Matrix: Jurnal Manajemen Teknologi dan Informatika
Volume 14 Issue 1 Year 2024 Pages 25-32

DOI: http://dx.doi.org/10.31940/matrix.v14i1.25-32
URL: http://ojs2.pnb.ac.id/index.php/matrix

Identification of top influence users in disseminating information on the 2024 Indonesian National Election

***Neny Sulistianingsih* [1]*, *Galih Hendro Martono* [2]**

[1,2] *Engineering Faculty, Universitas Bumigora, Indonesia*

*Corresponding Author: *neny.sulistianingsih@universitasbumigora.ac.id*

Abstract: Social media has a vital role in general elections in Indonesia because social media is one of the platforms used by presidential candidates for campaigns to gain public support. General elections in Indonesia occur every five years. Many tweets talk about presidential candidates approaching the national election period. Not least, some buzzers deliberately use Twitter to carry out propaganda against a candidate or to bring down other presidential candidates with their opinions because information can spread widely and quickly on Twitter. Based on this, it is necessary to identify influential users in disseminating information related to the 2024 National Election, especially on Twitter. Various centrality methods were used in this study to identify influence users in sharing information about the 2024 National Election such us Degree Centrality, Closeness Centrality, Harmonic Centrality, Eigenvector Centrality, and Load Centrality. For the evaluation in this study, the results of each method were compared to one another to measure the similarity and correlation between the ranking lists of users who were influential in disseminating information about the 2024 National Election.

Keywords: centrality, disseminating information, Indonesia, identification influence user, national election 2024

History Article: Submitted 29 August 2023 | Revised 30 April 2024 | Accepted 21 May 2024

How to Cite: N. Sulistianingsih and G. H. Martono, "Identification of top influence users in disseminating information on the 2024 Indonesian National Election," *Matrix: Jurnal Manajemen Teknologi dan Informatika,* vol. 14, no. 1, pp. 25-32, 2024.

Introduction

Social media has been widely used for political communication with the aim of political campaigns or verbally attacking political opponents. One of the social media that is widely used for political communication is Twitter [1]. Twitter is a microblogging that allows users to convey opinions or opinions of up to 280 characters. With the number of tweets reaching 500 million a day [2], many researchers are interested in studying data on Twitter. One area of Twitter analysis is identifying users who influence a particular topic. In politics, identifying influential actors can be used to identify actors who are buzzers or not. The buzzer in political communication on Twitter can cause political discussion to be not neutral. For certain parties, the identification of influential actors can be used as a way to conduct campaigns quickly and effectively.

In Indonesia, general elections was in 2024 to elect to be held for all lines of government, from regents, governors, and legislative bodies such as the DPR and DPRD to the President of Indonesia for the next five years. Based on this, since 2023, various parties, especially interested parties, have disseminated information and news related to the election. The dissemination of information on Twitter is unlimited and anonymous, causing the dissemination of information via social media cannot be filtered. News that is real or a hoax can spread without a filter. These stories greatly influence their readers and can further influence people's decisions to vote in the 2024 National Election [3]. The public needs to filter the information presented, especially on social media, so the information they read comes from accurate sources.

State of the art in this research encompasses various approaches undertaken to analyze the influence of social media users, particularly on Twitter, in the context of politics and elections. Previous studies have extensively explored the use of Twitter for political communication, including campaigns and verbal attacks against political opponents. Identifying influential users has become a focal point in this analysis to reveal actors acting as "buzzers" in political discussions.

These buzzers can significantly skew political discussions, which certain parties may exploit for rapid and effective campaigning.

In different contexts, research has also examined sentiment analysis related to national elections and the prediction of election outcomes using Twitter data. Studies like [4] provide a general overview of the process for predicting election results, although without specific data. Additionally, [5], [6] compare election prediction results in the US and the UK, while [1] focuses on election data from Turkey. Further, research by [7], [8], [9] has looked into identifying influential users in information dissemination, using varied data contexts such as rumor spread about floods in Indonesia [8] data from kcore-analytics.com [9], and random data [5].

However, there has been no specific research aimed at detecting influential users in disseminating information regarding the 2024 National Election in Indonesia. This study addresses this gap by focusing on identifying influential Twitter users involved in spreading information about the 2024 National Election, thus helping the public access accurate and reliable information sources. This study aims to identify social media users disseminating the 2024 National Election information. This identification process can then become a filter for the public to find accurate sources of information—data related to the 2024 National Election by crawling from Twitter. Furthermore, the data is processed to produce the required data format to identify the top-k users who disseminate information related to the 2024 National Election. Based on the results of the top-k users, it can determine who influences sharing information about the 2024 National Election.

Methodology

The data used in this study was data crawled from Twitter from 2023. Twitter data that is crawled is Twitter data that contains #Pemilu2024 in the tweet and only uses Indonesian. Furthermore, the attributes that match the needs were selected from the data: user screen name, tweet, and Retweet-Count. Then, data processing was continued with tweet processing. This tweet processing stage removed unnecessary symbols such as #, URLs, and emoticons so that the tweet data only consisted of strings. The tweet data processing stage was continued by searching for tweet data that contains Retweet interactions, which RT symbolizes as a representation of the interaction between the user who wrote the tweet and represented by the user-screen-name attribute with other users. The process was continued by separating the tweet's content with an RT that mentions the user-screen-name of another user. The final process was to delete user interactions that do not have a Retweet-Count or Retweet-Count = 0. This data row was deleted because interactions that do not have a Retweet-Count mean that they do not have interaction in their distribution, so they are considered not included in the information dissemination.

The diagram of tweet data cleaning, depicted in Figure 1, begins with gathering relevant tweets related to the 2024 election from Twitter using Twitter API. Subsequently, pertinent data attributes such as tweet text, date, user-screen-name, and tweet ID are selected for further analysis. Unnecessary symbols such as hashtags, URLs, and irrelevant emoticons are removed to streamline the analysis process. Retweeted content followed by another user's username is identified and separated to provide sentiment and public opinion analysis insights. User data rows lacking retweet interactions and possessing a Retweet-count of 0 are subsequently deleted, ensuring data integrity. Once all cleaning steps were completed, the refined data was stored in an appropriate format for subsequent analysis. This meticulous tweet data cleaning process was essential to ensure the accuracy and reliability of the data for various analytical purposes, including sentiment analysis and tracking public opinion trends during the election period. The amount of data successfully obtained in crawling Twitter data using #Pemilu2024 totaled 43,549. However, after going through the data processing process, the final amount of data that can be used in the following process is 9,303 data. This data is Twitter user interaction data related to the 2024 National Election.

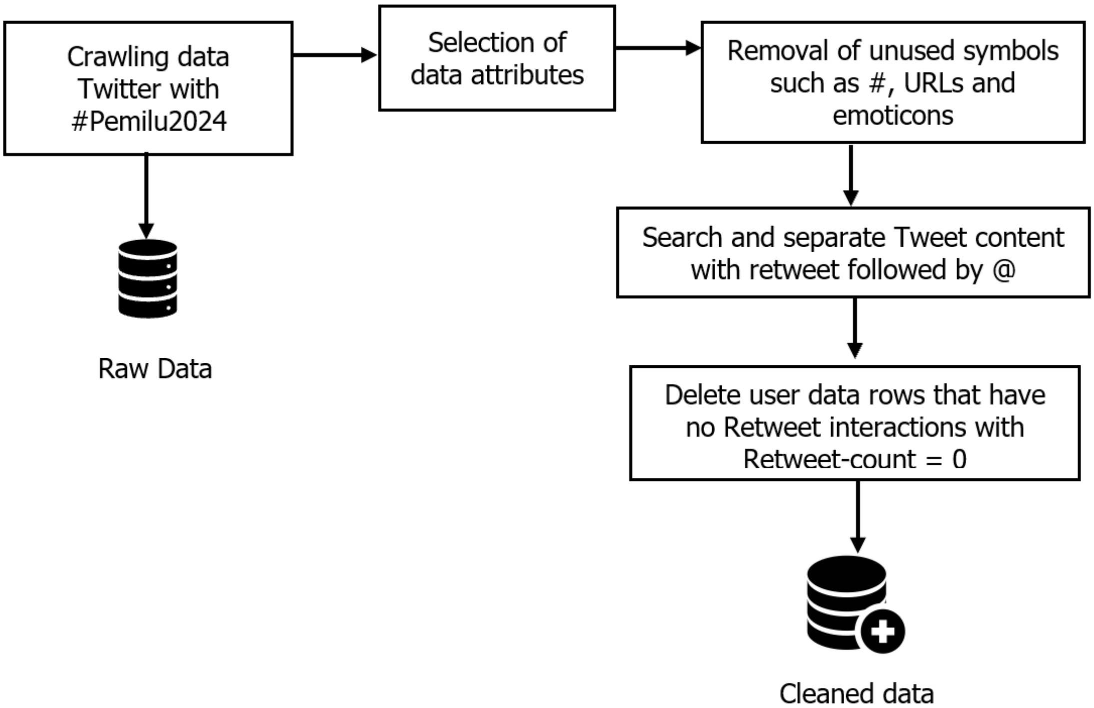

Figure 1. The steps of data processing

In addition to data, this research used several ways to identify influential users in disseminating the 2024 National Election information. These methods include Degree Centrality [10], Closeness Centrality [11], Harmonic Centrality [12], Eigenvector Centrality [13], and Load Centrality [14]. These methods are used to assess the centrality of the user on the Twitter communication network. The centrality of the user on the communication network describes the user's relationship with other users. This measure of centrality differs from one method to another. As in the Degree Centrality method, node centrality is measured by the degree that a node has [8]. Whereas Closeness Centrality looks at the centrality of a node on a network, so to calculate Closeness Centrality, we use Equation 1 [9].

$$CC(u) = \frac{1}{\sum_v d(u,v)} \tag{1}$$

Where $CC(u)$ is the closeness centrality value from the node u, meanwhile $d(u,v)$ is the shortest path between nodes u and v.

Meanwhile, one of the weaknesses of the Closeness centrality method is the inability of the way to calculate the centrality of nodes that are not connected. So, based on this, the Harmonic Centrality method was developed [15]. The equations used in calculating closeness centrality and harmonic centrality are not much different. The difference lies in the addition of condition v≠u. Thus, Harmonic centrality can be calculated as in Equation 2.

$$HC(u) = \frac{1}{\sum_{v \neq u} d(u,v)} \tag{2}$$

Where $HC(u)$ is the harmonic centrality value from the node u.

Furthermore, Eigenvector Centrality determines the centrality of a node by looking at the centrality of that node based on the centrality of its neighbors [13]. Another method that can be used to assess the centrality of a node is the Load Centrality method. The method was first developed by [16] and is known as equal betweenness because, in the process, each node sends the same amount of commodity to every other node, and then from a source, the commodity is sent to the closest node to the target. So that all the total commodities passed by a node during all exchanges are defined as the node's load [14].

An evaluation of the 2024 National Election was carried out by measuring several evaluation criteria. The test criteria include the Jaccard Similarity Index and Kendall's Rank Correlation Coefficient. The Jaccard Similarity Index is used to measure the similarity of the top-k list of influential users from the data used based on each method used. The Jaccard Similarity Index value ranges from 0 to 1 with a value close to 1, meaning that the two lists have a higher similarity. Jaccard Similarity Index can be calculated by Equation 3.

$$J(A,B) = \frac{|A \cap B|}{|A \cup B|} \tag{3}$$

The following evaluation criterion used is Kendall's Rank Correlation Coefficient. Kendall's Rank Correlation Coefficient measures the association between two sets of rank lists. Measurements with Kendall's Rank Correlation Coefficient were carried out to see the similarity of the ranking order of the existing data.

Results and Discussions

Results

Tests on the 2024 National Election data produced a list of top-ranking users on Twitter that are influential in disseminating information about the 2024 National Election in Indonesia. All data users were assessed for their centrality in sharing information regarding the 2024 National Election. Furthermore, the list results are sorted from most to least influential users. The intended user in this study is the user-screen-name used by the user on Twitter. The test was also carried out with each method used in this study. Table 1 shows an example of the ranking results of the ten most influential users or the so-called Top-10 users of each technique.

The results of tests in identifying influential users in disseminating information about the 2024 National Election vary from one method to another. For example, a Twitter user with user-screen-name rmp_officials is ranked first as the most influential user in disseminating information from all the centrality evaluation methods. Whereas in the second rank, the user-scree-name of the user that appears is different from one another. For example, the FixedPuan user-screen-name ranks second for evaluation with the Degree Centrality and Harmonic Centrality methods, while for the Closeness Centrality and Eigenvector Centrality methods, and the Load Centrality method, Ganjaran_app's user-screen-name ranks second. Users who occupy the ranking order of each technique can differ.

Based on the user ranking list generated from each of these methods, calculations are then carried out to measure the similarity and association between the rankings of each user. The similarity of the user ranking lists as measured using the Jaccard Similarity Index displays a comparison of the similarity of the ranking lists from a combination of one method with another. For example, the user ranking list from the Degree Centrality method is compared to the Closeness Centrality, Harmonic Centrality, Eigenvector Centrality, and Load Centrality methods. This process is carried out for each ranking list result. The test results using the Jaccard Similarity Index can be seen in Figure 2.

Table 1. Top 10 influential users of each method

Rank	Degree Centrality	Closeness Centrality	Harmonic Centrality	Eigenvector Centrality	Load Centrality
1	rmp_officials	rmp_officials	rmp_officials	rmp_officials	rmp_officials
2	TetapPuan	DPC PKB KOTA BATU	TetapPuan	DPC PKB KOTA BATU	Ganjaran_app
3	Ganjaran_app	Ricky Habibulloh	Ganjaran_app	Ricky Habibulloh	TetapPuan
4	GanjaranApp	stepanus subay	DPC PKB KOTA BATU	stepanus subay	catatanganjar
5	KPU_ID	Babul Fatih	Ricky Habibulloh	Babul Fatih	sangiranflying
6	muhammadiyah	Khairuddin	stepanus subay	Khairuddin	SERATUS_P
7	golkarpedia	Hendra ningrat Al bantani	Babul Fatih	Hendra ningrat Al bantani	DPP_PPP
8	maryshelparaiso	Asmi	Khairuddin	Petrus Riwoe	Up Ganjar Pranowo 2024
9	mypresidentid	zonapkbblitar	Hendra ningrat Al bantani	Rudolf Blasin	Diwangga Diwangga
10	Tempat Berita	Petrus Riwoe	Asmi	Asmi	GanjaranApp

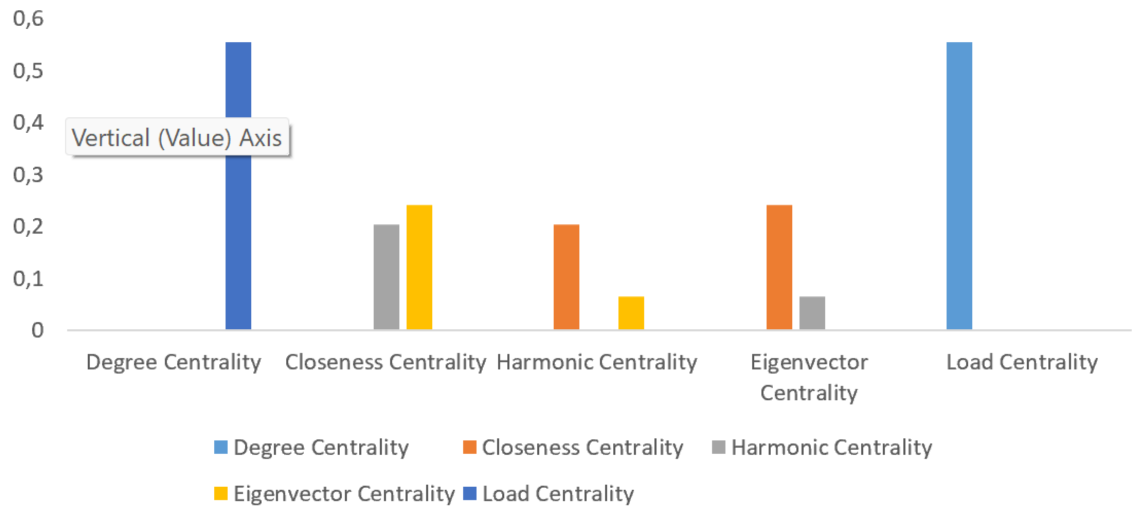

Figure 2. Jaccard Similarity Index

In the results shown in Figure 2 related to the similarity of the top user ranking list in disseminating information about the 2024 National Election, it is known that the complete ranking list produced by the Degree Centrality method has the highest similarity with the top ranking list produced by the Load Centrality method. The similarity value is 0.555. The highest value from the calculation of the Jaccard Similarity Index is bolded in Figure 2. In comparing top-ranking lists from other methods, the resulting similarity values range from 0.001 to 0.242. Furthermore, in the results of the Jaccard Similarity Index, several method comparisons have a value of 1. This value is generated because the method is compared with the technique itself, so this value is not considered in the evaluation related to the similarity of the resulting top list rank.

Furthermore, evaluation was also carried out using Kendall's Rank Correlation Coefficient to measure the correlation between the resulting top-ranking lists. Same as in evaluation using the Jaccard Similarity Index. The top ranking list of each method is compared to one another for evaluation using Kendall's Rank Correlation Coefficient. Evaluation results using Kendall's Rank Correlation Coefficient can be seen in Figure 3.

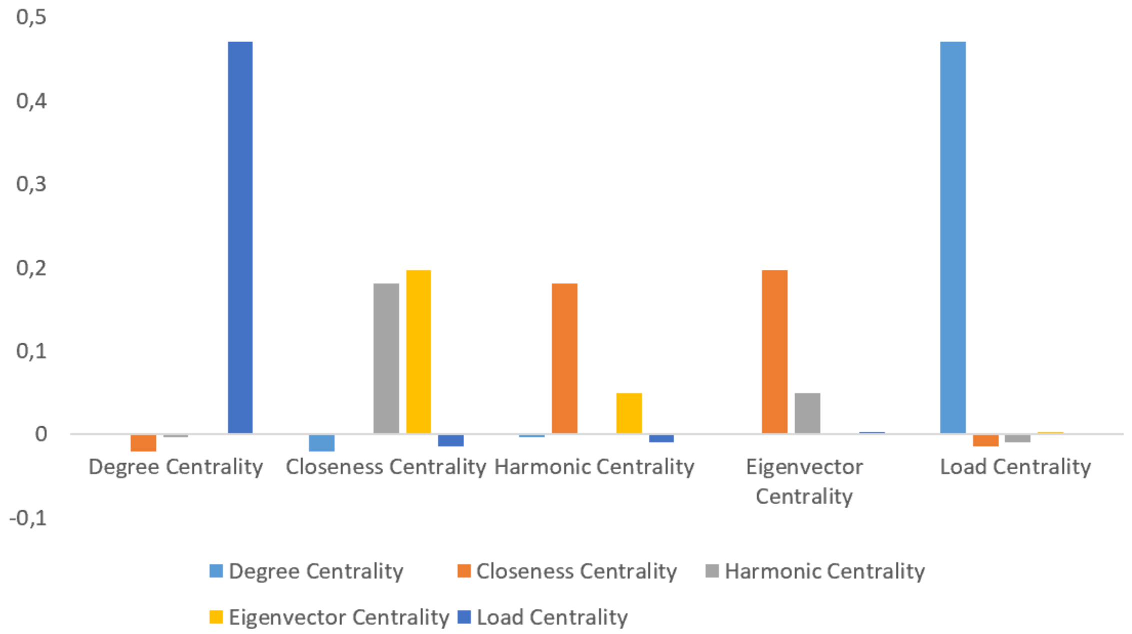

Figure 3. Kendall's Rank Correlation Coefficient

Evaluations carried out using Kendall's Rank Correlation Coefficient show the correlation of one method with the other techniques used in this study. The evaluation results using Kendall's Rank Correlation Coefficient show that the top user ranking list generated by the Degree Centrality method has the highest correlation with the complete user ranking list generated by the Load Centrality method. The resulting Kendall's Rank Correlation Coefficient is 0.471, the highest compared to other values ranging from -0.003 to 0.197. There is also a negative correlation in the evaluation of Kendall's Rank Correlation Coefficient, which means that the resulting top user ranking lists are very dissimilar. In Figure 3, the highest Kendall's Rank Correlation Coefficient is bolded.

Discussions

The evaluation results of the 2024 General Election information dissemination data on Twitter show various results. The results of implementing the method used in this study show that the ranking of influential users in disseminating information about the 2024 General Election differs from one way to another because each method uses a different concept in calculating the centrality of a user in the network, such as degree centrality which uses the degree of user connectivity, closeness centrality which measures the centrality of a node on the network and other methods.

Furthermore, the evaluation to measure the similarity of the top-ranking lists generated using the Jaccard Similarity Index shows that the Degree Centrality and Load Centrality methods have the highest similarity compared to other methods. This similarity value is 0.555. These results are in line with research conducted by Borgatti. Borgatti shows that the variation between degree centrality and load centrality at each node in a social network is relatively small, so it can be concluded that the centrality results of the two have high similarity.

The results of evaluating the similarity of the top-ranking lists using the Jaccard Similarity Index are also supported by the evaluation results with Kendall's Rank Correlation Coefficient. Evaluation using Kendall's Rank Correlation Coefficient also shows results consistent with the Jaccard Similarity Index; the top ranking list correlation of the Degree Centrality and Load Centrality methods is the highest of the top ranking list correlation combinations of other ways.

Conclusion

Research on identifying influential Twitter users in disseminating information about the 2024 General Election has been carried out in this study. This study has also used several methods of measuring user centrality in information dissemination networks. The test results of each method produce a top-ranking list that differs from one method to another. These different results are due to the concept of calculation methods that are different from one another. The evaluation showed that the Degree centrality and Load centrality methods had the most similar top-ranking lists among the methods used in this study. However, these results may change if other data and methods are used in similar studies.

Further research development can be done by adding the methods used in this research because several ways are used to measure centrality that requires modification to calculate the data used in this study. In addition, further analysis can also be carried out to determine whether Twitter users are indeed humans or bots. The research needs to be done because many information found during the preprocessing process results from bots.

References

[1] S. Network, A. Overview, T. Political, C. Competition, R. Sosyal, and B. Bak, "A Social Network Analysis Overview of Turkish Political Communication Competition on Twitter," *Çalışır, Vahit*, vol. 11, no. 1, pp. 181–197, 2023.

[2] N. Gupta, H. Crosby, D. Purser, S. Javis, and W. Guo, "Twitter usage across the industry: A spatiotemporal analysis," in *Proceedings - IEEE 4th International Conference on Big Data Computing Service and Applications, BigDataService 2018*, 2018, pp. 64–71, doi: 10.1109/BigDataService.2018.00018.

[3] C. Siahaan and M. Tampubolon, "Electoral Manipulation in Indonesia's 2019 National Election," *J. Posit. Psychol. &Wellbeing*, vol. 5, no. 4, pp. 1943–1955, 2021.

[4] K. Myilvahanan, P. Yashas, S. Pasha, M. Ismail, and V. Tharun, "A Study on Election Prediction using Machine Learning Techniques," *Proc. 3rd Int. Conf. Artif. Intell. Smart Energy, ICAIS 2023*, no. Icais, pp. 1518–1520, 2023, doi: 10.1109/ICAIS56108.2023.10073693.

[5] M. H. Tsai, Y. Wang, M. Kwak, and N. Rigole, "A machine learning based strategy for election result prediction," in *Proceedings - 6th Annual Conference on Computational Science and Computational Intelligence, CSCI 2019*, 2019, pp. 1408–1410, doi: 10.1109/CSCI49370.2019.00263.

[6] A. E. Murr, S. Munzert, and A. Murr, "Forecasting National Turnout at British General Elections: Combining Polls and Structural Models," *Polit. Sci. Econ.*, 2018.

[7] F. Morone, B. Min, L. Bo, R. Mari, and A. Makse, "Collective Influence Algorithm to Find Influencers via Optimal Percolation in Massively Large Social Media," *Sci. Rep.*, vol. 6, pp. 1–22, 2016.

[8] F. K. Dewi, S. B. Yudhoatmojo, and I. Budi, "Identification of Opinion Leader on Rumor Spreading in Online Social Network Twitter Using Edge Weighting and Centrality Measure Weighting," in *The Twelfth Interrnational Conference on Digital Information Management*, 2017, pp. 243–248, doi: 10.1109/ICACSIS.2013.6761583.

[9] X. Teng, S. Pei, F. Morone, and H. A. Makse, "Collective Influence of Multiple Spreaders Evaluated by Tracing Real Information Flow in Large-Scale Social Networks," *Sci. Rep.*, vol. 6, no. October, 2016, doi: 10.1038/srep36043.

[10] J. Zhang and Y. Luo, "Degree Centrality, Betweenness Centrality, and Closeness Centrality in Social Network," in *2nd International Conference on Modelling, Simulation and Applied Mathematics (MSAM 2017)*, 2017, vol. 132, pp. 300–304, doi: 10.2991/msam-17.2017.68.

[11] J. Guan, Y. Li, L. Xing, Y. Li, and G. Liang, "Closeness centrality for similarity-weight network and its application to measuring industrial sectors' position on the Global Value Chain," *Pyhysica A*, p. 103778, 2019, [Online]. Available: https://doi.org/10.1016/j.compedu.2019.103778.

[12] J. M. E. Ortega and R. G. Eballe, "Harmonic Centrality and Centralization of Some Graph Products," *Asian Res. J. Math.*, no. May, pp. 42–51, 2022, doi: 10.9734/arjom/2022/v18i530377.

[13] L. Lorenzini *et al.*, "Eigenvector centrality dynamics are related to Alzheimer's disease pathological changes in non-demented individuals," *Brain Commun.*, vol. 5, no. 3, pp. 1–13, 2023, doi: 10.1093/braincomms/fcad088.
[14] L. Maccari, L. Ghiro, A. Guerrieri, A. Montresor, and R. Lo Cigno, "Exact Distributed Load Centrality Computation: Algorithms, Convergence, and Applications to Distance Vector Routing," *IEEE Trans. Parallel Distrib. Syst.*, vol. 31, no. 7, pp. 1693–1706, 2020, doi: 10.1109/TPDS.2020.2973960.
[15] P. Boldi and S. Vigna, "Axioms for centrality," *Internet Math.*, vol. 10, no. 3–4, pp. 222–262, 2014, doi: 10.1080/15427951.2013.865686.
[16] K. I. Goh, B. Kahng, and D. Kim, "Universal Behavior of Load Distribution in Scale-Free Networks," *Phys. Rev. Lett.*, vol. 87, no. 27, pp. 278701-278701–4, 2001, doi: 10.1103/PhysRevLett.87.278701.