

Implementation of Social Network Analysis in the Spread of Natuna Issues on Twitter

Ashif Dzilfiqar Thayyibi^{1*}, Juliana Mansur²

¹Program Studi Magister Ilmu Komputer, Fakultas Teknologi Informasi, Universitas Budi Luhur

²Program Studi Magister Ilmu Komputer, Fakultas Teknologi Informasi, Universitas Budi Luhur

email: ¹ashifdzilfiqar@mail.com, ²julianabisnis@gmail.com

Abstract – Currently, the growth of internet users has been accompanied by the development of applications that support interaction among users, which is called social media. One of the popular social media in society today is twitter. Data on Twitter can be presented in a graph structure visualization in nodes that represent actors and edges that represent relationships between actors. In an effort to find the most influential actors and actors who interact the most in spreading the Natuna topic on social media twitter, an analysis will be carried out using the Social Network Analysis method using the Degree Centrality approach. The data used in this study were taken from December 20, 2019 at 00.00 WIB to January 7, 2020 at 10.00 WIB consisting of 71,477 nodes and 147066 edges. The results of this study can be concluded that the @susipudjiastuti account is the most influential actor and plays an important role in social networking because the @susipudjiastuti account is the most linked account with 29755 links. Meanwhile, the @shaktia704 account was the most active account during the data collection period, which reached 259 links.

Keywords – social network analysis; directed graph; degree centrality; twitter;

I. INTRODUCTION

Currently, the growth of internet users has been accompanied by the development of applications that support interaction among users, which is called social media. One of the popular social media in society today is twitter. According to a survey conducted by eMarketer, there were 22.8 million active users of social media twitter in Indonesia in 2019 [1]. Based on the survey conducted by eMarketer, it can be concluded that Twitter is a valuable source of social media data for analyzing opinions or opinions [2] and is an option in conducting research for social network analysis (social network analysis).

In Twitter, data can be visualized in a graphical structure formed by nodes that represent actors and edges that represent relationships between actors. The use of a graphical structure in the Twitter social media network makes it easy to visualize the nodes that have relations with other nodes, whether few or many relationships within the network or outside the group's network. The role of nodes in a central position in the group has an important function as control and stability in the group [3], which in turn shows the influence of these actors in the group. Edge and node relationships in twitter are referred to as follower (number of other actors who follow) and following (number of actors followed) between actors also affect how much popularity of actors in a group.

This research took the topic of Social Network Analysis on the Natuna Case on the Twitter Social Network. Natuna is one of the waters in the Riau Islands, which is one of the outer areas of the Unitary State of the Republic of Indonesia, which is an issue that is contested between Indonesia and neighboring countries. Having an area with abundant resources is one of the reasons Natuna is being contested. Each state claims that these waters belong to them with proof of their claim. In this study, data obtained from social media twitter was used on December 20, 2019 at 00.00 WIB to January 7, 2020 at 10.00 WIB.

Social Network Analysis is a study that studies human relations by utilizing graph theory. Looking at the problems above, the application of Social Network Analysis in an application that is able to describe the relationships between individuals by visualizing in the form of graphs is likely to help the process of solving existing problems. In addition, a calculation process will be carried out on each relationship between individuals to find the centrality of a social network based on the position of each related individual in the network structure [4].

Network according to Kadushin is defined as a collection of objects or nodes and a mapping or description of the relationships between these nodes. The relationship that occurs between one node and another is an edge or link [5].

According to Marin and Wellman, a social network is a group of nodes connected relevantly by one or more relationships. Nodes or network members are units that are connected by relations in the patterns studied [6]. Wasserman and Faust also expressed their opinion about social networks as a social structure consisting of individuals or organizations called "nodes", which are connected by one or more specific types of interdependence, such as friendship, kinship, common interest, exchange of money, dislike, knowledge or prestige [7].

This study aims to find centrality based on the degree centrality approach of a social network based on the position of each individual linked in the network structure. In addition, this study also conducted graphical visualization of existing relationships using the Gephi version 0.92 application to find the most influential actors and those with the most interaction.

II. RESEARCH METHODOLOGY

The interactions that occur in the "Natuna" case on the social media network twitter form an information

dissemination, then visualized and analyzed will produce information that is useful for improving the process of information dissemination by the government. Social network analysis itself is a science that studies the relationship between one entity unit and other entity units with the help of graph theory [8].

The SNA method and technique were chosen because this method can provide an overview or visualization down to the smallest relationship that occurs only in one individual to another in the network, this SNA method can also be used to find the nodes, communities, and informal hierarchies that have the most influence in the network [9].

There are several concepts in the social network analysis approach, apart from describing the patterns formed from the relationships between nodes or actors, SNA is more often used to determine the central node in a network by calculating several centrality values, among which the commonly calculated ones are:

A. *Degree centrality* [10] calculates the number of interactions a node has. To calculate the degree centrality value of this node, it can be done using the following formula:

$$CD (ni) = d(ni) \quad (1)$$

Information:

$d(ni)$ = the number of interactions this node has with other nodes on the network.

B. *Betweenness centrality* [11] calculate how often a node is passed by other nodes to go to a particular node in the network. This value serves to determine the role of the actor who is the bridge connecting interactions in the network. To calculate the degree centrality value of a node, it can be done using the following formula:

$$CB (ni) = \sum g_{jk} (ni) / g_{jk} \quad (2)$$

Information:

$g_{jk} (ni)$ = the number of shortest paths from node js at node k passing through node i .

g_{jk} = the number of shortest paths between 2 nodes in the network.

C. *Closeness centrality* [12], calculating the average distance between a node and all other nodes in the network or in other words measuring the closeness of a node to other nodes. In a network with g nodes, the closeness centrality of these nodes is as follows:

$$CC (ni) = [N-1 / \sum d(ni, nj)] \quad (3)$$

Information:

N = the number of nodes in the network $d (ni, nj)$ = the number of shortest paths connecting nodes ni and nj .

D. *Eigenvector centrality* [13], performs measurements that give higher weight to nodes connected to other nodes that also have high centrality values. To calculate

the eigenvector centrality value of a node, it can be done using the following formula:

$$\begin{aligned} C_i (\beta) &= \sum (\alpha + \beta c_j) A_{ji} \\ C (\beta) &= \alpha (I - \beta A)^{-1} A \mathbf{1} \end{aligned} \quad (4)$$

Information:

α = normalization constant (scale vector).

B = symbolizes how much a node has a centrality weight in a node that also has a high centrality value.

Where A is the adjacency matrix, I is the identity matrix and $\mathbf{1}$ is the matrix. The amount of β is the radius of power of a node. If β is positive, then it has high centrality bonds and is connected with people who are central. Meanwhile, if β is negative, then it has high centrality bonds but is connected to people who are not central. If $\beta = 0$, you will get degree centrality.

The method in this research is carried out in several stages as shown in the figure 1 [14].

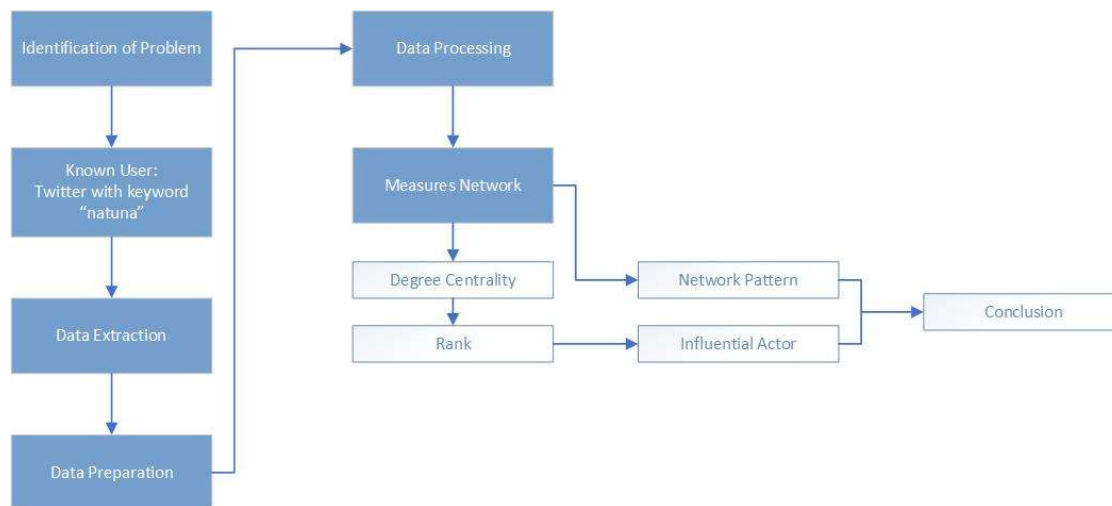


Figure 1. Research Methods

1. Identification of problems

At this stage, the identification of the research problem is carried out. The process of identifying this problem is carried out by observing the phenomena behind the research.

2. Known user

At this stage, determine the object of research. The object of research is the interaction carried out by accounts on the Twitter social network, which interact on the topic Natuna and with the keyword Natuna.

3. Data extraction

At this stage, the data collection process is carried out on the twitter social network. The data taken contains the keyword natuna.

4. Data Processing and Measures Networks

At this stage, network interaction data processing is carried out. The pattern used for graph visualization interaction uses directed type.

5. Centrality Value Calculation

At this stage, the centrality value is calculated, namely the degree centrality node or actor to identify influential actors with a high number of interactions.

6. Rank

This stage is the stage of ranking the calculated centrality values of actors in the network.

7. Draw conclusions

In this stage, conclusions are drawn from the SNA analysis using the centrality approach.

Data processing in this study was carried out using Gephi software version 0.9.2. Gephi software is an open source application for network exploration and manipulation. A network module to be developed can be imported, visualized, mapped, filtered, manipulated and exported in the Gephi software [15]. Data processing in the software is carried out in the following steps:

(1) Import network data sets that have been previously created using the help of spreadsheets in Microsoft Excel. The data set that can be used is only the dataset with a text-based .csv extension. The dataset is separated into two parts, first import data set nodes and second step import data set edges. (The nodes dataset contains a list of actor names on the network, and the edges dataset contains data on relationships or interactions that occur between nodes on the network.

(2) Choose which visualization algorithm to use. This algorithm functions to determine the layout of the nodes to be visualized in the sociogram. In addition, the selection of the algorithm also affects the form of network visualization that will be produced. In this study, the algorithm used is the Fruchterman Reingold algorithm [15].

(3) Set the algorithm configuration by changing the attribute column such as area, gravity and speed available in the properties window according to the desired configuration, then click the run button.

(4) Personalize the visualized network. In this process, you will adjust the appearance of the color, shape, labeling the nodes in the network and you can also adjust the thickness and thickness of a line edge between the nodes and give a name to the edges.

(5) Calculates network property values. In this study, the value of network property attributes was calculated in the

form of the value of Total Node, Total Edges, Average Degree, Average Weighted Degree, Network Diameter, and Number of Communities. All these attributes can be calculated by clicking one button at a time in the setting column of the statistics window.

(6) Displays the ranking of nodes that have the highest influence or interaction value on the network. This step can be done in two ways, first by looking directly at the data table window, and the second is by configuring the display by changing the size of the node or node label in the network visualization image in accordance with the order of values owned by the nodes (the larger value owned by the node, the greater the appearance of the node in the network visualization image). The second way can be done by changing the configuration in the appearance window.

(7) Export the visualization image with the .pdf, .png and .svg file extensions. The results of data calculation can also be exported by accessing the menu in the data table window. The results of data processing will be a file with the extension .csv format.

III. RESULTS AND DISCUSSION

The research was conducted by visualizing the interaction data of Natuna's information dissemination on Twitter using Gephi software version 0.9.2. Twitter data with the keyword natuna taken from December 20, 2019 at 00.00 WIB to January 7, 2020 at 10.00, consists of 71,477 actors and 147,066 interactions.

The visualization in Figure 2 below shows the formation of several groups that have been differentiated based on the color of the network or so-called modularity class. From the visualization, it can be seen that there are 3 large groups that have the most interactions on the topic of Natuna, which are shown in purple, green, and blue. The formation of these groups uses a modularity approach in the Gephi application. This degree centrality approach calculates the number of interactions a node has (indegree and outdegree).

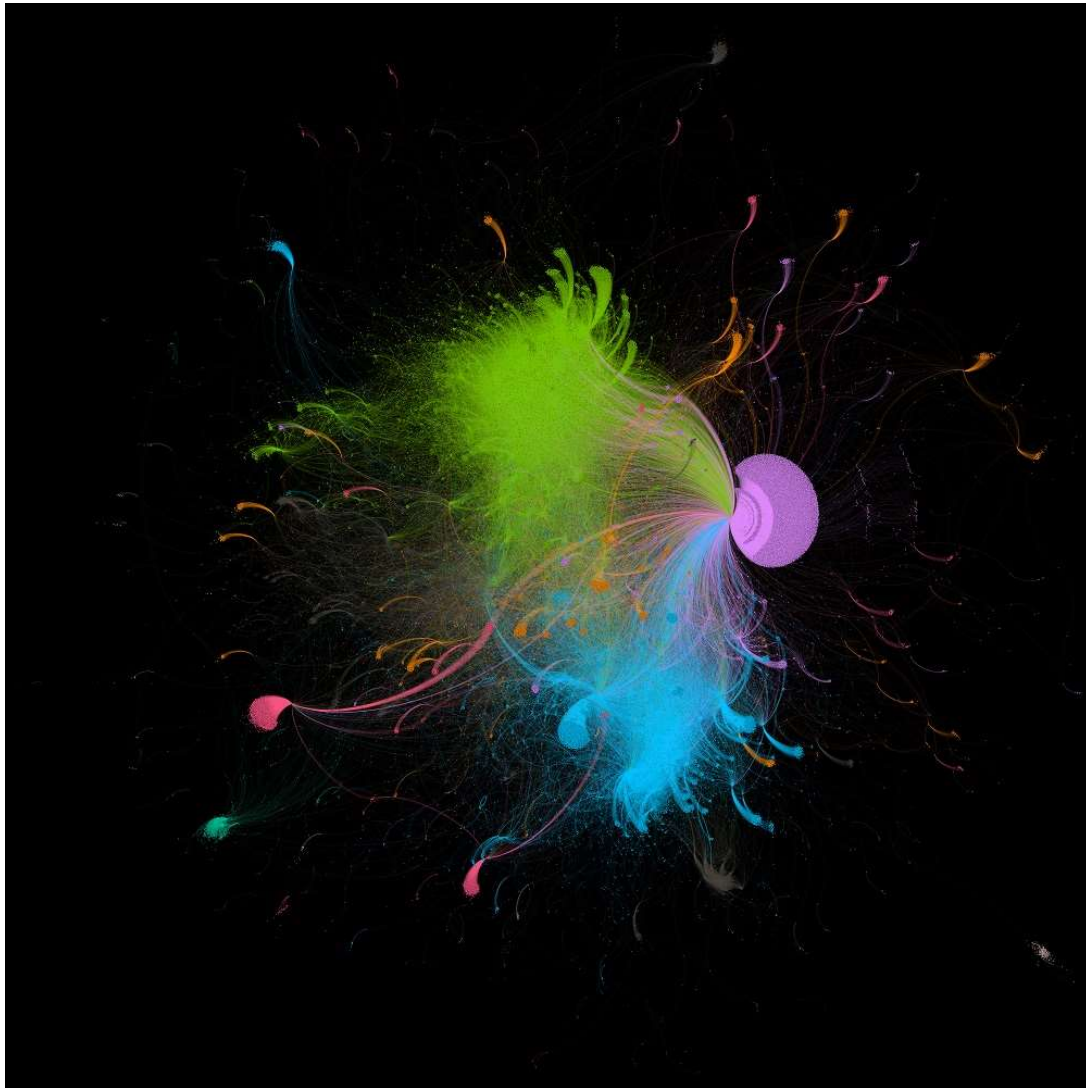


Figure 2. Visualization of the formation of issue groups

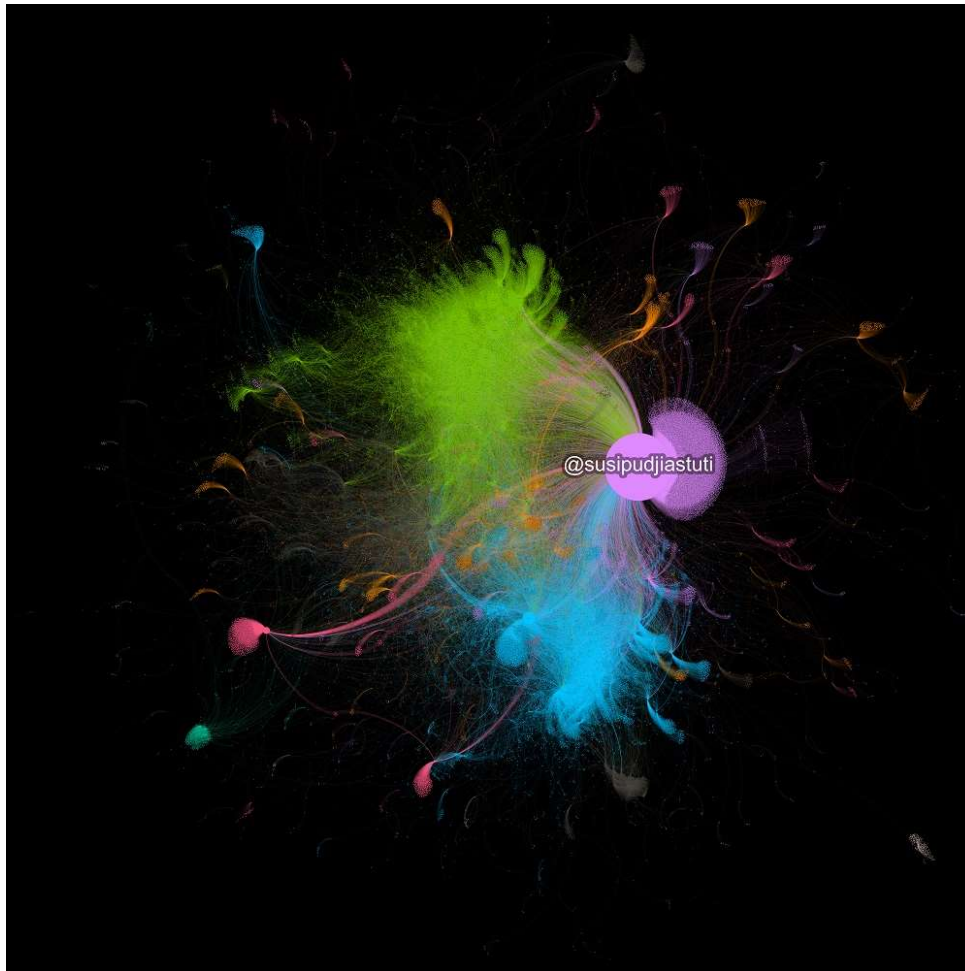


Figure3. Indegree Visualization

So for indegree is the total interaction received by a node or can be called to find which influencer / account has an important role in the network or the account that is most linked by other accounts, visualization can be independent can be seen in Figure 3. While outdegree is the total interaction which is made by a node / account which interacts most actively, outdegree visualization can be seen in Figure 4.

From Figure 3, namely indegree visualization, it can be seen that the @susipudjastuti account is the actor who plays the most role in the spread of the Natuna topic because the @susipudjastuti account is the most targeted. Notice table 1 below shows the top 10 influencers, a huge difference from the top 10 most linked accounts. The @susipudjastuti account is in the first place linked by 29755 accounts. The big difference is that the visualization only shows @susipudjastuti, while other accounts are so small that they are not visible on the visualization.

Table 1. Indegree value top 10 accounts

| No | Label | Node type | In-Degree |
|----|-----------------|------------|-----------|
| 1 | @susipudjastuti | TW_Account | 29755 |
| 2 | @jokowi | TW_Account | 5648 |
| 3 | @D4tuk_T4mburin | TW_Account | 3519 |
| 4 | @JeromePolin | TW_Account | 2817 |

| | | | |
|----|-----------------|------------|------|
| 5 | @ustadtengkuzul | TW_Account | 2381 |
| 6 | #natuna | Hashtag | 2191 |
| 7 | @geloraco | TW_Account | 2165 |
| 8 | @hnurwahid | TW_Account | 1837 |
| 9 | @do_ra_dong | TW_Account | 1621 |
| 10 | @_TNIAU | TW_Account | 1553 |

Figure 4 shows the outdegree visualization, which is the account that is most actively interacting, namely the @shaktia704 account, in this visualization only the @shaktia704 account is visible because of the difference in the value of interactions on that account. See table 1 below for a breakdown of the top 10 account outdegree values.

Table 2. Outdegree value top 10 accounts

| No | Label | Node type | Out-Degree |
|----|------------------|------------|------------|
| 1 | @shaktia704 | TW_Account | 259 |
| 2 | @GunGunG49169853 | TW_Account | 124 |
| 3 | @ANDRE OCTA | TW_Account | 90 |
| 4 | @AlvaroDeBazan2 | TW_Account | 88 |
| 5 | @AliDavala99 | TW_Account | 86 |
| 6 | @zenoldgonzalez | TW_Account | 73 |
| 7 | @margaretaputr13 | TW_Account | 62 |
| 8 | @stevansixciokre | TW_Account | 60 |
| 9 | @AriestaRiico | TW_Account | 59 |
| 10 | @MaskuMasku1 | TW_Account | 59 |

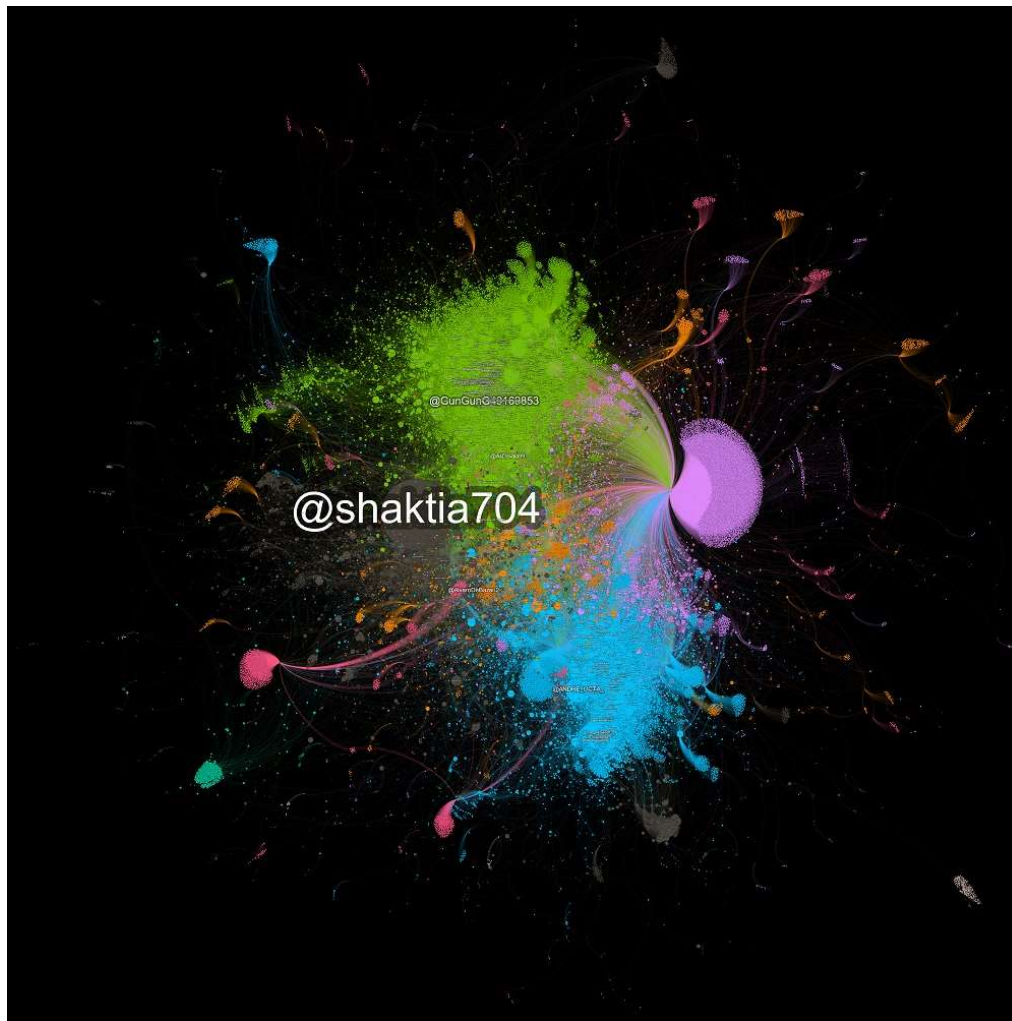


Figure 4. Outegree Visualization

In table 1, it can be seen from the results of in-degree calculations that the @susipudjiastuti account has the highest value, this is because the @susipudjiastuti account is the main actor who plays a role in the topic of Natuna. Then in table 2, it can be seen that the @shaktia704 account has the highest out-degree value because the @shaktia704 account is the most actively interacting.

Table 3. Betweenness Centrality scores top 10 accounts

| No | Label | Node type | Betweenness |
|----|----------------------|------------|----------------------|
| 1 | @Aryprasetyo85 | TW_Account | 6.387338011628019E7 |
| 2 | @susipudjiastuti | TW_Account | 6.2843006970944755E7 |
| 3 | @D4tuk_T4mburi n | TW_Account | 1.3921774279551372E7 |
| 4 | @Na_T1N4 | TW_Account | 1.3921774279551372E7 |
| 5 | @DonAdam68 | TW_Account | 1.3457628354689362E7 |
| 6 | @Zahrah4029166 0 | TW_Account | 1.2423828393732902E7 |
| 7 | @jr_kw19 | TW_Account | 8955050.328199675 |
| 8 | @7intaPutih | TW_Account | 6906324.800256199 |
| 9 | @johhhnygudhel | TW_Account | 6619723.706064895 |
| 10 | @Luana01194115 77 | TW_Account | 5196291.787102646 |

Table 3 shows the top 10 rankings betweenness centrality calculations obtained from Gephi. From these calculations it can be seen that there are 2 accounts that choose the highest value, namely the @Aryprasetyo85 and @susipudjiastuti accounts. This shows that the two accounts become a bridge or meeting point for netizens in their discussion of Natuna on social media twitter.

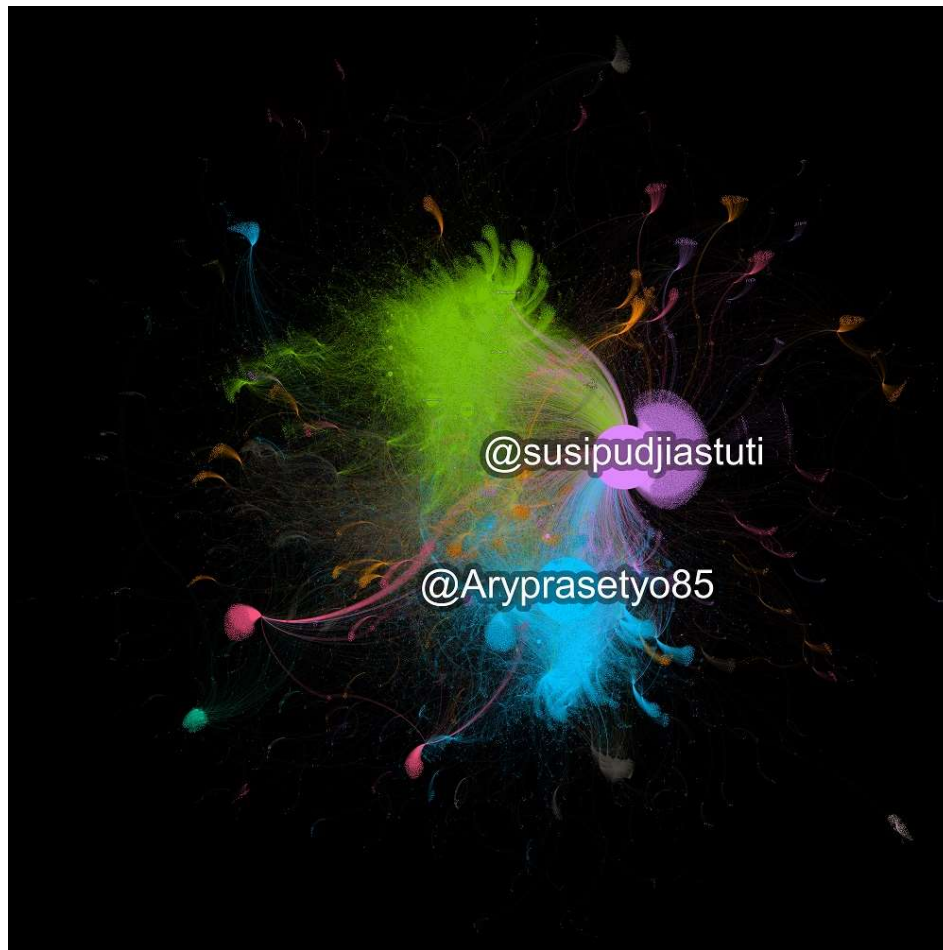


Figure 5. Betweenness Visualization

Figure 5 visualizes the results of betweenness centrality, which shows that the @Aryprasetyo85 and @susipudjiastuti accounts are meeting points for netizens on issues about Natuna on Twitter.

Table 4. Eigenvector Centrality scores top 10 accounts

| No | Label | Node type | Eigenvector |
|----|--------------------|------------|----------------------|
| 1 | @susipudjiastuti | TW_Account | 1.0 |
| 2 | @jokowi | TW_Account | 0.19216753986353677 |
| 3 | @D4tuk_T4mburi | TW_Account | 0.11837080026996827 |
| 4 | @JeromePolin | TW_Account | 0.09457074633210014 |
| 5 | @ustadtengkuzul | TW_Account | 0.08003242969610756 |
| 6 | #natuna | Hashtag | 0.07804578730451751 |
| 7 | @hnurwahid | TW_Account | 0.061940705758999355 |
| 8 | @do_ra_dong | TW_Account | 0.055108563151993695 |
| 9 | @_TNIAU | TW_Account | 0.052416368798993716 |
| 10 | #jokowikawalnatuna | Hashtag | 0.049589729591855675 |

Table 4 shows the top 10 ranking results of the Eigenvector centrality calculation obtained from Gephi. From these calculations it can be seen that the account with the highest value is the @susipudjiastuti account. This shows that the @susipudjiastuti account is the center of conversation in this case is the Natuna case, because all influencers are related to the @susipudjiastuti account.

The results of eigenvector centrality can be seen in Figure 6 which visualizes that the @susipudjiastuti account is the center of conversation among netizens in the Natuna case on Twitter.

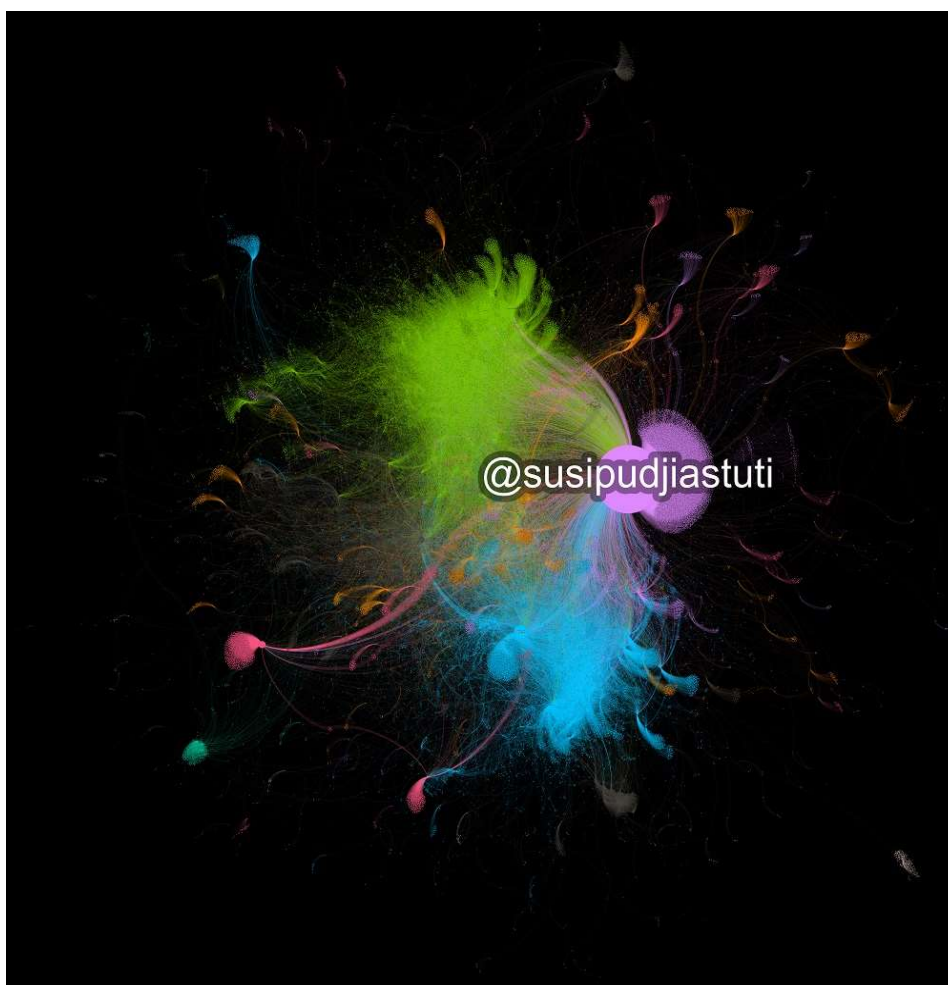


Figure 6. Eigenvector Visualization

IV. CONCLUSION

In the discussion of the results of the research that has been carried out, in order to answer the research questions that have been formulated, it can be concluded as follows:

1. The @susipudjiastuti account has an important role in the network because the account is often linked, reaching 29755.
2. The @shaktia704 account is the most active in the period of data taken, which reached 259.
3. The @susipudjiastuti and @ Aryprasetyo85 accounts become a bridge or meeting point for netizens in discussions about Natuna.
4. The @susipudjiastuti account became the center of discussion in the Natuna case.

REFERENCES

- [1] R. Nugraha, "Trending Topik Twitter sebagai Sarana Penyebaran Isu Kontroversial," *www.kompasiana.com*, 2020. <https://www.kompasiana.com/nugraha88/5e1087aed541df79920f6b22/trending-topik-twitter-sebagai-sarana-penyebaran-isu-kontroversial?page=all> (accessed Jan. 17, 2021).
- [2] L. Zhang, R. Ghosh, M. Dekhil, M. Hsu, and B. Liu, "Combining lexicon-based and learning-based methods for twitter sentiment analysis," *HP Lab. Tech. Rep.*, no. 89, 2011.
- [3] S. Fortunato, "Community detection in graphs," *Phys. Rep.*, vol. 486, no. 3–5, pp. 75–174, 2010, doi: 10.1016/j.physrep.2009.11.002.
- [4] B. Susanto, H. Lina, and A. R. Chrismanto, "Penerapan Social Network Analysis dalam Penentuan Centrality Studi Kasus Social Network Twitter," *J. Inform.*, vol. 8, no. 1, 2012, doi: 10.21460/inf.2012.81.111.
- [5] C. Kadushin, "Kadushin, Charles. Understanding social networks: theories, Kadushin, Charles. Understanding social New York, NY: Oxford

- University Press , Keen On Mass Comms ?,” vol. 18, no. April, pp. 2013–2015, 2012.
- [6] A. Marin and B. Wellman, *Handbook of Social Network Analysis*. 2009.
- [7] S. Wasserman and K. Faust, *Social Network Analysis: Methods and applications*. Cambridge University Press, 1994.
- [8] M. Tsvetovat and A. Kouznetsov, *Social Network Analysis for Startups*. 2011.
- [9] A. Bohn, I. Feinerer, K. Hornik, and P. Mair, “Content-based social network analysis of mailing lists,” *R J.*, vol. 3, no. 1, pp. 11–18, 2011, doi: 10.32614/rj-2011-003.
- [10] F. Bloch and M. O. Jackson, “Centrality Measures in Networks,” *SSRN Electron. J.*, no. January, 2016, doi: 10.2139/ssrn.2749124.
- [11] F. Y. Pratama, “Simulasi Jejaring Jalan Kota Pontianak Dengan Betweenness Centrality dan Degree Centrality,” *Progr. Stud. Tek. Ind. Fak. Tek. Univ. Tanjungpura*, pp. 1–6, 2018.
- [12] J. Zhang and Y. Luo, “Degree Centrality, Betweenness Centrality, and Closeness Centrality in Social Network,” vol. 132, no. Msam, pp. 300–303, 2017, doi: 10.2991/msam-17.2017.68.
- [13] G. Lohmann *et al.*, “Eigenvector centrality mapping for analyzing connectivity patterns in fMRI data of the human brain,” *PLoS One*, vol. 5, no. 4, 2010, doi: 10.1371/journal.pone.0010232.
- [14] 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, 2017, doi: 10.21108/indojc.2017.2.2.183.
- [15] T. M. J. Fruchterman and E. M. Reingold, “Graph Drawing by Force-directed Placement,” *Softw. Pract. Exp.*, vol. 21, pp. 1129–1164, 1991.