COVID-19 Contact Tracing Application Model Using Graph Theory

Shifa Salsabiila 13519106
Program Studi Teknik Informatika
Sekolah Teknik Elektro dan Informatika
Institut Teknologi Bandung, Jl. Ganesha 10 Bandung 40132, Indonesia
13519106@std.stei.itb.ac.id

Abstract—**The endeavor to eradicate COVID-19 is still ongoing in most countries. Optimization of contact tracing may be key to helping countries return to pre-pandemic situations. This paper shows that contact tracing can be improved through automation and appropriately modeled using graph theory. This paper shows how transmission probabilities can be calculated using an undirected, weighted graph model depending on the duration of interaction between two people. Results of comprehensive contact tracing data can be used to help countries and individuals determine the best decision to help end the pandemic.**

Keywords—**Contact tracing, COVID-19, Disease transmission, Graph theory**

I. INTRODUCTION

2020 came as a shock to probably most parts of the world, due to the unprecedented existence of COVID-19. The virus that first emerged in China back in September 2019 has now struck the world and became a source of chaos and panic. The virus was first officially identified and confirmed in Indonesia on the 2nd of March 2020. At the time, many Indonesians expected the outburst to be temporary and will pass in a short period of time. Many expected that situations would have gone back to normal by July of 2020 as models and forecasts of the virus growth started emerging. However, by the end of 2020 statistics still show an increasing trend of the virus growth in Indonesia.

Transmission of COVID-19 virus was originally believed to only be attainable through direct contact and droplet exchange, however more recent studies have shown that transmission can also occur airborne. Respiratory secretions and droplets such as infected saliva or exhaled air can be transmitted whenever an infected person coughs, sneezes, or simply talks. To make matters worse, it had been discovered that not all carriers of the virus display symptoms. Hence, anyone that may seem to be perfectly healthy may be a carrier and still has the potential to infect others.

This whole pandemic has affected many Indonesian citizens negatively, since they cannot go back to their normal lives, especially those who rely on normal conditions to make a living. The problem is, Indonesia seems to not have yet found an optimal solution that can help bring this pandemic to an end. A lot of cases where people are ignoring the health protocols and are conducting group gatherings can still be found all over the country. It needs to be brought to people's attention and be understood that this virus will not go away by itself, without active participation and action from all elements of the country, whether it be the government or the citizens.

One of the core disease control measure that should be implemented effectively in all countries is contact tracing. Perhaps, one of the main reasons why some Indonesian people still seem ignorant of the health protocols is that they feel as if the virus is a distant concept to them, and that it is unlikely for them to catch it. Here, contact tracing results can be a key to warn people, especially those who have high risks of being exposed, which will then hopefully stop chains of transmissions. Contact tracing itself can be modeled and implemented using a discrete mathematics concept: Graph Theory.

II. GRAPH THEORY

Graph is defined to be a structure that consists of 'nodes' and 'edges'. Nodes, often also known as vertices, represent certain objects in the structure and edges are lines that connect a node to another naode in a graph. A graph G is notated by G = (V,E) where V is a non-empty set of nodes in G. E is a set of edges in G and E can be empty.

A. Types of Graphs

Graphs can be categorized based on multiple classifications. Some of the main classifications will be explained here. This includes simple graphs, multigraphs, null graphs, undirected graphs, directed graphs, and weighted graphs. A simple graph is one where a maximum of one edge lies between two nodes and there is no edge that connects a node to itself (forming a loop).
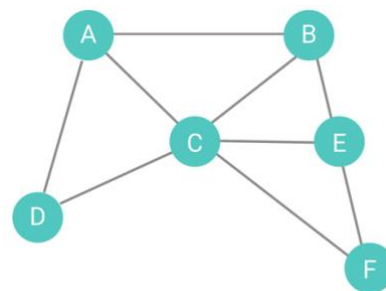

Fig.2.1 Simple graph

A multigraph has either multiple edges connecting two of the same nodes or an edge that forms a loop.
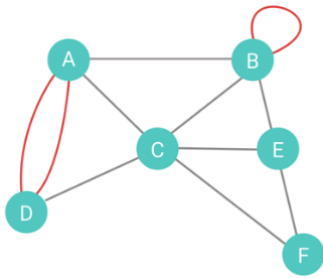
Fig.2.2 Multigraph or non-simple graph

A directed graph is one where the edges are in the form of an arrow, coming out of one node and into its destination node. All of the examplese shown above have been undirected graphs. An undirected graph assumes that the startind and end nodes do not matter, as it only represents the fact that there is a connection between the two nodes.
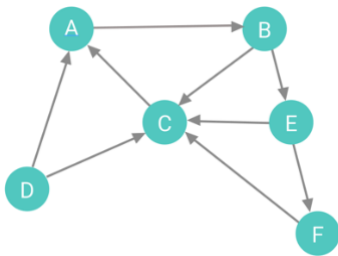

Fig.2.3 Directed graph

It was mentioned above that the set E in a graph G can be empty. This shows the case of an empty graph, where there are no edges connecting any nodes.
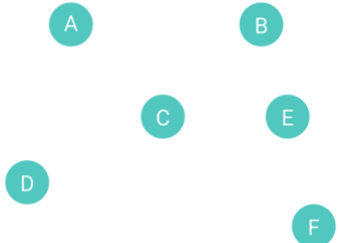

Fig.2.4 Null graph

Lastly, a weighted graph is one where the edges connecting two nodes are assigned a certain value. This can be used to represent multiple things, for example, a graph used to depict a map where the nodes represent places and the edges represent a road from on place to another. Here, a weight in the graph may represent the distance of the 'road' from place A to place B. Graphs that do not explicitly denote the weights of the edges are by default assumed to have homogenous edges of weight 1.
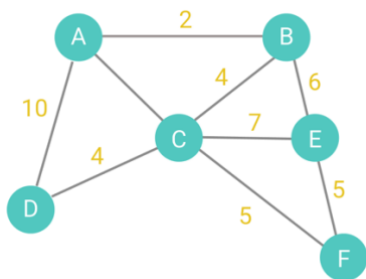

Fig.2.5 Weighted graph

B. Graph Degrees

The degree of a node is the number of edges that are incident on the node. In (*Fig.2.1*) Node A has a degree of 3, whereas node C has a degree of 5. This can be explained by the fact that C is incident on the undirected edges {(A,C), (B,C), (D,C), (E,C), (F,C)}. A loop is accounted for a number of 2 edges. In the case of a directed graph, the degree of a node is classified into two: indegree and outdegree.

The total number of degrees of a graph can be calculated according to the *Handshake Lemma*. The lemma states that the total number of degrees of a graph is twice the number of edges. Hence, it is also derived that a graph will always have an even total number of degrees.

C. Paths and Circuits

A path is what you may call a route that starts at a node and ends at a node. A graph is said to be connected if for any two nodes, there is at least one edge between them. If this requirement is not fulfilled, a graph is said to be disconnected. An example of a path from (*Fig.2.1*) is the path A-B-E-F. This path is said to have a length of 3. In a directed graph, there can only be a path between a node A and a node B if there is an edge coming out of A and into B. For example, in (*Fig.2.3*), there exists a path of length 1 A-B, but the same cannot be said about B-A.

A circuit is a type of path with a special condition, that is the initial node and the final node must be the same. The graph from (*Fig.2.1*) has a circuit B-A-C-B with a length of 3.

In graph theory, there exists special kinds of paths and circuits. These are known as the Euler path and circuit and also the Hamilton path and circuit. A graph G is said to have a Euler path (also known as a semi-euler graph) if there is a path that consists of all of its edges exactly once. Graph G is known to have a Euler circuit (also known as a euler graph) if it has a euler path that start and end at the same node. The graph in (*Fig.2.1*) does not have a euler path or a euler circuit, as there is no possible way to travel through all of its edges exactly once.

Similarly, a Hamilton path is a path that consists of all of its nodes exactly once, except for its starting and end nodes, if they happen to be the same node in the case of a Hamilton circuit. One thing to note here is that a Hamilton path does not have to travel throguh all of the graph's edges. A Hamilton circuit is a Hamilton path that start and end both at the same node. (*Fig.2.1*) does have a Hamilton circuit, that is A-B-E-F-C-D-A. A graph can have be both a Euler and a Hamilton graph. In (*Fig.2.6*), it can be seen that the graph has a Euler circuit: A-B-C-D-E-C-F-E-B-F-A. It also has a Hamilton circuit: A-B-C-D-E-F-A.
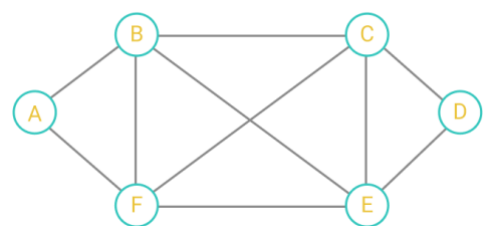

Fig.2.6 Euler and Hamilton graph

D. Graph Representation

Graphs can be represented in several forms. Here, three most common ways of graph representation will be discussed. The first way is using an *Adjacency matrix*. Here, the graph is represented as an n x n matrix, M, with n being the number of vertices that the graph has. In an unweighted graph, if there is an edge present between vertex a and vertex b, then M_{ab} will be set to 1. Otherwise, it will be set to 0. In a weighted graph, then the value set will be equal to the weighted value of the edge. Here is the adjacency matrix for the graph in (*Fig.2.1*).

$$
\begin{array}{c}
 \quad A \; B \; C \; D \; E \; F \\
\begin{array}{c} A \\ B \\ C \\ D \\ E \\ F \end{array}
\left[
\begin{array}{cccccc}
0 & 1 & 1 & 1 & 0 & 0 \\
1 & 0 & 1 & 0 & 1 & 0 \\
1 & 1 & 0 & 1 & 1 & 1 \\
1 & 0 & 1 & 0 & 0 & 0 \\
0 & 1 & 1 & 0 & 0 & 1 \\
0 & 0 & 1 & 0 & 1 & 0
\end{array}
\right]
\end{array}
$$

Fig.2.7 Adjacency matrix of graph (Fig.2.1)

A different way of representing the graph using a matrix is using the *Incidency matrix*. Here, the rows will represent all the nodes of the graph, and the colums will represent all the edges. The matrix element M_{ne} will be set to 1 if node n is incident with edge e and set to 0 otherwise. Here is the incidency matrix of the graph in (*Fig.2.1*).

$$
\begin{array}{c}
 \quad AB \; AC \; AD \; BC \; BE \; CD \; CE \; CF \; EF \\
\begin{array}{c} A \\ B \\ C \\ D \\ E \\ F \end{array}
\left[
\begin{array}{ccccccccc}
1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\
1 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 \\
0 & 1 & 0 & 1 & 0 & 1 & 1 & 1 & 0 \\
0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 1 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1
\end{array}
\right]
\end{array}
$$

Fig.2.8 Incidency matrix of graph (Fig.2.1)

Another common way of representing graphs is using an *Adjacency list*. An advantage of using an adjacency list, especially for representing undirected graphs is that it is not redundant, as opposed to how the adjacency matrix is, as in the case of an undirected graph, the resulting adjacency matrix will simply be a mirror matrix. The adjacency list for the graph in (*Fig.2.1*) is as such.

Table 2.1 Adjacency list of graph (Fig.2.1)

Node (Vertex)	Trailing Nodes
A	B, C, D
B	A, C, E
C	A, B, D, E, F
D	A, C
E	B, C, F
F	C, E

III. METHODOLOGY – CONTACT TRACING MODELLING

A. Contact Tracing Applications

Contact tracing is the process of identification of people who may have come into contact with a person who is or potentially infected by a disease. In this case, by the COVID-19 virus. Currently, in Indonesia contact tracing is only being done whenever a person has been confirmed to be infected with the virus and heavily relies on human memory. Consider cases where an infected person had gone out to a public place prior to being infected by the virus and there, had performed close interaction with strangers whom he/she will not be able to identify later on. Hence, the accuracy of the current performed method of contact tracing in Indonesia can be claimed to be relatively low.

Here, the author proposes a look into a modernized method of contact tracing in Indonesia, that is contact tracing applications. The idea of contact tracing applications have emerged since the early stages of the pandemic outburst, however, not many countries have successfully implemented the solution. The idea of implementing contact tracing applications is to widen the scope of contact tracing, which will provide more sufficient data that can be used in more accurately deciding who should be tested for the virus and who should conduct self isolations at a certain time.

With the implementation of contact tracing through applications, tracking can be done at all times, without having to wait until a person is positively confirmed to be infected or rely on man memory.

B. Contact Tracing and Graph Structure

The data accumulated from the application can be organized using a graph data structure. Here, the most discernible characteristic of the graph would be the nodes, that will represent each individual using the app. Two nodes will be connected through an undirected edge if the two people it represents had been in close contact with each other for a sufficient amount of time. Perhaps, the proposed distance here can be 2 meters and the duration being 15 minutes. Hence, two people simply passing by each other will not considered to be adjacent.

Here, an undirected graph is the ideal type to choose, because whenever two people are in contact with each other for a certain amount of time, then they both have an equal amount of chance of acquiring or transferring the virus, especially in situations where the virus may be present but is still within incubation phase. From this, it is chosen that the most sufficient method to represent the data is using an *adjacency list*. The main reason for this is that since we have decided the graph to be undirected, the use of an adjacency matrix will seemingly be redundant, since the input will be repeated twice, for example on account A-B and B-A.

C. Weighted or Unweighted Edges

Up to now, all forms of interaction above a certain amount of minimum time and minimum distance are all considered equally probable in terms of enabling virus transmission. However, in

reality, time of contact or interaction may play a considerable role in the significance of virus transmission probability between people. For example, two people interacting for 5 hours should have higher probabilities in virus transmission compared to two people interacting for only 15 minutes.
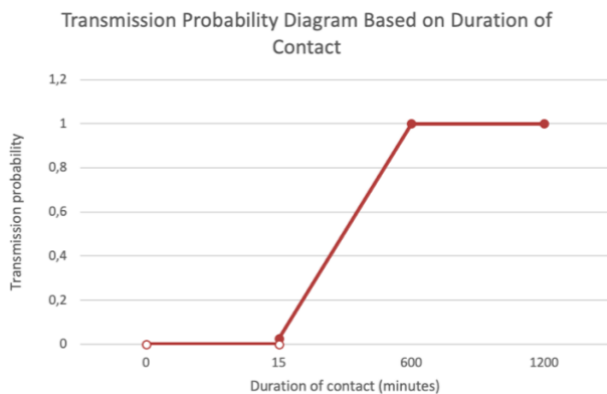

Fig.3.1 Transmission probability diagram based on duration of contact

Therefore, this design proposes that the contact tracing app should also take into account the duration of the interaction. This duration can then be translated to a certain probability value, ranging from 0 to 1. For the sake of simplicity, let the probability calculation model be follow linear function, as shown in (*Fig.3.1*), with any duration of 10 hours or longer be equal to 1 and any duration shorter than 15 minutes be equal to 0 (or simply not accounted). Of course, this model may not be the most accurate way of representing the situation, however, a more accurate probability model will still work with the overall concept.

Hence, within the graph representation of the data, it would be more appropriate to use graph with weighted edges. The value of the weights here can denote the probability of virus transmission between the two nodes connected. Although usually weighted graphs may be used to calculate the shortest path or lowest cost between nodes, here, the analysis implementation will have to differ. Instead, in order to get the probability of infection transmission from a confirmed infected node to a certain other node connected by a path, then all the weights of the edge leading upto the node can simply be multiplied.

Of course, there may be more than one path connecting two nodes, and in this case, the path that yield the highest probability should be taken into account, because in this context, we should always have our guards up for the worst case scenarios. Again, this method of analysis may not be the most representative of solutions, but is still more representative than current methods.

IV. CASE STUDIES

A. Case 1: Calculating Probability of Infection

One of the main reasons why this solution should be implemented is to give individuals a depiction of their chances of getting infected by the virus based on the people they have met. Letting people know of their chances will most likely cause

them to be more considerate in making their daily decisions, especially if it involves meeting and interacting with other people. A simulation of calculating the probability of a person being infected can be seen below.

Consider the graph in (*Fig.4.1*). For this case, we will only focus on nodes A, B, C, and D. Assume that other edges are also weighted, though not shown in the depiction below. In this case, A is confirmed to be infected by COVID-19. From the result of contact tracing, it can be seen that A had been in contact with B for 6 hours, yielding the tranmission probability between them to be 0.6. Other observed interactions include A and C, B and D, and C and D, each yielding transmission probabilities of 0.8, 0.4, and 0.99 respectively.
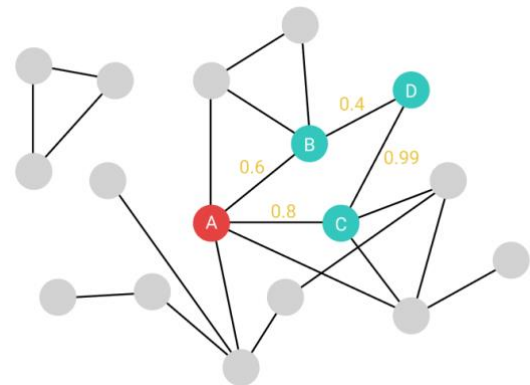

Fig.4.1 Simplified contact tracing graph example

At this point, it can be calculated that from these interactions alone, B has a probability of catching the disease of 0.6 and C has the probability of 0.8. Meanwhile, there are two paths that lead to D, one being A-B-D and the other being A-C-D. From the path A-B-D, the total probability is 0.6 x 0.4 = 0.24. From the path A-C-D, the total probability is 0.8 x 0.99 = 0.792. Here, D should consider taking into account not the lowest probability, but the highest.
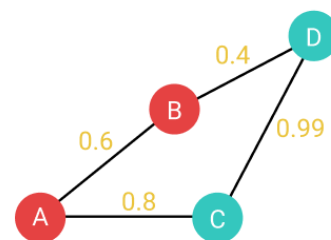

Fig.4.2 Change of closest known infected source in path

Now consider the case in (*Fig.4.2*). Here, it is depicted that perhaps a short-time after A had been confirmed to be infected, B is now confirmed to be infected. Hence, the transmission probability calculation for D will also change. Now, the two paths that lead to D from the closest known infected sources are A-C-D and B-D. The probability from path A-C-D is still the same as before, since C is not a confirmed carrier, however, the probability from the path that was once A-B-D now shortens to B-D and the probability increases from 0.24 to 0.4. Although in this example, the overall highest probability is still yielded from the path A-C-D, in other cases, the highest probability might also change.

B. Case 2: Euler Path and Circuit Significance

In this application of graph theory, the existence of a Eulerian path or circuit is not the main focus and hence does not have that high a significance in the results. The reason for this is that Eulerian paths and circuits are often utilized to find paths that are non-redundant for cases such as the travelling salesman problem. However here, a path's significance is more reliant on its weight and longest path. Perhaps a more suitable element to discuss is the connectivity of the graph (which will be dicussed further at a later part).

Moreover, it is highly unlikely to achieve a Eulerian circuit or even a path as a result of contact tracing data. It is known that in order to achieve a Eulerian path, there can only be either zero or two number of edges with an odd number of degree. Meanwhile here, the number of degree of each node cannot be predicted or modelled in a certain way, because human interaction is very difficult to predict. Hence, it is asserted that in the case of a contact tracing data model using graphs, Eulerian paths and circuits are unlikely to be found and have minimum significance on the overall result.

C. Case 3: Hamilton Path and Circuit Significance

Contradictory to the minimum significance of a Eulerian path and circuit, a Hamiltonian path and circuit has a relatively high significance in this context. The presence of a Hamiltonian path or circuit in a contact tracing model suggests that all nodes (people) that took place in the contact tracing are subject to a risk in transmission. This means that even if only one person in the sample is confirmed to be positively infected by COVID-19, then the entire sample has a probability of being infected as well. This is because in a Hamiltonian path or circuit, it can be infered that there is at least one path that connects every two nodes.

Even if this is so, the chances of acquiring a Hamiltonian path or circuit in this context is still very low. The higher the number of subjects involved in the contact tracing, the lower the chance of acquiring a Hamilton path or circuit. However, if even with a large amount of subjects involved, a Hamiltonian path or circuit is still managed to be acquired, then it should be clear that the situation achieved is extremely alarming and further actions must be taken.

D. Case 4: Regional Lockdowns and Connectivity

A graph is said to be connected if there is at least one path that connects every two nodes. Otherwise, a graph is known to be disconnected. In a disconnected graph, all connected subgraphs are also called connected components of the graph. Consider the graph in (*Fig.4.3*). It can be seen that in the example below that the yellow nodes have no path to the blue nodes and vice versa. Contextually, it means that if there is a confirmed case within the blue connected components, the yellow nodes will have a theoretical transmission probability of zero from the same source.
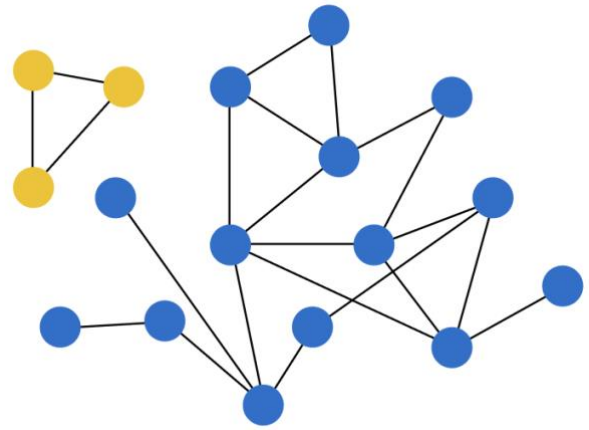

Fig.4.3 Disconnected graph

Such results is actually attainable through successful implementations of regional lockdown. The purpose behind implementing regional lockdowns is to maximize the number of disconnected subgraphs. Theoretically, it will definitely be easier to supress the virus spread in small, controlled regions, compared to trying to manage in one large region. However, it is always easier said than done.

V. IDEAL RESULTS AND SUGGESTIONS

When it comes to ideal results in the case of disease transmission models using graphs, the best case scenario will unquestionably be a null graph, where there is no interaction between subjects that are prone to being infected. However, also without question, this is an impossible result, as human interaction cannot be entirely abolished. Then, moving on to the second most ideal result according to the author is the maximization of disconnected subgraphs within the graph. This, as discussed in subsection D of chapter IV is a more realistic and attainable result.

In the case of disease transmission, there will be individuals who serve as 'bridges', connecting one community to another. These individuals will widen the scope of the graph and are dangerous to the objective of virus control. Perhaps it may be interesting to come up with strategies to educate these individuals regarding their actions. It is possible that one of the reasons that there still exists such individuals out there, especially in Indonesia is that they are not aware that what they are doing is playing a significant role in the spread and sustainability of the virus.

A possible solution to educate such bridges is to show them directly possibly through the application that they are playing a major role in the spread of the virus. Adding features such as action suggestions can also be possible. For example, if a person is calculated to have a current probability of transmission of over 0.5, then perhaps the application can suggest what the person should do next, whether it be self isolation or visiting a local testing center to get themselves tested.

VI. Model Limitations

Though it may be an improvement from the current methods of contact tracing in Indonesia, this method of contact tracing is still far from perfection. There are of course still many cases that cannot be taken into account by this model. One of these is the fact that transmission of COVID-19 may not occur directly. This means that there are cases that have been proven where the virus remains on inanimate objects for a certain period of time and transmission can also occur from this without two people even having to physically meet. This situation cannot be represented on the graph, because it is as if the edge between these two people do not actually exist.

As far as the discussion of this model here, it has not been taken into account how long an edge should be maintained in the data structure. This should be considered and decided, hence if two nodes had gone by a certain period of time without being confirmed positive of the virus, then it should be safe enough to remove the edge between them. Otherwise the relevance of the validity of the graph will diminish. However, further research regarding the virus itself needs to be conducted to be able to decide a certain amount of time.

Another limitation of such solution is more closely related to its real life implementation. In reality, and may also be one of the reasons why contact tracing applications have seemingly not been successful in any country right now is that it is quite difficult to manage a large amount of citizens and expect each and every one of them to strictly use the contact tracing apps. Policies as such most likely require years of socialization and practices until it can work as intended.

VII. Conclusion

Applications of contact tracing should be optimized to help countries put an end to the ongoing pandemic. The current most-used method of manual contact tracing has many flaws and loop holes that can be improved by automation, although even so the implementation itself will still be faced with several obstacles. This paper suggests an enhanced method of contact tracing by utilizing mobile applications. This paper presents an alternative method of modelling contact tracing data results using graph data structure. It was suggested that an undirected, weighted graph to be used to represent more accurate results. It was discussed that the value of the edge weights represent the probability of transmission between two people, depending on the duration of their interaction.

This paper suggests that countries should continue to strive for disconnectivity regions. Results from a comprehensive contact tracing data can be used for multiple purposes, among them are making better decisions on who to select for undergoing testings, suggestions on who should conduct self isolations and simply to remind people to be more mindful of their actions amidst this pandemic.

The illustrations and models in this paper is deliberately kept very simple to depict the core concept of contact tracing that incorporates graph thory. The author believes that this topic still has room for a lot of improvement, especially within the area of more sophisticated epidemiological models.

VIII. Acknowledgment

References

[1] A. Chakraborty, A. Nath, S. Mondal, T. Dutta, "Application of Graph Theory in Social Media"

[2] A. D. Filippo, A. D. Matteo, F. Blanchini, G. Giordano, M. Colaneri, P. Colaneri, R. Bruno, "Modelling the COVID-19 epidemic and implementation of pupulation-wide interventions in Italy"

[3] A. Dietrich, B. Born, G. Muller, "The effectiveness of lockdowns: Learning from the Swedish experience"

[4] A. Gimma, A. J. Kucharski, C. I. Jarvis,, J. D. Munday, J. Hellewell, N. I. Bosse, S. Abbott, T. W. Russell, W. J. Edmunds, "Feasibility of controlling COVID-19 outbreaks by isolation of cases and contacts"

[5] A. Johnson, A. Paleyes, B. Elesedy, B. He, G. Harling, M. Hutchinson, S. Zaidi, Y. W. The, "Technical Document 3: Effectiveness and Resource Requirements of Test, Trace and Isolate Strategies"

[6] D. Rorres, D. E. Zelner, G. Smith, J. A. Miller, J. M. Mossey, M. Romano, T. H. Grubesic, "Contact tracing for the control of infectious disease epidemics: Chronic Wasting Disease in deer farms"

[7] D. Sridhar, D. Fanian, G. Fernandes, M. S. Pacheco, "A Review of International Approaches to Test, Trace, Isolate"

[8] E. Crisostomi, E. Dudkina, L. Stone, M. Bin, P. Ferraro, R. Murray-Smith, R Shorten, S. Yilmaz, T. Parisini, "Kemeny-based testing for COVID-19"

[9] F. Saleh, S. Kharroubi, "Are Lockdown Measures Effective Against COVID-19?"

[10] L. Ferretti, C. Wymant, M. Kendall, L. Zhao, A. Nurtay, L. Abeler-Dörner, M. Parker, D. Bonsall, and C. Fraser. "Quantifying SARS-CoV-2 transmission suggests epidemic control with digital contact tracing"

[11] Munir, Rinaldi."http://informatika.stei.itb.ac.id/~rinaldi.munir/"

[12] Q. Bi, Y. Wu, S. Mei, C. Ye, X. Zou, Z. Zhang, X. Liu, L. Wei, S. A. Truelove, T. Zhang, W. Gao, C. Cheng, X. Tang, X. Wand, Y. Yu, B. Sun, S. Huang, Y. Sun, T. Feng, "Epidemiology and transmission of COVID-19 in 391 cases and 1286 of their close contacts in Shenzen, Chine: a retrospective cohort study"

Pernyataan

Dengan ini saya menyatakan bahwa makalah yang saya tulis ini adalah tulisan saya sendiri, bukan saduran, atau terjemahan dari makalah orang lain, dan bukan plagiasi.

Jakarta, 4 Desember 2020

Shifa Salsabiila - 13519106