

Hashtag Analysis of Indonesian COVID-19 Tweets Using Social Network Analysis

Muhammad Habibi*¹, Adri Priadana², Muhammad Rifqi Ma'arif³

¹Department of Informatics, FTTL, Universitas Jenderal Achmad Yani Yogyakarta, Indonesia

^{2,3}Center of Data Analytic Research and Services, Universitas Jenderal Achmad Yani Yogyakarta, Indonesia

e-mail: *¹habibi@unjaya.ac.id, ²dana@unjaya.ac.id, ³rifqi@unjaya.ac.id

Abstrak

Media sosial saat ini menjadi wadah yang paling banyak digunakan oleh masyarakat untuk mengungkapkan persepsi maupun pendapatnya terhadap isu-isu terkini isu-isu mengenai pandemi COVID-19. Dalam unggahan pesan di media sosial, hashtag merupakan anotasi sosial yang digunakan untuk menunjukkan konten pesan. Hashtag berfungsi sebagai alat yang intuitif dan fleksibel untuk membuat koleksi terbitan yang sangat besar sehingga dapat digunakan sebagai dasar pencarian pada Twitter. Melalui penggunaan beberapa hashtag pada suatu terbitan membuat masing-masing hashtag memiliki hubungan. Penelitian ini bertujuan untuk menganalisis hashtag Tweet COVID-19 di Indonesia menggunakan Social Network Analysis (SNA). Kami menggunakan teknik SNA untuk memvisualisasikan model jaringan dan mengukur beberapa centrality untuk menemukan hashtag yang paling berpengaruh di jaringan. Kami mengumpulkan dan menganalisis 500.000 tweet publik dari Twitter berdasarkan kata kunci COVID-19. Berdasarkan hasil pengukuran centrality, hashtag #corona merupakan hashtag yang paling banyak memiliki keterkaitan dengan hashtag lainnya. Hashtag #COVID19 adalah hashtag yang paling erat kaitannya dengan semua hashtag lainnya. Hashtag #corona merupakan hashtag yang paling berperan sebagai jembatan yang dapat mengontrol arus informasi terkait COVID-19. Hashtag #coronavirus adalah hashtag terpenting berdasarkan ketetanggaannya. Hashtag #covid19 dan #wabah memiliki hubungan yang substansial dengan hashtag keagamaan berdasarkan visualisasi jaringan.

Kata kunci—COVID-19, Twitter, Social Network Analysis, SNA, Hashtag

Abstract

Social media has become more critical for people to communicate about the pandemic of COVID-19. In social media, hashtags are social annotations which often used to denote message content. It serves as an intuitive and flexible tool for making huge collections of posts searchable on Twitter. Through practices of hashtagging, user representations of a given post also become connected. This study aimed to analyze the hashtag of Indonesian COVID-19 Tweets using Social Network Analysis (SNA). We used SNA techniques to visualize network models and measure some centrality to find the most influential hashtag in the network. We collected and analyzed 500.000 public tweets from Twitter based on COVID-19 keywords. Based on the centrality measurement result, the hashtag #corona is a hashtag with the most connection with other hashtags. The hashtag #COVID19 is the hashtag that is most closely related to all other hashtags. The hashtag #corona is the hashtag that most acts as a bridge that can control the flow of information related to COVID-19. The hashtag #coronavirus is the most important of hashtags based on their link. Our study also found that the hashtag #covid19 and #wabah have a substantial relationship with religious-related hashtags based on network visualization.

Keywords—COVID-19, Twitter, Social Network Analysis, SNA, Hashtag

1. INTRODUCTION

Social media has been growing to become an essential part of people's lives in recent years. Several big social media platforms like Facebook, Instagram and Twitter have become more critical for people to communicate almost of everything with no exception to pandemic of COVID-19 related issues [1]. Those make how much information we receive through this platform influences how we perceive and deal with the current COVID-19 epidemic. Even before the outbreak, the public and scientists used social media as an essential information source. The efficient use of social media also affects the character and communication way of modern society [2]. One kind of social media that is commonly used today is Twitter [3]. Previous research has noted the potential for Twitter to provide real-time content analysis so that public health authorities can quickly respond to anxieties grown by the public [4].

Hashtags, which are social annotations used to denote message content, serve as an intuitive and flexible tool for making huge collections of posts searchable on Twitter. The inherent ability of hashtags to bypass the network structure's boundaries during the communication process gives the easy dissemination of information outside the user's network. Hashtags take a double role in the communication process, functioning both as metadata for archiving purposes and information retrieval on social media platforms [5], such as Twitter. Hashtags written with the symbol # are even used on other social media platforms to index keywords or topics [6], [7].

Several previous studies have carried out hashtag analysis on social media for various purposes. Xing et al., in 2016 [8], analyzed hashtags to find sub-shows on Twitter. They model the relationship between the hashtag and the tweet's topic and highlight the hashtag's role as a semantic representation of the appropriate tweet. In 2016 [9], Yılmaz and Hero studied to detect a multimodal event in Twitter hashtag networks. They introduced a new unsupervised event detection approach to Twitter.

Through practices of hashtagging, user representations of a given post also become connected [10]. Simultaneously, we can define the relationship between hashtags as an organization within a topic area, especially related to the COVID-19 outbreak. In this study, we aimed to analyze the hashtag of Indonesian COVID-19 Tweets using Social Network Analysis (SNA). We used SNA techniques to visualize network models and measure some centrality to find the most influential hashtag in the network.

Several previous studies have applied SNA for various purposes. Iswandhani and Muhajir, in 2017 [11], used SNA to identify the central popularity of tourist destinations in an Instagram account. In 2017 [12], Setatama and Tricahyono implemented Social Network Analysis to analyze the most influential actors in the spread of the country branding "Wonderful Indonesia." Tahalea and Azhari, in 2019 [13], used SNA to identify the central actor of crimes done by several people using five centrality measurements, such as degree, betweenness, closeness, and eigenvector centrality. This study showed 80.39% accuracy from 102 criminal cases gathered with at least three actors involved in each case.

In the other study, Hung et al., in 2020 [14], used SNA to determine the social network of dominant topics related to COVID-19 on Twitter. The study successfully identified five dominant issues related to COVID-19. They are social change, health care environment, business economy, emotional support, and psychological stress. In 2020 [15], Tahalea used SNA to analyze the relationships between heroes and find out the roles of heroes in DOTA2 online games. He used the measurement of degree centrality to see the popularity of a hero, betweenness centrality to see the role as a liaison, and closeness centrality to see a hero's closeness to other heroes.

2. METHODS

The social network analysis process in this study consists of three stages. They are data extraction and preprocessing, building a network model, and measuring centrality value.

2.1 Data Extraction and Preprocessing

This study implemented a web scraping technique to extract tweet data related to COVID-19 from Twitter web-based. Online data extraction refers to the routine extraction of data from a web data source that can evolve [16]. After we collected the tweet data, the next step is preprocessing. In this step, we pulled only the hashtag from every tweet data.

2.2 Network Model

The second stage of this experiment is builds a network model. In the network model hashtags are represented as nodes, and the connections between them are represented as edges. When a user publishes more than one hashtags in a single tweet, the connection between hashtag happens. For example, if user post a tweet contains hashtag #Covid19, #wabah, and #lockdown, then there will be a connection between #Covid10 and #wabah, #Covid19 and lockdown, and also #wabah and #lockdown. **Figure 1** shows the network model representation of this case.

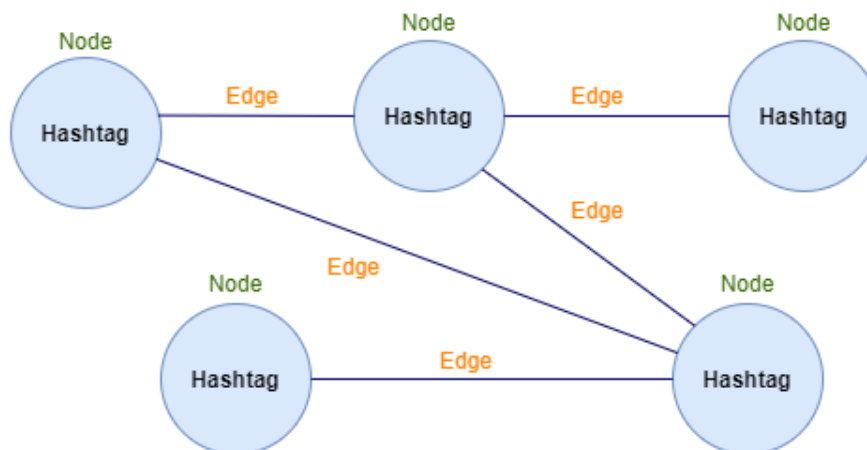


Figure 1 The network model representation

2.3 Centrality Measure

This study used five centrality measurements, such as degree, betweenness, closeness, and eigenvector centrality, to analyze the Indonesian COVID-19 Tweets' hashtag. In some fields such as psychological networks [17], social networks [18], criminal networks [13], and even trust networks [19], these centrality measures can determine the network's central or highly influential nodes.

2.3.1 Degree Centrality

The degree of centrality is determined by the total quantity of direct connections to other nodes in a network graph [20]. The node degree centrality is a key parameter representing the community centrality in networks [21]. We used degree centrality to identify the most connected hashtags in a tweet that can be measured using equation (1),

$$C_D(i) = \sum \frac{d_i}{n-1} \quad (1)$$

where hashtags represent as i and total nodes (hashtags) in the network represent as n .

2.3.2 Betweenness Centrality

Betweenness centrality is to measure the role of a node as a mediator in the network. We use this centrality to show how significant a hashtag is to act as a bridge in the network, which can be measured using **equation (2)**. In this case, the hashtags represent as i , the number of shortest paths from actor j to actor k shown as $g_{j,k}$, and the number of shortest paths from actor j to actor k through actor i shown as $g_{j,k}(i)$.

$$C_B(i) = \sum \frac{g_{j,k}(i)}{g_{j,k}} \quad (2)$$

2.3.3 Closeness Centrality

Closeness centrality measures one node to the other nodes' sum distances by measuring the average distance of a node from other nodes in the network [22]. We use this centrality to show the closeness of the connection between hashtags, which can be measured using equation (3). $C_C(i)$ is the closeness centrality of node i , and D_{ij} is the shortest path from node i to node j .

$$C_C(i) = \frac{n-1}{\sum D_{ij}} \quad (3)$$

2.3.4 Eigenvector Centrality

The eigenvector centrality measures the number of connections of a given node and its relevance in information movement [23]. We used eigenvector centrality to shows the importance of hashtags based on their link, which can be measured using equation (4). The hashtags represent as i , constant represents as λ , and $a_{i,j}$ is shown adjacency matrix of the network.

$$C_e(i) = \frac{1}{\lambda} \sum_{j \in G} a_{i,j} \quad (4)$$

3. RESULTS AND DISCUSSION

In this section, we discussed the results of the calculations. For the first time, we did data profiling from the data extraction. After that, we found the most influential hashtags that play an essential role in disseminating COVID-19 information on Twitter using SNA.

3.1 Data Profiling

We have collected 500000 public tweets from Twitter based on COVID-19 keywords. We gathered from March to May 2020, i.e., three months. We selected those months since President Joko Widodo reported the first confirmed two cases of COVID-19 infection in March 2020, in Indonesia [24]. Then, the data collection was closed in May 2020, when this research analysis begins.

3.2 Centrality Measure

At this stage, we will discuss the centrality measure of Twitter data about COVID-19. Centrality measures that are calculated include degree centrality, closeness centrality,

betweenness centrality, and eigenvector centrality. Degree centrality is a simple measure of centrality which calculates how many connections or neighbors a node has [25]. The hashtag regarding COVID-19, which has the most number of neighbors, can be seen in Table 1.

Table 1 Degree Centrality

No	Hashtag	Degree
1	#corona	0,0894
2	#Corona	0,0717
3	#VirusCorona	0,0660
4	#COVID19	0,0655
5	#coronavirus	0,0634
6	#dirumahaja	0,0608
7	#covid19	0,0597
8	#RememberingKhilafah	0,0430
9	#Covid19	0,0409
10	#viruscorona	0,0363

Based on Table 1, it is known that people often use ten hashtags in uploading tweets on Twitter. These ten hashtags are important because they have the highest number of connections to other hashtags. The 10 hashtags include #corona, #Corona, #VirusCorona, # COVID19, #coronavirus, #dirumahaja, # covid19, #RememberingKhilafah, # Covid19, and #viruscorona. The use of lowercase and capital letters in a hashtag is a sensitive matter because the hashtag #corona with a lowercase and hashtag #Corona with capital is considered different.

The next discussion is about closeness centrality. Closeness centrality indicates how close a hashtag is to all other hashtags in the network. The closeness centrality data of hashtags can be seen in Table 2. The 10 hashtags that have the most closeness to all hashtags in the network are # COVID19, #dirumahaja, #Corona, #coronavirus, #corona, #VirusCorona, # Covid_19, # Covid19, #Indonesia , and #COVID. To get the maximum speed of information flow related to COVID-19 data, you can use these hashtags.

Table 2 Closeness Centrality

No	Hashtag	Closeness
1	#COVID19	0,3519
2	#dirumahaja	0,3507
3	#Corona	0,3487
4	#coronavirus	0,3480
5	#corona	0,3433
6	#VirusCorona	0,3411
7	#Covid_19	0,3333
8	#Covid19	0,3279
9	#Indonesia	0,3279
10	#COVID	0,3260

Betweenness centrality is used to show how often other nodes pass a node to go to a particular node in the network. This value serves to determine the role of the hashtag as a bridge connecting interactions in the network. The hashtag data that has the highest betweenness centrality value can be seen in Table 3. The 10 hashtags are #corona, #VirusCorona, #dirumahaja, # COVID19, #Corona, #coronavirus, #Corona, # covid19, #dirumahaja, and # Covid19. These hashtags are located in communication channels and can control the flow of information related to COVID-19 data on Twitter.

Table 3 Betweenness Centrality

No	Hashtag	Betweenness
1	#corona	0,0785
2	#VirusCorona	0,0681
3	#dirumahaja	0,0678
4	#COVID19	0,0622
5	#Corona	0,0615
6	#coronavirus	0,0513
7	#Corona	0,0502
8	#covid19	0,0346
9	#dirumahaja	0,0341
10	#Covid19	0,0305

Eigenvector Centrality is used to measure a node's importance by considering the importance of its neighbors [26]. The hashtag data that has the highest eigenvector centrality value can be seen in Table 4. The 10 hashtags are #coronavirus, #corona, #Corona, # COVID19, #dirumahaja, #VirusCorona, # covid19, # Covid_19, #virus, and #COVID.

Table 4 Eigenvector Centrality

No	Hashtag	Eigenvector
1	#coronavirus	0,2204
2	#corona	0,2137
3	#Corona	0,2095
4	#COVID19	0,2025
5	#dirumahaja	0,1785
6	#VirusCorona	0,1674
7	#covid19	0,1613
8	#Covid_19	0,1315
9	#virus	0,1307
10	#COVID	0,1231

In this section, the data visualization of the information dissemination network interaction on COVID-19 is carried out on the Twitter social networking site. The visualization of social network analysis can be seen in Figure 2. Visualization of the interaction network is made up of 12,906 nodes and 50,349 edges. From the visualization of the network of interactions, it is known that the network interaction patterns of information dissemination related to COVID-19 are strongly influenced by the hashtags #corona, #Corona, #VirusCorona, # COVID19, #coronavirus, and #dirumahaja. The central hashtag for the spread of COVID-19 information is the hashtag #corona.

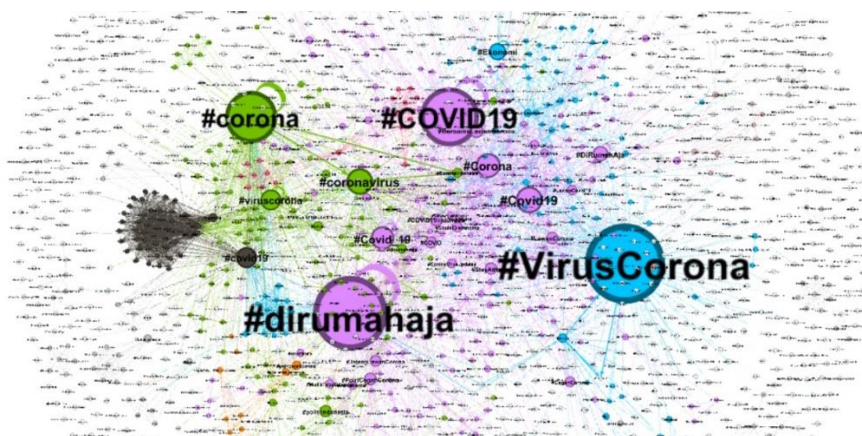


Figure 2 The network model representation

In our experiment we found some interesting fact that the the network interaction with the dissemination of information about COVID-19 on the social networking site Twitter has strong correlation with the hashtag form discussions related to religious issues. This can be seen in Figure 3, and in detail, can be seen in Figure 4. Based on the network visualization, it can be seen that the hashtag #covid19 and #wabah have a substantial relationship with religious hashtags. These hashtags include #islam, #muslim, #tauhid, #nikah, #kajianislam, #sunnah, #hijrah, #dakwah, and #ramadhan.

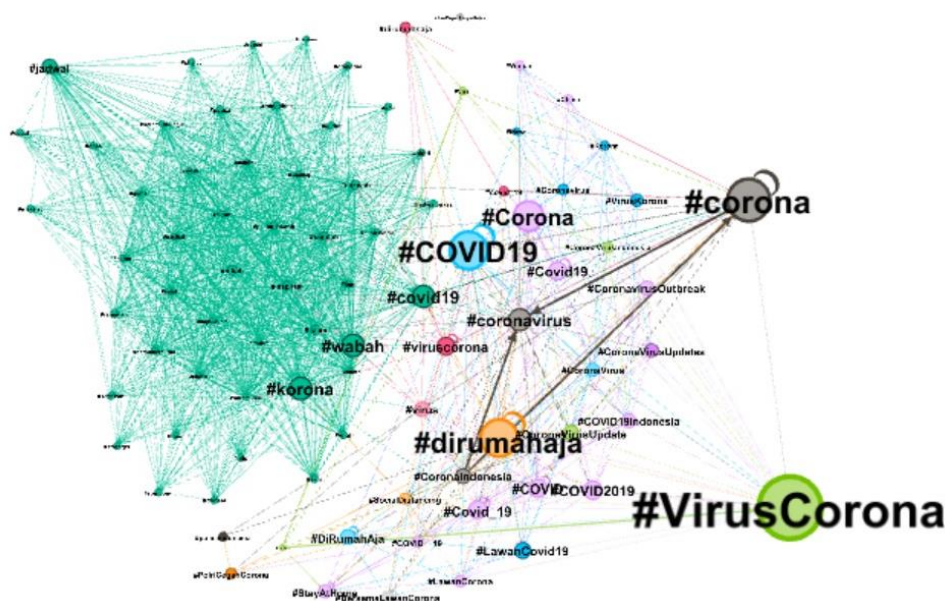


Figure 3 The network model representation

We reveal this network structure as represented in figure 4 by configuring the node size and network layout based on eigenvector centrality. Referring to the philosophy of eigenvector calculation that the nodes which connect to popular nodes will becoming popular, we can assume that religious post on Twitter take the benefits of #covid19 hashtag popularity to increase the visibility of their posts. This is the common practice found in social media marketing.

As we knew, Social Network Analysis is usually used to determine the influence of actors in social media as in studies [27], [28] and most previous studies. According to this study result, we could know that Social Network Analysis also can be used to determine the influence of hashtags in social media. Moreover, some studies used Social Network Analysis based on a particular hashtag, but it was only used to define the topic as in studies [29], [30]. Eventually, the Social Network Analysis is used to determine the influence of actors. In this study, we used Social Network Analysis to determine the influence of hashtags on the other hashtags or the connection between hashtags as in the study [31]. However, the study was implemented on the Instagram platform, whereas this study was implemented on the Twitter platform. With the approach defined in this study, we can understand which hashtag has an essential role in disseminating information on Twitter. Therefore, we can control the stream of information dissemination, mainly if it negatively influences news.

- [2] P. W. Cahyo and M. Habibi, "Entity Profiling to Identify Actor Involvement in Topics of Social Media Content," *IJCCS (Indonesian J. Comput. Cybern. Syst.*, vol. 14, no. 4, Oct. 2020.
- [3] Y. Yunitasari, A. Musdholifah, and A. K. Sari, "Sarcasm Detection For Sentiment Analysis in Indonesian Tweets," *IJCCS (Indonesian J. Comput. Cybern. Syst.*, vol. 13, no. 1, p. 53, Jan. 2019.
- [4] W. Ahmed, J. Vidal-Alaball, J. Downing, and F. L. Seguí, "COVID-19 and the 5G conspiracy theory: Social network analysis of twitter data," *J. Med. Internet Res.*, vol. 22, no. 5, p. e19458, May 2020.
- [5] K. G. Kapanova and S. Fidanova, "Generalized nets: A new approach to model a hashtag linguistic network on Twitter," in *Studies in Computational Intelligence*, vol. 793, Springer Verlag, 2019, pp. 211–221.
- [6] A. Priadana and M. Habibi, "Face Detection using Haar Cascades to Filter Selfie Face Image on Instagram," in *2019 International Conference of Artificial Intelligence and Information Technology (ICAIT)*, 2019, pp. 6–9.
- [7] M. Habibi and P. W. Cahyo, "Clustering User Characteristics Based on the influence of Hashtags on the Instagram Platform," *IJCCS (Indonesian J. Comput. Cybern. Syst.*, vol. 13, no. 4, pp. 399–408, 2019.
- [8] C. Xing, Y. Wang, J. Liu, Y. Huang, and W. Ma, "Hashtag-Based Sub-Event Discovery Using Mutually Generative LDA in Twitter," *undefined*, 2016.
- [9] Y. Yılmaz and A. O. Hero, "Multimodal Event Detection in Twitter Hashtag Networks," *J. Signal Process. Syst.*, vol. 90, no. 2, pp. 185–200, Feb. 2018.
- [10] T. Highfield and T. Leaver, "A methodology for mapping instagram hashtags," *First Monday*, vol. 20, no. 1, 2015.
- [11] N. Iswandhani and M. Muhajir, "K-means cluster analysis of tourist destination in special region of Yogyakarta using spatial approach and social network analysis (a case study: Post of @explorejogja instagram account in 2016)," in *Journal of Physics: Conference Series*, 2018, vol. 974, no. 1, p. 12033.
- [12] M. S. Setatama and D. Tricahyono, "Implementasi Social Network Analysis dalam Penyebaran Country Branding 'Wonderful Indonesia,'" *Indones. J. Comput.*, vol. 2, no. 2, pp. 91–104, 2017.
- [13] S. P. Tahalea and A. SN, "Central Actor Identification of Crime Group using Semantic Social Network Analysis," *Indones. J. Inf. Syst.*, vol. 2, no. 1, p. 24, Aug. 2019.
- [14] M. Hung *et al.*, "Social network analysis of COVID-19 sentiments: Application of artificial intelligence," *J. Med. Internet Res.*, vol. 22, no. 8, p. e22590, Aug. 2020.
- [15] S. Prian Tahalea, "Identifikasi Peran Hero DOTA2 Menggunakan Social Network Analysis," May 2020.
- [16] E. Ferrara, P. De Meo, G. Fiumara, and R. Baumgartner, "Web data extraction, applications and techniques: A survey," *Knowledge-Based Syst.*, vol. 70, pp. 301–323, Nov. 2014.
- [17] L. F. Bringmann *et al.*, "What Do Centrality Measures Measure in Psychological Networks?," *J. Abnorm. Psychol.*, 2019.
- [18] A. K. Yadav, R. Johari, and R. Dahiya, "Identification of Centrality Measures in Social Network using Network Science," in *Proceedings - 2019 International Conference on Computing, Communication, and Intelligent Systems, ICCIS 2019*, 2019, vol. 2019-Janua, pp. 229–234.
- [19] P. De Meo, K. Musial-Gabrys, D. Rosaci, G. M. L. Sarne, and L. Aroyo, "Using centrality measures to predict helpfulness-based reputation in trust networks," *ACM Trans. Internet Technol.*, vol. 17, no. 1, pp. 1–20, Feb. 2017.
- [20] J. Zhang and Y. Luo, "Degree Centrality, Betweenness Centrality, and Closeness Centrality in Social Network," 2017, pp. 300–303.
- [21] Cai, Zeng, Wang, Li, and Hu, "Community Detection Method Based on Node Density, Degree Centrality, and K-Means Clustering in Complex Network," *Entropy*, vol. 21, no.

- 12, p. 1145, Nov. 2019.
- [22] M. S. Setatama and D. Tricahyono, Ir., M.M., Ph.D., “Implementasi Social Network Analysis pada Penyebaran Country Branding ‘Wonderful Indonesia,’” *Indones. J. Comput.*, vol. 2, no. 2, p. 91, Nov. 2017.
- [23] C. F. A. Negre *et al.*, “Eigenvector centrality for characterization of protein allosteric pathways,” *Proc. Natl. Acad. Sci. U. S. A.*, vol. 115, no. 52, pp. E12201–E12208, Dec. 2018.
- [24] R. Djalante *et al.*, “Review and analysis of current responses to COVID-19 in Indonesia: Period of January to March 2020,” *Prog. Disaster Sci.*, vol. 6, p. 100091, Apr. 2020.
- [25] D. L. Hansen, B. Shneiderman, M. A. Smith, and I. Himelboim, “Calculating and visualizing network metrics,” in *Analyzing Social Media Networks with NodeXL*, Elsevier, 2020, pp. 79–94.
- [26] J. Golbeck, “Network Structure and Measures,” in *Analyzing the Social Web*, Elsevier, 2013, pp. 25–44.
- [27] F. C. Tsai, M. C. Hsu, C. T. Chen, and D. Y. Kao, “Exploring drug-related crimes with social network analysis,” in *Procedia Computer Science*, 2019, vol. 159, pp. 1907–1917.
- [28] D. Suárez, J. M. Díaz-Puente, and M. Bettoni, “Risks Identification and Management Related to Rural Innovation Projects through Social Networks Analysis: A Case Study in Spain,” *L. 2021, Vol. 10, Page 613*, vol. 10, no. 6, p. 613, Jun. 2021.
- [29] S. W. Wiiava and I. Handoko, “Examining a Covid-19 Twitter Hashtag Conversation in Indonesia: A Social Network Analysis Approach,” *Proc. 2021 15th Int. Conf. Ubiquitous Inf. Manag. Commun. IMCOM 2021*, Jan. 2021.
- [30] E. Carnia, B. Fermadona, H. Napitupulu, N. Anggriani, and A. K. Supriatna, “Implementation of centrality measures in graph represented information spreads with hashtag #bersatulawancorona in Twitter,” *J. Phys. Conf. Ser.*, vol. 1722, no. 1, p. 012068, Jan. 2021.
- [31] A. Priadana and S. P. Tahalea, “Hashtag activism and message frames: social network analysis of Instagram during the COVID-19 pandemic outbreak in Indonesia,” *J. Phys. Conf. Ser.*, vol. 1836, no. 1, p. 012031, Mar. 2021.