e-ISSN:2582-7219

INTERNATIONAL JOURNAL OF MULTIDISCIPLINARY RESEARCH IN SCIENCE, ENGINEERING AND TECHNOLOGY

Volume 7, Issue 13, April 2024

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

Impact Factor: 7.521

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

| ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 7.521| | Monthly Peer Reviewed & Refereed Journal |

|| Volume 7, Issue 13, April 2024 ||

International Conference on Intelligent Computing & Information Technology (ICIT-24)

Organized by

Erode Sengunthar Engineering College, Erode, Tamilnadu, India

Graph-Based Analysis of Twitter Data Using Social Network Visualization and Analysis Tool

Tharani.A[1], S.B. Mahalakshmi[2], E. Chandra Blessi[3], R. Banupriya[4]

PG Scholar, Dept. of. Computing (Artificial Intelligence and Machine Learning) Coimbatore Institute of Technology, Coimbatore, Tamil Nadu, India[1]

Assistant Professor, Department of Computing (Artificial Intelligence and Machine Learning), Coimbatore Institute of Technology, Coimbatore, Tamil Nadu, India[2,3,4]

ABSTRACT: Twitter is the most frequently used social media for sharing the messages among the users. It provides networking facilities such as information sharing, communication structures within the network, and exploration of user influence. Social Network Analysis on Twitter data involves nodes' importance in the broader network, node clustering, and coloring based on the features and nodes interaction among the users for finding similar patterns. Many research in this area is done using a popular tool called Gephi, a Social Network Visualization and Analysis tool. In this paper, degree distribution analysis, interaction pattern, link influence analysis, community detection, temporal trends analysis, path analysis for communication effectiveness are carried out for the twitter data. The analysis demonstrates the effectiveness of Gephi for exploring Twitter data. The visual result and the findings shows the better understanding of the user dynamics and invaluable insights into the intricate dynamics of social interactions and the diffusion of information within Twitter networks.

KEYWORDS: Twitter, Visualization, Network analysis, Gephi and Degree distribution.

I. INTRODUCTION

Twitter stands out as one of the most pervasive social media platforms, serving as a hub for message sharing among its users. Twitter data offers numerous benefits across diverse fields. It provides access to real-time data, enabling businesses, researchers, and policymakers to monitor current events and communications, businesses can gain insights into users behaviors. Twitter data [8] enables the analysis of competitors' activities, the assessment of market trends, and the identification of opportunities for differentiation and industry growth. Researchers can leverage Twitter data to study various social phenomena, conduct sentiment analysis, track the spread of information, and explore network dynamics, contributing to advancements in fields such as sociology, psychology, and communication studies. Twitter data offers a wealth of opportunities for businesses, researchers, and organizations to gain valuable insights, drive informed decision-making, and enhance their overall performance and effectiveness in the digital landscape.

Gephi stands as a robust and user-friendly solution tailored for visualizing and dissecting intricate networks. As open-source software, it empowers researchers, analysts, and enthusiasts alike to delve into, manipulate, and interpret network data in various domains. Gephi provides an intuitive interface and comprehensive features, offering a variety of tools for network visualization, layout algorithms, statistical analysis, and community detection. These functionalities render it essential for comprehending the structure, dynamics, and patterns inherent in networks spanning social, biological, transportation, and other domains.

Gephi's visualizations enable users to reveal insights that may not be immediately discernible from raw data alone. Through visual representation of nodes (entities) and edges (connections), users can pinpoint clusters, central nodes, and other critical network properties. This facilitates the exploration of intricate relationships and the detection of emerging patterns and structures. Gephi emerges as an indispensable and adaptable tool for visualizing and analyzing networks. It equips users with the capacity to acquire valuable insights into the structure, dynamics, and behaviors of intricate systems spanning diverse domains.

Gephi is a powerful tool for conducting social network analysis on Twitter data [10]. With Gephi, individuals have the capability to import Twitter data including user profiles, tweets, retweets, mentions, and follower relationships, and depict them as nodes and edges in a network graph. Gephi presents a range of layout algorithms for visualizing the network structure, simplifying the identification of clusters, influential users, and communication patterns.

Moreover, Gephi equips users with tools for community detection, centrality analysis, and clustering coefficient calculation, facilitating a deeper understanding of the network's topology and dynamics. This enables researchers to explore information diffusion, user influence, sentiment analysis, and various other aspects of Twitter interactions through Gephi's analytical capabilities.

This paper is organized as follows: Chapter 1 gives the introduction followed by Chapter 2 with the survey on related works. Chapter 3 describes the social network analysis techniques. Chapter 4 introduces the methodologies used followed by the experimental analysis with results and discussion in Chapter 5. Chapter 6 concludes with a summary.

II RELATED WORKS

In [1], the authors offered a thorough examination of the topological characteristics of the Twitter follow graph, encompassing metrics like degree distributions, connected components, shortest path lengths, clustering coefficients, and degree distribution. The analysis yields a set of authoritative statistics that can serve as reference points for the research community.

The authors in [2] reviewed several variants, contextualized them, and demonstrated their computability using simplified versions of the algorithm commonly employed for the standard case. The contribution in this paper is the compilation of various shortest-path betweenness variants and the computation of efficient algorithms.

Paper [3] conducts an analysis of tweets with the hashtag #onlineclasses from Filipino users, utilizing network analysis techniques with Gephi and NodeXL software. The outcomes of the network analysis conducted with Gephi and NodeXL offer insights into how users, primarily students, perceive online classes and their experiences with this novel mode of learning.

The article [4] introduces a novel conceptual and practical model for classifying topical Twitter networks based on their network-level structures. While existing literature primarily focuses on classifying individual users into key positions, this study proposes an approach that examines overall network structures to classify Twitter conversations based on their information flow patterns. By leveraging four network-level metrics—density, modularity, centralization, and the fraction of isolated users—a three-step classification model is developed. This model identifies six distinct structures of information flow: divided, unified, fragmented, clustered, in-hub-and-spoke, and out-hub-and-spoke networks.

In [5], The MGephi framework which was proposed serves as an integral component of the workflow, facilitating the visualization of output with enhanced information display, including node labels, IDs, edges, weights, and more. During the filtering process, specific portions of the graph are visualized with clarity. The overall graph organization is achieved by employing multiple algorithms such as ForceAtlas, Fruchterman-Reingold, Harel-Koren Fast Multiscale, and Kamada-Kawai, aimed at identifying clusters, nearest nodes, measuring centrality, betweenness, and other metrics. Permutation involves reading the network, creating partitions, and converting vectors into the graph. Additionally, fisheye transformation enables graph rotation in Cartesian or polar coordinates with adjustable movement parameters.

The study [7] focuses on text mining and social network analysis methods applied to Twitter data analysis. The study investigates techniques for deriving insights from Twitter data, which is notable for its abundance and lack of organization. The authors use entity recognition, topic modeling, sentiment analysis, and other text-mining techniques to glean useful information from tweets. They also use social network analysis to find impact and interaction trends among Twitter users. The study addresses the opportunities and difficulties involved in studying data from Twitter, emphasizing the value of combining text mining and social network research techniques to provide a thorough

| ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 7.521| | Monthly Peer Reviewed & Refereed Journal |

|| Volume 7, Issue 13, April 2024 ||

International Conference on Intelligent Computing & Information Technology (ICIT-24)

Organized by

Erode Sengunthar Engineering College, Erode, Tamilnadu, India

understanding of user behavior and trends on the network. Overall, the study contributes to the field of data mining and analytics by demonstrating effective methods for analyzing Twitter data and deriving actionable insights from it.

III. SOCIAL NETWORK ANALYSIS TECHNIQUES

Social Network Analysis (SNA) is a systematic method for comprehending social structures and connections by investigating the interactions among individuals, groups, or organizations within a network. It entails scrutinizing the patterns of relationships or links between these entities to reveal fundamental structures, behaviors, and dynamics. The main components of analysis are the nodes which represent entities, edges which represent relationships, network structure which includes connection patterns and association within the network, network measures and metrics which include properties of the network, Community detection which identifies the communities and visualization which gives a better understanding of the pattern and dynamics within the network.

A. Distribution measures:

1. Degree Distribution:

Degree distribution within a social network characterizes the arrangement of links between nodes based on their degrees, which represent the number of connections each node possesses. This distribution sheds light on how engagement is dispersed among users within the social context. Typical distributions include the normal distribution, where degrees are evenly distributed among users; the power-law distribution, where a minority of users possess a disproportionate number of connections (acting as hubs); and the exponential distribution, signifying a rapid decline in connectivity.

The degree distribution of a network can be represented by a probability mass function P(k), where k is the degree of a node. In mathematical terms, the degree distribution formula can be expressed as:

$P(k)= \frac{Nk}{N}$

Where Nk is the number of nodes with degree K and N is the total number of nodes in the network.

By scrutinizing degree distribution, analysts can unveil the underlying structure of the network, pinpoint influential users, identify potential subgroups, and understand the overall dynamics of connectivity. Such analysis offers valuable insights into the nature and resilience of the social network as a whole.

2. Centrality Measures

Centrality metrics in social network analysis assess the significance or prominence of nodes within a network, centrality measures are computed, including degree, betweenness, and closeness centralities. These centrality measures direct additional research and interpretation by offering insights into the significance, influence, and connectivity of particular nodes.

Additionally, We use Gephi's Timeline tool to examine how interactions change over time provided the dataset has temporal information. By identifying temporal patterns, trends, and fluctuations in network activity, temporal analysis gives a better understanding of network dynamics. If the dataset has weighted edges, weighted network analysis can also be performed, allowing to evaluation the frequency or intensity of interactions between nodes. An iterative process for analysis is used, continuously improving analyses and visualizations based on new information. Various facets of the network are investigated, and necessary modifications are made to visualization parameters or analytic methods as shown in figure 1. Lastly, the results are interpreted, with conclusions that meet the goals of the study and take the consequences of the results into account. The approach is thoroughly documented, guaranteeing transparency and repeatability in reporting results. Information on data sources, preprocessing stages, analytic methods, and result interpretation are all
included.

| ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 7.521| | Monthly Peer Reviewed & Refereed Journal |

|| Volume 7, Issue 13, April 2024 ||

International Conference on Intelligent Computing & Information Technology (ICIT-24)

Organized by

Erode Sengunthar Engineering College, Erode, Tamilnadu, India

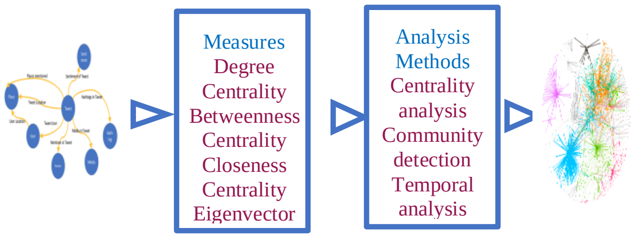


Figure 1: Social Network measures and Analysis methods

IV. EXPERIMENTAL ANALYSIS

A. DATASET DESCRIPTION:

The dataset needed for analysis is scraped from Twitter. Table 1 describes the number of nodes and edges. The node details include usernames or individuals in the network, number of followers, the time of user account was created, number of friends connected to the users, URL or path to the user's profile image, and geographical location associated with the user.

Table 1: Dataset description

Dataset	Number of Rows	Number of columns
EDGE CSV	5390	5
NODE CSV	2174	10

The edge details include an identifier for the source node and the target node, the nature of the interaction, the frequency of the relationship, and the interaction type. The dataset that is being made available depicts a directed graph that records interactions on a social media site—possibly Twitter. Number indices link each node to a user or other object, and directed actions are indicated by the edges. A network of interactions is created when an action is taken by the "Source" node and received by the "Target" node. While the "Weight" column, which is universally set to 1, indicates an equal value for each contact, the "Type" column describes the type of interactions, such as mentions, retweets, or quotes. Further background is provided by the "Relationship" column, which describes particular types of interactions. This dataset allows for examination of user influence, information transmission, and the structure of communication within the network, offering insightful information about the dynamics of social media participation.

B. RESULT AND DISCUSSION

Distribution Measures:
Degree distribution
Average Degree: 1.474

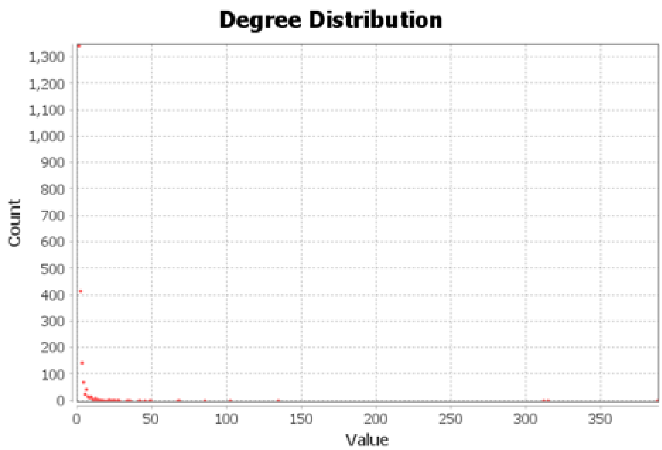


Figure 2: Degree distribution for twitter data

- **Out-Degree and In-Degree Distribution:** In social network analysis, out-degree and in-degree distributions are crucial metrics that shed light on how nodes in a directed network initiate and receive interactions. Higher out-degrees indicate active initiators, and out-degree distribution shows how nodes spread actions or information. On the other hand, nodes that draw interactions are highlighted by the in-degree distribution, where more in-degrees indicate more mentions or references from other people. These distributions help identify significant initiators and nodes that draw a lot of interaction by revealing information flow dynamics, influence patterns and the general structure of social network as given in figure 3.

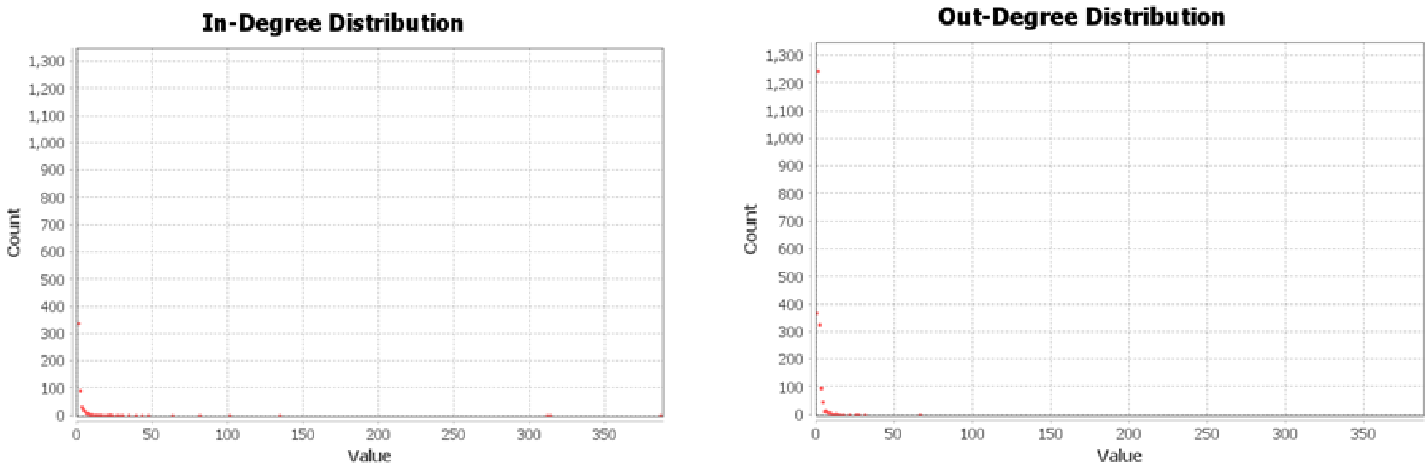


Figure 3: In-degree and out-degree distribution

V. INTERACTION PATTERNS

Interaction Type Distribution: The analysis of the frequency and distribution of various types of interactions among nodes is referred to as interaction type distribution in social networks shown in figure 4. These exchanges can take the shape of quotes, mentions, retweets, or other types of engagement on social media. Analyzing the distribution of interaction types can provide information about the most common types of interaction in the network.

| ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 7.521| | Monthly Peer Reviewed & Refereed Journal |

|| Volume 7, Issue 13, April 2024 ||

International Conference on Intelligent Computing & Information Technology (ICIT-24)

Organized by

Erode Sengunthar Engineering College, Erode, Tamilnadu, India

Through the identification of more prevalent or significant interaction types, this research contributes to a more complex understanding of the dynamics and communication patterns within the social network. It makes it possible for analysts or researchers to comprehend the various ways that users interact with one another, leading to a more thorough comprehension of the effect and flow of information inside the network.

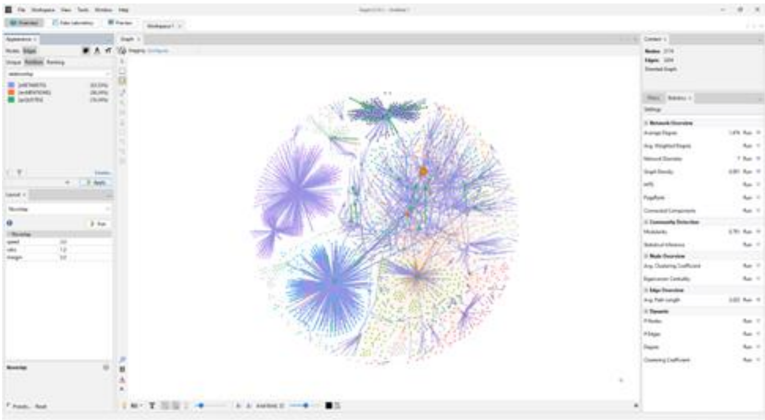
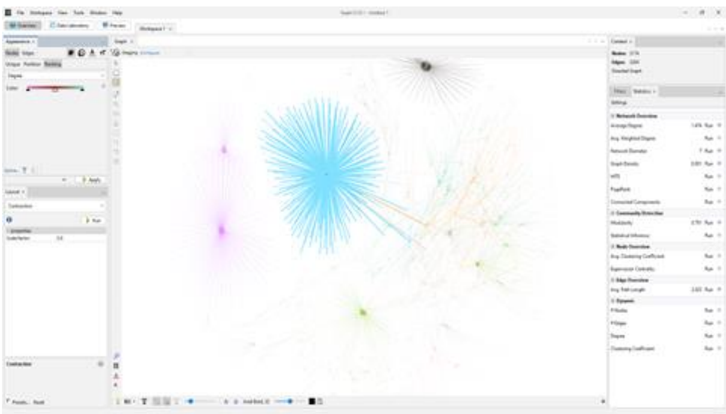


Figure 4: Interaction type distribution in social networks

Weighted Network Analysis: Examining the number or strength of connections between nodes in a network while taking edge weights into account is known as weighted network analysis as shown in figure 5. These weights can stand in for a variety of metrics in social network analysis, including communication frequency, interaction intensity, and other pertinent variables. Compared to a binary (unweighted) network, analyzing a weighted network offers a more detailed knowledge of the connections between nodes. Calculating metrics like average edge weight, finding highly weighted connections, and displaying the network with edges scaled according to their weights are important tasks in weighted network research. This method is useful for determining the significance and magnitude of contacts, identifying impact patterns, and comprehending the weighted structure of the social network as a whole.



VI. INFLUENCE ANALYSIS

PageRank: In network theory, PageRank shown in figure 6 is a link analysis technique that is used to assess a node's value in a directed graph, specifically webpages within the setting of the World Wide Web. PageRank, created by Larry Page and Sergey Brin at Google, gives each node in the network a numerical weight determined by the quantity and caliber of links pointing to it. The technique works on the basis that a node is considered more important if it receives

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

| ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 7.521| | Monthly Peer Reviewed & Refereed Journal |

|| Volume 7, Issue 13, April 2024 ||

International Conference on Intelligent Computing & Information Technology (ICIT-24)

Organized by

Erode Sengunthar Engineering College, Erode, Tamilnadu, India

more incoming links from other high-quality nodes. PageRank is used in the context of social network analysis to determine the number and importance of interactions that significant persons or nodes within the network get.
Epsilon = 0.001
Probability = 0.85

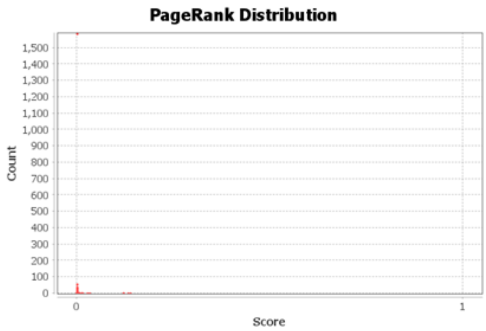
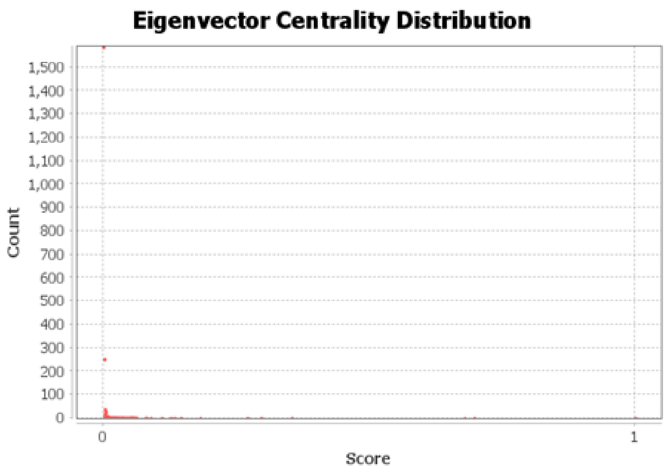


Figure 6: PageRank result for twitter data

Eigenvector Centrality: In network analysis, eigenvector centrality shown in figure 7 is a metric that evaluates a node's significance by considering the caliber of its connections rather than their amount. Every node is given a centrality score based on the weighted connections a node has in addition to the quantity of connections it has. Eigenvector centrality scores are essentially greater for nodes that are connected to other highly central nodes. The idea behind this metric is that a node may only be considered influential if it has connections to other influential nodes. Eigenvector centrality is a useful tool in social network analysis to find people who are not only highly linked but also related to other people who are important to the network as a whole.
Network Interpretation: directed
Number of iterations: 100
Sum change: 0.016238352597184317



VII. COMMUNITY DETECTION

Modularity Analysis: In social network analysis, modularity analysis is a technique used to find communities or groups of nodes in a network that have more connections amongst them than they do with nodes outside of the group. The degree to which the network can be split up into these distinct communities is measured by its modularity. Finding node subgroups with stronger internal connections—which may indicate a certain level of cohesion or specialization within

| ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 7.521| | Monthly Peer Reviewed & Refereed Journal |

|| Volume 7, Issue 13, April 2024 ||

International Conference on Intelligent Computing & Information Technology (ICIT-24)

Organized by

Erode Sengunthar Engineering College, Erode, Tamilnadu, India

those clusters—is the aim. Within social networks, modularity analysis is used to identify discrete user communities that engage in mutual interactions, have similar interests, or take part in related conversations as shown in figure 8.
Modularity: 0.791
Modularity with resolution: 0.791
Number of Communities: 20

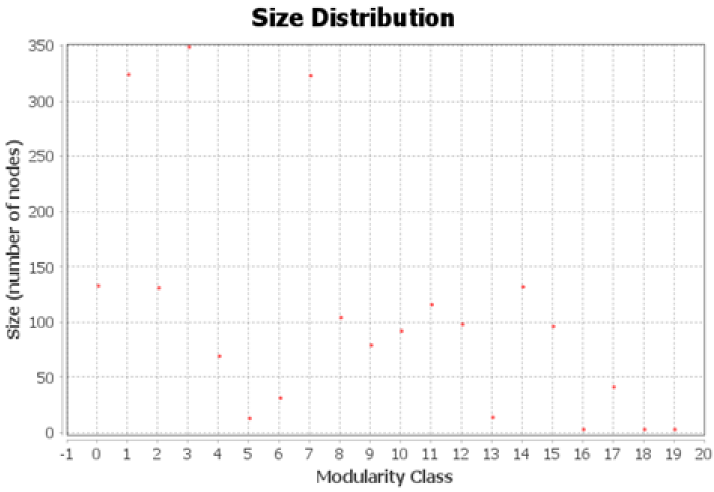


Figure 8: Modularity Analysis for the twitter data

VIII. TEMPORAL ANALYSIS

Temporal Patterns: In social network analysis, temporal patterns refer to the study of how links and interactions change over time. When examining temporal trends, timestamps connected to interactions—like posts, mentions, retweets, or other activities—in a social network are taken into account. Through the evaluation of temporal dynamics, trends, oscillations, or recurrent patterns in user engagement can be found. Peak activity times, the dates of significant events, and the temporal behavior of certain nodes or communities are all revealed by this study. Comprehending temporal patterns is essential to understanding the dynamics of information flow, pinpointing times of high or low activity, and appreciating the network's evolution over time. Strategic planning, content scheduling, and general understanding of the temporal dynamics inside a social network all benefit from this knowledge.

IX. PATH ANALYSIS

Shortest Paths: The most direct channels between pairs of nodes in a social network are referred to as the "shortest paths," and they show how few connections are required to move from one node to another. Finding the shortest pathways is a key idea in graph theory and network analysis, with practical implications in comprehending how information or influence can move through a network efficiently. Finding the network's important links and possible bottlenecks is made easier by doing shortest path analysis. The shortest pathways in social networks can be used to determine how fast information can move between members, providing insight on communication effectiveness and possible obstacles to information flow. This research offers insightful information about the connectivity, structure, and channels via which interactions spread within the network. Figure 9 shows the temporal pattern obtained for the twitter data.

| ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 7.521| | Monthly Peer Reviewed & Refereed Journal |

|| Volume 7, Issue 13, April 2024 ||

International Conference on Intelligent Computing & Information Technology (ICIT-24)

Organized by

Erode Sengunthar Engineering College, Erode, Tamilnadu, India

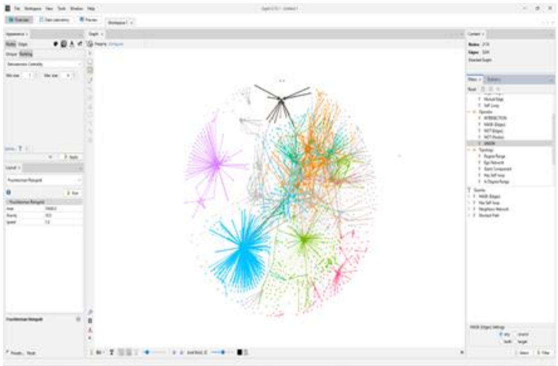


Figure 9: Temporal pattern of the network

Betweenness and Closeness Centrality: Critical metrics in social network analysis for assessing nodes' relative importance according to their placements within a network are betweenness centrality and closeness centrality as shown in figure 10. Betweenness centrality is a technique used to identify nodes that serve as important links in a network, allowing communication between various parts. High betweenness centrality nodes frequently have a crucial function to perform in preserving general connection and information flow. Conversely, closeness centrality tracks how fast a node may reach others in the network, emphasizing nodes that are central and easily reachable. Information can be quickly shared to a large audience by nodes with high proximity centrality. When taken as a whole, these centrality metrics offer a complex picture of the functional and expressive roles played by nodes in social networks.

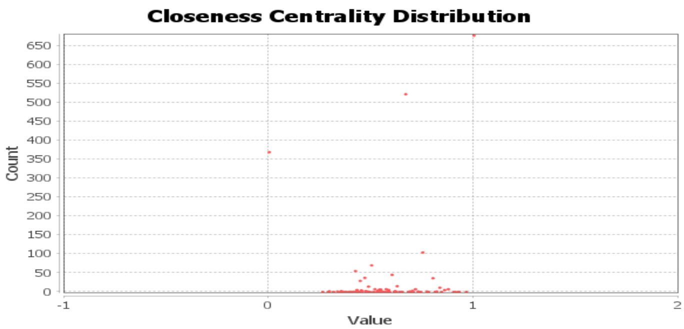


Figure 10: Closeness centrality distribution

X. DISCUSSION

Gephi's intricate method of social network analysis reveals a thorough comprehension of the network's dynamics, key nodes, and structure. Gephi guarantees data consistency and accuracy by importing and examining the dataset. The intricate relationships within the social network are clearly and intuitively represented in the following visualization, which makes use of sophisticated layouts and features. The visual analysis is improved by the tool's ability to modify node sizes, colors, and layouts based on metrics like degree centrality, emphasizing important nodes and their importance. The community detection techniques in Gephi aid in the modularity analysis by revealing discrete communities or clusters inside the network. This makes subgroup structures and interaction patterns visible, which helps find coherent node groups. By defining the significance of nodes in terms of their connectedness and influence, centrality measures like degree, betweenness, and closeness centrality are applied, further enhancing the study.

Gephi's dynamic and interactive features let academics explore temporal trends and track how interactions evolve over time. Finding patterns, pinpointing times of high activity, and comprehending the temporal evolution of the network all depend on this temporal analysis. A more comprehensive picture of the network's connectedness is provided by Gephi's support for weighted network analysis, which also sheds light on the frequency and strength of interactions. Gephi's cross-platform compatibility and open-source design encourage accessibility and teamwork all along the way. Because the tool can be extended through plugins, it can be customized and its capabilities can be expanded to meet a wide range of research needs. The network's iterative exploration, along with the export and save functions, make it easier to share research results and provide assistance for later use.

To put it briefly, the deep inference from the Gephi social network analysis entails figuring out the network's underlying structure, recognizing important nodes and clusters, deciphering temporal patterns, and learning more about the general dynamics of interaction and information flow. Researchers, analysts, and decision-makers who want to understand the complex relationships found in social networks and gain valuable insights for a range of applications—from community management to marketing strategies—need this information.

XI. CONCLUSION

The insights derived from social network analysis grow more and more pertinent as social networks develop and become more important in today's world. Researchers may continue to investigate and comprehend the complex dynamics of social networks by utilizing the capabilities of network analysis techniques and tools like Gephi. This will open up new avenues for creative solutions, significant interventions, and a greater understanding of human behavior and interaction.

REFERENCES

[1] S. A. Myers, A. Sharma, P. Gupta, and J. Lin, “Information network or social network?: the structure of the twitter follow graph,” in Proceedings of the 23rd International Conference on World Wide Web. ACM, 2014, pp. 493–498.
[2] U. Brandes, “On variants of shortest-path betweenness centrality and their generic computation,” Social Networks, vol. 30, no. 2, pp. 136–145, 2008.
[3] Joje Mar P Sanchez, Blanca A Alejandro, Michelle Mae Jugasan Olvido, Isidro Max Veloso Alejandro, “An Analysis of Online Classes Tweets Using Gephi: Inputs for Online Learning”, International Journal of Information and Education Technology 11(12):583-589, October 2021, DOI:10.18178/ijiet.2021.11.12.1568.
[4] Itai Himelboim, Marc A. Smith, Lee Rainie, Ben Shneiderman, and Camila Espina, “Classifying Twitter Topic-Networks Using Social Network Analysis”, Social Media + Society, pp.1-13, March 2017, DOI:10.1177/2056305117691545.
[5] Dr. M.Thangaraj, S.Amutha, “Mgephi : Modified Gephi for Effective Social Network Analysis”, International Journal of Scientific Research in Computer Science, Engineering and Information Technology © 2018 IJSRCSEIT | Volume 3 | Issue 1 | ISSN : 2456-3307.
[6] Niyati Aggrawal; Anuja Arora, Visualization, analysis and structural pattern infusion of DBLP co- authorship network using Gephi, 2016 2nd International Conference on Next Generation Computing Technologies (NGCT), Pages: 494 -500.
[7] Yanchang Zhao “Analysing Twitter Data with Text Mining and Social NetworkAnalysis” Conference: the 11th Australasian Data Mining & Analytics Conference (AusDM 2013)
[8] M. Cha, H. Haddadi, F. Benevenuto, and K. Gummadi, “Measuring user influence in twitter: The million follower fallacy”, Fourth International AAAI Conference on Weblogs and Social Media, vol. 4, no. 1, pp. 10-17, 2010.
[9] A. Alcalde. (Nov. 2018). An analysis and visualization of my twitter account with Gephi. The Programmer Chest . [Online]. Available: https://elbauldelprogramador.com/en/social-mining-gephi/.
[10] Gephi. (2017). About Gephi. [Online]. Available: https://gephi.org/about/

INTERNATIONAL JOURNAL OF MULTIDISCIPLINARY RESEARCH IN SCIENCE, ENGINEERING AND TECHNOLOGY

| Mobile No: +91-6381907438 | Whatsapp: +91-6381907438 | ijmrset@gmail.com |

www.ijmrset.com