Contact Tracking with Social Network Analysis Graph

Alvida Mustika Rukmi*, Wildan Zakky, M. Lutfhi Shahab

Department of Mathematics, Institut Teknologi Sepuluh Nopember
Campuss ITS, Sukolilo, Surabaya, 60111, Indonesia
Corresponding author. Email: alvidamr@gmail.com

ABSTRACT
In 2020, the world is facing a Covid-19 virus pandemic. The fields of epidemiology and networks are needed in dealing with its spread. Individual (contact) tracing is an important control measure in the spread of infectious diseases. The network of contacts describes the potential pathways for the spread of the disease. Social Network Analysis (SNA) is the study of structural construction based on graph theory with Degree Centrality, Betweenness Centrality, and Closeness Centrality. In this paper, the SNA graph provides a description of the contact network in a resident cluster. Visualization of the formed SNA graph provides information on the pattern of contact relationships in the cluster. The measurement of centrality in the SNA method identifies who has a high value of connectedness and closeness in the cluster. The calculation of degree centrality shows that node-F391 is the node with the highest interaction, which is 3,203. This indicates that node interacts the most with residents in its environment within 14 days before and after being positive for COVID-19 based on the date of the swab test. The calculation of Betweenness Centrality shows that nodes-B041 and C371 are important actors with the highest value of 0.006. The two node becomes an important link. The use of graphs on SNA provides visualization of the contact network to distinguish potential relationships between close contacts, so that contact tracing can be carried out based on disease path mapping. Compared to the conventional way of reviewing individual contact records, graph visualization on the SNA is a good adjunct tool for healthcare workers in contact tracing investigations.

Keywords: *Social network analysis, Contact tracing, SNA graph.*

1. INTRODUCTION

Social Network Analysis (SNA) as a method that expresses the structural relationship between interacting objects and their effect on the network. The SNA method has a concentration on interaction research and is often used to measure a relationship and describe some individual information [1]. In 1985, Klovdahl [2] formally published the relationship between contact tracing and SNA using a dataset from Auerbach et al (the AIDS dataset in the United States). In 2018, Adegboye used the UCINET program to analyze the spread of MERS from the Kingdom of Saudi Arabia between June 2012 and September 2016[3].

Social Network Analysis can be used in analyzing the occurrence of disease spread. When applied to epidemiology, social networks are called contact networks. Visualization of Social Network Analysis in the form of a graph containing vertices (nodes) and edges (lines) in its structure, makes it easier to analyze contacts that have the potential to transmit. Based on the graph formed, it is possible to examine how contacts are connected to each other, and the impact of these connections is how the network of contacts spreads disease. This will be followed up with contact tracing.

Contact tracing is a method used to combat infectious diseases, and is based on the assumption that disease is transmitted through close contact[4]. Tracing the interaction history of the exposed person can identify the main spreader of the infectious disease. Based on the contact history, action can be taken to break the chain of transmission by first tracing the source of the infection. This tracking can provide an overview of who is connected to whom, know the location of the transmission center and how the virus spreads from the network center, as well as the link that transmits the disease.

With the SNA method, centrality is measured, to identify potential individuals in an outbreak. In this study, the SNA graph provides a description of the contact network in a housing cluster. Visualization of the formed

SNA graph provides information on the pattern of contact relationships in the cluster. The results of the centrality measurement identify who has high connectedness and closeness values in the cluster.

A graph is a set of objects called vertices (or nodes) that are connected by edges[5]. A graph $G(V, E)$ consists of the set of vertices represented by $V = \{v_1, v_2, v_3, \ldots, v_n\}$ and the set of edges represented by $E = \{e_1, e_2, e_3, \ldots, e_n\}$ where $e_i = (v_i, v_j)$ is the edge that connects vertex v_i with vertex v_j. Set of objects called vertices (or nodes) that are connected by edges. A graph $G(V, E)$ consists of the set of vertices represented by $V = \{v_1, v_2, v_3, \ldots, v_n\}$ and the set of edges represented by $E = \{e_1, e_2, e_3, \ldots, e_n\}$ where $e_i = (v_i, v_j)$ is the edge that connects vertex $v_i$with vertex v_j. Graph with five vertices and eight edges can be seen in Figure 1.

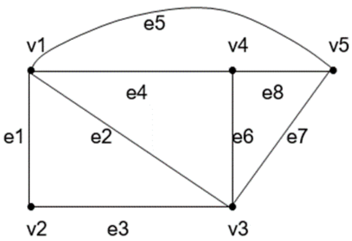

Figure 1 Example graph.

In mathematics and computer science, graph theory is a branch of science that studies the properties of graphs that represent objects and the relationships between these objects. Graph is also a diagram that contains certain information to describe various kinds of existing structures. Visualization of objects and their relationships in the form of graphs, is easier to understand.

A digraph with weighted edges is called a network. Network analysis is often used in the practical branch of graph theory.

Social network analysis is a technique to map and measure the relationships and communications that occur between humans, groups, organizations, and represent the organizational structure in the network, and the impact of transmission in the network. These relationships are visualized in the form of an SNA graph graph

The basic nature of this methodology is the ability to measure the connectedness of actors (nodes) to each other in the network, provide a visualization of a node connected to other nodes, represent variations and changes in these relationships, and pattern relationships based on node linkages.

In any society, each individual develops relationships with others and forms his personal network through social activities[6]. By reconstructing social networks, where individuals are represented as nodes and connected to each other by edges that describe the bonds of their relationship. Figure 2 shows a network of friends consisting of 10 people. In this network, Persons A and G are considered "popular" because they are connected to a large group. Whereas F is also important even though he doesn't have as many connections as A and G, because F bridges two different groups of friends. Without Person F, these two groups of people have no chance of establishing a relationship with one another. In SNA, these three people are said to be the center in the network.

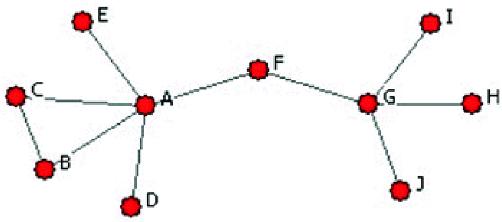

Figure 2 A network of friends consisting of ten people.

Various measures of centrality, such as degree and betweeness, can be used to define nodes (as central) in the network. Degree is the number of edges that a node has. Betweeness measures the extent to which a node lies between various other nodes. The higher the intermediateness of a node, the greater the potential it has to control other connections[7].

From a network science perspective, SNA provides a mathematical approach to measuring relationships between network members and as a tool to show patterns. Nodes and vertices have their own measurement meanings which are explained by the following terminology in graph theory:

a) *Degree centrality* is the number of connections a node has. The actor with the most relationships is the most important actor. It can be calculated using the following:

$$C_D = \frac{d(n_i)}{g} \quad (1)$$

where C_D = *degree centrality,* $d(n_i)$ = *degree* of actor (*node*) *I, and* g = total number of actor.

b) *Closeness centrality* is the closeness of an actor to other actors. Actors are said to be important if they are close to other actors. Closeness is based on the reciprocal of the distance from each actor to the other actors on the network. It can be calculated using the following:

$$C_c(n_i) = \left[\sum_{j=1}^{g} d(n_i, n_j)\right]^{-1} \quad (2)$$

where $C_c(n_i)$ = *closeness centrality* actor (*node*) *i*, $d(n_i, n_j)$= distance of actor *i* to *j, and* g = total number of actor.

c) *Betweeness centrality* is calculates the number of shortest paths between actors j and k where actor i is. It can be calculated using the following:

$$C_B(n_i) = \sum_{j<k} \frac{g_{jk}(n_i)}{g_{jk}} \quad (3)$$

where $C_B(n_i)$ = *betweeness centrality* actor (*node*) *i*, $g_{jk}(n_i)$ = number of paths where actor *i*, g_{jk} = number of paths connecting actor *j*, g = total number of actor.

Table 1. Degree and betweenness node in Figure 2

Node	Degree	Betweenness
A	5	25
G	4	21
F	2	20
B, C	2	0
D, E, H, I, J	1	0

Table 1 lists the degree and betweeness nodes in the sample network of friends in Figure 2, so you can see.

2. DISCUSSION

2.1 Data retrieval

Data retrieval using a questionnaire method, where respondents answered questions with a Yes or No choice technique based on their own circumstances[8]. The respondent's location was chosen in the form of a resident cluster during the period of July 2, 2021. Data collection from this questionnaire is qualitative in nature to analyze the interaction of family clusters in a real estate in Sidoarjo. In the process of collecting data during the COVID-19 pandemic, researchers sent questionnaire sheets to the homes of people who were positively exposed to COVID-19 who were undergoing self-isolation. The results of filling out the questionnaire were photographed and sent to the researchers via whatshapp message.

This study aims to create a platform that provides information on the spread of Covid-19 in a resident cluster in a neighborhood community scale (Rukun Tetangga, RT). This platform will produce a graph that can describe the spread of covid-19 in a housing consisting of 5 RT. Each RT has different numbers of people who have tested positive for COVID-19. Data collected were 99 people who were tracked to interact within 14 days before and after with positive people with COVID-19 based on the swab tests they carried out until August 1, 2021. The variables used were: address, age, gender, history of comorbid diseases.

Furthermore, the data that has been obtained will be processed so as to produce a display in the form of a graph to make it easier to read the data and speed up the process of finding the source of the spread and spread of COVID-19 in the family cluster.

2.2 Data Processing

At this stage, the questionnaire data was processed by tabulating the questionnaire in the form of a matrix using the Excel file format. The column and row sections of the matrix represent individual residents of housing. Giving a value of 0 in the matrix cell states that there is no relationship between 2 residents, while giving a value of 1 if there is a relationship between 2 residents. Figure 3 presents the connectedness matrix from the close contact tracing of one of the many RT's in the resident cluster.

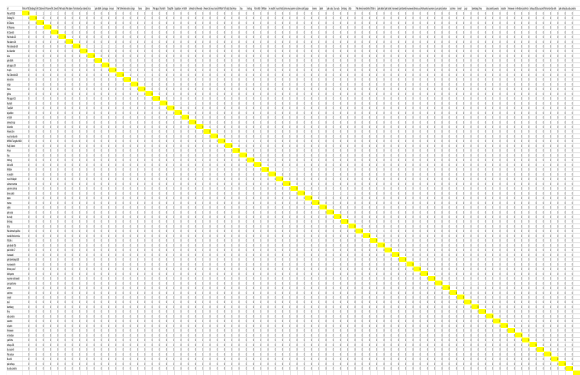

Figure 3 Neighborhood connectivity matrix.

2.3 Social Network Analysis Method and Visualization

The SNA graph is made up of nodes that are closely related. At this stage, the data of people exposed to Covid-19 is entered in an Excel file into the UCINET program so that they can form an SNA graph. UCINET will measure the system level of the network structure and network characteristics by using the density component, namely the measurement of centrality.

SNA is measured through three centralities, namely Degree Centrality from Equation (1), Closeness Centrality from Equation (2), Betweeness Centrality from Equation (3).

Based on the results of the calculation of centrality in Table 2, identify which residents have high connectedness and closeness values in the cluster. The calculation of Degree Centrality shows that node F391 is the node with the highest interaction, which is 3,203. This indicates that the F391 node interacts the most with residents in its environment within 14 days before and after being positive for COVID-19 based on the date of the swab test.

Table 2. Centrality measurement (degree, betweenness, closeness)

Rank	Node	Degree
1	F391	3,203
2	G471	2,135
3	I221	2,135
4	D091	2,135
5	K151	2,135
6	C351	2,135
Rank	**Node**	**Betweenness**
1	B041	0,006
2	C371	0,006
3	C411	0,005
4	J101	0,005
5	C361	0,005
6	K171	0,005
Rank	**Node**	**Closeness**
1	O102	0,358
2	C362	0,358
3	C363	0,358
4	F022	0,357
5	D102	0,357
6	C372	0,357

The calculation of Betweenness Centrality shows that nodes B041 and C371 are important actors with the highest value of 0.006. Actor B041 becomes an important link. While the calculation of Closeness Centrality there is an error which can only calculate two-way interactions, while the contact tracing data is in the form of a digraph or there is only one-way interaction. (closeness result with ucinet program)

Then the data on the connectedness matrix will be processed into the NetDraw program into a graph (Figure 4 and 5).

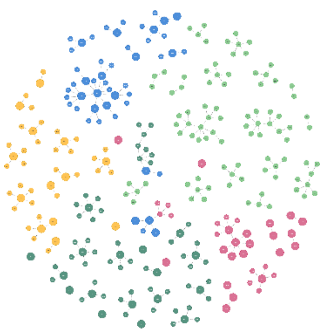

Figure 4 Graph visualization by location. (in the Neighborhood cluster)

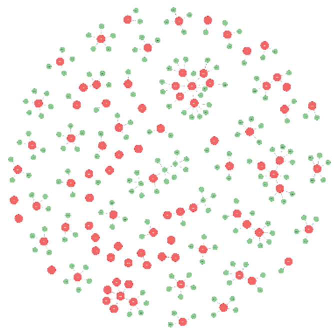

Figure 5 Visualization of contact tracing.

Based on the results of the SNA graph, it can provide an overview of who is indicated to be a close contact. This SNA graph assists health workers in testing close contacts so that it can be immediately identified who are close contacts who have tested positive for Covid-19 and who have not been exposed to Covid-19.

3. CONCLUSION

The use of graphs on SNA provides visualization of the contact network to distinguish potential relationships between close contacts, so that contact tracing can be carried out based on disease path mapping. Compared to the conventional way of reviewing individual contact records, graph visualization on the SNA is a good adjunct tool for healthcare workers in contact tracing investigations. Tissue measurements, particularly measures of centrality, allow researchers to identify key individuals who may be causing or exacerbating outbreaks. In addition, several studies have used SNA to study the dynamics of disease transmission, indicating that the structure of the contact network is a more accurate indicator of the epidemic phase.

AUTHORS' CONTRIBUTIONS

Alvida, conceived and designed the analysis using SNA graph to apply for contact tracking, and wrote paper to submit in international conference

Wildan, collected and contributed data from the resident cluter, simulated data using neo4j to make SNA graph

Lutfi Shahab, other contribution, like supported references and edited writing in paper

ACKNOWLEDGMENTS

The research is supported by the Departement of Mathematics, Institut Teknologi Sepuluh Nopember, Ministry of Education and Culture. The number of reseach contract is 1827/PKS/ITS/2021

REFERENCES

[1] Y. Wu, Z. Duan, Social network analysis of international scientific collaboration on psychiatry research, in International Journal Ment, Health Syst, 9,1, 2015.

[2] A. Klovdahl S., Social networks and the spread of infectious deseases: the AIDS, in: Example, Soc. Sci. Med. 21, 1985, pp 1203–1216.

[3] O. Adegboye A., F. Elfaki, Network analysis of MERS coronavirus within households, communities, and hospitals to identify most centralized and super-spreading in the arabian peninsula, 2012 to 2016, in Hindawi : Canadian Jurnal of Infection Diseases and Medical Microbiology, 9, 2018.

[4] C. Yi-da, C. Hsinchun, K. Chwan-Chuen, Social network analysis for contact tracing, in : Springer Science+Business Media, LLC. Ch. 15, 2011.

[5] N. Hartsfield, G. Ringel, Pearl in graph theory in : a Comprehensive Introduction, Academic Press, Inc., 1990.

[6] C. Prell, Becoming familiar with social networks, in : Social Network Analysis: History, Theory and Methodology, 2012, pp 7–18.

[7] J. Scott, Social network analysis, in : SAGE Publications Ltd, 2000.

[8] I. Etikan, Comparision of snowball sampling and sequential sampling technique, in : Biometrics Biostat, Int.J., 3,1, 2017.