujithra

[*1, 2] M. Sc. Data Science (Integrated), Department of Computing - Data Science, Coimbatore Institute of Technology, Coimbatore.
[*3]Assistant Professor, Department of Computing - Data Science, Coimbatore Institute of Technology, Coimbatore.

ABSTRACT

Techniques for social network analysis (SNA) have become one of the internet's most popular uses. There are several factors that call for a deeper comprehension of social network structures, support the need for their analysis, and call for researching how these networks will affect the internet of the future. Finding similar interests and trust, for instance, can be one of the motivations to investigate social networks. Furthermore, just like current peer-to-peer content distribution networks, distributed online social networks in the future could significantly affect Internet traffic if they become popular and bandwidth-intensive. Whatever one's opinion of these phenomena, a deeper comprehension of the structure of social networks will probably help us better comprehend the opportunities, constraints, and threats connected to these concepts. The study provides a detailed explanation of the necessity for social network analysis and its ramifications. It is learned that there are numerous factors to consider while examining any social network.

Keywords *Social Network, NetworkX, Clustering coefficient, Network distance measures, Centrality measures.*

1. INTRODUCTION

Network science, commonly referred to as social network analysis (SNA), is a branch of data analytics that makes use of networks and graph theory to comprehend social systems. SNA methods can be used on networks that are not part of society. Actors and relationships are the two essential elements needed to develop SNA graphs.

On the internet, pages frequently connect to other pages on other websites or on their own website. These connections could be viewed as actor relationships (web pages). In fact, this is an important part of search engine architecture. A social network is any network containing links between people that reflect their relationship to one another. By analyzing these networks, a lot about the network's users can be analyzed, including who the true influencers are and who is most connected. To create and evaluate these various networks, NetworkX has been utilized here.

2. DATASET

For this analysis the Facebook combined ego networks dataset has been utilized, which contains the networked list of ten people's Facebook friends, for this investigation. 4,039 nodes make up the network, which is connected by 88,234 edges.

Dataset Description

```
[Out]: Name:
       Type: Graph
       Number of nodes: 4039
       Number of edges: 88234
       Average degree:  43.6910
```

Sample Dataset

	0 1
88228	4026 4030
88229	4027 4031
88230	4027 4032
88231	4027 4038
88232	4031 4038

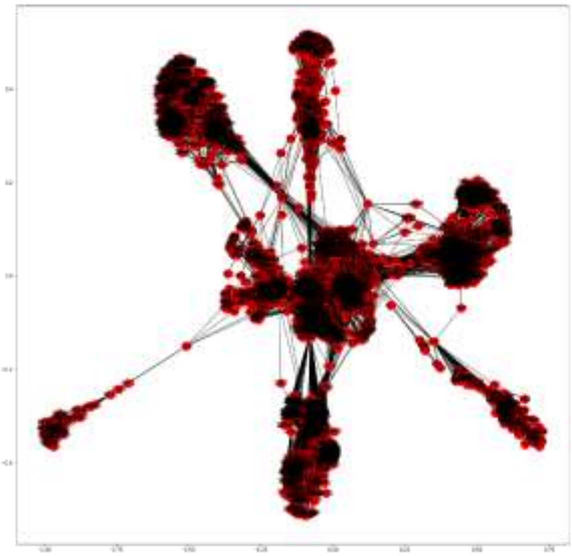

Fig 2.1 Visualization of Facebook dataset

3. FRAMEWORK USED

NetworkX is a set of Python-based tools for building, modifying, and researching the composition, dynamics, and operation of complicated networks. Large complicated networks that are represented as graphs with nodes and edges are studied using this method.

NetworkX can be used load and store complex networks. Also can able to create a variety of random and conventional networks, study their structure**,** create network models, create new network methods, and even sketch them.

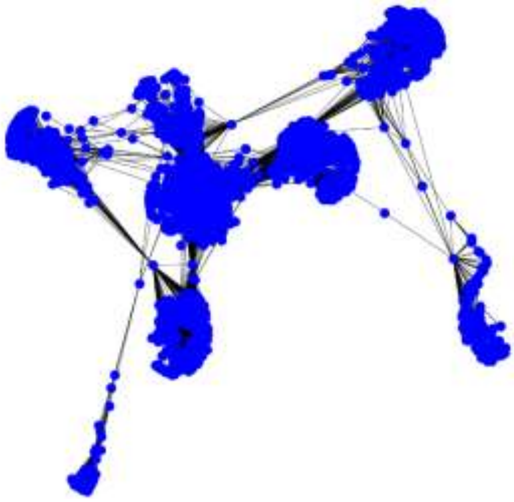

Fig 3.1 NetworkX Graph

4. IMPLEMENTATION

Utilizing networkX, social network analysis has been put into practise. It is a Python tool for building, modifying, and researching the composition, dynamics, and purposes of complex networks.

4.1 Social Network Basic

Each network consists of,

- Nodes: The individuals in the network.

- Edges: The connection between the nodes. It represents a relationship between the nodes of the network.

4.1.1 Weighted Network

Networks may contain weights; for instance, if a weight is assigned to the number of tasks completed concurrently in our initial network, and weighted network can be obtained.

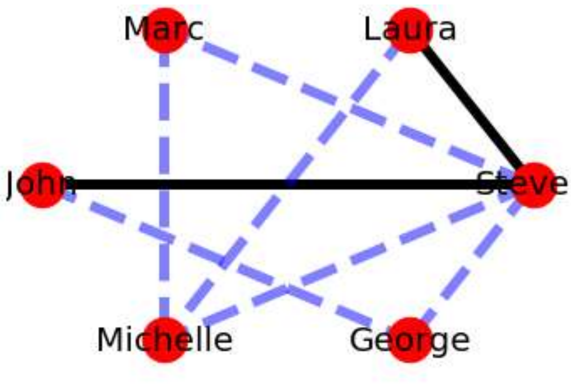

Fig 4.1 Weighted Graph

4.2 Clustering Coefficient

In a social network, it has been found that individuals who have connections in common are more likely to associate. In other words, a social network has a propensity to develop clusters. A node's clusters can be identified using the local clustering coefficient, which measures the proportion of connections between pairs of the node's friends. The average clustering coefficient for the symmetric employee-network is equal to the sum of all local clustering coefficients divided by the number of nodes.

```
nx.average_clustering(G_symmetric)
```

```
0.8277777777777778
```

4.3 Network Connectivity

4.3.1 Degree

The degree of a node determines how many connections it has. NetworkX's degree function can used to find out a node's degree in the network.

4.3.2 Distance

The functions nx.shortest path(Graph, Node1, Node2) and nx.shortest path length(Graph, Node1, Node2) in NetworkX can also be used to find the shortest path between two nodes and its length.

4.3.3 Breadth first search

By starting from that node and utilizing the breadth-first search strategy, how far apart each node in the network is from every other node can be determined. For this, networkX offers the function bfs tree.

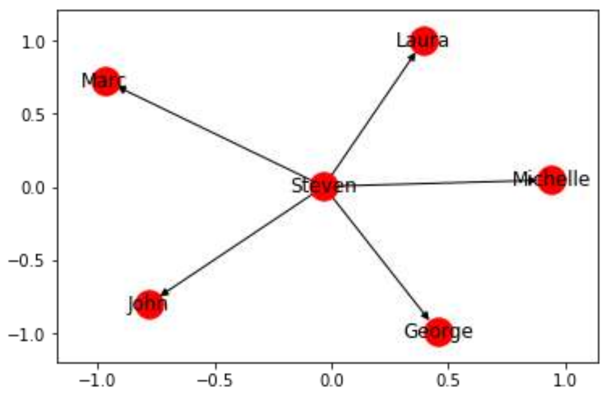

Fig 4.3 Breadth first search graph

4.4 Centrality measures

Centrality measures can find the most well-liked, influential, and popular members in the network.

4.4.1 Degree Centrality

Degree centrality is a way to quantify how many connections a specific node has within the network. Its foundation is the idea that significant nodes have a lot of connections. The degree centrality() function in NetworkX can be used to determine the degree of centrality of each network node.

```
{'Steven': 1.0,
 'Laura': 0.4,
 'Marc': 0.4,
 'John': 0.4,
 'Michelle': 0.6000000000000001,
 'George': 0.4}
```

4.4.2 Eigenvector Centrality

By considering how strongly a node is related to other significant nodes, eigenvector centrality calculates a node's importance.

```
{'Steven': 0.6006686104947806,
 'Laura': 0.3545677660798074,
 'Marc': 0.3545677660798074,
 'John': 0.30844592433424667,
 'Michelle': 0.4443904166426225,
 'George': 0.30844592433424667}
```

4.4.3 Closeness Centrality

Closeness Each node's significance is determined by how close it is to every other node in a measure called centrality.

```
{'Steven': 1.0,
 'Laura': 0.625,
 'Marc': 0.625,
 'John': 0.625,
 'Michelle': 0.7142857142857143,
 'George': 0.625}
```

4.4.4. Betweeness Centrality

It displays the frequency of occurrence of a spot along the shortest pathways connecting two points. It counts the number of times a specific node appears on the shortest route between two other nodes. The communication and information flow inside the network are significantly influenced by the nodes with high betweenness centrality.

High betweenness centrality nodes are able to strategically control and affect others. A person in such a key position has the ability to affect the entire group by either withholding or skewing the information being transmitted.

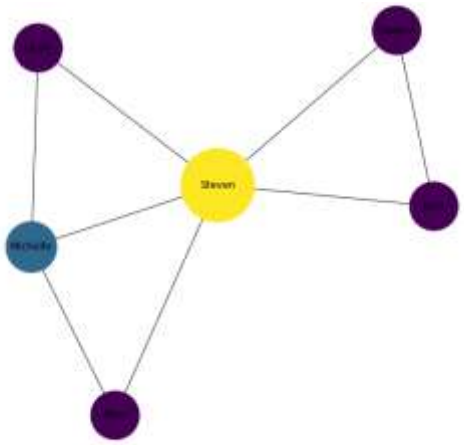

Fig 4.4 Visualization of Betweeness Centrality

5. INFERENCE

The graph fig 5.1 shows that some nodes are shared by Degree Centrality, a measure of degree, and Betweenness Centrality, a factor in information flow regulation.

Fig 5.1 visualization of the network such that the node color varies with Degree and node size with Betweenness Centrality

It is normal for nodes that are more related to one another to be on the shortest paths connecting them. The node 1912 is a significant node since all three centrality indicators that is considered place a high value on it.

6. CONCLUSION

In this paper, A useful technique that can accurately track the interactions in an online Problem Based Learning environment is social network analysis has been discussed. The general activity and the active groups in the course can be learned about via social network analysis. It provides a more complete view of the group and its members at the group level. assisting in the identification of those who are active, inactive, or isolated. A precise technique to add insights about the level of activity is by using mathematical parameters. Additionally, social network analysis's insights might be helpful in the context of learning analytics to track people's activities.

REFERENCES

[1] Jose Antonio Iglesias, Aaron García Cuerva, Agapito Ledezma, Araceli Sanchis, Social network analysis: Evolving Twitter mining, IEEE, 2019

[2] David Bright, Garry Robins, Patrycja Stys, Laurin Weissinger, DATA COLLECTION FOR SOCIAL NETWORKS RESEARCH, Science Direct, 2022.

[3] Garry Robins, Laurin Weissinger, Patrycja Stys, David Bright, Data Collection for Social Network Research: Challenging Contexts, Ethical Concerns, and New Approaches, Science Direct, 2021.

[4] Paola Tubaro, Louise Ryan, Antonio Casilli, Alessio D'Angelo, Recent ethical challenges in social network analysis, Science Direct, 2021.

[5] Nicholas M. Harrigan, Giuseppe (Joe) Labianca, Filip Agneessens, Social Network Research on Negative Ties and Signed Graphs, Science Direct, 2021.