



Faculty of Computer Science and Information Technology

Identifying COVID-19 Hotspots using Bipartite Network Approach

Hong Boon Hao

**Master of Science
2023**

Identifying COVID-19 Hotspots using Bipartite Network Approach

Hong Boon Hao

A thesis submitted

In fulfillment of the requirements for the degree of Master of Science

(Computational Science)

Faculty of Computer Science and Information Technology

UNIVERSITI MALAYSIA SARAWAK

2023

DECLARATION

I declare that the work in this thesis was carried out in accordance with the regulations of Universiti Malaysia Sarawak. Except where due acknowledgements have been made, the work is that of the author alone. The thesis has not been accepted for any degree and is not concurrently submitted in candidature of any other degree.



.....

Signature

Name: Hong Boon Hao

Matric No: 19020107

Faculty of Computer Science and Information Technology

Universiti Malaysia Sarawak

Date: 01 Mar 2023

ACKNOWLEDGEMENT

I would like to thank my supervisor, Professor Dr Jane Labadin for her constructive advices, guidance and encouragements throughout my years of study. Also, I am indebted to her supportive and excellent supervision. She is there for me every time I need advices. Our countless sessions of discussion have always been enlightening, motivating, assuring, and fruitful in provoking new thoughts and evaluations to existing ideas. I am thankful for the invaluable experiences I gained from her through conducting the research writing of scientific papers, coordinates the projects, preparing presentations and attending conferences. Thanks for accepting me as who I am, helping me to accomplish my Master endeavour and provided me precious opportunities in this journey.

Special thanks to Universiti Malaysia Sarawak for the Zamalah Graduate Scholarship which has provided me with monthly allowance. Besides, special thanks are also extended to Ministry of Higher Education Malaysia as appointed me as research assistant for PGRS/1/2020/ICT05/UNIMAS/0201 and GL/F08/KPT/02/2020. This made my dream to pursue a Master's degree a reality. My heartfelt thanks go to the Faculty of Computer Science and Information Technology for providing me with a conducive environment. Thanks to Dean and Deputy Deans for the wholehearted support. Thanks to all lecturers especially lecturers from Computational Science programs for your insightful advices. I also acknowledge the administrative staff members of the faculty and the Centre for Graduate Studies for their constant help. I would also like to also deliver my sincere appreciation to the Bintulu Division Health Office and Dr Melvin for the help and support in providing the research data generously. Thank you.

Thanks to my thoughtful friends for spending their precious time proofreading the draft of this thesis. Kian Soon, thank you for your generosity in proofreading my literature reviews, advises and comments on my writing. Yun Ta and Hwee Jia, thank you for your help in proofreading my draft and for providing me with constructive suggestions. Thanks for your friendship, warm advice and great support. Vei Ling, my best roommate, thank you for always giving me support during my research journey. Yong Xian, Ting Sun, Mei Wen, Chee Xian and many others, thank you for your time, understanding and motivation. I am blessed to have all of you along with me throughout my research journey. Thanks to all my friends who greet me happy and give me support and encouragement. I am unable to list everyone but you are always on my mind.

Last but not least, I am deeply grateful to my dearest family members. Thanks to my parents for raising me with a love of knowledge and lovingly encouraging me to strive for my master's dream. Also, thanks to both my sisters for their endless support in motivating me. My family contributed significantly to the accomplishment of my master's degree. Thank you. Dedicated to the memory of my grandmother, who always believed in my ability to be successful in the academic arena. You are gone but your belief in me has made this journey possible.

ABSTRACT

The COVID-19 pandemic has affected countries worldwide, causing major disruptions in both health and economic systems. While lockdowns have been effective in controlling the spread of the virus, their negative impact on the economy has prompted the need for alternative, cost-balanced control measures. Contact tracing has emerged as a promising solution in identifying community outbreaks of COVID-19. To improve the efficacy of contact tracing, this study aimed to formulate a contact network model for COVID-19 transmission. Conventional approaches were found to be inadequate in modelling the transmission of COVID-19, particularly in identifying the source of infection. To address this, the study utilized a bipartite network modelling approach to account for the heterogeneity of transmission routes, human hosts, and visited locations. The human host and visited location were identified as the two discrete entities in the research scenario. Using data from the Bintulu Health Office's contact tracing investigation forms, six network models were formulated. The link weight between the human host and location nodes was quantified using the summation rule, taking into consideration various factors such as environmental properties, building characteristics, human and pathogen characteristics, and transmission modes. The location and human nodes were then ranked using a web-based search algorithm based on their respective ranking values. The results of the study showed that the bipartite network modelling approach was successful in formulating the contact network model. Verification analysis revealed a root mean square error of 0.0002244 and 0.001419 for the location and human nodes, respectively, which were well within the threshold value of 0.05. The ranking between the target and validated models was found to have strong similarity with a good Spearman's rank correlation coefficient of above 0.70 ($p < 0.001$), indicating a high degree of relevance in improving contact tracing for COVID-19. The study also found

that all parameters used in the model were relatively significant, and that the model had the ability to predict potential hotspots with 90% accuracy within a 600m radius for the subsequent week. These findings highlight the potential of the bipartite network modeling approach in improving contact tracing for COVID-19 and reducing the spread of the virus in high-risk areas.

Keywords: Bipartite network modelling, COVID-19, hotspot, ranking

Mengenal pasti Titik Panas COVID-19 menggunakan Pendekatan Rangkaian Dwipartit

ABSTRAK

Pandemik COVID-19 telah menjejaskan negara di seluruh dunia, menyebabkan gangguan besar dalam kedua-dua sistem kesihatan dan ekonomi. Walaupun penguncian telah berkesan dalam mengawal penyebaran virus, kesan negatifnya terhadap ekonomi telah mendorong keperluan untuk langkah kawalan alternatif yang seimbang kos. Pengesanan kenalan telah muncul sebagai penyelesaian yang menjanjikan dalam mengenal pasti wabak komuniti COVID-19. Untuk meningkatkan keberkesanan pengesanan kenalan, kajian ini bertujuan untuk merumuskan model rangkaian hubungan untuk penularan COVID-19. Pendekatan konvensional didapati tidak mencukupi dalam memodelkan penularan COVID-19, terutamanya dalam mengenal pasti punca jangkitan. Untuk menangani perkara ini, kajian itu menggunakan pendekatan pemodelan rangkaian dwipartit untuk mengambil kira kepelbagaian laluan penghantaran, hos manusia dan lokasi yang dilawati. Hos manusia dan lokasi yang dilawati dikenal pasti sebagai dua entiti diskret dalam senario penyelidikan. Menggunakan data daripada borang siasatan pengesanan kenalan Pejabat Kesihatan Bintulu, enam model rangkaian telah dirumuskan. Berat pautan antara hos manusia dan nod lokasi dikira menggunakan peraturan penjumlahan, dengan mengambil kira pelbagai faktor seperti sifat persekitaran, ciri bangunan, ciri manusia dan patogen, dan mod penghantaran. Lokasi dan nod manusia kemudiannya ditarafkan menggunakan algoritma carian berasaskan web berdasarkan nilai kedudukan masing-masing. Hasil kajian menunjukkan pendekatan pemodelan rangkaian dwipartit berjaya merumuskan model rangkaian kenalan. Analisis pengesanan mendedahkan ralat purata kuasa dua akar 0.0002244 dan 0.001419 masing-masing untuk lokasi dan nod manusia, yang berada dalam nilai ambang 0.05. Kedudukan antara sasaran dan model yang disahkan didapati

mempunyai persamaan yang kukuh dengan pekali korelasi kedudukan Spearman yang baik melebihi 0.70 ($p < 0.001$), menunjukkan tahap perkaitan yang tinggi dalam meningkatkan pengesanan kenalan untuk COVID-19. Kajian juga mendapati bahawa semua parameter yang digunakan dalam model adalah agak ketara, dan model itu mempunyai keupayaan untuk meramalkan potensi titik panas dengan ketepatan 90% dalam radius 600m untuk minggu berikutnya. Penemuan ini menyerlahkan potensi pendekatan pemodelan rangkaian dwipartit dalam meningkatkan pengesanan kenalan untuk COVID-19 dan mengurangkan penyebaran virus di kawasan berisiko tinggi.

Kata kunci: *Permodelan rangkaian bipartite, COVID-19, hotspot, kedudukan*

TABLE OF CONTENTS

	Page
DECLARATION	i
ACKNOWLEDGEMENT	ii
ABSTRACT	iv
ABSTRAK	vi
TABLE OF CONTENTS	viii
LIST OF TABLES	xii
LIST OF FIGURES	xiv
LIST OF ABBREVIATIONS	xv
CHAPTER 1 INTRODUCTION	1
1.1 Introduction	1
1.2 Background of the Study	2
1.3 Problem Statement	8
1.4 Research Objectives	9
1.5 Research Scope	10
1.6 Research Significance	10
1.7 Thesis Outline	11
CHAPTER 2 LITERATURE REVIEW	14

2.1	Introduction	14
2.2	Hotspot Detection Studies	14
2.3	Hotspot Detection Studies in Disease Aspect	17
2.3.1	Main Factors of Hotspot Detection in Disease Studies	17
2.3.2	Approaches in Disease Model Studies	22
2.4	Network Modelling Approaches	25
2.4.1	Graph Theory and Network Theory	25
2.4.2	Network Modelling Approaches in Epidemiology Studies	27
2.5	Summary	34
CHAPTER 3 BIPARTITE NETWORK MODEL FOR COVID-19 CONTACT		
	TRACING	36
3.1	Introduction	36
3.2	Bipartite Network Model-Research Methodology Framework	36
3.2.1	Problem Characterization	38
3.2.2	Basic Building Block of the Bipartite COVID-19 Contact Network Model	39
3.2.3	Research Data and Potential Parameters	42
3.2.4	Bipartite Graph Formulation	44
3.3	Model Construction	49
3.3.1	Data Pre-processing	49
3.3.2	Pre-Process of Parameter	54

3.3.3	Quantification of Location Node Parameters	56
3.3.4	Quantification of Human Node Parameters	69
3.3.5	Quantification of Link Weight	78
3.4	Summary	84
CHAPTER 4 IMPLEMENTATION, ANALYSIS AND EVALUATION		86
4.1	Introduction	86
4.2	Search Algorithm	86
4.3	HITS Algorithm Implementation	88
4.3.1	Implementation Platform	89
4.3.2	Generation of Hub and Authority Matrix	90
4.3.3	Generation of Hub and Authority Principal Eigenvector	95
4.3.4	Assignment of Node's Label to the Corresponding Principal Eigenvectors	96
4.3.5	Ranking of Bipartite Nodes in BCC Network	98
4.3.6	Implementation Issues	99
4.4	Model Verification	101
4.4.1	Benchmark Verification	101
4.4.2	Analytical Verification	107
4.5	Model Validation	109
4.5.1	Validation with Raw Contact Tracing Data	109
4.5.2	Validation with Public Tracing Data	116

4.6	Model Analysis and Evaluation	119
4.6.1	Parameter Significance Analysis	119
4.6.2	Predictive Power Analysis	123
4.7	Summary	131
CHAPTER 5 CONCLUSION		133
5.1	Introduction	133
5.2	Conclusion of Research Activities	133
5.3	Research Contributions	136
5.4	Future Research	138
REFERENCE		141
APPENDICES		163

LIST OF TABLES

	Page
Table 2.1: Summary of Environmental Parameters in Hotspot Detection Studies	18
Table 2.2: Related Indirect Transmission Disease Studies Applying Contact Network	31
Table 3.1: Attributes of Research Raw Data	42
Table 3.2: Potential Parameters Used	44
Table 3.3: Detailed Information of BCC Network Model	45
Table 3.4: Movement of Each Human Host	46
Table 3.5: Location Node with GPS Coordinates for First BCC network	53
Table 3.6: List of building types with the air change rate (Q) according to the ASHRAE guideline	55
Table 3.7: Stability of the Virus in Aerosol, Ka of Each Location Node calculated based on Equation 3.5	60
Table 3.8: Stability of the Virus on Surface, Ks of Each Location Node calculated based on Equation 4.8	63
Table 3.9: Link Matrix of BCC Network for human	65
Table 3.10: Parameters Values of 21 Location Nodes in BCC Network	68
Table 3.11: P_{wr} for Each Human Node based on Location Node caculated based on Equation 3.15	71
Table 3.12: Total Duration of a human stay at one location, D	74
Table 3.13: Parameter Values of $Fh_{j:i}$	76
Table 3.14: Complete Normalized Parameter of Location Nodes	80
Table 3.15: COVID-19 Contact Strength (CCS) Matrix of BCC Network	81
Table 4.1: Hub Matrix	91
Table 4.2: Authority Matrix	92
Table 4.3: Principal Eigenvector and CHR of the Ranked Location Node	97
Table 4.4: Principal Eigenvector and CRR of the Human Node	98

Table 4.5: RMSE Analysis of Location Nodes	105
Table 4.6: RMSE Analysis of Human Nodes	106
Table 4.7: SRCC Indicators	108
Table 4.8: BCC Network Model Validation	111
Table 4.9: The Ranking of Location Nodes (First Validated Model)	112
Table 4.10: SRCC Calculation of Location Nodes between the Group 1	113
Table 4.11: SRCC of Location Nodes between the Models in Group 2	115
Table 4.12: SRCC of Human Nodes between the Models in Group 2	115
Table 4.13: SRCC of Location Nodes between the Models in Group 3	115
Table 4.14: BCC Network Model with Tracing period	116
Table 4.15: BCC Network Model Validation with Public Data	118
Table 4.16: SRCC for “Leave-one-out” Parameter Analysis	121
Table 4.17: SRCC for “Leave-multiple-out” Parameter Analysis	121
Table 4.18: Ranking of radius 200 meters cluster group 1	126
Table 4.19: Top 10 ranked locations in group 1 validated model (A)	126
Table 4.20: List of Targeted Models and Validated for predictive power analysis	128
Table 5.1: Summary of the thesis	136

LIST OF FIGURES

	Page
Figure 2.1 Three Main Components of Hotspot Detection Studies Structure (Kok, 2017)	21
Figure 2.2 Complexity of the epidemiological models (Dimitrov & Meyers, 2010)	24
Figure 3.1 Bipartite Network Modeling Research Methodology Framework (BNM-RMF) (Liew, 2016)	37
Figure 3.2 Epidemiological Triangle (ET)	39
Figure 3.3 Basic Building Block of BCC Network	40
Figure 3.4 An example of a BCC network	41
Figure 3.5 Complete Bipartite COVID-19 Contact Graph for COVID-19 patient	48
Figure 3.6 Pre-processing of the Location Node	50
Figure 3.7 Manual Searching GPS Coordinates	52
Figure 3.8 Parameters of BCC Network Location Node at Process 2.2 in BNM-RMF	57
Figure 3.9 Bipartite COVID-19 contact network with clustered human node	84
Figure 4.1 Process of the Adapted HITS Search Algorithm (Liew, 2016).	89
Figure 4.2 Benchmark Verification Process (Liew, 2016)	102
Figure 4.3 Benchmark Ranking and BCC Network Model Ranking	104
Figure 4.4 Implementation Design of Predictive Power Analysis	125
Figure 4.5 Accuracy of each targeted model within the same weeks	127
Figure 4.6 Accuracy of each targeted model for subsequent one week	129
Figure 4.7 Accuracy of each targeted model for subsequent two weeks	130

LIST OF ABBREVIATIONS

ABM	Agent-based modelling
ASHRAE	American Society of Heating, Refrigerating and Air-Conditioning Engineers
API	Application programming interface
ANN	Artificial neural network
BCC	Bipartite COVID-19 Contact
BNM	Bipartite Network Model
BNM-RMF	Bipartite network-based research methodology framework
CCS	COVID-19 contact strength
CHR	COVID-19 hotspot ranking
CRR	COVID-19 risk infection ranking
ET	Epidemiology Triangle
GPS	Global Positioning System
HCW	Health-care workers
HITS	Hypertext Indexed Topic Selection
ID	Identification
MCO	Movement control order
PDE	Partial differential equation
RMSE	Root-Mean-Square-Error
STDs	Sexual transmission diseases
SRCC	Spearman's Rank Correlation Coefficient
SI	Susceptible-infectious
SIR	Susceptible-infectious-removal

WHO

World Health Organization

CHAPTER 1

INTRODUCTION

1.1 Introduction

Contact tracing is one of the pillars of control measures to contain the outbreak of infectious diseases (Eames & Keeling, 2003). Herein, identifying high-risk areas is crucial to the effectiveness of contact tracing. This study focused on high-risk areas of COVID-19 transmission which can be denoted as hotspots or source of infection (Lessler et al., 2017). Many traditional approaches for contact tracing focused on the interaction between humans; however, the locations visited by the infected individuals are not factored. Moreover, conventional modelling, for instance, compartmental modelling, may not be suitable when integrating the spatial components because it is non-linear; hence, incorporating location or environmental properties becomes more complicated (Garnett, 2002; Nelson & Perelson, 2002). Accordingly, there is a need to identify and use an appropriate approach to detect hotspots of COVID-19 transmission.

The graph-theoretic network modelling approach that was recognised as a heuristic framework has the potential (Bunn, Urban & Keitt, 2000 in Liew, 2016, p. 1). Heuristics is a problem-solving technique that applies practical methods which do not promise an optimal solution but can generally produce solutions that are sufficient for practical use (Wahid & Hussin, 2016). This supports the present study to employ a bipartite network-based research methodology framework (BNM-RMF) (Liew, 2016) to address the research problem. To use that framework, the domain problem must be formulated into network models where the nodes and edges are well defined for location detection. The network structure is converted into a suitable format that can be used as input into the network model. Afterward, the model

is verified and validated, and analytical results were generated. It identifies the locations with high risks of COVID-19 transmission and rank those based on the estimated virus density. Such findings may serve an important cognizance for the authorities working on contact tracing and control measurements.

This chapter discusses the background of the problem and presents the problem statement; research questions, objectives, scope, methodology and significance; and the thesis outline.

1.2 Background of the Study

The first case of COVID-19 was reported in Wuhan, China, in December 2019. Accordingly, China notified the outbreak to the World Health Organization (WHO) as pneumonia of unknown aetiology. The virus had been identified as a novel coronavirus (2019-nCoV) and further named COVID-19 on 11 February 2020. Chinese government's attempts to contain the virus have failed, and it rapidly spread across the globe due to the high volume of air traffic and insufficiency of control measures. As many countries were already affected by the outbreak, WHO declared it as a pandemic on 11 March 2020, 3 months after the first case was reported. According to an investigation, the origin of the pathogen of COVID-19 was from a bat (Jiancheng Zhang et al., 2020) and escalate to the human population. COVID-19-infected individuals can transmit the virus directly and indirectly. Direct transmission involves, whereas indirect transmission includes airborne and fomite means (WHO, 2020). In addition, mass gathering, especially religious gathering events, is the main cause of outbreak in many countries (Ebrahim & Memish, 2020; Jeong et al., 2020; Mat et al., 2020). In Malaysia, the second wave of the outbreak originated from a Muslim Tablighi event in Sri Petaling, Selangor, on 27 February 2020. This event consists attendees from many countries, which amplifies the transmission (Mat et al., 2020; Dass et

al., 2021). Moreover, this event became the cause of infection in Malaysia and other countries in Southeast Asia. Consequently, a nationwide lockdown had been implemented in Malaysia on 18 March 2020, named Movement Control Order, and slowed down the outbreak at that time (Gill et al., 2020). This Movement Control Order caused a huge impact on Malaysia's economy. To curb the outbreak, control measures such as contact tracing, early detection of potential infection and quarantine are vital to limit the transmission (Xiao & Torok, 2020).

Contact tracing is a disease surveillance approach to obtain the possible transmission path by identifying the person who may have contact with an infected person. The main purpose of contact tracing is to break the current transmission chain (WHO, 2017). Contact tracing has been proven effective and widely used in several communicable diseases, i.e. droplet-transmitted diseases such as SARS, Middle East respiratory syndrome (MERS) and H1N1 (Shankar et al., 2014; Kwok et al., 2019); airborne-transmitted diseases such as tuberculosis and measles (Begun et al., 2013; Thole et al., 2019); and sexually transmitted diseases such as HIV/AIDS (Hyman et al., 2003). The common characteristic of the abovementioned diseases is that they transmit via close contact; thus, contact tracing can identify the index patient and the possible exposed individuals and avoid the disease from spreading further. Besides, some studies of vector-borne diseases such as dengue and malaria are using contact tracing (Kok, 2017; Vazquez-Prokopec et al., 2017). Those studies focused on the location the patient visited to identify the place of infection.

Generally, the contact tracing will be initiated when a patient is admitted to a hospital or a public health officer received a notification. The public health officer will interview the patient to obtain the contact and activity history. After identifying the close contacts, they will be asked to be home quarantined if they are identified as low risk; otherwise, they will

be referred to isolate in a hospital if they are categorised as high risk (WHO, 2017). The procedure mentioned is a manual contact tracing because it requires human resources to interview patients and trace their activities. The cost of manual contact tracing is high in terms of manpower and human lead error. The needed activity and contact history of the patient is up to 14 days (MOH, 2020); thus, patients tend to forget their activity/contact history, which will affect the efficiency of the procedure. To address the issues of manual contact tracing, a digital approach has been introduced. Digital contact tracing will make use of a mobile application to capture the activity and contact history of every citizen. This approach has been used in many countries during the COVID-19 pandemic (Jalabneh et al., 2020). According to Jalabneh, 15 countries developed a mobile application to conduct contact tracing, which showed the importance of this approach in controlling the spread of COVID-19. The application will collect several personal information such as name, age and post-code, with some applications further recording global positioning system (GPS) and Bluetooth contact data.

However, early transmission studies did not well integrate the contact tracing data collected manually or digitally because they focused on the prediction of outbreaks and simulation of the epidemic curve (Gill et al., 2020; Kucharski et al., 2020; Shen et al., 2020; Wu et al., 2020). These studies used the conventional approach to formulate a mathematical model, assuming the population is well-mixed. Due to this assumption, the model is difficult to incorporate with the mobility data (contact tracing) heterogeneously. Some studies considered the contact rate among the population distinctive based on age group (Prem et al., 2020; Zaplotnik et al., 2020; Juanjuan Zhang et al., 2020). These studies are beneficial in implementing control measures, i.e. school closing, work from home, or public place closure. Nonetheless, these studies are focused on the macro-level of the population, that is,

the national level, wherein contact tracing data were not utilised. In addition, transmission of the disease became complex when it involves transmission route, physical properties of the sites, human behaviour and characteristics of the pathogen (Killingley & Nguyen - Van - Tam, 2013; Sze-To et al., 2014). These factors highly affect the stability of the pathogen of COVID-19. It is found viable for more than 1 day on the surface of an object (Bueckert et al., 2020; Chin et al., 2020) and in the air (Ahlawat et al., 2020; Anderson et al., 2020; van Doremalen et al., 2020), resulting in onerous tracing of COVID-19 transmission compared with other diseases and difficulties in employing conventional approaches. Although other approaches, i.e. partial differential equation, agent-based modelling, or deep-learning approaches, may solve this issue, these are non-linear and of high dimension and need more computational power (Cheng et al., 2003). Somehow, the outputs generated from deep-learning approaches lack sense, making it hard for decision-makers to interpret (Zohuri & Moghaddam, 2020).

Recent researchers utilised network modelling approaches to address the limitations of the abovementioned approaches for the epidemiology domain (Eubank et al., 2004; Eubank, 2005; Eze et al., 2011; Gao et al., 2016; Kok et al., 2017). The network modelling approach originated from graph theory, which is a study to identify the relationship between the vertices and edges (Pavlopoulos et al., 2011; Gao et al., 2018). Any problems that can be simplified into vertex and edge are suitable for graph theory use. In terms of contact tracing, researchers used network modelling approaches to identify the transmission clusters between individuals (Eames & Keeling, 2003; Kiss et al., 2008; Begun et al., 2013). Eames and Keeling (2003) highlighted the effectiveness of the network model utilised with real-life data to identify the disease clusters, promptly eliminate the outbreak and reduce the burden on health resources. Although Begun et al. (2013) claimed that data on the effect of individual

behaviours and location visited by the close contacts are required to improve the efficacy of contact tracing, however, most of the network modelling of COVID-19 studies focused on the transmission between humans rather than their locations (Benlloch et al., 2020; Karaivanov, 2020; Xue et al., 2020). This is because formulating a contact network for human-to-human transmission is more convenient (Danon et al., 2011).

The lack of a location component in the contact network makes the study of indirect transmission of diseases insufficient, especially for vector-borne and airborne diseases (Danon et al., 2011). As the host and location are discrete components, researchers employed a bipartite network model (BNM) to incorporate these into a contact network. Early studies of BNM focused on the interaction between the host and location in the network by examining the network properties such as node degree, centrality and connectivity (Ancel Meyers et al., 2003; Eubank et al., 2004). Although these studies only emphasise the network properties of BNM, they provided insights for other studies to use BNM in identifying the risk of infection for a host and the risk of transmission for a location. Eze et al. (2011) adopted BNM to create a contact network for malaria transmission and successfully replaced the mosquito node with a location node. The location node quantified by Eze et al. integrated the environmental properties of a location with the behaviour of mosquitoes such as life cycle, survival rate and biting rate. According to this study, a mosquito normally flies within a certain distance from a particular location; hence, the location node is more suitable in the network instead of the mosquito node. Similarly, Kok et al. (2017) used BNM to identify the hotspot of dengue in Malaysia by integrating the human mobility into the network. Both of these studies demonstrate the capability of BNM to address vector-borne diseases. One of their contributions is resolving the issue raised by Pellis et al. (2015), that is, to merge the

link between the nodes with the epidemiology data by quantifying the mosquito behaviour into their network and to compute the contact strength between the nodes.

For airborne diseases, it is crucial to identify the public places with high risks of transmission to prevent further community transmission (Goscé & Johansson, 2018). In addition, the conventional contact tracing method should consider the indirect transmission in those public places. Public health authorities collecting contact tracing data should utilise these data to lower the risk of indirect transmission in public places. Although vaccines were already provided to the community to lower the risk of infection, contact tracing should still be improved and always on standby due to the higher variability of SARS-CoV-2.

Herein, we aim to model the transmission of COVID-19 by using Bintulu as a case study to assist the public health authorities in improving the efficacy of contact tracing. Instead, of conducting the study on a large area, early targeting of a potential transmission hub is more appropriate to lessen the impact on residents' livelihood during lockdown (Eubank et al., 2004; Goscé & Johansson, 2018). However, there is no method that aids decision-makers in identifying locations with a high risk of transmission.

COVID-19 is transmitted either directly or indirectly; hence, the network model considers the environmental and physical properties of the buildings. These factors are used to quantify the stability of the pathogen in several modes and the risk of infection of at-risk individuals in the network. The Bintulu Health Office provided the contact tracing data for the network to process real data; the details of the data are presented in Chapter 3. This study aims to establish a new approach to model a district-level COVID-19 transmission by identifying the risk of transmission at the location and the risk of infection on humans. In addition, this study aims to resolve one of the challenges raised by Pellis et al. (2015), that is, to develop a network-based intervention to control the outbreak of diseases.

In summary, contact tracing is one of the effective control measures to prevent further outbreak especially for human-to-human transmission (Eames & Keeling, 2003; Vazquez-Prokopec et al., 2017; Ferretti et al., 2020). Researchers extensively used network modelling approaches to study disease transmission and incorporate contact tracing (Hyman et al., 2003; Kiss et al., 2008; Begun et al., 2013). However, contact tracing is difficult to integrate in indirect transmission diseases such as vector-borne, airborne and vehicle-borne (Danon et al., 2011). Therefore, BNM is proposed to study vector-borne (Bisanzio et al., 2010; Eze et al., 2011; Kok et al., 2017) and airborne diseases (Ancel Meyers et al., 2003; Eubank et al., 2004; Gao et al., 2016). Nevertheless, the BNM used in these diseases lacks other factors such as the variations in transmission routes, stability of pathogens and the risk of infection. Thus, the contact network model used in this study will combine these factors to understand the impact of heterogeneity of humans and pathogens on COVID-19 transmission. The research problem of this study is presented in the next section.

1.3 Problem Statement

This study aims to identify hotspots of COVID-19 and the risk of infection to increase the accuracy of contact tracing of COVID-19 by using a network modelling approach. The two pillars of the contact tracing are the human (infectious and close contact) and location. However, one of the critical component in epidemiology of triangle (ET) is agent (virus). Thus, the transmission routes of the disease will affect the stability of the pathogen. Accordingly, the problem statement of this study will be:

How to relate the location and virus in ET to formulate a contact network that consists of two entities—human and location—while taking into account the transmission modes of direct and indirect?

Based on the problem statement, three main research questions are derived as below:

1. How to integrate the relationship between virus and location into existing ET?
2. How to formulate a contact network for a disease consisting of multiple transmission modes?
3. How to quantify the contact network link weight for a disease can be transmitted directly and indirectly?

To formulate a network model that increases the accuracy of contact tracing, this network will be a two-type node network as contact tracing covered two discrete entities (human and location), hence bipartite network will be adopted in this study. The relationship between virus, environment and location must be integrated to obtain the link weight between the human and location in the bipartite network. A quantification process will be conducted using significant parameters related to these components. The hotspot of COVID-19 will then be identified based on the ranking of the link weight by using implemented search algorithm.

1.4 Research Objectives

Based on the research questions mentioned above, the objectives of this research are as follows:

- (a) To integrate the characteristics of the COVID-19 virus with the location visited by the patient of COVID-19.
- (b) To formulate a contact network model based on the interactions between humans and the location visited by the patient of COVID-19.
- (c) To identify hotspots of COVID-19 by ranking the link weight between human and location in the contact network using quantification rule and a search algorithm.

1.5 Research Scope

This study aims to formulate a contact network model to increase the accuracy of contact tracing by identify the COVID-19 hotspot. The hotspot in this study is a location node that has a higher rank compared with other location nodes in the network. The higher-ranked location nodes imply a relatively higher density of COVID-19 and hence a higher risk of transmission. The bipartite network model will be used to simulate the contact tracing of COVID-19 and hypertext induced topic search (HITS) will be used based on the optimistic outcomes from previous studies.

The raw data used here are the real-world data provided by the Sarawak Bintulu Health Office from Epidemiology week (EW) 3 to 7 of year 2021. However, the raw data are sloppy and not suitable as model parameters. Hence, a pre-processing technique of the raw data has been conducted and will be discussed in Chapter 3. Moreover, the network model can predict the hotspot 1 week after the incidence case based on predictive power analysis.

The target users for the project outcome are public health and medical officers, which would assist them in managing the resources for contact tracing and planning control measures to contain the COVID-19 outbreak in Malaysia.

1.6 Research Significance

The significance of this study lies in its ability to successfully utilize the BNM approach to identify COVID-19 hotspots. The ongoing COVID-19 pandemic has underlined the need for effective contact tracing as a means of controlling the spread of the virus. However, traditional manual contact tracing is costly in terms of manpower and prone to human error, which is why digital contact tracing using mobile applications has been

introduced. While early transmission studies have focused on predicting outbreaks and simulating epidemic curves, they have not well integrated contact tracing data. This study aims to fill that gap by incorporating mobility data obtained through contact tracing into mathematical models to understand COVID-19 transmission dynamics at a micro-level, taking into account the transmission route, physical properties of sites, human behavior, and pathogen characteristics.

The study presents a solution by using BNM, to addresses the challenge of modelling COVID-19 transmission networks with multiple modes, while taking into account the heterogeneity of human health status and transmission routes. This approach contributes to the field by providing a solution to designing network-based interventions, as suggested by Pellis et al. (2005), and by improving the understanding of COVID-19 transmission dynamics. The BNM model has the potential to enhance contact tracing accuracy and identify COVID-19 hotspots more accurately, leading to better public health outcomes. The use of the epidemiology triangle and network-based modelling in the BNM approach demonstrates the importance of considering the spatial and social context of COVID-19 transmission, which can inform the development of more effective prevention and control strategies.

1.7 Thesis Outline

This thesis has five chapters. The content of the chapters is summarised as follows: Chapter 1 (Introduction) discusses the background of this study. It presents the main problem statement of the case study and accordingly defines the objectives and scope of this study.

Chapter 2 (Literature review) discusses the review of related literature on different methods used to detect hotspots, particularly on the network approaches employed, limitations and opportunities that can support the solution of this study.

Chapter 3 (Bipartite network model for covid-19 contact tracing) discussed the research methodology used in this study – Bipartite Network Modelling-Research Methodology Framework (BNM-RMF). This chapter presents the stage 1 and 2 of the BNM-RMF. The first stage is to understand the research scenario by presenting the dataset, research region (Bintulu) and potential variables and assumptions. The second process in stage 1 is to identify the basic building blocks of the graph. Finally, the last process is to formulate the bipartite graph from the basic building blocks. There are four processes involved in stage 2 of BNM-RMF are covered in this chapter while the last process of stage 2 will be discussed in next chapter. The first activity in stage 2 is to pre-process the research data such as COVID-19 contact tracing data, weather data and building ventilation rate data. The second and third processes are to quantify the parameters of location and human nodes. The fourth process is to quantify the link weight between location and human nodes.

Chapter 4 (Implementation, Analysis and Evaluation) presents the implementation of a search algorithm to rank the hotspot value of the location and the infection risk of human in model. The implementation platform used in this study is introduced and issues encountered in implementation are discussed. Model verification, model validation and network evaluation, which serves as the stage 3 of the BNM-RMF will be covered in this chapter. The design of the parameter significance analysis and predictive power analysis will be demonstrated to evaluate the performance of the model.

Chapter 5 (Conclusion) concludes the overall research and findings. This chapter presents the achievements of this study and ends with some recommendations for future work that may improve the research.

CHAPTER 2

LITERATURE REVIEW

2.1 Introduction

The objective of this literature review is to examine the existing knowledge and research on hotspot identification and the significant factors affecting disease transmission. Previous studies have attempted to identify the relationship between environmental predictors and disease transmission. This chapter will review the main predictors that will be used to quantify human and location nodes in this study. Furthermore, the current methods employed for modelling disease transmission and identifying the source of infections will be critically evaluated. While these methods have been successful in incident forecasting, they lack the ability to address the heterogeneity of human data, particularly in regard to daily activities and locations visited. The literature review will begin by discussing hotspot identification in different domains and focusing on disease hotspot detection studies and the important factors involved. Subsequently, existing approaches for modelling disease transmission and identifying the source of infections will be reviewed. Finally, the proposed bipartite network modelling approach will be justified to detect hotspots of transmission diseases with multiple transmission routes and address the limitations of previous network modelling methods.

2.2 Hotspot Detection Studies

Hotspot identification refers to the process of identifying areas or regions that have high concentration of a particular phenomenon or activity. This can refer to a variety of things such as crime hotspots, disease hotspots, or hotspots of biodiversity (Braga & Weisburd, 2010; Myers et al., 2000; Ying et al., 2015). Hence, hotspot identification is often

used in fields such as criminology, epidemiology, and conservation biology to target resources and interventions in areas of highest need. In criminology, crime hotspots refer to specific geographic locations where crime is concentrated and is likely to occur. These areas can be identified by analysing crime data, such as incidents reported to the police, to identify patterns and clusters of crime.

In criminology, crime hotspots can include areas such as city streets, neighbourhoods, or specific buildings or businesses. Once crime hotspots are identified, law enforcement and other criminal justice agencies can target their resources and interventions to those specific areas. This can include increasing patrols, implementing community policing programs, or addressing underlying social and economic issues that contribute to crime (Braga & Weisburd, 2010). Additionally, researchers also use crime hotspots data to understand the underlying causes of crime and to develop evidence-based crime prevention strategies. In perspective of ecology, hotspots of biodiversity refer to specific geographic regions that have a high concentration of unique and diverse species of plants and animals (Myers et al., 2000). Due to their high concentration of diverse and unique species, coupled with increasing human activities such as pollution, habitat destruction, and climate change, these regions are of great importance for biodiversity conservation. Therefore, they are considered as areas of high priority for conservation efforts (Mittermeier et al., 2011).

The main methods used to identify crime hotspots is spatial analysis. This method involves using Geographic Information Systems (GIS) and spatial statistics to map crime data and identify patterns and clusters of crime (Andresen & Lauzon, 2020; Duan et al., 2021). This can include techniques such as kernel density estimation, which calculates the density of crime incidents in a specific area, or spatial autocorrelation, which measures the similarity of crime incidents in relation to their geographic location. While domain of

biodiversity, spatial analysis will be focusing on distribution of and abundance of plant and animal species by applying similar techniques in crime hotspot identification. In addition to identifying hotspots of biodiversity, spatial analysis can also be used to identify areas that are under significant threat from human activities such as habitat destruction, pollution and climate change (De Albanesi & Di Paola, 2021). This can be done by analysing data on human land use, such as land cover change, and overlaying this with data on species and habitats. This can help to identify areas where conservation efforts are most needed.

Network analysis also be applied to identifying hotspots of biodiversity is by analysing the connections between different habitats in a landscape. For example, Renner et al. (2015) used a network-based approach to identify and manage threatened plant species in Australia. They constructed a spatially explicit network model that incorporated information on the geographic distribution and co-occurrence patterns of plant species, as well as their ecological traits and conservation status. The model was then used to predict the potential occurrence of threatened species and to prioritize areas for conservation management. In the context of crime hotspot identification, Andresen and Lauzon (2020) applied a network model to analyze the spatial and temporal patterns of crime incidents in Vancouver, Canada during the COVID-19 pandemic. They constructed a network of crime events based on their geographic proximity and temporal co-occurrence, and used the network to identify clusters of crime that were most likely to persist over time.

Overall, the use of network models in both crime hotspot and biodiversity hotspot identification has shown great potential for improving our understanding of the underlying spatial processes and relationships that drive the distribution and abundance of crime incidents and plant and animal species.

2.3 Hotspot Detection Studies in Disease Aspect

From a public health perspective, disease hotspot detection is the analysis of the spreading dynamics and spatial distribution across an area for a disease. Hotspots can be defined from different perspectives in disease transmission studies. According to Lessler et al. (2017), hotspots can be categorised into several types, for instance, an area having a high number of cases and high risk of disease transmission. In this study, a hotspot is a location that has a high risk of disease transmission. For vector-borne disease, hotspot detection is vital because when the correct hotspot is targeted, the spread of a disease can be controlled. Kok et al. (2017) showed that by identifying hotspots of dengue, health authorities can effectively perform fogging action instead of fogging all locations visited by the patient. This will lower dengue transmission and reduce the burden on public health resources. As mentioned by Goscé and Johansson (2018), identifying hotspots of airborne diseases plays an important role in curbing the disease transmission. Hence, it is vital to identify the source of infection and alert the public about the risk of transmission in public places.

If an area has agreeable conditions for the emergence or re-emergence of a disease, it can be classified as a hotspot as there is a risk of transmission. In order to determine the conditions of disease re-emergence, the environmental properties and its connection with the disease transmission need to be identified and understood.

2.3.1 Main Factors of Hotspot Detection in Disease Studies

From an epidemiological perspective, epidemiologists and public health officers must determine whether there is a cluster of incidents of diseases (Kulldorff & Nagarwalla, 1995). This can help them to promptly detect the outbreak of the diseases and decide on managing resources for control measures. When the cluster of incidents is unusual and relates to spatiality, researchers can identify the location with a higher risk of disease

transmission. Hence, hotspot detection studies are correlated with the transmission route of the diseases, physical properties of the sites and the distribution of the disease occurrence (Brundage et al., 1988; Killingley & Nguyen-Van-Tam, 2013; Sze-To et al., 2014).

This section presents the aim of the study to discover the structural similarities among the reviewed hotspot detection studies of infectious diseases. As aforementioned, physical characteristics and environmental properties of the location are crucial in hotspot detection studies. Hence, a summary of the environmental predictors of infectious diseases is presented on Table 2.1. Researchers usually predict the transmission risk based on the distribution of the incidents, abundance of pathogens and environmental predictors that are highly related to the stability, infectivity, and transmission of the pathogen (Masuda & Holme, 2013; Pica & Bouvier, 2012; Sundell et al., 2016; Tang & Loh, 2014). However, early COVID-19 modelling studies focused on temporal studies, such as time series analysis, predicting the outbreak size or determining the relationship of the incidents with the environmental predictors, thereby neglecting the characteristics of disease's pathogen.

Table 2.1: Summary of Environmental Parameters in Hotspot Detection Studies

Source	Diseases studied	Study area	Parameters included	Approach(es)
Demongeot et al. (2020)	COVID-19	21 countries	Temperature	ARIMA
Azuma et al. (2020)	COVID-19	Japan	Temperature, precipitation, relative humidity	Weighted random-effects regression

Table 2.1 continued

Runkle et al. (2020)	COVID-19	United States	Temperature, humidity, solar radiation	Conditional Poisson regression
Shi et al. (2020)	COVID-19	China	Temperature	Distributed lag non-linear models
Ma et al. (2020)	COVID-19	China	Temperature, humidity	Generalized additive model (GAM)
Kuddus et al. (2019)	Tuberculosis	Bangladesh	Temperature, humidity, rainfall	Generalized linear Poisson regression models
Wang et al. (2017)	Influenza	China, Canada	Mean temperature, relative humidity, absolute humidity	Poisson generalized additive models
Wu et al. (2014)	Hand, foot, and mouth disease (HFMD)	China	Temperature	Poisson generalized additive model

According to Killingley and Nguyen-Van-Tam (2013), transmission routes of the diseases significantly contribute to the risk of disease transmission. It was stated in Section 1.2 that human-to-human direct and indirect transmission is the prevalent transmission route of COVID-19 (WHO, 2020). Details of the transmission modes can be found in Appendix B (Defination of direct and indirect transmission modes)

As summary, environmental properties play a significant role in the stability of pathogens in indirect transmission modes. Several experiments have been conducted to assess the relationship between the environmental properties and the stability of pathogens in different scenarios. Chapter 3 discusses the quantification process of stability of the pathogen.

Other than environmental predictors, an indoor setting is another critical factor in disease transmission risk (Brundage et al., 1988). Numerous studies have claimed that the ventilation of buildings plays a significant role in airborne-transmitted diseases (Dai & Zhao, 2020; Escombe et al., 2007; Morawska et al., 2020; Nicas et al., 2005). According to Luongo et al. (2016), the air change rate is highly related to the incidence of infectious diseases in an indoor environment. For instance, for respiratory diseases, a lower rate of ventilation leads to a higher number of infections (Amoatey et al., 2020; Sun et al., 2011). According to Meselson (2020), an indoor setting with poor ventilation allows the pathogen in aerosol form to stay longer. To resolve this issue, researchers have used the ventilation rate of an indoor building to calculate the risk of infection using Wells–Riley equation (Chen & Liao, 2008; Rudnick & Milton, 2003; Sun & Zhai, 2020; Zemouri et al., 2020). The Wells–Riley equation, as shown in Equation 2.1, is based on the concept of ‘quantum of infection’ proposed by Wells (1955) and formulated by Riley et al. (1978) and integrated with the ventilation rate.

$$P = 1 - e^{\left(-\frac{Iqpt}{Q}\right)} \quad \text{Equation 2.1}$$

where e is Euler’s number, I is the number of source patients (infectors), q is the rate of generation of infectious quanta, p is the pulmonary ventilation rate of each susceptible person per hour, t is the duration of exposure, and Q is the room ventilation rate. The

‘quantum of infection’ is a hypothetical sufficient number of infectious airborne particles expelled by an infected person to infect others. The Wells–Riley equation assumes that these particles are randomly distributed in a room where the air is well-mixed (Riley et al., 1978) and the half-life of the pathogen can be neglected. The applications of Wells-Riley equation are demonstrated in Appendix B (Applications of Well-Riley equation).

According to the studies reviewed above, the significant environmental properties were identified and used to quantify the nodes in the network model. Temperature and relative humidity evidently affect the stability of pathogen in airborne and fomite transmissions. The ventilation of a building highly influenced the risk of infection via direct contact or within the same space in an indoor environment. These factors will be raw parameters to further establish their relationship to quantify the human and location nodes used in this study.

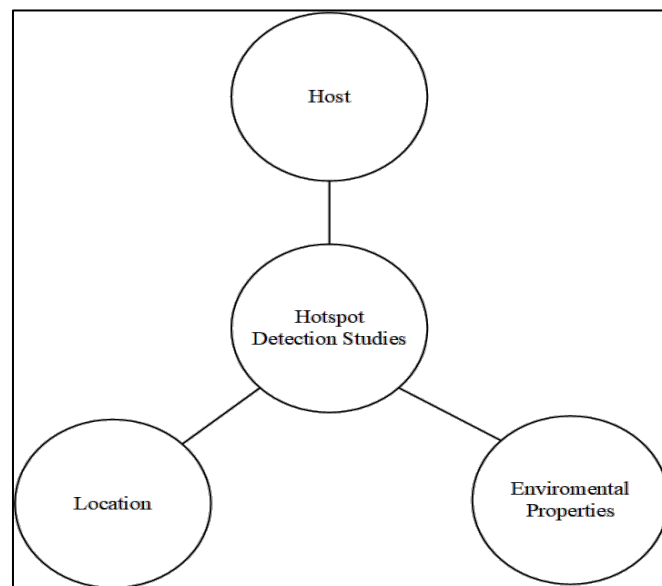


Figure 2.1: Three Main Components of Hotspot Detection Studies Structure (Kok, 2017)

In summary, based on the literature presented above, it was found that there always exist three pillars in the research structure for the reviewed studies of respiratory diseases (including COVID-19): host, location, and environmental properties. According to Kok (2017), these three components are the basic structures of hotspot detection studies, as illustrated in Figure 2.1. Moreover, Chapter 3 will further discuss the structure shown in Figure 2.1 to develop a fundamental framework for the COVID-19 contact network.

2.3.2 Approaches in Disease Model Studies

This section presents the motivation for employing the network modelling approach in this study. The approaches reviewed in this section will focus on disease modelling, specifically the risk of transmission due to location and human factors.

As mentioned in Section 1.2, early studies of disease modelling were mostly homogeneous, which assumes that the population are well-mixed and the contact rate between individuals is constant. Due to its simplicity, the parameters can be easily estimated, and the model can be simulated and frequently updated. Accordingly, researchers can promptly build a disease model and quickly provide useful information for the policymakers to implement control measures. One of the conventional approaches used by researchers is the compartmental model. During the initial outbreak of COVID-19, most of the models focused on the prediction of occurrences and outbreak sizes (Gill et al., 2020; Kucharski et al., 2020; Shen et al., 2020; Wu et al., 2020). The output of these studies are the simulations of the number of infected individuals and the predicted peak of the outbreak. However, the outbreak of COVID-19 is still ongoing, resulting in more issues that need to be resolved by researchers. One of these issues is the heterogeneity of components such as the environmental properties, human behaviours and stability of pathogen. The bottleneck of the conventional approach is the homogeneity assumptions that do not comprehensively capture

the outbreak in the real world. Even though the partial differential equation is one of the alternative approaches for researchers to replace the conventional approaches (Viguerie et al., 2020; Wang et al., 2020), the model is non-linear and has a higher dimensionality that makes it computationally complex (Cheng et al., 2003; Crandall et al., 1992).

Agent-based modelling is another approach used by researchers to resolve the real-world data heterogeneity issue. It has been widely used in studies that deal with interactions between people, locations, times and things (Galea et al., 2010). Cuevas (2020) used this approach to identify the transmission risk in different facilities by incorporating human behaviour and status (background and health status) in the model. The decision made by each agent in the model (human) is defined by a list of rules designed by researchers. According to Epstein (2009), ‘Agents can be made to behave something like real people: prone to error, bias, fear’. The complexity of the human decision process is programmable using agent-based modelling. However, Bansal et al. (2010) argued that some evident, algorithmic and computational issues are found in agent-based modelling as compared with the network modelling approach. According to Dimitrov and Meyers (2010), the computational complexity of the network modelling approach is between the compartmental model and agent-based model, as shown in Figure 2.2.

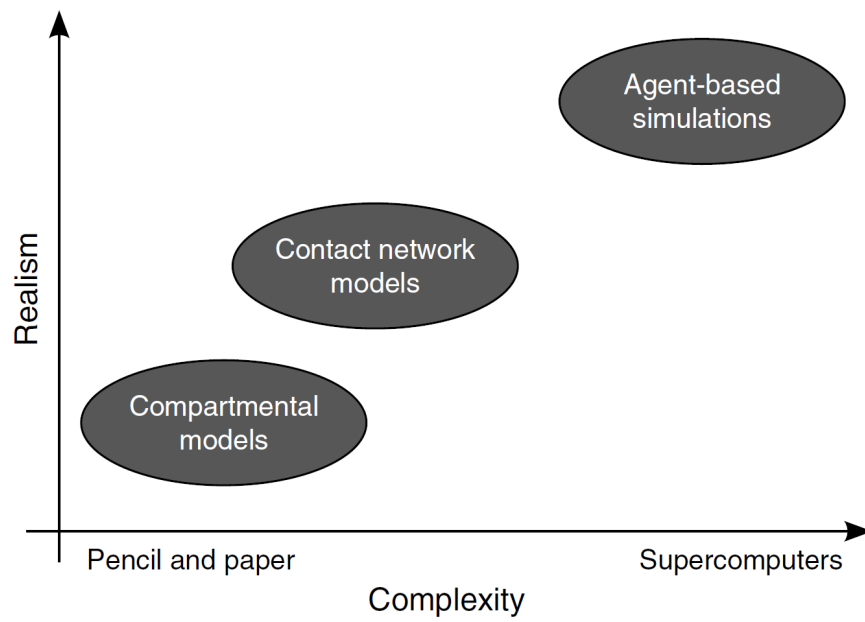


Figure 2.2: Complexity of the epidemiological models (Dimitrov & Meyers, 2010)

In Figure 2.2, the x-axis indicates the required computational complexity of models, and the y-axis indicates the ability of the model to capture the real-world complexities that are significant in disease transmission (Dimitrov & Meyers, 2010). As the network model lies between the compartmental and agent-based modelling, it can modify compartmental modelling by capturing the heterogeneous data and resolve the heterogeneity by integrating agent-based modelling (Bansal et al., 2007).

Several researchers had proposed other scientific methods to resolve the heterogeneity issues. For instance, Arora et al. (2020) had adopted a deep-learning approach to predict the daily incidents of COVID-19 for every state in India and used the predicted incidents as an indicator to classify the infection risk of each state. Moreover, Shaffiee Haghshenas et al. (2020) utilised a similar method to specifically identify the role of environmental predictor in the transmission of COVID-19 using an artificial neural network to build a model with environmental properties (average temperature, relative humidity and

wind speed) and population density and successfully predicted the number of incidents in Italy. However, these studies only focused on a large spatial distribution, neglecting the risk of transmission to the human entity. Besides, the main limitation of deep-learning approaches is the interpretability of the model. According to Zohuri and Moghaddam (2020), the generated data of a trained model are not always interpretable by humans and lack sense. This can lead to many issues, especially for policymakers because they need to provide proper grounds and basis before enforcing new policies.

As illustrated in Figure 2.1, each pillar in the hotspot detection studies can be delineated as a graph structure with three vertices and three edges (Liew, 2016). The network model can be visualised in a graph structure given that the network theory is derived from graph theory. There is a strong reason to employ the network modelling approach in addressing the research problem. Hence, the next section will discuss the epidemiology research, i.e. hotspot detection, contact tracing and risk of infection identification that applied network modelling.

2.4 Network Modelling Approaches

This section discusses the emergence of graph theory and its relationship with network theory. The literature on the network modelling approaches in epidemiology studies that are related to the research problem of this study will be reviewed.

2.4.1 Graph Theory and Network Theory

According to Danon et al. (2011), network theory became a popular tool to resolve the issues in epidemiology studies. Network theory is a branch of graph theory that is used to study the relationship between discrete entities (Harary, 1969). Graph theory originated from a problem—Seven Bridges of Königsberg—in 1735 and the concept of the Eulerian

graph was introduced in this problem (Shirinivas et al., 2010). Using this concept, many researchers applied the same theory to address similar problems and these efforts continue to grow in various domains.

Graph theory is used extensively in the biological domain (Gao et al., 2018) to abstract the epidemiology and ecology data (Mason & Verwoerd, 2007) because of the complexity of the biology system (Pavlopoulos et al., 2011). In ecology, the vertex of the graph represents the location of the species, whereas the edges represent the movement or relationship between the locations (Rayfield et al., 2011; Urban et al., 2009). This allows the employment graph theory to understand the transmission of diseases (Shirinivas et al., 2010). The contact between each individual can be captured by a graph where the vertices represent each individual and the edges represent the interaction (contact) (Eubank, 2005). These examples show how graph theory can be employed as a proxy of a network to illustrate the relationship between the discrete entities. Therefore, a research problem that can be abstracted as nodes and the nodes are connected by edges motivates to solve it with network theory.

As illustrated in Figure 2.1, three pillars must be present in order to identify the hotspot of COVID-19. These components can be seen as the nodes and the relationship between the components can be identified with the links. Hence, the graph is a visual representation of the solution to the research problem of this study where the functional definitions of the nodes and links are pre-determined. In this study, a network refers to the weighted graph wherein the weight is determined via quantification of the properties of the nodes and links.

2.4.2 Network Modelling Approaches in Epidemiology Studies

Danon et al. (2011) emphasised the popularity of network theory in epidemiology studies by comparing with other domains. The network structures can be used to formulate the transmission paths and take part in disease control measures such as contact tracing and hotspot detection (Danon et al., 2011; Kok et al., 2017). An epidemiologist using network modelling can simulate the interaction between individuals more naturally than other approaches, especially in direct transmission diseases (Danon et al., 2011). This is known as the contact network model where the nodes represent the host and an edge between the nodes represents an interaction between the host and the contacts, which can be either people, animals, or locations in the network (Bansal et al., 2007; Masuda & Holme, 2013; Meyers et al., 2006). The formulation of a contact network depends on real-life contact data (Meyers et al., 2005). Collecting large-scale contact data is an onerous task during the initial outbreak; hence, most of the contact networks are formulated using small datasets such as contact tracing data or respondent-driven sampling (Barrat et al., 2014; Danon et al., 2011; Robineau et al., 2020), whereas large scale data are gathered for longitudinal studies (Klov Dahl et al., 1994; Wylie & Jolly, 2001). According to Salathé et al. (2010), a small amount of data is sufficient to construct a contact network and obtain accurate results. Thus, contact network is the most suitable model to identify the transmission of disease when only a small amount of data is available, for instance, contact tracing data.

One of the classic examples of infectious diseases that can be simulated by employing the network model is sexually transmitted diseases (STDs) (Eames & Keeling, 2002; Klov Dahl, 1985; Newman, 2002). As STDs have a direct transmission mode, it can be easier to model with contact networks compared with other types of diseases. Network models of STDs focused on the sexual interaction between the individuals, with each individual

represented by a node in the network and sexual interaction between the individuals represented by the link formed between the nodes (Eames & Keeling, 2002). According to Danon et al. (2011), several studies on employing contact networks to investigate STDs have been conducted (Klov Dahl et al., 1994; Wylie & Jolly, 2001). These studies analyse the contact and network properties such as the degree of the nodes and centrality of the network to provide an intuitive and simple interpretation of the disease transmission. However, the contact network abovementioned is less informative as it was not made for prediction.

Tracing of STDs can be more specific using real-life data (Rocha et al., 2011). Rocha et al. (2011) have used a public dataset in Brazil consisting of two entities, namely, heterosexual males and female prostitutes, to construct a bipartite network where the sex buyers are represented by one type of node, whereas the sex sellers are represented by another type of node. The bipartite network consists of 6,642 sex sellers and 10,106 sex buyers; the edge is formed when there is a sexual contact between the sex sellers and buyers with a timestamp. Hence, a total of 50,185 edges are formed in the bipartite graph based on the dataset and the transmission probability used in the model is assumed constant. To understand the temporal effect of the network model, recent studies use the network model to solve the compartmental modelling, i.e. susceptible–infected–recovered and susceptible–infected model (Newman, 2002; Rocha et al., 2011). The network model captures the temporal effect of disease transmission after being incorporated with these temporal models.

Droplet-borne is another type of direct transmission that can also be employed with the contact network model (Bansal et al., 2006; Mossong et al., 2008; Yang & Jung, 2020; Zaplotnik et al., 2020) for formulating the contact and close contact network. Bansal et al. (2006) have used the contact network model to evaluate the efficacy of interventions for avian influenza in Canada. Using census data, each vertex in the network represents each

individual in Vancouver and the interactions among the vertices are based on social activities (school, work, hospital and others), which are categorised by their age. If an individual is vaccinated, the node and the edges that are connected to them will be removed from the network. The Bansal model assumes the healthcare workers get infected by a certain group of people who frequently visit the hospital. Based on this setting, healthcare workers have a higher risk of infection than other individuals. Similarly, Mossong et al. (2008) have formulated a large-scale contact network based on census and survey data. The population was divided into 15 age groups to obtain the number of contacts based on their age group and social activities. They found that the age group influences the speed of disease transmission especially for younger age groups as their contact duration and frequency are longer and higher than other age groups. Hence, contact duration and frequency are highly correlated to the risk of disease transmission.

Yang and Jung (2020) devised a contact network to study the spread of MERS-CoV in Korea. Small-world property is found on the network based on the network centrality as the contact of individuals can be clustered. Zapolnik et al. (2020) introduced the uncertainty to the contact network, which is known as perturbation factors. These perturbation factors are the incubation period, attack rate, mortality and other clinical parameters. Similarly, Yang and Jung (2020) and Zapolnik et al. (2020) used the census data to build the contact network. Moreover, So et al. (2020) have used the network properties—connectedness of the network—to analyse the risk of COVID-19 in several regions in China. Unlike others, the So model focused on the connectivity of the incidents between the regions instead of human contacts.

Together, these studies indicate the ability of the network model in terms of contact tracing and predict the risk of infection for each individual in the network. As mentioned

above, directly transmitted diseases can be easily applied in a network model than indirectly transmitted diseases (Danon et al., 2011) because the latter need to consider more factors to formulate the network. Regarding transmission risk, these studies had assumed the transmission probability in the edges is constantly based on assumptions (Bansal et al., 2006; Mossong et al., 2008) or depends on the degree of the nodes (Rocha et al., 2011; Yang & Jung, 2020; Zaplotnik et al., 2020). According to Danon et al. (2011), the edge between the vertices shall be included in other factors to provide comprehensive information on the transmission path of diseases. This statement is consistent with the discussion in Section 2.3, especially for indirectly transmitted diseases.

The common indirect transmission modes are vector-borne, airborne and vehicle-borne (fomite). Hence, the intuitive contact network is not suitable for these modes because the transmission path will involve the host and other components, as shown in Figure 2.1. To better understand the transmission of these modes, some researchers had used the bipartite graph or BNM to formulate the network for vector-borne (Bisanzio et al., 2010; Eze, 2013; Kok, 2017; Zhang et al., 2015) and airborne diseases (Ancel Meyers et al., 2003; Eubank et al., 2004; Gao et al., 2016).

Table 2.2: Related Indirect Transmission Disease Studies Applying Contact Network

Source	Disease	Nodes
Kok et al. (2017)	Dengue	Node 1: Patient Node 2: Visited locations Edges: Contact strength
Gao et al. (2016)	Influenza	Node 1: Human Node 2: Visited locations Edges: Probability of visit
Eze et al. (2011)	Malaria	Node 1: Patient Node 2: Visited locations Edges: Contact strength
Bisanzio et al. (2010)	Vector-borne disease	Nodes 1 and 2: Human Edge: Transmission
Eubank et al. (2004)	Smallpox	Node 1: Human Node 2: Visited locations Edges: Connected if the human visit the locations
Ancel Meyers et al. (2003)	<i>Mycoplasma pneumoniae</i>	Node 1: Patient and caregiver Node 2: Ward Edges: Connected if the patient resides in the ward or caregiver works in the ward

Two discrete nodes are introduced in these studies: human and location. Ancel Meyers et al. (2003) have built a bipartite graph to study the transmission of *Mycoplasma pneumoniae*; however, the location node was only focused on hospital wards, which linked with the patient and caregivers. Hence, the human node in the Ancel model consists of the patient and caregiver. Based on this model, the transmission between location and humans can occur. Although the Ancel model is empirical and only focused on the transmission within the hospital, this provided a foundation for other researchers to formulate a network

to study the transmission between humans and location. In addition, Eubank et al. (2004) have formulated a bipartite graph using the census, land-use and population mobility data in the USA to simulate the disease transmission. Unlike the Ancel model, the Eubank model extended the coverage of the location node in the network, which encompassed all locations in the study area. This study found that the small-world behaviour is shown in the contact network among the people connected by location as a hub. Moreover, the early detection of an outbreak can be identified from the location using the bipartite graph. Consequently, other researchers can be encouraged to use the bipartite graph to detect the early outbreak based on location, that is, hotspots (Lessler et al., 2017).

According to Eze (2013), for vector-borne diseases, for instance, mosquito-borne disease, the mosquitoes fly within certain distances from a particular location, and the behaviour of the mosquitoes is affected by the environmental properties of the location; thus, the location entity is used in BNM instead of mosquito entity. Kok (2017) used this finding to identify the location node in BNM to identify dengue hotspots in Malaysia. The breakthrough of these studies is quantifying the weight of the edge which has been denoted as link weight in their network. The link weight for each pair of linked nodes is computed based on the quantified values of these nodes. Moreover, environmental properties such as temperature, relative humidity, precipitation and altitude are used to quantify the location node in BNM.

Although the network shown in these studies focused more on the spatial component, it also emphasised on the effect of temporal component. In the location node, these studies defined three parameters to capture the behaviour of mosquitoes in the location: life cycle, survival rate and biting rate. The quantification of these parameters required the temperature of the location and the temperature is taken precisely based on the mobility data. For

instance, the biting rate of mosquitoes depends on the temperature by the time an individual visit that location; thus, the biting rate is varying from time to time. Therefore, this necessitates the employment of BNM for hotspot detection study of network model, which resolves the abovementioned issues too raised by Danon.

The primary concern of vector-borne diseases is the behaviour of the vector, whereas airborne diseases need to consider the behaviour of the host, agent and their simultaneous interaction. As highlighted in Section 2.3.1, the ventilation rate affects the risk of transmission of airborne diseases where it is varied based on the environmental properties and setting of a particular location. Hence, BNM is used to relate the movement of the host and the transmission path. In addition, Gao et al. (2016) have formulated an indoor contact network with a revised Eubank model (Eubank et al., 2004), as discussed above. Likewise, the two unique entities defined in this network are human and location. An indoor contact network is an indoor location visited by two or more people at the same time. Unlike other contact networks of airborne diseases that are formulated based on interaction among hosts, the Gao model formulates the network based on host and location. The Gao model used the census data of Hong Kong to construct an indoor contact network, that is, the population of Hong Kong was divided into eight groups based on age and occupation. The location node was categorised into seven groups according to the main usage of the location and assigned with the size of the places, density of occupants and ventilation rate of the location. To pair the human node with the location node, a fixed daily routine was established for each population group to simulate the visiting time and location for each individual based on a pre-defined probability.

The ventilation rate is used to estimate the risk of infection using Well–Riley equation (Equation 2.1) and integrated with an indoor contact network. The Gao model

focused on the risk of infection of each individual based on their visited location simulated using a modified susceptible–exposed–infected–recovered model. Based on the findings of the simulation, the ventilation rate played an important role in the airborne transmission of diseases, as mentioned in Section 2.3.1. This work justified the role of BNM in airborne-transmitted diseases and provides a solution to resolve the issue raised by Danon, that is, how to formulate the contact network for airborne-transmitted disease (Danon et al., 2011). However, the transmission risk in outdoor locations is also important in airborne-transmitted diseases and the stability of the pathogen is one of the crucial factors to identify disease hotspots, which is neglected in the Gao model.

According to Kok (2017), hotspot detection studies in the epidemiology domain are scarce. Hence, in this study, BNM will be employed to identify hotspots of COVID-19. As COVID-19 can be transmitted in several modes such as droplet-borne, airborne and fomite, the network formulated in this study need to integrate these modes and the factors listed in Section 2.3.1. We model the COVID-19 transmission based on the contact tracing data and stability of the pathogen, and risk of infection of susceptible individuals will be integrated into the network as well to capture the transmission path of COVID-19.

2.5 Summary

This chapter discussed the review of related literature of this study, which is hotspot detection studies in epidemiology in terms of identifying the critical factors, i.e. environmental predictors, building properties, human mobility and the approaches used. After the reviews, three pillars are always existing in the research structure of all risks of transmission of diseases, i.e. host, location and environmental properties, as shown in Figure 2.1. These three pillars are interacting with each other; thus, the structure of these three pillars forms the basis of the network modelling approach which will be discussed in Chapter 3.

Additionally, the emergence of network modelling is discussed by reviewing the strengths, gaps and weaknesses of common approaches used in the disease transmission domain. Various studies that employed network modelling approaches to address disease transmission are presented based on the disease transmission mode. These studies demonstrate the strong ability of network modelling approaches in disease transmission issues but lack consideration for the risk of infection in the human population. One of the tough challenges for all researchers in this domain is integrating more than one transmission mode in a network model. It is insufficient to study the transmission of certain diseases, especially COVID-19, which can be transmitted directly and indirectly between humans.

According to the reviews on the application of BNM, it was efficient in studying the transmission of vector-borne and airborne diseases. Thus, BNM is selected as a potential solution to address the limitation of current network modelling approaches. It is employed in this study because our research problem is to resolve the heterogeneity of humans and locations, integrate different transmission routes of the diseases, and incorporate all important predictors to identify the hotspot and risk of infection of susceptible individuals. In this study, a BNM-RMF is used to formulate the bipartite network model for COVID-19, as shown in Figure 3.1. The formulation of BNM will be discussed in the next chapter.

CHAPTER 3

BIPARTITE NETWORK MODEL FOR COVID-19 CONTACT TRACING

3.1 Introduction

This chapter presents the methodology employed in the study, which is the Bipartite Network Model-Research Methodology Framework (BNM-RMF). The activities of both stage 1 and 2 of BNM-RMF will be described and justified. In stage 1, the primary data used to construct the bipartite graph is obtained from the contact tracing data provided by the Bintulu Health Office. The problem is defined by combining assumptions related to the study, hotspot detection studies, and the epidemiology triangle from the literature review found in Chapter 2. These components are used to formulate the basic building blocks of the bipartite graph, which describe the relationship between the host, pathogen, location, and environmental properties of COVID-19. The outcome of stage 1 is a bipartite graph of COVID-19. In stage 2 of BNM-RMF, the network formulation is introduced by quantifying the parameters of the nodes and the weights of the links between connected nodes. The result is a complete bipartite network model for COVID-19 contact tracing, which will be utilized to determine the hotspots in Chapter 4.

3.2 Bipartite Network Model-Research Methodology Framework

This study models the spread of COVID-19 using a bipartite contact network approach to increase the accuracy of COVID-19 contact tracing. To achieve this goal, several stages are involved, as shown in Figure 3.1. The Figure shows the main activities of BNM-RMF approach with a web-based graph search algorithm in identifying malaria hotspots by Eze (2013), detecting dengue hotspots by Kok (2017), and using species habitat suitability by Liew (2016). The framework encompasses three stages: problem characterisation, model

construction and analysis and evaluation. Stage 1 (problem characterisation) and stage 2 (model construction) will be introduced latterly in this chapter; stage 3 (model analysis and evaluation) will be presented in chapter 4.

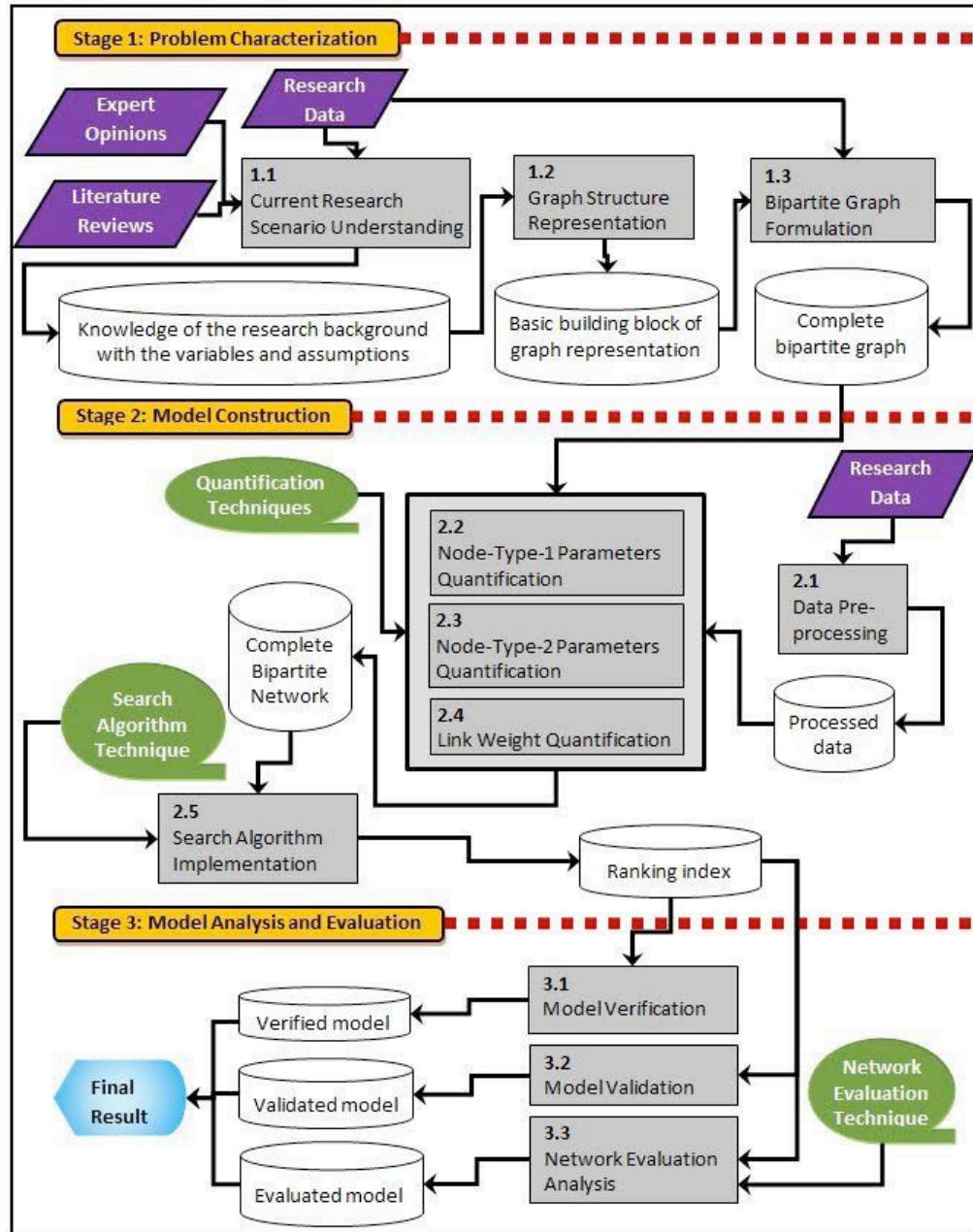


Figure 3.1: Bipartite Network Modeling Research Methodology Framework (BNM-RMF) (Liew, 2016)

3.2.1 Problem Characterization

The previous chapter concluded that a bipartite network is the best representation for the research problem. A bipartite network consists of two different node types and links that connect these nodes. Before forming the basic graph structure, the assumptions of this study were produced to ensure that the research problem is best depicted by a bipartite network.

- I. COVID-19 is a disease that spreads through multiple transmission modes and has a human host. It can be transmitted directly from an infected person through droplets, or indirectly through droplets and airborne particles.
- II. Airborne transmission can occur in indoor settings with crowded or poorly ventilated environments.(Godri Pollitt et al., 2020).
- III. The virus expelled from an infected individual can transmit indirectly through the air or remain on various surfaces.
- IV. The stability of the virus can be affected by environmental properties such as temperature and relative humidity.

Based on these assumptions, only locations visited by infected individuals are considered significant in contact tracing. These location nodes are assumed to be hubs where all possible surfaces may have been infected with the virus and remain so for a certain duration. Hence, this study hypothesizes that only human nodes and location nodes are significant in formulating the network model. This study aims to analyze the association between humans and the virus in the locations visited by infected individuals by quantifying the link weight.

The graph edges between two node types represent the correlation, that is, the contact between humans and the location in this study. Therefore, the transmission rate of COVID-

19 was assumed to be proportional to the intensity of the contacts between the adjacent nodes.

3.2.2 Basic Building Block of the Bipartite COVID-19 Contact Network Model

Eze (2013) mentioned that the epidemiological triangle is a visual representation graph that consists of three vertices which are agent, host, and environment (Dembek, Kortepeter & Pavlin, as cited in Eze, 2013; Huerta & Leventhal, as cited in Eze, 2013, p. 22). The triangle is used by Eze and Kok (2017) as one of the bases of formulating their work on malaria and dengue transmission as shown in Figure 3.2. The three epidemiologic factors in the epidemiological triangle are interdependent and used to be identified as the basic building blocks. As stated in assumptions, the main host is human, and the location is assumed contain the virus which act as agent. Therefore, the basic building block for COVID-19 would be similar to dengue and malaria mentioned above.

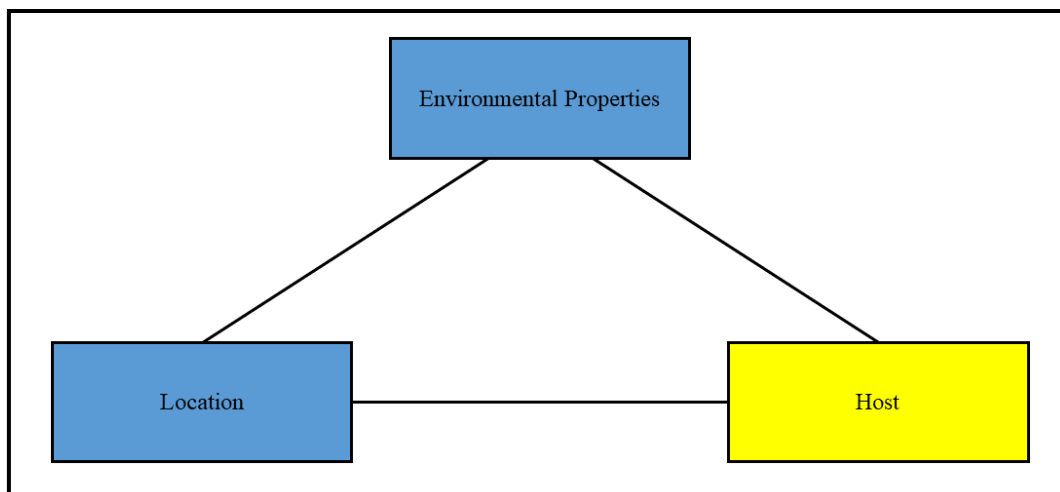


Figure 3.2: Epidemiological Triangle (ET)

Location and environmental properties are two risk factors that are strongly related. Liew (2016) had indicated that the location component specifies the spatial features of certain environmental properties. Simultaneously, environmental properties characterise a

specific location and differentiate one location from another. Thus, the location component can be viewed as a component containing the environmental properties and ET is further modified, as shown in Figure 3.3.

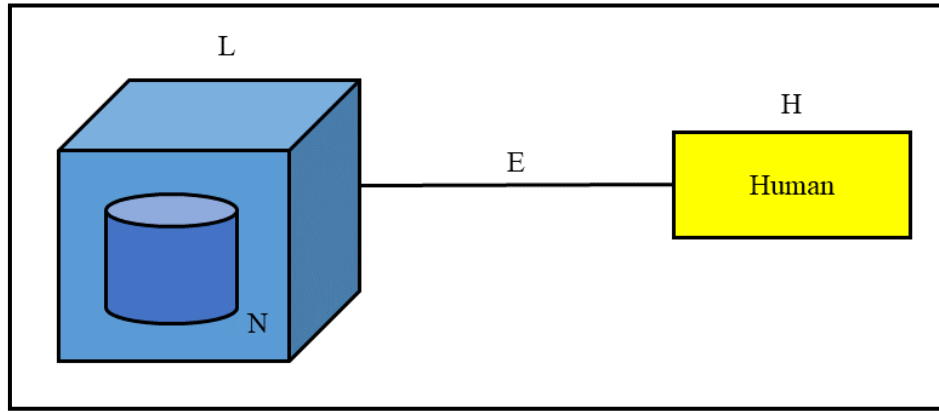


Figure 3.3: Basic Building Block of BCC Network

As shown in Figure 3.3, the location is denoted as L, the component of environmental properties is denoted as N, and the component of the human is denoted as H. The former three-vertex graph structure (Figure 3.2) of ET is modified to a two-vertex graph structure (Figure 3.3). As mentioned in Section 3.2.1, the location is strongly related to the environmental properties. Figure 3.2 illustrated the graph structure that can be used as the basic building block of the BNM, which is one of the research objectives of this study. Therefore, the two different nodes are human and location nodes. In this study, the host is a human; From this point onward, the term ‘human’ used in this thesis refers to the COVID-19 patient (infected) and possible contact (susceptible) unless otherwise specified.

As an outcome, a network structure is formed which consists of the set of location nodes (L), set of human nodes (H), and set of edges (E), as shown in Figure 3.3. The set of edges is the link between the location and human nodes. This network can be referred as a bipartite network, which is a heterogeneous network of two node types. The graph structure is termed as bipartite COVID-19 contact (BCC) graph, whereas the network model is termed

as BCC network. The difference between the BCC graph and network is that the graph is an unweighted bipartite graph for a visual representation of the BCC network, whereas BCC network is a weighted bipartite graph that indicates both the topological and functional relationship of its elements—the bipartite nodes and their correspondiv links (Rayfield et al., 2011 in Liew (2016, p. 38)). The relationship between nodes in the BCC network is evaluated by its link weight. To quantify the contact strength values, appropriate parameters related to location and human nodes need to be considered, as these nodes are highly associated with the increased probability of COVID-19 transmission.

An example of a BCC network can be formulated by using a basic building block. Figure 3.4 illustrates the completed bipartite network. Component sets are denoted as C , that is, $C = C(H, L, E)$ where L and H depict the set of nodes and E represents the set of edges. Set H is the set of human consisting of six elements, that is, $H = \{H1, H2, H3, H4, H5, H6\}$; set L is the set of locations consisting of five elements, that is, $L = \{L1, L2, L3, L4, L5\}$; and set E is the set of links that connect the elements of L and H consisting 11 elements, that is, $E = \{e1, e2, e3, e4, e5, e6, e7, e8, e9, e10, e11\}$, which can be expressed as $\{L1H1, L1H2, L2H1, L3H1, L3H3, L4H2, L4H3, L4H4, L5H4, L5H5, L5H6\}$.

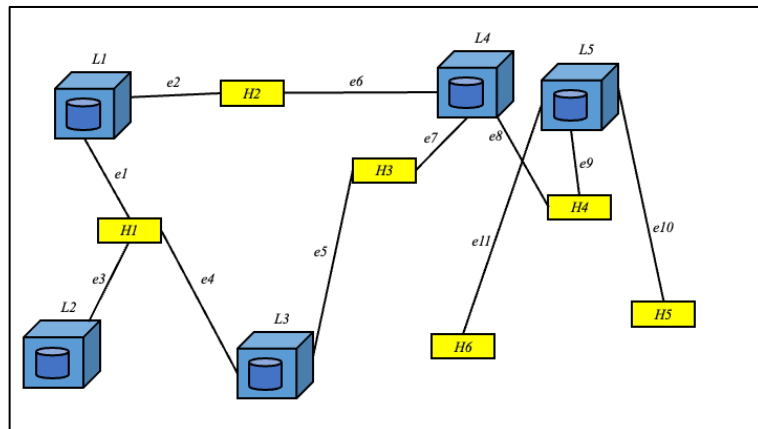


Figure 3.4: An example of a BCC network

In summary, the basic building block of the BCC network is formulated from the ET (Figure 3.1) which is then modified into a graph structure that consists of two node types (Figure 3.2). The graph structure influences the use of the bipartite network as there is no link from a node to another of the same type; the link is only formed between two different node types. The definition of the link between the nodes will be discussed in the next section.

3.2.3 Research Data and Potential Parameters

As explained in the previous section, it is clear that the link between two node types is the set of edges in this study. The edges can be considered as the contact strength of the human and location nodes. In this section, the representation of the link according to the research data will be derived. Further, this section introduces the potential parameters mentioned in Chapter 4 that will be used to quantify the link weight.

The raw data are real contact tracing data of COVID-19 obtained by the public health authorities in Bintulu. For the contact tracing of COVID-19, it began once a person tested positive. The raw data consist of a patient's 14-day history before the date of onset, that is, the date and time the patient visits a particular location, the details of activity in that location, the contact, and the contact details. Table 3.1 displayed each of these raw data attributes.

Table 3.1: Attributes of Research Raw Data

No	Dataset	Attributes
1	The contact tracing investigation form	i. Day (number of nth days before onset) ii. Date (the date of the nth day before onset) iii. Details of activities (place went) iv. Contact (number of contacts and the category of contact)
2	Line listing of contact	i. Date of performing contact tracing ii. Category of contact iii. Name of the contact iv. IC number

Table 3.1 continued

		v. Age vi. Gender vii. Nationality viii. District of contact The current address of the contact
3	Building ventilation dataset	i. Ventilation rate based on building/premises type

First, the contact tracing investigation form is provided to the patient; the patient then fills it out with the necessary details. The location listed in the investigation will be a subset of a new dataset, which is the location dataset. The location dataset includes the location visited by the patient and the residential housing address of the patient, for instance, ‘Masjid Daerah Bintulu’ or ‘MDS MART’. These locations will be converted into GPS coordinates (latitude and longitude) as it is not asked in the investigation form.

Second, the line listing for those close contacts with the patient is attached as sub-datasets of the raw data. All of the contacts identified by the public health authorities will be listed in this dataset. Finally, these contacts will be categorised according to their relationship with the index patient. For instance, ‘workplace’ means that the contact and the patient are a colleague or meet in the workplace, and ‘social’ means that the contact and the patient meet in a public or private gathering event. In addition, the ventilation dataset is the minimum air change rate of each type of building recommended by the American Society of Heating, Refrigerating and Air-Conditioning Engineers (ASHRAE). The air change rate is retrieved from ASHRAE 2013 guideline (American Society of Heating & Engineers, 2013).

Based on the discussion mentioned in the previous sections and Chapter 2, potential parameters can be derived after analysing the raw data in Table 3.1. The potential parameters

listed in Table 3.2 with their respective symbols and units of measurement have no unit if there is no notation in brackets. The human and location categories are derived from the raw dataset discussed before. For the environmental category, an online weather data application programming interface (API) is used. The next chapter will discuss the data processing for these potential parameters. Based on the research data, the link between the nodes was formed when the human visited the locations. The formulation of the BCC graph is explained in the following section.

Table 3.2: Potential Parameters Used

No.	Category	Potential parameter
1	Human	Human status Frequency of patient visit in a location, Fh The total duration of stay in a location, D (h)
2	Location	Latitude, x Longitude, y Location type, L_{type} The air change rate of the location, Q (h^{-1}) Frequency of a location visited by a patient, Fl
3	Environmental	Mean surface air temperature, T ($^{\circ}\text{C}$) Mean relative humidity, RH (%)

3.2.4 Bipartite Graph Formulation

This section summarises the third process in stage 1 of the BNM-RMF, which explains the formulation of the complete BCC graph using the basic building block derived from Figure 3.3. The human and location nodes and the connections between these nodes are retrieved from the research data from Section 3.3.1.

This study collected a total of six sets of contact tracing investigation forms provided randomly by the Bintulu Health Office. In this study, the network is formed based on a single

contact tracing investigation form. To understand the effect of removing the residential contacts from the network towards the location ranking, one dataset is used to formulate two different networks. Thus, a total of seven sets of network models will be formulated, as shown in Table 3.3. Noticed that the seventh BCC network model listed in Table 3.3 is extended from the first BCC network model by including the family contacts (11 humans) using dataset D6.

Table 3.3: Detailed Information of BCC Network Model

Dataset	BCC network model	Onset date	Epidemiology week	Number of human nodes	Number of location nodes
D6	1	18 February 2021	7	1 Patient (index patient) 30 contacts	21
D1	2	21 January 2021	3	11 humans (1 index patient and 10 contacts)	6
D2	3	25 January 2021	4	10 human (1 index patient and 9 contacts)	3
D3	4	27 January 2021	4	39 human (1 index patient and 38 contacts)	11
D4	5	1 February 2021	5	30 human (1 index patient and 29 contacts)	12
D5	6	15 February 2021	7	36 human (1 index patient and 35 contacts)	2
D6	7	18 February 2021	7	1 Patient (Index patient) 42 contacts	22

The first BCC network model, as shown in Table 3.3, is used for a case study to demonstrate the formulation process in this section. A total of 31 individuals are identified

in this network, which included 1 index patient. The 30 contacts are workplace, social and religious contacts. These 31 individuals are given a unique code which is consequently identified as the 31 human nodes of the BCC network. They are labelled as $H1$ to $H31$, where H is the symbol used to indicate the human node and the numbers 1 to 31 are used to differentiate one human from another. $H1$ is the index patient of this BCC network model.

Based on the mobility data of $H1$ in the first BCC network, 21 locations have been visited by $H1$. These locations are labelled as $L1$ to $L21$, where L is the symbol used to indicate the location node and the numbers 1 to 21 are used to differentiate one location from another. The identification of the location is determined by using Algorithm 3.1 which will be discussed in the next section.

Table 3.4: Movement of Each Human Host

Human node (H)	Location node (L)	Degree of human node
$H1$	$L1, L2, L3, L4, L5, L6, L7, L8, L9, L10, L11, L12, L13, L14, L15, L16, L17, L18, L19, L20, L21$	21
$H2$	$L1$	1
$H3$	$L1$	1
$H4$	$L1$	1
$H5$	$L1$	1
$H6$	$L1$	1
$H7$	$L1$	1
$H8$	$L1$	1
$H9$	$L1$	1
$H10$	$L1$	1
$H11$	$L1$	1
$H12$	$L1$	1
$H13$	$L1$	1
$H14$	$L1$	1
$H15$	$L8, L11, L12, L17, L19$	5
$H16$	$L8, L11, L12, L17, L19$	5
$H17$	$L8, L11, L12, L17, L19$	5
$H18$	$L8, L11, L12, L17, L19$	5

Table 3.4 continued

<i>H19</i>	<i>L8, L11, L12, L17, L19</i>	5
<i>H20</i>	<i>L8, L11, L12, L17, L19</i>	5
<i>H21</i>	<i>L8, L11, L12, L17, L19</i>	5
<i>H22</i>	<i>L16</i>	1
<i>H23</i>	<i>L16</i>	1
<i>H24</i>	<i>L16</i>	1
<i>H25</i>	<i>L16</i>	1
<i>H26</i>	<i>L3</i>	1
<i>H27</i>	<i>L3</i>	1
<i>H28</i>	<i>L3</i>	1
<i>H29</i>	<i>L3</i>	1
<i>H30</i>	<i>L3</i>	1
<i>H31</i>	<i>L3</i>	1

As shown in Table 3.4, one human node can visit more than one location; the contacts of the index patient are connected via a common location. Hence, the location acts as a hub to connect the human node in the network. A total of 79 unique links were formed when a human visited a location. These 79 links are eventually the input that will be used in the formulation of the BCC network model that will be introduced in the next chapter.

Based on the nodes and links identified, the complete BCC graph of the chosen network is formulated, as shown in Figure 3.5. BCC graph is the bipartite graph for the fifth network. The set of bipartite nodes is denoted by H and L subsets and the set of links formed between set H and L is denoted by E . Set H is the human node consisting of 31 elements:

$$H = \{H1, H2, H3, H4, H5, H6, H7, H8, H9, H10, \dots, H31\}$$

and set L is the location node consisting of 21 elements:

$$L = \{L1, L2, L3, L4, L5, L6, L7, L8, L9, L10, \dots, L21\}.$$

Set E is the set of links that join H and L and has 79 elements:

$$E = \{e1, e2, e3, e4, e5, e6, e7, e8, e9, e10, \dots, e79\}.$$

As shown in Figure 3.5, the links are formed when one location is visited by one human node. For instance, the ‘ $H1L1$ ’ represents the link formed when the index patient ($H1$) visited the location ($L1$) node on February 4 to 12, 2021, and contacted with the human node with labels $H2$ to $H14$, as shown in Table 3.4.

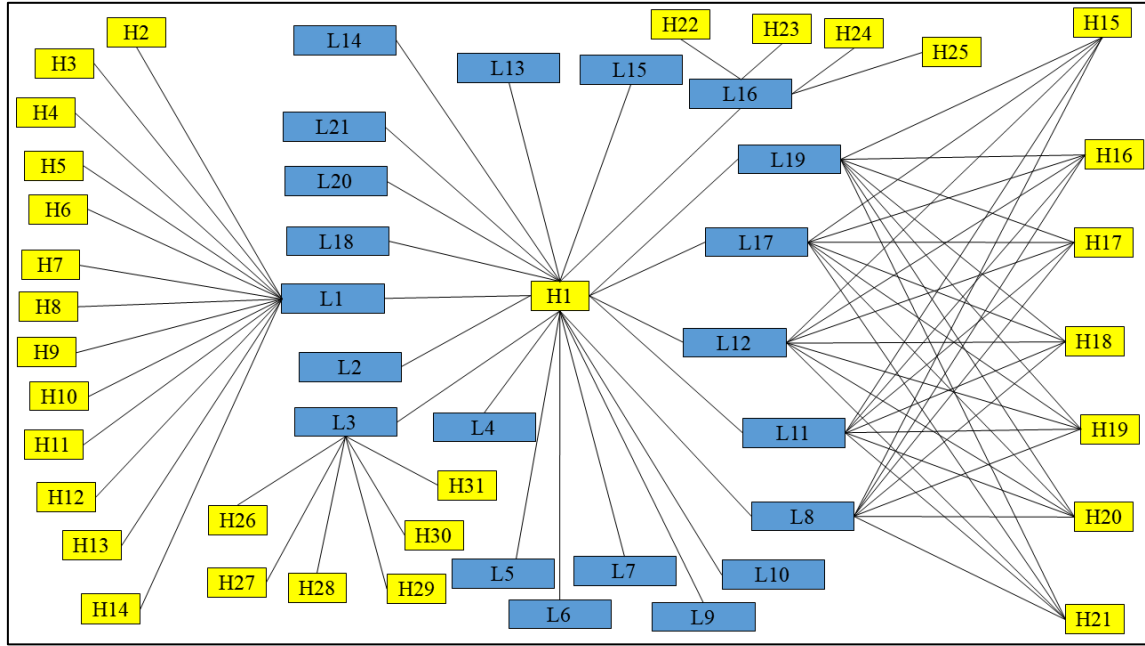


Figure 3.5: Complete Bipartite COVID-19 Contact Graph for COVID-19 patient

In summary, the BCC graph of the network is a graph of $31H$ by $21L$ with 79 edges. The degree of each location is $\{14, 1, 7, 1, 1, 1, 1, 8, 1, 1, 8, 8, 1, 1, 1, 4, 8, 1, 8, 1, 1\}$. On the contrary, the degree of each human node is listed in Table 3.4. The sum of the degrees of the location or human nodes is 79, which is also the total number of edges in the BCC graph. Based on Figure 3.5, there is no direct connection between human nodes or location nodes themselves, fulfilling the definition of a bipartite graph.

In the next stage, the BCC graph will be formulated into a complete BCC network by computing the values of the link weight, which is a process described in stage 2 of BNM-RMF. The BCC graph will be used as an input to stage 2 which can quantify the node

parameters (in Process 2.2 and 2.3) and the link weight (in Process 2.4). Chapter 4 discusses the process of stage 2.

3.3 Model Construction

The second stage has five main processes: data pre-processing, type 1 node parameter quantification, type 2 node parameter quantification, link weight quantification and search algorithm implementation. The purpose of the first process is to ensure that the data are complete, especially when real data are used, and to prepare the data for the next four processes. Data normalisation, interpolation and transformation will be utilised where appropriate. The secondary data will be aggregated and analysed. As the data are raw and unformatted, they would need to be pre-processed. For instance, the location address needs to be pre-processed into GPS coordinates including longitude and latitude. Based on the coordinates, environmental properties such as temperature and humidity can be accordingly identified and quantified. The search algorithm implementation will be covered in next chapter.

3.3.1 Data Pre-processing

As mentioned in Section 3.3.1, the location addresses are retrieved using the contact tracing investigation form, but the GPS coordinates are not provided. The details required on this study will be in a text file format labelled as ‘contact_tracing_ (network number)’ with the ‘.txt’ extension. The network number is listed in Table 3.3. The format of the text file after being pre-processed is similar to Table 3.4 by adding the date column. The names in the investigation form will be replaced with a unique ID which will be explained in the next section.

A location without a GPS coordinate in the text file is a raw data and is not suitable in this study. Therefore, the location data need to be digitised and the distance between each location node needs to be computed to avoid data redundancy in the BCC network, as shown in Figure 3.6. As shown in Figure 3.6, the first step is to compare the new location address or locality with the existing location node in the database. This study assumes every location name (point of interest) is different from each other; hence, the address is directed to the same existing location node that needs to be removed. Second, if the location does not exist in the database, the location will be converted into a GPS coordinate using a generator. For instance, a location reported in the investigation form is ‘Masjid Daerah Bintulu’; afterwards, the location search is performed using the GPS coordinate generator to obtain the respective coordinates, as illustrated in Figure 3.6 and Algorithm 3.1.

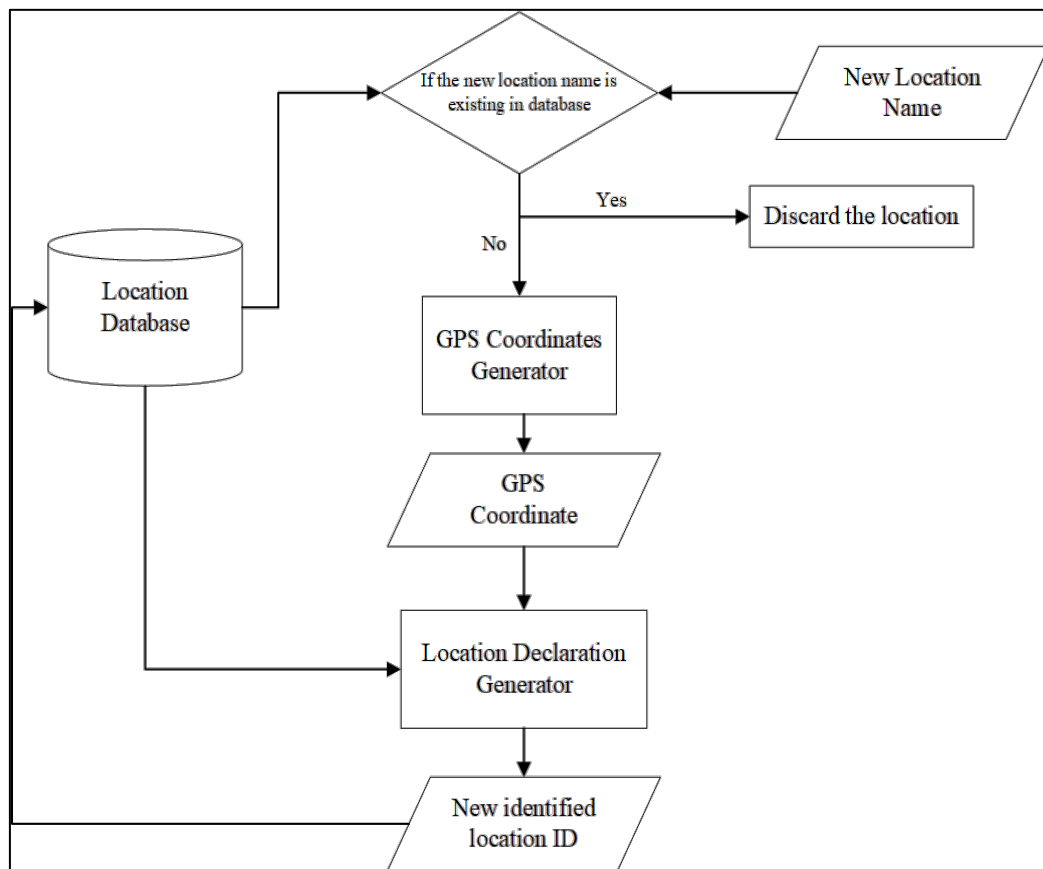


Figure 3.6: Pre-processing of the Location Node

Algorithm 3.1: GPS Coordinates Generator

FUNCTION GPS Coordinates Generator

Input: list of new locations

API_KEY $\leftarrow$ 'SECRET KEY'

res $\leftarrow$ CALL google_geocode (list of new locations, API_KEY)

gps_coordinates $\leftarrow$ List (new locations, res\$geometry\$location)

Output: list of new locations with gps_coordinates

END

Google Maps API is the service utilised for the implementation of GPS coordinate generator using R language. The google_geocode() is a function provided by an open-source library 'googleway' in R (Cooley, 2020). Hence, the process of API is executed through google_geocode(). The generator then produced a list of the new locations with GPS coordinates in LatLng format (latitude, longitude) to be processed in the next step.

However, GPS coordinates of some localities recorded in the investigation form cannot be generated using the automated Algorithm 3.2; hence, the manual way of retrieving these coordinates will be conducted. For instance, one location retrieved from the investigation form was 'RUMAH ADIK DI TAMAN KEMENA'. This location name is human-readable but not recognisable by the Google Maps API. Therefore, the first step is extracting the address component in the string which is 'TAMAN KEMENA'. The district name will be concatenated to 'TAMAN KEMENA', resulting in 'TAMAN KEMENA, Bintulu'. The next step is passing the string 'TAMAN KEMENA, Bintulu' to the search column on the Google Maps website, as shown in Figure 3.7.

The red rectangle box in Figure 3.7 indicated the GPS coordinates for 'TAMAN KEMENA, Bintulu' which are '3.17385, 113.05123'. The geocoded location will be coded with a new identified location ID using a location declaration generator, as shown in Figure

3.6. The distance between each location node is ignored because this study considers every location as unique if the location name/address is different from other locations in the database. Algorithm 3.2 details the implementation process of the location declaration generator

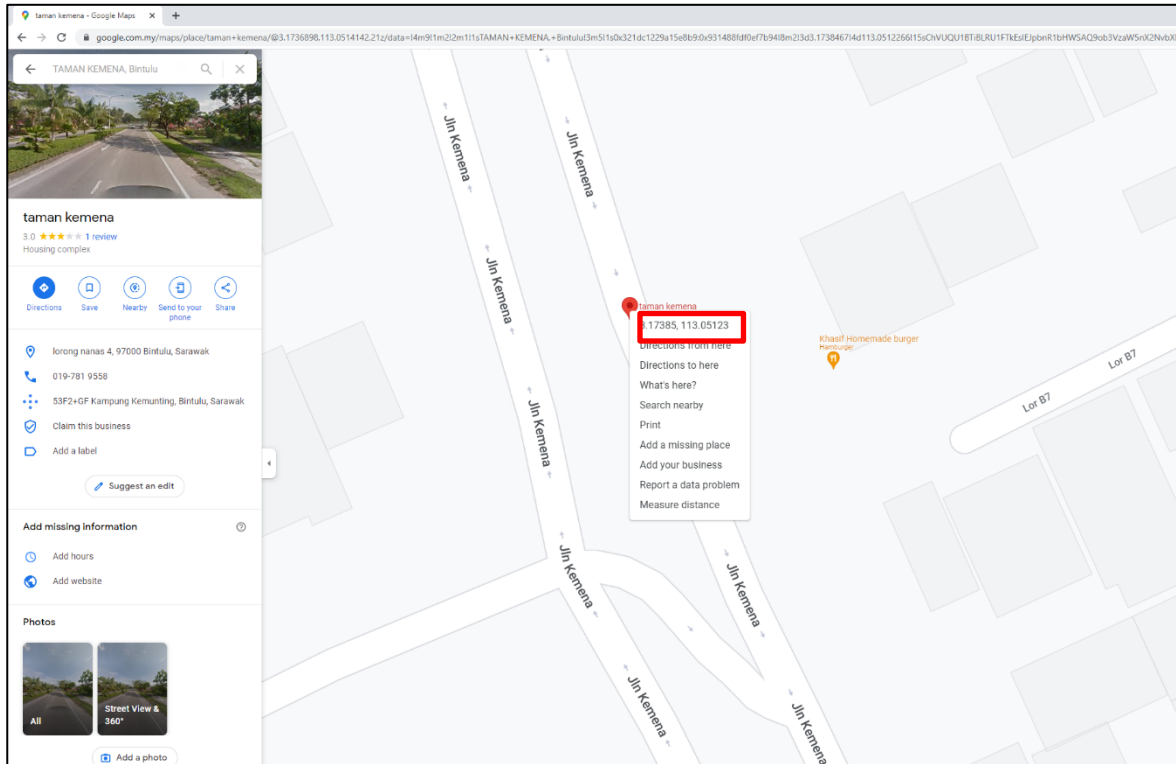


Figure 3.7: Manual Searching GPS Coordinates

Algorithm 3.2: Location Node Declaration Generator

FUNCTION Location Node Declaration Generator

Input: list of new locations with GPS coordinates, existing location node

$new_loc_node \leftarrow$ list of new locations with GPS coordinates

$exist_loc_node \leftarrow$ existing location node

$lastID \leftarrow CHECK (Row(exist_loc_node))$

$newID \leftarrow CONCATENATE("L", lastID+1 : lastID+CHECK(Length(new_loc_node)))$

Output: list of new locations with $gps_coordinates$ and node label

END

The expression of the location node in this study is prefixed with ‘*L*’ followed by an integer. The integer after the prefix is the row number of each location in the database. As for the list of geocoded locations, they are numbered after the last row number in the location database. For instance, the last number in the location database is 20; hence, the integer for the first item in the list of geocoded locations is 21, which is labelled as ‘*L21*’. All new location nodes are updated to the location database after being labelled with a node ID.

Table 3.5: Location Node with GPS Coordinates for First BCC network

Location node	GPS coordinates
<i>L1</i>	3.195955, 113.054506
<i>L2</i>	3.200107, 113.060860
<i>L3</i>	3.167168, 113.047160
<i>L4</i>	3.199246, 113.057647
<i>L5</i>	3.199179, 113.060837
<i>L6</i>	3.197694, 113.056076
<i>L7</i>	3.196108, 113.053640
<i>L8</i>	3.199040, 113.060836
<i>L9</i>	3.200794, 113.058691
<i>L10</i>	3.168762, 113.051625
<i>L11</i>	3.198104, 113.055793
<i>L12</i>	3.197320, 113.055444
<i>L13</i>	3.169877, 113.049386
<i>L14</i>	3.191315, 113.047658
<i>L15</i>	3.180072, 113.036974
<i>L16</i>	3.173860, 113.051216
<i>L17</i>	3.181154, 113.038469
<i>L18</i>	3.171220, 113.043080
<i>L19</i>	3.171955, 113.042263
<i>L20</i>	3.179512, 113.039457
<i>L21</i>	3.171449, 113.047822

The human node is the identity of patients and close contacts and this can be retrieved from the contact tracing investigation mentioned in Section 3.2.3. To ensure the identity of patients and close contacts is confidentially protected, an algorithm-generated ID has been used. The investigation form is attached with the line listing of the patient and contacts such as name, age, gender, occupation and residential address. This information can help in identifying the human node ID to avoid redundancy. The BCC network model used consists of 31 human nodes (1 patient with 30 close contacts). Thus, the human nodes are denoted as $H1$ to $H31$, where $H1$ is the patient and $H2$ to $H31$ are the close contacts.

3.3.2 Pre-Process of Parameter

Pre-processing of the parameters such as location type, premise category and human health status is required for node quantification in the following sections. The contact investigation form records patients' movement 2 weeks before the date of onset along with their health status. Hence, two types of human nodes can be identified from the form according to their health status: positive or negative (asymptomatic) during the contact tracing period. The parameter H_s in Equation 3.1 is the health status of an individual in the contact tracing investigation form. As the contact tracing form used in this chapter only consists of one patient ($H1$), then the H_s of $H1$ is 1 and for $H2$ to $H32$ are 0.

$$H_s = \begin{cases} 1, & \text{if the individual is positive of COVID - 19} \\ 0, & \text{if the individual is negative of COVID - 19} \end{cases} \quad \text{Equation 3.1}$$

The activities conducted by the human at the location visited need to be determined to check whether it is indoor or outdoor. The parameter L_{type} in Equation (3.2) is the location that is either indoor or outdoor.

$$L_{type} = \begin{cases} 1, & \text{if the location is indoor} \\ 0, & \text{if the location is outdoor} \end{cases} \quad \text{Equation 3.2}$$

Next, the type of building (B_{type}) of the location visited by the patient also needs to be identified. This can be carried out according to the activities performed by the patient that are recorded in the contact investigation form. If the activities were not mentioned in the form, the manual approach explained in Section 3.3.1 will be used. Based on the location retrieved from the form, the B_{type} can be summarised into several categories, as shown in Table 3.6.

Table 3.6: List of building types with the air change rate (Q) according to the ASHRAE guideline

B_{type}	List of building	Q
1	Places of religious worship	5
2	Restaurant dining rooms	5
3	Residential	5
4	Office space	5
5	Supermarket	7.5
6	Mall common areas	7.5
7	Bank	7.5
8	Retail	7.5
9	School	10
10	Beauty and nail salons	20
11	Gym, sports arena (play area)	20

The B_{type} is used to obtain the air change rate (Q) of the building for each location. The values of Q are based on the recommended air change rate by the ASHRAE guideline (American Society of Heating & Engineers, 2013). According to ASHRAE, the value of Q

was the required minimum ventilation rate for indoor air quality, that is, the ideal air quality level for humans.

The temperature and relative humidity data of each location node are retrieved from Word Weather Online, as shown in Algorithm 3.3. If the patient visits the location more than once on a certain date, the average value of the temperature and relative humidity will be used as the final value for the location node on that date.

Algorithm 3.3: Environmental Properties Extractor

FUNCTION Environmental Properties Extractor

 Input: Details of location node visited by the patient

 loc_details $\leftarrow$ List (GPS coordinate, date of visit, time of visit, time of leave)

 API_KEY $\leftarrow$ "SECRET KEY"

 FOR j = 1 to CHECK(ROW(loc))

 res $\leftarrow$ CALLING (URL, API_KEY, loc_details)

 T[j] $\leftarrow$ res\$temperature

 H[j] $\leftarrow$ res\$humidity

 END FOR

 Output: Mean value of temperature and humidity for the location nodes with date

END

3.3.3 Quantification of Location Node Parameters

This section discusses the Process 2.2 of BNM-RMF (Figure 3.1) that quantifies the parameter values of type 1 node. The type 1 node is the location nodes. The potential parameters of location and environmental properties displayed in Table 3.2 will be considered. Based on Table 3.2, two derived parameters will be introduced in the location node, as shown in Figure 3.8.

As mentioned in Chapter 2, the virus expelled by an infected individual usually remain in a location either in aerosol form (Anderson et al., 2020; Kohanski et al., 2020; Zuo

et al., 2020) or on surfaces of an object (Fiorillo et al., 2020; Kim et al., 2020; Luo et al., 2020). The behaviour of the virus in these two forms will influence the COVID-19 transmission in a particular location. Thus, the stability of the virus is used to measure the level of infection in a particular location.

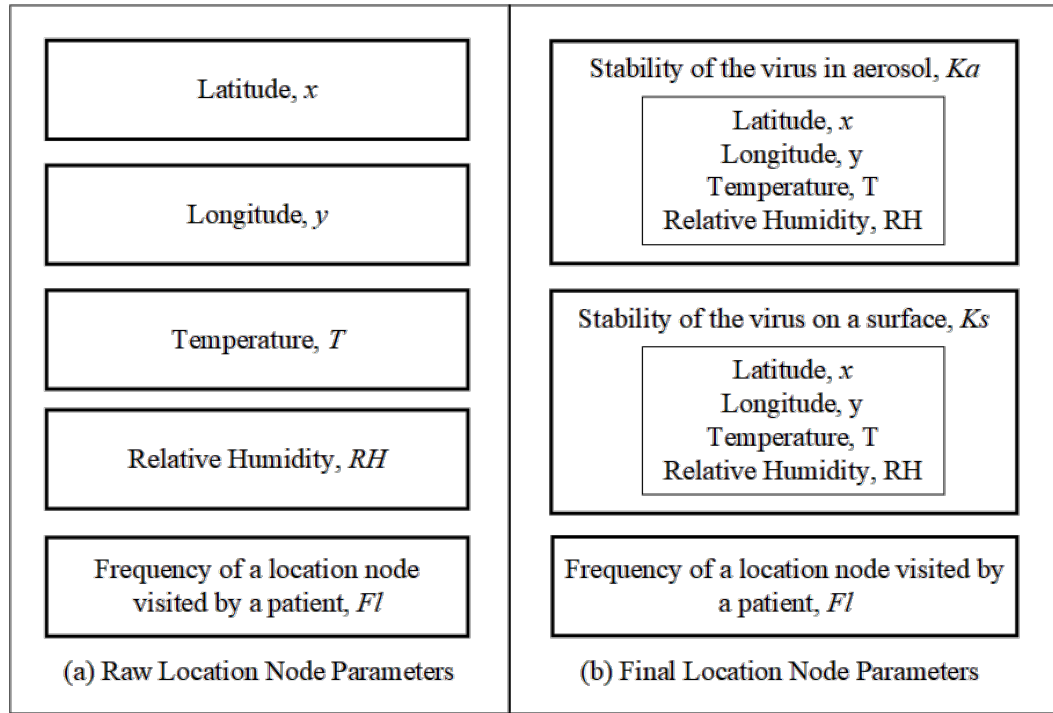


Figure 3.8: Parameters of BCC Network Location Node at Process 2.2 in BNM-RMF

Quantification of Stability of the virus in aerosol, Ka

As mentioned in Chapter 2, one of the indirect transmission modes of COVID-19 is airborne transmission. Hence, most of the casual contact will be identified if an infected individual is present within the same space, regardless of social distancing. However, indirect transmission is found in several case studies even though those cases visit the same location at different period (Cai et al., 2020; Zhang et al., 2021). The aerosol is found to play an important role in these cases.

In this study, the stability of the COVID-19 pathogen in aerosol form depends on the decay rate of the virus, whereas the decay rate is determined by the decay constant of the

virus particles. Thus, the decay constant is the quantity of interest used to quantify the stability of the virus. As aforementioned, the stability of the virus in aerosol form is influenced by several environmental factors such as temperature and relative humidity. Hence, a quantification of the correlation between the stability of the virus and temperature and relative humidity is required.

Unfortunately, in Malaysia, there are no direct laboratory data on the abovementioned correlation. Hence, this study uses the available model developed by Dabisch et al. (2021) that illustrated the influence of temperature (T) and relative humidity (RH) on the stability of virus in aerosol form. According to Dabisch et al., an experiment was conducted to determine the decay constant of the virus by measuring the stability of the virus under several environmental conditions. The temperature was set from 10 °C to 40 °C and relative humidity from 20% to 70%. The effect of intensity of sunlight was taken into account, which was simulated using ultraviolet B from darkness to 1.9W/m².

As the experiment investigates the various combinations of the impact of environmental properties, Dabisch et al. used a stepwise regression analysis to obtain a regression model, as shown in Equation 3.3. Equation 3.3 defined the correlation of the decay constant of infectivity of SARS-CoV-2 ($k_{infectivity}$) with the effect of temperature (T), relative humidity (RH) and sunlight (S).

$$k_{infectivity} = 0.16030 + 0.04018 \left(\frac{T - 20.615}{10.585} \right) + 0.02176 \left(\frac{RH - 45.235}{28.665} \right) + 0.14369 \left(\frac{S - 0.95}{0.95} \right) + 0.2636 \left(\frac{T - 20.615}{10.585} \right) \left(\frac{S - 0.95}{0.95} \right) \quad \text{Equation 3.3}$$

However, obtaining the temperature and relative humidity for an indoor environment is an arduous task. The environmental properties of all indoor places varied, and no available data can be used in this study. According to Nguyen et al. (2014), a strong correlation

between indoor and outdoor temperatures exists when the outdoor temperature is ≥ 12.7 °C. Because the average temperature in Malaysia is 26.4 °C (Jamaludin et al., 2015), which is twice than 12.7 °C, the statement supports this study to use the outdoor temperature as the value of temperature.

However, the correlation between the indoor and outdoor relative humidity is < 0.6 using the Pearson correlation coefficient, indicating a moderate correlation. In addition, the absolute humidity is not available in this study; hence, relative humidity will be the only option to measure the humidity of the location. According to de Dear et al. (1991), In Singapore, the relative humidity is 73.5% for indoors without air-conditioning and 55.5% for indoors with air-conditioning. Moreover, in Malaysia, the average relative humidity recorded is from 74% to 86% (Jamaludin et al., 2015), which is close to the values found by de Dear, Leow and Foo. Thus, the outdoor relative humidity will be used as the value of relative humidity in this study.

Similarly, the intensity of the sunlight is difficult to estimate in this study due to the lack of study data. Most of the location nodes in this study are indoor buildings; thus, there is no presence of sunlight. The ultraviolet (UV) in indoor buildings was usually emitted by incandescent filament lamps, halogen lamps, electric discharge lamps, and others (Quill et al., 2007). According to Quill et al., these light sources only contributed a fraction of UV with a wavelength of above 300 nm or 320 nm. Hence, we assume the wavelength of indoor UV to be 310 nm ($0.1 \text{ Wm}^{-2}\text{nm}^{-1}$). Thus, using Equation 3.3, the argument S is assigned 0.1 for indoor building, as shown in Equation 3.4:

$$k_{infectivity} = 0.16030 + 0.041018 \left(\frac{T - 20.615}{10.585} \right) + 0.02176 \left(\frac{RH - 45.235}{28.665} \right) - 0.15457 \left(\frac{T - 20.615}{10.585} \right) \quad \text{Equation 3.4}$$

where the value of S of outdoor location depends on the time of visit. Three values of S (0, 0.9, 1.9 W/m²) provided in the experiment by Dabisch et al. (2021) are used for the outdoor location. Hence, Equation 3.5 was used to compute the stability of the virus in aerosol form in this study.

$$Ka = \begin{cases} \text{Equation 4.4,} & \text{if } L_{type} = 1 \\ \text{Equation 4.3,} & \text{otherwise} \end{cases} \quad \text{Equation 3.5}$$

Finally, the values of temperature and relative humidity retrieved using Algorithm 3.3 are substituted to Equation 4.5 to obtain the stability of virus in aerosol form, Ka . All the location nodes visited by H1 are recorded, as shown in Table 3.4. The result of Ka is displayed in Table 3.7 and will be used in the link weight quantification later. Note that higher decay constant indicates a fast reduction in the infection rate of the virus in aerosol form in a location.

Table 3.7: Stability of the Virus in Aerosol, Ka of Each Location Node calculated based on Equation 3.5

<i>L node</i>	<i>Ka</i>
<i>L1</i>	0.072647
<i>L2</i>	0.070790
<i>L3</i>	0.072906
<i>L4</i>	0.071592
<i>L5</i>	0.071382
<i>L6</i>	0.074714

Table 3.7 Continued

<i>L7</i>	0.071408
<i>L8</i>	0.071956
<i>L9</i>	0.070117
<i>L10</i>	0.071931
<i>L11</i>	0.071678
<i>L12</i>	0.070919
<i>L13</i>	0.070919
<i>L14</i>	0.072437
<i>L15</i>	0.073067
<i>L16</i>	0.072688
<i>L17</i>	0.071170
<i>L18</i>	0.070031
<i>L19</i>	0.072308
<i>L20</i>	0.073067
<i>L21</i>	0.070919

Quantification of Stability of the virus on surface, *Ks*

Indirect transmission of COVID-19 not only happens in the form of aerosol but also via contaminated surfaces (Rahman et al., 2020). As mentioned in Chapter 2, several studies claim that the active virus remains on the surface from an hour to several days (Guo et al., 2020; Santarpia et al., 2020; van Doremalen et al., 2020). The droplets expelled by the infected patient can carry the active virus (Bueckert et al., 2020) and droplets of larger size can fall on surfaces of the object (Fernstrom & Goldblatt, 2013) and are touched or inhaled by the susceptible. Hence, surfaces with high touch frequencies provide suitable conditions that allow the virus to thrive longer, resulting in more people being infected.

Likewise, the term stability used here is consistent with the stability of the virus in aerosol form discussed in Section 3.2.1. The decay constant of the virus on the surface represents its stability. The stability of the virus on the surface of the object is influenced by the temperature, relative humidity, and types of the surfaces.

However, there are no available data and experiments conducted in Malaysia that investigate the stability of the virus on various surfaces; thus, an available experiment based on a literature review is employed in this study. To identify the correlation of the effect of temperature and relative humidity with K_s , the experiment must consider the highlighted effect. Moreover, Biryukov et al. (2020) conducted this experiment to examine the effect of temperature and relative humidity on the stability of the virus on surfaces.

The main purpose of the experiment conducted by Biryukov et al. is to further extend the studies conducted by van Doremalen et al. (2020) and Chin et al. (2020). These two previous studies identified the behaviour and stability of the virus on several surfaces using a single combination of indoor environmental properties. Both of the studies proved that the transmission of COVID-19 via contaminated surfaces is possible. However, Biryukov et al. stated that the result of those studies is hard to generalise; more combinations of temperature, relative humidity, and types of surfaces should be subsumed in the experiment.

Therefore, the type of surfaces used in the experiment carried out by Biryukov et al. were stainless steel, acrylonitrile butadiene styrene plastic, and nitrile rubber glove—the common materials often found in public places and used to create personal protective equipment. In the experiment, the temperature (T) is from 24 °C to 35 °C, whereas the relative humidity (RH) is from 20% to 80%. These environmental conditions and types of surfaces provide numerous combinations to obtain the half-life of the virus.

Biryukov et al. adopted a regression model to obtain a predictive model that estimates the half-life of the virus by utilising the data collected from the experiment with different surfaces and environmental conditions, as shown in Equation 3.6:

$$t_{\frac{1}{2}} = 32.426272 - 0.622108T - 0.153707RH \quad \text{Equation 3.6}$$

where $t_{1/2}$ is the half-life of the virus. As mentioned earlier, the decay constant is used to quantify the stability of the virus on the surface. Hence, Equation 3.8 is refined from Equation 3.6 based on the correlation of half-life and decay constant defined in Equation 3.7, where λ represents the decay constant and is replaced with Ks in Equation 3.8.

$$t_{\frac{1}{2}} = \frac{\ln(2)}{\lambda} \quad \text{Equation 3.7}$$

$$Ks = \frac{\ln(2)}{32.426272 - 0.622108T - 0.153707RH} \quad \text{Equation 3.8}$$

As mentioned in Section 4.3.1, the values of T and RH are based on the outdoor value obtained using Algorithm 3.3 (Table A.2). The last step is to substitute the data points displayed in Table 3.4 in Equation 3.8 to quantify the stability of the virus on the surface, Ks . Table 3.8 presents the results of Ks and will be used in link weight quantification later.

Table 3.8: Stability of the Virus on Surface, Ks of Each Location Node calculated based on Equation 3.8

L node	Ks
$L1$	0.26665
$L2$	0.221751
$L3$	0.265519
$L4$	0.240808
$L5$	0.25565

Table 3.8 Continued

$L6$	0.329577
$L7$	0.238638
$L8$	0.255846
$L9$	0.226249
$L10$	0.26012
$L11$	0.255024
$L12$	0.241374
$L13$	0.241374
$L14$	0.270311
$L15$	0.254343
$L16$	0.247553
$L17$	0.224144
$L18$	0.207524
$L19$	0.240764
$L20$	0.254343
$L21$	0.241374

Quantification of Number of Times a Location is visited by human infectious, Fl

The number of times a location is visited by an infected person (Fl) is one of the location node parameters of this study. As mentioned before, only the locations visited by the infected person in the network are identified as the location node. The number of visiting times implemented by other people is not considered because only the infected person is assumed to carry and spread the virus. Hence, Fl can be denoted as the frequency or number of times a location is visited by an infected person.

Fl denotes the frequency of a location node visited by an infected human node of the BCC network. As discussed in Section 3.2.2, an infected human node can be indicated by H_s

when $H_s = 1$, and a link matrix is formed to count the Fl (Eze, 2013; Kok, 2017; Liew, 2016), as calculated in Equation 3.9 and presented in Table 3.9.

$$Link_{Matrix_{BCC\ network}(L_i H_j)} = \begin{cases} 1, & \text{if } L_i \text{ is linked to } H_j \\ 0, & \text{if } L_i \text{ is not linked to } H_j \end{cases} \quad \text{Equation 3.9}$$

Where L_i = location node i of BCC network

H_j = human node j of BCC network

$i \in \{1, 2, \dots, 21\}$ and $j \in \{1, 2, \dots, 31\}$

Table 3.9: Link Matrix of BCC Network for human

(a) $H1$ to $H15$

	<i>H1</i>	<i>H2</i>	<i>H3</i>	<i>H4</i>	<i>H5</i>	<i>H6</i>	<i>H7</i>	<i>H8</i>	<i>H9</i>	<i>H10</i>	<i>H11</i>	<i>H12</i>	<i>H13</i>	<i>H14</i>	<i>H15</i>
<i>L1</i>	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0
<i>L2</i>	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
<i>L3</i>	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
<i>L4</i>	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
<i>L5</i>	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
<i>L6</i>	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
<i>L7</i>	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
<i>L8</i>	1	0	0	0	0	0	0	0	0	0	0	0	0	0	1
<i>L9</i>	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
<i>L10</i>	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
<i>L11</i>	1	0	0	0	0	0	0	0	0	0	0	0	0	0	1
<i>L12</i>	1	0	0	0	0	0	0	0	0	0	0	0	0	0	1
<i>L13</i>	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
<i>L14</i>	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
<i>L15</i>	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
<i>L16</i>	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
<i>L17</i>	1	0	0	0	0	0	0	0	0	0	0	0	0	0	1
<i>L18</i>	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
<i>L19</i>	1	0	0	0	0	0	0	0	0	0	0	0	0	0	1

Table 3.9 (a): Continued

L20	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
L21	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0

(b) *H16 to H31*

	H16	H17	H18	H19	H20	H21	H22	H23	H24	H25	H26	H27	H28	H29	H30	H31
L1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
L2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
L3	0	0	0	0	0	0	0	0	0	0	1	1	1	1	1	1
L4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
L5	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
L6	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
L7	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
L8	1	1	1	1	1	1	0	0	0	0	0	0	0	0	0	0
L9	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
L10	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
L11	1	1	1	1	1	1	0	0	0	0	0	0	0	0	0	0
L12	1	1	1	1	1	1	0	0	0	0	0	0	0	0	0	0
L13	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
L14	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
L15	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
L16	0	0	0	0	0	0	1	1	1	1	0	0	0	0	0	0
L17	1	1	1	1	1	1	0	0	0	0	0	0	0	0	0	0
L18	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
L19	1	1	1	1	1	1	0	0	0	0	0	0	0	0	0	0
L20	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
L21	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

Thus, Fl can be calculated in Equation 3.10 based on information listed in Table A.1 where the number of times that a human visited a location is denoted as Fh . The next section will discuss the process of quantification for Fh .

$$Fl_i = \sum Link_{Matrix_{BCC network}(L_i H_k)} \times Fh_{k:i} \quad \text{Equation 3.10}$$

Where L_i = location node i of BCC network

H_k = human node k of BCC network

$$k = \begin{cases} j, & \text{if } H_j H_s = 1 \\ 0, & \text{otherwise} \end{cases}$$

$H_j H_s = H_s$ of human node j of BCC network

$$i \in \{1, 2, \dots, 21\} \text{ and } j \in \{1, 2, \dots, 31\}$$

There is only one infected person in this BCC network, hence, $k = 1$. In addition, Fl for the location node LI , Fl_i , can be calculated as follows:

$$\begin{aligned} Fl_1 &= Link_{Matrix_{BCC network}(L_1 H_1)} \times Fh_{1:1} \\ &= 9 \end{aligned} \quad \text{Equation 3.11}$$

Moreover, the Fl values for the other location nodes are calculated as [9, 3, 7, 6, 3, 1, 8, 5, 3, 3, 1, 1, 1, 1, 1, 2, 2, 1, 1, 1, 1] according to the order of the location nodes. Furthermore, the next section will summarise all parameters for the location node of the BCC network.

Summarize of Parameter of Location Node in BCC Network

Three location parameters have been quantified in the previous section. Ka and Ks are quantified through the regression with four attributes: latitude (x), longitude (y), temperature (T), and relative humidity (RH). Fl is quantified using Equation 3.10.

The raw parameters initially considered for the location node of the BCC network, as shown in Figure 3.8 (a), are then further refined after the process of quantifying Ka and

K_s parameters. The latitude, longitude, temperature, and relative humidity are excluded from the location node parameters as their effect is considered in the K_a and K_s .

Consequently, three parameters are determined as the BCC network location node parameters: stability of the virus in aerosol form, stability of the virus on a surface, and the number of times a location is visited by humans, as shown in Figure 3.8 (b). Table 3.10m displayed the values of these three parameters for all 21 location nodes in the BCC network dataset.

Table 3.10: Parameters Values of 21 Location Nodes in BCC Network

Location Node	K_a	K_s	Fl
$L1$	0.072647	0.26665	9
$L2$	0.07079	0.221751	3
$L3$	0.072906	0.265519	7
$L4$	0.071592	0.240808	6
$L5$	0.071382	0.25565	3
$L6$	0.074714	0.329577	1
$L7$	0.071408	0.238638	8
$L8$	0.071956	0.255846	5
$L9$	0.070117	0.226249	3
$L10$	0.071931	0.26012	3
$L11$	0.071678	0.255024	1
$L12$	0.070919	0.241374	1
$L13$	0.070919	0.241374	1
$L14$	0.072437	0.270311	1
$L15$	0.073067	0.254343	1
$L16$	0.072688	0.247553	2
$L17$	0.07117	0.224144	2
$L18$	0.070031	0.207524	1
$L19$	0.072308	0.240764	1

Table 3.10 Continued

$L20$	0.073067	0.254343	1
$L21$	0.070919	0.241374	1

3.3.4 Quantification of Human Node Parameters

As summarised in Chapter 2, three parameters are identified for the human node of the BCC network:

- a) Indoor infection risk of a non-infectious human in a location, P_{wr}
- b) Total Duration time the human stay at one location, D
- c) Frequency of a human visiting a location, Fh

Quantification of indoor infection risk, P_{wr}

As presented in Chapter 2, the Wells–Riley equation is a well-known model adopted by many studies to estimate the risk of infection for the susceptible. It is a simple predictive model that estimates the risk of infection of the susceptible when sharing a room with one or more infected persons. A few important assumptions are defined by the Wells–Riley equation (Noakes et al., 2006; G. N. Sze To & C. Y. H. Chao, 2010):

- a. The room air is well-mixed; the quanta are distributed uniformly in the room.
- b. The air is at a steady-state which only varies with the ventilation rate.
- c. The droplets expelled by infected individuals contain the virus at a constant rate and stay in the room for a certain duration.
- d. The biological decay of the virus can be neglected (Riley et al., 1978).

Equation 3.12 defined the Wells–Riley equation where the quantum generation rate is denoted as q . However, the value of q is difficult to obtain; thus, this study adopted the

function developed by Dai and Zhao (2020). Dai and Zhao used the epidemiological data to quantify basic reproduction number, R_0 to estimate the value of q . Equation 3.13 defined the function of R_0 formulated by Dai and Zhao (2020) to obtain the value of q .

$$P = 1 - \exp\left(-\frac{Iqpt}{Q}\right) \quad \text{Equation 3.12}$$

$$q = A + B_1 R_0 + B_2 R_0^2 \quad \text{Equation 3.13}$$

Where $A = -30.27958$

$$B_1 = -44.81536$$

$$B_2 = 19.67934$$

The value of R_0 used in this study is based on the estimation made by Gill et al. (2020), which is 3.08 in the early outbreak in Malaysia in 2020. Hence, the following shows the value of q that will be used in Equation 3.12:

$$\begin{aligned} q &= -30.27958 + (-44.81536 \times 3.08) + (19.67934 \times 3.08^2) \\ &= 18.38 \end{aligned} \quad \text{Equation 3.14}$$

The value of q is corrected into four significant Figures. The parameter p in Equation 3.12 indicates the pulmonary ventilation rate of susceptible. In Equation 3.12, the parameter I indicates the number of infectors in a location. Thus, in this study, the number of infected in the BCC network is 1 (HI).

The pulmonary ventilation rate of a person is influenced by various factors such as body size, metabolic rate, gender and age (Ai et al., 2019). In this study, the body size and metabolic rate are excluded from the raw data; hence, the pulmonary ventilation rate cannot be measured efficiently. Hence, in this study, the value of p is assumed as a constant value, that is, 0.3 m^3 per hour based on the estimation of Duan (2015).

The parameter t in Equation 3.12 indicates the exposure time interval of the susceptible in a location. This is compared with one of the human node parameters D — the total duration time of the human stay at one location. Hence, the values of t for each human node (except $H1$) used in Equation 3.12 are equal to the values of D which will be covered in the next section. The last parameter of Equation 3.12, Q , indicates the room ventilation rate of a location. Thus, the value of Q will be based on the building type, as explained in Section 3.3.2.

Further, Equation 3.15 is the final equation used in the quantification of indoor infection risk which is denoted as P_{wr} in this study. P_{wr} will be computed for indoor environment and susceptible human nodes:

$$P_{wr_{H_j:L_i}} = \begin{cases} 0, & \text{if } H_j H_s = 1, \\ \left[1 - \exp\left(\frac{-Iqpt}{Q}\right)\right] \times L_{type L_i} \times Link_{Matrix_{BCC network}(H_j L_i)}, & \text{otherwise} \end{cases} \quad \text{Equation 3.15}$$

The data points $H_j:L_i$ are substituted to Equation 3.15 along with the link matrix shown in Table 3.9. The values of P_{wr} are displayed in Table 3.11 in matrix form and will be used in the quantification of link weight later.

Table 3.11: P_{wr} for Each Human Node based on Location Node calculated based on

Equation 3.15

(a) H1 to H10

	H1	H2	H3	H4	H5	H6	H7	H8	H9	H10
L1	0	0.9879	0.9879	0.9879	0.9879	0.9879	0.9879	0.9879	0.9879	0.9879
L2	0	0	0	0	0	0	0	0	0	0
L3	0	0	0	0	0	0	0	0	0	0
L4	0	0	0	0	0	0	0	0	0	0

Table 3.11 (a): Cotinued

<i>L5</i>	0	0	0	0	0	0	0	0	0	0
<i>L6</i>	0	0	0	0	0	0	0	0	0	0
<i>L7</i>	0	0	0	0	0	0	0	0	0	0
<i>L8</i>	0	0	0	0	0	0	0	0	0	0
<i>L9</i>	0	0	0	0	0	0	0	0	0	0
<i>L10</i>	0	0	0	0	0	0	0	0	0	0
<i>L11</i>	0	0	0	0	0	0	0	0	0	0
<i>L12</i>	0	0	0	0	0	0	0	0	0	0
<i>L13</i>	0	0	0	0	0	0	0	0	0	0
<i>L14</i>	0	0	0	0	0	0	0	0	0	0
<i>L15</i>	0	0	0	0	0	0	0	0	0	0
<i>L16</i>	0	0	0	0	0	0	0	0	0	0
<i>L17</i>	0	0	0	0	0	0	0	0	0	0
<i>L18</i>	0	0	0	0	0	0	0	0	0	0
<i>L19</i>	0	0	0	0	0	0	0	0	0	0
<i>L20</i>	0	0	0	0	0	0	0	0	0	0
<i>L21</i>	0	0	0	0	0	0	0	0	0	0

(b) H11 to H20

	<i>H11</i>	<i>H12</i>	<i>H13</i>	<i>H14</i>	<i>H15</i>	<i>H16</i>	<i>H17</i>	<i>H18</i>	<i>H19</i>	<i>H20</i>
<i>L1</i>	0.9879	0.9879	0.9879	0.9879	0	0	0	0	0	0
<i>L2</i>	0	0	0	0	0	0	0	0	0	0
<i>L3</i>	0	0	0	0	0	0	0	0	0	0
<i>L4</i>	0	0	0	0	0	0	0	0	0	0
<i>L5</i>	0	0	0	0	0	0	0	0	0	0
<i>L6</i>	0	0	0	0	0	0	0	0	0	0
<i>L7</i>	0	0	0	0	0	0	0	0	0	0
<i>L8</i>	0	0	0	0	0.5206	0.5206	0.5206	0.5206	0.5206	0.5206
<i>L9</i>	0	0	0	0	0.5206	0.5206	0.5206	0.5206	0.5206	0.5206
<i>L10</i>	0	0	0	0	0	0	0	0	0	0
<i>L11</i>	0	0	0	0	0	0	0	0	0	0

Table 3.11 (b): Cotinued

L12	0	0	0	0	0.5206	0.5206	0.5206	0.5206	0.5206	0.5206
L13	0	0	0	0	0	0	0	0	0	0
L14	0	0	0	0	0	0	0	0	0	0
L15	0	0	0	0	0	0	0	0	0	0
L16	0	0	0	0	0	0	0	0	0	0
L17	0	0	0	0	0.5206	0.5206	0.5206	0.5206	0.5206	0.5206
L18	0	0	0	0	0	0	0	0	0	0
L19	0	0	0	0	0.5206	0.5206	0.5206	0.5206	0.5206	0.5206
L20	0	0	0	0	0	0	0	0	0	0
L21	0	0	0	0	0	0	0	0	0	0

(c) H21 to H31

	H21	H22	H23	H24	H25	H26	H27	H28	H29	H30	H31
L1	0	0	0	0	0	0	0	0	0	0	0
L2	0	0	0	0	0	0	0	0	0	0	0
L3	0	0	0	0	0	0.996	0.996	0.996	0.996	0.996	0.996
L4	0	0	0	0	0	0	0	0	0	0	0
L5	0	0	0	0	0	0	0	0	0	0	0
L6	0	0	0	0	0	0	0	0	0	0	0
L7	0	0	0	0	0	0	0	0	0	0	0
L8	0.5206	0	0	0	0	0	0	0	0	0	0
L9	0.5206	0	0	0	0	0	0	0	0	0	0
L10	0	0	0	0	0	0	0	0	0	0	0
L11	0	0	0	0	0	0	0	0	0	0	0
L12	0.5206	0	0	0	0	0	0	0	0	0	0
L13	0	0	0	0	0	0	0	0	0	0	0
L14	0	0	0	0	0	0	0	0	0	0	0
L15	0	0	0	0	0	0	0	0	0	0	0
L16	0	0.8088	0.8088	0.8088	0.8088	0	0	0	0	0	0
L17	0.5206	0	0	0	0	0	0	0	0	0	0
L18	0	0	0	0	0	0	0	0	0	0	0

Table 3.11 (c): Continued

L19	0.5206	0	0	0	0	0	0	0	0	0	0
L20	0	0	0	0	0	0	0	0	0	0	0
L21	0	0	0	0	0	0	0	0	0	0	0

Quantification of Total Duration of a human stay at one location, D

The parameter of a total duration of a human stay at one location is denoted as D , which is identified from the contact tracing investigation. To align the definition of exposure time interval which is the values of t value used in Equation 3.12, it is equivalent to the values of D in this section; the unit of D used in this study is hours. For instance, human node $H1$ visited location node $L2$ three times, 30 minutes for each stay, which is a total of 90 minutes. Therefore, the value of D for human node j visited location node i ($D_{Hj:Lj}$) is calculated as follows:

$$D_{H1:L2} = \frac{90 \text{ minutes}}{60} = 1.5 \text{ hours} \quad \text{Equation 3.16}$$

Similarly, the total duration for other human and location nodes are calculated and the results of D are displayed in Table 3.12 for the first BCC network.

Table 3.12: Total Duration of a human stay at one location, D

Human Node	Location node	D (h)
$H1, H2, H3, H4, H5, H6, H7, H8, H9, H10, H11, H12, H13, H14$	$L1$	72
$H1$	$L2$	1.5
$H1$	$L3$	35
$H26, H27, H28, H29, H30, H31$	$L3$	15
$H1$	$L4$	3

Table 3.12 Continued

<i>H1</i>	<i>L5</i>	1.5
<i>H1</i>	<i>L6</i>	0.5
<i>H1</i>	<i>L7</i>	24
<i>H1, H15, H16, H17, H18, H19, H20, H21</i>	<i>L8</i>	5
<i>H1</i>	<i>L9</i>	1.5
<i>H15, H16, H17, H18, H19, H20, H21</i>	<i>L9</i>	1
<i>H1</i>	<i>L10</i>	1.5
<i>H1</i>	<i>L11</i>	1
<i>H1, H15, H16, H17, H18, H19, H20, H21</i>	<i>L12</i>	1
<i>H1</i>	<i>L13</i>	0.5
<i>H1</i>	<i>L14</i>	0.5
<i>H1</i>	<i>L15</i>	1
<i>H1, H22, H23, H24, H25</i>	<i>L16</i>	1.5
<i>H1, H15, H16, H17, H18, H19, H20, H21</i>	<i>L17</i>	2
<i>H1</i>	<i>L18</i>	0.5
<i>H1, H15, H16, H17, H18, H19, H20, H21</i>	<i>L19</i>	1
<i>H1</i>	<i>L20</i>	1
<i>H1</i>	<i>L21</i>	0.5

Quantification of Frequency of a human visiting a location, Fh

The last parameter included in the human node of the BCC network is the frequency of a human visiting a location, which is denoted as Fh . Human and location here refer to each of the corresponding human and location nodes. Hence, Fh is calculated as follows to record the number of times the human node j visited location node i :

$$Fh_{j:i} = \begin{cases} n, & \text{if } H_j \text{ visited } L_i \text{ } n \text{ times where } n \in \mathbb{Z}^+ \\ 0, & \text{if } H_j \text{ not visit } L_i \end{cases} \quad \text{Equation 3.17}$$

As the first BCC network formulated in Section 3.2.4 was used as a demonstration example in this study, the values of n and Fh for every human corresponding to each of the 21 location nodes are displayed in Table 3.4, which summarises the movement of the patient ($H1$) and the close contacts ($H2$ to $H31$). Table 3.13 displays the values of Fh for each human node to each location node. Note that the sum of Fh for $H1$ is the same as the sum of Fl recorded in Table 3.9. Hence, the Fh of $H1$ to each location node is the value of Fl for the particular location node.

Table 3.13: Parameter Values of $Fh_{j:i}$

(a) H1 to H16

Fh		j (Human Node)															
		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
i (Location Node)	1	9	9	9	9	9	9	9	9	9	9	9	9	9	9	0	0
	2	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	3	7	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	4	6	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	5	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	6	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	7	8	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	8	5	0	0	0	0	0	0	0	0	0	0	0	0	0	5	5
	9	3	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1
	10	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	11	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	12	1	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1
	13	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	14	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	15	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

Table 3.13: Continued

	16	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	17	2	0	0	0	0	0	0	0	0	0	0	0	0	0	2	2
	18	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	19	1	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1
	20	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	21	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

(b) H17 to H31

<i>Fh</i>		<i>j</i> (Human Node)														
		17	18	19	20	21	22	23	24	25	26	27	28	29	30	31
<i>i</i> (Location Node)	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	3	0	0	0	0	0	0	0	0	0	0	3	3	3	3	3
	4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	5	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	6	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	7	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	8	5	5	5	5	5	0	0	0	0	0	0	0	0	0	0
	9	1	1	1	1	1	0	0	0	0	0	0	0	0	0	0
	10	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	11	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	12	1	1	1	1	1	0	0	0	0	0	0	0	0	0	0
	13	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	14	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	15	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	16	0	0	0	0	0	1	1	1	1	0	0	0	0	0	0
	17	2	2	2	2	2	0	0	0	0	0	0	0	0	0	0
	18	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	19	1	1	1	1	1	0	0	0	0	0	0	0	0	0	0
	20	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	21	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

Summarize of Node Parameters in BCC Network

Three parameters are identified and quantified for the human node, namely, P_{wr} , D , and Fh , and the location node is also determined by three parameters, namely, Ka , Ks , and Fl . So, far, execution of Processes 2.2 and 2.3 in stage 2 of BNM-RMF has been achieved. The next section will discuss Process 2.4 of BNM-RMF, which is the quantification of link weight, resulting in the completion of the final BCC network.

3.3.5 Quantification of Link Weight

According to the first BCC graph formulated in Section 3.4.4, a total of 79 edges are formed from 31 human and 21 location nodes. Consequently, there are 79 link weights to be computed, which corresponded to the 79 data points. After computing the link weights, these values will be denoted as COVID-19 contact strength (CCS). CCS values constitute the link affinity between the human and location nodes. The higher CCS value indicates a stronger attachment between the human and the particular location, which results in a higher degree of human contact with the specific location.

Accordingly, Eze (2013) and Kok (2017) applied the summation rule to quantify the link weight between the bipartite nodes of malaria (Eze, 2013, p. 83) and dengue transmission networks (Kok, 2017, p. 79). Both of these studies aim to identify disease hotspots. On the contrary, Liew (2016) employed the product rule (multiplication rule) to quantify the link weight between the bipartite nodes of a dolphin habitat network (Liew, 2016, p. 122). As the main concern of this study is identical to Eze (2013) and Kok (2017), the summation rule is adopted to quantify the contact strength between the human and location nodes to obtain the CCS of the BCC network. Quantify by summation rule can capture the real effect of the value if the value are less than one. Based in the research from Eze and Kok, the summation rule demonstrated an ability in term of quantify the bipartite

network, which believed will increase the ability of BCC network to enhance the performance of contact tracing of COVID-19.

Equation 3.18 shows the quantification technique applied in this study, where $CCS_{i:j}$ is the CCS of the link between location node i and human node j :

$$\begin{aligned} CCS_{i:j} &= \left(\sum Location_{NodeParameters_i} \right) + \left(\sum Human_{NodeParameters_{j:i}} \right) \\ &= (Ka_i + Ks_i + Fl_i) + (P_{wr_{Hj:Li}} + D_{j:i} + Fh_{j:i}). \end{aligned} \quad \text{Equation 3.18}$$

Normalisation of the parameter value is performed before computing the CCS wherein each parameter value is transformed into a numerical value that ranges between 0 and 1. The maximum value of each parameter is identified and used to transform all values into a normalised value and then multiplied by 0.9 to ensure all values are between 0 and 1 but not equal to 1. In this section, parameter Fl is used to demonstrate the normalisation, as shown in Equation 3.19. After normalisation, the values of Fl have transformed from [9, 3, 7, 6, 3, 1, 8, 5, 3, 3, 1, 1, 1, 1, 1, 2, 2, 1, 1, 1, 1] into [0.9, 0.3, 0.7, 0.6, 0.3, 0.1, 0.8, 0.5, 0.3, 0.3, 0.1, 0.1, 0.1, 0.1, 0.1, 0.2, 0.2, 0.1, 0.1, 0.1, 0.1]. Likewise, the other five parameter values (Ka , Ks , P_{wr} , Fh and D) undergone the same process. Table 3.14 displays the normalised values for location node parameters.

$$Fl_{i_{norm}} = \frac{Fl_i}{Max(Fl_1, Fl_2, Fl_3, \dots, Fl_{19}, Fl_{20}, Fl_{21})} \times 0.9 \quad \text{Equation 3.19}$$

Table 3.14: Complete Normalized Parameter of Location Nodes

Location Node	Ka	Ks	Fl
$L1$	0.875099	0.728161	0.9
$L2$	0.852729	0.605551	0.3
$L3$	0.878213	0.725071	0.7
$L4$	0.86239	0.657591	0.6
$L5$	0.859858	0.698122	0.3
$L6$	0.9	0.9	0.1
$L7$	0.860168	0.651665	0.8
$L8$	0.866771	0.698657	0.5
$L9$	0.844618	0.617835	0.3
$L10$	0.866471	0.710327	0.3
$L11$	0.863423	0.696412	0.1
$L12$	0.854279	0.659136	0.1
$L13$	0.854279	0.659136	0.1
$L14$	0.872567	0.738157	0.1
$L15$	0.880161	0.694552	0.1
$L16$	0.875589	0.676011	0.2
$L17$	0.857301	0.612086	0.2
$L18$	0.843584	0.566702	0.1
$L19$	0.871017	0.65747	0.1
$L20$	0.880161	0.694552	0.1
$L21$	0.854279	0.659136	0.1

Normalised parameter values of location and human nodes are employed for obtaining CCS in this study. For instance, the link weight between location node $L2$ and human node $H1$ ($CCS_{2:1}$) is 2.077, obtained via Equation 4.18 and shown in Equation 3.20 which is corrected to four decimal places.

$$CCS_{2:1} = \left(\sum Location_{NodeParameters_1} \right) + \left(\sum Human_{NodeParameters_{2:1}} \right)$$

$$\begin{aligned}
&= (Ka_2 + Ks_2 + Fl_2) + (P_{wr_{H1:L2}} + D_{1:2} + Fh_{1:2}) \\
&= (0.8527288 + 0.6055511 + 0.3) + (0 + 0.01875 + 0.3) \\
&= 2.077
\end{aligned}$$

Equation 3.20

Similarly, the remaining CCS was computed using Equation 3.20. Table 3.15 displays the complete CCS denoted as the CCS matrix.

Table 3.15: COVID-19 Contact Strength (CCS) Matrix of BCC Network

(a) H1 to H11

	<i>H1</i>	<i>H2</i>	<i>H3</i>	<i>H4</i>	<i>H5</i>	<i>H6</i>	<i>H7</i>	<i>H8</i>	<i>H9</i>	<i>H10</i>	<i>H11</i>
<i>L1</i>	4.3033	5.1959	5.1959	5.1959	5.1959	5.1959	5.1959	5.1959	5.1959	5.1959	5.1959
<i>L2</i>	2.077	0	0	0	0	0	0	0	0	0	0
<i>L3</i>	3.4408	0	0	0	0	0	0	0	0	0	0
<i>L4</i>	2.7575	0	0	0	0	0	0	0	0	0	0
<i>L5</i>	2.1767	0	0	0	0	0	0	0	0	0	0
<i>L6</i>	2.0063	0	0	0	0	0	0	0	0	0	0
<i>L7</i>	3.4118	0	0	0	0	0	0	0	0	0	0
<i>L8</i>	2.6279	0	0	0	0	0	0	0	0	0	0
<i>L9</i>	2.0812	0	0	0	0	0	0	0	0	0	0
<i>L10</i>	2.1955	0	0	0	0	0	0	0	0	0	0
<i>L11</i>	1.7723	0	0	0	0	0	0	0	0	0	0
<i>L12</i>	1.7259	0	0	0	0	0	0	0	0	0	0
<i>L13</i>	1.7197	0	0	0	0	0	0	0	0	0	0
<i>L14</i>	1.817	0	0	0	0	0	0	0	0	0	0
<i>L15</i>	1.7872	0	0	0	0	0	0	0	0	0	0
<i>L16</i>	1.9703	0	0	0	0	0	0	0	0	0	0
<i>L17</i>	1.8944	0	0	0	0	0	0	0	0	0	0
<i>L18</i>	1.6165	0	0	0	0	0	0	0	0	0	0
<i>L19</i>	1.741	0	0	0	0	0	0	0	0	0	0
<i>L20</i>	1.7872	0	0	0	0	0	0	0	0	0	0
<i>L21</i>	1.7197	0	0	0	0	0	0	0	0	0	0

(b) H12 to H22

	<i>H12</i>	<i>H13</i>	<i>H14</i>	<i>H15</i>	<i>H16</i>	<i>H17</i>	<i>H18</i>	<i>H19</i>	<i>H20</i>	<i>H21</i>	<i>H22</i>
<i>L1</i>	5.1959	5.1959	5.1959	0	0	0	0	0	0	0	0
<i>L2</i>	0	0	0	0	0	0	0	0	0	0	0
<i>L3</i>	0	0	0	0	0	0	0	0	0	0	0
<i>L4</i>	0	0	0	0	0	0	0	0	0	0	0
<i>L5</i>	0	0	0	0	0	0	0	0	0	0	0
<i>L6</i>	0	0	0	0	0	0	0	0	0	0	0
<i>L7</i>	0	0	0	0	0	0	0	0	0	0	0
<i>L8</i>	0	0	0	3.0984	3.0984	3.0984	3.0984	3.0984	3.0984	3.0984	0
<i>L9</i>	0	0	0	2.3454	2.3454	2.3454	2.3454	2.3454	2.3454	2.3454	0
<i>L10</i>	0	0	0	0	0	0	0	0	0	0	0
<i>L11</i>	0	0	0	2.2428	2.2428	2.2428	2.2428	2.2428	2.2428	2.2428	0
<i>L12</i>	0	0	0	2.1963	2.1963	2.1963	2.1963	2.1963	2.1963	2.1963	0
<i>L13</i>	0	0	0	0	0	0	0	0	0	0	0
<i>L14</i>	0	0	0	0	0	0	0	0	0	0	0
<i>L15</i>	0	0	0	0	0	0	0	0	0	0	0
<i>L16</i>	0	0	0	0	0	0	0	0	0	0	2.6012
<i>L17</i>	0	0	0	2.3648	2.3648	2.3648	2.3648	2.3648	2.3648	2.3648	0
<i>L18</i>	0	0	0	0	0	0	0	0	0	0	0
<i>L19</i>	0	0	0	2.2114	2.2114	2.2114	2.2114	2.2114	2.2114	2.2114	0
<i>L20</i>	0	0	0	0	0	0	0	0	0	0	0
<i>L21</i>	0	0	0	0	0	0	0	0	0	0	0

(c) H23 to H31

	<i>H23</i>	<i>H24</i>	<i>H25</i>	<i>H26</i>	<i>H27</i>	<i>H28</i>	<i>H29</i>	<i>H30</i>	<i>H31</i>
<i>L1</i>	0	0	0	0	0	0	0	0	0
<i>L2</i>	0	0	0	0	0	0	0	0	0
<i>L3</i>	0	0	0	3.6908	3.6908	3.6908	3.6908	3.6908	3.6908
<i>L4</i>	0	0	0	0	0	0	0	0	0
<i>L5</i>	0	0	0	0	0	0	0	0	0
<i>L6</i>	0	0	0	0	0	0	0	0	0
<i>L7</i>	0	0	0	0	0	0	0	0	0

Table 4.11 (c) continued

<i>L8</i>	0	0	0	0	0	0	0	0	0
<i>L9</i>	0	0	0	0	0	0	0	0	0
<i>L10</i>	0	0	0	0	0	0	0	0	0
<i>L11</i>	0	0	0	0	0	0	0	0	0
<i>L12</i>	0	0	0	0	0	0	0	0	0
<i>L13</i>	0	0	0	0	0	0	0	0	0
<i>L14</i>	0	0	0	0	0	0	0	0	0
<i>L15</i>	0	0	0	0	0	0	0	0	0
<i>L16</i>	2.6012	2.6012	2.6012	0	0	0	0	0	0
<i>L17</i>	0	0	0	0	0	0	0	0	0
<i>L18</i>	0	0	0	0	0	0	0	0	0
<i>L19</i>	0	0	0	0	0	0	0	0	0
<i>L20</i>	0	0	0	0	0	0	0	0	0
<i>L21</i>	0	0	0	0	0	0	0	0	0

To transform the BCC graph displayed in Figure 3.5 into a complete BCC network, each CCS will be replaced with the corresponding link, as shown in Figure 3.9. To make the network visualization more comprehensive, the human nodes from *H2* to *H31* will be grouped based on the location they came into contact with *H1*. As depicted, those grouped human nodes share the same link weight with their paired location nodes, similar to the grouping identified by the public health officers. Hence, the COVID-19 contact tracing can be simulated using the BNM with chosen node parameters.

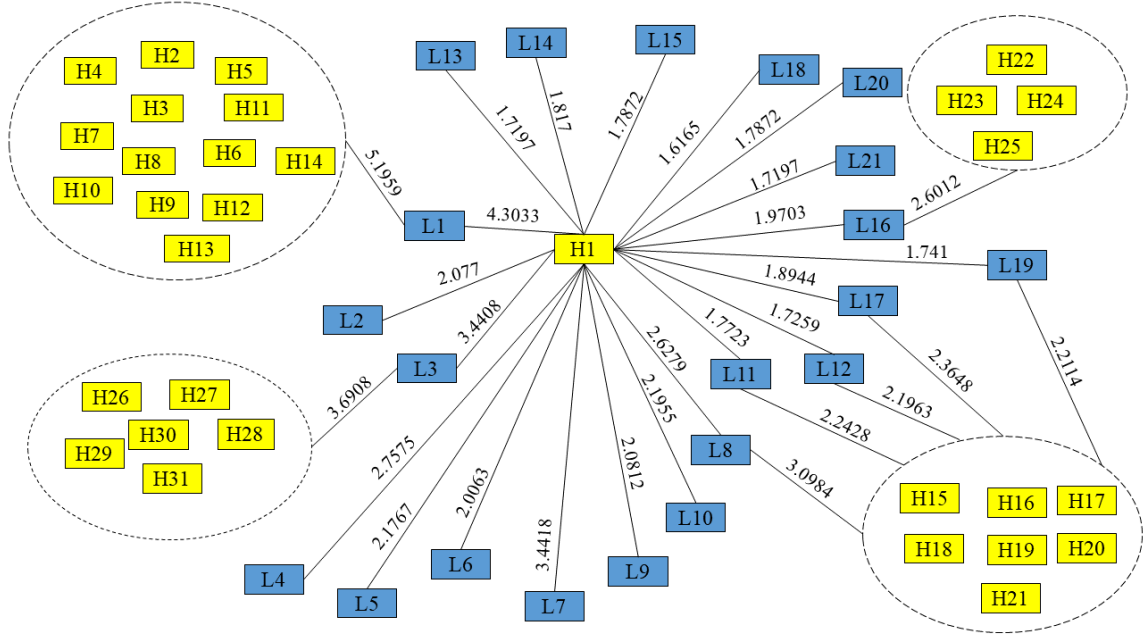


Figure 3.9: Bipartite COVID-19 contact network with clustered human node

3.4 Summary

This chapter discussed the execution of stage 1 BNM-RMF, which covered the sub-processes such as current research scenario understanding, graph structure representation and formulation of bipartite graph for the BCC network model that this study intends to formulate. As an outcome, the quantity of interest in variables and assumptions concerned is identified. From the formulation of the basic building block, a graphical representation is constructed. Finally, the BCC graph is formulated and the formal definition of the graph is established by using the research data. This chapter also demonstrated the execution of Process 2.1 until Process 2.4 of stage 2 in BNM-RMF (Figure 1.1) by using the BCC graph as input. Initially, the pre-processing of the raw data for humans, location, and required parameters was presented followed by determining the parameters for the location node. Existing studies were adopted to quantify the stability of the virus in aerosol form (Ka) and on a surface (Ks). The number of times a location is visited by a patient (Fl) was quantified using the link matrix. The three parameters used in the human node were identified: the

indoor infection risk of susceptible, P_{wr} ; total duration of a human stay at one location, D ; and frequency of a human visiting a location, Fh . The quantification of P_{wr} was based on the Well–Riley equation. Finally, the link weight of the BCC network was quantified using the summation rule. The link weight of the BCC network is termed CCS.

CHAPTER 4

IMPLEMENTATION, ANALYSIS AND EVALUATION

4.1 Introduction

This chapter presented the implementation details of a search algorithm to rank the network formulated in previous chapter which are the BCC network. This network will be presented in contact matrix and served as input for the search algorithm. Hyperlink-induced topic search (HITS) is used as search algorithm to generate the ranking of the location and human nodes by using the BCC network as search spaces. The outcome of the location and human nodes ranking are discussed. Finally, the implementation issues encountered in model construction and HITS algorithm are presented.

The design of model analysis and evaluation is discussed in this chapter. The model verification was carried out to ensure that the modelling procedure had been executed and implemented correctly. Benchmark verification was then carried out by calculating the RMSE between the ranking values produced from the benchmark system and the BCC network. The analytical verification is conducted by calculating the SRCC between the hub matrix and the CHR values. The validation process was conducted by calculating the SRCC between the targeted and validated networks. Finally, the parameter significance analysis and predictive power analysis are discussed to evaluate the performance of the network model.

4.2 Search Algorithm

The web-based search algorithm is a well-known algorithm used to rank the items based on the link details within different items. Page-rank and HITS algorithms are commonly used in network-related studies. PageRank is a well-known algorithm used to

rank web pages in search engines. It was developed by Google and is based on the idea that a page with many incoming links is likely to be important. PageRank works by treating each link as a vote for the destination page and distributing the importance of each page to the pages it links to. PageRank can be applied to bipartite networks, where there are two types of nodes, such as user-item networks in recommendation systems. HITS, on the other hand, is designed for bipartite networks and calculates two scores for each node: authority score and hub score. The authority score reflects the importance of a node as a destination for links, while the hub score reflects the node's importance as a source of links to other authoritative nodes. HITS is useful for identifying the most influential nodes in a network, such as authoritative sources of information in a topic or highly connected users in a social network. While both PageRank and HITS are used to rank nodes in networks, PageRank is better suited for general networks, while HITS is better suited for bipartite networks where there are two types of nodes.

Eze (2013) and Kok (2017) implemented an adapted HITS algorithm in their studies to rank the location node for malaria (Eze, 2013) and dengue transmissions (Kok, 2017). Liew (2016) also employed the HITS algorithm to identify the habitat suitability of dolphins. The HITS algorithm used in information retrieval and web mining to rank the importance of web pages. The HITS algorithm determines the hub and authority scores of a set of web pages, where hubs are pages that have many outgoing links to other web pages (authorities), and authorities are pages that have many incoming links from hub pages. The idea behind HITS is that a hub page that links to many authoritative pages will serve as a good starting point for a search query, and authoritative pages that are linked to by many hub pages are likely to contain valuable information related to the search query. The HITS algorithm

iteratively updates the hub and authority scores of each page until the scores converge to a stable ranking.

The main reason HITS is chosen is due to its ability to measure the importance of nodes in a network. As mentioned, the algorithm uses two measures, authority and hub scores, to identify the important nodes. In a bipartite network with two sets of nodes, such as human nodes and location nodes in a COVID-19 contact tracing network, the HITS algorithm can be used to rank the importance of both sets of nodes in terms of their relationship to each other. For example, the hub score of a location node would reflect its importance as a potential source of COVID-19 transmission, while the authority score of a human node would reflect its potential as a spreader of the virus. The HITS algorithm can provide valuable insights into the relationships between human nodes and location nodes in a COVID-19 network, and help to identify potential hotspots of transmission.

4.3 HITS Algorithm Implementation

This section discusses the implementation details of the search algorithm. The search algorithm is the ranking algorithm in this study and involves the computation of principal eigenvalues and eigenvectors. The output of the HITS algorithm for the location node is the COVID-19 hotspot ranking (CHR) values, which is a measure of the virus density of the location, whereas the output of the HITS algorithm for the human node is the COVID-19 risk-infection ranking (CRR), which indicates the risk of infection of susceptible in the network and the super-spreader if there are more than one infected individual present in the same network. The locations are ranked based on the value of CHR whereby a high-rank location indicates a high transmission risk that needs to be addressed first. In addition, a high rank for susceptible human node indicates a high chance of infection in the network, hence the need to perform early isolation and contact tracing for the susceptible. Both CHR and

CRR values range between 0 and 1, which is approximated and dominated by the values of principal eigenvectors generated from the implementation of the HITS algorithm onto the BCC network by employing the power iteration method.

The implementation process of the HITS algorithm in this study is based on the studies of Eze (2013) and Liew (2016), which focused on malaria hotspot ranking and dolphin suitability ranking, respectively. Both studies were referred by Kok (2017) to identify dengue hotspot ranking.

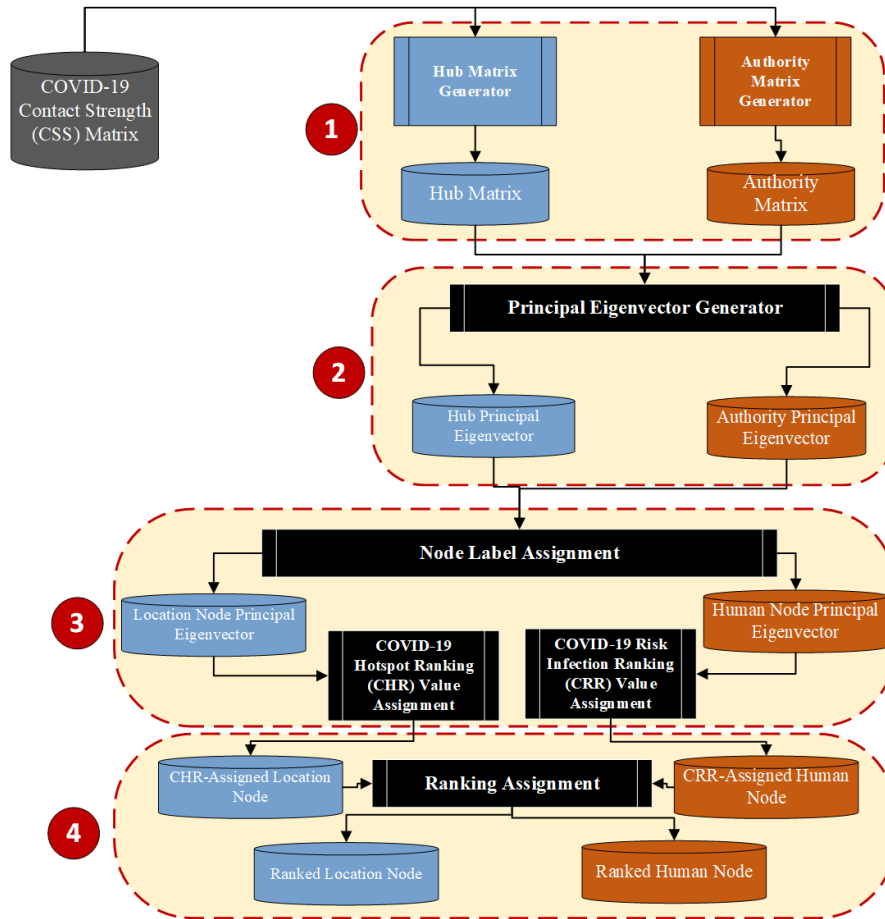


Figure 4.1: Process of the Adapted HITS Search Algorithm (Liew, 2016).

4.3.1 Implementation Platform

As mentioned before, the implementation process of HITS algorithm is adapted from Eze (2013) and Liew (2016) by using Matlab. R language is used in this study to implement

all algorithms presented in Chapter 3. Hence, the implementation of the HITS algorithm will use R in this study. R is a programming language and software environment that is widely utilized in data analysis, statistical modelling, and graphical visualization. Originating in 1993, R was created by Ross Ihaka and Robert Gentleman at the University of Auckland, New Zealand. It has since gained popularity because of its open-source nature, active user community, and extensive libraries. With various functions and packages for data manipulation, cleaning, statistical analysis, and visualization, R is a potent tool for researchers and analysts across diverse fields. The language's adaptability enables users to customize and extend existing packages and functions according to their requirements. R is compatible with multiple operating systems and can run in command-line mode or through an integrated development environment (IDE) such as RStudio.

4.3.2 Generation of Hub and Authority Matrix

The first step in Figure 4.1 is to generate a hub and authority matrices using a matrix generator. The hub and authority matrix generator, as shown in Figure 4.1, is based on Equations 4.1 and 4.2, where T is the transpose operator. These equations are modified from the studies of Eze (2013) and Liew (2016).

$$hub = CCS \times CCS^T \quad \text{Equation 4.1}$$

$$authority = CCS^T \times CCS \quad \text{Equation 4.2}$$

The input of the hub and authority matrix generator is the CCS matrix from Table 3.15 which is the step 1 in Figure 4.1. To ensure the real eigen pair has existed, the CCS matrix must be a real and non-negative matrix. The dimension of the CCS matrix for the first BCC network model is 21 (location nodes) by 31 (human nodes), which is a real matrix. The hub and authority matrices generated by Equations 4.1 and 4.2 are symmetric to ensure the

existence of eigen pair. The output of the hub and authority matrix generator is presented in Tables 4.1 and Table 4.2 respectively. The dimension of the hub matrix is 21 by 21, which represents the 21 location nodes, whereas the authority matrix is 31 by 31, which represents the 31 human nodes in the BCC network model.

Table 4.1: Hub Matrix

(a) Column 1 to 10

369.49	8.94	14.81	11.87	9.37	8.63	14.68	11.31	8.96	9.45
8.94	4.31	7.15	5.73	4.52	4.17	7.09	5.46	4.32	4.56
14.81	7.15	93.57	9.49	7.49	6.9	11.74	9.04	7.16	7.55
11.87	5.73	9.49	7.6	6	5.53	9.41	7.25	5.74	6.05
9.37	4.52	7.49	6	4.74	4.37	7.43	5.72	4.53	4.78
8.63	4.17	6.9	5.53	4.37	4.03	6.84	5.27	4.18	4.4
14.68	7.09	11.74	9.41	7.43	6.84	11.64	8.97	7.1	7.49
11.31	5.46	9.04	7.25	5.72	5.27	8.97	74.1	56.34	5.77
8.96	4.32	7.16	5.74	4.53	4.18	7.1	56.34	42.84	4.57
9.45	4.56	7.55	6.05	4.78	4.4	7.49	5.77	4.57	4.82
7.63	3.68	6.1	4.89	3.86	3.56	6.05	53.3	40.51	3.89
7.43	3.58	5.94	4.76	3.76	3.46	5.89	52.17	39.65	3.79
7.4	3.57	5.92	4.74	3.74	3.45	5.87	4.52	3.58	3.78
7.82	3.77	6.25	5.01	3.96	3.65	6.2	4.77	3.78	3.99
7.69	3.71	6.15	4.93	3.89	3.59	6.1	4.7	3.72	3.92
8.48	4.09	6.78	5.43	4.29	3.95	6.72	5.18	4.1	4.33
8.15	3.93	6.52	5.22	4.12	3.8	6.46	56.27	42.77	4.16
6.96	3.36	5.56	4.46	3.52	3.24	5.52	4.25	3.36	3.55
7.49	3.62	5.99	4.8	3.79	3.49	5.94	52.54	39.93	3.82
7.69	3.71	6.15	4.93	3.89	3.59	6.1	4.7	3.72	3.92
7.4	3.57	5.92	4.74	3.74	3.45	5.87	4.52	3.58	3.78

(b) Column 11 to 21

7.63	7.43	7.4	7.82	7.69	8.48	8.15	6.96	7.49	7.69	7.4
3.68	3.58	3.57	3.77	3.71	4.09	3.93	3.36	3.62	3.71	3.57
6.1	5.94	5.92	6.25	6.15	6.78	6.52	5.56	5.99	6.15	5.92
4.89	4.76	4.74	5.01	4.93	5.43	5.22	4.46	4.8	4.93	4.74
3.86	3.76	3.74	3.96	3.89	4.29	4.12	3.52	3.79	3.89	3.74
3.56	3.46	3.45	3.65	3.59	3.95	3.8	3.24	3.49	3.59	3.45
6.05	5.89	5.87	6.2	6.1	6.72	6.46	5.52	5.94	6.1	5.87
53.3	52.17	4.52	4.77	4.7	5.18	56.27	4.25	52.54	4.7	4.52
40.51	39.65	3.58	3.78	3.72	4.1	42.77	3.36	39.93	3.72	3.58
3.89	3.79	3.78	3.99	3.92	4.33	4.16	3.55	3.82	3.92	3.78
38.35	37.54	3.05	3.22	3.17	3.49	40.48	2.87	37.8	3.17	3.05
37.54	36.75	2.97	3.14	3.08	3.4	39.63	2.79	37	3.08	2.97
3.05	2.97	2.96	3.12	3.07	3.39	3.26	2.78	2.99	3.07	2.96
3.22	3.14	3.12	3.3	3.25	3.58	3.44	2.94	3.16	3.25	3.12
3.17	3.08	3.07	3.25	3.19	3.52	3.39	2.89	3.11	3.19	3.07
3.49	3.4	3.39	3.58	3.52	30.95	3.73	3.19	3.43	3.52	3.39
40.48	39.63	3.26	3.44	3.39	3.73	42.74	3.06	39.91	3.39	3.26
2.87	2.79	2.78	2.94	2.89	3.19	3.06	2.61	2.81	2.89	2.78
37.8	37	2.99	3.16	3.11	3.43	39.91	2.81	37.26	3.11	2.99
3.17	3.08	3.07	3.25	3.19	3.52	3.39	2.89	3.11	3.19	3.07
3.05	2.97	2.96	3.12	3.07	3.39	3.26	2.78	2.99	3.07	2.96

Table 4.2: Authority Matrix

(a) Column 1 to 15

113.5 8	22.3 6	22.3 6	22.3 6	22.3 6	22.3 6	22.3 6	22.3 6	22.3 6	22.3 6	22.3 6	22.3 6	22.3 6	22.3 6
22.36	27	27	27	27	27	27	27	27	27	27	27	27	27
22.36	27	27	27	27	27	27	27	27	27	27	27	27	27
22.36	27	27	27	27	27	27	27	27	27	27	27	27	27
22.36	27	27	27	27	27	27	27	27	27	27	27	27	27

Table 4.2 (a) continued

22.36	27	27	27	27	27	27	27	27	27	27	27	27	27
22.36	27	27	27	27	27	27	27	27	27	27	27	27	27
22.36	27	27	27	27	27	27	27	27	27	27	27	27	27
22.36	27	27	27	27	27	27	27	27	27	27	27	27	27
22.36	27	27	27	27	27	27	27	27	27	27	27	27	27
22.36	27	27	27	27	27	27	27	27	27	27	27	27	27
22.36	27	27	27	27	27	27	27	27	27	27	27	27	27
22.36	27	27	27	27	27	27	27	27	27	27	27	27	27
22.36	27	27	27	27	27	27	27	27	27	27	27	27	27
29.12	0	0	0	0	0	0	0	0	0	0	0	0	0
29.12	0	0	0	0	0	0	0	0	0	0	0	0	0
29.12	0	0	0	0	0	0	0	0	0	0	0	0	0
29.12	0	0	0	0	0	0	0	0	0	0	0	0	0
29.12	0	0	0	0	0	0	0	0	0	0	0	0	0
29.12	0	0	0	0	0	0	0	0	0	0	0	0	0
29.12	0	0	0	0	0	0	0	0	0	0	0	0	0
29.12	0	0	0	0	0	0	0	0	0	0	0	0	0
5.13	0	0	0	0	0	0	0	0	0	0	0	0	0
5.13	0	0	0	0	0	0	0	0	0	0	0	0	0
5.13	0	0	0	0	0	0	0	0	0	0	0	0	0
5.13	0	0	0	0	0	0	0	0	0	0	0	0	0
12.7	0	0	0	0	0	0	0	0	0	0	0	0	0
12.7	0	0	0	0	0	0	0	0	0	0	0	0	0
12.7	0	0	0	0	0	0	0	0	0	0	0	0	0
12.7	0	0	0	0	0	0	0	0	0	0	0	0	0
12.7	0	0	0	0	0	0	0	0	0	0	0	0	0
12.7	0	0	0	0	0	0	0	0	0	0	0	0	0

(b) Column 16 to 31

29.12	29.12	29.12	29.12	29.12	29.12	29.12	5.13	5.13	5.13	5.13	12.7	12.7	12.7	12.7	12.7	12.7
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
35.44	35.44	35.44	35.44	35.44	35.44	35.44	0	0	0	0	0	0	0	0	0	0
35.44	35.44	35.44	35.44	35.44	35.44	35.44	0	0	0	0	0	0	0	0	0	0
35.44	35.44	35.44	35.44	35.44	35.44	35.44	0	0	0	0	0	0	0	0	0	0
35.44	35.44	35.44	35.44	35.44	35.44	35.44	0	0	0	0	0	0	0	0	0	0
35.44	35.44	35.44	35.44	35.44	35.44	35.44	0	0	0	0	0	0	0	0	0	0
35.44	35.44	35.44	35.44	35.44	35.44	35.44	0	0	0	0	0	0	0	0	0	0
35.44	35.44	35.44	35.44	35.44	35.44	35.44	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	6.77	6.77	6.77	6.77	0	0	0	0	0	0
0	0	0	0	0	0	0	6.77	6.77	6.77	6.77	0	0	0	0	0	0
0	0	0	0	0	0	0	6.77	6.77	6.77	6.77	0	0	0	0	0	0

Table 4.2 (b) continued

0	0	0	0	0	0	0	6.77	6.77	6.77	6.77	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	13.62	13.62	13.62	13.62	13.62	13.62
0	0	0	0	0	0	0	0	0	0	0	13.62	13.62	13.62	13.62	13.62	13.62
0	0	0	0	0	0	0	0	0	0	0	13.62	13.62	13.62	13.62	13.62	13.62
0	0	0	0	0	0	0	0	0	0	0	13.62	13.62	13.62	13.62	13.62	13.62
0	0	0	0	0	0	0	0	0	0	0	13.62	13.62	13.62	13.62	13.62	13.62
0	0	0	0	0	0	0	0	0	0	0	13.62	13.62	13.62	13.62	13.62	13.62
0	0	0	0	0	0	0	0	0	0	0	13.62	13.62	13.62	13.62	13.62	13.62

4.3.3 Generation of Hub and Authority Principal Eigenvector

The step 2 in Figure 4.1 aims to identify the principal eigenvector for the hub and authority matrices using Equation 4.3, where Hu is the hub matrix mentioned earlier, as shown in Table 4.1; A is the authority matrix mentioned earlier, as shown in Table 4.2; λ_H is the eigenvalues of the hub matrix; λ_A is the eigenvalues of authority matrix; v is an eigenvector of hub matrix; and η is an eigenvector of authority matrix.

$$Huv = \lambda_H v$$

$$A\eta = \lambda_A \eta \quad \text{Equation 4.3}$$

To calculate the principal eigenvectors and eigenvalues, the power iteration method in Equation 4.3 is adopted (Eze, 2013; Kok, 2017; Liew, 2016). In Equation 4.4, the initial vector is denoted as v_0 , which is an approximation of the principal eigenvector, whereas $\|Huv_k\|$ and $\|A\eta_k\|$ are the length of the product of the matrix and its k^{th} eigenvector. In this study, for a linear vector of 21 elements, the initial guess value, v_0 , is 1's row vector, and for a linear vector of 31 elements, the initial guess value, η_0 , is also 1's row vector. The k is a linear vector that records the step of iteration that starts from 0 until the iteration stops.

$$v_{k+1} = \frac{Hu v_k}{||Hu v_k||}$$

$$\eta_{k+1} = \frac{A \eta_k}{||A \eta_k||} \quad \text{Equation 4.4}$$

Vectors v_k and η_k are multiplied by the respective matrix Hu or A which is executed at every iteration in Equation 4.24 and normalised until the stopping criteria are met. The principal eigenvector is produced when the iteration stops. Two stopping criteria are used in this study: (a) the estimation error between the current and previous eigenvalues and (b) the maximum number of iterations allowed. The stopping criteria for the iteration are (1) the threshold value of estimation error (1E-07) and (2) the maximum number of iterations (1000). Using the hub matrix in Table 4.1 and authority in Table 4.2 as input, the output is the eigen pair. The eigenvector will be used as an input for the next step.

4.3.4 Assignment of Node's Label to the Corresponding Principal Eigenvectors

As shown in Figure 4.1, the third step has three processes: (i) assigning the node label to the corresponding principal eigenvector, (ii) assigning the CHR values to each location node, and (iii) assigning the CRR values to each human node. First, the location and human node labels will be assigned with each element in the principal eigenvectors. The corresponding elements in the hub principal eigenvector and the authority principal eigenvector are labelled with location and human nodes, respectively. Table 4.3 displayed the hub principal eigenvector with the corresponding location node label. Table 4.4 displayed the authority principal eigenvectors with the corresponding human node labels. The human node labels in Table 4.4 are clustered based on Figure 3.9. To assign the CHR and CRR values to each node, a normalisation process is conducted to convert the principal

eigenvector in hub and authority matrices. Equation 4.5 computed the value of CHR, where v_i is the i th eigenvector in v and n is the number of elements in v .

$$CHR = \frac{v_i}{Max(v_i, v_{i+1}, v_{i+2}, \dots, v_n)} \quad \text{Equation 4.5}$$

Table 4.3: Principal Eigenvector and CHR of the Ranked Location Node

Location node label	Hub principal eigenvector	CHR value	Rank
<i>L1</i>	0.975249	1	1
<i>L8</i>	0.094208	0.09659884	2
<i>L3</i>	0.072006	0.07383377	3
<i>L17</i>	0.070077	0.07185522	4
<i>L11</i>	0.066062	0.06773823	5
<i>L19</i>	0.06503	0.06668079	6
<i>L12</i>	0.064535	0.06617241	7
<i>L7</i>	0.055977	0.05739758	8
<i>L4</i>	0.045241	0.04638934	9
<i>L10</i>	0.036022	0.03693591	10
<i>L5</i>	0.035713	0.03661934	11
<i>L9</i>	0.034146	0.03501226	12
<i>L2</i>	0.034077	0.03494206	13
<i>L16</i>	0.03305	0.03388918	14
<i>L6</i>	0.032916	0.03375132	15
<i>L14</i>	0.029811	0.03056712	16
<i>L15</i>	0.029322	0.03006645	17
<i>L20</i>	0.029322	0.03006645	18
<i>L13</i>	0.028214	0.02893009	19
<i>L21</i>	0.028214	0.02893009	20
<i>L18</i>	0.026522	0.02719513	21

Similarly, the computation of CRR used Equation 4.5 by replacing the v with η . The ranking of location and human nodes is based on the CHR and CRR values. High CHR

values indicated high ranking in location nodes and higher CRR values indicated high ranking in human nodes. The next section will discuss the ranking of location and human nodes in the BCC network.

Table 4.4: Principal Eigenvector and CRR of the Human Node

Human node label	Authority principal eigenvector	CRR value	Rank
<i>H1</i>	0.319135187	1	1
<i>H2, H3, H4, H5, H6, H7, H8, H9, H10, H11, H12, H13, H14</i>	0.260510833	0.924232	2
<i>H15, H16, H17, H18, H19, H20, H21</i>	0.045822581	0.806119	3
<i>H26, H27, H28, H29, H30, H31</i>	0.013662691	0.702686	4
<i>H22, H23, H24, H25</i>	0.004419703	0.643522	5

4.3.5 Ranking of Bipartite Nodes in BCC Network

As shown in Table 4.3, location node *L1* has the highest CHR value, which is 1. Thus, the *L1* is ranked at the top among all the location nodes. As the CHR value is a measurement of the COVID-19 virus density in a location, this implies that *L1* is the location with the highest virus density of COVID-19 in Bintulu for the Epi week 5 (the first BCC network). The *L8* is the location node with the second-highest CHR, which is 0.0966, and is ranked as the second-highest virus density in Bintulu during that time.

As shown in Table 4.4, human node *H1* has the highest CRR value, which is 1. As *H1* is the only infectious human in the first BCC network, *H1* is the only possible super-spreader. For the non-infectious human node, *H2* until *H14* hold the same CRR values and are clustered into the same rank. They contact with *H1* in the location node *L1*, and as *L1*

was ranked first in the location node, this cluster of humans have a high risk of being infected compared with other human nodes in this network.

In this final process of stage 2 in BNM-RMF (stage 2.5), this study produced the COVID-19 hotspot index to rank the location node and the risk of infection index to rank the human node. Both indices aim to assist the public health authorities in managing the resources of COVID-19 contact tracing. At the end of stage 2, a set of ranked nodes are generated for location nodes and human nodes. These ranked nodes will be served as input of stage 3 which involved model evaluation (verification and validation) and analysis (parameters significance analysis and predictive power analysis).

4.3.6 Implementation Issues

Implementation issues were encountered in this study, as there was limited data available to quantify the vaccination effect in the network model. The raw data provided by the Bintulu Health Office were collected prior to the launch of the mass vaccination programme, making it impossible to measure the impact of vaccinations on the transmission dynamics. This may affect the accuracy of the results and limit the conclusions that can be drawn about the effectiveness of vaccinations in reducing COVID-19 transmission.

Another implementation issue in this study was the neglect of the degree of compliance with the standard operating procedure (SOP). The hypothesis in this study is that COVID-19 can transmit indirectly via aerosol and contaminated surfaces. Therefore, the effect of SOP compliance was not considered in capturing the ranking of location and human nodes. This may have resulted in an underestimation of the transmission risk associated with certain locations or human nodes, particularly those with poor compliance with SOPs. Future studies could consider including compliance with SOPs in their analysis to improve the accuracy of the results.

The density of the crowd in a location was also neglected in this study. The raw data used were retrieved from a contact tracing investigation form recorded by a health officer. Although all non-infectious humans who visited the same place at the same time as the infectious individual were considered close contacts, they were not included in the raw data. Estimating the size of the location or building is a challenging task, and as a result, the estimation of crowd density was not included in this study. Therefore, the susceptible human node did not reflect the casual contact of the location in this study, which may limit the accuracy of the results. Future studies could consider incorporating crowd density data to improve the representation of the transmission dynamics in crowded locations.

Another implementation issue that may arise from this study is the potential lack of access to MySejathera data. While using MySejathera data would be a useful way to gather information about the visitors of exposed locations and validate the BNM model's ability to identify the risk of infection and hotspot of COVID-19, access to this data may be restricted due to privacy concerns or other reasons. This limitation may reduce the scope and validity of the study's findings, as the model may not capture the other transmission dynamics of COVID-19 in the area under investigation. To address this issue, alternative methods for gathering data on the visitors of exposed locations could be explored. The details of the alternative method are discussed in Section 4.5.2.

4.4 Model Verification

This section depicts the first process in stage 3 of BNM-RMF. Verification is an approach used to determine whether the model has been correctly implemented (Davis, 1992) and to increase the reliability of the model (Robinson, 1997). In other words, the verification procedure determines whether the mode is appropriately constructed (Cook & Skinner, 2005) and fits the requirements and assumptions of the study (Carson, 2002). Thus, the verification process used in this study need to make sure that the ways to implement the search algorithm – HITS algorithm are correct and verified with a benchmark system. There are two verification techniques used in this study: i) benchmark verification and ii) analytical verification.

4.4.1 Benchmark Verification

Benchmark verification aims to verify the research process by comparing the results generated by the BCC network model and from a benchmark system (Liew, 2016). This can be achieved by calculating the RMSE between the ranking values obtained from this study and the values generated by a benchmark system. RMSE is the well-known measure of goodness-of-fit of data. A model is verified as ‘built right’ when the value of RMSE is ≤ 0.05 (Albright & Park, 2009; Browne & Cudeck, 1992). Figure 4.2 illustrated the flow of the benchmark verification of this study, which is adapted from the studies of Eze (2013) and Liew (2016) and is used by Kok (2017).

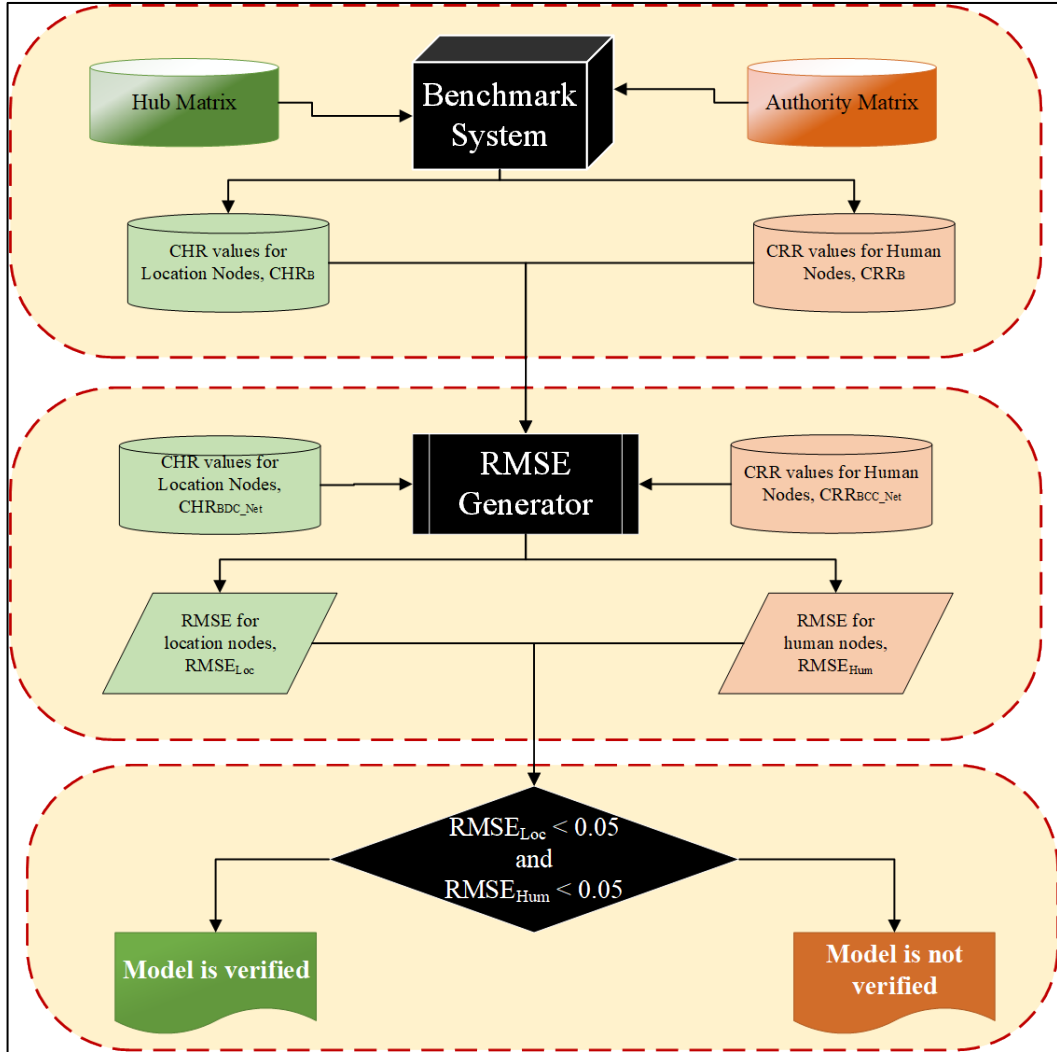


Figure 4.2: Benchmark Verification Process (Liew, 2016)

UCINET has been used to generate a benchmark model to verify the BCC network model in this study. UCINET 6 is a network analysis software that can support many types of network analyses. It can process 2 million nodes, which includes type 2 node networks (Borgatti, Everett, & Freeman, 2002), and is one of the most powerful and reliable network analysis software. In addition, many comprehensive analysis functions are provided (Boyd, Fitzgerald, & Beck, 2006). Moreover, the strong matrix analysis routines of UCINET are suitable for this research because the input of the benchmark model is the hub and authority matrices. The centrality measure analysis determines the node centrality of a network. Based

on the centrality, the relative important node within the network is identified. In addition, Bonacich eigenvector centrality provided in UCINET is one of the centrality measures, which can determine the node that is more central if they are connected to other nodes that are themselves central. Hence, it relates the number of connections to other nodes and measures how well the neighbours are connected.

First, the hub matrix generated in the previous section is converted to a text file format (.txt) and is loaded into UCINET, as shown in Figure 4.2. For UCINET to read the hub matrix, the hub matrix will then be converted into UCINET system files, which are header (in ##H file) and data files (##D file). Second, the header file is employed for centrality analysis in UCINET. UCINET generates the principal eigenvector wherein the elements are determined as the ranking values for the location nodes and summarised in a log file. These ranking values produced by the benchmark system are denoted as benchmark values of CHR , CHR_B . The ranking of the location nodes is based on CHR_B , which is the benchmark ranking of location nodes.

$$Calibrated\ CHR = \frac{CHR_{Each\ Location}}{maximum\ CHR_B} \times 1 \quad \text{Equation 4.6}$$

Figure 4.3 illustrates similar results of benchmark ranking and BCC network model ranking of the location nodes. Nonetheless, the range of the CHR_B , which is 0–0.975, is different from the range of CHR_{BCC_Net} , which is 0–1. Hence, calibration needs to be conducted to make two different rates or units similar. Equation 4.6 is used to calibrate CHR_B values. The calibrated CHR values are generated and termed as CHR_B , as shown in the first step in Figure 4.3.

Benchmark Ranking		BCC Network model Ranking	
Location Node	CHR_B	Location Node	CHR_{BCC_Net}
L1	0.975	L1	1
L8	0.094	L8	0.096599
L3	0.072	L3	0.073834
L17	0.07	L17	0.071855
L11	0.066	L11	0.067738
L12	0.065	L12	0.066172
L19	0.065	L19	0.066681
L7	0.056	L7	0.057398
L4	0.045	L4	0.046389
L5	0.036	L5	0.036619
L10	0.036	L10	0.036936
L2	0.034	L2	0.034942
L9	0.034	L9	0.035012
L6	0.033	L6	0.033751
L16	0.033	L16	0.033889
L14	0.03	L14	0.030567
L15	0.029	L15	0.030066
L20	0.029	L20	0.030066
L13	0.028	L13	0.02893
L21	0.028	L21	0.02893
L18	0.027	L18	0.027195

Figure 4.3: Benchmark Ranking and BCC Network Model Ranking

The second step is to perform RMSE analysis. The inputs of the RMSE generator are CHR_B and CHR_{BCC_Net} , as shown in Figure 4.7. Equation 4.7 calculated the RMSE for location nodes, where N_{Loc} is the number of location nodes (21 in this study) and i is a regular number starting from 1, as shown in Table 4.5 The sum of squares difference of CHR_B and CHR_{BCC_Net} obtained from the last row in Table 4.5 is substituted into Equation 4.7. As a result, RMSE for location node 0.0002244 is shown in Equation 4.8, corrected to four decimal places, which is much lesser than the threshold value of 0.05.

$$RMSE_{Loc}(CHR_B, CHR_{BCC_{Net}}) = \sqrt{\frac{1}{N_{Loc}} \left[\sum_{i=1}^{N_{Loc}} (CHR_{B_i} - CHR_{BCC_{Net}_i})^2 \right]} \quad \text{Equation 4.7}$$

Table 4.5: RMSE Analysis of Location Nodes

Location Node	CHR _B	CHR _{BCC_Net}	CHR _B - CHR _{BCC_Net}	(CHR _B - CHR _{BCC_Net}) ²
<i>L1</i>	1	1	0	0
<i>L2</i>	0.034872	0.034942	-7E-05	4.93719E-09
<i>L3</i>	0.073846	0.073834	1.24E-05	1.5336E-10
<i>L4</i>	0.046154	0.046389	-0.00024	5.54574E-08
<i>L5</i>	0.036923	0.036619	0.000304	9.22561E-08
<i>L6</i>	0.033846	0.033751	9.48E-05	8.99346E-09
<i>L7</i>	0.057436	0.057398	3.83E-05	1.46823E-09
<i>L8</i>	0.09641	0.096599	-0.00019	3.55638E-08
<i>L9</i>	0.034872	0.035012	-0.00014	1.97305E-08
<i>L10</i>	0.036923	0.036936	-1.3E-05	1.64688E-10
<i>L11</i>	0.067692	0.067738	-4.6E-05	2.10886E-09
<i>L12</i>	0.066667	0.066172	0.000494	2.4429E-07
<i>L13</i>	0.028718	0.02893	-0.00021	4.50039E-08
<i>L14</i>	0.030769	0.030567	0.000202	4.08488E-08
<i>L15</i>	0.029744	0.030066	-0.00032	1.04239E-07
<i>L16</i>	0.033846	0.033889	-4.3E-05	1.85125E-09
<i>L17</i>	0.071795	0.071855	-6E-05	3.64191E-09
<i>L18</i>	0.027692	0.027195	0.000497	2.47186E-07
<i>L19</i>	0.066667	0.066681	-1.4E-05	1.99469E-10
<i>L20</i>	0.029744	0.030066	-0.00032	1.04239E-07
<i>L21</i>	0.028718	0.02893	-0.00021	4.50039E-08
Sum of (CHR _B - CHR _{BCC_Net}) ² :				1.05734E-06

$$RMSE_{Loc}(CHR_B, CHR_{BCC_{Net}}) = \sqrt{\frac{1}{21} [1.05734E - 06]} = 0.0002244 \quad \text{Equation 4.8}$$

The RMSE for the human node must be calculated. The authority matrix is the input of the benchmark system, as shown in the first step in Figure 4.2. The ranking value of each human node computed by the UCINET is denoted as CRR_B , whereas the principal eigenvector of the human node that returned from the BCC network model is denoted as

CRR_{BCC_Net} . Equation 4.9 calculates the RMSE for humans, that is, $RMSE_{Hum}$, where N_{Hum} is the number of human nodes (31 in this network) and j is a regular number starting from 1, as shown in Table 4.6.

$$RMSE_{Hum}(CRR_B, CRR_{BCC_Net}) = \sqrt{\frac{1}{N_{Hum}} \left[\sum_{j=1}^{N_{Hum}} (CRR_{Bj} - CRR_{BCC_Netj})^2 \right]} \quad \text{Equation 4.9}$$

Table 4.6: RMSE Analysis of Human Nodes

Human Node	CRR_B	CRR_{BCC_Net}	$CRR_B - CRR_{BCC_Net}$	$(CRR_B - CRR_{BCC_Net})^2$
$H1$	1	1	0	0
$H2$	0.818181818	0.816302	0.001879	3.53206E-06
$H3$	0.818181818	0.816302	0.001879	3.53206E-06
$H4$	0.818181818	0.816302	0.001879	3.53206E-06
$H5$	0.818181818	0.816302	0.001879	3.53206E-06
$H6$	0.818181818	0.816302	0.001879	3.53206E-06
$H7$	0.818181818	0.816302	0.001879	3.53206E-06
$H8$	0.818181818	0.816302	0.001879	3.53206E-06
$H9$	0.818181818	0.816302	0.001879	3.53206E-06
$H10$	0.818181818	0.816302	0.001879	3.53206E-06
$H11$	0.818181818	0.816302	0.001879	3.53206E-06
$H12$	0.818181818	0.816302	0.001879	3.53206E-06
$H13$	0.818181818	0.816302	0.001879	3.53206E-06
$H14$	0.818181818	0.816302	0.001879	3.53206E-06
$H15$	0.144200627	0.143584	0.000617	3.8071E-07
$H16$	0.144200627	0.143584	0.000617	3.8071E-07
$H17$	0.144200627	0.143584	0.000617	3.8071E-07
$H18$	0.144200627	0.143584	0.000617	3.8071E-07
$H19$	0.144200627	0.143584	0.000617	3.8071E-07
$H20$	0.144200627	0.143584	0.000617	3.8071E-07
$H21$	0.144200627	0.143584	0.000617	3.8071E-07
$H22$	0.012539185	0.013849	-0.00131	1.71562E-06
$H23$	0.012539185	0.013849	-0.00131	1.71562E-06
$H24$	0.012539185	0.013849	-0.00131	1.71562E-06
$H25$	0.012539185	0.013849	-0.00131	1.71562E-06
$H26$	0.043887147	0.042812	0.001076	1.15678E-06
$H27$	0.043887147	0.042812	0.001076	1.15678E-06
$H28$	0.043887147	0.042812	0.001076	1.15678E-06
$H29$	0.043887147	0.042812	0.001076	1.15678E-06

Table 4.6 Continued

<i>H30</i>	0.043887147	0.042812	0.001076	1.15678E-06
<i>H31</i>	0.043887147	0.042812	0.001076	1.15678E-06
Sum of $(CRR_B - CRR_{BCC_Net})^2$:				6.23849E-05

Based on the sum of square differences, as shown in Table 4.6, $RMSE_{Hum}$ is calculated as in Equation 4.10. The RMSE of CRR_B and CRR_{BCC_Net} is 0.001419, corrected to four decimal places, which is lesser than the threshold value of 0.05.

$$RMSE_{Hum}(CRR_B, CRR_{BCC_Net}) = \sqrt{\frac{1}{31}[6.23849E - 05]} = 0.001419 \quad \text{Equation 4.10}$$

As shown in Figure 4.2, the last step is to input the value of RMSE of location and human nodes to the comparator. The RMSE of the location (0.0002244) and human (0.001419) nodes are much lesser than the threshold RMSE value of 0.05 in this study. Thus, based on the RMSE values of both nodes, it can be concluded that the BCC network model in this study is a verified model.

4.4.2 Analytical Verification

In this study, SRCC is used for analytical verification to measure the closeness of the ranking values for each location and human node and the degree of relationship between a pair of rankings. As the data size used in this study is relatively small, SRCC or ρ is used to verify the network model analytically (Bouzerdoun et al., 2004; Lim et al., 2007; Mukaka, 2012). Table 4.7 displays the indicators of SRCC value. The SRCC values range from -1 to $+1$, where $+1$ indicates a very strong positive association of ranks, 0 indicates no associations between ranks, and -1 indicates a very strong negative association of ranks. Thus, SRCC with a value of 1 indicates perfect similarity, that is, the same ranking, whereas value 0 indicates the absence of similarity, that is, a different ranking.

Table 4.7: SRCC Indicators

SRCC coefficient (Positive and Negative)	Indicator
$\pm (0.00-0.19)$	Very weak
$\pm (0.20-0.39)$	Weak
$\pm (0.40-0.59)$	Moderate
$\pm (0.60-0.79)$	Strong
$\pm (0.80-1.00)$	Very strong

According to Liew (2016), the value of SRCC between the CHR and hub matrix of each location node should not be <0.7 to verify the model analytically. Consequently, the hypothesis in this study is a positive and high correlation with an SRCC coefficient of not <0.7 . Hence, SRCC for location nodes is denoted as $\rho_{CHR_HubMatrix}$ and calculated in Equation 5.6, where the N_{Loc} is 21 and I is a regular number starting from 1. Table A.3 displays the calculation of $d_{CHR_HubMatrix}$ in Equation 4.11 (attached in Appendix A). By using $d_{CHR_HubMatrix}$ obtained in Table A.3, the calculation of $\rho_{CHR_HubMatrix}$ is shown in Equation 4.12.

$$\rho_{CHR_HubMatrix} = 1 - \frac{6 \sum_i^{N_{Loc}} [d_{CHR_HubMatrix}]_i^2}{N_{Loc}(N_{Loc}^2 - 1)}$$

$$where \{d_{CHR_HubMatrix}\}_i = \{RankCHR\}_i - \{RankHubMatrix_{sum}\}_i \quad \text{Equation 4.11}$$

$$\rho_{CHR_HubMatrix} = 1 - \frac{6 \times 40}{21(21^2 - 1)} = 0.974 \quad \text{Equation 4.12}$$

The value of $\rho_{CHR_HubMatrix} = 0.974$ shows that the CHR of each location node is positively and strongly correlated with the sum of a hub matrix of the edges linked to it. The ρ value is much higher than the value of 0.70. Thus, this has further strengthened the conclusion made in benchmark verification about the processes involved in formulating the

BCC network model and implementing the network. Therefore, the BCC network model is verified, and the model validation will be discussed in the next section.

4.5 Model Validation

Model validation aims to answer the question of ‘are you building the right thing’ and to ensure the model can represent the real problem. The validation in this study is a process to ensure the accuracy of the outputs obtained from the model. This section discusses process 3.2 of BNM-RMF, which is to validate the first BCC network model to achieve the objectives. To validate the model, the conventional approach is to compare the simulation results with real data. This is to ensure that the model is capable to capture reality with sufficient exactness. In Section 4.5.1, the model validation is conducted via analysis using the other sets of models, as shown in Table 3.3. In Section 4.5.2, the locations collected from the public tracing data for Bintulu are included in each set of models based on the epidemiology weeks, as shown in Table 3.3. The result of the targeted model should agree with the result obtained from the validated model.

In the validation analysis, SRCC (ρ) is computed to compare the location rank generated from the targeted model and the validated model which utilised the real data. As mentioned in Section 4.4.2, the threshold value is 0.70, which means that there is a high correlation between the targeted model ranking and the ranking generated by the validated model for the SRCC reported. This implies that if the correlation coefficient is >0.70 , the BCC network is validated.

4.5.1 Validation with Raw Contact Tracing Data

This analysis appears to be focused on evaluating BNM for infectious diseases, with a specific mention of COVID-19 which been defined. The research uses data collected

through contact tracing efforts to validate models that aim to identify the spread of the COVID-19 within networks of infected individuals and the locations they visit. The models are validated by comparing rankings of locations and individuals within the models with rankings generated by actual contact tracing data. If the correlation between the rankings in the models and the actual data is higher than a specified threshold, the models are validated. The analysis could be useful in improving the efficiency and accuracy of contact tracing efforts, and in better understanding the spread of infectious diseases like COVID-19.

In Chapter 3, the first BCC network model consists of 1 index patient and 30 individuals who have met the patient, as well as 21 different locations visited by the index patient and the contacts. However, as mentioned in the same chapter, the seventh BCC network model was expanded from the first BCC network model to include the patient's family members and close contacts. As a result, the first BCC network model served as the initial targeted model, while the seventh BCC network model was the first model to undergo validation in the analysis. The validation analysis utilized the combination of the targeted and validated models, based on whether the networks shared any identical location or human nodes. Networks that did not share any identical nodes, such as BCC networks 4 and 5, were excluded from the analysis. Table 4.8 displays the groups of BCC network models that were used for validation analysis, along with their SRCC values.

Table 4.8: BCC Network Model Validation

Group	BCC Network	Model	SRCC values (location)	SRCC values (human)
1	1	Targeted model	0.74	1.00
	7	Validated model		
2	2	Targeted model	1.00	1.00
	3	Validated model		
3	6	Targeted model	1.00	No Applicable
	7	Validated model		

As shown in Table 4.8, the first group consists of two models: the targeted and validated models. The first targeted model used the 6th dataset listed in Table 3.3, consisting of 1 index patient and 30 contacts (without family members). Section 3.3 presents the process of formulating and quantifying the first targeted model, which is the first BCC network model. The ranking process of location and human nodes has been verified in Sections 4.4.1 and 4.4.2. Hence, for the first validation group, if the closeness of the location and human rankings is >0.70 (of SRCC values) between the comparison of the seventh BCC network (including family members), then the model is validated for the first BCC network and the ranking is still highly correlated. Thus, the seventh BCC network is a validated model to test whether the location and human rankings in the targeted model correlate to the location and human in the validated model. The seventh BCC network is extended from the first BCC network by including new 11 human nodes labelled as *H150*, *H151*, *H152*, *H153*, *H154*, *H155*, *H156*, *H157*, *H158*, *H159*, and *H160* and a new location node labelled as *L52*.

Next, the location node parameters, namely, *Ka*, *Ks*, and *Fl*, need to be quantified using the method described in Sections 3.3.3. All attributes required by the parameters are retrieved from the dataset. The human node parameters, namely, *P_{wr}*, *Fl*, and *Du*, are quantified using the method introduced in Sections 3.3.4. Table A.4 illustrated all the values of the parameters for location and human nodes for the first validated model. The CCS is computed using normalised parameter values. The CCS is calculated as in Equation 3.12

introduced in Section 3.3.5. Table A.5 presents the CCS for all links in the first validated model.

The CCS matrix in Table A.5 is the input of the HITS search algorithm. Figure 4.1 depicts the implementation of HITS. Table 4.9 presents the ranking of location nodes in the validated model with the CHR. As mentioned earlier, one new location node is added to the network, which means only 21 location nodes are identical to the first targeted model. Hence, the ranking of these 21 location nodes will be compared between the targeted and validated models in the first group validation analysis. The correlation coefficient of the ranking of the validated and targeted models is denoted as $\rho_{TestLoc}$ and the equation used for the calculation of $\rho_{TestLoc}$ is Equation 4.13 where N_{Loc} is 21 and a is a regular number starting from 1. Table 4.10 presents the calculation of various terms in Equation 4.13 and Expression 4.14 presents the calculation of $\rho_{TestLoc}$ based on Equation 4.14. The result of this validation $\rho_{TestLoc}$ is 0.74, which is higher than the threshold value of 0.70; thus, the model is validated.

$$\rho_{TestLoc} = 1 - \frac{6 \sum_a^{N_{Loc}} [\{d_{TargetValidate}\}_a]^2}{N_{Loc}(N_{Loc}^2 - 1)}$$

$$\text{where } \{d_{TargetValidate}\}_a = \{RankCHR_{Target}\}_a - \{RankCHR_{Validate}\}_a \quad \text{Equation 4. 13}$$

$$\rho_{TestLoc} = 1 - \frac{6 \times 399}{21(21^2 - 1)} = 0.74 \quad \text{Equation 4. 14}$$

Table 4.9: The Ranking of Location Nodes (First Validated Model)

Rank	Location Node	CHR _{Validate}
1	<i>L52</i>	1
2	<i>L1</i>	0.215865
3	<i>L8</i>	0.084612
4	<i>L3</i>	0.070784
5	<i>L17</i>	0.067683

Table 4.9 continued

6	<i>L11</i>	0.066373
7	<i>L19</i>	0.06473
8	<i>L7</i>	0.055997
9	<i>L4</i>	0.047775
10	<i>L16</i>	0.04166
11	<i>L10</i>	0.041135
12	<i>L5</i>	0.040738
13	<i>L6</i>	0.040622
14	<i>L9</i>	0.038721
15	<i>L2</i>	0.038633
16	<i>L14</i>	0.036625
17	<i>L15</i>	0.035944
18	<i>L21</i>	0.035865
19	<i>L20</i>	0.034968
20	<i>L12</i>	0.03465
21	<i>L13</i>	0.034571
22	<i>L18</i>	0.032393

Table 4.10: SRCC Calculation of Location Nodes between the Group 1

Location Node	Rank CHR_{Target}	Rank $CHR_{Validate}$	$d_{TargetValidate}$	$d_{TargetValidate}^2$
<i>L1</i>	1	2	-1	1
<i>L8</i>	2	3	-1	1
<i>L3</i>	3	4	-1	1
<i>L17</i>	4	5	-1	1
<i>L9</i>	5	14	-9	81
<i>L19</i>	6	7	-1	1
<i>L12</i>	7	20	-13	169
<i>L7</i>	8	8	0	0
<i>L4</i>	9	9	0	0
<i>L16</i>	10	10	0	0
<i>L10</i>	11	11	0	0
<i>L5</i>	12	12	0	0
<i>L2</i>	13	15	-2	4
<i>L6</i>	14	13	1	1
<i>L15</i>	15	17	-2	4
<i>L20</i>	15	19	-4	16
<i>L11</i>	16	6	10	100
<i>L14</i>	17	16	1	1
<i>L13</i>	18	21	-3	9

Table 4.10 Continued

<i>L21</i>	18	18	0	0
<i>L18</i>	19	22	-3	9
Sum of $d_{TargetValidate}^2$:				399

Table 4.11 displays the second group of BCC networks. For the targeted and validated model in the second group, both models are built by using datasets 1 and 2 mentioned in Table 3.3, respectively. The second targeted model comprises of nine new human nodes and five new location nodes. Hence, these 9 new human nodes are denoted as *H32*, *H33*, *H34*, *H35*, *H36*, *H37*, *H38*, *H39*, and *H40*, whereas the new location nodes are labelled as *L35*, *L37*, *L38*, *L39*, and *L40*. For the second validated model, all the humans in dataset 2 are identical to the human nodes in the targeted model and only one new location node is identified from second dataset, which been denoted as *L36*.

Similarly, the CCS values for the targeted and validated models in groups 2 and 3 are computed and listed in Tables A.6 and A.7. Tables A.8 present the ranking of human and location nodes for the targeted and validated models for group 2. As shown in Table A.6, there are nine identical human nodes and two location nodes in the targeted and validated models, which are used to compute the SRCC value as shown in Tables 4.11 and 4.12, respectively. The SRCC value ($\rho_{TestLoc}$) is calculated as in Equation 4.13 where N_{Loc} is 2 and a is a regular number starting from 1. For the SRCC value for a human node ($\rho_{TestHum}$), the formula is similar to Equation 4.13 but N_{Loc} is replaced with N_{Hum} and RankCHR with RankCRR for both the targeted and validated models. Thus, N_{Hum} is 9 and a is a regular number starting from 1. Expressions 4.15 and 4.16 calculate the $\rho_{TestLoc}$ and $\rho_{TestHum}$, respectively.

Table 4.11: SRCC of Location Nodes between the Models in Group 2

Location node	Rank CHR _{Target}	Rank CHR _{Validate}	$d_{TargetValidate}$	$d_{TargetValidate}^2$
<i>L34</i>	2	2	0	0
<i>L35</i>	1	1	0	0
Sum of $d_{TargetValidate}^2$:				0

$$\rho_{TestLoc} = 1 - \frac{6 \times 0}{2(2^2 - 1)} = 1 \quad \text{Equation 4.15}$$

Table 4.12: SRCC of Human Nodes between the Models in Group 2

Location node	Rank CHR _{Target}	Rank CHR _{Validate}	$d_{TargetValidate}$	$d_{TargetValidate}^2$
<i>H32</i>	1	1	0	0
<i>H33</i>	1	1	0	0
<i>H34</i>	1	1	0	0
<i>H35</i>	1	1	0	0
<i>H36</i>	1	1	0	0
<i>H37</i>	1	1	0	0
<i>H38</i>	1	1	0	0
<i>H39</i>	1	1	0	0
<i>H40</i>	1	1	0	0
Sum of $d_{TargetValidate}^2$:				0

$$\rho_{TestHum} = 1 - \frac{6 \times 0}{9(9^2 - 1)} = 1 \quad \text{Equation 4.16}$$

Table 4.13: SRCC of Location Nodes between the Models in Group 3

Location Node	Rank CHR _{Target}	Rank CHR _{Validate}	$d_{TargetValidate}$	$d_{TargetValidate}^2$
<i>L51</i>	1	1	0	0
<i>L1</i>	2	2	0	0
Sum of $d_{TargetValidate}^2$:				0

$$\rho_{TestLoc} = 1 - \frac{6 \times 0}{2(2^2 - 1)} = 1 \quad \text{Equation 4.17}$$

The values of $\rho_{TestLoc}$ and $\rho_{TestHum}$, are much higher than the threshold value of 0.70. Therefore, the BCC network models in groups 2 and 3 are validated. In a nutshell, based on the SRCC value of the three groups of models, it can be concluded that the BCC network model is validated by using the contact tracing investigation forms.

4.5.2 Validation with Public Tracing Data

As mentioned in Section 3.4.1, the network is formed for each set of contact tracing forms provided by the Bintulu Health Office. Hence, the location nodes in the network are visited by the index patient and the contacts. However, the small number of datasets was insufficient to capture the comprehensive full picture of the transmission source in the real world and lack of access to MySejathera data. As alternative method, the exposed locations identified by public health authorities are retrieved from the daily press statement released by the Sarawak Disaster Management Committee on their official website (<https://sdmc.sarawak.gov.my/>). The press statements were collected from 18 January 2021 to 27 February 2021, to match with the tracing period for every contact tracing data provided by Bintulu Health Office, as shown in Table 4.14.

Table 4.14: BCC Network Model with Tracing period

Dataset	BCC Network Model	Epi Week of Contact Tracing
1	1	1-3
2	2	1-3
3	3	2-4
4	4	2-4
5	5	5
6	6, 7	5-7

The press statements provide information on the date and time of visit and exposed locations, which will be used for validation analysis in this section. However, the press statements do not include information about other possible contacts at each exposed location. As discussed in Chapter 3, only locations visited by infectious individuals are significant in contact tracing. Thus, the exposed locations are assumed to have been visited by an infectious individual (represented by a human node with $Hs = 1$) and are aligned with the location nodes in this study. The data pre-processing of these exposed locations follows the process outlined in Chapter 3. The date of visit is used to retrieve environmental data for computing the Ka and Ks for each exposed location. Since these locations must have been visited by an infectious individual at least once, the Fl value for each exposed location is assigned as 1, with cumulative Fl values computed if exposed locations have the same name or GPS coordinates. All non-repeated exposed locations are denoted as location nodes and labelled with the prefix 'P' to distinguish them from location nodes (L) identified from the raw contact tracing forms.

These location nodes are linked to a set of dummy human nodes labelled with the prefix 'HD', with Hs values assigned as 1 to indicate that they are infectious. Consequently, the P_{wr} values for each dummy human node are equal to 0, as per Equation 3.15. The time of visit is used to quantify the human node parameter, D . For a location declared as a location node in the network, it must have been visited by at least one dummy human node, so the Fh values for these locations are the same as their Fl values. These location nodes and dummy human nodes are included in each network in Table 4.15 based on epidemiology weeks. To reduce bias, human nodes with $Hs = 0$ are removed, resulting in networks formulated based only on infectious human nodes. In this validation, the combination of the targeted and validated models is based only on the tracing week in terms of EW, as listed in Table 4.14.

Table 4.15 displays the details of the new targeted and validated models, along with the SRCC values for the ranking of location nodes.

Table 4.15: BCC Network Model Validation with Public Data

Group	Data	Model	Number of location nodes	SRCC values (location)
1	BCC 1	Targeted	21	0.76
	D6 with data in Epi week 5 to 7	Validated	128 (107 added)	
2	BCC 2	Targeted	6	0.90
	D1 with data in Epi week 1 to 3	Validated	117 (111 added)	
3	BCC 3	Targeted	3	0.50
	D2 with data in Epi week 1 to 3	Validated	115 (112 added)	
4	BCC 4	Targeted	11	1.00
	D3 with data in Epi week 2 to 4	Validated	148 (137 added)	
5	BCC 5	Targeted	12	0.993
	D3 with data in Epi week 2 to 4	Validated	149 (137 added)	
6	BCC 6	Targeted	2	1.00
	D5 with data in Epi week 2 to 4	Validated	39 (37 added)	

As only the ranking of location nodes is compared within every targeted and validated model, as shown in Table 4.15, Equation 4.11 is used to compute the SRCC values, where N_{Loc} is the number of location nodes of the targeted model. Based on the result presented in Table 4.15, only one group of models is below the threshold value of 0.70, which is group 3 ($\rho_{TestLoc} = 0.50$). This outcome is expected because only three location nodes are identified from the 2nd BCC network model, making it difficult to capture the entire picture of mobilities for 3 weeks; hence $\rho_{TestLoc}$ is lower than the threshold value.

The other five groups of models obtain a value >0.70 , which indicated that using the locations listed in the single contact tracing form can formulate a network to identify the hotspot in the period. However, $\rho_{TestLoc}$ of group 6 is >0.70 but only consists of three location

nodes in the sixth BCC network model. This result is due to the limited time of dataset tracing used by the sixth BCC network model (D5), which is only one week; hence, the new location nodes added to the network are less than another group. Thus, a lower number of location nodes is still sufficient to identify hotspots for a single EW.

Based on the correlation coefficient, the majority showed that after removing the non-infectious human node from the network, the ranking was still highly positively correlated with the original network. Hence, a further analysis has been conducted to identify the significant parameters of location and human nodes in the following section. In a summary, by combining the results produced from the verification and validation analyses of the research results, the BCC network model is now verified and validated by using the contact tracing investigation forms.

4.6 Model Analysis and Evaluation

This section discussed the last activity of stage 3 in BNM-RMF which is model analysis and evaluation. The focus of the network evaluation analysis is to perform extended evaluations and analyses of the model formulated. The evaluation results obtained will further strengthen the research findings and provide auxiliary illustrations for the research findings. Two primary analyses are presented in this section which are parameter significance analysis and predictive power analysis.

4.6.1 Parameter Significance Analysis

This section presents the parameter significance analysis to identify the relative significance of each parameter used in the BCC network model. One of the objectives of this study is to identify the critical parameters, which have a greater impact on the ranking of the location and human nodes. A total of six parameters are used: Ka , Ks , and Fl for the location

node and P_{wr} , Fh , and D for the human node. In addition, two types of analysis are conducted: (i) ‘leave-one-out’ analysis for each of the parameters and (ii) ‘leave-multiple-out’ for more than one parameter. In the ‘leave-one-out’ analysis, the parameter will be excluded one at a time from the BCC network model in obtaining CHR and CRR and conducting the subsequent process of the research methodology listed in the BNM-RMF. Table 4.16 displays the results. For ‘leave-multiple-out’ analysis, at least two parameters will be removed from the BCC network model and the results are tabulated in Table 4.17. The ranking generated from the modified model will be compared with the ranking generated from the original model by calculating the SRCC. The SRCC value (ρ) will be elucidated based on the general rule of thumb (Hinkle et al., 2003; Mukaka, 2012). These outputs are used to suggest further analyses to identify the relationship between the parameter(s) that influences the research result and the performance of the BCC network model.

Equations 4.18 and 4.19 compute the SRCC related to the ranking of the location and human nodes, respectively, where $RankCHR_{BCC}$ and $RankCRR_{BCC}$ represent the original ranking of location and human nodes and $RankCHR_{New}$ and $RankCRR_{New}$ represent the new ranking of the location and human nodes computed from the modified model. $N = 21$ for Equation 4.18, $N = 31$ for Equation 4.19, and a is a regular number that starts from 1 for both equations.

$$\rho_{ParaLoc} = 1 - \frac{6 \sum_a^N [d_{ParaLoc}]_a^2}{N(N^2 - 1)}$$

$$\text{where } \{d_{ParaLoc}\}_a = \{RankCHR_{BCC}\}_a - \{RankCHR_{New}\}_a \quad \text{Equation 4.18}$$

$$\rho_{ParaHum} = 1 - \frac{6 \sum_a^N [d_{ParaHum}]_a^2}{N(N^2 - 1)}$$

$$\text{where } \{d_{ParaHum}\}_a = \{RankCRR_{BCC}\}_a - \{RankCRR_{New}\}_a \quad \text{Equation 4.19}$$

Table 4.16: SRCC for “Leave-one-out” Parameter Analysis

Analysis	Parameter excluded	ρ_{ParaHum}	ρ_{ParaLoc}
1	Fh	1	0.9636
2	D	1	0.987
3	P_{wr}	1	0.9812
4	Ka	1	0.9747
5	Ks	1	0.9812
6	Fl	1	0.961

Table 4.17: SRCC for “Leave-multiple-out” Parameter Analysis

Analysis	Parameter excluded	ρ_{ParaHum}	ρ_{ParaLoc}
1	Fh, D	1	0.939
2	Ka, Ks	0.9938	0.7929
3	Ka, Fl	1	0.9519
4	Ks, Fl	1	0.9779
5	Fh, P_{wr}	1	0.9682
6	P_{wr}, D	1	0.9682
7	P_{wr}, Fl	1	0.9656
8	Fh, Fl	1	0.863
9	D, Fl	1	0.9364
10	Ka, Ks, Fl	1	0.7227
11	WR, Ka, Ks	1	0.7487
12	Fh, P_{wr}, Fl	1	0.8786
13	D, P_{wr}, Fl	1	0.937
14	Fh, D, Fl	1	0.8331
15	Fh, Ka, Ks, Fl	0.9938	0.7338
16	D, Ka, Ks, Fl	0.9938	0.8325
17	P_{wr}, Ka, Ks, Fl	1	0.7571
18	D, P_{wr}, Ka, Ks	1	0.7208
19	Fh, P_{wr}, Ka, Ks	1	0.7487
20	Fh, D, Ka, Ks	0.9938	0.8266
21	D, P_{wr}, Ka, Fl	1	0.9519
22	Fh, P_{wr}, Ka, Fl	1	0.8851
23	Fh, D, Ka, Fl	1	0.8214
24	Fh, D, Ks, Fl	1	0.9779
25	D, P_{wr}, Ks, Fl	1	0.9305
26	Fh, P_{wr}, Ks, Fl	1	0.826
27	D, P_{wr}, Ka, Ks, Fl	1	0.7227
28	Fh, P_{wr}, Ka, Ks, Fl	1	0.7929
29	Fh, D, Ka, Ks, Fl	0.9938	0.4955

Table 4.16 presents the results of the ‘leave-one-out’ analysis. All the coefficients reported a value >0.95 , which signifies a positive and very high correlation between the new and original ranking and is almost the same as the original ranking for human and location nodes. As only one parameter is removed from the model, the output will remain almost the same. Thus, a ‘leave-multiple-out’ analysis is needed to further identify the significant parameters in this study. Table 4.17 displays the results of this analysis.

Table 4.17 shows that by excluding five parameters, retaining only P_{wr} in the model (analysis 29), the correlation coefficient reported a value <0.50 which is less than the threshold value of 0.70. Analyses 26 and 27 show that if only Fh or D remains in the model, whereas the other five parameters are removed from the model, the correlation coefficient still entails a strong and positive correlation with the original location node ranking. It suggested that Fh and D might be the significant parameters; thus, they are maintained in the model, whereas the other four are excluded in analysis 17. Similarly, the correlation coefficient still indicates a strong and positive correlation with the original location node ranking. Hence, the combination of Fh and D is significant in the BCC network. Analyses 10 and 11 show that by maintaining Fh and D in the network and only three parameters are removed from the network, the correlation coefficient reported a value lower than in analysis 17. Consequently, Ka and Ks might be another combination of the parameter that is crucial in the ranking of the location node. To further examine the significance of Ka and Ks , both of them are removed from the network in analyses 2, 7, 15–16, 18–20, and 27–28. In these eight analyses, only two correlation coefficients were reported above 0.8, which are analyses 16 and 20, and the other five analyses reported below 0.8.

When the parameters of similar nature (analyses 11 and 14) are excluded from the model, Table 4.17 shows that exclusion of location-related parameters (P_{wr} , Ka , and Ks)

reports a lower correlation with the original location node ranking than exclusion of frequency-related parameters (Fh , Fl , and D). This suggests that the location-related parameters are more significant than frequency-related parameters in the ranking of location nodes. At this stage, the conjectures proposed in Chapter 3 has been validated based on this analysis, hence location node and human node are shown the significance in the network model to determine the hotspot of COVID-19.

As shown in Table 4.17, the ranking of the human node in the analysis remains the same as the original ranking of the human node except for analyses 2, 15, 16, and 29 because the human node that exists in the network for this study can be grouped based on the contact type. Hence, the ranking of the human node will be the same if the human node is in the same category. Only the ranking of the index patient in the network (human node with $Hs = 1$) will change if Ka and Ks are removed but P_{wr} is retained in the model because the link weight value of the index patient is related to Ka , Ks , and Fl and P_{wr} is not computed for the index patient. Therefore, the link weight of the index patient will be decreased if only P_{wr} is maintained in the model and both Ka and Ks are excluded at the same time. In addition, the parameters in the human node cannot be empty, because it can indicate that a human has not visited the location, discarding the analysis of exclusion for Fh , D , and P_{wr} .

In a nutshell, a statistical study of the different parameters involved has been conducted to identify the interrelations between the model parameters. All six parameters are relatively important to the performance of the BCC network model and need to be taken into account in order to generate an optimal performance of the BCC network model.

4.6.2 Predictive Power Analysis

The aim of predictive power analysis is to determine the potential predictive ability of BNM. According to Shmueli (2010), the purpose of a predictive model is to predict new

observations or future trends for given values. The new observations or future trends included a predictive distribution, an interval or points, and the new rankings based on new observations. Hence, the ability of BNM to predict a future value is exceptionally crucial to model a complex real-world problem with abstractive information. As this study involves the spatial component, namely, location in the model, the predictive power analysis will determine the ability of BNM from temporal and spatial perspectives.

As shown in Tables 4.8 and 4.15, the targeted model has the potential to predict other exposed locations in the week(s) of tracing the index patient in the model. The SRCC values indicated that the ranking of the targeted model is strongly positively correlated with the ranking of the validated model. Therefore, this section extends the analysis to further examine the potential of the predictive power of BNM in hotspot prediction. According to Benavides et al. (2012), ‘When disease transmission does not occur at the hotspot, it can still occur at a certain distance from the hotspot’ (p. 1). To investigate the predictive power of BNM from a spatial perspective, all the location nodes in each targeted and validated models need to be clustered based on the given radius.

Figure 4.4 illustrates the implementation design of predictive power analysis. In this analysis, five radii are used as input to the clustered location node generator to cluster all the location nodes into five datasets: 200, 400, 600, 800, and 1000 m. All the clustered location nodes are labelled with an ID with the prefix ‘C’. All the link weights of the location nodes in the targeted model need to be re-calculated.

The BCC network dataset is used as the input of the BCC network model process, as shown in Figure 4.4, to formulate the targeted model. Afterwards, summation is applied to combine the parameter values in the location node (Ka , Ks , and Fl) that belong to the same clustered location (ID). For the human node, the summation is applied to parameters Fh and

D , and the mean value of Pwr is obtained for all the human nodes that visited the same clustered location node. After calculating the link weight for the new cluster network, the CCS matrix is formed, which will be the input to the HITS search algorithm.

Table 4.18 presents the new ranking of the cluster network (radius = 200 m). The public exposed location dataset is used to formulate a validated model using the same process as in Figure 4.4. The formulation of the bipartite graph for public exposed location is identical to the process mentioned in Section 4.5.2. Table 4.19 displays the ranking of the top 10 location nodes in the validated model.

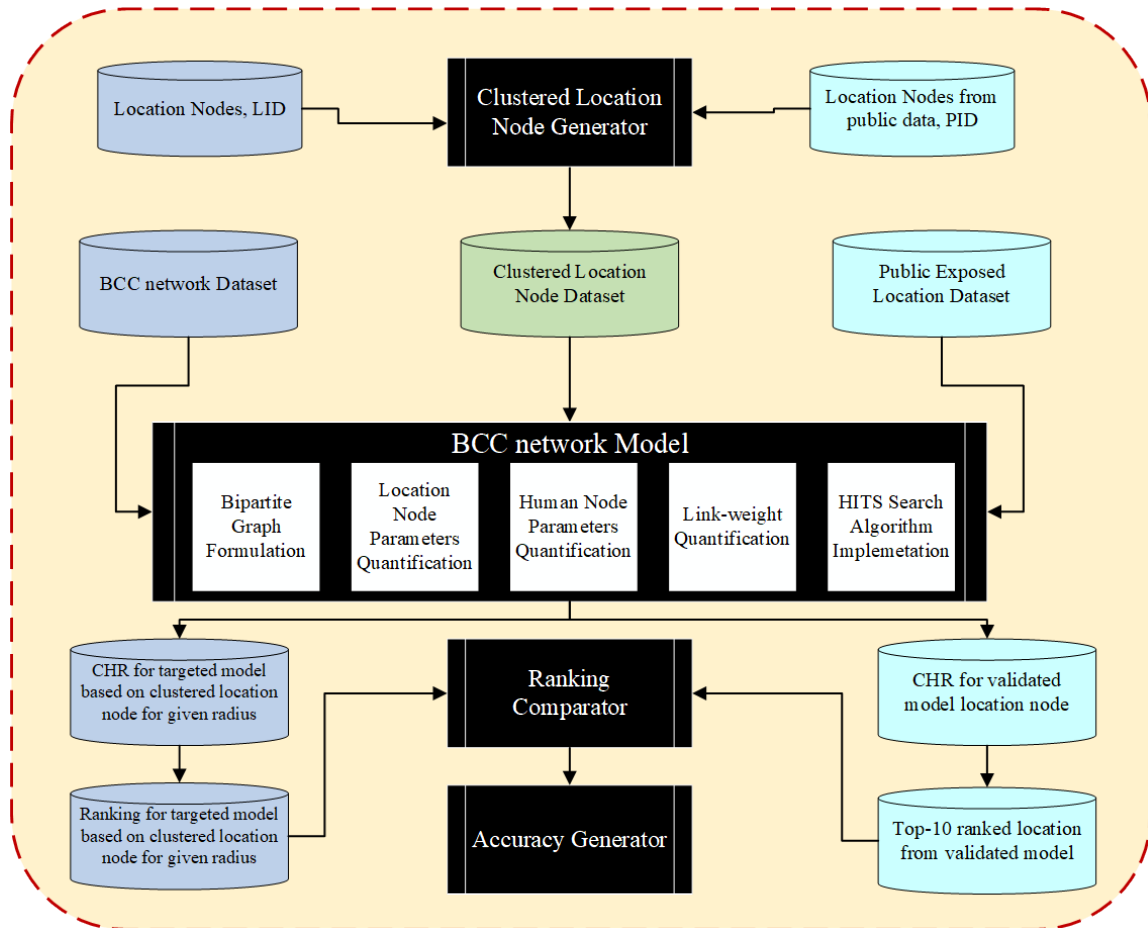


Figure 4.4: Implementation Design of Predictive Power Analysis

Table 4.18: Ranking of radius 200 meters cluster group 1

CID	CHR	Original location node
<i>C1</i>	1	<i>L1, L7</i>
<i>C2</i>	0.2135	<i>L2, L5, L8, P19, P22, P29, P30, P42, P50, P51, P52, P186, P209, P210, P211, P212, P219, P249, P251, P255, P258, P263, P265, P282, P300, P301</i>
<i>C5</i>	0.1712	<i>L6, L11, L12, P7, P8, P38, P96, P110, P150, P241</i>
<i>C13</i>	0.1160	<i>L18, L19, L49, P23, P33, P91, P129, P193, P217, P221, P283, P289, P292, P319, P322</i>
<i>C12</i>	0.08450	<i>L17, L29, L30, L44, P20, P82, P89, P161, P225, P226, P227, P257, P288, P297, P307, P314</i>
<i>C6</i>	0.08161	<i>L9, P4, P146, P293</i>
<i>C3</i>	0.07442	<i>L3</i>
<i>C11</i>	0.03934	<i>L16, P76</i>
<i>C4</i>	0.03861	<i>L4, P1, P35, P37, P84, P99, P152, P155, P182, P185, P190, P220, P232, P236, P286, P309, P310</i>
<i>C7</i>	0.03065	<i>L10</i>
<i>C10</i>	0.03000	<i>L15, L45, P100, P159, P176, P204, P208, P215, P275, P281, P308, P320, P321</i>
<i>C14</i>	0.02998	<i>L20</i>
<i>C9</i>	0.0253	<i>L14, P266, P305</i>
<i>C15</i>	0.02432	<i>L21, P93, P162, P177, P312</i>
<i>C8</i>	0.02432	<i>L13, P90, P153, P169</i>

Table 4.19: Top 10 ranked locations in group 1 validated model (A)

Location node	CHR	Ranking
<i>P173</i>	1	1
<i>P180</i>	2.50E-07	2
<i>P182</i>	2.98E-18	3
<i>P191</i>	1.50E-18	4
<i>P151</i>	1.56E-19	5
<i>P177</i>	5.81E-21	6
<i>P159</i>	1.87E-26	7
<i>P229</i>	1.02E-26	8
<i>P200</i>	1.14E-27	9
<i>P201</i>	1.14E-27	10

Based on the lower accuracy, the ability of the targeted model in 200 m is insufficient to predict the exposed location within the same duration. Hence, the process flow, as shown in Figure 4.4, will be repeated for another radius.

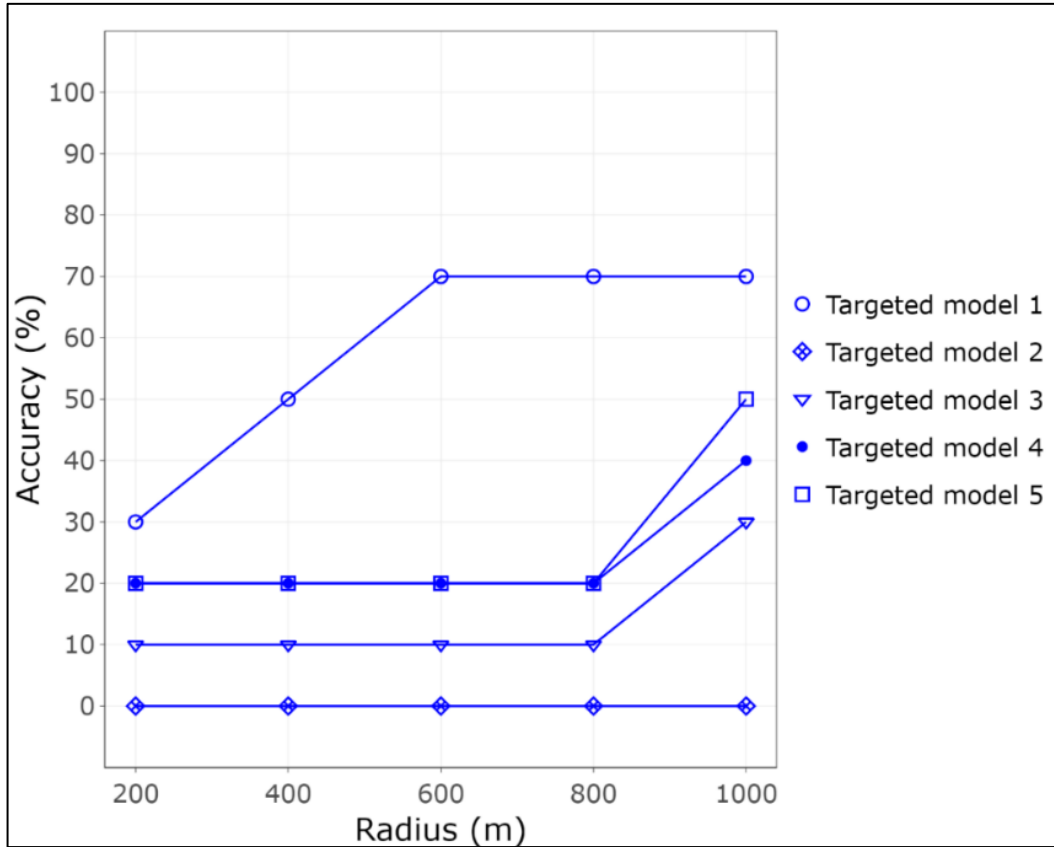


Figure 4.5: Accuracy of each targeted model within the same weeks

Figure 4.5 illustrated the summary of each target model and the accuracy based on the cluster group. The validated models that used to compare ranking for each targeted model utilised the data within the same contact tracing with the datasets used for each targeted model. The highest accuracy is 70% from the targeted model in 600 m after comparing to the corresponding validated model (weeks 5 to 7). This result shows that the BCC network in targeted model 1 can predict the top 10 ranked locations from the public exposed location with 70% accuracy. These results are likely to be related to the number of location nodes

included in the targeted model; the highest number of location nodes for a targeted model has high accuracy for every cluster group.

As mentioned above, the predictive power of a model in temporal perspective needs to be examined. Hence, this study further analysed the accuracy of each target model by comparing the ranking generated for the validated model i) one week after and ii) 2 weeks later. The process is identical as is in Figure 4.4 by changing the datasets used to formulate the validated model. Table 4.20 displays the group of models and the data used to formulate the targeted and validated models. Figures 4.6 and 4.7 illustrate the accuracy for each targeted model on the cluster group to predict the ranking 1 week after and 2 weeks later.

Table 4.20: List of Targeted Models and Validated for predictive power analysis

Group	Data	Model
1	BCC network 1	Targeted Model
	Public exposed location in Epi week 8	Validated Model (A)
	Public exposed location in Epi week 8 and 9	Validated Model (B)
2	BCC network 2	Targeted Model
	Public exposed location in Epi week 4	Validated Model (A)
	Public exposed location in Epi week 4 and 5	Validated Model (B)
3	BCC network 3	Targeted Model
	Public exposed location in Epi week 4	Validated Model (A)
	Public exposed location in Epi week 4 and 5	Validated Model (B)
4	BCC network 4	Targeted Model
	Public exposed location in Epi week 5	Validated Model (A)
	Public exposed location in Epi week 5 and 6	Validated Model (B)
5	BCC network 5	Targeted Model
	Public exposed location in Epi week 5	Validated Model (A)
	Public exposed location in Epi week 5 and 6	Validated Model (B)

Figure 4.6 illustrates the accuracy of the targeted model in predicting the top 10 ranked locations of the validated model. In addition, targeted model 1 obtained the highest accuracy, which is 90%, from 600 m to 1000 m one week later. The accuracy of predicting

the top 10 ranked location of validated model for targeted models 4 and 5 increased from 20% to 50% in 600 m. Both models obtained the highest accuracy (60% to 70%) in 1000 m.

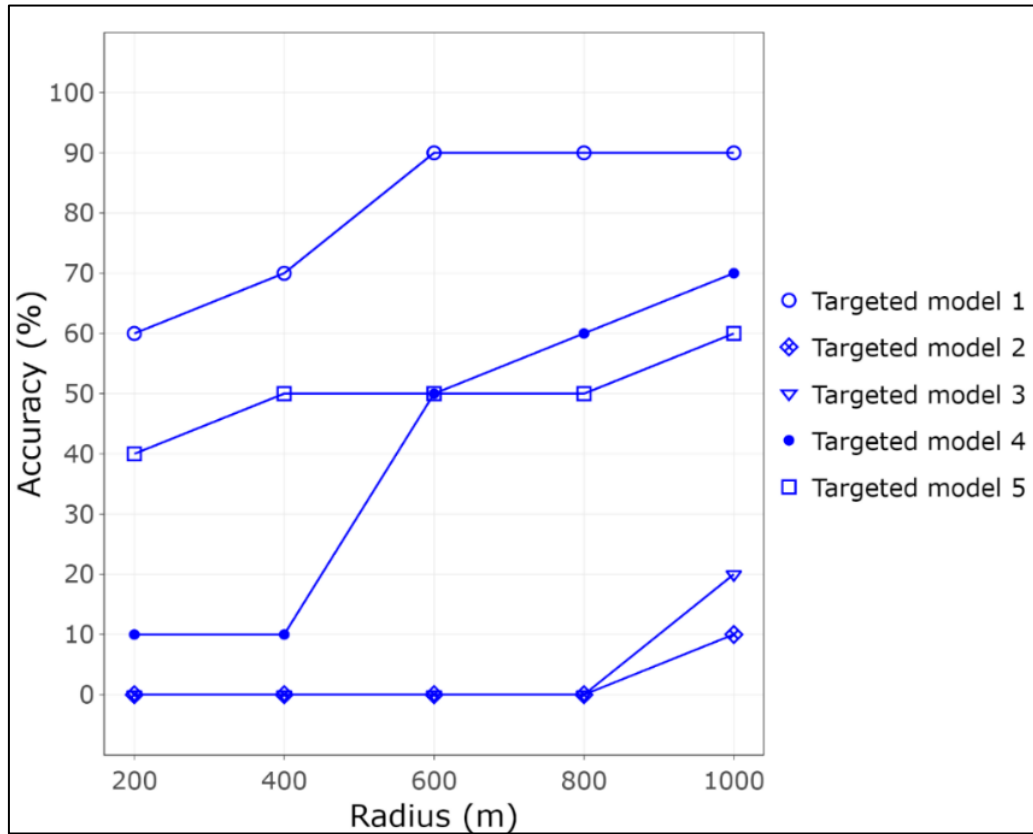


Figure 4.6: Accuracy of each targeted model for subsequent one week

In terms of accuracy of predicting the top 10 ranked locations in the validated model 2 weeks later, the accuracy of targeted model 1 slightly decreased to 60%, as shown in Figure 4.7 compared with Figure 4.6. The highest accuracy in Figure 4.7 is 70%, which was obtained by targeted models 4 and 5. However, the overall highest accuracy decreased from 90% in 600 m to 70% in 1000 m. As shown in Figures 4.6 and 4.7, the ability of the targeted model to predict the top 10 ranked locations of the validated model 1 week after is better than 2 weeks later. In terms of spatiality, the 600 m cluster group returned a better accuracy

compared with other radii. This indicated that most of the high-ranked exposed locations are within 600 me of hotspots identified by the BCC network.

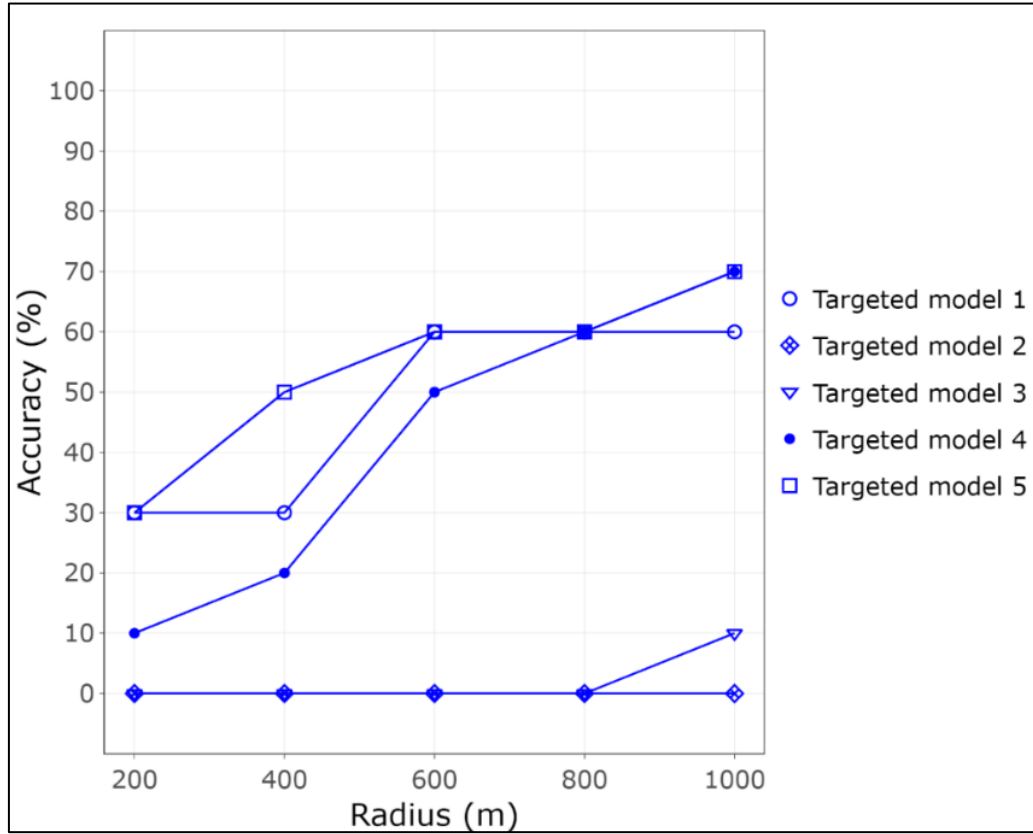


Figure 4.7: Accuracy of each targeted model for subsequent two weeks

In summary, the results of this study indicate that the BNM model can effectively predict the potential hotspots of COVID-19 in terms of both temporal and spatial components. Specifically, the exposed locations in the following week are closely related to the locations visited by infected individuals in the previous week. Additionally, the public health officers should focus on the 600-meter area visited by the infected individual to promptly identify the possible exposed locations and populations and prevent a further outbreak. This demonstrates the improved accuracy of COVID-19 contact tracing achieved through the use of BNM.

4.7 Summary

This chapter presented the last step of stage 2 and stage 3 of BNM-RMF (Figure 3.1) which are the results, including the ranking of location nodes and human nodes based on their CHR and CRR values respectively. The location node with the highest CHR value (*LI*) is ranked as the location with the highest virus density of COVID-19 in BCC network 1 during the studied period. Meanwhile, the human node with the highest CRR value (*HI*) is considered as the only possible super-spreader, while a cluster of non-infectious human nodes with the same CRR values are also identified as having a high risk of being infected.

A benchmark model generated by UCINET 6.0 was used to verify the result generated by the BCC network. An RMSE of 0.0002244 was reported for the ranking of location nodes and 0.001419 for the ranking of human nodes between the benchmark model and BCC network. RMSE values are lower than the threshold value of 0.05 which indicates that the BCC network is verified. Moreover, an analytical verification was conducted to examine the correlation between the CHR and a hub matrix of location nodes. Further, the SRCC was used to determine the correlation, and a reported value of 0.974 signifies that CHR and hub matrix of location nodes are strongly correlated. Model validation was conducted by comparing the ranking result between the targeted and validated models to ensure the model is not bias. SRCC of 0.74 is obtained for location nodes in group 1, and SRCC of 1 is obtained for groups 2 and 3, which indicates that the BCC network is positively and strongly correlated with the result even using difference datasets. The validation of the model by collecting the published exposed location is accomplished, and all the SRCC values for each group are much higher than the threshold value, which demonstrated the ability of BCC network model in predicting COVID-19 hotspots is stable when more location data are used.

The hypothesis of this study has been proven in parameter significance analysis. The parameters capture the integration of human and location are shown strongly significance, lowest SRCC value is obtained when location-related parameters (Ka , Ks , P_{wr}) are excluded from the model. Predictive power analysis was designed in this chapter to determine the ability of the model to predict possible hotspots in perspective of spatial and temporal. Based on the optimize results, the model was able to identify the potential hotspot of future one week with 90% accuracy in 600 m. These findings justified the ability of BNM in formulating a human-location contact network to increase the accuracy of contact tracing of COVID-19.

The next chapter will provide a summary of this research and the contribution of the research and will cover the limitations, suggestions for improvement of this study, and recommendations for future research too.

CHAPTER 5

CONCLUSION

5.1 Introduction

This chapter summarises the entire research activities conducted in this study, the contribution of this study to the related body of knowledge, and the potential future work that can be extended from this study.

5.2 Conclusion of Research Activities

The purpose of this study is to determine the ability of the bipartite network modelling approach in identifying the high-risk location of infectious diseases to increase the accuracy of contact tracing by using COVID-19 as a case study. The primary control measures to break the transmission chain of COVID-19 are performing contact tracing, promptly detecting possible infectious, and isolating the possible infected individuals. Thus, contact tracing is crucial and is conducted when a person tested positive for COVID-19. However, conventional contact tracing is forward tracing and only focused on the interaction between the contacts while neglecting the effect of the location visited by the patient. Only the close contacts with the patient were required to self-monitoring at the initial stage of the outbreak. Most of the existing studies of COVID-19 transmission focused on the homogenous level of prediction, neglecting the heterogeneous factors between humans and the location visited. Hence, the bipartite network modelling approach was presented to integrate patient mobility to identify the source of infection.

This study intends to formulate a network model that integrates the contact between the human and the location visited. The human in this context is categorised into two groups: infectious and susceptible. Hence, the location visited by an infectious is taken into network

formulation. To achieve the primary research objective, three pillars in disease transmission studies have been identified: host, agent, and environmental properties. In this study, the virus can stay in a location in aerosol form or on surfaces of objects for a certain period of time; thus, the agent was replaced by location to match the research objective. Therefore, ET was formulated and used as the basic building block for bipartite graphs.

In stage 2, the algorithms were designed to convert the location into GPS coordinates and code the location node. Due to the insufficient data for indoor temperature and relative humidity, an algorithm was developed to retrieve these data using an API service. The location and building types were identified in the process of location node declaration. The human node ID was generated and the human health status was coded based on the status extracted from the contact tracing investigation form. Three parameters were introduced for the location node: the stability of the virus in aerosol form (Ka), stability of the virus on surfaces (Ks), and number of times a location is visited by the patient (Fl). Existing experiment models have been adopted for the quantification of Ka and Ks . These two parameters are used to capture the effect of the pathogen for indirect transmission modes. While Fl is quantified via the frequency of times a human visited a location (Fh), Fh for the patient corresponds to Fl of the locations. For the human node, three parameters were subsumed: the indoor infection risk of the susceptible (P_{wr}), total stay duration at one location (D), and Fh . The Well–Riley equation is the technique used to quantify P_{wr} by encompassing the effect of the ventilation rate of the location. The unit of D in this study is in hours and is also used in the quantification of P_{wr} . Therefore, the first objective of this study is achieved after identified the significance parameters in network model.

The summation rule was employed to quantify the link weights (CCS). As a case study, a BCC network was formulated using a single contact tracing data in EW 5. The

network comprises of 31 human nodes, 21 location nodes, and 79 weighted edges. The BCC network acts as an input to an adapted HITS search algorithm to produce CHR values for each location node and CRR values for each human node. This CHR value was used to rank the location according to the virus density, whereas the CRR value indicates the risk of infection. At the end of this stage, the second and third objectives of this study were achieved since the BCC network was formulated and ranked.

In stage 3, UCINET was used to verify the model. The CHR and CRR values obtained in the study were compared with the estimated CHR and CRR values generated via UCINET. The RMSE was computed for CHR and CRR, and the RMSE of the location node is 0.0002244, whereas the RMSE of the human node is 0.001419. Both values are much lower than the threshold value of 0.05. Consequently, analytical verification was conducted by computing the correlation coefficient (SRCC) between CHR values and the sum of the CCS for location nodes, resulting in a value of 0.974, which is higher than the threshold value of 0.70 and thus indicates the model is a high correlation. The impact of removing the human node was determined via model validation. The SRCC for the first targeted and validation models of group 1 is 0.74 for the location node which exceeds the threshold value of 0.70. This implied that removing some human nodes does not affect the model that much. Hence, the model validation in group 2 included all human node identified from the data, and the SRCC value for these two groups was reported as 1, which implies a strong positive correlation.

Moreover, parameter significance analysis was conducted to understand the relative significance of each parameter. Based on the analysis, all parameters were relatively important to the ranking of the location and human nodes. Lastly, the predictive power analysis was performed to examine the ability of BNM. Predictive power analysis has shown

that BNM is able to predict the top 10 ranked exposed locations within 600 m a week after with an accuracy of 90%.

5.3 Research Contributions

This section outlines the contributions of this study upon addressed the research statement proposed in Chapter 1. The main deliverables and main research contributions corresponded to each research objective and research question are summarised in Table 5.1

Table 5.1: Summary of the thesis

Research Questions	Research Objectives	Research Contributions	Deliverables
How to integrate location into existing epidemiology triangle (ET)?	To integrate the characteristics of the COVID-19 virus with the location visited by the patient of COVID-19.	Virus only carries by infected, hence only locations visited by infected are significant. Therefore, agent component in ET can be replaced by location component	Modified ET
How to formulate a bipartite network for a disease consisting of multiple modes and interactions between humans?	To formulate a contact network model based on the interactions between humans and the location visited by the patient of COVID-19.	Introduced node parameters that covered different transmission modes to capture the characteristic of the virus in different scenarios.	1. Basic building block 2. Location node parameters and human node parameters 3. Complete bipartite COVID-19 contact (BCC) network

Table 5.1 Continued

How to quantify the contact network link weight for a disease can be transmitted directly and indirectly?	To identify hotspots of COVID-19 by ranking the link weight between human and location in the contact network using quantification rule and a search algorithm.	The summation rule and HITS adopted in previous research are justified sufficiently to deal with this study.	CHR and CRR to indicate the risk of location and human that can uses in contact tracing to increase the accuracy of contact tracing,
---	---	--	--

- (a) A contact network model of COVID-19 transmission where the heterogeneity of humans and transmission modes are considered was developed.
- (b) This study successfully illustrated the ability of BNM-RMF to formulate a network model of infectious diseases with multiple transmission modes by following its processes.
- (c) This study demonstrated that the agent component in ET can be replaced by the location component for any airborne or droplet-borne disease.
- (d) This study defined the CHR value for the location to integrate the interaction between the human mobility, environmental properties of the location, and stability of the virus.
- (e) This study defined the CRR value for the human that captures the properties of the location and interaction between humans via the same location.
- (f) This study has provided a new sight to identify the risk of location and risk of humans based on the ranking. The ranking of the location node identified the high risk of the source of infection for COVID-19, whereas the ranking of a human node can be used

to compute the human infection risk and the super-spreader if more the one infectious is present.

- (g) This study identifies the significant parameters to predict the possible source of infection and risk of humans, which encompass stability of the virus in aerosol form, stability of the virus on surfaces, indoor infection risk of susceptible, duration of stay at one location, and number of times a human visited a location.
- (h) The spatial and temporal effects are considered for the formulation of the contact network model where the data one week after and distance between each location are used to determine the ability of the network model.

5.4 Future Research

Considerably more work will need to be performed to determine, based on the limitations outlined above, the recommendations for future research to improve and expand this research.

- (a) The status of vaccination for the human and the type of vaccine received can be included in the parameter of the human node. This can increase the heterogeneity for the human node to capture the effect of vaccination in the ranking of location and human nodes.
- (b) The quantification of the location or human node should comply with the SOP, which can be estimated by calculating the effective reproduction number (R_t).
- (c) This study collected the exposed location not included in the raw data to validate the ability of the model. To further determine the potential of BNM, the visitor of these exposed locations can be used to validate its ability in identifying the risk of infection. The data is suggested to be gathered via a contact tracing application, i.e. MySejathera, or by public repositories of Minister of Health (MoH) in Malaysia.

- (d) Integrate the biological properties of the virus, especially the mutation of the virus (variance) in the network. As the properties of each variance are different, this can be introduced to the network model to increase its reliability.
- (e) The model can be extended to other infectious diseases by changing the quantification of virus stability in the location node, for instance, measles, influenza, hand–foot–mouth disease, and MERS. All these diseases are airborne or droplet-borne which can spread indirectly via aerosol and contaminated surfaces.
- (f) Bipartite network modelling can be applied in another domain.

Economy

In the economy, a bipartite network can be used to model the relationship between two distinct types of entities such as firms and consumers or firms and products (Bruno et al., 2018). By constructing a bipartite network, researchers can understand the relationships between firms and the products they produce, as well as the relationships between firms and the consumers that purchase their products.

In the context of the economy, the firm nodes can be connected to product nodes if the firm produces that product. Similarly, the firm nodes can be connected to consumer nodes if the firm has sold a product to that consumer. By analysing the links between firms, products, and consumers, researchers can gain insights into the behaviour of firms, such as which products are popular, which consumers are loyal, and which firms have a large market share. Additionally, this type of analysis can be used to identify patterns in consumer behaviour and to understand the factors that influence consumer purchasing decisions.

In conclusion, the application of bipartite network modelling in the economy can provide valuable insights into the relationships between firms, products, and consumers. This

information can be used to inform business strategies and to understand the underlying mechanisms driving economic activity. Additionally, by using a BNM-RMF, researchers can gain a more comprehensive understanding of the economy and the relationships between different entities within it.

REFERENCE

- Ahlawat, A., Wiedensohler, A., & Mishra, S. K. (2020). An Overview on the Role of Relative Humidity in Airborne Transmission of Sars-Cov-2 in Indoor Environments. *Aerosol Air Quality Ressearch*, 20(9), 1856-1861.
- Ai, Z., Hashimoto, K., & Melikov, A. K. (2019). Influence of Pulmonary Ventilation Rate and Breathing Cycle Period on the Risk of Cross-Infection. *Indoor Air*, 29(6), 993-1004. <https://doi.org/https://doi.org/10.1111/ina.12589>
- Albright, J. J., & Park, H. M. (2009). Confirmatory Factor Analysis Using Amos, Lisrel, Mplus, Sas/Stat Calis.
- Amoatey, P., Omidvarborna, H., Baawain, M. S., & Al-Mamun, A. (2020). Impact of Building Ventilation Systems and Habitual Indoor Incense Burning on Sars-Cov-2 Virus Transmissions in Middle Eastern Countries. *Science of The Total Environment*, 733, 139356. <https://doi.org/https://doi.org/10.1016/j.scitotenv.2020.139356>
- Ancel Meyers, L., Newman, M. E. J., Martin, M., & Schrag, S. (2003). Applying Network Theory to Epidemics: Control Measures for Mycoplasma Pneumoniae Outbreaks. *Emerging Infectious Diseases*, 9(2), 204-210. <https://doi.org/10.3201/eid0902.020188>
- Anderson, E. L., Turnham, P., Griffin, J. R., & Clarke, C. C. (2020). Consideration of the Aerosol Transmission for Covid-19 and Public Health. *Risk Analysis*, 40(5), 902-907. <https://doi.org/https://doi.org/10.1111/risa.13500>
- Andresen, M. A., & Lauzon, E. (2020). Crime hotspots and change points in the COVID-19 pandemic: the case of Vancouver, BC. *Crime Science*, 9(1), 19.

- Arora, P., Kumar, H., & Panigrahi, B. K. (2020). Prediction and Analysis of Covid-19 Positive Cases Using Deep Learning Models: A Descriptive Case Study of India. *Chaos, Solitons & Fractals*, 139, 110017.
- ASHRAE. (2013). Ansi-Ashrae Standard 62.1-2013 Ventilation for Acceptable Indoor Air Quality. In: American Society of Heating, Refrigerating and Air-Conditioning Engineers.
- Azuma, K., Kagi, N., Kim, H., & Hayashi, M. (2020). Impact of Climate and Ambient Air Pollution on the Epidemic Growth During Covid-19 Outbreak in Japan. *Environmental Research*, 190, 110042-110042. <https://doi.org/10.1016/j.envres.2020.110042>
- Bansal, S., Grenfell, B. T., & Meyers, L. A. (2007). When Individual Behaviour Matters: Homogeneous and Network Models in Epidemiology. *Journal of the Royal Society Interface*, 4(16), 879-891. <https://doi.org/doi:10.1098/rsif.2007.1100>
- Bansal, S., Pourbohloul, B., & Meyers, L. A. (2006). A Comparative Analysis of Influenza Vaccination Programs. *PLOS Medicine*, 3(10), e387.
- Bansal, S., Read, J., Pourbohloul, B., & Meyers, L. A. (2010). The Dynamic Nature of Contact Networks in Infectious Disease Epidemiology. *Journal of Biological Dynamics*, 4(5), 478-489.
- Barrat, A., Cattuto, C., Tozzi, A. E., Vanhems, P., & Voirin, N. (2014). Measuring Contact Patterns with Wearable Sensors: Methods, Data Characteristics and Applications to Data-Driven Simulations of Infectious Diseases. *Clinical Microbiology and Infection*, 20(1), 10-16. <https://doi.org/https://doi.org/10.1111/1469-0691.12472>

- Begun, M., Newall, A. T., Marks, G. B., & Wood, J. G. (2013). Contact Tracing of Tuberculosis: A Systematic Review of Transmission Modelling Studies. *PLOS One*, 8(9), e72470.
- Benavides, J., Walsh, P. D., Meyers, L. A., Raymond, M., & Caillaud, D. (2012). Transmission of Infectious Diseases En Route to Habitat Hotspots. *PLOS One*, 7(2), e31290. <https://doi.org/10.1371/journal.pone.0031290>
- Benlloch, J.-M., Cortés, J.-C., Martínez-Rodríguez, D., Julián, R.-S., & Villanueva, R.-J. (2020). Effect of the Early Use of Antivirals on the Covid-19 Pandemic. A Computational Network Modeling Approach. *Chaos, Solitons & Fractals*, 140, 110168. [https://doi.org/https://doi.org/10.1016/j.chaos.2020.110168](https://doi.org/10.1016/j.chaos.2020.110168)
- Biryukov, J., Boydston, J. A., Dunning, R. A., Yeager, J. J., Wood, S., Reese, A. L., . . . Zeitouni, N. E. (2020). Increasing Temperature and Relative Humidity Accelerates Inactivation of Sars-Cov-2 on Surfaces. *MSphere*, 5(4).
- Bisanzio, D., Bertolotti, L., Tomassone, L., Amore, G., Ragagli, C., Mannelli, A., . . . Provero, P. (2010). Modeling the Spread of Vector-Borne Diseases on Bipartite Networks. *PLOS One*, 5(11), e13796. <https://doi.org/10.1371/journal.pone.0013796>
- Bouzerdoun, A., Havstad, A., & Beghdadi, A. (2004). Image Quality Assessment Using a Neural Network Approach. Proceedings of the Fourth IEEE International Symposium on Signal Processing and Information Technology, 2004.,
- Braga, A. A., & Weisburd, D. L. (2010). *Policing Problem Places: Crime Hot Spots and Effective Prevention*. Oxford University Press. <https://doi.org/10.1093/acprof:oso/9780195341966.001.0001>
- Browne, M. W., & Cudeck, R. (1992). Alternative Ways of Assessing Model Fit. *Sociological Methods & Research*, 21(2), 230-258.

- Brundage, J. F., Scott, R. M., Lednar, W. M., Smith, D. W., & Miller, R. N. (1988). Building-Associated Risk of Febrile Acute Respiratory Diseases in Army Trainees. *JAMA*, 259(14), 2108-2112. <https://doi.org/10.1001/jama.1988.03720140028029>
- Bruno, M., Saracco, F., Squartini, T., & Dueñas, M. (2018). Colombian Export Capabilities: Building the Firms-Products Network. *Entropy*, 20(10), 785.
- Bueckert, M., Gupta, R., Gupta, A., Garg, M., & Mazumder, A. (2020). Infectivity of Sars-Cov-2 and Other Coronaviruses on Dry Surfaces: Potential for Indirect Transmission. *Materials (Basel, Switzerland)*, 13(22), 5211. <https://doi.org/10.3390/ma13225211>
- Cai, J., Sun, W., Huang, J., Gamber, M., Wu, J., & He, G. (2020). Indirect Virus Transmission in Cluster of Covid-19 Cases, Wenzhou, China, 2020. *Emerging Infectious Diseases*, 26(6), 1343.
- Carson, J. S. (2002). Model Verification and Validation. Proceedings of the winter simulation conference,
- Chan, K.-H., Sridhar, S., Zhang, R. R., Chu, H., Fung, A.-F., Chan, G., . . . Cheng, V.-C. (2020). Factors Affecting Stability and Infectivity of Sars-Cov-2. *Journal of Hospital Infection*, 106(2), 226-231.
- Chen, S.-C., & Liao, C.-M. (2008). Modelling Control Measures to Reduce the Impact of Pandemic Influenza among Schoolchildren. *Epidemiology & Infection*, 136(8), 1035-1045.
- Cheng, A. H.-D., Golberg, M. A., Kansa, E. J., & Zammuto, G. (2003). Exponential Convergence and H-C Multiquadric Collocation Method for Partial Differential Equations. *Numerical Methods for Partial Differential Equations*, 19(5), 571-594. <https://doi.org/https://doi.org/10.1002/num.10062>

- Chin, A. W. H., Chu, J. T. S., Perera, M. R. A., Hui, K. P. Y., Yen, H.-L., Chan, M. C. W., . . . Poon, L. L. M. (2020). Stability of Sars-Cov-2 in Different Environmental Conditions. *The Lancet Microbe*, *1*(1), e10.
[https://doi.org/https://doi.org/10.1016/S2666-5247\(20\)30003-3](https://doi.org/https://doi.org/10.1016/S2666-5247(20)30003-3)
- Cook, D. A., & Skinner, J. M. (2005). How to Perform Credible Verification, Validation, and Accreditation for Modeling and Simulation. *The Journal of Defense Software Engineering*, *18*(5), 20-24.
- Cooley, D. (2020). *Googleway: Accesses Google Maps Apis to Retrieve Data and Plot Maps*. In <https://CRAN.R-project.org/package=googleway>
- Crandall, M. G., Ishii, H., & Lions, P.-L. (1992). User's Guide to Viscosity Solutions of Second Order Partial Differential Equations. *Bulletin of the American Mathematical Society*, *27*(1), 1-67.
- Cuevas, E. (2020). An Agent-Based Model to Evaluate the Covid-19 Transmission Risks in Facilities. *Computers in Biology and Medicine*, *121*, 103827.
- Dabisch, P., Schuit, M., Herzog, A., Beck, K., Wood, S., Krause, M., . . . Hooper, I. (2021). The Influence of Temperature, Humidity, and Simulated Sunlight on the Infectivity of Sars-Cov-2 in Aerosols. *Aerosol Science Technology*, *55*(2), 142-153.
- Dai, H., & Zhao, B. (2020). Association of the Infection Probability of Covid-19 with Ventilation Rates in Confined Spaces. *Building Simulation*, *13*(6), 1321–1327.
<https://doi.org/10.1007/s12273-020-0703-5>
- Danon, L., Ford, A. P., House, T., Jewell, C. P., Keeling, M. J., Roberts, G. O., . . . Vernon, M. C. (2011). Networks and the Epidemiology of Infectious Disease. *Interdisciplinary Perspectives on Infectious Diseases*, *2011*, 284909.
<https://doi.org/10.1155/2011/284909>

- Dass, S. C., Kwok, W. M., Gibson, G. J., Gill, B. S., Sundram, B. M., & Singh, S. (2021). A Data Driven Change-Point Epidemic Model for Assessing the Impact of Large Gathering and Subsequent Movement Control Order on Covid-19 Spread in Malaysia. *PLOS One*, 16(5), e0252136.
- Davis, P. (1992). Generalizing Concepts of Verification, Validation, and Accreditation (Vv&a) for Military Simulation. *RAND*, October 1992a (to be published as R-4249-ACQ).
- de Dear, R., Leow, K. G., & Foo, S. C. (1991). Thermal Comfort in the Humid Tropics: Field Experiments in Air Conditioned and Naturally Ventilated Buildings in Singapore. *International Journal of Biometeorology*, 34, 259-265.
<https://doi.org/10.1007/BF01041840>
- Demongeot, J., Flet-Berliac, Y., & Seligmann, H. (2020). Temperature Decreases Spread Parameters of the New Covid-19 Case Dynamics. *Biology*, 9(5), 94.
- Dimitrov, N. B., & Meyers, L. A. (2010). Mathematical Approaches to Infectious Disease Prediction and Control. In *Risk and Optimization in an Uncertain World* (pp. 1-25). INFORMS.
- Duan, X. (2015). *Highlights of the Chinese Exposure Factors Handbook*. Academic Press.
- Eames, K. T. D., & Keeling, M. J. (2002). Modeling Dynamic and Network Heterogeneities in the Spread of Sexually Transmitted Diseases. *Proceedings of the National Academy of Sciences*, 99(20), 13330-13335.
<https://doi.org/doi:10.1073/pnas.202244299>
- Eames, K. T. D., & Keeling, M. J. (2003). Contact Tracing and Disease Control. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, 270(1533), 2565-2571.
<https://doi.org/doi:10.1098/rspb.2003.2554>

- Ebrahim, S. H., & Memish, Z. A. (2020). Covid-19 – the Role of Mass Gatherings. *Travel Medicine and Infectious Disease*, 34, 101617.
<https://doi.org/https://doi.org/10.1016/j.tmaid.2020.101617>
- Epstein, J. M. (2009). Modelling to Contain Pandemics. *Nature*, 460(7256), 687-687.
<https://doi.org/10.1038/460687a>
- Escombe, A. R., Oeser, C. C., Gilman, R. H., Navincopa, M., Ticona, E., Pan, W., . . . Moore, D. A. J. (2007). Natural Ventilation for the Prevention of Airborne Contagion. *PLOS Medicine*, 4(2), e68.
- Eubank, S. (2005). Network Based Models of Infectious Disease Spread. *Japanese Journal of Infectious Diseases*, 58(6), S9-13.
- Eubank, S., Guclu, H., Anil Kumar, V., Marathe, M. V., Srinivasan, A., Toroczkai, Z., & Wang, N. (2004). Modelling Disease Outbreaks in Realistic Urban Social Networks. *Nature*, 429(6988), 180-184.
- Eze, M., Labadin, J., & Lim, T. (2011). Contact Strength Generating Algorithm for Application in Malaria Transmission Network. 2011 7th International Conference on Information Technology in Asia,
- Eze, M. O. (2013). *Web Algorithm Search Engine Based Network Modeling of Malaria Transmission* [Doctoral dissertation, Universiti Malaysia Sarawak (UNIMAS)].
- Fears, A. C., Klimstra, W. B., Duprex, P., Hartman, A., Weaver, S. C., Plante, K. S., . . . Fernández, D. (2020). Persistence of Severe Acute Respiratory Syndrome Coronavirus 2 in Aerosol Suspensions. *Emerging infectious diseases*, 26(9), 2168.
- Fernstrom, A., & Goldblatt, M. (2013). Aerobiology and Its Role in the Transmission of Infectious Diseases. *Journal of Pathogens*, 2013, 493960.
<https://doi.org/10.1155/2013/493960>

- Ferretti, L., Wymant, C., Kendall, M., Zhao, L., Nurtay, A., Abeler-Dörner, L., . . . Fraser, C. (2020). Quantifying Sars-Cov-2 Transmission Suggests Epidemic Control with Digital Contact Tracing. *Science*, 368(6491), eabb6936. <https://doi.org/10.1126/science.abb6936>
- Fiorillo, L., Cervino, G., Matarese, M., D'Amico, C., Surace, G., Paduano, V., . . . Cicciù, M. (2020). Covid-19 Surface Persistence: A Recent Data Summary and Its Importance for Medical and Dental Settings. *International Journal of Environmental Research and Public Health*, 17(9), 3132. <https://www.mdpi.com/1660-4601/17/9/3132>
- Galea, S., Riddle, M., & Kaplan, G. A. (2010). Causal Thinking and Complex System Approaches in Epidemiology. *International Journal of Epidemiology*, 39(1), 97-106.
- Gao, W., Wu, H., Siddiqui, M. K., & Baig, A. Q. (2018). Study of Biological Networks Using Graph Theory. *Saudi Journal of Biological Sciences*, 25(6), 1212-1219. <https://doi.org/https://doi.org/10.1016/j.sjbs.2017.11.022>
- Gao, X., Wei, J., Lei, H., Xu, P., Cowling, B. J., & Li, Y. (2016). Building Ventilation as an Effective Disease Intervention Strategy in a Dense Indoor Contact Network in an Ideal City. *PLOS One*, 11(9), e0162481.
- Garnett, G. P. (2002). An Introduction to Mathematical Models in Sexually Transmitted Disease Epidemiology. *Sexually Transmitted Infections*, 78(1), 7-12.
- Gill, B. S., Jayaraj, V. J., Singh, S., Mohd Ghazali, S., Cheong, Y. L., Md Iderus, N. H., . . . Labadin, J. (2020). Modelling the Effectiveness of Epidemic Control Measures in Preventing the Transmission of Covid-19 in Malaysia. *International Journal of Environmental Research and Public Health*, 17(15), 5509.

- Godri Pollitt, K. J., Peccia, J., Ko, A. I., Kaminski, N., Dela Cruz, C. S., Nebert, D. W., . . . Vasiliou, V. (2020). Covid-19 Vulnerability: The Potential Impact of Genetic Susceptibility and Airborne Transmission. *Human Genomics, 14*, 1-7.
- Goscé, L., & Johansson, A. (2018). Analysing the Link between Public Transport Use and Airborne Transmission: Mobility and Contagion in the London Underground. *Environmental Health, 17*(1), 1-11.
- Guo, Z.-D., Wang, Z.-Y., Zhang, S.-F., Li, X., Li, L., Li, C., . . . Chen, W. (2020). Aerosol and Surface Distribution of Severe Acute Respiratory Syndrome Coronavirus 2 in Hospital Wards, Wuhan, China, 2020. *Emerging Infectious Disease journal, 26*(7), 1583. <https://doi.org/10.3201/eid2607.200885>
- Harary, F. (1969). Graph Theory Addison-Wesley Reading Ma Usa. In (pp. 1975): Wiley New York.
- Hertzberg, V. S., Weiss, H., Elon, L., Si, W., Norris, S. L., Baker, H., . . . Wu, H. (2018). Behaviors, Movements, and Transmission of Droplet-Mediated Respiratory Diseases During Transcontinental Airline Flights. *Proceedings of the National Academy of Sciences, 115*(14), 3623-3627. <https://doi.org/doi:10.1073/pnas.1711611115>
- Hinkle, D. E., Wiersma, W., & Jurs, S. G. (2003). *Applied Statistics for the Behavioral Sciences* (Vol. 663). Houghton Mifflin College Division.
- Hyman, J. M., Li, J., & Stanley, E. A. (2003). Modeling the Impact of Random Screening and Contact Tracing in Reducing the Spread of HIV. *Mathematical Biosciences, 181*(1), 17-54.
- Jalabneh, R., Zehra Syed, H., Pillai, S., Hoque Apu, E., Hussein, M. R., Kabir, R., . . . Azim Majumder, M. (2020). Use of Mobile Phone Apps for Contact Tracing to Control the Covid-19 Pandemic: A Literature Review. *Anwarul, Use of Mobile Phone Apps for*

Contact Tracing to Control the COVID-19 Pandemic: A Literature Review (July 1, 2020).

- Jamaludin, N., Mohammed, N. I., Khamidi, M. F., & Wahab, S. N. A. (2015). Thermal Comfort of Residential Building in Malaysia at Different Micro-Climates. *Procedia-Social and Behavioral Sciences*, 170, 613-623.
- Jeong, E., Hagose, M., Jung, H., Ki, M., & Flahault, A. (2020). Understanding South Korea's Response to the Covid-19 Outbreak: A Real-Time Analysis. *International Journal of Environmental Research and Public Health*, 17(24), 9571.
- Jones, R. M., & Brosseau, L. M. (2015). Aerosol Transmission of Infectious Disease. *Journal of Occupational and Environmental Medicine*, 57(5), 501-508.
<https://doi.org/10.1097/jom.0000000000000448>
- Karaivanov, A. (2020). A Social Network Model of Covid-19. *PloS one*, 15(10), e0240878.
- Killingley, B., & Nguyen - Van - Tam, J. (2013). Routes of Influenza Transmission. *Influenza and Other Respiratory Viruses*, 7, 42-51.
- Kim, U. J., Lee, S. Y., Lee, J. Y., Lee, A., Kim, S. E., Choi, O.-J., . . . Jang, H.-C. (2020). Air and Environmental Contamination Caused by Covid-19 Patients: A Multi-Center Study. *Journal of Korean Medical Science*, 35(37), e332-330.
<https://doi.org/10.3346/jkms.2020.35.e332>
- Kiss, I. Z., Green, D. M., & Kao, R. R. (2008). The Effect of Network Mixing Patterns on Epidemic Dynamics and the Efficacy of Disease Contact Tracing. *Journal of the Royal Society Interface*, 5(24), 791-799. <https://doi.org/doi:10.1098/rsif.2007.1272>
- Klov Dahl, A. S. (1985). Social Networks and the Spread of Infectious Diseases: The Aids Example. *Social Science & Medicine*, 21(11), 1203-1216.

- Klov Dahl, A. S., Potterat, J. J., Woodhouse, D. E., Muth, J. B., Muth, S. Q., & Darrow, W. W. (1994). Social Networks and Infectious Disease: The Colorado Springs Study. *Social Science & Medicine*, 38(1), 79-88.
- Kohanski, M. A., Lo, L. J., & Waring, M. S. (2020). Review of Indoor Aerosol Generation, Transport, and Control in the Context of Covid-19. *International Forum of Allergy & Rhinology*, 10(10), 1173-1179. <https://doi.org/https://doi.org/10.1002/alr.22661>
- Kok, W. C. (2017). *A Computational Approach to Predict the Spread of Dengue* [Unpublished master's dissertation, Universiti Malaysia Sarawak, (UNIMAS)].
- Kok, W. C., Labadin, J., & Perera, D. (2017). Modeling Dengue Hotspot with Bipartite Network Approach. International Conference on Computational Science and Technology,
- Kucharski, A. J., Russell, T. W., Diamond, C., Liu, Y., Edmunds, J., Funk, S., . . . Munday, J. D. (2020). Early Dynamics of Transmission and Control of Covid-19: A Mathematical Modelling Study. *The Lancet Infectious Diseases*.
- Kuddus, M. A., McBryde, E. S., & Adegboye, O. A. (2019). Delay Effect and Burden of Weather-Related Tuberculosis Cases in Rajshahi Province, Bangladesh, 2007–2012. *Scientific Reports*, 9(1), 1-13.
- Kulldorff, M., & Nagarwalla, N. (1995). Spatial Disease Clusters: Detection and Inference. *Statistics in Medicine*, 14(8), 799-810. <https://doi.org/https://doi.org/10.1002/sim.4780140809>
- Kutter, J. S., Spronken, M. I., Fraaij, P. L., Fouchier, R. A., & Herfst, S. (2018). Transmission Routes of Respiratory Viruses among Humans. *Current Opinion in Virology*, 28, 142-151.

- Kwok, K. O., Tang, A., Wei, V. W., Park, W. H., Yeoh, E. K., & Riley, S. (2019). Epidemic Models of Contact Tracing: Systematic Review of Transmission Studies of Severe Acute Respiratory Syndrome and Middle East Respiratory Syndrome. *Computational and Structural Biotechnology Journal*, 17, 186-194.
- Kwon, K.-S., Park, J.-I., Park, Y. J., Jung, D.-M., Ryu, K.-W., & Lee, J.-H. (2020). Evidence of Long-Distance Droplet Transmission of Sars-Cov-2 by Direct Air Flow in a Restaurant in Korea. *Journal of Korean Medical Science*, 35(46).
- Lessler, J., Azman, A. S., McKay, H. S., & Moore, S. M. (2017). What Is a Hotspot Anyway? *The American Journal of Tropical Medicine and Hygiene*, 96(6), 1270-1273.
- Liew, C. Y. (2016). *Bipartite-Network-Based Modeling of Habitat Suitability* [Unpublished doctoral dissertation, Universiti Malaysia Sarawak (UNIMAS)].
- Lim, W. K., Wang, K., Lefebvre, C., & Califano, A. (2007). Comparative Analysis of Microarray Normalization Procedures: Effects on Reverse Engineering Gene Networks. *Bioinformatics*, 23(13), i282-i288.
- Lindsley, W. G., Noti, J. D., Blachere, F. M., Thewlis, R. E., Martin, S. B., Othumpangat, S., . . . Beezhold, D. H. (2015). Viable Influenza a Virus in Airborne Particles from Human Coughs. *Journal of Occupational and Environmental Hygiene*, 12(2), 107-113. <https://doi.org/10.1080/15459624.2014.973113>
- Lu, J., Gu, J., Li, K., Xu, C., Su, W., Lai, Z., . . . Yang, Z. (2020). Covid-19 Outbreak Associated with Air Conditioning in Restaurant, Guangzhou, China, 2020. *Emerging Infectious Diseases*, 26(7), 1628.
- Luo, L., Liu, D., Zhang, H., Li, Z., Zhen, R., Zhang, X., . . . Mao, C. (2020). Air and Surface Contamination in Non-Health Care Settings among 641 Environmental Specimens

- of 39 Covid-19 Cases. *PLOS Neglected Tropical Diseases*, 14(10), e0008570.
<https://doi.org/10.1371/journal.pntd.0008570>
- Luongo, J. C., Fennelly, K. P., Keen, J. A., Zhai, Z. J., Jones, B. W., & Miller, S. L. (2016). Role of Mechanical Ventilation in the Airborne Transmission of Infectious Agents in Buildings. *Indoor Air*, 26(5), 666-678.
<https://doi.org/https://doi.org/10.1111/ina.12267>
- Ma, Y., Zhao, Y., Liu, J., He, X., Wang, B., Fu, S., . . . Luo, B. (2020). Effects of Temperature Variation and Humidity on the Death of Covid-19 in Wuhan, China. *Science of The Total Environment*, 724, 138226.
- Mason, O., & Verwoerd, M. (2007). Graph Theory and Networks in Biology. *IET Systems Biology*, 1(2), 89-119.
- Masuda, N., & Holme, P. (2013). Predicting and Controlling Infectious Disease Epidemics Using Temporal Networks. *F1000Prime Reports*, 5, 6-6.
<https://doi.org/10.12703/P5-6>
- Mat, N. F. C., Edinur, H. A., Razab, M. K. A. A., & Safuan, S. (2020). A Single Mass Gathering Resulted in Massive Transmission of Covid-19 Infections in Malaysia with Further International Spread. *Journal of Travel Medicine*.
- Meselson, M. (2020). Droplets and Aerosols in the Transmission of Sars-Cov-2. *New England Journal of Medicine*, 382(21), 2063-2063.
<https://doi.org/10.1056/NEJMc2009324>
- Meyers, L. A., Newman, M., & Pourbohloul, B. (2006). Predicting Epidemics on Directed Contact Networks. *Journal of Theoretical Biology*, 240(3), 400-418.
- Meyers, L. A., Pourbohloul, B., Newman, M. E. J., Skowronski, D. M., & Brunham, R. C. (2005). Network Theory and Sars: Predicting Outbreak Diversity. *Journal of*

<https://doi.org/https://doi.org/10.1016/j.jtbi.2004.07.026>

Mittermeier, R. A., Turner, W. R., Larsen, F. W., Brooks, T. M., & Gascon, C. (2011).

Global biodiversity conservation: the critical role of hotspots. *Biodiversity hotspots: distribution and protection of conservation priority areas*, 3-22.

MOH. (2020). *How Contact Tracing Is Done for Covid-19*. The Ministry of Health Malaysia.

Retrieved 5 Aug 2020 from

<https://www.infosihat.gov.my/index.php/multimedia/infografik/item/how-contact-tracing-is-done-for-covid-19>

Morawska, L., Tang, J. W., Bahnfleth, W., Bluyssen, P. M., Boerstra, A., Buonanno, G., . .

. Querol, X. (2020). How Can Airborne Transmission of Covid-19 Indoors Be Minimised?

Mossong, J., Hens, N., Jit, M., Beutels, P., Auranen, K., Mikolajczyk, R., . . . Edmunds, W.

J. (2008). Social Contacts and Mixing Patterns Relevant to the Spread of Infectious Diseases. *PLOS Medicine*, 5(3), e74. <https://doi.org/10.1371/journal.pmed.0050074>

Mukaka, M. M. (2012). A Guide to Appropriate Use of Correlation Coefficient in Medical

Research. *Malawi Medical Journal*, 24(3), 69-71.

Myers, N., Mittermeier, R. A., Mittermeier, C. G., da Fonseca, G. A. B., & Kent, J. (2000).

Biodiversity Hotspots for Conservation Priorities. *Nature*, 403(6772), 853-858.

<https://doi.org/10.1038/35002501>

Nelson, P. W., & Perelson, A. S. (2002). Mathematical Analysis of Delay Differential

Equation Models of Hiv-1 Infection. *Mathematical Biosciences*, 179(1), 73-94.

Newman, M. E. (2002). Spread of Epidemic Disease on Networks. *Physical Review E*, 66(1),

016128.

- Nguyen, J. L., Schwartz, J., & Dockery, D. W. (2014). The Relationship between Indoor and Outdoor Temperature, Apparent Temperature, Relative Humidity, and Absolute Humidity. *Indoor Air*, 24(1), 103-112.
<https://doi.org/https://doi.org/10.1111/ina.12052>
- Nicas, M. (1996). An Analytical Framework for Relating Dose, Risk, and Incidence: An Application to Occupational Tuberculosis Infection. *Risk Analysis*, 16(4), 527-538.
- Nicas, M., Nazaroff, W. W., & Hubbard, A. (2005). Toward Understanding the Risk of Secondary Airborne Infection: Emission of Respirable Pathogens. *Journal of Occupational and Environmental Hygiene*, 2(3), 143-154.
<https://doi.org/10.1080/15459620590918466>
- Noakes, C. J., Beggs, C. B., Sleight, P. A., & Kerr, K. G. (2006). Modelling the Transmission of Airborne Infections in Enclosed Spaces. *Epidemiology and Infection*, 134(5), 1082-1091. <https://doi.org/10.1017/S0950268806005875>
- Pavlopoulos, G. A., Secrier, M., Moschopoulos, C. N., Soldatos, T. G., Kossida, S., Aerts, J., . . . Bagos, P. G. (2011). Using Graph Theory to Analyze Biological Networks. *BioData Mining*, 4(1), 10. <https://doi.org/10.1186/1756-0381-4-10>
- Pellis, L., Ball, F., Bansal, S., Eames, K., House, T., Isham, V., & Trapman, P. (2015). Eight Challenges for Network Epidemic Models. *Epidemics*, 10, 58-62.
<https://doi.org/https://doi.org/10.1016/j.epidem.2014.07.003>
- Peng, X., Xu, X., Li, Y., Cheng, L., Zhou, X., & Ren, B. (2020). Transmission Routes of 2019-Ncov and Controls in Dental Practice. *International Journal of Oral Science*, 12(1), 1-6.
- Pica, N., & Bouvier, N. M. (2012). Environmental Factors Affecting the Transmission of Respiratory Viruses. *Current Opinion in Virology*, 2(1), 90-95.

- Prem, K., Liu, Y., Russell, T. W., Kucharski, A. J., Eggo, R. M., Davies, N., . . . Klepac, P. (2020). The Effect of Control Strategies to Reduce Social Mixing on Outcomes of the Covid-19 Epidemic in Wuhan, China: A Modelling Study. *The Lancet Public Health*, 5(5), e261-e270. [https://doi.org/https://doi.org/10.1016/S2468-2667\(20\)30073-6](https://doi.org/https://doi.org/10.1016/S2468-2667(20)30073-6)
- Rahman, H. S., Aziz, M. S., Hussein, R. H., Othman, H. H., Omer, S. H. S., Khalid, E. S., . . . Abdullah, R. (2020). The Transmission Modes and Sources of Covid-19: A Systematic Review. *International Journal of Surgery Open*, 26, 125-136.
- Rayfield, B., Fortin, M.-J., & Fall, A. (2011). Connectivity for Conservation: A Framework to Classify Network Measures. *Ecology*, 92(4), 847-858.
- Renner, I. W., Warton, D. I., Ramp, D., & Wintle, B. A. (2015). Identifying and managing threatened Australian plant species using a Bayesian hierarchical modelling framework for spatially correlated data. *Diversity and Distributions*, 21(10), 1192-1205.
- Riley, E., Murphy, G., & Riley, R. (1978). Airborne Spread of Measles in a Suburban Elementary School. *American Journal of Epidemiology*, 107(5), 421-432.
- Robineau, O., Gomes, M. F. C., Kendall, C., Kerr, L., Périssé, A., & Boëlle, P.-Y. (2020). Model-Based Respondent-Driven Sampling Analysis for Hiv Prevalence in Brazilian Msm. *Scientific Reports*, 10(1), 2646. <https://doi.org/10.1038/s41598-020-59567-2>
- Robinson, S. (1997). Simulation Model Verification and Validation: Increasing the Users' Confidence. Proceedings of the 29th conference on Winter simulation,
- Rocha, L. E. C., Liljeros, F., & Holme, P. (2011). Simulated Epidemics in an Empirical Spatiotemporal Network of 50,185 Sexual Contacts. *PLOS Computational Biology*, 7(3), e1001109. <https://doi.org/10.1371/journal.pcbi.1001109>

- Rudnick, S. N., & Milton, D. K. (2003). Risk of Indoor Airborne Infection Transmission Estimated from Carbon Dioxide Concentration. *Indoor Air*, 13(3), 237-245.
<https://doi.org/10.1034/j.1600-0668.2003.00189.x>
- Runkle, J. D., Sugg, M. M., Leeper, R. D., Rao, Y., Matthews, J. L., & Rennie, J. J. (2020). Short-Term Effects of Specific Humidity and Temperature on Covid-19 Morbidity in Select Us Cities. *Science of The Total Environment*, 740, 140093.
- Salathé, M., Kazandjieva, M., Lee, J. W., Levis, P., Feldman, M. W., & Jones, J. H. (2010). A High-Resolution Human Contact Network for Infectious Disease Transmission. *Proceedings of the National Academy of Sciences*, 107(51), 22020-22025.
<https://doi.org/doi:10.1073/pnas.1009094108>
- Santarpia, J. L., Rivera, D. N., Herrera, V. L., Morwitzer, M. J., Creager, H. M., Santarpia, G. W., . . . Broadhurst, M. J. (2020). Aerosol and Surface Contamination of Sars-Cov-2 Observed in Quarantine and Isolation Care. *Scientific Reports*, 10(1), 1-8.
- Shaffiee Haghshenas, S., Pirouz, B., Shaffiee Haghshenas, S., Pirouz, B., Piro, P., Na, K.-S., . . . Geem, Z. W. (2020). Prioritizing and Analyzing the Role of Climate and Urban Parameters in the Confirmed Cases of Covid-19 Based on Artificial Intelligence Applications. *International Journal of Environmental Research and Public Health*, 17(10), 3730.
- Shankar, A. G., Janmohamed, K., Olowokure, B., Smith, G. E., Hogan, A. H., De Souza, V., . . . Cleary, P. (2014). Contact Tracing for Influenza a (H1n1) Pdm09 Virus–Infected Passenger on International Flight. *Emerging Infectious Diseases*, 20(1), 118.
- Shen, M., Peng, Z., Xiao, Y., & Zhang, L. (2020). Modelling the Epidemic Trend of the 2019 Novel Coronavirus Outbreak in China. *bioRxiv*, 2020.2001.2023.916726.
<https://doi.org/10.1101/2020.01.23.916726>

- Shi, P., Dong, Y., Yan, H., Zhao, C., Li, X., Liu, W., . . . Xi, S. (2020). Impact of Temperature on the Dynamics of the Covid-19 Outbreak in China. *Science of The Total Environment*, 728, 138890.
- Shirinivas, S., Vetrivel, S., & Elango, N. (2010). Applications of Graph Theory in Computer Science an Overview. *International Journal of Engineering Science and Technology*, 2(9), 4610-4621.
- Shmueli, G. (2010). To Explain or to Predict? *Statistical Science*, 25(3), 289-310.
- So, M. K. P., Tiwari, A., Chu, A. M. Y., Tsang, J. T. Y., & Chan, J. N. L. (2020). Visualizing Covid-19 Pandemic Risk through Network Connectedness. *International Journal of Infectious Diseases*, 96, 558-561.
<https://doi.org/https://doi.org/10.1016/j.ijid.2020.05.011>
- Sun, C., & Zhai, Z. (2020). The Efficacy of Social Distance and Ventilation Effectiveness in Preventing Covid-19 Transmission. *Sustainable Cities and Society*, 62, 102390.
- Sun, Y., Wang, Z., Zhang, Y., & Sundell, J. (2011). In China, Students in Crowded Dormitories with a Low Ventilation Rate Have More Common Colds: Evidence for Airborne Transmission. *PLOS One*, 6(11), e27140.
<https://doi.org/10.1371/journal.pone.0027140>
- Sundell, N., Andersson, L.-M., Brittain-Long, R., Lindh, M., & Westin, J. (2016). A Four Year Seasonal Survey of the Relationship between Outdoor Climate and Epidemiology of Viral Respiratory Tract Infections in a Temperate Climate. *Journal of Clinical Virology*, 84, 59-63.
- Sze To, G. N., & Chao, C. Y. H. (2010). Review and Comparison between the Wells-Riley and Dose-Response Approaches to Risk Assessment of Infectious Respiratory Diseases. *Indoor Air*, 20(1), 2-16. <https://doi.org/10.1111/j.1600-0668.2009.00621.x>

- Sze To, G. N., & Chao, C. Y. H. (2010). Review and Comparison between the Wells – Riley and Dose - Response Approaches to Risk Assessment of Infectious Respiratory Diseases. *Indoor Air*, 20(1), 2-16.
- Sze-To, G. N., Yang, Y., Kwan, J. K. C., Yu, S. C. T., & Chao, C. Y. H. (2014). Effects of Surface Material, Ventilation, and Human Behavior on Indirect Contact Transmission Risk of Respiratory Infection. *Risk Analysis*, 34(5), 818-830. <https://doi.org/10.1111/risa.12144>
- Tang, J. W., & Loh, T. P. (2014). Correlations between Climate Factors and Incidence—a Contributor to Rsv Seasonality. *Reviews in Medical Virology*, 24(1), 15-34.
- Tellier, R. (2006). Review of Aerosol Transmission of Influenza a Virus. *Emerging Infectious Diseases*, 12(11), 1657-1662. <https://doi.org/10.3201/eid1211.060426>
- Teunis, P. F., Moe, C. L., Liu, P., E. Miller, S., Lindesmith, L., Baric, R. S., . . . Calderon, R. L. (2008). Norwalk Virus: How Infectious Is It? *Journal of Medical Virology*, 80(8), 1468-1476.
- Thole, S., Kalhoefer, D., an der Heiden, M., Nordmann, D., Daniels-Haardt, I., & Jurke, A. (2019). Contact Tracing Following Measles Exposure on Three International Flights, Germany, 2017. *Eurosurveillance*, 24(19), 1800500.
- Urban, D. L., Minor, E. S., Treml, E. A., & Schick, R. S. (2009). Graph Models of Habitat Mosaics. *Ecology letters*, 12(3), 260-273.
- van Doremalen, N., Bushmaker, T., Morris, D. H., Holbrook, M. G., Gamble, A., Williamson, B. N., . . . Munster, V. J. (2020). Aerosol and Surface Stability of Sars-Cov-2 as Compared with Sars-Cov-1. *The New England Journal of Medicine*, 382(16), 1564-1567. <https://doi.org/10.1056/NEJMc2004973>

- Vazquez-Prokopec, G. M., Montgomery, B. L., Horne, P., Clennon, J. A., & Ritchie, S. A. (2017). Combining Contact Tracing with Targeted Indoor Residual Spraying Significantly Reduces Dengue Transmission. *Science Advances*, 3(2), e1602024.
- Viguerie, A., Veneziani, A., Lorenzo, G., Baroli, D., Aretz-Nellesen, N., Patton, A., . . . Auricchio, F. (2020). Diffusion–Reaction Compartmental Models Formulated in a Continuum Mechanics Framework: Application to Covid-19, Mathematical Analysis, and Numerical Study. *Computational Mechanics*, 1-22.
- Wahid, J., & Hussin, N. M. (2016). Combination of Graph Heuristics in Producing Initial Solution of Curriculum Based Course Timetabling Problem. AIP Conference Proceedings,
- Wang, X.-L., Yang, L., He, D.-H., Chiu, A. P., Chan, K.-H., Chan, K.-P., . . . Hu, W. (2017). Different Responses of Influenza Epidemic to Weather Factors among Shanghai, Hong Kong, and British Columbia. *International Journal of Biometeorology*, 61(6), 1043-1053.
- Wang, Y., Zhou, Y., Xie, H., Liu, C., & Zhao, S. (2017). Identifying potential hotspots of endemic birds in China through spatial conservation prioritization. *PeerJ*, 5, e2975.
- Wang, Z., Zhang, X., Teichert, G., Carrasco-Teja, M., & Garikipati, K. (2020). System Inference for the Spatio-Temporal Evolution of Infectious Diseases: Michigan in the Time of Covid-19. *Computational Mechanics*, 1-24.
- Wells, W. F. (1955). Airborne Contagion and Air Hygiene. An Ecological Study of Droplet Infections. *Airborne Contagion and Air Hygiene. An Ecological Study of Droplet Infections*.
- WHO. (2017). *Contact Tracing*. World Health Organization. Retrieved 1 August from <https://www.who.int/news-room/q-a-detail/contact-tracing>

- WHO. (2020). *Modes of Transmission of Virus Causing Covid-19: Implications for Ipc Precaution Recommendations*. World Health Organization. Retrieved 11 November 2020 from <https://www.who.int/news-room/commentaries/detail/modes-of-transmission-of-virus-causing-covid-19-implications-for-ipc-precaution-recommendations>
- Wu, H., Wang, H., Wang, Q., Xin, Q., & Lin, H. (2014). The Effect of Meteorological Factors on Adolescent Hand, Foot, and Mouth Disease and Associated Effect Modifiers. *Global Health Action*, 7(1), 24664. <https://doi.org/10.3402/gha.v7.24664>
- Wu, J. T., Leung, K., & Leung, G. M. (2020). Nowcasting and Forecasting the Potential Domestic and International Spread of the 2019-Ncov Outbreak Originating in Wuhan, China: A Modelling Study. *The Lancet*, 395(10225), 689-697.
- Wylie, J. L., & Jolly, A. N. N. (2001). Patterns of Chlamydia and Gonorrhea Infection in Sexual Networks in Manitoba, Canada. *Sexually Transmitted Diseases*, 28(1), 14-24. <http://www.jstor.org/stable/44965398>
- Xiao, Y., & Torok, M. E. (2020). Taking the Right Measures to Control Covid-19. *The lancet infectious diseases*, 20(5), 523-524. [https://doi.org/10.1016/S1473-3099\(20\)30152-3](https://doi.org/10.1016/S1473-3099(20)30152-3)
- Xue, L., Jing, S., Miller, J. C., Sun, W., Li, H., Estrada-Franco, J. G., . . . Zhu, H. (2020). A Data-Driven Network Model for the Emerging Covid-19 Epidemics in Wuhan, Toronto and Italy. *Mathematical Biosciences*, 326, 108391. <https://doi.org/https://doi.org/10.1016/j.mbs.2020.108391>
- Yang, C. H., & Jung, H. (2020). Topological Dynamics of the 2015 South Korea Mers-Cov Spread-on-Contact Networks. *Scientific Reports*, 10(1), 1-11.

- Zaplotnik, Ž., Gavrić, A., & Medic, L. (2020). Simulation of the Covid-19 Epidemic on the Social Network of Slovenia: Estimating the Intrinsic Forecast Uncertainty. *PLOS One*, 15(8), e0238090. <https://doi.org/10.1371/journal.pone.0238090>
- Zemouri, C., Awad, S., Volgenant, C., Crielaard, W., Laheij, A., & De Soet, J. (2020). Modeling of the Transmission of Coronaviruses, Measles Virus, Influenza Virus, Mycobacterium Tuberculosis, and Legionella Pneumophila in Dental Clinics. *Journal of Dental Research*, 99(10), 1192-1198.
- Zhang, J., Litvinova, M., Liang, Y., Wang, Y., Wang, W., Zhao, S., . . . Yu, H. (2020). Changes in Contact Patterns Shape the Dynamics of the Covid-19 Outbreak in China. *Science*, 368(6498), 1481-1486. <https://doi.org/doi:10.1126/science.abb8001>
- Zhang, J., Xie, B., & Hashimoto, K. (2020). Current Status of Potential Therapeutic Candidates for the Covid-19 Crisis. *Brain, Behavior, and Immunity*, 87, 59-73.
- Zhang, R., Li, D., & Jin, Z. (2015). Dynamic Analysis of a Delayed Model for Vector-Borne Diseases on Bipartite Networks. *Applied Mathematics and Computation*, 263, 342-352. <https://doi.org/https://doi.org/10.1016/j.amc.2015.04.074>
- Zhang, X., Ji, Z., Yue, Y., Liu, H., & Wang, J. (2021). Infection Risk Assessment of Covid-19 through Aerosol Transmission: A Case Study of South China Seafood Market. *Environmental Science & Technology*, 55(7), 4123-4133. <https://doi.org/10.1021/acs.est.0c02895>
- Zohuri, B., & Moghaddam, M. (2020). Deep Learning Limitations and Flaws. *Modern Approaches on Material Science*, 2, 241-250.
- Zuo, Y. Y., Uspal, W. E., & Wei, T. (2020). Airborne Transmission of Covid-19: Aerosol Dispersion, Lung Deposition, and Virus-Receptor Interactions. *ACS Nano*, 14(12), 16502-16524. <https://doi.org/10.1021/acsnano.0c08484>

APPENDICES

Appendix A: List of Tables

Table A.1: Visited date of human node with correspond location node

H node labelled (Contacts)	Date	L node visited
<i>H1, H2, H3, H4, H5, H6, H7, H8, H9, H10, H11, H12, H13, H14</i>	04-Feb, 05-Feb, 06-Feb, 07-Feb, 08-Feb, 09-Feb, 10-Feb, 11-Feb, 12-Feb-2021	<i>L1</i>
<i>H1</i>	04-Feb, 14-Feb, 15-Feb-2021	<i>L2</i>
<i>H1, H26, H27, H28, H29, H30, H31</i>	05-Feb, 12-Feb, 18-Feb, 04-Feb, 06-Feb, 11-Feb, 13-Feb-2021	<i>L3</i>
<i>H1</i>	05-Feb 08-Feb 09-Feb 10-Feb 14-Feb 15-Feb-2021	<i>L4</i>
<i>H1</i>	6-Feb, 10-Feb, 15-Feb-2021	<i>L5</i>
<i>H1</i>	06-Feb-2021	<i>L6</i>
<i>H1</i>	07-Feb, 08-Feb, 09-Feb, 10-Feb, 14-Feb, 15-Feb, 16-Feb, 17-Feb-2021	<i>L7</i>
<i>H1, H15, H16, H17, H18, H19, H20, H21</i>	14-Feb, 10-Feb, 12-Feb, 07-Feb, 09-Feb-2021	<i>L8</i>
<i>H1</i>	09-Feb, 11-Feb, 15-Feb-2021	<i>L9</i>
<i>H1</i>	09-Feb, 10-Feb, 12-Feb-2021	<i>L10</i>
<i>H1, H15, H16, H17, H18, H19, H20, H21</i>	10-Feb-2021	<i>L11</i>
<i>H1</i>	11-Feb-2021	<i>L12</i>
<i>H1</i>	11-Feb-2021	<i>L13</i>
<i>H1</i>	12-Feb-2021	<i>L14</i>
<i>H1</i>	13-Feb-2021	<i>L15</i>
<i>H1, H22, H23, H24, H25</i>	13-Feb, 17-Feb-2021	<i>L16</i>
<i>H1, H15, H16, H17, H18, H19, H20, H21</i>	16-Feb, 17-Feb-2021	<i>L17</i>
<i>H1</i>	16-Feb-2021	<i>L18</i>
<i>H1, H15, H16, H17, H18, H19, H20, H21</i>	11-Feb-2021	<i>L19</i>

<i>H1</i>	17-Feb-2021	<i>L20</i>
<i>H1</i>	18-Feb-2021	<i>L21</i>

Table A.2: Enviromental parameters raw values

Location Node	Date of Visit	Temperature (°C)	Relative Humidity (%)
<i>L1</i>	04-Feb, 05-Feb, 06-Feb, 07-Feb, 08-Feb, 09-Feb, 10-Feb, 11-Feb, 12-Feb-2021	26, 26, 27, 27, 26, 27, 27, 27, 27	88, 88, 88, 85, 89, 84, 84, 83, 85
<i>L2</i>	04-Feb, 14-Feb, 15-Feb-2021	26, 26, 26	88, 86, 81
<i>L3</i>	05-Feb, 12-Feb, 18-Feb, 04-Feb, 06-Feb, 11-Feb, 13-Feb-2021	26, 27, 26, 26, 27, 27, 26	88, 85, 88, 88, 88, 83, 88
<i>L4</i>	05-Feb 08-Feb 09-Feb 10-Feb 14-Feb 15-Feb-2021	26, 26, 27, 27, 26, 26	88, 89, 84, 84, 86, 81
<i>L5</i>	6-Feb, 10-Feb, 15-Feb-2021	27, 27, 26	88, 84, 81
<i>L6</i>	06-Feb-2021	27	88
<i>L7</i>	07-Feb, 08-Feb, 09-Feb, 10-Feb, 14-Feb, 15-Feb, 16-Feb, 17-Feb-2021	27, 26, 27, 27, 26, 26, 26, 26	85, 89, 84, 84, 86, 81, 84, 87
<i>L8</i>	14-Feb, 10-Feb, 12-Feb, 07-Feb, 09-Feb-2021	26, 27, 27, 27, 27	86, 84, 85, 85, 84
<i>L9</i>	09-Feb, 11-Feb, 15-Feb-2021	27, 27, 26	84, 83, 81
<i>L10</i>	09-Feb, 10-Feb, 12-Feb-2021	27, 27, 27	84, 84, 85
<i>L11</i>	10-Feb-2021	27	84
<i>L12</i>	11-Feb-2021	27	83
<i>L13</i>	11-Feb-2021	27	83
<i>L14</i>	12-Feb-2021	27	85
<i>L15</i>	13-Feb-2021	26	88
<i>L16</i>	13-Feb, 17-Feb-2021	26, 26	88, 87
<i>L17</i>	16-Feb, 17-Feb-2021	26, 26	84, 87
<i>L18</i>	16-Feb-2021	26	84
<i>L19</i>	11-Feb-2021	27	83
<i>L20</i>	17-Feb-2021	26	87
<i>L21</i>	18-Feb-2021	26	88

Table A.3: Calculation of SRCC for Location Nodes (Analytical Verification)

L Node	CHR	HubMat	RankCHR	RankHub	$d_{CHR_HubMatrix}$	$[d_{CHR_HubMatrix}]^2$
<i>L1</i>	1	551.19755	1	1	0	0
<i>L2</i>	0.0349	96.642721	13	14	-1	1
<i>L3</i>	0.0738	241.8286	3	7	-4	16
<i>L4</i>	0.0464	128.3036	9	9	0	0
<i>L5</i>	0.0366	101.28173	11	12	-1	1
<i>L6</i>	0.0338	93.349383	15	15	0	0
<i>L7</i>	0.0574	158.75017	8	8	0	0
<i>L8</i>	0.0966	385.00312	2	2	0	0
<i>L9</i>	0.035	96.836882	12	13	-1	1
<i>L10</i>	0.0369	102.1573	10	11	-1	1
<i>L11</i>	0.0677	272.64236	5	4	1	1
<i>L12</i>	0.0662	266.54629	7	6	1	1
<i>L13</i>	0.0289	80.014816	19	19	0	0
<i>L14</i>	0.0306	84.542506	16	16	0	0
<i>L15</i>	0.0301	83.157755	17	17	0	0
<i>L16</i>	0.0339	114.09048	14	10	4	16
<i>L17</i>	0.0719	288.67074	4	3	1	1
<i>L18</i>	0.0272	75.216271	21	21	0	0
<i>L19</i>	0.0667	268.52554	6	5	1	1
<i>L20</i>	0.0301	83.157755	18	18	0	0
<i>L21</i>	0.0289	80.014816	20	20	0	0
Sum of $[d_{CHR_HubMatrix}]^2$						40

Table A.4: Parameter Values of Nodes in Validated Model in Group 1

(a) Location Node

Location.ID	Fl	Ka	Ks
<i>L1</i>	9	0.072647	0.26665
<i>L2</i>	3	0.07079	0.221751
<i>L3</i>	7	0.072906	0.265519
<i>L4</i>	6	0.071592	0.240808
<i>L5</i>	3	0.071382	0.25565
<i>L6</i>	1	0.074714	0.329577
<i>L7</i>	8	0.071408	0.238638
<i>L8</i>	5	0.071956	0.255846
<i>L9</i>	3	0.070117	0.226249
<i>L10</i>	3	0.071931	0.26012
<i>L11</i>	1	0.071678	0.255024
<i>L12</i>	1	0.070919	0.241374
<i>L13</i>	1	0.070919	0.241374
<i>L14</i>	1	0.072437	0.270311
<i>L15</i>	1	0.073067	0.254343
<i>L16</i>	2	0.072688	0.247553
<i>L17</i>	2	0.07117	0.224144
<i>L18</i>	1	0.070031	0.207524
<i>L19</i>	1	0.070919	0.241374
<i>L20</i>	1	0.072308	0.240763
<i>L21</i>	1	0.073067	0.254343
<i>L52</i>	15	0.072107	0.251182

(b) Human Node

Location.ID	Human.ID	<i>Fh</i>	<i>D</i>	<i>Pwr</i>
<i>L1</i>	<i>H1, H10, H11, H12, H13, H14, H2, H3, H4, H5, H6, H7, H8, H9</i>	9	72	0.917298
<i>L2</i>	<i>H1</i>		1.5	0
<i>L3</i>	<i>H1, H26, H27, H28, H29, H30, H31</i>	3.5 714 29	17. 857 14	0.853689

<i>L4</i>	<i>H1</i>	6	3	0
<i>L5</i>	<i>H1</i>	3	1.5	0
<i>L6</i>	<i>H1</i>	1	0.5	0
<i>L7</i>	<i>H1</i>	8	24	0
<i>L8</i>	<i>H1, H15, H16, H17, H18, H19, H20, H21</i>	5	5	0.455517
<i>L9</i>	<i>H1</i>	3	1.5	0
<i>L10</i>	<i>H1</i>	3	1.5	0
<i>L11</i>	<i>H1, H15, H16, H17, H18, H19, H20, H21</i>	1	1	0.455517
<i>L12</i>	<i>H1</i>	1	1	0
<i>L13</i>	<i>H1</i>	1	0.5	0
<i>L14</i>	<i>H1</i>	1	0.5	0
<i>L15</i>	<i>H1</i>	1	1	0
<i>L16</i>	<i>H1, H22, H23, H24, H25</i>	2	3	0.647004
<i>L17</i>	<i>H1, H15, H16, H17, H18, H19, H20, H21</i>	2	2	0.455517
<i>L18</i>	<i>H1</i>	1	0.5	0
<i>L19</i>	<i>H1, H15, H16, H17, H18, H19, H20, H21</i>	1	1	0.455517
<i>L20</i>	<i>H1</i>	1	1	0
<i>L21</i>	<i>H1</i>	1	0.5	0
<i>L52</i>	<i>H1, H150, H151, H152, H153, H154, H155, H156, H157, H158, H159, H160</i>	15	120	0.916532

Table A.5: CCS Matrix of First Validated Model

(a) *H1* to *H11*

	<i>H1</i>	<i>H2</i>	<i>H3</i>	<i>H4</i>	<i>H5</i>	<i>H6</i>	<i>H7</i>	<i>H8</i>	<i>H9</i>	<i>H10</i>	<i>H11</i>
<i>L1</i>	3.2233	4.1125	4.1125	4.1125	4.1125	4.1125	4.1125	4.1125	4.1125	4.1125	4.1125
<i>L2</i>	1.8295	0	0	0	0	0	0	0	0	0	0
<i>L3</i>	2.7058	0	0	0	0	0	0	0	0	0	0
<i>L4</i>	2.2625	0	0	0	0	0	0	0	0	0	0
<i>L5</i>	1.9292	0	0	0	0	0	0	0	0	0	0
<i>L6</i>	1.9238	0	0	0	0	0	0	0	0	0	0
<i>L7</i>	2.6518	0	0	0	0	0	0	0	0	0	0
<i>L8</i>	2.2029	0	0	0	0	0	0	0	0	0	0

<i>L9</i>	1.8337	0	0	0	0	0	0	0	0	0	0
<i>L10</i>	1.948	0	0	0	0	0	0	0	0	0	0
<i>L11</i>	1.6873	0	0	0	0	0	0	0	0	0	0
<i>L12</i>	1.6409	0	0	0	0	0	0	0	0	0	0
<i>L13</i>	1.6372	0	0	0	0	0	0	0	0	0	0
<i>L14</i>	1.7345	0	0	0	0	0	0	0	0	0	0
<i>L15</i>	1.7022	0	0	0	0	0	0	0	0	0	0
<i>L16</i>	1.8141	0	0	0	0	0	0	0	0	0	0
<i>L17</i>	1.7244	0	0	0	0	0	0	0	0	0	0
<i>L18</i>	1.534	0	0	0	0	0	0	0	0	0	0
<i>L19</i>	1.6409	0	0	0	0	0	0	0	0	0	0
<i>L20</i>	1.656	0	0	0	0	0	0	0	0	0	0
<i>L21</i>	1.6985	0	0	0	0	0	0	0	0	0	0
<i>L52</i>	4.2545	0	0	0	0	0	0	0	0	0	0

(b) *H12 to H22*

	<i>H12</i>	<i>H13</i>	<i>H14</i>	<i>H15</i>	<i>H16</i>	<i>H17</i>	<i>H18</i>	<i>H19</i>	<i>H20</i>	<i>H21</i>	<i>H22</i>
<i>L1</i>	4.1125	4.1125	4.1125	0	0	0	0	0	0	0	0
<i>L2</i>	0	0	0	0	0	0	0	0	0	0	0
<i>L3</i>	0	0	0	0	0	0	0	0	0	0	0
<i>L4</i>	0	0	0	0	0	0	0	0	0	0	0
<i>L5</i>	0	0	0	0	0	0	0	0	0	0	0
<i>L6</i>	0	0	0	0	0	0	0	0	0	0	0
<i>L7</i>	0	0	0	0	0	0	0	0	0	0	0
<i>L8</i>	0	0	0	2.6715	2.6715	2.6715	2.6715	2.6715	2.6715	2.6715	0
<i>L9</i>	0	0	0	0	0	0	0	0	0	0	0
<i>L10</i>	0	0	0	0	0	0	0	0	0	0	0
<i>L11</i>	0	0	0	2.1559	2.1559	2.1559	2.1559	2.1559	2.1559	2.1559	0
<i>L12</i>	0	0	0	0	0	0	0	0	0	0	0
<i>L13</i>	0	0	0	0	0	0	0	0	0	0	0
<i>L14</i>	0	0	0	0	0	0	0	0	0	0	0
<i>L15</i>	0	0	0	0	0	0	0	0	0	0	0
<i>L16</i>	0	0	0	0	0	0	0	0	0	0	2.5421
<i>L17</i>	0	0	0	2.193	2.193	2.193	2.193	2.193	2.193	2.193	0
<i>L18</i>	0	0	0	0	0	0	0	0	0	0	0
<i>L19</i>	0	0	0	2.1095	2.1095	2.1095	2.1095	2.1095	2.1095	2.1095	0
<i>L20</i>	0	0	0	0	0	0	0	0	0	0	0
<i>L21</i>	0	0	0	0	0	0	0	0	0	0	0
<i>L52</i>	0	0	0	0	0	0	0	0	0	0	0

(c) *H23 to H151*

	H23	H24	H25	H26	H27	H28	H29	H30	H31	H150	H151
<i>L1</i>	0	0	0	0	0	0	0	0	0	0	0
<i>L2</i>	0	0	0	0	0	0	0	0	0	0	0
<i>L3</i>	0	0	0	3.2123	3.2123	3.2123	3.2123	3.2123	3.2123	0	0
<i>L4</i>	0	0	0	0	0	0	0	0	0	0	0
<i>L5</i>	0	0	0	0	0	0	0	0	0	0	0
<i>L6</i>	0	0	0	0	0	0	0	0	0	0	0
<i>L7</i>	0	0	0	0	0	0	0	0	0	0	0
<i>L8</i>	0	0	0	0	0	0	0	0	0	0	0
<i>L9</i>	0	0	0	0	0	0	0	0	0	0	0
<i>L10</i>	0	0	0	0	0	0	0	0	0	0	0
<i>L11</i>	0	0	0	0	0	0	0	0	0	0	0
<i>L12</i>	0	0	0	0	0	0	0	0	0	0	0
<i>L13</i>	0	0	0	0	0	0	0	0	0	0	0
<i>L14</i>	0	0	0	0	0	0	0	0	0	0	0
<i>L15</i>	0	0	0	0	0	0	0	0	0	0	0
<i>L16</i>	2.5421	2.5421	2.5421	0	0	0	0	0	0	0	0
<i>L17</i>	0	0	0	0	0	0	0	0	0	0	0
<i>L18</i>	0	0	0	0	0	0	0	0	0	0	0
<i>L19</i>	0	0	0	0	0	0	0	0	0	0	0
<i>L20</i>	0	0	0	0	0	0	0	0	0	0	0
<i>L21</i>	0	0	0	0	0	0	0	0	0	0	0
<i>L52</i>	0	0	0	0	0	0	0	0	0	5.1545	5.1545

(d) *H152 to H160*

	<i>H152</i>	<i>H153</i>	<i>H154</i>	<i>H155</i>	<i>H156</i>	<i>H157</i>	<i>H158</i>	<i>H159</i>	<i>H160</i>
<i>L1</i>	0	0	0	0	0	0	0	0	0
<i>L2</i>	0	0	0	0	0	0	0	0	0
<i>L3</i>	0	0	0	0	0	0	0	0	0
<i>L4</i>	0	0	0	0	0	0	0	0	0
<i>L5</i>	0	0	0	0	0	0	0	0	0
<i>L6</i>	0	0	0	0	0	0	0	0	0
<i>L7</i>	0	0	0	0	0	0	0	0	0
<i>L8</i>	0	0	0	0	0	0	0	0	0
<i>L9</i>	0	0	0	0	0	0	0	0	0
<i>L10</i>	0	0	0	0	0	0	0	0	0
<i>L11</i>	0	0	0	0	0	0	0	0	0
<i>L12</i>	0	0	0	0	0	0	0	0	0
<i>L13</i>	0	0	0	0	0	0	0	0	0
<i>L14</i>	0	0	0	0	0	0	0	0	0
<i>L15</i>	0	0	0	0	0	0	0	0	0

<i>L16</i>	0	0	0	0	0	0	0	0	0
<i>L17</i>	0	0	0	0	0	0	0	0	0
<i>L18</i>	0	0	0	0	0	0	0	0	0
<i>L19</i>	0	0	0	0	0	0	0	0	0
<i>L20</i>	0	0	0	0	0	0	0	0	0
<i>L21</i>	0	0	0	0	0	0	0	0	0
<i>L52</i>	5.1545	5.1545	5.1545	5.1545	5.1545	5.1545	5.1545	5.1545	5.1545

Table A.6: CCS Matrix of Targeted and Validated Model in Group 2

(a) Targeted Model

	<i>H32</i>	<i>H33</i>	<i>H34</i>	<i>H35</i>	<i>H36</i>	<i>H37</i>	<i>H38</i>	<i>H39</i>	<i>H40</i>	<i>H42</i>	<i>H43</i>
<i>L34</i>	0.3606	0	0	0	0	0	0	1.2606	1.2606	0	0
<i>L35</i>	4.5	5.4	5.4	5.4	5.4	5.4	5.4	5.4	5.4	0	0
<i>L37</i>	0	0	0	0	0	0	0	0	2.3274	0	0
<i>L38</i>	0	0	0	0	0	0	0	0	1.3529	1.3529	1.3529
<i>L39</i>	0	0	0	0	0	0	0	0	0.5935	0	0
<i>L40</i>	0	0	0	0	0	0	0	0	0.945	0	0

(b) Validated Model

	<i>H32</i>	<i>H33</i>	<i>H34</i>	<i>H35</i>	<i>H36</i>	<i>H37</i>	<i>H38</i>	<i>H39</i>	<i>H40</i>	<i>H41</i>
<i>L34</i>	0.5579	1.4579	1.4579	1.4579	1.4579	1.4579	1.4579	1.4579	1.4579	0
<i>L35</i>	4.5	5.4	5.4	5.4	5.4	5.4	5.4	5.4	5.4	0
<i>L36</i>	3.6023	0	0	0	0	0	0	0	0	4.4997

Table A.7: CCS Matrix of Targeted Model in Group 3

(a) *H114* to *H123*

	<i>H114</i>	<i>H115</i>	<i>H116</i>	<i>H117</i>	<i>H118</i>	<i>H119</i>	<i>H120</i>	<i>H121</i>	<i>H122</i>	<i>H123</i>
<i>L1</i>	3.78	0	0	0	0	0	0	0	0	0
<i>L51</i>	4.5	5.4	5.4	5.4	5.4	5.4	5.4	5.4	5.4	5.4

(b) *H124* to *H133*

	<i>H124</i>	<i>H125</i>	<i>H126</i>	<i>H127</i>	<i>H128</i>	<i>H129</i>	<i>H130</i>	<i>H131</i>	<i>H132</i>	<i>H133</i>
<i>L1</i>	0	0	0	0	0	0	0	0	0	0
<i>L51</i>	5.4	5.4	5.4	5.4	5.4	5.4	5.4	5.4	5.4	5.4

(c) *H134 to H143*

	<i>H134</i>	<i>H135</i>	<i>H136</i>	<i>H137</i>	<i>H138</i>	<i>H139</i>	<i>H140</i>	<i>H141</i>	<i>H142</i>	<i>H143</i>
<i>L1</i>	0	0	0	0	0	0	0	0	0	0
<i>L51</i>	5.4	5.4	5.4	5.4	5.4	5.4	5.4	5.4	5.4	5.4

(d) *H144 to H149*

	<i>H144</i>	<i>H145</i>	<i>H146</i>	<i>H147</i>	<i>H148</i>	<i>H149</i>
<i>L1</i>	0	0	0	0	0	0
<i>L51</i>	5.4	5.4	5.4	5.4	5.4	5.4

Table A.8: Ranking of Location and Human Node of the Targeted and Validated Model
in Group 2

(a) Ranking of Location Node (Targeted Model)

Ranking	Location ID	CHR
1	<i>L35</i>	1
2	<i>L34</i>	0.061364
3	<i>L37</i>	0.051617
4	<i>L38</i>	0.030442
5	<i>L40</i>	0.020958
6	<i>L39</i>	0.013163

(b) Ranking of Human Node (Targeted Model)

Ranking	Human ID	CRR
1	<i>H40</i>	1
2	<i>H39</i>	0.966656
3	<i>H33</i>	0.953005
3	<i>H34</i>	0.953005
3	<i>H35</i>	0.953005
3	<i>H36</i>	0.953005

3	<i>H37</i>	0.953005
3	<i>H38</i>	0.953005
4	<i>H32</i>	0.798075
5	<i>H42</i>	0.007269
5	<i>H43</i>	0.007269

(c) Ranking of Location Node (Validated Model)

Ranking	Location ID	CHR
1	<i>L35</i>	1
2	<i>L34</i>	0.258143
3	<i>L36</i>	0.070186

(d) Ranking of Human Node (Validated Model)

Ranking	Human ID	CRR
1	<i>H33</i>	1
1	<i>H34</i>	1
1	<i>H35</i>	1
1	<i>H36</i>	1
1	<i>H37</i>	1
1	<i>H38</i>	1
1	<i>H39</i>	1
1	<i>H40</i>	1
2	<i>H32</i>	0.847741
3	<i>H41</i>	0.054675

Appendix B: Supplementary Information

Definition of direct and indirect transmission modes

Direct transmission involves direct contact and droplet spread, whereas indirect transmission involves airborne, fomite and vector-borne spread. The common transmission modes of respiratory diseases (including COVID-19) are droplet, airborne and fomite (Hertzberg et al., 2018; Kutter et al., 2018; Peng et al., 2020). Droplet transmission can happen within 1 m from where the respiratory droplets (droplet particles with a diameter between 5 and 10 μm) were expelled by the infected individuals (WHO, 2014). These droplets can be expelled when speaking, coughing and sneezing and can then enter the mucous membranes of susceptible individuals. Some of these droplets land on objects' surfaces and remain viable for more than a day (Chin et al., 2020). Consequently, susceptible individuals can possibly touch the contaminated surfaces—an example of indirect transmission (also called as fomite transmission) (Fernstrom & Goldblatt, 2013). Several studies had collected samples from public places and claim that the active virus remains on the surface from an hour to several days (Guo et al., 2020; Santarpia et al., 2020; van Doremalen et al., 2020). According to the existing studies mentioned above, the stability of the pathogen on the object's surface is affected by the temperature, relative humidity and types of the surfaces.

Particles with a diameter $<5 \mu\text{m}$ are called droplet nuclei. These droplet nuclei are able to linger in the air for a longer duration and across longer distances ($>1 \text{ m}$) than respiratory droplets and cause airborne transmission to susceptible individuals (Kwon et al., 2020; Lu et al., 2020; The Lancet Respiratory, 2020). Airborne transmission is an indirect transmission mode and can spread via aerosol form. The term aerosol commonly refers to particles in the air/gas that are small enough to remain airborne for an extended duration due

to their low settling velocity (Tellier, 2006). These aerosols are produced when talking, sneezing and coughing (Nicas et al., 2005). Consequently, if such actions were displayed by an infected person, the aerosols can carry the virus (Lindsley et al., 2015). Thus, these infected aerosols have an active pathogen and remain viable within a certain period. The stability of the pathogen in aerosol form is affected by several factors, for instance, temperature, relative humidity and sunlight (Jones & Brosseau, 2015; Fears et al., 2020; van Doremalen et al., 2020).

Applications of Well-Riley equation

Dai and Zhao (2020) extended the Wells–Riley equation by integrating the effect when a COVID-19-affected individual wear a mask. They introduced two new parameters, namely, η_I and η_s , which represent exhalation filtration efficiency and respiration filtration efficiency, respectively. These two parameters are used to observe the efficiency of wearing a mask as the study assumes that an infected individual wearing a mask can dilute the concentration of the pathogen (Dai & Zhao, 2020). Equation B.1 shows the modified Wells–Riley equation after the introduction of the two new parameters:

$$P = 1 - e^{\left(-\frac{Iqpt(1-\eta_I)(1-\eta_s)}{Q}\right)} \quad \text{Equation B.1}$$

As opposed to those assumptions, the dose-response model had been proposed by other researchers to access the risk of infection (Nicas, 1996; Nicas et al., 2005; Teunis et al., 2008). Dose-response model integrates the infectious dose data in the model where the dose refers to the quantity of the pathogen. As stated by Gin Nam Sze To and Christopher Yu Hang Chao (2010), the dose-response model is able to take other disease transmission routes into account to identify the strength of the infectious source in the indoor building. Based on the review made by Gin Nam Sze To and Christopher Yu Hang Chao (2010), the

earliest study using the dose-response model was done by Nicas in 1996. Nicas (1996) used a dose-response model as shown in Equation B.2 to identify the risk of infection for tuberculosis.

$$P = 1 - e^{\left(-\frac{IGp\beta t}{Q}\right)} \quad \text{Equation B.2}$$

where β is the deposition fraction of infectious particles in the alveolar region and G is the number of airborne tuberculosis bacilli released per infector per unit time. While the definitions of I , p , t and Q are identical to that of Equation 2.1, infectious dose data are required to employ a dose-response model that can recognise the infection risk.