STRUCTURE COMMUNITY ANALYSIS ON SOCIAL NETWORK

Diah Novita Sari[1], Dedy Syamsuar[2], Edi Surya Negara[3]

Informatics Engineering, Universitas Bina Darma
Informatics Engineering, Universitas Bina Darma
Jalan Ahmad Yani No.3, Plaju, Palembang, Indonesia
[1] diahnvtsr@gmail.com, [2] dedysyamsuar@binadarma.ac.id,
[3] e.s.negara@binadarma.ac.id

ABSTRACT

Social network analysis is one field of science that has its own theory, method, and research. Social network analysis becomes one way that can be used to know the structure of community on social networks. Analysis of this community structure can provide information about the centrality of actors in a community network. The purpose of this research is to analyze the community structure contained in Twitter's social network and whoever the centrality in the community. The object of this research is twitter social network with data limit of 6000 data. The method used in this research is the Katz algorithm. The result of this research is the size of the community network with the keyword "jokowi" of 4052 nodes. Low network density is 0 (zero). The reciprocity of this research is two-way. Network diameter in this research is 6 and distance / Average Path length is 2.23 and there are 12 accounts as important users.

Keywords: Social Network, Phyton, Community Structure, Twitter

1. Introduction

Twitter is a social networking and microblogging service with the facility of texting info on user accounts with a maximum length of 140 characters via SMS, instant messaging or electronic mail [1]. Since its launch in 2006, Twitter has developed rapidly. This service is rapidly becoming popular in the world with 500 million registered users in 2012 and collecting approximately 58 million tweets per day. In Twitter's network of friends there are various virtual communities. Data network of friends on twitter obtained can form a network graph of friendship that will be analyzed its community structure. The use of twitter data has been used to do sentiment analysis, extracting geospatial data and conducting community detection analysis by utilizing several detection algorithms [2][3][4] [5]. The centrality in a virtual community can be used for marketing, criminal information, and so on. Based on the above statement, how the community structure on the social network Twitter. In addition, this social network data has also been used to measure friendship and detect the research community on the Researchgate researchers' social network [6][7]. In this study, to identify the community structure contained in the social network of

Twitter and whoever the centrality in the community used a method that is the algorithm Katz.

2 Literatur Review

2.1 Social Media

Social media is an online medium, along users easily participating, sharing and creating content including blogs, social networks, wikis, forums and virtual worlds. Social media as a group of internet-based applications built on the foundation of ideology and Web 2.0 technology, and enables the creation and exchange of user-generated content [8]. Blogs, social networks and wikipedia are the most common forms of social media used by people around the world. Others say that social media is an online medium that supports social interaction and social media using web-based technologies that turn communication into interactive dialogue [9].

2.2 Social Media Twitter

Brian J. Dixiaon (2012: 40) [10] describe that Twitter is a microblogging messaging service that limits you to 140 characters per message, including spacs and punctuation, to update your content. Twitter is part of a microblog where it can help users to post any statements in 140 characters including spaces and punctuation. Twitter is basically an instant messaging service that allows users to post anytime and from anywhere to read by anyone.

2.4 Katz Centrality

Leo Katz introduced a measure of his version of centrality which would later be known as Katz's centrality [10]. To calculate the degree of centrality of node i, we sum up nodes that are one hop away from node i (so this value equals the degree node i). However, in Katz's centrality, the n n nodes can be found in the proximity force of the matrix to n. However, less-hop nodes, to achieve a centralized value available over a distant node. The centrality of Katz can be expressed as the equation below

$$K_i = \left| \left(\sum_{k=0}^{\infty} \alpha^k A^k \right) \right|_i \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (1)$$

Where K_i is the central Katz of node i, A^k represents the strength of the k^{th} maktrik proximity, α is the attenuation factor and I is the vector column where all values are 1. Since 2 algorithms contain many overlapping parts, we choose to merge only the length of 3 centrality subgraphs with Katz centrality to create our centrality value. The formula of centrality values is shown below

$$K_i = \left| \left(\sum_{k=0}^{\infty} \alpha^k A^k \right) I \right|_i + (cA^3)_{ii} \ldots\ldots\ldots\ldots\ldots\ldots\ldots (2)$$

The other point to be discussed is how many hops can be used [11]. The Katz centrality value (ckatz) calculates the number of paths weighted in the graph starting from node i, punishing the longer path with the user choice parameter α, where α Є (0,1 / || A || 2). Walking in the graph across the edge between a series of vertices v1, v2, v3, ..., vk, where the points and edges are allowed to calculate the weighted weights of different lengths in the graph, we compute the power of adjacencymatrix using infinite series [12].

$$\sum_{k=1}^{\infty} \alpha^{k-1} A^k = A + \alpha A^2 + \alpha A^3 + \alpha A^4 + \cdots + \alpha^{k-1} A^k \ldots\ldots\ldots (3)$$

Where k (i, j) is the number of long paths k from node i to node j. As long as it is chosen to be within the right range, this infinite series integrates with the resolvent matrix A $(1-\alpha A)^{-1}$ 1, where 1 is the n x 1 vector of all 1s. This is referred to as the global Katz value and calculates the total number of weighted paths with different lengths starting from each angle

2.5 Community Structure

Some researchers who analyze the community structure on a social media are as follows:

1. Analysis and Design Tools for Finding Communities on Twitter Using Social Network Analysis and Visualizations Methods. This research was conducted in 2012 by Y. Sigit Purnomo WP, S.T., M.Kom and Th. Devi Indriasari, S.T., M.Sc. The results of this study indicate that the analysis and design of software tools to find community on social networking site Twitter using Social Network Analysis and Visualizations method has been successfully done and can be the basis for software development. [8]
2. Matching Community Structure Across Online Social Networks (2014). The research was conducted in 2014 by Lin Li, W.M.Campbell. The results of the study, there are three methods of detection of multilayer communities - Aggregation, Linking, and Relaxed Random Walk. Researchers point out that for the preparation of a researcher problem, the Aggregation approach finds a community's most loyal task on the underlying community structure [9].
3. A Novel Algorithm for Community Detection and In fl uence Ranking in Social Networks. This study was conducted in 2011 by Wenjun Wang and W. Nick Street. The result of this research is the centralistic action, the centrality that influences the correlation of the degree of centrality. All the top 10 rankings of influential centralities are very high in the degree of their respective rank. [16].

3. Result and Discussion

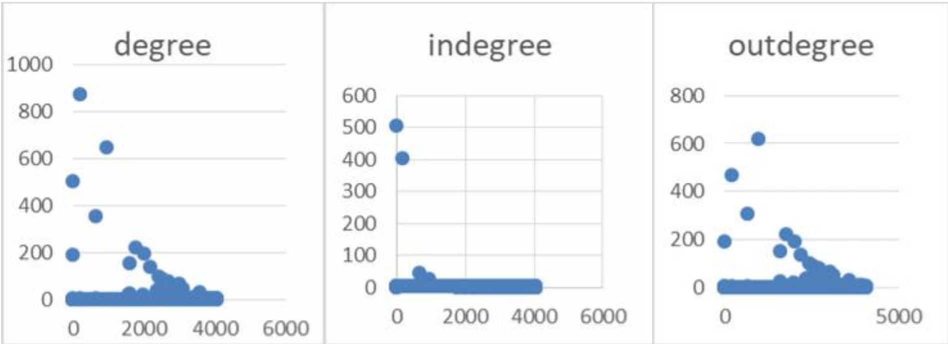
3.1 Degree Centrality

The centrality of this study can be calculated as follows:

$$C_D = \frac{4552}{4052 - 1} = 1.124 = 1$$

From the above calculation is obtained centrality level (degree centrality) is worth 1, this indicates that all the actors contacted or contacted (Valente, 2010: 82; Prell, 2012: 97).

Here is a graph of *degree distibution, indegree distribution, and outdegree distribution*.



Gambar 5.1*Degree distribution*

In Figure 5.1 above shows the spread of degree throughout the network. The highest score of the degree is 871 and its lowest value is 1. The average rating of the degree (Average Degree) is 1,123.

Figure 5.1 above shows the distribution of indegree throughout the network. Indegree is a link that leads to the actor and the actor's position as the receiving object. The highest value of the indegree is 503 and the lowest value is 0. In Figure 5.3 above shows the distribution of outdegree throughout the network. Outdegree is the number of links that come out of the actor actor and the actor's position as the subject. The highest outdegree value is 619 and the lowest value is 0.

3.2 Proximity Centrality

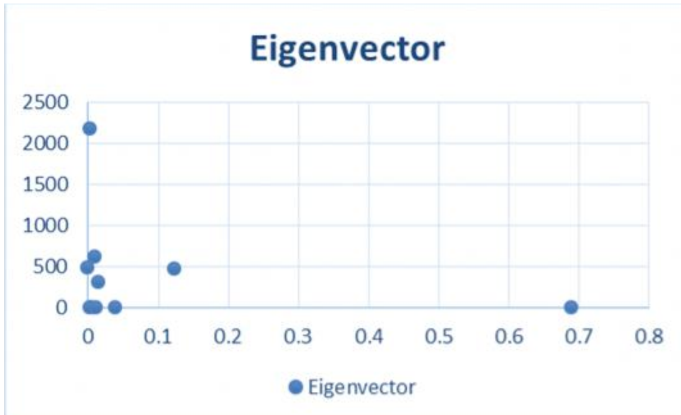
The calculation of the centrality of proximity can be seen below:

$$C_c = \frac{4052 - 1}{0.00} = 0$$

Based on the above calculation, the proximity number is 0 (zero). This shows that the average distance of the actor with all other actors in the network is very close (good).

3.3 Eigenvector Centrality

The Eigenvector describes how important a person has a network with actors by looking at how much network a person / organization with a relationship with an actor has.



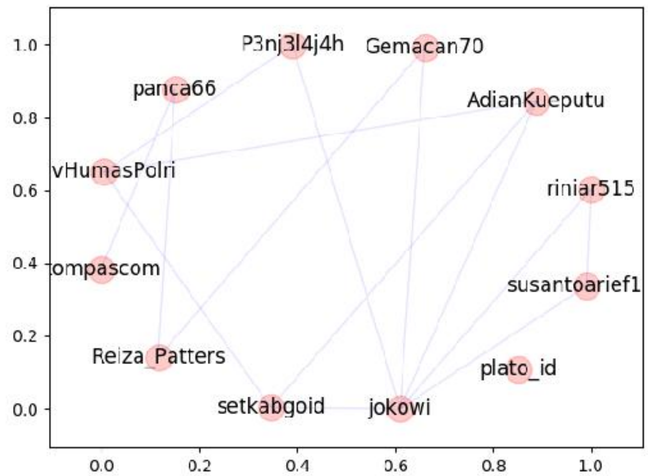
Gambar 5.4*Eigenvector Centrality Distribution*

Figure 5.4 shows the distribution of Eigenvector Centrality. The central eigenvector has a normal value of 0 to 1, which means that the highest actor has the highest eigenvector as well. Here is a visualization of the user with the highest eigenvector.

Tabel 5.1*Important User*

No	*User*	*Eigenvector*
1	Jokowi	0.666
2	Setkabgoid	0.090
3	DivHumasPolri	0.082
4	Reiza_Patters	0.079
5	Kompascom	0.055
6	panca66	0.049
7	riniar515	0.044
8	susantoarief141	0.044
9	AdianKueputu	0.042
10	Gemacan70	0.042
11	plato_id	0.039
12	P3nj3l4j4h	0.039

In Table 5.1 there are 12 accounts as important users. The account is "Jokowi" account with eigenvector 0.666, "setkabgoid" account with eigenvector 0.090, account "DivHumasPolri" with eigenvector 0.082, "Reiza_Patters" account with eigenvector 0.079, "kompascom" account with eigenvector 0.055, panca66 account with degreeeigenvector 0.049, account "riniar515" with eigenvector 0.044, account "susantoarief141" with eigenvector 0.044, "AdianKueputu" account with eigenvector 0.042, account "Gemacan70" with eigenvector 0.042, "plato_id" account with eigenvector 0.039, and "P3nj3l4j4h" account with eigenvector 0.039.



Gambar 5.5 Graph *Important User*

4. Conclussions and Recomendations

4.1 Conclussions

1. The size of the community network with the keyword "ahok" is a number of 4052 nodes. Low network density is 0.0002 if rounded to 0 (zero). The reciprocity of this research is two-way. The network diameter in this research is 6 and the distance / Average Path length is 2.23.
2. There are 12 accounts as important users. The account is a "Jokowi" account with a degree of 0.666, a "setkabgoid" account with a 0.090 degree, a "DivHumasPolri" account with a 0.082 degree, a "Reiza_Patters" account with a 0.079 degree, a "kompascom" account with a 0.055 degree, a "panca66" account with a 0.049 degree, account "riniar515" with 0.044 degrees, "susantoarief141" account with 0.044 degree, "AdianKueputu" account with 0.042 degree, "Gemacan70" account with 0.042 degree, "plato_id" account with 0.039 degree, and "P3nj3l4j4h" account with 0.039 degree.

4.2 Recomendations

This study only analyzes the existing community structure on a network of friends on Twitter social networks. More research on text mining analysis is required.

Reference

[1] Kusuma, Octavia, dkk .2014. *Analisis Penggunaan Twitte Sebagai Media Komunikasi Selebritis di Jakarta*. Jakarta : Universitas Taruma Negara.

[2] Retnawiyati, E. and Fatoni, M.M., M. Kom.,/Surya, Edi Negara, M. Kom. Analisis Sentimen Pada Data Twitter dengan Menggunakan Text Mining terhadap Suatu Produk.

[3] Negara, E.S., Andryani, R. and Saksono, P.H., 2016. Analisis Data Twitter: Ekstraksi dan Analisis Data G eospasial. INKOM Journal, 10(1), pp.27-36.

[4] Antoni, D., Negara, E.S. and Suweno, S., 2015. Ekstraksi Data Geo-Spatial Twitter (Studi Kasus: Badan Penyelenggara Jaminan Sosial Kesehatan). Seminar Nasional Teknologi Komputer, 22-24 Oktober 2015.

[5] Negara, E.S. and Andryani, R., 2018. A Review on Overlapping and Non-Overlapping Community Detection Algorithms for Social Network Analytics. Far East Journal of Electronics and Communications, 18 (1), 1-27

[6] Negara, E.S., Kerami, D., Wiryana, M., Kusuma, TM. 2017. Researchgate data analysis to measure the strength of Indonesian research. Far East Journal of Electronics and Communications, 17 (5), 1177-1183

[7] Ria Andryani, M.M., Kom, M., Ria Andryani, M.M., Kom, M. and Edi, S.N., 2017. Network of Friends to The Other Friends by Social Media on Facebook. The Turkish Online Journal of Design, Art and Communication, 12(12), pp.1363-1378.

[8] Surya Negara, Edy, Andriani, Ria, dan Saksono, Prihambodo Hendro. 2016. *Analisis Data Twitter : Ekstrasi dan Analisis Data Geospasial*. Palembang : Univeresitas Bina Darma.

[9] Kaplan, M. Andreas and Halein, Michele. 2009. *The challenges and opportunities of Social Media*. Indiana University.

[10] Dixiaon, Brian. J. 2012. *Social Media for School Leader*. United Stated : Jossey Bass A Wiley Inprint.

[11] Sheppard, Kevin. 2014. *Introduction to Python for Econometrics, Statistics and Data Analysis.* University of Oxford.

[12] Jarukasemratana, Sorn. 2014. *Community Detection Algoritm based on Centrality and Node Closeness in Scale-Free Networks*. Japan : Tokyo Institute for Technology

[13] Nathan, Eisha. 2017. *Local Community Detection in Dynamic Graphs Using Personalized Centrality*. Atlanta : Georgia Tech.

[14] Purnomo, Sigit Y, dkk. 2012. *Analisis dan Perancangan Alat Bantu untuk Menemukan Komunitas di Twitter Menggunakan Metode Sosial Network Analysis and Visualizations*. Yogyakarta : Universitas Atma Jaya.

[15] Li, Lin dan Champbell, W.M. 2014. *Matching Community Structure Across Online Social Networks*. United Stated : MIT Lincoln Laboratory.

[16] Wang, Wenjun dan Stree, W.Nick. 2011. *A Novel Algorithm for Community Detection and Influence Ranking in Social Networks*. USA : Universiti of IOWA