SOCIAL NETWORK ANALYSIS OF SPATIAL HUMAN MOBILITY BEHAVIOUR IN INFECTIOUS DISEASE INTERACTION: AN EXPLORATORY EVIDENCE OF TUBERCULOSIS IN MALAYSIA

Ilham Abdul Jalil[1], Abdul Rauf Abdul Rasam[1]

[1] Centre of Studies for Surveying Science and Geomatics, Faculty of Architecture, Planning and Surveying, University Teknologi MARA, Malaysia

KEY WORDS: Social Network Analysis, Geospatial, Human Mobility, Infectious Disease, Tuberculosis

ABSTRACT:

The movement of individuals between specific locations and the different group contacts of people is essential to predict the future movement and interaction pattern of infectious diseases. Previous studies have shown major factor of infectious disease spread comes from human mobility because a complex and dynamic network of spatial interactions between locations such as the mobility formed by the daily activity of people from place to place. To better understand the such human mobility behaviour, innovative methods are required to depict and analyse their structures by using social network analysis (SNA). This paper aims to investigate the social network structure of selected tuberculosis (TB) case in Klang, Selangor as actors (nodes), and then human mobility (home-work place) data as edge generally used to investigate social network mobility structures and analyse relation among the nodes and study their edges in term of their network centrality. The main finding has revealed that the higher the centrality (degree and betweenness) of a node in the network structure, the higher the chance the node influencing the TB spread in the whole network, after comparing the network graph result with the geographic information system (GIS) mapping approach. Most of the result share the similar result where most of high infection of TB are located in urban and crowded areas. The SNA is a practical knowledge of the social system and contact structure of a community that can therefore provide crucial information to predict outbreaks of infectious diseases in a dynamic spatial phenomena.

1. INTRODUCTION

1.1 Background

The epidemic dynamics of infectious diseases vary among cities, but it is unclear how this is caused by patterns of infectious contact among individuals. Here, we ask whether systematic differences in human mobility patterns are sufficient to cause inter-city variation in epidemic dynamics for infectious diseases spread by casual contact between hosts. According to an individual-based model of airborne pathogen transmission parametrized with mobility data, systematic variation in mobility patterns is sufficient to trigger significant differences in infectious disease dynamics among cities, even among cities of similar size (Dalziel, Pourbohloul, & Ellner, 2013). This suggests that differences among cities in host contact patterns are sufficient to drive differences in infectious disease dynamics and provides a framework for testing the effects of host mobility patterns in city-level disease data (Dalziel, Pourbohloul, & Ellner, 2013).

The movements of individuals between specific locations and the contacts between different groups of people are essential in modeling disease spread (Eubank, et al., 2004). The daily activity of people from place to place forms a complex and dynamic network of spatial interactions between locations. Many pathogens spread through host populations via social interactions (Altizer, Nunn, & Thrall, 2003) thus, knowledge of a community's social system and contact structure can provide crucial information for predicting infectious disease outbreaks. Human mobility patterns generate the proximity between individuals prerequisite for the transmission of many infectious diseases. This suggests that cities with different mobility patterns may also differ in the rate at which their inhabitants have infectious contact, leading to variation among cities in the risk of an epidemic (Merler & Ajelli, 2010).

TB is the world's leading cause of infectious mortality, responsible for 1.6 million annual deaths, according to the World Health Organization (WHO) (WHO, 2018). As reported in the news, TB cases in Malaysia recorded higher mortality rates than expected by the world health organization (WHO). (Marzuki, 2019) there were 26,168 active TB cases in the country involving the death of 2098 cases. This suggests that the need for strengthening in controlling and more comprehensive case detection to detect new cases. The Malaysian health ministry has outlined some guidance in controlling the disease but it needs to be strengthened with more relevant and up-to-date methods.

To better understand the such human mobility behavior, innovative methods are required to depict and analyse their structures by using SNA (Pagan, 2019). The use of social network analysis in the study of human mobility is still lacking in current studies. In fact, the human relationship of mobility to the spread of infectious diseases is the key to any type of disease that belongs to human-borne illness such as TB and Covid-19. One way to control or analysis the spread of this disease is by knowing the behaviour of human mobility. SNA focuses on the structure of ties within a set of social actors or nodes.

Inter-individual connections leading to pathogen transmission can be expressed by networks where each node represents an individual and node edges reflect interactions, enabling pathogen transmission (Rushmore, Caillaud, & Matamba, 2013). Network analysis creates a realistic strategy for mathematically formalizing host contact variation transmission pathways. Furthermore, SNA may identify potential super spreaders, individuals with disproportionately high rates of interaction who may be targeted for vaccine, care or alienation (Lloyd-Smith, Schreiber, & Kopp, 2005).

The main problem of this study was to improve the quality of TB case detection by using SNA from TB human mobility pattern integrating with geographical feature. Geospatial is a crucial tool where to analyst and visualize the TB cases in efficiency compared to current method which is the clinical diagnosis of individuals in a population. Then creating TB risk mapping hotspots using GIS to identify potential TB risk area. In order to prove this is true and significant, this study needs to be carried out with the help of SNA and integrated with the GIS to make this study more realistic and easier to interpret. The majority of social network studies use either whole socio-centric networks or egocentric study designs. Whole network studies examine the relationships between individuals or actors that are viewed as bounded or closed for analytical purposes, despite the fact that the network's boundaries are permeable or ambiguous in reality (Scoot, 2000).

For this study, network analysis is being applied especially in egocentric perspective, sociology theory or other social and sciences may offer reasons for believing that the structure of ties is related to the actions, beliefs or social position of the network participants. Egocentric is relevant to study each individual actor and understanding their whole network system (De Brun & McAuliffe, 2018). Centrality is an important concept to study degree, betweenness and closeness within SNA. Understanding the formation of social ties, the structure and function of networks and the associated mechanisms that link them to health or health behaviours can be extremely relevant when allocating limited resources or targeting interventions in low and middle income countries for public health and economic development (Christakis, 2004).

1.2 Study Area

This study is limited to the TB cases in the area which is in Klang district of Selangor, Malaysia. Klang is one of the most populous city besides Kuala Lumpur with the total number 879,867 (Geonames, 2019) and density 1,298/km^2 (Wikipedia, Klang (City), 2019). Klang was selected as the study area because it recorded a constant number of TB cases in Selangor. Within the Klang district, various demographic features are unbalanced and there are various types of residential areas which are the rural, sub-urban and industrial areas. The rapid economic growth in these areas is leading to increased population density, while inadequate hygiene and sanitation, as well as poor ventilation quality, are the key factors contributing to the district's high number of tuberculosis cases (Jalil, 2021).

2. METHODOLOGY

2.1 Data Collection

Datasets of TB cases collected from the Selangor State Health Department in excel format in the year of 2017 and 2018. The data give an information about the total number and list of cases, demographic status, patient permanent address and patient work address where this address will pin point into google map in order to know their geographic location on a map. Analysis of TB patient mobility based on the patient's home address and work address. Then the two addresses will grouped in with several locality areas to easier to see and analyse. Human mobility needs origin-destination (OD) to analyse the effect of outside exposure to the infectious disease. Knowing the patient home location and work location is preliminary assumption to know how the geographical condition can influence to this spread of the disease.

2.2 Data Processing

2.2.1 Geographic Information System (GIS)

Data will process by using ArcGIS to create a shapefile, digitize to create TB distribution maps and visualize and map out the potential hotspot area of TB and displayed in Choropleth map. The permanent patient address and patient working address were geocoded in google map before importing it into ArcGIS software 10.3.1. Point shapefile layer was created as TB case distribution. The basemap was digitized from ArcGIS online basemap. The basemap digitized covered only Klang district was created as a polygon shapefile layer and this district then was separated by several localities areas.

Since this study was only conducted in Klang district, some of the point shapefiles is outside the Klang district had went through the geoprocessing clip process to remove irrelevant data points. Intersect geoprocessing was done between polygon and point layer where this process to compute the geometric intersection of the input features. Features or portions of the features which overlap in all layers and was written to the output feature class. To represent geographic features on the basemap (localities) shapefile layer, symbology was carried out according to values selected from the layer attribute table as shown in figure 1 below.
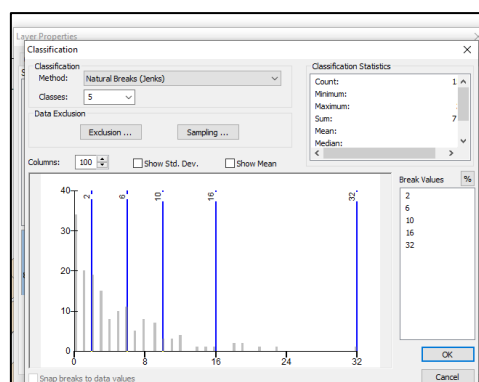

Figure 1. Classification of TB localities cases.

The choropleth map of TB hotspot area as shown in figure 2 below are mostly focused on the central part of Klang district. The blue point represents TB distribution cases and raster grouping values was created into classes in the form of 5 quantile classes. Darker colour grouping represent a higher risk area of TB cases where light colour represent a low risk area of TB cases.
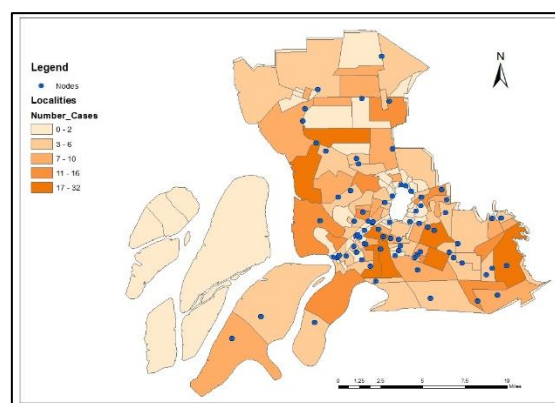

Figure 2. Choropleth Map of TB Hotspot Area.

The level of risk for point (nodes) shapefile layer which fall under specific localities classification was ranked into five scales from level 1 which is the low risk to level 5 indicates high risk as shown in table 1 below. This level of risk was used as weightage data in SNA. Locality falls under 0-2 TB cases category was ranked 1 for low risk area while locality falls under 17-32 TB cases category was ranked 5 for high risk area.

Number of TB Cases Categories	Rank
0-2	1
3-6	2
7-10	3
11-16	4
17-32	5

Table 1. Weightage indicator for SNA.

Figure 3 below shows the process to classify the weightage based on the number of cases falls into several TB case localities.

Figure 3. Weightage Classification.

2.2.2 Social Network Analysis (SNA)

SNA is used to quantify relationships between social entities and is especially useful for determining social relationships that affect disease outcomes or health interventions (Wasserman & Faust, 1994). Social network used to measure power and network activity for a node by using analyse their network centrality. Centrality is a very important indicator because it shows which node occupies a key position in the entire network. The concept of degrees which can be describe like the number of direct connections a node has.

The central role node is invariably associated with widespread recognition and a strong network reputation. Degree centrality, betweenness centrality and closeness centrality are among the popular detailed measurements. As a social actor's centrality increase, node moves closer to the network's core, gaining more strength, leverage, and convenience of the network (Sparrowe, Liden, Wayne, & Wayne, 2001) (Hochberg, Ljungqvist, & Lu, 2007).

Degree Centrality

Degree centrality is measured by the total amount of direct links to the other nodes in a network graph. The nodes with higher degree is more central. Figure 4 below depicts network degree centrality, where red node has a high degree centrality.
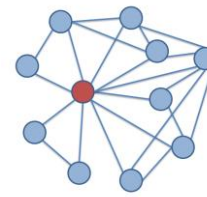

Figure 4. Degree Centrality Network Graph.

Betweenness Centrality

Betweenness centrality is the variation in the betweenness centrality of nodes divided by the maximum variation in betweenness centrality scores possible in a network of the same size. If one node is the only way for other nodes to communicate, bind, transport, or transact, then it is likely that this node is essential and has a high betweenness centrality (Freeman, 1977) as shown in figure 5 for nodes A and B.
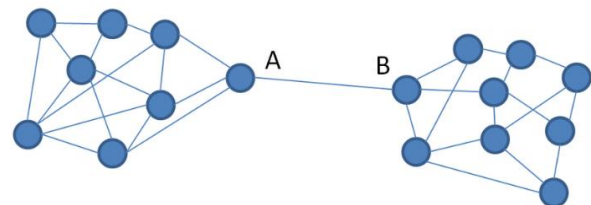

Figure 5. Betweenness Centrality Network Graph.

The OD from excel file will sort by following SNA software format. In this SNA study, Gephi software was applied. In order to import file, the file must separated with two files where one is for nodes file and the other is in edge file. The node edge consists of 149 localities. Then the process of loading up that network. With open two different files in the Gephi software and getting data in the right format. The software importing an adjacency list so it can populate the edges correctly.

The imported data is in a comma separated list where have a comma between words. Figure 6 below shows a table of edges after import into the Gephi where consists of the human mobility data. Source column represents the patient's permanent home address while target column represents patient work address. Those data is compulsory in creating OD for analysing human mobility in SNA.

Figure 6. Human Mobility in Edges Data Table.

2.3 Data Visualisation of Network Graph

Create a social network structure of the selected TB case as actors(nodes) by using SNA open-source software then human mobility (OD) data as edge generally uses to investigate social network mobility structures and analyse relation among the actors (nodes) and study their edges in term of their network centrality. Figure 7 below shows black dots where this dots represents TB distribution after completed the importing process from the node file. While figure 8 below shows the data imported from the edge file emit a line between the nodes. This line has a direction since this edge represents human mobility (OD).
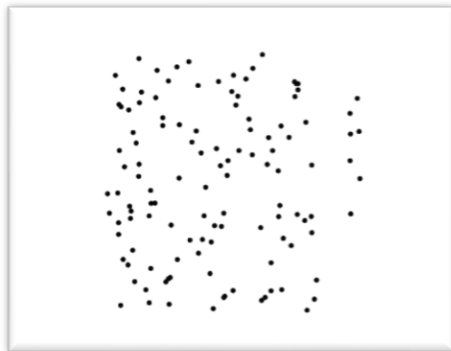

Figure 7. Example of Nodes Graph.

Figure 8. Example of Edges Graph.

3. RESULTS AND DISCUSSION

Figure 9 illustrate the whole sociogram of TB case relations with TB patient mobility. The nodes positions are random at first it was a slightly different representation of set layout algorithms. Graphs are layouted with "Force Atlas" algorithms for clear visualisation where linked nodes attract each other and non-linked are pushed apart. The figures depicts the sociogram of TB case where nodes are represented by circles and lines connecting the node represent edges. The network structure consists of 120 nodes and edges consists of 148 lines with all in a directed graph. The node size presents the value centrality attribute of highest TB locality while the red colour of node represent the high centrality and a blue node represents lowest centrality. The colour of edges based on mobility attributes weightage where red is the highest weightage and blue is lowest weightage. Edge weightage is represented when there are several trips between these two nodes(localities).
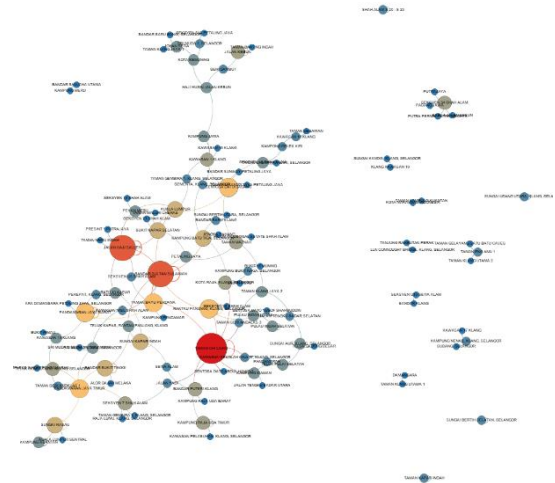

Figure 9. Network graph of TB patient mobility.

3.1 High Degree Centrality of TB Mobility Localities

Network graph as shown in figure 10 below illustrate the localities that have higher degree centrality where Taman Chi Liung, Jalan Haji Salleh and Bandar Sultan Sulaiman are among the top three localities with high degree centrality with bigger nodes and in red and dark orange colour.
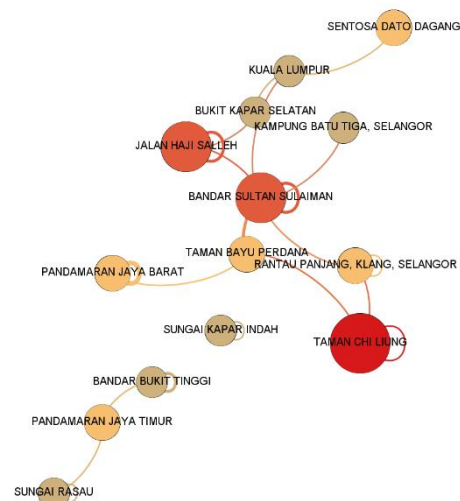

Figure 10. Degree Centrality Network Graph.

The table 2 shows, 10 out of 120 TB localities which have highest degree centrality. The table also shows the value of degree centrality for each of the top ten localities in decending order from the top is the highest value to the lowest value.

TB Localities	No. of Case Categories	Rank	Degree Centrality
Taman Chi Liung	7-10	3	11
Jalan Haji Salleh	11-16	4	9
Bandar Sultan Sulaiman	11-16	4	9
Pandamaran Jaya Barat	3-6	2	6
Pandamaran Jaya Timur	0	0	6
Taman Bayu Perdana	17-32	5	6
Rantau Panjang, Klang, Selangor	3-6	2	6
Sentosa Dato Dagang	11-16	4	6
Sungai Kapar Indah	7-10	3	5
Bandar Bukit Tinggi	17-32	5	5

Table 2. Comparison between Rank and Degree Centrality.

Figure 11 depicts, the cartesian graph to visualize degree centrality distribution. The horizontal axis represents value for each distribution and the vertical axis represent the total amount of distribution value.
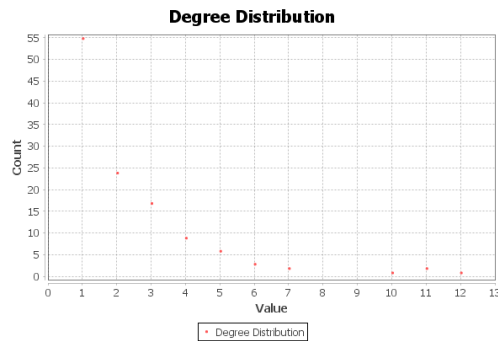

Figure 11. Degree Distribution in Network Structure.

The integration of degree centrality node and TB hotspot area was depicts in figure 12 below. The node consists of top ten localities with highest degree centrality value. The nodes mostly located in the darker colour boundary which in the central of the map.
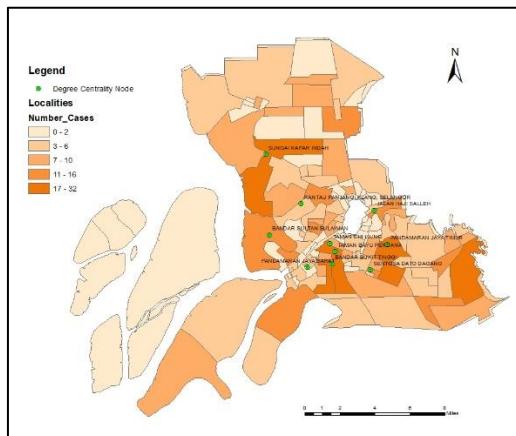

Figure 12. The relationship of degree centrality node and TB hotspot area.

3.2 High Betweenness Centrality of TB Mobility Localities

Network graph as shown in figure 13 below depicts the localities that have higher betweenness centrality where Taman Chi Liung, Batu 10 Kapar and Jalan Haji Salleh are among the top three localities with high betweenness centrality with bigger nodes and in red, dark orange and light orange colour.
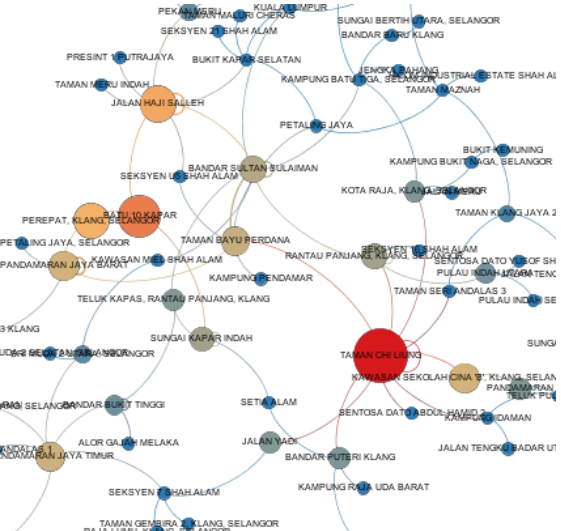

Figure 13. Betweenness Centrality Network Graph.

The table 3 below shows the value of betweenness centrality for each of the top ten localities in decending order from the top is the highest value to the lowest value.

TB Localities	No. of Case Categories	Rank	Betweenness Centrality
Taman Chi Liung	7-10	3	68.5
Batu 10 Kapar	0-2	1	48
Jalan Haji Salleh	11-16	4	39
Perepat, Klang, Selangor	7-10	3	37
Kawasan Sekolah Cina 'B', Klang, Selangor	7-10	3	28.5
Pandamaran Jaya Barat	0	0	0
Pandamaran Jaya Timur	3-6	2	28
Taman Bayu Perdana	0	0	27
Bandar Sultan Sulaiman	17-32	5	26
Rantau Panjang, Klang, Selangor	11-16	4	23
	3-6	2	20

Table 3. Comparison between Rank and Betweenness Centrality.

Figure 14 depicts, the cartesian graph to visualize betweenness centrality distribution. The horizontal axis represents value for each distribution and the vertical axis represent the total amount of distribution value.
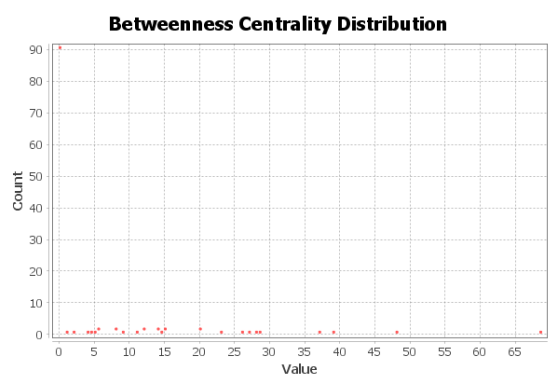

Figure 14. Betweenness Centrality Distribution in Network Structure.

The integration of betweenness centrality node and TB hotspot area was depicts in figure 15 below. The node consists of top ten localities with highest betweenness centrality value. The nodes mostly located in the darker colour boundary which in the central of the map. There are also several nodes located in the light colour boundary. It mean that these areas acts as a bridge along the shortest path between two other nodes and it

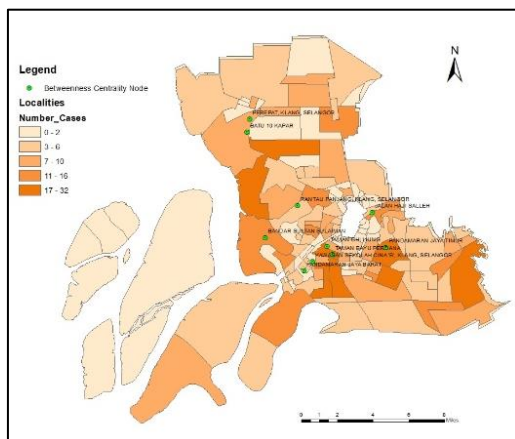acts as a link to other nodes that are far away in other boundary that influential to transmit the TB disease.

Figure 15. The relationship of betweenness centrality node and TB hotspot area.

3.3 Discussion

The higher degree and betweenness centrality in these localities are getting closer to the centre of network that higher power to influence the TB disease spread or become a potential superspreaders. By comparing degree and betweenness centrality by using SNA method with GIS, most localities proves the hypothesis that the higher the degree and betweenness centrality of a node in the network structure, the higher the chance the node influencing the TB spread in the whole network.

4. CONCLUSIONS

SNA is based on theoretical constructs of sociology and mathematical foundations of graph theory where it is used to study of structure and to know how it influence health. The fundamental premise in social network research is that network structure and its properties have a direct impact on the outcome of interest, especially the case study of TB in Selangor, Malaysia. Klang was chosen as the study area as this place is one of the areas with a high distribution of TB cases in Selangor. Besides, it is one of the most rapidly developing areas and there are factors that lead to the occurrence of TB disease in the area such as geography and socioeconomic factor. The study designs are focused on individual characteristics/behaviours and how those characteristics affect health typically collect and analyse attribute data. The higher the centrality (degree and betweenness) of a node in the network structure, the higher the chance the node influencing the TB spread in the whole network. Demonstrate geovisual analytics and exploratory analysis of mobility data include the use of static maps. The findings of the SNA of human mobility stated factor is to proof that human mobility can influence infectious disease by integrating with GIS mapping.

ACKNOWLEDGEMENTS

REFERENCES

Altizer, S., Nunn, C., & Thrall, P. (2003). Social Organization and Parasite Risk in Mammals: Integrating Theory and Empirical Studies. *Annual Review of Ecology, Evolution, and Systematics*.

Christakis, N. (2004). Social networks and collateral health effects. *British Medical Journal*.

Dalziel, B. D., Pourbohloul, B., & Ellner, S. P. (2013). Human mobility patterns predict divergent epidemic dynamics among cities. *Proceedings of the Royal Society B: Biological Sciences*.

Dalziel, B., Pourbohloul, B., & Ellner, S. (2013). Human mobility patterns predict divergent epidemic dynamics among cities. *Proceedings of the Royal Society B: Biological Sciences*.

De Brun, A., & McAuliffe, E. (2018). Social network analysis as a methodological approach to explore health systems: A case study exploring support among senior managers/executives in a hospital network. *International Journal of Environmental Research and Public Health*.

Eubank, S., Guclu, H., Kumar, V., Marathe, M., Srinivasan, A., & Toroczkai, Z. (2004). Modelling disease outbreaks in realistic urban social networks. *Nature*.

Freeman, L. C. (1977). A Set of Measures of Centrality Based on Betweenness. *Sociometry*, 35-41.

GeoNames. (2019). *World Population Reveiw*. Retrieved from World Population Review: http://worldpopulationreview.com/countries/malaysia-population/

Geonames. (2019, November 3). *World Population Review*. Retrieved from World Population Review: http:/worldpopulationreview.com/countries/malaysia-population/

Hochberg, Y., Ljungqvist, A., & Lu, Y. (2007). Whom you know matters: Venture capital networks and investment performance. *Journal of Finance*, 251–301.

Jalil, I. A. (2021). Susceptibility Risk Index of Tuberculosis Epidemics: A Socio-Spatial Risk Factor Analysis in Klang, Selangor. *Malaysian Journal of Public Health Medicine*.

Lloyd-Smith, J., Schreiber, S., & Kopp, P. (2005). Superspreading and the effect of individual variation on disease emergence. *Nature*.

Marzuki, N. M. (2019, March 13). (A. Sahib, Interviewer)

Merler, S., & Ajelli, M. (2010). The role of population heterogeneity and human mobility in the spread of pandemic influenza. *Proceedings of the Royal Society B: Biological Sciences*.

Pagan, N. (2019, December 3). *Behavioural & Social Sciences*. Retrieved from social sciences nature: https://socialsciences.nature.com/posts/55507-can-social-network-structure-reveal-human-behavior

Rushmore, J., Caillaud, D., & Matamba, L. (2013). Social network analysis of wild chimpanzees provides insights for predicting infectious disease risk. *Journal of Animal Ecology*.

Scoot, J. (2000). Social network analysis: a handbook. In J. Scoot, *Social network analysis: a handbook* (pp. 120-124). Newbury Park: Sage.

Sparrowe, R. T., Liden, R., Wayne, S., & Wayne, S. J. (2001). Social networks and the performance of individuals and groups. *Academy of Management Journal*, 316-325.

Wasserman, S., & Faust, K. (1994). *Social Network Analysis Methods and Application.* Cambridge University Press .

WHO, W. (2018). WHO - The top 10 causes of death. *24 Maggio*.

Wikipedia. (2019, November 5). *Klang (City)*. Retrieved from www.wikipedia.com:
https://en.wikipedia.org/wiki/Klang_(city)

Wikipedia. (2019, November 5). *Klang (City)*. Retrieved from www.wikipedia.com:
https://en.wikipedia.org/wiki/Klang_(city)