Implementation of centrality measures in graph represented information spreads with hashtag #bersatulawancorona in Twitter

E Carnia*, B Fermadona, H Napitupulu, N Anggriani, and A K Supriatna

Department of Mathematics, Universitas Padjadjaran, Jl. Raya Bandung-Sumedang km 21 Jatinangor, Sumedang 45363, Indonesia

*ema.carnia@unpad.ac.id

Abstract. Twitter is a social network that is very popular among the people that allows the users to send and read text-based messages up to 140 characters, known as tweets. Since it was first launched until now, this social network has become one of the ten most visited sites on the internet. Because there are so many people use this social network, Twitter has become a source of information that can disseminate information in a short time. Dissemination of information on Twitter is very preferred by the accounts of Twitter users. The most important account in the dissemination of information on Twitter can be searched by calculating the centrality measurement. In this paper, four centrality measures are used to find the most important accounts, that are degree centrality, closeness centrality, betweenness centrality, and eigenvector centrality. Before calculating the centrality of the Twitter accounts, data must be collected first, wherein this paper the data to be used is tweeted data from Twitter using hashtag #BersatuLawanCorona obtained using Python with the help of Twitter API. In calculating this centrality value, we can get the most influential account in distributing information related to hashtag #BersatuLawanCorona. Based on the data, obtained one account that became the most influential account because it obtained the highest value for each centrality measurement, which is SapawargaSby account.

1. Introduction
Coronavirus is a virus that attacks the respiratory system. This disease due to viral infection is called Covid-19. The Coronavirus can cause minor disturbances to the respiratory system, severe lung infections, and death. Covid-19 is currently a worldwide disaster. Currently, Covid-19 is being researched by scientists from all scientific fields according to their expertise, such as the health sector. Many researchers are developing and demonstrating in a real case study a methodology for supporting Occupational Health and Safety Services in the design and assessment of preventive measures to reduce the risks of Covid-19 outbreaks within their entities [1,2]. In the economic field, the influence of Covid-19 is very large [3]. On the other hand, mathematicians view this problem, among others, from the modeling and graph aspects, including analysis of social networks [4-6].

Twitter as a social media that is widely used by the community plays a role in the dissemination of information to prevent the virus. Twitter as a social media that is widely used by the community plays a role in the dissemination of information to prevent the virus. The dissemination of information through Twitter social networks can be done quickly and can be spread in a very short time through the posts of Twitter users themselves. The information provided by these users will be visible to other users and may be reposted by that user via retweet. Some researchers use Twitter in their research related to social network analysis [7,8]. This research is intended to find out which accounts have the
most influence in the dissemination of information on the coronavirus in tweets using the hashtag #BersatuLawanCorona based on the calculation of Degree centrality, Betweenness Centrality, Closeness Centrality, and Eigenvector Centrality. The calculation of eigenvector centrality is done using characteristic equations and by using power iteration. In this research data is obtained using Python software with the help of Twitter API and the data retrieved is in the period from March 1, 2020 to June 20, 2020.

2. Method
In this paper, the data used is tweet data from the Twitter social network using the hashtag #BersatuLawanCorona, the data obtained is then represented in graph form and then processed and analyzed by Centrality Measurement using the Python application, so that later on which accounts will influence the most dissemination of information provided.

Many theories are supporting the calculation of centrality measures used for the search of the most influential accounts in the graph of the dissemination of #BersatuLawanCorona information on the social network Twitter, that can be found in [5,9-11].

2.1 Centrality
The idea of centrality as applied to human communication was introduced by Bavelas in 1948. This study will be used the calculation of four kinds of centrality, namely degree centrality, closeness centrality, betweenness centrality, and eigenvector centrality. The following is the calculation formula of the four centralities on a G-directionless graph.

2.1.1 Degree Centrality
Degree centrality is used to search for the accounts that have the most influence on the dissemination of information on Twitter by looking at the number of direct relationships an account has with another account. The higher the degree centrality value, the more relationships an account has with another. To calculate the degree centrality value, $C_D(v)$, the following formula can be used:

$$C_D(v) = \frac{d(v)}{n - 1}$$  \hspace{1cm} (1)

where $n$ is the sum of all vertices in the graph and $d(v)$ is the degree of vertex $v$.

2.1.2 Closeness Centrality
Closeness centrality is used to search for the most influential accounts by looking at how close an account is to another based on the shortest distance obtained. Closeness centrality, $C_C(v)$, is calculated using the following formula:

$$C_C(v) = \frac{n - 1}{\sum_{i=1}^{n} d(v, t_i)}$$  \hspace{1cm} (2)

where $n$ is the number of vertices in the graph, and $d(v, t_i)$ is the shortest distance connecting vertices $v$ and $t_i$.

2.1.3 Betweenness Centrality
Betweenness centrality is used to search for the most influential accounts in the dissemination of information based on the extent to which they are required as a link in the dissemination of information on Twitter social networks. To calculate the value betweenness centrality, $C_B(v)$, the following formula can be used:

$$C_B(v) = \frac{2}{(n - 1)(n - 2)} \sum_{s,t\in V} \frac{\sigma_{s,t}(v)}{\sigma_{s,t}}$$  \hspace{1cm} (3)
where:
\( \sigma_{st} = \) number of shortest paths between vertices s and t
\( \sigma_{st}(v) = \) the number of shortest paths between vertices s and t that pass through vertex v.
\( n = \) number of vertices on the graph

2.1.4 Eigenvector Centrality
Eigenvector centrality is used to search for the most influential accounts by identifying the influence of those accounts across the network, not just their influence on directly connected nodes. To find the eigenvector centrality value used the following formula:

\[
C_E(v) = x_i = \frac{1}{\lambda} \sum_{j=1}^{n} A_{ij} x_j
\]

where \( A_{ij} \) neighboring matrix, \( n \) is the number of vertices in the graph, and \( \lambda \) is the dominant eigenvalue. To search for the dominant eigenvector, the power iteration method can be used.

2.2 Power Iteration Algorithm
The power iteration algorithm also known as the power method is an algorithm to search for the dominant eigenvalue in the matrix. For example, given a matrix that can be recognized \( A \), the algorithm will generate \( \lambda \) which is the largest eigenvalue (of absolute value) of \( A \) and a non-zero vector \( v \) which is an eigenvector corresponding to the Eigen \( \lambda \) value, thus \( Av = \lambda v \). The Power Iteration algorithm starts with the \( b_0 \), which is the initial value of the dominant eigenvector or random vector. The iteration steps of this method are as follows:

\[
b_{k+1} = \frac{Ab_k}{\|Ab_k\|}
\]

with \( \|Ab_k\| \) is the norm of multiplication of matrix \( A \) with vector \( b_k \) on the iteration to-\( k \). The iteration will continue and stop when \( \|Ab_{k+1}\| \) convergent to \( \|Ab_k\| \). In this case, convergence is achieved when \( \|Ab_{k+1}\| - \|Ab_k\| \leq \epsilon \), with \( \epsilon \) is a positive small number of tolerances.

2.3 Twitter and API Twitter
Twitter is a free social networking tool that is widely used and allows people to share information, in real-time newsfeeds to people who have the same thoughts [7]. Application Programming Interface (API) twitter is a program or application provided by Twitter to make it easier for other developers to access the information on the Twitter website

3. Result and Discussion
The data used in this article is tweet data posted from the 1st of March to the 20th of June 2020. From the data, it’s determined whether the tweet is the result of a retweet or not. If the tweet was a result of a retweet, then find who is the original writer of the tweet. This data then represented in a simple graph where the nodes represent the Twitter account and if an account retweets from another account, then the two accounts will be linked by side.

By using tweet data with hashtags #BersatuLawanCorona obtained using Python software and Twitter API help, the following graphs are shown in Figure 1.
Figure 1. Representation of tweet data graph with hashtag #BersatuLawanCorona

3.1 Centrality Measures from tweet data with hashtag #BersatuLawanCorona

From the data obtained, we represented the data into a graph, and from the graph, the centrality value is calculated for each account using Python software to find the most influential accounts and obtained 5 accounts with the highest value in each centrality as follows:

Table 1. Comparison of the value of the four Centrality accounts with #BersatuLawanCorona with the top five rankings

| No | Username          | DC   | Username     | CC   | Username | BC   | Username     | Power Iteration | Char. Eq. |
|----|-------------------|------|--------------|------|----------|------|--------------|-----------------|-----------|
| 1  | Sapawarga Sby     | 0.15758 | Sapawarga    | 0.15758 | Sapawarga Sby | 0.02402 | SapawargaSby | 0.70710         | 0.70710   |
| 2  | Humassijn Jung    | 0.06667 | Humassiju    | 0.08033 | Gebns    | 0.00407 | Gebns        | 0.14142         | 0.14142   |
| 3  | Humaspol Dajbr    | 0.04242 | humaspol     | 0.08033 | BabyOnKey93 | 0.00200 | BabyOnK93  | 0.14142         | 0.14142   |
| 4  | MadeBudi Yasa9    | 0.01818 | polsekibat   | 0.08033 | Ppxid    | 0.00052 | Ppxid       | 0.14142         | 0.14142   |
| 5  | Bpblbatam         | 0.01818 | bpblbatam    | 0.08033 | Siemak   | 0.00022 | siemak      | 0.14142         | 0.14142   |

From Table 1, it can be seen that according to each centrality calculation, SapawargaSby account always gets the highest value, which means that the account is the account that has the most connection with other accounts, has the closest relationship with many other accounts, becomes the contact of an account with another account, and has the most interaction with other important accounts in the graph. If the calculation result of each centrality is represented in a graph where the centrality value is comparable to the size of the account node itself, the graph is obtained as follows:
The SapawargaSby accounts have the highest centrality measures, meaning it has the most influential accounts based on degree centrality measures, closeness centrality measures, betweenness centrality measures, and eigenvector centrality measures.

3.2 Comparison of Eigenvector Centrality Search Using Characteristic Equation and Power Iteration

The eigenvector centrality value of a Twitter account that has been represented in the form of a graph can be searched using the characteristic equation of the eigenvalue, i.e. by searching for the eigenvalue and its eigenvector, or it can be by using the power iteration method. However, each of these means has its differences and disadvantages, so it will be matched by the search results of eigenvector centrality using both ways through the following examples.

Given five sample graphs as in Table 2. For the above five graphs are searched for centrality values using characteristic equations and power iterations where each step is as follows, and obtained results as in Table 2.
Table 2. Comparison of eigenvector centrality results using characteristic equations and power iteration algorithms

| No | Graph | Eigenvector Centrality ($C_E$) | Error $|C_E(pk) - C_E(pi)|$ |
|----|-------|--------------------------------|-----------------------------|
|    |       | characteristic eq. $C_E(pk)$ | Power Iteration $C_E(pi)$ |                               |
| 1  |       | [0.57]                         | [0.57]                      | [0]                           |
|    |       | 0.57                           | 0.57                        | 0                             |
|    |       | 0.57                           | 0.57                        | 0                             |
| 2  |       | 0.5                            | 0.5                         | 0                             |
|    |       | 0.5                            | 0.5                         | 0                             |
|    |       | 0.5                            | 0.5                         | 0                             |
| 3  |       | 0.43                           | 0.44                        | 0.01                          |
|    |       | 0.55                           | 0.54                        | 0.01                          |
|    |       | 0.43                           | 0.44                        | 0.01                          |
|    |       | 0.55                           | 0.54                        | 0.01                          |
| 4  |       | 0.36                           | 0.36                        | 0                             |
|    |       | 0.58                           | 0.58                        | 0                             |
|    |       | 0.58                           | 0.58                        | 0                             |
|    |       | 0.36                           | 0.36                        | 0                             |
|    |       | 0.22                           | 0.22                        | 0                             |
| 5  |       | 0.41                           | 0.40                        | 0.01                          |
|    |       | 0.41                           | 0.40                        | 0.01                          |
|    |       | 0.54                           | 1.59                        | 0.05                          |
|    |       | 0.41                           | 0.40                        | 0.01                          |
|    |       | 0.41                           | 0.40                        | 0.01                          |

From the table above it can be seen that the results obtained using characteristic equations are not much different than the results obtained when using power iteration. However, working on using characteristic equations will be more complicated because many steps have to be worked on, especially if you have to find the dominant eigenvalue of a graph that has many vertices, it will certainly take more time than using a power iteration that only takes a few iterations.
4. Conclusion
Centrality's calculations in the study were used to look for an account that was most influential in the dissemination of information from tweets that used #BersatuLawanCorona hashtags on Twitter social networks based on four centrality measurement: degree centrality, closeness centrality, betweenness centrality, and eigenvector centrality. From the calculation of degree centrality value obtained the most influential account in a graph based on the number of direct relationships owned by an account with another account. While with the calculation of closeness centrality obtained the most influential account in the graph based on how close the account is to other accounts according to the shortest distance, and from the calculation betweenness centrality obtained the most influential account based on how far the account is needed as a link in the dissemination of information on social media Twitter. Meanwhile, the calculation of eigenvector centrality obtained the most influential account in the spread of tweets with hashtag #BersatuLawanCorona based on how many relationships the account had with other accounts across the network.

From all centrality calculations, the account with the highest value for each centrality calculation is sapawargasby account. So, it can be said based on the calculation of the value of centrality, then this account is the most influential account in the dissemination of information through tweets with hashtag #BersatuLawanCorona on the social network Twitter.

The calculation of eigenvector centrality using characteristic equations and power iteration will give a not much different result. However, using characteristic equations for the search for eigenvector centrality values for graphs that have multiple vertices will take a considerable time as they will have to take several steps to find the eigenvector centrality value.

Acknowledgments
The authors would like to thanks Universitas Padjadjaran for the full support of the research conducted through Academics Leadership Grant 2020 under Prof. Dr. Asep. K. Supriatna with Contract no. 1427/UN6.3.1/LT/2020. We also thanks to the Departement of Mathematics and Faculty of Mathematics and Natural Sciences for research opportunity.

References
[1] Gallardo E C, Arroyabe J, and Arranz N 2020 Preventing Internal Covid-19 Outbreaks within Businesses and Institutions: A Methodology Based on Social Networks Analysis for Supporting Occupational Health and Safety Services Decision Making Sustainability 12 4655
[2] Wang, Y, Hu Z, Feng Y, Wilson A, and Chen R 2020 Changes in network centrality of psychopathology symptoms between the Covid-19 outbreak and after peak Molecular psychiatry 1-10
[3] Jackson, J K, Weiss M A, Schwarzenberg A B, and Nelson R 2020 Global Economic Effects of Covid-19 Congressional Research Service 2020 21 September (Washington, USA) R46270
[4] Zhang J and Luo Y 2017 Degree Centrality, Betweenness Centrality, and Closeness Centrality in Social Network MSAM 2017: Advances in Intelligent Systems Research 2017 26-27 March (Bangkok, Thailand) 132 300-3
[5] Freeman L C 1978 Centrality in social networks conceptual clarification Social networks 1 215-39
[6] Dekker A H, 2008 Centrality in social networks: Theoretical and simulation approaches Proceedings of SimTecT 2008 12-15 May (Melbourne, Australia)
[7] Mistry V 2011 Critical care training: using Twitter as a teaching tool British Journal of Nursing 20 1292-1296
[8] Priyanta S, Trisna I N P, and Prayana N 2019 Social Network Analysis of Twitter to Identify Issuer of Topic using PageRank International Journal Of Advanced Computer Science And Applications 10 107-111
[9] Newman M E J 2010 Network: An introduction 1st edition (Oxford University Press)
[10] Wasserman S and Faust K 1994 Social network analysis: methods and applications (Cambridge University Press)

[11] Golbeck J 2013 Analyzing the social web (USA, Elsevier) 67-101