Social Network Analysis for Understanding Online Information Diffusion Patterns

Antara Dessai[1], Prof. Siddhi Naik[2]

[1]Student, Department of Information Technology and Engineering, Goa College of Engineering, Farmagudi, Goa, India
[2]Assistant Professor, Department of Information Technology and Engineering, Goa College of Engineering, Farmagudi, Goa, India

---***---

Abstract - *Social Network Analysis (SNA) with a focus on unraveling the intricate patterns of online information diffusion. This study focuses on Social Network Analysis (SNA) techniques to gain insights into how information propagates through social network by means of data analysis of information shared on network by Visualizing the graph, measuring Degree Centrality, Betweenness Centrality, Closeness Centrality, Clustering Effects, Bridges, Assortativity, Network Communities and Plot Word Cloud, Bar Chart Representation for Sentiment Analysis.*

The study employs a multidisciplinary approach, combining elements data science, and social network to provide a comprehensive understanding of online information diffusion. Utilizing real-world datasets from prominent social media platforms, we analyze the interactions, connections, and behavioral patterns of users involved in the dissemination process.

***Key Words*:- Social Network Analysis,Influence Maximization, Sentiment Analysis**

1.INTRODUCTION

Social Network Analysis (SNA) is a powerful method used to study the relationships and interactions among individuals, groups, or organizations within a social system. It provides a valuable framework for understanding the structure, dynamics, and patterns of social networks.

Social Network Analysis stands as a pivotal tool in unraveling the intricate dynamics of online information diffusion patterns, providing a nuanced lens through which researchers can comprehend the complexities inherent in the digital dissemination of information. In the realm of project reports, a comprehensive overview of SNA's application in understanding online information diffusion is imperative for unlocking key insights.

SNA delves into the intricate web of relationships within social networks, enabling the identification of influential nodes and pathways that govern the flow of information. This analytical approach facilitates a granular examination of the factors influencing the speed, reach, and impact of information dissemination, shedding light on the underlying structures that shape online communication. Through the analysis of connections between nodes, SNA unveils the role of key influencers, mapping out the network elements that drive the diffusion process. Moreover, the exploration of network properties, such as density, centrality, and clustering, adds depth to the understanding of how information spreads. An illustration of Social Network is depicted in Figure 1.

Fig -1: Social Network

By adopting an SNA perspective, It can offer valuable insights into the interplay between social connections and the dynamics of information diffusion in the digital landscape. This holistic understanding has far-reaching implications for academics, policymakers, and businesses, providing actionable knowledge for optimizing communication strategies, enhancing online engagement, and fostering informed decision-making in an increasingly interconnected world.

2. Related works

While existing literature surveys predominantly focus on state-of-the-art methodologies within specific perspectives, our survey uniquely addresses the gap in social network analysis by incorporating privacy preservation. To the best of our knowledge, this marks the inaugural attempt to provide foundational insights into social network analysis issues, serving as an essential resource for emerging researchers in this field. Notably, our survey differs from previous ones in various aspects. For instance, Reference [10] primarily delves into information diffusion associated with community detection, concentrating solely on community detection algorithms within an information diffusion context. Recent studies, including [6, 8, 15, 17], have explored the influence maximization problem, offering insights into problem-solving approaches, performance evaluations, and theoretical approximation guarantees. Our work, therefore, stands out by providing a comprehensive and foundational overview of social network analysis, bridging critical knowledge gaps in the literature.

3. Methodology

1.Data Collection

Data Source is based on data collected from facebook and twitter. Both of them have a board specifically for information diffusion topic discussion.

facebook_combined Dataset- Facebook network is undirected and has no weights because one user can become friends with another user just once. Each node represents an anonymized facebook user that belongs to one of those ten friends lists. Each edge corresponds to the friendship of two facebook users that belong to this network. In other words, two users must become friends on facebook in order for them to be connected in the particular network.

Twitter Sentiments Dataset - A Twitter Sentiments Dataset typically consists of labeled data where each entry represents a tweet along with information about its sentiment. The objective of this task is to detect hate speech in tweets. ID is field usually contains a unique identifier for each tweet. It is a numerical or alphanumeric code assigned to differentiate one tweet from another. The ID serves as a reference point and helps in tracking and organizing the dataset. Label field indicates the sentiment associated with the tweet. Where label '1' denotes the tweet is racist/sexist and label '0' denotes the tweet is not racist/sexist and Tweet field contains the actual text of the tweet.

2. SOCIAL NETWORKS ANALYSIS MEASURES

The properties and measures outlined in this discussion serve as valuable tools for analyzing and quantifying various tasks within social networks. Researchers and analysts can leverage these measures to scrutinize network structures, pinpoint influential users, predict missing links, conduct behavioral analyses of product adoption, and more. The application of these measures enables the characterization of networks in a comprehensive manner.

2.1. Describing Nodes and Edges

A Facebook network is undirected and has no weights because one user can become friends with another user just once. Each node represents an anonymized facebook user that belongs to one of those ten friends lists. Each edge corresponds to the friendship of two facebook users that belong to this network. In other words, two users must become friends on facebook in order for them to be connected in the particular network.

Degree Centrality: Easily computed by assessing the number of connections a node has, higher degree nodes are considered more central. However, in dense graphs, the assumption that the highest degree node has the best reachability may not always hold true. In such cases, a chain of lower degree nodes can be beneficial for maximizing influence spread.

Betweenness Centrality: This measure evaluates a node's importance in all shortest paths between pairs of nodes in the network, computing the fraction of these paths a node is present in. It is particularly suited for applications involving information diffusion.

Closeness Centrality: Defined as a node's proximity to the rest of the network, closeness centrality is calculated as the average of the shortest distances between a node and other nodes. A node with a lower closeness score is considered more central, indicating better connectivity to most nodes in the network. This centrality measure is useful for influence maximization, as a node's influence is confined to its local region.

Fig-2: Degree Centrality

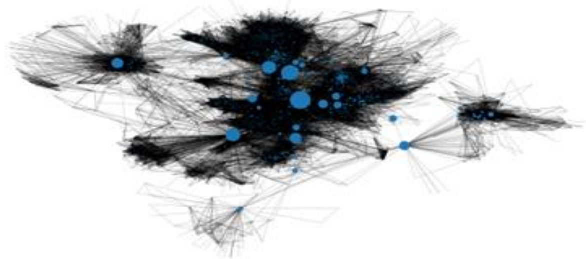

Fig -3: Betweenness Centrality

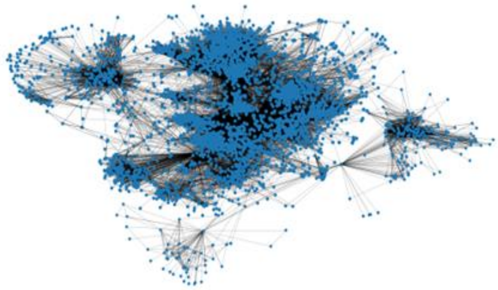

Fig -4: Closeness Centrality

3. Clustering Effect

The Clustering Effect: is a measure that reflects a node's tendency to form coalitions within a social network, determined by the total number of triplets present. A higher clustering coefficient indicates dense interconnections among nodes in the network. Essentially, it is the average of neighborhood densities for all nodes, illustrating the extent to which nodes are embedded in the network. In other words, a network with tightly connected nodes will exhibit a higher clustering coefficient, highlighting the cohesive nature of the relationships within the network. This measure provides valuable insights into the structural cohesion and interdependence of nodes in the social network.

4. Community Detection

Plays a crucial role in the dissemination of information as it establishes connections among nodes with similar ideas. The spread of information tends to be slower when the consensus is reached within a community. Members of a community receive confirmation that a majority supports a particular idea, boosting their confidence in propagating the information. Determining the key node facilitates faster message propagation through the network. The density of a community serves as an indicator of how well-connected its members are, shedding light on the overall cohesion within the community.

Fig -5: Community Detection

5. Sentiment Analysis

Understand the emotional tone associated with the spread of information. Positive or negative sentiments can significantly influence the diffusion patterns. Sentiment analysis plays a crucial role in understanding information diffusion within social networks. It involves the use of natural language processing and machine learning techniques to determine the sentiment or emotional tone expressed in textual data.

Fig -4: The 80 most frequent tokens in the database as a word cloud. Those tweet contains hate speech if it has a racist or sexist sentiment associated with it. So, the task is to classify racist or sexist tweets from other tweets.

3. CONCLUSIONS

Social Network Analysis stands as a valuable and multidimensional tool for unraveling the intricate dynamics of online information diffusion patterns. Visualization serves as a powerful tool for unraveling the complex network patterns, enabling a more intuitive comprehension of the relationships and clusters within the data. Centrality measures, including degree centrality and betweenness centrality highlight the pivotal nodes and individuals that play crucial roles in information flow and network cohesion. Identifying these central nodes aids in understanding key influencers and potential points of control or influence within the network.

Clustering effects reveal the presence of communities within the facebook network. Analyzing bridges and assortativity provides further depth by exposing connections between different clusters and uncovering patterns of similar or different among nodes. Network communities, identified through community detection algorithms help to identify existence of groups with shared interests. Sentiment analysis, when applied to the content shared within the network, enriches our understanding by guessing the sentiment prevalent among users.

REFERENCES

[1] Shanshan Feng , Kaiqi Zhao , Lanting Fang , Kaiyu Feng, Wei Wei , Xutao Li , and Ling Shao . "H-Diffu: Hyperbolic Representations for Information Diffusion Prediction". IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING, VOL. 35, NO. 9, SEPTEMBER 2023.

[2] Bhawna Saxena, Vikas Saxena."Influence Maximization in Social Networks using Hurst exponent based Diffusion Model".IEEE 10th International Conference on Cloud Computing, Data Science & Engineering ,2020.

[3] Liangtian Wan , Xiaona Li , Jin Xu , Lu Sun , Xianpeng Wang , and Kaihui Liu. "Application of Graph Learning With Multivariate Relational Representation Matrix in Vehicular Social Networks". IEEE TRANSACTIONS ON INTELLIGENT TRANSPORTATION SYSTEMS, VOL. 24, NO. 3, MARCH 2023.

[4] A. Susi, Dr. V. Akila, Dr. V. Govindasamy."ANALYSIS OF INFORMATION DIFFUSION IN SOCIAL NETWORKS BASED ON GRAPH REPRESENTATION". IEEE 23 International Conference on Networking and Communications (ICNWC) ,2023.

[5] Hui Li, Hui Li, Sourav S Bhowmick. "CHASSIS: Conformity Meets Online Information Diffusion". In Proceedings of the 2020 ACM SIGMOD International Conference on Management of Data (SIGMOD'20), JUNE 2020.

[6] Shiori Furukawa, Sho Tsugawa."Limitations of Link Deletion for Suppressing Real Information Diffusion on Social Media". 2021 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining, NOVEMBER 2021.

[7] B. Liu, Opinions, Sentiment, and Emotion in Text. Sentiment Analysis: Mining Opinions, Sentiments, and Emotions, Cambridge University Press, 2015.

[8] J. Jansen, M. Zhang, K. Sobel, and A. Chowdury, "Twitter power: Tweets as electronic word of mouth," JASIST, vol. 60, pp. 2169–2188, 11 2009.

[9] K. Curran, K. O'Hara, and S. O'Brien, "The role of twitter in the world of business," IJBDCN, vol. 7, pp. 1–15, 07 2011.

[10] G. Ranco, D. Aleksovski, G. Caldarelli, M. Grcar, and I. Mozeti ˇ c, "The ˇ effects of twitter sentiment on stock price returns," PloS one, vol. 10, p. e0138441, 09 2015.

[11] X. Chen, Y. Cho, and S. Y. Jang, "Crime prediction using twitter sentiment and weather," in 2015 Systems and Information Engineering Design Symposium, pp. 63–68, 2015.

[12] B. R. Prathap and K. Ramesha, "Twitter sentiment for analysing different types of crimes," in 2018 International Conference on Communication, Computing and Internet of Things (IC3IoT), pp. 483–488, 2018.

[13] Nicolas Claidière and Andrew Whiten. 2012. Integrating the study of conformity and culture in humans and nonhuman animals. Psycholog�ical bulletin 138, 1 (2012), 126.

[14] William D Crano. 1970. Effects of sex, response order, and expertise in conformity: A dispositional approach. Sociometry (1970), 239–252.

[15] Morton Deutsch and Harold B Gerard. 1955. A study of normative and informational social influences upon individual judgment. The journal of abnormal and social psychology 51, 3 (1955), 629.

[16] Kuntal Dey, Ritvik Shrivastava, and Saroj Kaushik. 2017. Twitter Stance Detection - A Subjectivity and Sentiment Polarity Inspired Two�Phase Approach. In 2017 IEEE International Conference on Data Mining Workshops, ICDM Workshops 2017, New Orleans, LA, USA, November 18-21, 2017. 365–372.

[17] Sainyam Galhotra, Akhil Arora, and Shourya Roy. 2016. Holistic Influence Maximization: Combining Scalability and Efficiency with Opinion-Aware Models. In Proceedings of the 2016 International Confer�ence on Management of Data, SIGMOD Conference 2016, San Francisco, CA, USA, June 26 - July 01, 2016. 743–758.

[18] Jacob K Goeree, Thomas R Palfrey, Brian W Rogers, and Richard D McKelvey. 2007. Self-correcting information cascades. The Review of Economic Studies 74, 3 (2007), 733–762.

[19] Alan G Hawkes. 1971. Spectra of some self-exciting and mutually exciting point processes. Biometrika 58, 1 (1971), 83–90.

[20] Adrien Guille, Hakim Hacid, Cécile Favre, and Djamel A. Zighed. 2013. Information diffusion in online social networks: a survey. SIGMOD Record 42, 2 (2013), 17–28

. .