



Faculty of Computer Science and Information Technology

A Computational Approach to Predict the Spread of Dengue

Kok Woon Chee

**Master of Science
2018**

A Computational Approach to Predict the Spread of Dengue

Kok Woon Chee

A thesis submitted
In fulfilment of the requirements for the degree of Master of Science
(Computational Science)

Faculty of Computer Science and Information Technology
UNIVERSITI MALAYSIA SARAWAK
2018

DECLARATION

I hereby declare that the thesis is based on my original work except for quotations and citation, which have been duly acknowledged. The thesis has not been accepted for any degree and is not concurrently submitted in the candidature of any other degree.

Name of the student: KOK WOON CHEE

Date:

ACKNOWLEDGEMENT

When you meet people, show real appreciation, then genuine curiosity –Martha Beck

I would like to thank my main supervisor, Associate 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. I am thankful for the invaluable experiences I gained from her through conducting the research writing of scientific papers, preparing presentations and attending conferences. Likewise, I thank my co-supervisor, Professor Dr David Perera for his expertise in providing valuable suggestion with regards to the respective aspects of my research.

Special thanks are also extended to Universiti Malaysia Sarawak for the Zamalah Vice Chancellor Research Award which has provided me with full scholarship and 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 the Dean and Deputy Deans for the wholehearted support. Thanks to all lecturers especially lecturers from Computational Science and Mathematics Department for your insightful advices. I also acknowledge the administrative staff members of the faculty and the Centre for Graduate Studies for their constant help. Also, special thanks to the faculty for entrusting me as the Honorary Treasurer for the FCSIT Postgraduate Society for the term of October 2015 to October 2016 of which I gained valuable experiences through many activities. Thank you to Institut de Recherche pour le Développement (IRD) for providing me a full grant to attend the summer school student exchange program in Vietnam. This invaluable experience helped me in sharpening my research and programming skills.

I would also like to also deliver my sincere appreciation to the Sarawak State Health Department and National Institute of Health for the help and support for providing the research data generously. Besides that, special thanks to Sarawak Meteorological Office for providing the relevant data. Thank you.

Thanks to my thoughtful lecturer, Mr Chuah Kee Man who lends a helping hand in proof-reading the draft of this thesis.

Also, thanks to my friends and research mates for spending their precious time in giving me their supports throughout the research. Yi Ming, thank you for your generosity in borrowing me your MacBook. Thanks for your friendship, warm advices and great support. Chia Lih, Chin Ying, Cynthia, Faisal, Faizol, Hui Sun, Ima, Jessica, Marcella, Nisa, Shian Li, Siew Ming, Sue, Wendy, Yee Chen 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 when I am far away from my family. Thanks to all my friends who greet me happily and give me support and encouragement. I obviously 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, Yeo Peng and Sam Mooi for raising me with a love of knowledge and lovingly encouraged me to strive for my master dream. Also, thanks to my brother, Wen Howe and both my sisters, Woon Jun and Woon Shi for their endless support in motivating me. My family contributed significantly to the accomplishment of my master degree. Thank you.

ABSTRACT

Dengue disease poses a large economic burden in Malaysia among other endemic countries. Malaysia had not adopted the dengue vaccine yet. In the absence of dengue vaccine, it is important to identify dengue hotspot for eradication. Several methods for example compartmental modeling, agent-based modelling and other scientific methods have been used to detect hotspot. Network modelling approach, in particular a bipartite network is currently the most appropriate method to solve the research problem. A contact network model of dengue transmission was formulated. Two distinct entities, human and location nodes are always interconnected. The relationship between these two nodes is quantified by using a summation rule. When the resources such as budget and manpower are limited, control strategies are prioritised according to the risk area of higher vector density. Dengue Hotspot Ranking (DHR) values are generated for location nodes in a network model to capture the human heterogeneity, environmental properties and the vector characteristic. The location nodes are then ranked according to DHR value which is also a measurement of vector density. In previous study, researcher attempted to model malaria hotspot by using the same method, but there is no adequate model validation with real data performed. Verification analysis reports a root mean square error (RMSE) of 6.419×10^{-4} and 9.188×10^{-4} for the location and human nodes respectively, which is much lower than the threshold value of 0.05. Validation analysis revealed a strong ranking similarity with good Spearman rank correlation coefficient of above 0.80 ($\rho > 0.80$; $p < 0.001$) which is greater than the threshold value set at 0.70. Through parameter significance analysis, this study identified significant parameters to detect possible dengue hotspot which include mosquito's life cycle, survival and biting rate, precipitation, humidity, altitude the number of times human visited a location, duration of human stay at one location

and the frequency of a human visited one location. This study also developed a polynomial model to represent the life cycle and survival parameters by using polynomial curve fitting due to the fact that there is no primary data available from any data source. Consequently, the bipartite network modelling approach has been successfully formulated and the ranked locations are believed to be applicable to help public health authorities to prioritise the locations for vector control. Eradication of dengue in the risk areas can help in the reduction of the spread of dengue disease.

Keywords: bipartite network modeling, dengue, hotspot, ranking

Kaedah Komputasi untuk Ramalan Penyebaran Denggi

ABSTRAK

Denggi menimbulkan bebanan ekonomi yang besar kepada Malaysia dan juga negara-negara endemik yang lain. Malaysia belum menggunakan vaksin denggi. Memang tidak dapat dinafikan bahawa pengenalpastian kawasan pembiakan vektor nyamuk sangat penting dalam ketiadaan vaksin denggi. Beberapa kaedah seperti permodelan kompartan, permodelan berasaskan ejen dan kaedah saintifik yang lain telah digunakan untuk mengesan hotspot. Kaedah permodelan rangkaian, terutamanya rangkaian bipartite adalah kaedah yang paling sesuai dalam mengatasi soalan penyelidikan ini. Satu model rangkaian hubungan telah dirumuskan. Dua entiti yang berlainan iaitu manusia dan lokasi nod saling berhubungkait. Hubungan antara kedua-dua nod diukur dengan petua penjumlahan. Apabila sumber-sumber seperti bajet dan tenaga kerja terhad, strategi kawalan perlu diutamakan mengikut kawasan risiko yang mempunyai kepadatan vektor yang lebih tinggi. Nilai Kedudukan Hotspot Denggi (DHR) dihasilkan untuk nod lokasi dalam model rangkaian untuk heterogeniti manusia, sifat persekitaran dan sifat vektor. Nod lokasi kemudiannya disusun mengikut nilai DHR yang juga pengukuran kepadatan vektor. Dalam kajian sebelum ini, penyelidik cuba mengesan hotspot malaria dengan kaedah yang sama, tetapi penyelidik tidak melakukan pengesahan model yang cukup dengan penggunaan data sebenar. Analisis pengesahan melaporkan kesilapan purata (RMSE) iaitu 0.0006419 dan 0.0009188 untuk lokasi dan nod manusia, masing-masingnya, yang lebih rendah daripada 0.05. Analisis pengesahan melaporkan persamaan kedudukan yang kuat dengan pekali korelasi pangkat Spearman yang baik iaitu 0.80 ke atas ($p > 0.80$; $p < 0.001$), keputusan ini lebih tinggi daripada nilai 0.70 iaitu nilai yang ditetapkan. Melalui analisis kepentingan parameter, analisis ini mengenal pasti parameter yang penting untuk

mengesan hotspot denggi yang mungkin termasuk kitaran hidup nyamuk, ketahanan hidup nyamuk dan kadar menggigit, pemendakan, kelembapan, bilangan manusia yang melawat lokasi, tempoh tinggal manusia di satu lokasi dan kekerapan manusia melawat sesuatu lokasi. Kajian ini juga merumuskan model polinomial untuk kitaran hidup nyamuk dan ketahanan hidup nyamuk kerana tiada data langsung telah direkodkan dari mana-mana sumber data. Oleh itu, kaedah pendekatan pemodelan rangkaian bipartite telah berjaya diformulasikan dan lokasi yang telah disusun mengikut nilai DHR diyakini boleh digunakan untuk membantu pihak berkuasa kesihatan awam untuk mengutamakan lokasi bagi mengawal vektor. Pemberantasan denggi di kawasan risiko dapat membantu dalam pengurangan penyebaran penyakit denggi.

Kata kunci: permodelan rangkaian bipartite, denggi, 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	7
1.4 Research Objectives	8
1.5 Research Scope	8
1.6 Research Methodology	9
1.7 Research Significance	13
1.8 Thesis Outline	14
CHAPTER 2: LITERATURE REVIEW	16
2.1 Introduction	16

2.2	Hotspot Detection Studies in Disease Aspect.....	16
2.2.1	Main Factors of Hotspot Detection Studies	17
2.2.2	Approaches of Hotspot Detection Studies	22
2.3	Network Modeling Approach.....	30
2.3.1	From Graph Theory to Network Theory	30
2.3.2	Network Modeling Approaches and Spatial Studies.....	31
2.4	Reviews on Different Web-based Search Algorithms.....	35
2.5	Summary.....	39
CHAPTER 3: RESEARCH PROBLEM CHARACTERISATION		40
3.1	Introduction	40
3.2	Current Research Scenario	40
3.2.1	Graph Structure Representation	41
3.2.2	Assumptions	41
3.2.3	Epidemiological Triangle	41
3.2.4	Basic Building Block of the Bipartite Dengue Contact Network Model	43
3.3	Research Data and Potential Parameters	45
3.3.1	Raw Data	45
3.3.2	Potential Parameters Used.....	47
3.4	Bipartite Graph Formulation	48
3.4.1	Identification of the BDC Network.....	48
3.4.2	Identification of Human Node.....	50

3.4.3	Identification of Location Node	51
3.4.4	Identification of BDC Network Links between Nodes	51
3.4.5	Complete Bipartite Dengue Contact Graph for BDC Network.....	52
3.5	Summary.....	54
CHAPTER 4: MODEL CONSTRUCTION		55
4.1	Introduction	55
4.2	Data Pre-processing.....	55
4.2.1	Pre-processing of Location Node	55
4.2.2	Pre-Process of Human Node	61
4.2.3	Pre-Process of Parameter	61
4.3	Quantification of Location Node Parameter.....	63
4.3.1	Quantification of Vector Life Cycle Duration Parameter, Lc	65
4.3.2	Quantification of Vector Survival Parameter, S	69
4.3.3	Quantification of Vector Biting Parameter, B	72
4.3.4	Quantification of Number of Times a Location is visited by Human, Fl	74
4.4	Quantification of Human Node Parameters.....	78
4.4.1	Quantification of Duration time the human stay at one location, Du	78
4.4.2	Quantification of Number of times a human visited a location, Fh	79
4.5	Quantification of Link Weight	81
4.6	HITS Search Algorithm.....	86
4.6.1	Generation of Hub and Authority Matrix.....	87

4.6.2	Generation of Hub and Authority Principal Eigenvector	88
4.6.3	Assignment of Nodes' Label and Dengue Hotspot Ranking Values	90
4.6.4	Ranking of Bipartite Nodes in BDC Network	91
4.7	Summary.....	93
CHAPTER 5: MODEL ANALYSIS AND EVALUATION		94
5.1	Introduction	94
5.2	Model Verification	94
5.2.1	Benchmark Verification	95
5.2.2	Analytical Verification.....	100
5.3	Model Validation.....	101
5.4	Parameter Significance Analysis	108
5.5	Predictive Power Analysis.....	112
5.6	Data Size Analysis.....	115
5.7	Summary.....	118
CHAPTER 6: CONCLUSION		120
6.1	Introduction	120
6.2	Conclusion of Research	120
6.3	Research Contributions.....	122
6.4	Future Work.....	124
REFERENCES		126
APPENDICES		141

LIST OF TABLES

	Page
Table 2.1 Summary of Environmental predictors in Hotspot Detection Studies.....	18
Table 2.2 Summary of Parameter Involved in Hotspot Detection Studies	19
Table 2.3 Related Disease Studies Applying Contact Network Modeling	34
Table 2.4 Related Disease Studies Applying Bipartite Network Modeling Approach	35
Table 2.5 Comparison of Web-based Search Algorithms	38
Table 3.1 Four Sub-datasets of Research Data	46
Table 3.2 Potential Parameters Used	48
Table 3.3 Detailed Information of Six BDC Network Models.....	49
Table 3.4 Movement of Each Human Host	52
Table 4.1 Distance Matrix between all Location Nodes in BDC Network	60
Table 4.2 Source data of Mean Development Duration of <i>Ae. Aegypti</i>	66
Table 4.3 RMSE Values for 7 Degrees in Quantifying Life Cycle Parameter	68
Table 4.4 Life Cycle Index, <i>Lc</i> of Each Location Nodes	69
Table 4.5 Source Data of Survival (%) of <i>Ae. Aegypti</i>	70
Table 4.6 RMSE Values for 6 Degrees in Quantifying Survival Index	71
Table 4.7 Survival Index, <i>S</i> of Each Location Node	72
Table 4.8 Biting Index, <i>B</i> of Each Location Nodes.....	74
Table 4.9 Link Matrix of BDC Network	75
Table 4.10 Parameter Values of 19 Location Nodes in BDC Network	77
Table 4.11 Duration Time the Human Stay at One Location, <i>Duij</i>	79
Table 4.12 Parameter Values of <i>Fhij</i>	80
Table 4.13 Complete Normalised Parameter of Location Nodes (First BDC network).....	83

Table 4.14	Dengue Contact Strength (DCS) Matrix of BDC Network	84
Table 4.15	Authority Matrix.....	88
Table 4.16	Hub Matrix.....	89
Table 4.17	Principal Eigenvector of the (a) Location Node and (b) Human Node	91
Table 4.18	The Ranking of (a) Location Nodes and (b) Human Node	92
Table 5.1	RMSE Analysis of Location Nodes.....	98
Table 5.2	RMSE Analysis of Human Nodes	99
Table 5.3	SRCC Indicators	100
Table 5.4	BDC Network Model Validation.....	102
Table 5.5	Dengue Contact Strength (DCS) Matrix of the First Validated Dataset.....	104
Table 5.6	The Ranking of Location Nodes (First Validated Model).....	105
Table 5.7	SRCC Calculation of Location Nodes between the Models in Group 1	106
Table 5.8	SRCC Calculation of Location Nodes between the Models in Group 2	107
Table 5.9	SRCC Calculation of Location Nodes between the Models in Group 3	108
Table 5.10	SRCC for “Leave-one-out” Parameter Analysis	110
Table 5.11	SRCC for “Leave-multiple-out” Parameter Analysis.....	110
Table 5.12	Result of Prediction	114
Table 5.13	Distance Matrix of Location Node versus the Real Data	115
Table 5.14	Summary of Characteristics of the New BDC Network.....	117
Table 5.15	SRCC for Comparison of BDC Networks’ Location Ranking.....	117

LIST OF FIGURES

	Page
Figure 1.1 Bipartite Network Modeling Research Methodology Framework (BNM-RMF).	10
Figure 2.1 Three Main Components of Dengue Location Detection Studies Structure	22
Figure 2.2 Complexity of the epidemiological models	32
Figure 3.1 Epidemiological Triangle (ET).....	42
Figure 3.2 Modified Epidemiological Triangle.....	43
Figure 3.3 Basic Building Block of BDC Network	44
Figure 3.4 An example of BDC network	45
Figure 3.5 Complete Bipartite Dengue Contact Graph for Dengue Patient.....	53
Figure 4.1 Pre-processing of the Location Node.....	56
Figure 4.2 Declare GPS Coordinates	57
Figure 4.3 Timeline for Human Node 1	63
Figure 4.4 Parameters of BDC Network Location Node at the Process 2.2 in BNM-RMF ..	64
Figure 4.5 Generation of Life Cycle Index, Lc	67
Figure 4.6 Quantifying the Link Weights	82
Figure 4.7 Bipartite Dengue Contact (BDC) Network.....	85
Figure 4.8 Process of the Adapted HITS Search Algorithm	87
Figure 5.1 Benchmark Verification Process	95
Figure 5.2 Benchmark Ranking and BDC Network Model Ranking	97

LIST OF ABBREVIATIONS

ABM	Agent-based Modeling
BDC	Bipartite Dengue Contact
BNM	Bipartite Network Modeling
BNM-RMF	Bipartite Network Modeling Research Methodology Framework
CDC	Centers of Disease Control and Prevention
DCS	Dengue Contact Strength
DHR	Dengue Hotspot Ranking
DENV	Dengue Virus
DT	Decision Tree
EIP	Extrinsic Incubation Period
ET	Epidemiological Triangle
GIS	Geographical Information System
GPS	Global Positioning System
HITS	Hypertext Indexed Topic Selection
ID	Identification
KNN	K-nearest Neighbor
MaxEnt	Maximum Entropy
PCA	Principal Component Analysis
PDE	Partial Differential Equations
RMSE	Root-Mean-Square-Error

SARS	Severe Acute Respiratory Syndrome (SARS)
SRCC	Spearman's Rank Correlation Coefficient
SVM	Support Vector Machine
VC	Vectorial Capacity
WHO	World Health Organisation

CHAPTER 1

INTRODUCTION

1.1 Introduction

Hotspot detection of mosquito-borne diseases for example dengue, malaria and zika is pivotal in ensuring eradication (Aziz et al., 2014). In this thesis, hotspot is defined as the prime location of mosquito breeding site. Public health authority targets the hotspots to eliminate the vector effectively (Nagao et al., 2003; Ritchie & Johnson, 2017). However, the use of a conventional modeling approach such as compartmental modeling may not be suitable due to the non-linearity of the compartmental model when spatial components are incorporated (Garnett, 2002; Nelson & Perelson, 2002). Therefore, this research attempted to identify and employ a suitable method to detect the possible location with a high density of infected mosquitoes which is referred as a dengue hotspot.

The graph-theoretic network modeling approach that was reported as a heuristic framework has the possibility to be a potential approach (Bunn, Urban & Keitt, 2000 in Liew (2016, p.1)). Heuristic is defined as any approach to problem solving that employs practical methods which do not guarantee an optimal solution but can usually generate solutions that are good enough for practical use (Wahid & Hussin, 2016). This motivates the present study to employ the network modeling approach in tackling our research problem. The domain problem needs to be properly constructed into network models where the nodes and edges are clearly defined for location detection. The network structure is transformed into an appropriate format in order to use it as an input into the network model. The model is run, validated and a number of analytical results were generated. It is potential to identify the location that encourages the

breeding of dengue vector and thus rank the location according to the estimated vector density. Such findings may serve as important insights for the authorities in control interventions.

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

1.2 Background of the study

Dengue is the world's fastest growing mosquito-borne disease, with a 30-fold increase in disease incidence over the last 50 years (Centers for Disease Control and Prevention [CDC], 2014). As reported by CDC (2014), about 2.5 billion people (or 40% of the world population) live in parts of the world where they are at risk of dengue infection. The total annual economic cost of dengue for eight countries in the Americas and Asia reached I\$ (international dollars) 587 million (Suaya et al., 2009). Dengue poses a large economic burden in endemic countries to date (Suaya et al., 2009; Viennet, Ritchie, Williams, Faddy, & Harley, 2016). Authorities attempt to reduce the disease burden using the population-level control methods. Despite such effort, these methods are limited and ineffective. To date, only one licensed dengue vaccine, Dengvaxia, provides moderate protection. Unfortunately, this vaccine is unable to prevent the spread of other mosquito-borne diseases such as zika, malaria and chikungunya viruses. Therefore, improvements in innovative approaches to prevent dengue outbreaks should continue. Furthermore, Malaysia is not adopting the vaccine yet. In the absence of vaccine, the main control strategy to eradicate dengue is to kill the vector mosquitoes in Malaysia (Lam, 1994; Packierisamy et al., 2015). Thus, it is important to identify the area where it is likely a breeding site for *Aedes* mosquitoes in order to eradicate the dengue instead of using a passive approach for instance traditional dengue surveillance. The passive approach practice is to remove the breeding sites after the dengue outbreak occurs at one particular location.

Typical studies of dengue transmission dynamics aim to examine the relationship between the dengue occurrences, distribution of the susceptible and infected at different locations, with the environmental properties (Johansson, Dominici, & Glass, 2009; Rueda, Patel, Axtell, & Stinner, 1990; Sarfraz et al., 2012; Suaya et al., 2009; Tran et al., 2013; Wijayanti et al., 2016). They mainly used compartmental approach combining the epidemiological data, demographic data and environmental predictors such as temperature and precipitation to predict the dengue outbreak (Jeefoo, Tripathi, & Souris, 2011; Rueda et al., 1990; Suaya et al., 2009; Wijayanti et al., 2016; Wu et al., 2009). Researchers used conventional modeling approaches to understand the long-term dengue trends in general include among others (Esteva & Vargas, 2003; Esteva & Yang, 2005; Focks DA, 1998; Syafruddin & Noorani, 2012). A conventional approach, compartmental modeling, employs a system of differential equations and divides the whole population into several sub-populations based on the assumption that sub-populations are homogenous and well-mixed (Hethcote, 2000). All individuals are assumed to have equal probability to have contact within the sub-populations. This assumption is simple but is crucial to formulate a simple model to analyse the disease dynamics. In reality, human contacts relate to different factors which involve age, social and travel patterns. The compartmental modeling is widely used to estimate the number of dengue cases or to study the abundance of a sub-populations. However, the compartmental model has several shortcomings when spatial elements are incorporated. The compartmental model cannot capture the heterogeneity of human interactions which serves as a basis for infectious disease transmission. Dimitrov and Meyers (2010) reported that the compartmental modeling has a lack in the support to detect locations of possibly high density of infected mosquitoes. Many attempted to combine population dynamics with the aspect of location detection by forming partial differential equations (PDE). The model eventually turned into more complicated models such as dispersal-

reaction models and dispersal-reaction-advection models that range from one- to two- and even three-dimensions of linear or non-linear PDE (Garnett, 2002; Kon & Labadin, 2013; Nelson & Perelson, 2002). Most of the time, these complicated models are hard to be solved because the crucial parameter values are not available (Garnett, 2002). On top of that, it is time consuming and computationally expensive to solve non-linear partial differential equations covering large problem domain (Cheng, Golberg, Kansa, & Zang, 2003; Crandall, Ishii, & Lions, 1992; Fehlberg, 1969; Jang, Chen, & Liu, 2001; Sincovec & Madsen, 1975).

In more recent studies, network modeling approaches offered endless possibilities in location detection of infectious diseases (Eze, Labadin, & Lim, 2011b; Malik, Mahesar, Abid, 2014; Salathé et al., 2010). This approach can depict the complexity of the real world problem (Craft & Caillaud, 2011). The basis of the approach is network theory, which is a subset of graph theory (Harary, 1969). Eze (2013) introduced a Bipartite Network Modeling (BNM) approach with a web-based graph search algorithm to identify Malaria hotspots. There are many similarities between both malaria and dengue diseases in terms of the clinical and biological metrics, as they are both mosquito-borne diseases (Selvaretnam, Sahu, Sahu, & Ambu, 2016; Wiwanitkit, 2011; Yong & Koh, 2013). Therefore we hypothesise that BNM can also be used to locate potential dengue hotspots. The BNM approach is used to identify species habitat suitability by Liew (2016). Irrawaddy dolphin has been used as a case study with an attempt to locate the preferred habitat of the dolphins (Liew, Labadin, Chai, Tuen, & Peter, 2014). On top of that, the suitability index of the preferred habitat has been quantified at the end of the study. Thus, we hypothesise that BNM can identify a hotspot ranking value for different levels of dengue hotspots and prioritise the hotspots according to the estimated vector density.

From the infectious disease aspect, network theory based approaches have been applied in several researches to assess the relationship between the disease and the human host with

different demographic data (Belik, Geisel, & Brockmann, 2009; Bian & Liebner, 2007; Gardner & Sarkar, 2015; Malik et al., 2014; Parikh, Youssef, Swarup, & Eubank, 2013; Perez & Dragicevic, 2009; Poletto, Tizzoni, & Colizza, 2013; Salathé et al., 2010). Belik et al. (2009) used the network modeling approach to understand the effect of the heterogeneity of human movements on disease transmission. Salathé et al. (2010) reported that the network theory based approach can obtain effective results with a little amount of data collected. Eze (2013) formulated a two-node type network, which is a network consisting of two different types of nodes, location and human host. The relationship between these two different nodes in malaria transmission was analysed to understand the human interactions (Eze et al., 2011b). This two different node types can also be referred as a bipartite network.

It is important to detect the locations having unusually high mosquito densities in order to curb the number of dengue cases (Lam, 1994; Packierisamy et al., 2015). If the resources, including budget and manpower are limited, control strategies should be targeted in priority towards the risk areas of higher vector densities. This can limit the resources spent in the least important area and focus on the areas of critical vector existence. The contribution of the research is to demonstrate a new approach to model dengue transmission at district-level by locating the hotspots and prioritising the locations according to vector density.

In this study, we intend to model dengue transmission in Kuching to assist public health professionals to conduct effective control interventions. Vector control through fogging and removal of mosquito breeding sites are the main intervention methods of dengue prevention in Malaysia (Lam, 1994; Packierisamy et al., 2015). However, there is no scientific way for public health officers to identify the hotspots or the locations with high vector densities.

The current practice of public health officers in controlling the spread of the dengue disease is by fogging the reported dengue patients' residential houses and their neighbourhood within a 400 meters radius. Thus, locating the vector hotspots and vector removal in the particular site is crucial. It is undeniable that patients may visit or may have visited many other places apart from staying in their houses. They would have many different activities, for example working, exercising in the park, buying groceries at the mall, catching flights at airports and many others. Reiner (2014) suggested that detecting, monitoring and recording such human movements would improve the efficiency and efficacy of dengue surveillance and disease prevention programs.

Human movement is one of the factors that contribute to a dengue incidence surged. With increased transport of material and people, the potential for vector transportation and virus transmission and spread has increased substantially (Sutherst, 2004; Viennet et al., 2016). In this research, the relationship between human movements and the spread of dengue is investigated. In Malaysia, public health officers keep track of the whereabouts of a dengue patients and suspected dengue patients two weeks before they show the symptoms. The locations and the duration they had stayed there would be extracted and implemented into the network model.

Dengue is a vector-borne disease, the network model also takes into account the biological data of the *Aedes* mosquito and the environmental data such as humidity, precipitation and temperature. The data on human movement is collected and aggregated from the Sarawak State Health Department. Environmental data is obtained from the meteorological department in Kuching, Sarawak. Dengue transmission is sensitive to environmental factors, where temperature would increase the mosquito survival and rate of mosquito development, and would also reduce the length of the DENV incubation period within the mosquito (Chan &

Johansson, 2012; Xiao et al., 2014). Precipitation is critical to the spatial and temporal distribution of the mosquito in respective locations (Biggerstaff, Gallagher, Marano, Staples, & Johansson, 2012; Graham & House, 2014; Johansson et al., 2009). The details of the data are discussed in Chapter 3.

The problem statement is presented in the next section.

1.3 Problem Statement

Identifying dengue hotspot is the main approach for dengue prevention (Aziz et al., 2014; Nagao et al., 2003; Ritchie & Johnson, 2017). The limitations of the conventional compartmental modelling approach have been addressed by introducing a bipartite contact network model for malaria, and also taking into account population heterogeneity (Eze et al., 2011b). However, there is no contact network model for dengue that incorporates human movement to assist public health authorities in eradicating dengue. There have been no adequate model validations with real data performed in previous studies (Eze, Labadin, & Lim, 2011a; Malik et al., 2014). In addition, there is no analysis in identifying the key parameters in formulating the contact network model. In previous study, researcher utilised the same method to detect preferred habitat of Irrawaddy dolphin (Liew, 2016). The study can only detect preferred habitat where the dolphin is sighted, thus, no new habitat can be detected. In this study, patients' movement is integrated so that any new possible hotspot can be detected. There are two important components of a contact network model observed which are the host (human) and the visited location. Thus, the research question is as below:

How do we formulate two components – the host and visited location – into a network model?

It is seemed that with two discrete components, this will result in a two-node type network. In addition, the relationship between the two components will be quantified by using appropriate parameters which are related to these two components.

1.4 Research Objectives

The objectives of this research are:

- (a) To formulate a contact network model based on the interactions between humans and *Aedes* mosquitoes in different locations.
- (b) To quantify the link weight between the human and location nodes in order to evaluate the relationship between humans and *Aedes* mosquitoes in different locations based on the contact network model.
- (c) To identify the critical parameters that affect the hotspot detection.

1.5 Research Scope

The research aimed to formulate a contact network model to detect dengue hotspot in Sarawak, Malaysia. Dengue hotspot will be identified based on the two-node types network which consists of human node and location node. The hotspot is defined as the location nodes which described in terms of 400 meters of radius further from another location node. The hotspot is presented in terms of the ranking of location nodes. The higher ranked location nodes would indicate relatively higher density of the infected mosquito and thus a higher risk of the hotspot for dengue transmission.

The raw data used here are the real – world data, where the dengue patient data obtained from Sarawak State Health Department while the environmental data obtained from the meteorological department in Kuching between July 2015 and October 2016. The raw data are

untidy and not ready used for model parameters. A pre-processing process of the raw data has been done and will be expounded in Chapter 4. In addition, it is observed that the network model possess the potential to predict the hotspot two weeks before the incidence case.

The targeted users for the project outcome are public health and medical authorities. The project outcome would help them in planning prevention measures to curb dengue in Malaysia.

1.6 Research Methodology

This study models the spread of dengue through a contact network approach. Achieving this objective involves a number of stages that as depicted in Figure 1.1. The figure shows a Bipartite Network Modeling Research Methodology Framework (BNM-RMF). The BNM approach has been applied with a web-based graph search algorithm in determining malaria hotspots by Eze (2013) and applied using species habitat suitability by Liew (2016). The framework involves three stages that include problem characterisation, model construction and model analysis, and evaluation.

Stage 1: Research Problem Characterisation

The first stage comprises three processes: current research scenario understanding, graph structure representation, and bipartite graph formulation. This stage focuses on formulating of the graph structure for dengue transmission. In the first process, the primary purpose is to understand the current state of the dengue research scenario through the expert opinion such as public health officer obtained from the discussion sessions held with the expert, and the results from the reviewing of the past literature about dengue research and studying the research data. The biological characteristics of dengue, such as the mosquitoes' flight range and vector characteristic, can be identified.

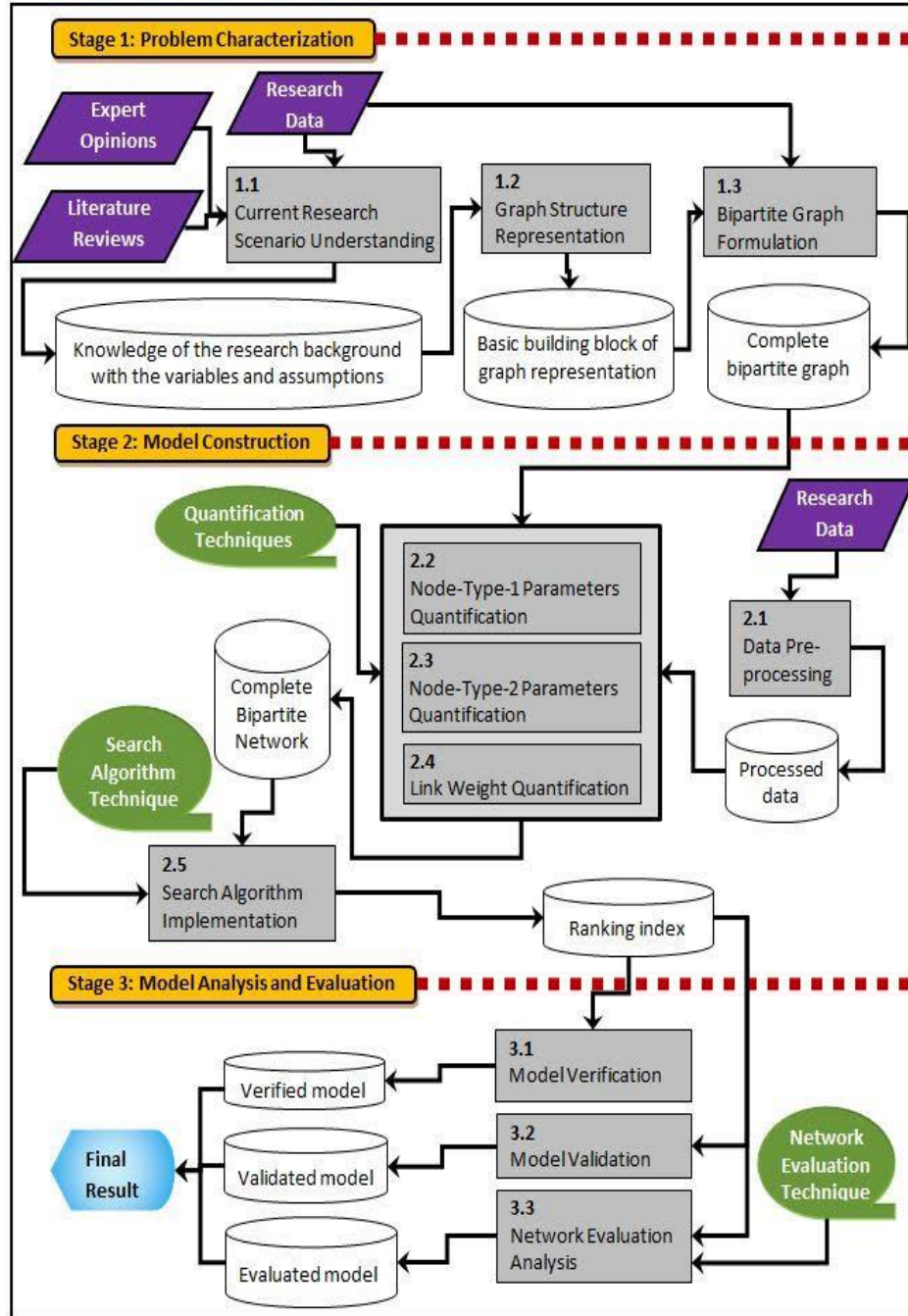


Figure 1.1: Bipartite Network Modeling Research Methodology Framework (BNM-RMF) (Liew, 2016)

In the second process, the output from first process was used to produce the basic building block of graph representation. The basic building block can be derived from the epidemiological triangle consists of environmental properties, location and host. Finally, the human and the location with environmental properties are identified as the basic building blocks.

In the third process, a complete Bipartite Dengue Contact (BDC) graph was constructed by using the basic building blocks (humans and locations). The model input is the dengue patient data, environmental data and weather data. This bipartite graph consists of two node types, the human and the location nodes and a single link between two nodes. The link is formed when a human visits a location. The bipartite graph from this stage will be the input to the next stage.

Stage 2: Model Construction

The second stage consists of five main processes: data pre-processing, node-type-1 parameters quantification, node-type-2 parameters quantification, link weight quantification and search algorithms implementation. The aim of the first process is to ensure that the data is complete especially when real data is used, and to prepare data for the execution of the next four processes. Data normalisation, data interpolation and data transformation will be utilised where appropriate. The secondary data will be aggregated and analysed. Since the data is raw and unformatted, the data would need to be pre-processed, for example, the location address needs to be processed into Global Positioning System (GPS) coordinates including longitude and latitude. Based on the coordinates, altitude; environmental properties such as temperature, precipitation and humidity; and distance between locations can be identified and quantified accordingly. The pre-processing processes will be further explained in Chapter 3 and 4.

The second and third processes involve quantification of the nodes parameters. Node-type-1 and node-type-2 refer to the respective bipartite nodes of the bipartite graph formulated in the previous stage. Node-type-1 in this study is human nodes while node-type-2 is location nodes. Quantification of the parameters of human nodes - life cycle duration (L_c), vector survival rate (S), vector biting rate (B), precipitation (Pre), altitude (Al), humidity (H) and

frequency of a location is visited by one human (Fl), and location nodes - duration of time of the human staying at one location (Du), and frequency of one human visiting one location (Fh) - are all to be determined.

The fourth process intends to quantify the link weight between human and location nodes. The decision on the quantification technique used is also chosen scientifically from a list of available techniques in the library. In this study, a summation rule is applied to quantify the link weight. The main output from the second to the fourth processes is the complete bipartite network.

The last process in this stage is the implementation of a search algorithm that determines the hotspot ranking value for the location and human nodes. The input for this process is the complete bipartite network produced from the previous process. The final result of this stage is the bipartite network model with ranking values of the human and location nodes.

Stage 3: Model Analysis and Evaluation

This final stage consists of three processes: model verification, model validation and network evaluation analysis. The goal of this stage is to ascertain that the network model formulated from the previous stage is verified, validated and evaluated. The model verification process will be carried out using UCINET as a benchmark platform in order to verify the research implementation processes performed. Another verification is the analytical verification by calculating the Spearman's Rank Correlation Coefficient (SRCC) of the location ranking values and the hub matrix. Model validation in the second process compares the SRCC values of targeted and validated network model results to validate the research results obtained in the previous stage. Error analysis will be performed to compare the actual model results and the verification or validation results. Should the model fail to pass the verification and/or

validation processes, the model will be further refined by returning to the earlier stage(s) of the methodology.

The verified and validated model will then be passed to the last process of this stage. The focus of the network evaluation analysis is to perform extended evaluations and analyses towards the model formulated. The evaluation results obtained will serve the purpose of further strengthening the research findings and also to provide auxiliary illustrations for the research findings. The techniques used may involve existing and/or new analytical method, and also visualisation tools which, together, are represented as the network evaluation technique library in Figure 1.1. The output of this stage is a verified, validated and evaluated bipartite network model.

These three stages of the BNM-RMF (Figure 1.1) will be expounded in detail as Chapters 3, 4 and 5, respectively.

1.7 Research Significance

The significance of this research is that the BNM approach has been used successfully in identifying the hotspots of the dengue vector. The model inputs are the location and the human nodes, and the link that relates these bipartite nodes. The model output is the locations ranked according to the density of the dengue vector by using a search algorithm. The output can assist public health officers and decision makers to prioritise control measures and to target interventions to the highest risk areas in order to eradicate the likely origins of the vector.

There are three main academic contributions of this research. First, an algorithm to identify the distance matrix to declare location node was developed. Second, a technique in quantifying the life cycle and survival rate parameters are developed. This technique is useful

when there is no direct data has been recorded from any data source. Third, the BNM approach is able to take into account the heterogeneity of the human population and capture the characteristics at the individual level.

Lastly, one main impact contributed by this research is that the BNM-RMF able to function as a generic methodology framework for researchers who intend to employ BNM approach.

A more detail discussion on the research contributions is presented in Section 6.3.

1.8 Thesis Outline

This thesis consists of six chapters. The content of the chapters are summarised as follows:

Chapter 1: Introduction, describes the background of this research. The main problem statement of the case study is discussed, followed by the objectives and scope of this research being defined accordingly. This chapter also presents the methodology employed in this research. Finally, the significance of this research is highlighted.

Chapter 2: Literature Review, provides the reviews of literature on different methods to detect hotspots, particularly on the main aspects of the various approaches employed, limitations and opportunities offered to the solution of this research study.

Chapter 3: Research Problem Characterisation - Bipartite Graph Formulation, explains the case study of dengue in Kuching, Sarawak, Malaysia. This chapter presents the three processes in the stage 1 of the BNM-RMF (Figure 1.1). The first process is to understand the research scenario by presenting the dataset, research region Kuching, potential variables and assumptions. The second process is to identify the basic building blocks of the graph, and finally the last process is to formulate the bipartite graph from the basic building blocks.

Chapter 4: Model Construction - Bipartite Network Formulation and Search Algorithm for Location Detection, presents the processes to construct a complete network model, being Stage 2 of the BNM-RMF (Figure 1.1). There are five processes involved and the first process is to pre-process the research data such as dengue patient data, weather data and human mobility 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. The last process discusses the implementation of a web-based search algorithm to rank the hotspot ranking value of the location.

Chapter 5: Model Analysis and Evaluation, presents three processes including model verification, model validation and network evaluation, being Stage 3 of the BNM-RMF (Figure 1.1).

Chapter 6: Conclusion, concludes the overall research and findings. The achievements of this research are presented. This chapter ends with some ideas for future work that may improve the solution.

CHAPTER 2

LITERATURE REVIEW

2.1 Introduction

This chapter presents the related review on modeling infectious disease in order to understand the theories, approaches, strengths, gaps and evolving opportunities in suggesting feasible solution to the current research problem. This review focuses on the approaches employed in vector-borne diseases particularly in dengue studies which incorporate with location detection. Since the network modeling approach stands out as motivation toward the solution of the research problem, the application of the network modeling approach in vector-disease modeling are also discussed.

2.2 Hotspot Detection Studies in Disease Aspect

Hotspot detection studies also referred to geospatial patterns modeling, area clustering modeling or spatial structure modeling. Typical studies of hotspots detection aim to map the disease occurrences with demographic and environmental properties to better understand the spread of disease. Hotspot detection studies are commonly carried out on infectious disease such as tuberculosis, dengue and many others (Hassarangsee, Tripathi, & Souris, 2015). Researchers are interested to identify specific environmental properties that are highly correlated to the distribution of the dengue patients in order to identify area where having unusually high response which can be categorised as hotspot. The following sections identify the main factors in hotspots detection studies, the main components of the study and the approaches employed.

2.2.1 Main Factors of Hotspot Detection Studies

In the disease domain, it is important for public health authorities to assess whether there is an unusual spatial aggregation of cases of diseases to make decisions in allocating resources for prevention. This unusual spatial aggregation motivates researchers to study hotspot detection in the disease domain. Environmental properties that relate to hotspot detection studies include the physical characteristics of a site and the distribution of the dengue occurrences (Rueda et al., 1990; Tran et al., 2013; Wijayanti et al., 2016). Researchers usually predict the abundance and the distribution of *Aedes* mosquitoes, where the environmental predictors that are closely linked to the survival, growth and development of the mosquitoes are reported (Arboleda & Peterson, 2009; Tran et al., 2013; Wang, Tang, & Cheke, 2016). Initial dengue modeling studies focus on temporal studies, for instance time series analysis, or to identify the relationship of the occurrences with the climate conditions. With advances in technology, for example Geographical Information Systems (GIS), more studies are conducted on modeling spatial-temporal patterns of dengue.

This section discovered the research structural similarities from the hotspot detection studies. As mentioned earlier, environmental properties are significant in hotspot detection studies. Thus, the summary of environmental predictors of mosquito-borne diseases is tabulated in Table 2.1.

From the studies listed in Table 2.1, significant environmental properties are identified and were used to quantify the nodes in the network model. Altitude, temperature, precipitation and humidity and temperature will be used in this study as important parameters to quantify the link between the nodes. Other than environmental properties, Patz, Martens, Focks & Jetten (1998) introduced the vectorial capacity equation as shown in Table 2.2 which indicates the number of contacts infected by a mosquito vector from an infected person per unit time.

Table 2.1: Summary of Environmental predictors in Hotspot Detection Studies.

Source	Disease studied	Study Area	Environmental Predictors	Approach(es) used
Wijayanti et al. (2016)	Dengue	Java, Indonesia	Elevation, precipitation, enhanced vegetation index, day-time and night time land surface temperature.	Poisson regression model
Tran et al. (2013)	Dengue and Chikungunya	Nice, France	Daily rainfall and daily temperature	Generic model with ordinary differential equations
Sarfraz et al. (2012)	Dengue	Phitsanulok, Thailand	Altitude, rainfall, mean temperature, humidity	Principal component analysis (PCA) and Bivariate Pearson's correlation
Jeefoo, Tripathi and Souris (2011)	Dengue	Chachoengsao, Thailand	Monthly rainfall, temperature and relative humidity	Spatial explicit GIS
Johansson, Dominici and Glass (2009)	Dengue	Puerto Rico	Monthly maximum temperature, mean minimum temperature and accumulated precipitation	Poisson regression and bayesian model
Wu et al. (2009)	Dengue	Taiwan	Temperature	GIS, PCA and Regression model
Scott et al. (2009)	Dengue	Thailand and Puerto Rico	Temperature	Linear regression model
Rueda et al. (1990)	Dengue	Raleigh, N.C.	Temperature	Generalized linear model (GLM)

The parameter involved in the equations is important in hotspot detection study as they are the significant predictors to describe the abundance and characteristics of the dengue vector at one hotspot from biological aspect. The vector parameters involved in the vectorial capacity equation (2.1 until 2.4) for dengue are summarised in Table 2.2.

Table 2.2: Summary of Parameter Involved in Hotspot Detection Studies

Source	Study Area	Equation/Model (VC: Vectorial Capacity)	Model Parameter(s)
Villela (2016)	Brazil	$VC = \frac{m\lambda^2 e^{-gv}}{g} \quad (2.1)$	m = vector density, v = duration required for a vector become infectious λ = biting rate g = survival rate
Liu-Helmersson et al. (2014)	Singapore	$VC = \frac{a^2 b_h e^{-\mu_m n}}{\mu_m} \quad (2.2)$	a = biting rate, b_h = transmission probability between vector and human, n = Extrinsic incubation period (EIP) μ_m = survival rate
Luz et al. (2011)	Rio de Janeiro, Brazil	None	vector density, vector survival, the extrinsic incubation period and biting rate
Lambrechts et al. (2011)	Nakhon Ratchasima Province, Thailand	$VC = \frac{ma^2 p^n b}{-\ln p} \quad (2.3)$	m = vector density, b = vector competence a = daily probability of a vector biting a human host p = vector survival rate, n = EIP duration
Focks (1998)	Bangkok & Puerto Rico & San Juan & Mexico & Athens & Greece	$VC = \frac{mbca^2 p^n}{-\ln p} \quad (2.4)$	m = vector density, b = transmission probability (infected mosquito to susceptible human), c = infection probability (susceptible mosquito bite infected human), a = number of bites p = vector survival rate, n = EIP duration

Villela (2016) redefined the vectorial capacity index in equation (2.1) of a mosquito to observe the expected number of infectious bites given by mosquito infected from biting one infected individual with the assumption of perfect transmission between human and vectors. The index can be seen as how capable the mosquitoes in spreading the dengue. From Table 2.2, vector density represents the mean number of female mosquitoes per person necessary for an infectious human case to give rise to one dengue case in a susceptible population (Focks, Patz, Martens, & Jetten, 1998; Lambrechts et al., 2011; Luz et al., 2011; Reiner, Stoddard, & Scott, 2014). In this study, the risk level of a hotspot is ranked according to different environmental factors that will contribute to mosquito density. Thus, this parameter will be used in this study.

Extrinsic incubation period (EIP) in the vectorial capacity equation represents the viral incubation period between the time when a mosquito takes a blood meal and the time when that mosquito becomes infectious (Carrington, Armijos, Lambrechts, & Scott, 2013; Chan & Johansson, 2012). In other words, little time is required for the virus to replicate after the mosquito took a blood meal from infected person and spread through the body of mosquito until the virus can be passed on to another host. These experiments on EIP are inconsistent and contradict because EIP is strongly dependent on temperature change. Also, different serotype of the dengue virus cause different EIPs (Rohani, Wong, Zamre, Lee, & Zurainee, 2009). In addition, knowledge about the EIP of dengue for *Ae. Aegypti* was scarce. Thus, life cycle duration is introduced to examine the risk of dengue occurring in a given area (Carrington, Armijos, Lambrechts, Barker, & Scott, 2013; Rueda et al., 1990; Tun-Lin, Burkot, & Kay, 2000). There are four life stages of *Aedes* mosquito, eggs, larva, pupa and adult. The life cycle duration refer to the entire life cycle it takes from eggs until adult (CDC, 2016). Several experiments have been done to access the relationship between the temperature and the life cycle duration. The fitting and quantification process will be discussed in Chapter 4.

The transmission of dengue is strongly dependent on mosquito biting behaviour and this biting behaviour is temperature-dependent (Focks DA, 1998; Maciel-de-Freitas, Eiras, & Lourenço-de-Oliveira, 2008; Phaijoo & Gurung, 2015; Sylvestre, Gandini, & Maciel-de-Freitas, 2013; Wesolowski et al., 2015). The biting rate can be defined as the number of blood-feeding of a mosquito taken per unit time (Scott et al. 2000). A linear regression model of vector biting which associated with temperature was applied in this study (Scott et al., 2012). The equation (2.5) from Scott et al. where $B(T)$ refers to biting rate at certain temperature T . The temperature range from 21°C to 32°C. The equation and the quantification process will be further explained in Chapter 4.

$$B(T) = 0.03T + 0.66 \quad (2.5)$$

Apart from that, vector survival has long been recognised as one of the most significant parameters in mosquito-borne disease transmission (Carrington et al., 2013; Rueda et al., 1990; Tun-Lin et al., 2000). Experiments conducted by Tun-Lin et al. (2000) on female *Aedes* mosquitoes over the temperature range of 15°C to 34°C found that the survival rate are ranged from 23.5% per day to 67% per day. The fitting and quantification process will be discussed in Chapter 4.

“Vector competence is a measure of the ability of a mosquito to become infected with, allow replication of, and transmit virus to a susceptible host” (Kramer & Ebel, 2003). This vector competence is a linear component in the equation (2.3), thus is a weak predictors to the overall vectorial capacity. In other words, even if it is a poorly competent mosquito might cause intense dengue transmission if other predictors included in the equation are favourable. High survival rate, for example, compensated the minimal vector competence. Therefore, the vector competence is not applied in this study.

As a summary, life cycle, survival, biting behaviour and frequency of mosquito visits the location are concluded as important model predictors. The quantification process of the environmental properties and vector-related parameters will be further discussed in Chapter 4.

Based on the literature presented above, it is observed that there always exist three main components in the research structure of all hotspot detection studies of dengue being reviewed, namely host, location, and environmental properties as shown in Figure 2.1.

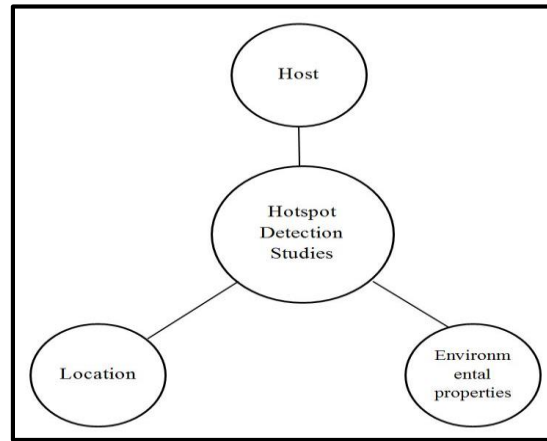


Figure 2.1: Three main components of dengue location detection studies structure.

2.2.2 Approaches of Hotspot Detection Studies

This section discusses the capability of the approach that can incorporate the important predictors and parameters as highlighted in previous section. As mentioned in section 1.2, compartmental modeling is one of the most used approaches in disease studies (Siettos & Russo, 2013). Besides that, another recently developed modeling approaches for instance agent-based modeling and several scientific methods such as machine learning and graph-theoretic network modeling approach are also employed to study the dynamics of infectious diseases. These recent approaches are applied to overcome the limitations of conventional approaches. This section outlines the existing methods and understands the gaps of these four approaches in hotspot detection.

a. Compartmental Modeling

The first dengue model used for public health decision developed by Newton & Reiter (1992) who introduced compartment model for dengue by dividing the human population into susceptible, exposed, infected and recovered compartments. Compartmental modeling is conventional approach in the mathematical study of infectious disease which employs a system of differential equations (Chitnis, 2011). The approach focuses on categorising the human population into different compartments and ignores heterogeneity and random effects, which are generally most influential in the initial or final stage of an epidemic (Garnett et. al., 2011).

Researchers started to incorporate the vector populations to examine the effects and relationship between mosquito and human (Focks, Haile, Daniels & Mount, 1993). Other than described the human and vector population, Esteva and Vargas (2003) introduced different dengue serotypes into the compartmental model to understand the vertical transmission conditions effect on the overall disease. In other words, they formulated non-linear differential equations to determine the relationship between two dengue serotypes. They obtained convincing basic reproduction numbers to suggest the coexistence of two different serotype viruses is possible. On the other hand, PDE are introduced when spatial component integrated with temporal component.

Here we will take one example of dengue study with spatial component. Maidana and Yang (2008) studied the spatial distribution of dengue in order to locate the outbreak region by using a system of partial differential reaction-diffusion equations with application of mass action law. This studies model 2 scenarios: (a) categorised the mosquito and human population into infected, susceptible and recovered sub-population and disregard the movement of human population and only mosquito can move around; and (b) categorised the mosquito and human population into infected, susceptible and recovered sub-population, human and mosquito can

move around. In other words, the human population distribute homogenously in scenario (a). This study suggested two groups mosquito are involved in the model which are adult mosquito and mosquito in aquatic stage (eggs, larvae and pupae), only adult mosquito is allowed movement in this model. Scenario (a) shows the occurrence of dengue infection is very intense due to human movements and parameter is obtained from the model such as minimum speed of transmission when only the mosquito can move around. The result can help public health authority to understand the contribution of the vector movement in spreading the disease, so they can make efficient control measure. Scenario (b) allows the movement of human and mosquito in the space. New parameter for human movement, diffusion of human population is introduced in the system of PDE. The solved reaction-diffusion equations show the human movement can increase the rate of transmission and shortens the duration of dengue. Maidana and Yang (2008) used 33 parameter such as survival time in aquatic phase, oviposition rate and others to solve the reaction-diffusion numerally and the respective parameter values are available from the paper (Table 1-5, 67-71). The compartmental modelling cannot incorporate the predictors such as the temperature, altitude, life cycle rate and others predictors in the different equations.

As a conclusion, human population dynamics studies incorporating spatial aspect that employed compartmental modeling approach has three characteristics: (a) the human and vector population are assumed to be well-mixed and homogenous; (b) a population dynamic model is formulated is formulated with a strong basis on physical laws; (c) detailed parameter values are essential to solve the differential equations. The limitations are expected if compartmental modeling is employed when the problem has no available physical law to base and the population heterogeneity need to be considered. This approach unable to incorporate the

important predictors that decided in previous section. Besides that, obstacle is expected when the detail parameter values are not available from other literature.

b. Agent-based Modeling

Researchers aim to overcome the limitation of compartmental model which assume the population interacts homogenously. Agent-based modeling emerged in the 1990s and it can be used to observe the propagation of the disease (Heath, Hill, & Ciarallo, 2009). Agent-based Modeling (ABM) aims to model different agents to understand interaction between agents (Galea, Riddle, & Kaplan, 2010; Garnett, Cousens, Hallett, Steketee, & Walker, 2011). In epidemiology, ABM is a stochastic model which can model every agent for example patients, buildings, transportation in one imaginary environment. Each individual is assigned with certain attributes and behaviour, for instance, human have the ability to move randomly in the imaginary environment to assure physical contacts or infected with the disease. It is important to include all related information of the disease for example the social behaviour patterns, individual attributes and disease characteristic to build a realistic and useful model (Bruch & Atwell, 2015; Miksch, Urach, Einzinger, & Zauner, 2014). The agents can be programmed to behave and interact with other agents in the imaginary environment.

Jacintho et al. (2010) build an agent-based model to focus micro-level behaviour of dengue in Brazil. The researchers build a model with mosquito agent behaviour (four stages-eggs, larva, pupa, and the landform), human agent behaviour (susceptible or infected), exterminator agent behaviour which represents the public health organizations that notify all hotspot area when planning control actions and environment (temperature, probability to find food and humidity). A conceptual model implemented in Swarm simulation platform to simulate real life situation. The authors validated the model by examining different climatic seasons, when winter season, the number of mosquitoes drops while when summer season, the

number of mosquito rises. The simulated scenarios were very similar to real life situation and they provide a framework for better understanding the roles of each entity in the global context of the spread of the disease. A hotspot detection study is done by using ABM to model the outbreak sites in urban area when human movement is incorporated in the model (Perez & Dragicevic, 2009). Hotspot can be identified when researchers model the human mobility and simulate the agent interact among themselves or move around from one location to another.

Different with compartmental modeling, ABM does not start the expression of the relationship with equations but the individual behaviour is based on rules set by modelers (Teoh et al., 2015). Individual heterogeneity and random events dominate the disease propagation and significant in eliminating the disease transmission (Arifin et al., 2015). For instance, Teoh et al. (2015) captured the feeling of fear with human behaviour, the feeling of fear associated with the vaccination program. The random events such as whether or not the susceptible take vaccine will affect the spread of the disease. The researchers model the scenario with NetLogo® (Wilensky, 1999) and their results showed that there are lesser infected cases for a population who received more disease information compared to those who received less information. The authors hypothesised the fearfulness level of the population who received more information is higher. GAMA modeling and simulation platform (Drogoul et al., 2013) is aimed at supporting the design of spatialised and multiple-scales models. GAMA can address the limitation of NetLogo when dealing with rich models relying on GIS data (Drogoul et al., 2013). Netlogo has the ability to integrate human heterogeneity by programmed the agent to make the agents moving and contact with others in random direction. This random event resulted in a stochastic effects that might cause different results after many simulation trials (Geoffrey P. Garnett et al., 2011; Willem et al., 2014). Run multiple times with the same parameter values give different results due to this randomness (Dommar, Lowe, Robinson, & Rodo, 2014). The stochasticity

of the agent-based model can increase the model divergence. However, the model stochasticity can be addressed by using the average results from a series of simulation trials. This approach is time-consuming as a number of simulation need to be run.

Besides, good parameterisation is required to prevent the result from distortion (Miksch et al., 2014). Similar to deterministic compartmental model, complete parameter values are required to model one agent-based model. In addition, the parameter values need to be corresponded to the model in term of time unit. For instance, the incubation periods of one disease is known to be one week and the model time step is one day, the recovery time has to be programmed as 7 time steps to make it one week. Wrong parameterisation will change the dynamic of the whole model.

From the technical perspective, ABM allows independent development of different modules for instance, population module and contacts module can be programmed and reuse if researcher want to simulate the same disease in different countries which allows reusability of the modules or combination of existing modules. Reuse the validated modules can decrease the effort and increase the accuracy (Miksch et al., 2014). However, due to the fact that ABM has the ability to incorporate with the population heterogeneity, large computation power is required since the model need to store and process each individual's attributes separately (Dommar et al., 2014; Garnett et al., 2011; Miksch et al., 2014). The realistic disease scenario makes a large-scale agent-based model and thus there is a need to increase hardware performance for example introduces multi-core or parallel processing.

As a summary, ABM offers great benefits and has three characteristics: (a) disease can be modelled based on individual's behaviour and interaction, this capture the population heterogeneity; (b) module is reusable; (c) related information of the disease and complete parameter values are required. However, the ability to simulate the population heterogeneity

can cause stochasticity that leads to model divergence. Also, the interaction of every individual in the model causes computational intensive and require large computation power.

c. Other Scientific Methods

Many hotspot detection studies have used other scientific methods to overcome the limitation of the conventional approaches mentioned in previous sections (Bian & Liebner, 2007; Kesorn et al., 2015; Malik et al., 2014; Poletto et al., 2013; Salathé et al., 2010). One of the scientific methods is machine learning approach. There are three types of machine learning algorithms: unsupervised learning, supervised learning, and reinforcement learning. Unsupervised learning used to understand the structure of data set and the inherent pattern of the data, used for clustering population in different groups. There is no dependent variable as all variables are treated equally with respect to one another, for example K-means. In contrast, supervised learning is looking for the relationship between independent variables or dependent variable to make predictions based on a given set of predictors for instance regression, decision tree (DT), support vector machine (SVM), K-nearest neighbour (KNN) and random forest. The reinforcement learning is used to make decision that it trains itself continuously by using trial and error basis. The algorithm can learn from the past experience and capture the best possible knowledge to make decision for example Markov decision process.

To date, the disease hotspot detection using machine learning has not been attempted yet. Nevertheless, Liew (2016) applied SVM and decision tree techniques in habitat suitability study to quantify the water suitability and food availability parameter due to the imbalanced dataset. Even though the machine learning method is not directly applied in the location detection study, but it has been applied in parameter quantification process of location detection study.

Kesorn et al. (2015) used SVM technique to forecast the morbidity rate of dengue in Thailand. The researchers take nine predictors includes temperature, rainfall, humidity, wind speed larvae infection rate, male mosquito infection rate, female mosquito infection rate, population density and morbidity rate. Different supervised classification algorithms for instances KNN, DT and SVM are deployed in the model. The result of the study shows the female mosquito infection rate is strongly related to the number of observed dengue cases where male mosquito infection rate is not linked to number of dengue cases. Last but not least, the high infection rate of larvae mosquito found in summer season proportional to high number of dengue cases.

Data sample size does affect the model accuracy, machine learning methods for instance SVM and maximum entropy (Maxent) runs effective only when the sample size is large (Wisz et al., 2008; Yu, Yang, & Han, 2003). Small sample size leads to overfitting and the noise becomes a real issue (Bishop, 2013). Bishop suggests regularisation technique to avoid overfitting when the typical problem encountered due to small sample size. Thus, small sample size is a big concern to the researchers to undergo location detection study. This is a research gap where the current study intends to fill. To conclude, machine learning method may not be the main approach applied in hotspot detection in this study, however the techniques are available in the quantification techniques library as shown in Figure 1.1.

One scientific approach that particularly stands out in handling the small size of research data and have the ability to include both the environmental predictors of the location and the human mobility data are graph-theoretic or network-based modeling. The three main components shown in Figure 2.1 can be represented as a graph structure with three vertices (Liew, 2016). Network model is a graph structure, in other words, the network theory is originated from graph theory. Hence, there exists a strong motivation that the network-based

approach can be employed in solving our research problem. Hotspot detection related studies that employ network modeling will be discussed in the next section. This network approach applied in many recent hotspot detection studies address the limitations of the conventional approaches mentioned earlier. A review of network modeling studies is presented in next section.

2.3 Network Modeling Approach

This section outlines the literatures of the graph-based and network-based approach in dengue modeling with spatial components.

2.3.1 From Graph Theory to Network Theory

Graph theory is the root of network theory, this section presents the introduction of graph theory and the progress to its application in network theory. The origin of graph theory started with the problem of Koinseberg bridge in 1735 and introduce the concept of Eulerian graph (Shirinivas, Vetrivel, & Elango, 2011). Graph theory is an important branch of mathematics which is highly utilised by computer science applications especially in research areas such as clustering, data mining and network topologies. Graph theory widely used in biology where a vertex represents the location where certain species/host exist and the edges represent the migration or movement between the locations (Shirinivas et al., 2011). This finding is significant and applied in our research problem to study the breeding patterns or tracking the spread of the dengue.

A graph is reported as a “versatile models for analysing a wide range of practical problems concerned with the properties and functions of networks” (Laita et al., 2011 in Liew (2016, p.38)). Network uses graph as a representation of relations between discrete objects (Barnes & Harary, 1983; Mason, & Verwoerd, 2007; Shirinivas et al., 2011). Therefore, a

research problem that can be viewed and represented as nodes and the nodes are linked by edges motivates to solve it with network theory.

As indicated in Figure 2.1, three main components of the location detection studies can be viewed as the nodes and the relationship between the components can be analysed with the links. Hence, in this research, graph is used as a graphical visualization representation of the solution to our research problem where the functional definitions of the nodes and links are pre-determined. A network in this research refers to the weighted graphs which the weight is determined through quantification of the properties for the nodes and links.

2.3.2 Network Modeling Approaches and Spatial Studies

In disease domain, the network theory is applied in different scenario which comprises analysis of spatial dispersion (Zhang et al., 2014), human sexual network in transmitting AIDS/HIV (Handcock & Jones, 2004; Potterat, Rothenberg, & Muth, 1999), heterogeneity network of host contact patterns (Bansal, Grenfell, & Meyers, 2007) and many others.

A network model is a graph structure where each node represents a person or a location while the edges represent the contacts among people or location in the network (Meyers, 2007). In disease epidemiology, researchers need to build a realistic network model that includes contact patterns at an appropriate spatial and temporal scale. This network which incorporates contact is known as contact network model that counts human interactions that lead to disease transmission. Rather than simply assuming that all individuals have equal possibility of contacting the disease, a contact network modeling takes note of the probability of disease transmission is proportional to the level of contacts (Newman, 2002). Infectious disease can be transmitted due to human contacts either with other infected human, visited high-risk locations, or contact with non-human infectious agents for example, dengue virus is transmitted to humans

via the bite of infected mosquito at infected location (Helmersson, 2012). The human contact takes places within a location where these infected mosquitoes thrive.

In a contact network, each host is represented into a node while the edges connecting the nodes can represents the contacts between the hosts (Dimitrov & Meyers, 2010). Contact network is good to use when researcher form a small-world network that can model the disease with local clustering. The complexity between compartmental model, agent-based model and contact network model is shown in Figure 2.2 where the x-axis indicates the required computational power and the y-axis indicates the extent to which the model capture the real world complexities that influence the disease transmission (Dimitrov & Meyers, 2010). Based on Figure 2.2, compartmental model is easy to analyse but it cannot capture the heterogeneity and the contact patterns of the population while agent-based model can model the real world realism, however it is difficult to parameterise and analyse the model. Only contact network model can model the disease with higher fidelity compared to compartmental model yet remain analytically tractable.

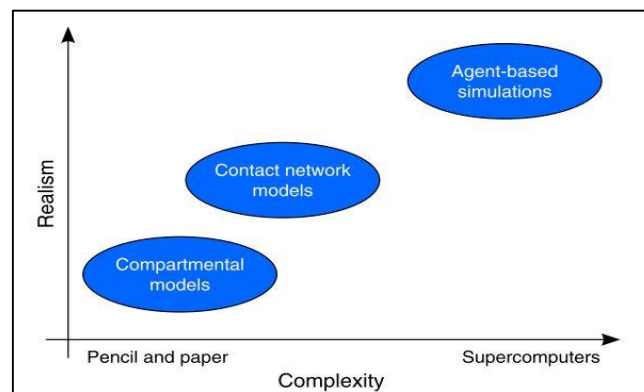


Figure 2.2: Complexity of the epidemiological models (Dimitrov & Meyers, 2010)

Besides that, Malik et al. (2014) formalised the dataset of dengue fever cases in Selangor, Malaysia into two-node type complex network to analyse the impact on the location. Dataset provided by Ministry of Health Malaysia involves number of infected cases per week. Researcher build two-node type network with two different sets of nodes which are locations

as primary node and number of infected cases in weeks as secondary node. The two nodes are considered as connected if they share same week of infected cases and the number of infected cases is considered as the links weight. The two-node type networks are converted into one-node type network by projection methods because it is rarely analyse in their original form as there are very few network metric available (Zhou et al., 2007). However, this projection makes the network a less informative network compared to the two-node type network. By using network analysis metrics this research has revealed that few locations have high impact of dengue. The graphical representation of the network plotted whereby the node with unusually links representing as hotspots of dengue. The results show these nodes are important to be focused for treatment. There is no data validation or analysis about the hotspot with real data.

The studies summarised in Table 2.3 suggest the nodes of the network represent the patients while the edges describe the contact between the nodes. The contact can be defined as the distance between the patients and hotspot (Benavides, Walsh, Meyers, Raymond, & Caillaud, 2012). These studies involve one-node type in the network model and the one-node type network formed through the projection process.

The projection process turns the two-node type network into one-node type network, therefore the network will loss important information and it turns into an unweighted network (Zhou et al., 2007). Two-node type network also known as bipartite network. A bipartite network refer to a network consisting of two different types of nodes with links joining only between unlike nodes (Ferrerri & Giacobini, n.d. ; Liew, 2016). In other words, there is no link from a node to another of the same node type. Therefore, bipartite network have important role to play in shaping the solution to this research problem and enhancing the understanding of the epidemiological process.

Table 2.3: Related Disease Studies Applying Contact Network Modeling

Source	Study focus	Disease	Region
Benavides et al. (2012)	Examine the effect of distances between hotspots	Measles	United States
Emch et al.(2012)	Spatial and social network analysis	Bacterial diarrheal	Bangladesh
Firestone, Christley, Ward and Dhand (2012)	Incorporate spatial and contact relationship between infected locations to predict hotspot	Equine influenza	Australia
Balcan et al. (2009)	Improve the performance of network modeling with large scale of mobility data	Influenza	29 countries in five continents
Zeng, Chen, Cao and Zhen (2009)	Improve the social-contact network by enriching the context of patient contacts with human mobility to predict SARS outbreak	Severe acute respiratory syndrome (SARS)	China
Massad et al. (2008)	Connectivity in heterogeneous contact	Dengue	Singapore

Hotspot detection studies in epidemiological field are very scarce. Table 2.4 shows the summary of the recent literature reviewed. The focus centres on identifying the impact between two distinct nodes and analysing the related issues on disease studies by applying the bipartite network modeling approach. The network-theory is reported that can obtain accurate result with little data used (Salathé et al., 2010).

Network is a flexible tool to represent the interactions of different components in a complex system. Thus, network has been used to model the spread of disease in many forms. The nodes can be used to represent one individual, group of individuals or locations to which individuals are connected. On the other hand, edges can be used to describe transmission events, movements of the individual, travel route or simply acquaintances between the individuals (Pellis et al., 2015). In other words, the environmental, epidemiological and the demographic predictors can be integrated into the network model by quantifying the node link.

Based on demographic, environmental, epidemiological and patient mobility data, we model dengue patients' interactions within residential houses, schools, workplaces, hospital, shops, restaurant, neighbourhoods and many others on a small scale to identify the possible dengue hotspots or any area with unusual spatial aggregation of dengue cases.

It is noticed that these contact network-based research on disease transmission share some common characteristics: (a) capture the population heterogeneity; (b) can support location detection; (c) it can be solved analytically; (d) can be used even the dataset is small; (e) some model parameter values are required such as transmission probability (Meyers, Newman, & Pourbohloul, 2006).

Table 2.4: Related Disease Studies Applying Bipartite Network Modeling Approach

Source	Disease	Nodes
Zhang, Li and Jin (2015)	Dengue	Node 1: Human Node 2: Sub-regions Edge: Connected when there is the possibility of transmission between them
Pillai et al. (2012)	Asthma	Node 1: Patient Node 2: Cytokines Edge: Each patient to each cytokine
Eze, Labadin and Lim (2011)	Malaria	Node 1: Patient Node 2: Visited locations Edge: Contact strength
Bisanzio et al. (2010)	Vector-borne disease	Node 1 and 2: Human Edge: Transmission

2.4 Reviews on Different Web-based Search Algorithms

A web-based search algorithm is implemented to rank the location and human nodes in the last process in Stage 2 as shown in Figure 1.1. Four web-based search algorithms are reviewed and comparison between these algorithms are expounded in detail in the following paragraphs.

a. Hypertext Induced Topic Search (HITS)

Hypertext Induced Topic Search (HITS) rank a web page by processing the in-links and out-links of the web pages (Dilip et al., 2010). The pointed web page namely authority and the web page that pointing namely as hub. A hub and authority score is assigned to a page that are query specific because the HITS algorithm is executed at query time with the associated hit on performance that accompanies query time processing (Pardakhe & Keole, 2013).

In HITS algorithm, decision on ranking node is made by analysing textual content against a given query. Hence, the returned pages have high relevancy and importance. The weight of hub and authority are computed by using an interactive algorithm. The HITS algorithm gives World Wide Web as directed graph $G(V,E)$, where V is a set of vertices representing pages and E is set of edges corresponds to link. The first step is sampling step while the second one is iterative step. Sampling step a set of relevant pages for the given query are collected, In iterative step, a hub and authority are found by using the output of sampling step. Expressions (2.6) are used to calculate the weight of Hub (H_p) and the weight of Authority (A_p) where H_q is the hub score of a page, A_q is the authority score of page, I_p is the set of reference pages of page p , Bp is the set of referrer pages of page p .

$$HUB(H_p) = \sum_{q \in I_p} A_q \quad (2.6)$$

$$AUTHORITY(A_p) = \sum_{q \in Bp} H_q$$

b. PageRank Algorithm

PageRank algorithm is dependent on link structure of the nodes and will consider the node towards other node as important node if node contain important node towards it (Dilip et al., 2010). The back link is taken in consideration by PageRank to decide the rank score. Hence, the rank is calculated on the basis of the importance of nodes. The limitation of this algorithm

is the results are generated at the indexing time and not at the query time. Assume page A has page $T1$ until Tn which point to it, PageRank equation for page A , $PR(A)$ is expressed in equation (2.7) where d is a damping factor, Tn is all the page that link to page A , $PR(Ti)$ is the PageRank of page Ti and $C(T1)$ is number of outgoing link from page $PR(T1)$.

$$PR(A) = (1 - d) + d(PR(\frac{T1}{C(T1)}) + \dots + PR(\frac{Tn}{C(Tn)})) \quad (2.7)$$

c. Eigen Rumor Algorithm

Fujimura et al. (2005) proposed Eigen Rumor Algorithm that rank each blog entry by weighting the score of hub and authorities based on eigenvector calculations. The high score to a blog entry is entered by a good blogger. Three vectors are considered in Eigen Rumor algorithm which are authority, hub and reputation. The PageRank and HITS are related with this algorithm that results are based on eigenvector calculation of the adjacency matrix of the links. The hub and authority scores can be calculated even though do not have any in-link yet using EigenRumor algorithm.

d. Tag Rank Algorithm

Jie et al. (2008) proposed TagRank algorithm or known as novel algorithm which the social annotations is the basis for ranking the web page. The time factor of data and annotations behaviour is used by TagRank algorithm to calculate the tag. The results produced by this algorithm is very accurate and the new information resources are indexed better.

Based on Table 2.5, HITS algorithm is used in this study because HITS relate both hub and authority pages and this structure is similar to this study which involves two node types. HITS has high relevancy from the returned pages and this can help to determine a critical location from a dengue patient.

Table 2.5: Comparison of Web-based Search Algorithms

Algorithm	HITS	Page Rank	Eiden Rumor	Tag Rank
Technique	Web Structure Mining, Web Content Mining	Web Structure Mining	Web Content Mining	Web Content Mining
Methodology	Computes the hubs and authority of the relevant pages and relevant as well as important pages as the result.	Computes the score for the pages at the time of indexing of the pages.	The adjacency matrix is used.	The sequential clicking is used for sequence vector calculation by using random surfing model.
Input Parameter	Content, Back and Forward links	Back links	Agent or Object	Popular tags and related bookmarks
Relevancy	More	Less	High	Less
Importance	Moderate as the hubs and authority scores are used.	High as the back links are considered	High	High
Advantages	Have high relevancy and importance from the returned pages	Rank calculation is based on importance of nodes.	High relevancy in blog ranking	The results are very exact and information resources are indexed more effectively.
Limitations	Topic drift and efficiency problem	The result is obtained at time of indexing instead of query time	It is only specifically used for blog ranking.	More sites are required to make comparison.

2.5 Summary

This chapter presented the reviews on the related work about hotspot detection studies from the aspect of identifying the main factors, for instance environmental predictors and human mobility and the common approaches employed. After the reviews, it is discovered that there always exists three main components (Figure 2.1) in the research structure of all hotspot detection studies of dengue, namely host, location and environmental properties. This structure serve as a basis of the network modeling approach and the network formulation will be discussed in next chapter. The formulation comes with a development framework, namely the BNM-RMF (Figure 1.1). In addition, the strengths, gaps and opportunities of various approaches that are employed in hotspot detection related studies are identified. Among the compartmental modeling, agent-based modelling, machine learning and bipartite network modeling approaches, the bipartite network modeling approach is applicable in this study when the researchers want to capture human heterogeneity and when the dataset is small and have the ability to incorporate all important predictors.

CHAPTER 3

RESEARCH PROBLEM CHARACTERISATION

3.1 Introduction

The chapter presents the case study of dengue transmission in Kuching, Sarawak. The formulation of the bipartite graph which is the input of bipartite network is discussed. Stage 1 of the methodology BNM-RMF shown in Figure 1.1 consists of three processes to formulate the bipartite graph. In order to identify the current research scenario, related assumptions are drawn to further ascertain the problem. Next, the basic building blocks are determined to formulate the bipartite graph. In order to formulate the bipartite graph, expert opinion, the assistant environmental health officer with grade U29 is collected. The expert opinion, raw data and parameters after pre-processed will be expounded. The outcome of stage 1 of BNM-RMF is the bipartite graph which serves as an input to formulate bipartite network and will be discussed in the next chapter.

3.2 Current Research Scenario

Previous chapter concluded the research problem is best represented by a bipartite network. A bipartite network consists of two node types and links between nodes of two different types, it never between nodes of the same type. Thus, the following sections will explained the graph representation and justify the basic building block of the bipartite network.

3.2.1 Graph Structure Representation

This section discusses the process in Stage 1 of BNB-NF which focuses the formulation of basic building block as the basic graph structure representation of the bipartite network model.

3.2.2 Assumptions

Before discussing the graph structure representation, the assumptions of this study are drawn to further ascertain the research problem is best represented by a bipartite network. Dengue is a mosquito-borne disease which the hosts are mosquito and human. However, the mosquitoes were not characterized as a node in the network model in this study as the mosquitoes are reported cannot fly very far and it can only stay within 400 meters from the particular location (Eze et al., 2011b). Therefore, the location nodes and the mosquito nodes are assumed the same.

As discussed in Section 1.4, this study aims to evaluate the relationship between human and mosquito in different locations through link weight quantification. The relationship was represented by using the graph edges between two node types. In this study, the edge represents the contacts between human host and the location. Therefore, the dengue transmission rate was assumed to be proportional to the intensity of the contacts between the adjacent nodes.

3.2.3 Epidemiological Triangle

Eze (2013) highlighted that the epidemiological triangle is a visual representation graph which consists of three vertices (Dembek, Kortepeter & Pavlin, as cited in Eze, 2013; Huerta & Leventhal, as cited in Eze, 2013, p. 22). The triangle is used as one of the basis of formulating his work on malaria transmission as shown in Figure 3.1. The three epidemiologic factors in epidemiological triangle are interdependent and used to be identified the basic building blocks. The focus disease in this study is dengue fever. Since dengue is also a mosquito-borne disease and dengue has the same clinical and biological metrics as malaria. Therefore, the basic

building blocks for dengue would be the same as malaria, which has been published by Eze (2011a).

Human and mosquitoes act as the main hosts of dengue transmission, according to the assumption in Section 3.2.2, the mosquitoes would be included within the location nodes due to the short flight range.

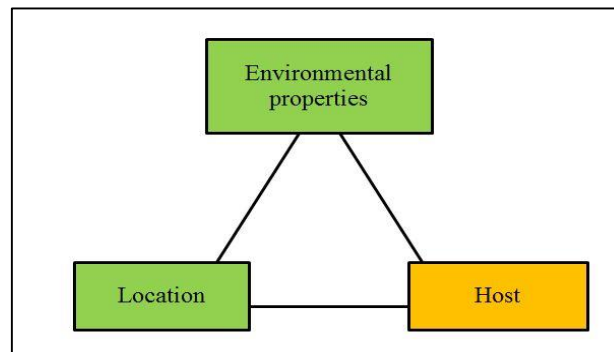


Figure 3.1: Epidemiological Triangle (ET)

However, the location and environmental properties are two risk factors that are strongly related. Liew (2016) pointed out that the location component specifies the spatial features of certain environmental properties. At the same time, environmental properties characterise a specific location and differentiate one location from another. Therefore, the location component can be viewed as a component housing the environmental properties. Hence ET is further modified as shown in Figure 3.2.

From Figure 3.2, the location denoted as L, component of environmental properties denoted as N and the component of host denoted as H. The previous three vertex graph structure (Figure 3.1) of ET is modified to a two-vertex graph structure (Figure 3.2).

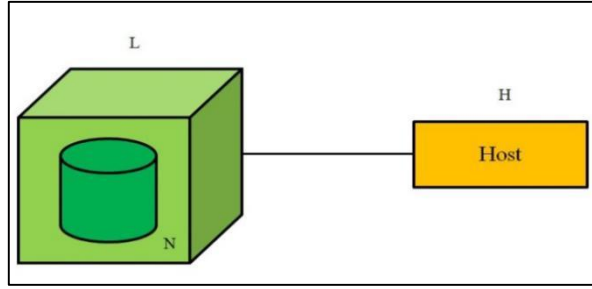


Figure 3.2: Modified Epidemiological Triangle

Based on the assumption in Section 3.2.2, the location component is strongly related with the environmental properties. It is clearly seen that the graph structure in Figure 3.2 can be used to form the basic building block of the bipartite network model, which is one of the research objectives. The two different nodes consist of human and location nodes. Discussion on this basic building block is discussed in next section.

3.2.4 Basic Building Block of the Bipartite Dengue Contact Network Model

The basic building block which derived from previous section is shown in Figure 3.3. The component of host is replaced with the human and denoted as H. From this point onward, the term “human” used in this thesis refers to the dengue patients in Kuching unless otherwise specified.

From Figure 3.3, there are two vertices L and H that imply the attributes for vertex location (L) and vertex human (H) are different. As an outcome, here we formed a network structure which consists of the sets of location nodes (L), the set of human nodes (H) and the 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 also be referred as bipartite network, which is a heterogeneous network of two node types. The graph structure is termed as Bipartite Dengue Contact (BDC) Graph while the network model is termed as Bipartite Dengue Contact (BDC) network.

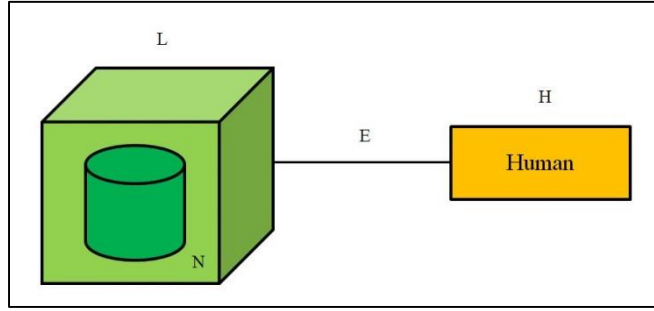


Figure 3.3: Basic Building Block of BDC Network

The difference between the graph and the network is the BDC graph is an unweighted bipartite graph for visual representation of BDC network. BDC network is weighted bipartite graph that gives both the topological and functional relationship of its elements – the bipartite nodes and their respective links (Rayfield et al., 2011 in Liew (2016, p. 38)). The link weight is a measure of affinity between nodes in BDC network. In order to quantify the contact strength values, appropriate parameters which are related to location and human nodes need to be taken into account as location and human nodes are associated with the increased probability of dengue spread.

An example of BDC network can be formulated by using the basic building block. The completed bipartite network is obtained in Figure 3.4, C , can be defined in terms of its component sets, $C = C(H, L, E)$ where H and L represent the set of nodes and E represents the set of edges. Set H is the set of human consists of five elements which $H = \{H1, H2, H3, H4, H5\}$ and set L is the set of locations consists of six elements which $L = \{L1, L2, L3, L4, L5, L6\}$ while set E , the set of links that join the elements of L and H , has nine elements, $E = \{e1, e2, e3, e4, e5, e6, e7, e8, e9\}$, which can be expressed as $\{H1L1, H1L2, H2L6, H5L6, H5L3, H5L5, H4L5, H3L3, H1L4\}$.

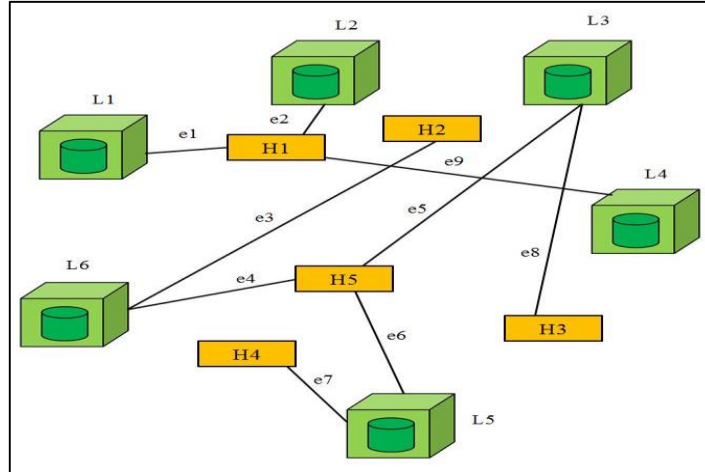


Figure 3.4: An example of BDC network

In conclusion, the basic building block for BDC 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 motivates the uses of the bipartite network as there is no link from a node to another of the same type, the link only formed between two different node types. Next section will discuss the definition of the link between the nodes.

3.3 Research Data and Potential Parameters

As mentioned in previous section, the sets of edges are the link between two node types. It can be seen as the contact strength of the human and location nodes. Here, we derived the representation of the link by looking at the research data. Besides, we derived potential parameters that will be used to quantify the link weight in Chapter 4.

3.3.1 Raw Data

The study area is in Kuching city, the capital city of the Malaysian state of Sarawak on the island of Borneo. The raw data from Kuching were collected and aggregated. The raw data was categorised into four sub-datasets: the epidemiological dataset; location dataset;

meteorological dataset; and mosquito dataset. Each of these sub-datasets is summarised in Table 3.1.

Table 3.1: Four Sub-datasets of Research Data

No	Sub- Dataset	Data Collected and Recorded
1	The epidemiological dataset	i. Onset date (Date when patient first show symptoms) ii. Notification date (Date when patient visit hospital) iii. Physical conditions (age, gender, home/work address) iv. Date and time the patient visit the particular location
2	The location dataset	i. Human mobility data (location visited by the patients) ii. The residential housing address
3	The meteorological dataset/weather data	i. Date ii. Daily maximum surface air temperature (degree Celsius) iii. Daily minimum surface air temperature (degree Celsius) iv. Daily total precipitation amount (millimetre) v. Daily mean relative humidity (percentage)
4	The mosquito dataset	i. Physical characteristics (species, living habitat) ii. Flight range

Firstly, the epidemiological dataset recorded all individual positive or negative dengue patients whom are identified by public health officers through serological tests (Fry et al., 2011; Parkash & Shueb, 2015; Wang & Sekaran, 2010). The epidemiological data were collected from a form named PBV (DD/DDB/JE/CHIKU) 201 Pindaan 4/13 as shown in Figure B.1 (in Appendix B). From this point of thesis onward, the investigation form is used to refer to this PBV (DD/DDB/JE/CHIKU) 201 Pindaan 4/13 form. Public health officers in Kuching use this investigation form to interview the patients to gather data. In this study, the patients from July 2015 until October 2016 or between Epi Week 28 until Epi Week 39 in 2015, total 12 Epi weeks were collected. Epi Week is epidemiological week which is also referred as CDC week (Arias, 2016). It is a standardised method of counting weeks to allow for the comparison of

data year after year. The Epi Week calendar of year 2015 is tabulated in Table A.1 (in Appendix A). All data recorded and collected from the investigation form are shown in Table 3.1.

Secondly, the location dataset was collected from investigation form. The human mobility data refer to the data of the patient's movement or whereabouts two weeks prior to their diagnoses date. The name or address of the location was recorded in the investigation form. For example, "3rd miles supermarket" or "Sarawak General hospital" as shown in Figure B.1 (in Appendix B). However, the GPS coordinates (latitude and longitude) were not provided. The coordinate identification process is further discussed in Section 4.2.1.

Thirdly, the meteorological dataset is the data provided by the Sarawak Meteorological Office. The daily total precipitation was recorded in the study area (Kuching weather station) through standard tipping bucket rain gauge (Jong, Lam, D'Aico & Din, 2016), the daily relative humidity, daily maximum surface air temperature and daily minimum surface air temperature were recorded in Kuching weather station through Automatic Weather Stations, which is updated at every minute without human intervention (Najid et al., 2009). The pre-processing for meteorological data is further discussed in Section 4.2.3.

Next, the mosquito dataset is the physical characteristic of the dengue vector, *Ae. Aegypti* mosquito including the species type, the habitat and the flight range, The dataset is collected through literature papers.

3.3.2 Potential Parameters Used

Potential parameters can be derived after analysed the raw data in Table 3.1. The potential parameters are outlined in Table 3.2. Table 3.2 shows the parameters, the respective parameters' symbol and their units of measurement that tied to the human and location nodes. The potential parameters have no unit if there is no notation in the bracket shown in Table 3.2. There are two parameters tied with human nodes, frequency of a patient visited a location, *Fh* and total

duration of stay in a location, Du (s). The parameter tied to location nodes involve the mosquito characteristic for instance, life cycle, survival rate and biting rate; environmental parameters for instance, precipitation, humidity; geographical parameter for instance altitude; and frequency of a location is visited by human. The pre-processing and the quantification of the parameter will be discussed in Section 4.2.

Table 3.2: Potential Parameters Used

No	Node	Potential Parameter
1	Human	Frequency of a patient visited a location, Fh Total duration of stay in a location, Du (s)
2	Location	Life cycle index, Lc Survival rate, S Biting rate, B Total Precipitation amount, Pre (m) Mean relative humidity, K (%) Altitude, Al (m) Frequency of a location is visited by patients, Fl

Based on the research data, the link between the nodes formed when the humans visiting the locations. The formulation of the BDC graph is discussed in the next section.

3.4 Bipartite Graph Formulation

This section outlines the third process in Stage 1 of the BNM-RMF, which discusses the formulation of the complete BDC graph by using the basic building block shown in Figure 3.3. The human and location nodes, and the link formed between human and location nodes are identified from the research data from Section 3.2.1.

3.4.1 Identification of the BDC Network

This study collected a total of 12 Epi week dengue patient data as mentioned in section 3.2.1. We aimed to build six BDC network models where every 2 Epi Week data is used to

formulate one network. Thus, we have six network models and the detailed information is tabulated in Table 3.3.

Table 3.3: Detailed Information of Six BDC Network Models

Group	BDC Network Model	Model	Epi Week	Number of Human Nodes	Number of Location Nodes
1	1	Targeted model	28-29	2 dengue positive patients 6 possible positive patients	19 locations
	2	Validated model	30-31	3 dengue positive patients only	27 locations with 8 new locations
2	3	Targeted model	32-33	9 dengue positive patients 3 possible positive patients	78 locations with 51 new locations
	4	Validated model	34-35	2 dengue positive patients only	81 locations with 3 new locations
3	5	Targeted model	36-37	3 dengue positive patients 7 possible positive patients	98 locations with 17 new locations
	6	Validated model	38-39	7 dengue positive patients only	100 locations with 2 new locations

From Table 3.3, targeted model is constructed to rank the location nodes according to their possible vector density in order to detect the possible risk area where spread dengue. On the other hand, the validated model utilised the data two weeks later from the targeted model to observe whether there is any correlation of the locations' ranking between the targeted and validated model and therefore these two models are grouped into one. In addition, the validated model is used for validation purpose to examine the capability of the targeted model in identifying possible locations with high vector density (to be discussed in chapter five). The number of human nodes in the targeted model includes dengue positive and possible positive patients. To recap, individual will be registered with the doctor when the patient visits the

hospital or clinic due to fever and IgM dengue serology test will be conducted to diagnose for dengue fever. The possible positive patient in this study represents the patient who has negative IgM dengue serology test result (obtained from the investigation form). However, this does not mean the patient rule out dengue infection. The immune response following a dengue infection varies depending on whether the individual has a primary or a secondary dengue infection. Therefore, the diagnosis of dengue is dependent on the phase of the infection as shown in the Figure B.2 (in Appendix B). From the figures, the primary infection show a slow and low titer antibody response compared to the secondary infection. Dengue IgM serology has low sensitivity during the early phase of dengue fever as the virus and IgM antibodies may be at undetectable levels for those who submit a day 5 acute specimen (CDC, 2017). In short, a negative IgM cannot rule out dengue fever. Therefore, the human mobility of the patient who has a negative IgM serology needs to be considered in this model. This individual with a negative IgM serology will be termed as possible positive dengue patient.

The location nodes were visited by both human nodes with dengue positive and possible positive capable to provide a possible new risk location to the targeted model. In validated model, only dengue positive human is included to construct a network model for validation purpose to see whether the location visited by the human with possible positive result is risk area.

3.4.2 Identification of Human Node

In this section, the targeted model of group 1 which utilise data of Epi Week 28 and 29 is used for a case study to demonstrate the formulation process. In this first BDC network model, there are eight unique individual dengue patients and each of them is given a unique code. These eight unique coded individual dengue patients are consequently identified as the eight human nodes of BDC network. They are labelled as H1, H2, H3, H4, H5, H6, H7, H8 where H is the

symbol used for human node (Figure 3.5) and the numerical number 1 and 8 is used to differentiate one human from another.

3.4.3 Identification of Location Node

Based on the 8 human nodes which has identified from section 3.4.2, there are total 19 location nodes has been visited by the 8 human hosts. The identification of the location is determined by the Algorithm 4.1 which will be discussed in the next chapter.

Thus, these locations are labelled as L1, L2, L3, L4, L5, L6, L7, L8, L9, L10, L11, L12, L13, L14, L15, L16, L17, L18 and L19 where L is the symbol used for the location node as shown in Figure B.2 and the numerical number 1 until 19 is used to differentiate one location from another. The actual location nodes are revealed in Figure B.2a and Figure B.2b. Each of the location nodes labelled with their latitude and longitude values.

3.4.4 Identification of BDC Network Links between Nodes

In the first BDC network model, 8 human nodes and 19 location nodes are identified from the research data. This section will discuss the identification of the link which joins the human and location nodes. As mentioned in previous section, the link is formed when H visits L. In order to trace all links, the movement of each unique human host visited each location are extracted from the investigation form. The information of links formed with reference to each location and each human can be identified and is revealed in Table 3.4 and Figure 3.5. As mentioned earlier, the distance between each location node is not provided from the investigation form, therefore the distance matrix to calculate the distance between each node is discussed in Algorithm 4.2 in Section 4.2.1. From Table 3.4, we declare the nodes with the

algorithm generated location identification (ID) code such as L1, L2, ..., L19 based on Algorithm 4.1 in Section 4.2.1.

From the Table 3.4, we observed that one human node can visit more than one location, for example, H2 visited three locations for example L2, L10 and 15. Also, one location can be visited by more than one human host, for example, L7 is visited by H6 and H7 on different dates. From Table 3.4, there is total 20 unique links formed when a human visited a location, in other words, one H visit one L. These 20 links are eventually the dataset which will be used in the formulation of the BDC network model that will be discussed in the next chapter.

Table 3.4: Movement of Each Human Host

H node labelled	Date	L node visited
H1	25-Jun-15	L1
	25-Jun-15	L9
H2	30-Jun-15	L2
	30-Jun-15	L10
	30-Jun-15	L15
H3	1-Jun-15	L3
	1-Jul-15	L11
	17-Jun-15	L16
H4	15-Jun-15	L4
	20-Jun-15	L12
	23-Jun-15	L17
	27-Jun-15	L19

H node labelled	Date	L node visited
H5	8- Jul-15	L5
H6	8- Jul-15	L6
	23-Jun-15	L7
	3- Jul-15	L18
H7	9-Jul-15	L7
	25-Jun-15	L13
H8	6-Jul-15	L8
	22-Jun-15	L14

3.4.5 Complete Bipartite Dengue Contact Graph for BDC Network

Based on the nodes and links identified, the complete bipartite dengue contact graph, BDC graph of the first BDC network model is formulated. This is presented in Figure 3.5. BDC

graph is the bipartite graph, BDC_{DEN_KCH} for the first BDC network. The set of bipartite nodes refers to H and L subsets and E refer to the set of links formed between set H and set L . Set H , the human nodes, consists of eight elements,

$$H = \{H1, H2, H3, H4, H5, H6, H7, H8\}$$

and set L , the set of location nodes, consists of nineteen elements,

$$L = \{L1, L2, L3, L4, L5, L6, L7, L8, L9, L10, L11, L12, L13, L14, L15, L16, L17, L18, L19\}$$

Set E , the set of links that join elements of H and D , has 20 elements,

$$E = \left\{ \begin{array}{l} H1L1, H1L9, H2L2, H2L10, H2L15, \\ H3L3, H3L11, H3L16, H4L4, H4L12, \\ H4L17, H4L19, H5L5, H6L6, H6L7, \\ H6L18, H7L7, H7L13, H8L8, H8L14 \end{array} \right\}$$

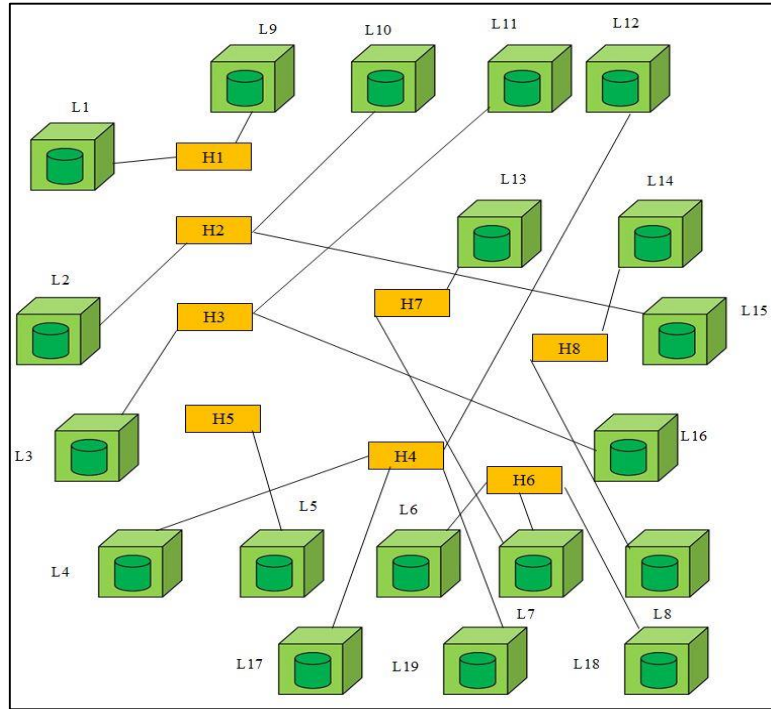


Figure 3.5: Complete Bipartite Dengue Contact Graph for Dengue Patient

From Figure 3.5, the links formed when one location node is visited by one human node. For illustration, '*H1L1*' represents the link formed when a dengue patient with label H1 visited the location with label L1 node on 25 June 2015 as shown in Table 3.4.

As a summary, BDC graph of BDC network is a graph of 8H by 19L with 20 edges. The degree of each location node is {1, 1, 1, 1, 1, 1, 2, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1}. On the other hand, the degree of each human node is {2, 3, 3, 4, 1, 3, 2, 2}. The sum of the degrees of the location or human nodes is 20 which is also the total number of edges in BDC graph.

In the next chapter, the BDC graph needs to be computed into values to formulate a complete BDC network as illustrated in Stage 2 of BNM-RMF. The BDC graph will act as an input to Stage 2 which can quantify the nodes' parameters (in Process 2.2 and 2.3) and the link weight (in Process 2.4). The execution of Stage 2 is discussed in Chapter four.

3.5 Summary

This chapter discussed the execution of Stage 1 of the research methodology, BNM-RMF which covered the sub-processes include current research scenario understanding, graph structure representation and formulation of bipartite graph for the BDC network model this study intends to formulate. The process of understanding the current research scenario achieved through the literature review, understanding and analyse the raw data and gathered opinions from experts. As an outcome, the quantity of interest with variables and assumptions concerned is identified. From the formulation of basic building block, a graphical representation is computed. Finally, BDC graph is formulated and the formal definition of the graph is established by using the research data. Next chapter will discuss the process of developing the BDC network model.

CHAPTER 4

MODEL CONSTRUCTION

4.1 Introduction

Having established the bipartite graph in Chapter 3, this chapter explains the processes of Stage 2 of BNM-RMF. This chapter now quantifies each node (location and human nodes), and quantifies the links between the nodes. The quantification will then enable the use of search algorithms for hotspot detection, in particular the HITS algorithm, which is based on a ranking of location and human nodes.

4.2 Data Pre-processing

This section presents the Process 2.1 as shown in Figure 1.1 which prepares the processed research data for quantification. The data obtained is raw and untidy, this data pre-process is important to determine the location and human nodes' values. There are three sub-processes in this data pre-processing process that includes pre-process of data used for location node; pre-process of data used for human node; and pre-process of the parameter used in link weight quantification.

4.2.1 Pre-processing of Location Node

The location addresses or names are collected through the investigation form and the GPS coordinate is not provided as mentioned in Section 3.2.1. A location without GPS coordinate is raw and impossible to be processed in this study. Therefore, the location data need to be digitalized and the distance between each location nodes need to be calculated in order to avoid data redundancy in BDC network. The pre-process of location nodes is summarised in Figure 4.1.

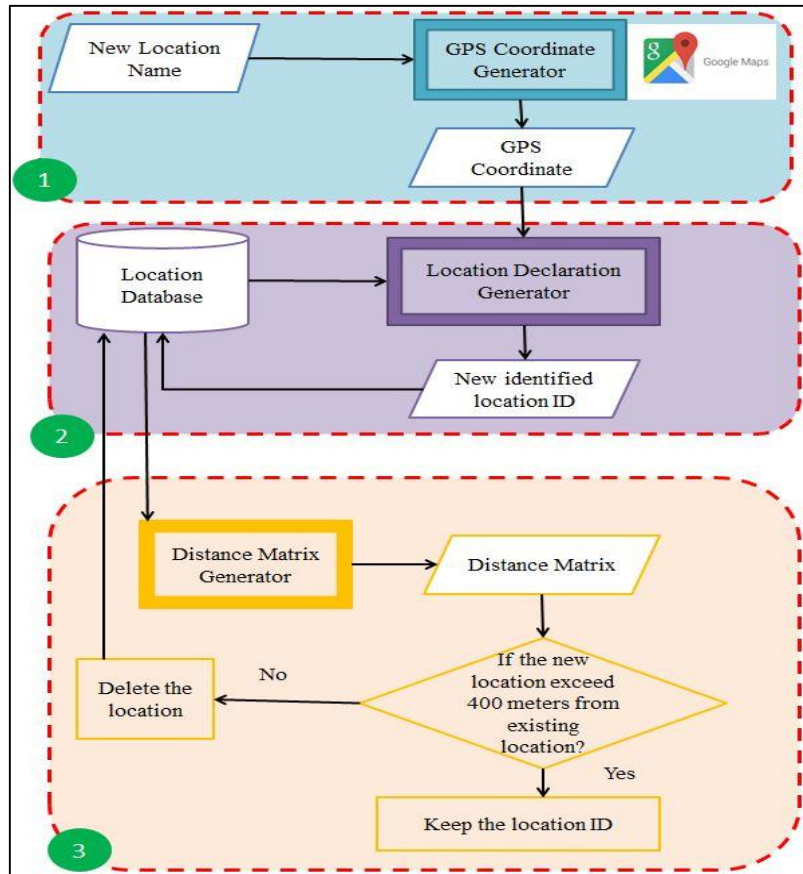


Figure 4.1: Pre-processing of the Location Node

The first step is to change the location name into GPS coordinates by using Google Maps (<https://www.google.com.my/maps?source=tldsi&hl=en>). For example, one location name is collected from the investigation form as “No 40A, Lorong 6D, Taman Malihah, Kuching”, the location search is done by using the Google Maps website and shown in Figure 4.2. The GPS coordinates can be retrieved in the red rectangle box as 1.5567, 110.2475 as shown in the Figure 4.2.

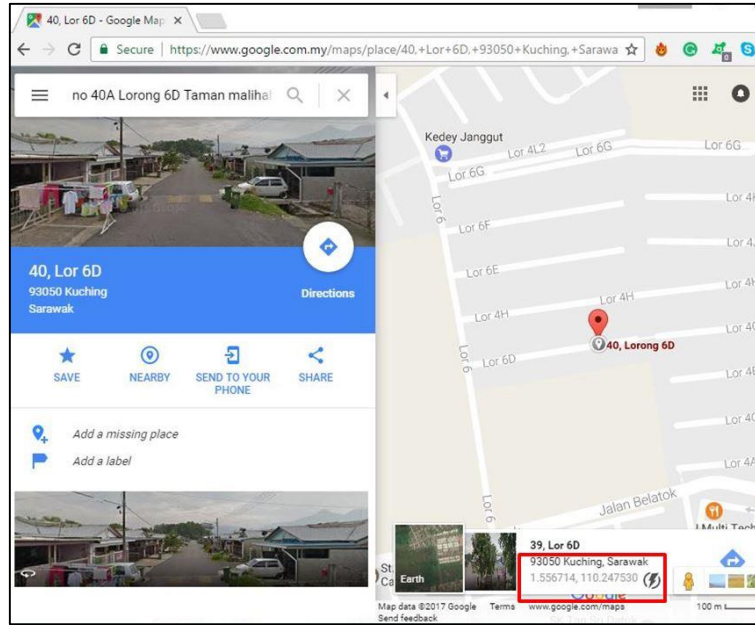


Figure 4.2: Declare GPS Coordinates

Secondly, declaration of location identification (ID) in the study is important. The output from the first step in Figure 4.1 is used in location declaration generator. This generator is developed using R programming and aims to declare new location ID in the database. Algorithm 4.1 creates a data frame named Kch_City to store the GPS coordinate of the location. When the database or the Kch_City data frame is empty, a data frame will be created and the new location's GPS coordinate store in the data frame else, the new location's GPS coordinate will be added into the existing data frame with existing location's GPS coordinates. Algorithm 4.1 will call a user-defined function named DistanceMatrix () to calculate the distance of the new incoming location and the existing locations from the data frame. The distance between the existing location in the data frame cannot overlap or redundant with the new incoming location node. According to the characteristic of the dengue vector discussed in Chapter 2, the maximum flight range of the *Aedes* mosquito is 400 meters, therefore, a new location can be declared in the database if the distance between new location node and current location (in database) is greater than 400 meter.

Algorithm 4.1: Declare location node

```
IF (Kch_City is empty)
  Create DataFrame Kch_City
  WRITE Kch_City[1]=Loc_GPS[1]
ELSE
  ASSIGN k=(CHECK Length(Kch_City))
  WRITE Kch_City[k+1]= new_Loc_GPS
ENDIF
  CALLING function DistanceMatrix (Kch_City)
IF (DistanceMatrix>400)
  DELETE Kch_City[k+1]
ELSE
  KEEP new_Loc_GPS Kch_City
EXIT
ENDIF
```

The third step in the Figure 4.1 reads all the locations' GPS coordinates of the data frame and passes it to the distance matrix generator. The generator was also developed by using R and it is summarised in Algorithm 4.2. If the user-defined function in Algorithm 4.1 returns the distance exceed 400 meters, the new incoming location ID will remain in the data frame and this creates a new location ID that extends the location database. If the distance of this new location didn't exceed 400 meters with any one of the locations in the data frame, this new location will be deleted from the data frame.

The Algorithm 4.2 explained the function DistanceMatrix(). This function returns a matrix of distances between geographic points. The row and column names are given as "L1" for location node 1, "L2" for location node 2 and so on. Another function named GeoLocDistance() within the function will returns a distance vector where the first value is the distance between g1[[1]] and g2[[1]], the second value is the distance between g1[[2]] and g2[[2]] and so on until all the GPS pairs is calculated. Each g1[[x]] or g2[[x]] is a list with named elements "index", "lat" and "lon" for example g1=list(list("index" = 1, "lat"=1.2301 and "lon"=110.3928), ("index" = 3, "lat"= 1.3999 and "lon"=110.3251)). This user-defined

function will return a matrix, therefore the g1 and g2 in the Algorithm 4.2 consists of identical GPS coordinates which the diagonal of the matrix are always 0 because the distance between location node L1 and L1 are 0. ASSIGN Kch_City\$index = 1 to num_geopoints is to ensure the calculation is done for the upper triangle of points.

Algorithm 4.2: Function to create a distance matrix to calculate the GPS distance

```
DistanceMatrix=function (Kch_City)
{
  GeoLocDistance=function (g1,g2)
  IF(g1$index > g2$index)
    Returns 0
  ELSE
    Return the distance vector
  ENDIFELSE
  ASSIGN num_geo=num_row(Kch_City)
  ASSIGN Kch_City$index = 1 to num_geopoints

  CREATE List list.geo = by(Kch_City[, c("index", "lat",
    "long")], 1 to num_geo, function (x){return(list(x))})

  CREATE Matrix mat.dist = LowerUpperTri(outer (list.geo,
    list.geo, GeoLocDistance), "lower")

  ASSIGN rownames(mat.dist)=Kch_City$name
  ASSIGN colnames(mat.dist)=Kch_City$name
  return (mat.dist)
}

LowerUpperTri=function (m, triangle)
{
  IF(nrow(m)not equal ncol(m) stop("Must be a square matrix"))
  IF(tolower(triangle equal "lower"))
    ASSIGN tri = lower.tri(m)
  ELSE IF (tolower(triangle equal "upper"))
    ASSIGN tri = upper.tri(m)
  ELSE stop("triangle must be set as lower or upper")
  ASSIGN m[tri]=t(m)[tri]
  return (m)
}
```

Next is to create a list of lists to keep all the lists into one list. After that, create a matrix to keep the distance, this matrix store the return value from LowerUpperTri() function. This function checks whether the triangle is upper or lower. If it is an upper triangle, replace

the upper triangle with its lower triangle and vice versa. This function replaces either the upper or lower triangle of the matrix. Next, assign the row name and column name in the matrix with the name provided by the data frame from Algorithm 4.1.

As mentioned in Section 3.4.2, the first BDC network, which utilises data during Epi Week 28 and 29 will be used to demonstrate the quantification process. Table 4.1 demonstrates the distance matrix created through Algorithm 4.1 and the distances of 19 location nodes is shown. The element located in the first row and second column indicates the distance between L1 and L2 which is 30.23 km. Once there is any element in the matrix greater than 400 meters (or 0.4 km), the new incoming location node will be deleted in the data frame. However, if the distance is less than 400m, the new incoming location will be declared as a new valid location node in this study. The Kch_City data frame is expanding in this study.

Table 4.1: Distance Matrix between all Location Nodes in BDC Network
a) Column 1-10

	L1	L2	L3	L4	L5	L6	L7	L8	L9	L10
L1	0	30.23	34.28	34.88	23.12	45.16	39.57	36.69	40.43	34
L2	30.23	0	4.26	4.75	8.51	14.95	13.47	6.81	10.24	5.38
L3	34.28	4.26	0	2.52	11.79	10.95	10.28	4.47	6.17	5.5
L4	34.88	4.75	2.52	0	13.13	10.42	12.41	2.21	5.94	3.23
L5	23.12	8.51	11.79	13.13	0	22.67	16.74	15.28	17.9	13.68
L6	45.16	14.95	10.95	10.42	22.67	0	12.62	9.04	4.78	12.32
L7	39.57	13.47	10.28	12.41	16.74	12.62	0	13.43	10.16	15.63
L8	36.69	6.81	4.47	2.21	15.28	9.04	13.43	0	5.15	3.32
L9	40.43	10.24	6.17	5.94	17.9	4.78	10.16	5.15	0	8.43
L10	34	5.38	5.5	3.23	13.68	12.32	15.63	3.32	8.43	0
L11	7.66	23.48	27.66	27.98	17.38	38.38	34.13	29.66	33.72	26.8
L12	13.99	18.13	22.39	22.38	13.66	32.8	29.96	23.89	28.25	20.88
L13	40.1	11.82	7.94	9.62	16.97	8.98	3.78	10.22	6.44	12.83
L14	13.49	30.66	33.8	35.3	22.16	44.42	35.51	37.45	39.73	35.65
L15	37.8	8.16	3.93	5.21	14.9	8.02	7.61	5.85	3.58	8.4
L16	24.93	8.08	12.11	11.46	9.72	21.72	21.44	12.69	17.4	9.55
L17	20.7	43.58	47.82	47.43	39.16	57.64	55.69	48.6	53.37	45.34
L18	15.85	17.93	21.19	22.59	9.45	31.97	24.31	24.73	27.22	22.94
L19	35.26	5.18	2.91	0.47	13.59	10.11	12.61	1.74	5.71	3.12

Table 4.1 continued
b) Column 11-19

	L11	L12	L13	L14	L15	L16	L17	L18	L19
L1	7.66	13.99	40.1	13.49	37.8	24.93	20.7	15.85	35.26
L2	23.48	18.13	11.82	30.66	8.16	8.08	43.58	17.93	5.18
L3	27.66	22.39	7.94	33.8	3.93	12.11	47.82	21.19	2.91
L4	27.98	22.38	9.62	35.3	5.21	11.46	47.43	22.59	0.47
L5	17.38	13.66	16.97	22.16	14.9	9.72	39.16	9.45	13.59
L6	38.38	32.8	8.98	44.42	8.02	21.72	57.64	31.97	10.11
L7	34.13	29.96	3.78	35.51	7.61	21.44	55.69	24.31	12.61
L8	29.66	23.89	10.22	37.45	5.85	12.69	48.6	24.73	1.74
L9	33.72	28.25	6.44	39.73	3.58	17.4	53.37	27.22	5.71
L10	26.8	20.88	12.83	35.65	8.4	9.55	45.34	22.94	3.12
L11	0	6.34	34.13	17	31.38	17.5	22.3	12.78	28.33
L12	6.34	0	29.43	21.29	26.24	11.41	25.74	12.94	22.69
L13	34.13	29.43	0	37.27	4.44	19.89	55.12	25.44	9.71
L14	17	21.29	37.27	0	36.43	28.71	33.88	12.73	35.75
L15	31.38	26.24	4.44	36.43	0	16.02	51.73	24.05	5.28
L16	17.5	11.41	19.89	28.71	16.02	0	35.97	16.65	11.7
L17	22.3	25.74	55.12	33.88	51.73	35.97	0	34.87	47.67
L18	12.78	12.94	25.44	12.73	24.05	16.65	34.87	0	23.04
L19	28.33	22.69	9.71	35.75	5.28	11.7	47.67	23.04	0

4.2.2 Pre-Process of Human Node

The human node refers to the patient identity and this can be obtained from the investigation form. In order to protect the patient confidentiality, patient identity replaced by an algorithm-generated ID. The investigation provided the patient's name, age, gender and personal details that include residential address. This information can help to identify the human node ID and avoid redundancy. The first BDC network model consists of 8 human nodes. Thus, the human nodes are identified as H1, H2, H3, H4, H5, H6, H7 and H8.

4.2.3 Pre-Process of Parameter

Parameters such as temperature, humidity, precipitation and altitude are collected for node quantification. Pre-processing of these parameter values is necessary to prepare the parameter in building the useful model to quantify the nodes in the following sections.

The human mobility data capture the patients' movement two weeks before the onset date. In order to observe the effect of the parameters towards the environment, average temperature two weeks ago at that particular location need to be calculated. For instance, the onset date for the first patient (denoted as H1) was on 9 July 2015, the human mobility data capture the movement within two weeks ago, which is 25 June - 8 July 2015. The H1 visited L1 on 25 June 2015. Thus, the average temperature 14 days ago, which was between 11 June and 24 June 2015 is calculated with the equation (4.1). In equation 4.1, the i represent the day before the human mobility date start, $i = 14$ represents the 14 days ago from the human mobility date start and decrease to $i = 1$ which is one day before the human mobility date. k in equation (4.1) represents temperature, humidity, precipitation or altitude. The detailed timeline is illustrated in Figure 4.3.

$$Average_k = Average (\sum_{i=14}^1 k_i) \quad (4.1)$$

$$\begin{aligned} Average_k &= Average (\sum_{i=14}^1 k_i) \\ &= Average (Temp_{14} + Temp_{13} + \dots + Temp_2 + Temp_1) \\ &= Average (34.2 + 34.3 + 34.8 + 33.3 + 29.9 + 33.9 + 34.9 \\ &\quad + 33.6 + 34.2 + 32.5 + 33.0 + 34.6 + 36.4) \\ &= 33.8 \end{aligned}$$

From equation (4.1), temperature is used to represent k which also an example to calculate average temperature. This average temperature, 33.8 for the location L1 which the onset date of human with node ID H1 is 9 July 2015. The average temperature from 11 June to 24 June 2015 has a significant effect on that particular location. Likewise, other parameters including humidity, precipitation and altitude can be calculated using equation (4.1) with k =humidity/ k =precipitation/ k =altitude. After processing the data, next section outlines the process to quantify the parameter of the location, human nodes and the edges that link the nodes.

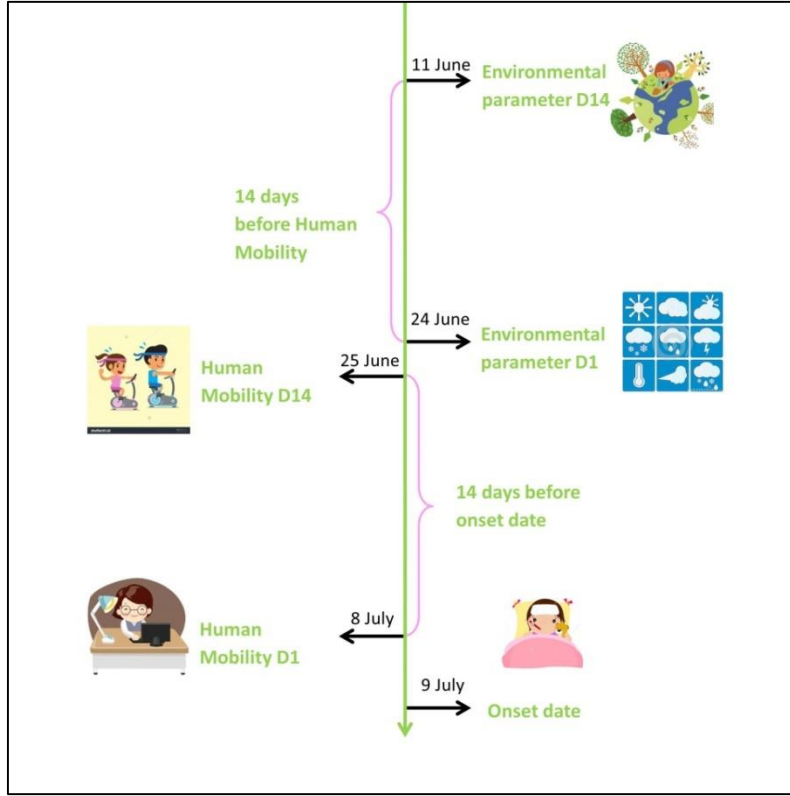


Figure 4.3: Timeline for Human Node 1

4.3 Quantification of Location Node Parameter

This section highlights the process 2.2 of BNM-RMF (Figure 1.1) which quantify the parameter values of node-type-1, the node refers to the location nodes. The potential parameters for the location and environmental properties listed in Table 3.1 will be considered. The possible parameters to be considered at the initial stage are listed in Figure 4.4(a). Majority of the parameters listed in Figure 4.4(a) have their data available in the raw data except for life cycle parameter, survival rate, biting rate and the number of times a location is visited by humans.

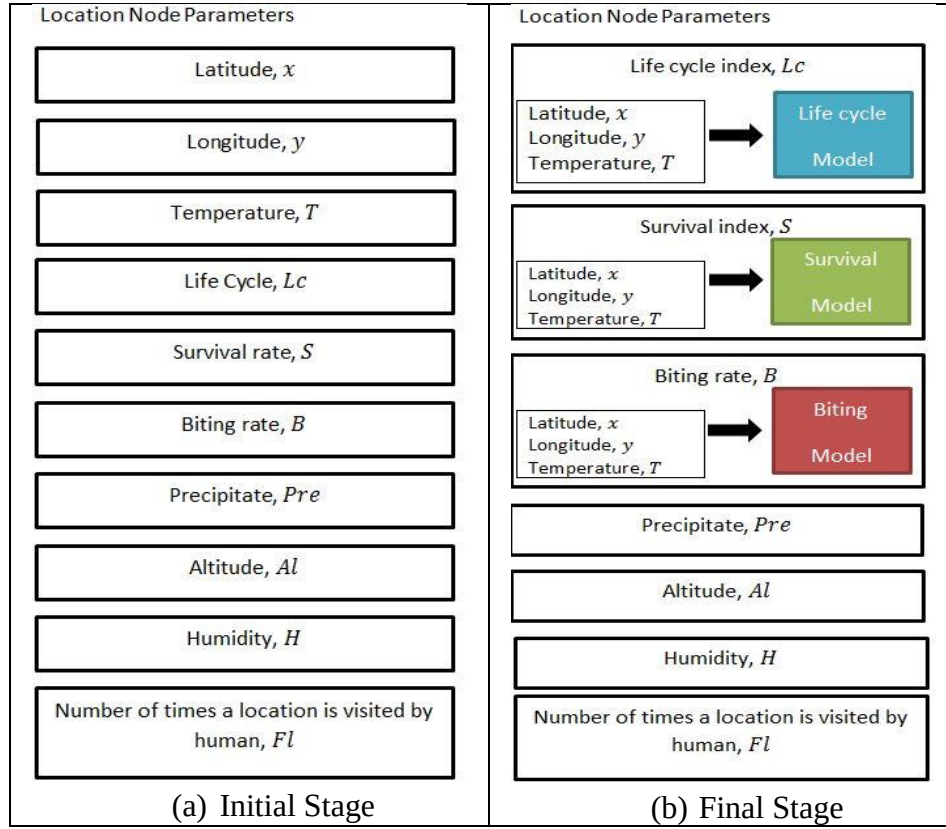


Figure 4.4: Parameters of BDC Network Location Node at the Process 2.2 in BNM-RMF

It is established that the mosquito vectors hardly move far away from the locality where breed them. Thus, these vector activities taken place in the particular locality will affect the dengue transmission. The activities as below have been taken into account in this study to model the dengue transmission.

- i. Number of days to complete a mosquito vector life cycle which affect the dengue transmission rate. Thus, vector life cycle duration model need to be constructed.
- ii. Mosquito survival rate also affects the dengue transmission. As the higher the survival rate, the higher the population and hence the higher the transmission rate. Thus, the vector survival rate model is incorporated in the contact network.

- iii. Mosquito biting rate affects the transmission of dengue, as the more frequent the mosquito bites the human host, the higher the probability of dengue spread. Thus, the vector biting rate model is constructed.

The quantification of these parameters will be presented in the Sections 4.3.1 until 4.3.4.

4.3.1 Quantification of Vector Life Cycle Duration Parameter, L_c

Vector life cycle duration parameter in this study measures the duration of development from mosquito egg hatching to adult at every locality in BDC network. Due to its known dependence on temperature, the life cycle of *Aedes* mosquito plays an important role in order to understand the effects of environmental property on the dengue transmission (Carrington, Armijos, Lambrechts, & Scott, 2013). However, there is no direct life cycle duration data that has been recorded from any data source. Nonetheless, the vector life cycle duration could naturally be implied by the attributes of a location node and the environmental properties enclosed. These attributes that are reflected through the location physical characteristics and condition, could be utilised to imply the life cycle duration of *Aedes* mosquito at one locality.

Life cycle will be quantified as a function of the temperature attribute with particular latitude and longitude. In the work of Eze (2013), polynomial curve fitting using polyfit tool in MATLAB on data gathered from the related experimental research was used to quantify the parameters. Thus, this study applies the same technique to quantify the life cycle duration parameter. Related experimental researches suggested that temperature is the main driver of the vector transition. The existing experimental research by Carrington et al. (2013) comprises of the number of days to complete one life cycle of *Aedes* mosquito temperature range from 16 – 39 degree Celsius as tabulated in Table 4.2. The experiment is conducted to examine the effects of daily fluctuations on larval development time of *Ae. Aegypti* collected from Thailand. This

mosquito is common in Asian countries including Malaysia (Benedict, Levine, Hawley, & Philip, 2008). This data is utilised to build the vector life cycle duration model. The life cycle duration is negatively associated with the temperature. Vector life cycle duration parameter is thus termed as the vector life cycle index with symbol L_c . This vector life cycle index which valued between 0 and 1, is defined as the measurement of life cycle duration of *Aedes* mosquito. In this study, the process to generate life cycle model is depicted in Figure 4.5.

Table 4.2: Source data of Mean Development Duration of *Ae. Aegypti* (Carrington et al., 2013)

Temperature, T (°C)	Mean development time from egg hatching to pupation, D (days)
16	31.704
20	16.0383
26	9.15326
30	6.45608
35	5.85143
37	7.79076
38	7.75892
39	12.9355

The generation of life cycle index shown in Figure 4.5 introduces the data normalisation as the first step. In order to reduce the problem intensity, the temperature, T in Table 4.2 is normalised by centring it at a zero mean ($\text{mean}(T)$) and scaling to a unit standard deviation($\text{std}(T)$) using the equation (4.2).

$$t = \frac{t - \text{mean}(T)}{\text{std}(T)} \quad (4.2)$$

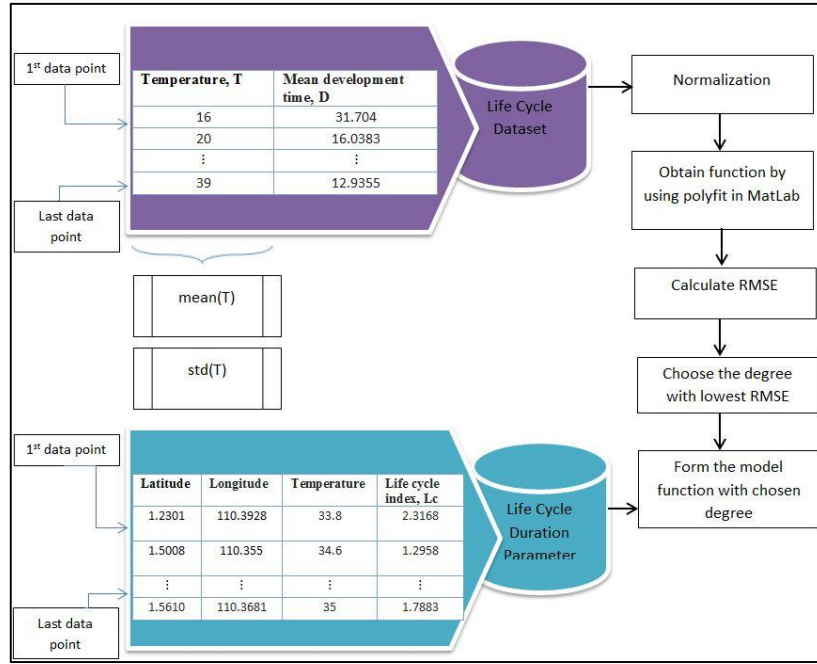


Figure 4.5: Generation of Life Cycle Index, L_c

After calculating the mean and standard deviation, the temperature data points can be normalised which the T consists of [16, 20, 26, 30, 35, 37, 38, 39], mean of the T is 28.8571 and the standard deviation is 8.5718. After normalisation, temperature data became [-1.4999, -1.0333, -0.3333, 0.1333, 0.7166, 0.9500, 1.0666].

Next, polynomial curve fitting using polyfit tool in MATLAB is done to return the coefficients for the polynomial $f(x)$ of degree n that is a best fit for the data temperature, T . In theory, the degree n is used for data points $n + 1$. This data consists of 8 data points, thus degree 7 will be the best fit. However, an iteration loop is implemented to fit the data from degree 1 until n to identify the best fit.

The third step is calculating the root-mean-square-error (RMSE) for all degrees. And choose the degree with the lowest RMSE value. The RMSE for all degree is listed in Table 4.3. Thus, degree of 6 is chosen to build the model in MatLab with polyfit function.

Table 4.3: RMSE Values for 7 Degrees in Quantifying Life Cycle Parameter

Degree	RMSE
1	3.2174
2	0.8218
3	0.5501
4	0.2547
5	0.1290
6	3.44×10^{-15}
7	4.11×10^{-15}

The life cycle duration model, L_c a temperature-dependent model is formulated as a function of t where t represents temperature with degree of 6 as in equation (4.3).

$$L_c(t) = -0.633t^6 - 0.786t^5 + 1.488t^4 + 1.153t^3 - 0.408t^2 - 0.758t - 0.504 \quad (4.3)$$

The final step is putting in the collected data points into the model to obtain the vector life cycle duration, L_c in equation (4.3). Table 3.4 from previous chapter records the location nodes visited by the human node. The average temperatures are the processed data according to the process explained in section 4.2.3 with respect to the particular date and location. The average temperature, Temp is substituted into the life cycle duration model to obtain the life cycle index, L_c . The results of the *life cycle*, L_c is tabulated in Table 4.4. The life cycle index is inversely proportional to the dengue transmission rate where the shorter the time taken for a complete life cycle leads to an increase in the dengue infection rate. This implies that the life cycle duration model is negatively proportional to the life cycle index. Therefore, $\frac{1}{L_c}$ will be used in the link weight quantification later.

Table 4.4: Life Cycle Index, *Lc* of Each Location Nodes

Latitude, Longitude	L node	Visited by	Temp(°C)	<i>Lc</i> (days)	$\frac{1}{Lc}$
1.2301, 110.3928	L1	H1	33.8	5.0134	0.1995
1.5008, 110.3550	L2	H2	34.6	5.5030	0.1817
1.5350, 110.3373	L3	H3	32.2	4.9800	0.2000
1.5436, 110.3583	L4	H4	33.7	4.9742	0.2008
1.4288, 110.3280	L5	H5	34.7	5.5845	0.1791
1.6338, 110.3311	L6	H6	34.7	5.5845	0.1791
1.5567, 110.2475	L7	H6	33.6	4.9400	0.2024
		H7	34.4	5.3528	0.1868
1.5610, 110.3681	L8	H8	35	5.8514	0.1709
1.5906, 110.3324	L9	H1	33.8	5.0134	0.1995
1.5375, 110.3867	L10	H2	34.6	5.5030	0.1817
1.2965, 110.4124	L11	H3	32.5	4.8942	0.2043
1.3525, 110.4245	L12	H4	33.9	5.0578	0.1977
1.5735, 110.2771	L13	H7	33.8	5.0134	0.1994
1.2365, 110.2718	L14	H8	33.7	4.9742	0.2010
1.5630, 110.3156	L15	H2	34.6	5.5030	0.1817
1.4548, 110.4114	L16	H3	32.5	4.8942	0.2043
1.1709, 110.5692	L17	H4	33.7	4.9742	0.2010
1.3456, 110.3084	L18	H6	34.7	5.5845	0.1791
1.5473, 110.3604	L19	H4	33.9	5.0578	0.1977

4.3.2 Quantification of Vector Survival Parameter, *S*

The survival parameter measures the survival probability at a locality as an indication of vector survival rate at one locality. This parameter is included to account the importance of dengue transmission as one of the significant contributors to the vector hotspot (Lambrechts et al., 2011). The higher vector survival contributes to higher vector capacity in one locality. Thus, vector survival rate in one locality directly influence the dengue transmission rate. This is because the mosquitoes lived longer and were more likely to be infected with the dengue virus

(Carrington et al., 2013; Lambrechts et al., 2011). *Aedes* mosquitoes are small-bodied ectotherms, which their survival depend on temperature. The mosquito survival rate is positively associated with the temperature.

There is no documented source about vector survival data. The vector survival data could naturally be implied by the attributes of a location node and the environmental properties enclosed. Temperature increases are associated with a higher survival rate within the *Aedes* mosquito (Rueda et al., 1990; Tun-Lin et al., 2000; Carrington et al., 2013). Similarly, polynomial curve fitting technique by using polyfit tool in MATLAB on data gathered from the related experimental researches is used to quantify the vector survival parameter. The existing experimental research by Tun-Lin, Burkot and Kay (2000) consists of the survival rate of *Aedes* mosquito and temperature range from 10 – 34 degree Celsius as tabulated in Table 4.5. This data is utilised to build the vector survival model. Vector survival parameter is thus termed as the *vector survival index* with symbol S . In this study, the process to generate *vector survival index* is same with the process to generate the *life cycle index* as depicted in Figure 4.5.

Table 4.5: Source Data of Survival (%) of *Ae. Aegypti* (Tun-Lin et al., 2000)

Temperature, T (°C)	Vector Survival, Sv (%)
10	0
15	23.5
20	90
25	88
27	93
30	88
34	67

The first step is to normalise the temperature, T in Table 4.5 as shown in Figure 4.5. After calculating the mean and standard deviation, the temperature data points can be normalised which the T consists of [10, 15, 20, 25, 27, 30, 34], the mean of the T is 23 and the

standard deviation is 8.4853. After normalisation, temperature data became [-1.5321, -0.9428, -0.3536, 0.2357, 0.4717, 0.8250, 1.2964].

Next, polynomial curve fitting using polyfit tool in MATLAB is done to return the coefficients for the polynomial $f(x)$ of degree n that is a best fit for the data temperature, T . In theory, the n degree is used for $n + 1$ data points. This data consists of 7 data points, thus degree of 6 will be the best fit. However, an iteration loop is implemented to fit the data from degree 1 until n to identify the best fit. The third step is calculating the RMSE for all degrees. And choose the degree with the lowest RMSE value. The RMSE for all degree is listed in Table 4.6. Thus, degree of 6 is chosen to build the model in MatLab with polyfit function.

Table 4.6: RMSE Values for 6 Degrees in Quantifying Survival Index

Degree	RMSE
1	0.6051
2	0.2610
3	0.2366
4	0.1752
5	0.0602
6	3.1264×10^{-15}

The vector survival is positively associated with the dengue transmission rate. Therefore, the vector survival model, a temperature-dependent model is formulated as a function of t with degree 6 as in equation (4.4).

$$S(t) = 1.3908t^6 - 0.2951t^5 - 3.8642t^4 + 1.3217t^3 + 1.2971t^2 - 0.1412t + 0.591 \quad (4.4)$$

The final step is to substitute the collected data points into the model in equation (4.4) to obtain the vector survival index, S . The average temperature with respect to the particular date and location is substituted into the vector survival model to obtain the vector survival index, S . The results of *Survival index*, S is tabulated in Table 4.7.

Table 4.7: Survival Index, S of Each Location Node

L node	Visited by	Temp(°C)	S (%)
L1	H1	33.8	65.1060
L2	H2	34.6	78.1297
L3	H3	32.2	68.8743
L4	H4	33.7	64.4377
L5	H5	34.7	80.9369
L6	H6	34.7	80.9369
L7	H6	33.6	63.9387
	H7	34.4	73.3948
L8	H8	35	91.3251
L9	H1	33.8	65.1060
L10	H2	34.6	78.1297
L11	H3	32.5	66.5415
L12	H4	33.9	65.9558
L13	H7	33.8	65.1060
L14	H8	33.7	64.4377
L15	H2	34.6	78.1297
L16	H3	32.5	66.5415
L17	H4	33.7	64.4377
L18	H6	34.7	80.9369
L19	H4	33.9	65.9558

4.3.3 Quantification of Vector Biting Parameter, B

An increase in the biting rate of *Aedes* mosquito also increases the force of dengue transmission within the community (Scott et al., 2000). The mosquitoes' behaviour can be affected by the climate change. One of the crucial activities like biting is significant in contributing to the dengue transmission (Maciel-de-Freitas et al., 2008; Phaijoo & Gurung, 2015; Wesolowski et al., 2015). Scott et al. (2000) associated the temperature and blood-feeding frequency of female *Ae. Aegypti*. This blood-feeding frequency indicates the number of blood meals of the mosquito takes, this can also be referred as the number of mosquito bites. A linear regression model of the blood feeding as in equation (4.5).

$$B(T) = 0.03T + 0.66 \quad (4.5)$$

The $B(T)$ is biting rate model and T in this model represents the average weekly temperature with the range from 21°C to 32°C in the study area, Thailand. Equation (4.5) represents the total mosquito biting rate per week. However, the unit of biting rate in this study is daily mosquito biting rate. Therefore the equation (4.5) is dividing by 7 to transform it into a daily biting rate. The biting parameter model is modified into the equation (4.6).

$$B(T) = \begin{cases} 0.004286T + 0.09429, & 21^\circ\text{C} \leq T \leq 32^\circ\text{C}, \\ 0.8, & \text{otherwise.} \end{cases} \quad (4.6)$$

The model from Scott et al. (2000) is applied when the temperature ranged from 21 until 34, if the temperature out of this range, the biting parameter will be considered as a baseline value, 0.8 (Scott et al., 2012 in Luz et al. (2011, Supplemental material p.1)).

Lastly, putting in the collected data points into the biting model to obtain the vector biting index, B . The average temperature with respect to the particular date and location is substituted into the biting model (equation (4.6)) to obtain the vector biting index, B . The results of *Biting index*, B is tabulated in Table 4.8. Table 4.8 shows the Biting index for all location nodes is 0.8 as the temperatures fall out of the temperature range from 21 to 32.

Table 4.8: Biting Index, B of Each Location Nodes

L node	Visited by	Temp(°C)	$B (\frac{1}{day})$
L1	H1	33.8	0.8
L2	H2	34.6	0.9
L3	H3	32.2	0.8
L4	H4	33.7	0.8
L5	H5	34.7	0.8
L6	H6	34.7	0.8
L7	H6	33.6	0.8
	H7	34.4	0.8
L8	H8	35	0.8
L9	H1	33.8	0.8
L10	H2	34.6	0.8
L11	H3	32.5	0.8
L12	H4	33.9	0.8
L13	H7	33.8	0.8
L14	H8	33.7	0.8
L15	H2	34.6	0.8
L16	H3	32.5	0.8
L17	H4	33.7	0.8
L18	H6	34.7	0.8
L19	H4	33.9	0.8

4.3.4 Quantification of Number of Times a Location is visited by Human, Fl

This study includes the number of times a location is visited by the human patient as one of the parameters for location node in the BDC network to capture the effect of the visiting of the dengue patient to a location. This parameter is termed as Fl which can be referred as the frequency or number of times a location is visited by one human.

Fl captures the number of times a location node is visited by any human node of BDC network. A link matrix, $Link_Matrix_{BDC\ network}$, is created to record the Fl (Eze, 2013; Liew, 2016). This link matrix is defined in the equation and presented in Table 4.9.

$$Link_Matrix_{BDC\ network}(L_i H_j) = \begin{cases} 1, & \text{if } L_i \text{ is linked to } H_j \\ 0, & \text{if } L_i \text{ is not linked to } H_j \end{cases} \quad (4.7)$$

where L_i =location node i of BDC network

H_j =human node j of BDC network

$i \in \{1,2, \dots, 19\}$ and $j \in \{1,2, \dots, 8\}$

Table 4.9: Link Matrix of BDC Network

	H1	H2	H3	H4	H5	H6	H7	H8
L1	1	0	0	0	0	0	0	0
L2	0	1	0	0	0	0	0	0
L3	0	0	1	0	0	0	0	0
L4	0	0	0	1	0	0	0	0
L5	0	0	0	0	1	0	0	0
L6	0	0	0	0	0	1	0	0
L7	0	0	0	0	0	1	1	0
L8	0	0	0	0	0	0	0	1
L9	1	0	0	0	0	0	0	0
L10	0	1	0	0	0	0	0	0
L11	0	0	1	0	0	0	0	0
L12	0	0	0	1	0	0	0	0
L13	0	0	0	0	0	0	1	0
L14	0	0	0	0	0	0	0	1
L15	0	1	0	0	0	0	0	0
L16	0	0	1	0	0	0	0	0
L17	0	0	0	1	0	0	0	0
L18	0	0	0	0	0	1	0	0
L19	0	0	0	1	0	0	0	0

Fh measures the number of times that a human visited a location and the details of quantification for Fh discussed in section 4.4.2. Fl can be defined as in the equation (4.8).

$$Fl_i = \sum_{j=1}^8 [Link_Matrix_{BDC\ network}(L_i H_j) \times Fh_{ji}] \quad (4.8)$$

where $i \in \{1,2, \dots, 19\}$ and $j \in \{1,2, \dots, 8\}$

For instance, Fl for the location node 1, Fl_1 can be determined through expression (4.9)

where the value computed is 1.

$$\begin{aligned}
Fl_1 &= [Link_Matrix_{BDC\ network}(L_1H_1) \times Fh\ 1:1] \\
&+ [Link_Matrix_{BDC\ network}(L_1H_2) \times Fh\ 2:1] \\
&+ [Link_Matrix_{BDC\ network}(L_1H_3) \times Fh\ 3:1] \\
&+ [Link_Matrix_{BDC\ network}(L_1H_4) \times Fh\ 4:1] \\
&+ [Link_Matrix_{BDC\ network}(L_1H_5) \times Fh\ 5:1] \\
&+ [Link_Matrix_{BDC\ network}(L_1H_6) \times Fh\ 6:1] \\
&+ [Link_Matrix_{BDC\ network}(L_1H_7) \times Fh\ 7:1] \\
&+ [Link_Matrix_{BDC\ network}(L_1H_8) \times Fh\ 8:1] \\
&= [1 \times 1] + [0 \times 0] + [0 \times 0] + [0 \times 0] + [0 \times 0] + \\
&\quad [0 \times 0] + [0 \times 0] + [0 \times 0] \\
&= 1
\end{aligned} \tag{4.9}$$

Likewise, the Fl values for the remaining location nodes are determined as [1, 1, 1, 1, 1, 1, 2, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1] according to the sequence of the locations. All parameters for location node will be summarised in section 4.4.

Parameter of Location Node in BDC Network

There are four significant location parameters include Lc, S, B and Fl_i have been quantified. Lc and S are quantified through polynomial curve fitting with three attributes: latitude (x), longitude (y) and temperature (T). B is quantified using equation with respect to temperature (T) and Fl_i is quantified through equation (4.8).

Parameters initially decided for the location node of the BDC network as shown in Figure 4.4 (a) are then further refined, particularly after the processes of quantifying Lc, S, B and Fl_i parameters. It is important to keep the model simple (Barns & Fulford, 2009, p. 3).

Thus, latitude, longitude and temperature are excluded from the location parameter as their effect have been accounted for in the life cycle model, survival model and biting rate.

As a result of process 2.2 (Figure 1.1), 7 parameters are finalised as the BDC network location node parameters: life cycle index, survival index, biting index, humidity, precipitation, altitude and number of times a location is visited by human as presented in Figure 4.4(b). The values of three parameters – *Al*, *Pre* and *K* – are directly obtained from the research data. Table 4.10 presents the values of all 7 parameters for all of the 19 location nodes in the BDC network dataset.

Table 4.10: Parameter Values of 19 Location Nodes in BDC Network

Location Node	<i>Lc</i>	<i>S</i>	<i>B</i>	<i>Al</i> (m)	<i>K</i> (%)	<i>Pre</i> (mm)	<i>Fl</i>
L1	0.1995	65.1060	0.8	13	76.3	2.2	1
L2	0.1817	78.1297	0.9	24	74.2	0.1	1
L3	0.2000	68.8743	0.8	7	85.3	13.1	1
L4	0.2008	64.4377	0.8	8	79.3	7.7	1
L5	0.1791	80.9369	0.8	20	78	1.1	1
L6	0.1791	80.9369	0.8	4	78	1.1	1
L7	0.2024	63.9387	0.8	9	78.8	5.4	2
L8	0.1709	91.3251	0.8	6	76.1	0.7	1
L9	0.1995	65.1060	0.8	13	76.3	2.2	1
L10	0.1817	78.1297	0.8	54	74.2	0.1	1
L11	0.2043	66.5415	0.8	9	82.6	11.6	1
L12	0.1977	65.9558	0.8	4	77.7	5.3	1
L13	0.1994	65.1060	0.8	21	76.3	2.2	1
L14	0.2010	64.4377	0.8	8	77.4	5.3	1
L15	0.1817	78.1297	0.8	28	74.2	0.1	1
L16	0.2043	66.5415	0.8	8	82.2	12.3	1
L17	0.2010	64.4377	0.8	10	76.9	5.3	1
L18	0.1791	80.9369	0.8	25	75.3	0.6	1
L19	0.1977	65.9558	0.8	9	76.1	0.1	1

4.4 Quantification of Human Node Parameters

Parameter for human node will be quantified. Based on the conclusion made from the related works presented in Chapter 2, two parameters are identified for the human node of BDC network. Two parameters are listed below.

- a) Duration time on the human stay at one location, Du
- b) Number of times a human visited a location, Fh

4.4.1 Quantification of Duration time the human stay at one location, Du

Duration time of human stay at one location is termed by symbol Du . Total duration of a human stay at one location across 14 days is recorded in the investigation form. These 14 days are the periods of dengue patient before emerge the first symptoms and the dengue patients' movement within these 14 days are important in dengue transmission (WHO, 2012). The duration is recorded in either days, hours or even minutes. In order to standardise the duration, all time taken of a human stay at one location across 14 days is calculated in seconds. The duration for human j visited location i across 14 days is denoted as Du_{ij} are calculated as in equation (4.10).

$$Du_{ij} = \text{total duration of human } j \text{ visited location } i \text{ in seconds} \quad (4.10)$$

As an illustration, location node with ID, L10 represents BDC market in Kuching with the latitude and longitude as 1.5117, 110.1386. This is the residential house of human, H2. The total duration of patient H2 stay at location L10 is from 7:30am until 4pm which is 8.5 hours. And H2 stayed in L10 everyday across 14 days before the first symptom emerge. Thus, the total duration of H2 stay at L2 across 14 days are obtained from the expression (4.11).

$$Du_{10,2} = 8.5 \text{ hour} \times 60 \text{ minute} \times 60 \text{ seconds} \times 14 \text{ day} = 428400 \text{ seconds} \quad (4.11)$$

Likewise, the duration for other locations are calculated and the results of Du_{ij} is tabulated in Table 4.11 according to respective human in the first targeted network model.

Table 4.11: Duration Time the Human Stay at One Location, Du_{ij}

L node	H Node	Du_{ij} (s)
L1	H1	655200
L2	H2	705600
L3	H3	990000
L4	H4	284400
L5	H5	756000
L6	H6	1146600
L7	H6	37800
	H7	997200
L8	H8	918000
L9	H1	529200
L10	H2	428400
L11	H3	162000
L12	H4	207000
L13	H7	126000
L14	H8	291600
L15	H2	75600
L16	H3	57600
L17	H4	14400
L18	H6	25200
L19	H4	86400

4.4.2 Quantification of Number of times a human visited a location, Fh

The parameter of the number of times a human visited a location is represented by symbol Fh . Human and location here refer to each of the respective location and human nodes.

As such, this parameter is further detailed as Fh_{ij} and is defined in equation (4.12), to record how many times has human node j visited location node i .

$$Fh_{ij} = \begin{cases} n & \text{if } H_j \text{ visited } L_i \text{ } n \text{ times where } n \in \mathbb{Z}^+ \\ 0 & \text{if } H_j \text{ not visit } L_i \end{cases} \quad (4.12)$$

Since the first targeted BDC network formulated in previous chapter serve as demonstration example, the value of n and thus Fh_{ij} for every human node with respect to each of the 19 location nodes is determined from Table 3.4 that traces the movement of each unique individual human. The values of Fh_{ij} agree with the number of link(s) formed between human node j and location node i in BDC network shown in Figure 3.5. Table 4.12 displays the value of Fh for each human node. The sum of each row in Table 4.12 is same with the degree of that particular location node, which is the value of parameter Fl .

Table 4.12: Parameter Values of Fh_{ij}

Fh		j (Human Node)							
		1	2	3	4	5	6	7	8
i (Location Node)	1	1	0	0	0	0	0	0	0
	2	0	1	0	0	0	0	0	0
	3	0	0	1	0	0	0	0	0
	4	0	0	0	1	0	0	0	0
	5	0	0	0	0	1	0	0	0
	6	0	0	0	0	0	1	0	0
	7	0	0	0	0	0	1	1	0
	8	0	0	0	0	0	0	0	1
	9	1	0	0	0	0	0	0	0
	10	0	1	0	0	0	0	0	0
	11	0	0	1	0	0	0	0	0
	12	0	0	0	1	0	0	0	0
	13	0	0	0	0	0	0	1	0
	14	0	0	0	0	0	0	0	1
	15	0	1	0	0	0	0	0	0
	16	0	0	1	0	0	0	0	0
	17	0	0	0	1	0	0	0	0
	18	0	0	0	0	0	1	0	0
	19	0	0	0	1	0	0	0	0

Seven parameters are identified and quantified for location node while the human node is determined by two parameters. The values for each of the parameters of every location and human node are presented in Figure 4.7.

By now, execution of Process 3.2 and Process 3.3 in Stage 3 of BNM-RMF are completed. In the next section, quantification of link weight will be presented and this leads to the completion of a complete BDC network.

4.5 Quantification of Link Weight

20 link edges are identified from the first BDC graph established in section 3.4.4 and the weights of these edges need to be computed. This link weight are termed as dengue contact strength (DCS) that represents the link affinity between the human and location nodes. The stronger the strength indicates the greater degree of attachment between the human and the specific location which contributes a higher degree of human contact with the specific location.

Eze (2013) introduced summation rule to compute the contact strength for the edge formed between the bipartite nodes of the malaria transmission network (Eze, 2013, p. 83). This applied in the study and the quantification technique explained in equation (4.13). Summation is used because the total of the individual parameter will contribute to a bigger value which indicate a stronger strength. However, individual parameters consist of rational number such as 0.7, the product of these rational numbers will contribute a smaller value which indicate a weaker strength. Thus, a summation rule is the most suitable one. DCS_{ij} refers to the dengue contact strength of the link formed between location i and human j .

$$\begin{aligned}
 DCS_{ij} &= \left(\sum Location_Node_Parameters_i \right) + \left(\sum Human_Node_Parameters_{ij} \right) \\
 &= (Lc_i + S_i + B_i + Al_i + K_i + Pre_i + Fl_i) + (Du_{ij} + Fh_{ij})
 \end{aligned} \tag{4.13}$$

Figure 4.6 shows the relationship between the parameter of the location and human nodes and the quantification of DCS of the BDC network. MATLAB R2014b is used as the platform for the computation of DCS.

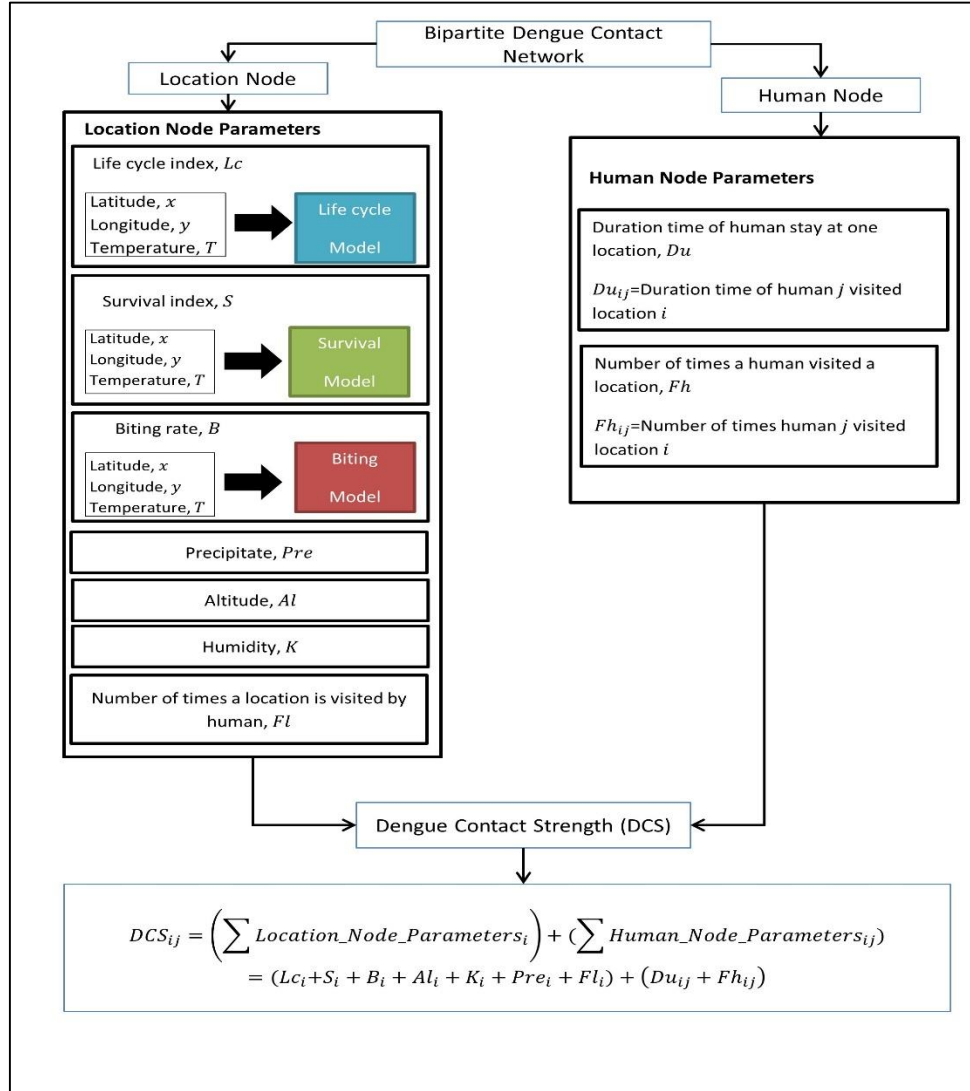


Figure 4.6: Quantifying the Link Weights

Normalisation is executed prior to DCS computation which convert each parameter values into a numerical value that range between 0 and 1. Normalisation process looks for the maximum value of a parameter, use this value to transform all values into a normalised value so all the normalised value lie between 0 and 1 (Liew, 2016, Normalisation Overview, 2004).

Equation (4.14) shows the transformation for values of parameter Al . The maximum value of Al is 54 m. Normalisation has transformed the values of Al which are [13, 24, 7, 8, 20, 4, 9, 6, 13, 54, 9, 4, 21, 8, 28, 8, 10, 25, 9] into {0.2407, 0.4444, 0.1296, 0.1481, 0.3704, 0.0741, 0.1667, 0.1111, 0.2407, 1.0, 0.1667, 0.07407, 0.3889, 0.1481, 0.5185, 0.1481, 0.1852, 0.4630, 0.1667}, correct to four decimal places. The same goes for the parameter values Lc , S , B , K , Pre , Fl , and Fd . The complete normalised parameter values of Lc , S , B , Al , K , Pre and Fl are tabulated in Table 4.13.

$$Al_{i_norm} = \frac{Al_i}{MAX(Al_1, Al_2, Al_3, \dots, Al_{17}, Al_{18}, Al_{19})} \quad (4.14)$$

Table 4.13: Complete Normalised Parameter of Location Nodes (First BDC network)

Location ID	Lc	S	B	Al	K	Pre	Fl
L1	0.9765	0.7129	0.8889	0.2407	0.8945	0.1679	1
L2	0.8894	0.8555	1	0.4444	0.8699	0.0076	1
L3	0.9790	0.7542	0.8889	0.1296	1	1	1
L4	0.9829	0.7056	0.8889	0.1481	0.9297	0.5878	1
L5	0.8767	0.8863	0.8889	0.3704	0.9144	0.0840	1
L6	0.8767	0.8863	0.8889	0.0741	0.9144	0.0840	1
L7	0.9907	0.7001	0.8889	0.1667	0.9238	0.4122	1
L8	0.8365	1	0.8889	0.1111	0.8921	0.0534	1
L9	0.9765	0.7129	0.8889	0.2407	0.8945	0.1679	1
L10	0.8894	0.8555	0.8889	1.0	0.8699	0.0076	1
L11	1	0.7286	0.8889	0.1667	0.9683	0.8855	1
L12	0.9677	0.7222	0.8889	0.0741	0.9109	0.4046	1
L13	0.9760	0.7129	0.8889	0.3889	0.8945	0.1679	1
L14	0.9838	0.7056	0.8889	0.1481	0.9074	0.4046	1
L15	0.8894	0.8555	0.8889	0.5185	0.8699	0.0076	1
L16	1	0.7286	0.8889	0.1481	0.9637	0.9389	1
L17	0.9838	0.7056	0.8889	0.1852	0.9015	0.4046	1
L18	0.8767	0.8863	0.8889	0.4630	0.8828	0.0458	1
L19	0.9677	0.7222	0.8889	0.1667	0.9144	0.0840	1

The DCS is computed by using the normalised parameter values of location and human nodes. For example, the weight of the link formed between node L7 and node H6, $DCS_{7,6}$ is obtained through the working shown in expression (4.15) where the normalised parameter value are used here.

$$\begin{aligned}
 DCS_{7,6} &= \left(\sum Location_Node_Parameters_7 \right) + \left(\sum Human_Node_Parameters_{7,6} \right) \quad (4.15) \\
 &= (Lc_7 + S_7 + B_7 + Al_7 + K_7 + Pre_7 + Fl_7) + (Du_{7,6} + Fh_{7,6}) \\
 &= 0.9907 + 0.7001 + 0.8889 + 0.1667 + 0.9238 + 0.4122 + 1 + 0.0253 + 1 \\
 &= 6.1077
 \end{aligned}$$

Likewise, the remaining of DCS parameters are computed with the same expression (4.15). The complete DCS is tabulated in Table 4.14 and termed as DCS matrix.

Table 4.14: Dengue Contact Strength (DCS) Matrix of BDC Network

	H1	H2	H3	H4	H5	H6	H7	H8
L1	6.4528	0	0	0	0	0	0	0
L2	0	6.6822	0	0	0	0	0	0
L3	0	0	7.6151	0	0	0	0	0
L4	0	0	0	6.4910	0	0	0	0
L5	0	0	0	0	6.6800	0	0	0
L6	0	0	0	0	0	6.7244	0	0
L7	0	0	0	0	0	6.1077	6.1154	0
L8	0	0	0	0	0	0	0	6.5826
L9	6.3429	0	0	0	0	0	0	0
L10	0	6.8849	0	0	0	0	0	0
L11	0	0	6.7793	0	0	0	0	0
L12	0	0	0	6.1489	0	0	0	0
L13	0	0	0	0	0	0	6.1390	0
L14	0	0	0	0	0	0	0	6.2927
L15	0	6.0957	0	0	0	0	0	0
L16	0	0	6.7184	0	0	0	0	0
L17	0	0	0	6.0822	0	0	0	0
L18	0	0	0	0	0	6.0655	0	0
L19	0	0	0	5.8193	0	0	0	0

The output of DCS for all links in BDC Graph (Figure 3.8) eventually resulted in the BDC network as shown in Figure 4.7.

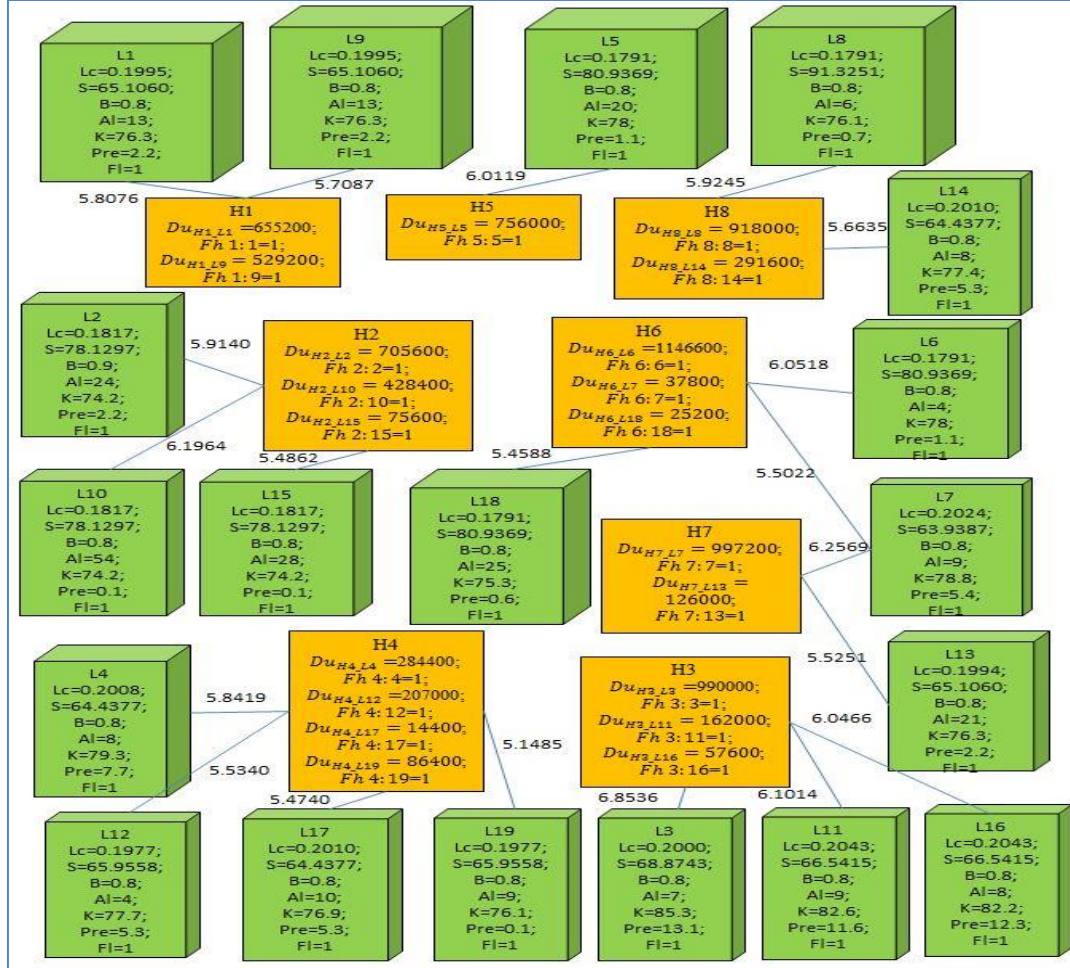


Figure 4.7: Bipartite Dengue Contact (BDC) Network

Web-based Search Algorithm

One of the processes in BNM-RMF is to rank the nodes with the ranking values. Web-based search algorithm is a popular algorithm to rank the documents based on the link information among different documents. The web-based search algorithm has been used to rank the nodes in a network (Eze et al., 2011a; Liew et al., 2014). As discussed in Section 2.4, Hypertext Induced Topic Selection (HITS) search algorithm has been chosen to rank the

location and human nodes in this study. HITS is implemented to rank the location nodes for malaria transmission (Eze, 2013) and determine the preferred habitat of the dolphin (Liew, 2016) in previous studies.

4.6 HITS Search Algorithm

This section discussed the process 2.5 in BNM-RMF, which is the process of implementing the search algorithm. This search algorithm is the ranking algorithm which involves the computation of principal eigenvalues and eigenvectors.

The output of this HITS search algorithm is the dengue hotspot ranking (DHR) values, which is a measure of the vector density. The locations are prioritised according to the DHR values whereby the higher ranking location represents the more critical location where need to be tackled first. This DHR value ranges between 0 and 1 which is approximated and indicated by the values of principal eigenvectors computed from the implementation of the HITS search algorithm onto BDC network by adopting power iteration method.

The implementation process of HITS algorithm is adapted from the work of Eze (2013) and Liew (2016) which involved malaria hotspots ranking and habitat suitability ranking research. The process involves four steps and is summarised in Figure 4.8. The input for this algorithm is the DCS matrix in Table 4.13. The row of this matrix refers to the location node while the column represents the human node.

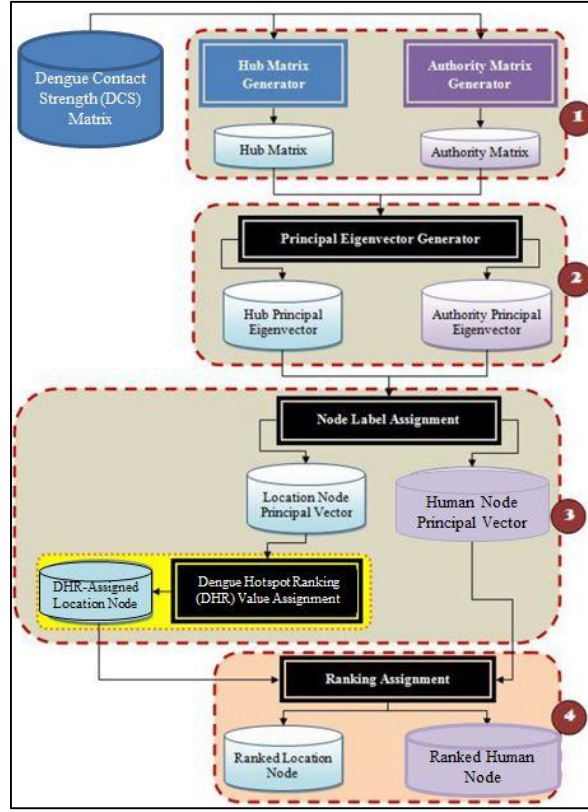


Figure 4.8: Process of the Adapted HITS Search Algorithm (Liew, 2016)

4.6.1 Generation of Hub and Authority Matrix

Step 1 from Figure 4.8 aims to produce a hub and authority matrices via respective matrix generator. The equations adopted to generate hub and authority matrices are given in equations (4.16) and (4.17) where T indicates a transpose operator. These equations are adapted from the work of Eze (2013) and Liew (2016).

$$hub = DCS \times DCS^T \quad (4.16)$$

$$authority = DCS^T \times DCS \quad (4.17)$$

The DCS matrix from Table 4.13 is the input to the hub and authority matrix generator in Step 1 as shown in Figure 4.8. The DCS matrix needs to be a real and non-negative matrix to ensure the exist of real eigen pair. The dimension DCS matrix for the first BDC network model is 19 (location nodes) by 8 (human nodes) and it is a real matrix. Equations (4.16) and

(4.17) produce the hub and authority matrices that are symmetric to ensure the existence of the eigen pair. The resulted authority and hub matrices are shown in Table 4.15 and Table 4.16, respectively. The dimension of authority matrix is 8 by 8 to represent 8 human nodes while the hub matrix is 19 by 19 to represent 19 location nodes in the network model.

Table 4.15: Authority Matrix

66.32	0	0	0	0	0	0	0
0	103.47	0	0	0	0	0	0
0	0	120.76	0	0	0	0	0
0	0	0	121.22	0	0	0	0
0	0	0	0	36.14	0	0	0
0	0	0	0	0	96.71	34.44	0
0	0	0	0	0	34.44	69.68	0
0	0	0	0	0	0	0	67.17

4.6.2 Generation of Hub and Authority Principal Eigenvector

The second step in the adapted HITS algorithm in Figure 4.8 aims to identify the principal eigenvector for the hub and authority matrices through equation (4.18) where Hu is hub matrix, A is authority matrix, λ_H is eigenvalues of hub matrix and λ_A Are eigenvalues of authority matrix, v is eigenvectors of hub matrix and η is an eigenvector of authority matrix.

$$Huv = \lambda_H v$$

$$A \eta = \lambda_A \eta \quad (4.18)$$

In order to calculate the principal eigenvectors and eigenvalues, the power iteration method in equation (4.18) is adopted (Choi, 2012; Eze, 2013; Liew, 2016). In equation (4.19), $k = \{0, 1, 2, \dots\}$; v_0 is the initial vector which is an approximation of the principal eigenvector; $\|Huv_k\|$ or $\|A\eta_k\|$ is the length of the product of the matrix and its k^{th} eigenvector. In this trial, the chosen initial vector, v_0 is a linear vector of nineteen elements of '1'.

Table 4.16: Hub Matrix

33.7	0	0	0	0	0	0	0	33.2	0	0	0	0	0	0	0	0	0
0	35.0	0	0	0	0	0	0	0	36.7	0	0	0	0	32.5	0	0	0
0	0	47.0	0	0	0	0	0	0	0	41.8	0	0	0	0	41.4	0	0
0	0	0	34.1	0	0	0	0	0	0	0	32.3	0	0	0	0	32.0	30.1
0	0	0	0	36.1	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	36.6	33.3	0	0	0	0	0	0	0	0	0	33.0	0
0	0	0	0	0	33.3	69.4	0	0	0	0	0	34.6	0	0	0	30.0	0
0	0	0	0	0	0	0	35.1	0	0	0	0	0	33.6	0	0	0	0
33.2	0	0	0	0	0	0	0	32.6	0	0	0	0	0	0	0	0	0
0	36.7	0	0	0	0	0	0	0	38.4	0	0	0	0	34.0	0	0	0
0	0	41.8	0	0	0	0	0	0	0	37.2	0	0	0	0	36.9	0	0
0	0	0	32.3	0	0	0	0	0	0	0	30.6	0	0	0	0	30.3	28.5
0	0	0	0	0	0	34.6	0	0	0	0	0	30.5	0	0	0	0	0
0	0	0	0	0	0	0	33.6	0	0	0	0	0	32.1	0	0	0	0
0	32.5	0	0	0	0	0	0	0	34.0	0	0	0	0	30.1	0	0	0
0	0	41.4	0	0	0	0	0	0	0	36.9	0	0	0	0	36.6	0	0
0	0	0	32.0	0	0	0	0	0	0	0	30.3	0	0	0	0	30.0	28.2
0	0	0	0	0	33.0	30.0	0	0	0	0	0	0	0	0	0	29.8	0
0	0	0	30.1	0	0	0	0	0	0	0	28.5	0	0	0	0	28.2	26.5

$$v_{k+1} = \frac{Huv_k}{||Huv_k||} \quad (4.19)$$

$$\eta_{k+1} = \frac{A\eta_k}{||A\eta_k||}$$

Iterations in equation (4.19) are executed where at every iteration, vector v_k and η_k are multiplied by the respective matrix H or A , and normalised until stopping criteria are met. Finally the principal eigenvector is produced when the iteration stop. There are two stopping criteria used in this research, (a) the estimation error between the current and the previous eigenvalues; (b) the maximum number of iterations allowed. The stopping criteria in is a threshold value 0.001 and the maximum number of iterations is 1000 as reported by Liew (2016). Taking hub matrix (Table 4.15) as input, the output is the eigen pair. The eigenvector product will be served as an input for the next step.

4.6.3 Assignment of Nodes' Label and Dengue Hotspot Ranking Values

The third step involves two processes: (i) assigning the node label to the principal eigen vectors and (ii) assign the DHR values to each location node. Firstly, each element in the principal eigenvectors will be assigned to the location node and human node labels. For the first BDC network, the location nodes are assigned to the respective elements of hub principal eigenvector (obtained from the previous step). In addition, the human nodes are assigned to the respective elements of authority principal eigenvectors. The outputs of this step are the hub principal eigenvector with respective location node label as tabulated in Table 4.17.

Table 4.17: Principal Eigenvector of the (a) Location Node and (b) Human Node

Location Node Label	Hub Principal Eigenvector
L1	0
L2	3.1E-116
L3	0.001617
L4	1
L5	0
L6	4.64E-07
L7	7.49E-07
L8	0
L9	0
L10	3.3E-116
L11	0.00144
L12	0.947297
L13	2.89E-07
L14	0
L15	2.9E-116
L16	0.001427
L17	0.937019
L18	4.19E-07
L19	0.881315

Human Node Label	Authority Principal Eigenvector
H1	0
H2	2.1E-107
H3	0.002599
H4	1
H5	0
H6	1.96E-06
H7	1.34E-06
H8	0

4.6.4 Ranking of Bipartite Nodes in BDC Network

From Table 4.17 (a), location L4 node has the highest principal eigenvector with a value of 1, this principal eigenvector of location node is also termed as a DHR values. DHR value is a measurement of the density of dengue vector which can be also seen as a measurement to prioritise different levels of dengue hotspot. A location node with a higher DHR value indicates the location with higher vector density and vice versa. Thus, the L4 is ranked the first among all the location nodes. This implies that L4 is the location with the highest vector density for *Aedes* mosquitoes in Kuching during Epi Week 28 and 29 (the first BDC network model). The L12 is the location node with the second highest DHR, i.e. 0.947297 and hence L12 is ranked

as the second highest vector density in Kuching. The ranking result for location nodes is tabulated in Table 4.18 (a).

Table 4.18: The Ranking of (a) Location Nodes and (b) Human Node

(a)		(b)	
Location Node Label	DHR Value	Human Node Label	Ranking Value
L4	1	H4	1
L12	0.947297	H3	0.002599
L17	0.937019	H6	1.96E-06
L19	0.881315	H7	1.34E-06
L3	0.001617	H2	2.1E-107
L11	0.00144	H1	0
L16	0.001427	H5	0
L7	7.49E-07	H8	0
L6	4.64E-07		
L18	4.19E-07		
L13	2.89E-07		
L10	3.3E-116		
L2	3.1E-116		
L15	2.9E-116		
L1	0		
L5	0		
L8	0		
L9	0		
L14	0		

Similarly, when authority matrix is passed to the search algorithm implementation system, the elements of the authority principal eigenvectors produced as shown in Table 4.17 (b) are taken as the ranking index for human nodes. These ranking indices are used to rank the human nodes of BDC network. The focus of ranking in this study is on the location nodes, the ranking of the human nodes is utilised by the research as an indicator of the relative importance role played by the human nodes in identifying the high risk areas. The highest ranking index of

the authority principal eigenvector is 1 which belongs to H4. This implies that, among all the unique individual humans, H4 is relatively the most influential patient in determining the location that is higher vector density or higher risk. A ranking of the human nodes is tabulated in Table 4.18 (b).

In this final process of the stage two in BNM-RMF, this study generated the dengue hotspot index to rank the location nodes to assist the public health authority to manage the resources in eradicating dengue. At the end outcome of stage 2, a complete bipartite dengue contact (BDC) network is formulated successfully.

4.7 Summary

This chapter discussed the execution of Process 2.1 until Process 2.5 of the model construction stage of BDC network (Figure 1.1). Firstly, the pre-processing of the raw data for location, human and parameters. Next, the parameters for the location node are determined, polynomial curve fitting is chosen as the quantification technique to build Life Cycle Model and Survival Model. Life cycle index (Lc) and survival index (S) are generated from the model, number of times a location is visited by human (Fl) is quantified by using the link matrix, and all parameters of location node are determined. After that, two parameters: duration time of a human stay at a location (Du) and number of times a human visited a location (Fh) are identified as the parameter of human node. Finally, link weight of BDC network is quantified using the summation rule to produce the dengue contact strength (DCS).

Last but not least, the implementation of web-based search algorithm in generating dengue hotspot ranking value for the BDC network model. The adapted HITS search algorithm techniques has been employed where DHR and ranking index are produced and used to rank the location node and the human nodes in the BDC network model.

CHAPTER 5

MODEL ANALYSIS AND EVALUATION

5.1 Introduction

This chapter presents the network model results obtained through a web-based search algorithm. The processes in Stage 3 (Model Analysis and Evaluation) which include model verification and validation will be presented in this chapter. The model verification was conducted to determine the modelling process has been executed correctly. Benchmark verification was then carried out by calculating the Root-Mean-Square-Error (RMSE) between the ranking values that produced from the benchmark system and the BDC network. The analytical verification is conducted by calculating the Spearman's Rank Correlation Coefficient (SRCC) between the hub matrix and the dengue hotspot ranking values. The validation process was executed by calculating the SRCC between the targeted and validated network. Finally, the parameter significance analysis, predictive power analysis and data size analysis are discussed.

5.2 Model Verification

This section outlines the first process in Stage 3 of BNM-RMF. Verification is defined as the procedure of determining if the implementation of a model has been done correctly (Davis, 1992). Verification also commented as the process to answer the question of "have we built the model right?" (Cook & Skinner, 2005). Therefore, this verification process executed in this research is to ensure the procedures to build the network model are correct and complies with a benchmark system. Two verification methods: benchmark verification and analytical verification are employed in this research.

5.2.1 Benchmark Verification

Benchmark verification aims to verify research procedures by comparing the result obtained from the BDC network model and the result produced by a benchmark system (Liew, 2016). This can be achieved by calculating the RMSE between the ranking values obtained from this research and the values produced by a benchmark system. RMSE is the most popular measures of goodness-of-fit, a good RMSE value to verify the model is built right when the values is not greater than 0.05 (Albright, 2008; Browne & Cudeck, 1992). The benchmark verification involved three steps as shown in Figure 5.1 which is adopted from the work of Eze (2013) and Liew (2016).

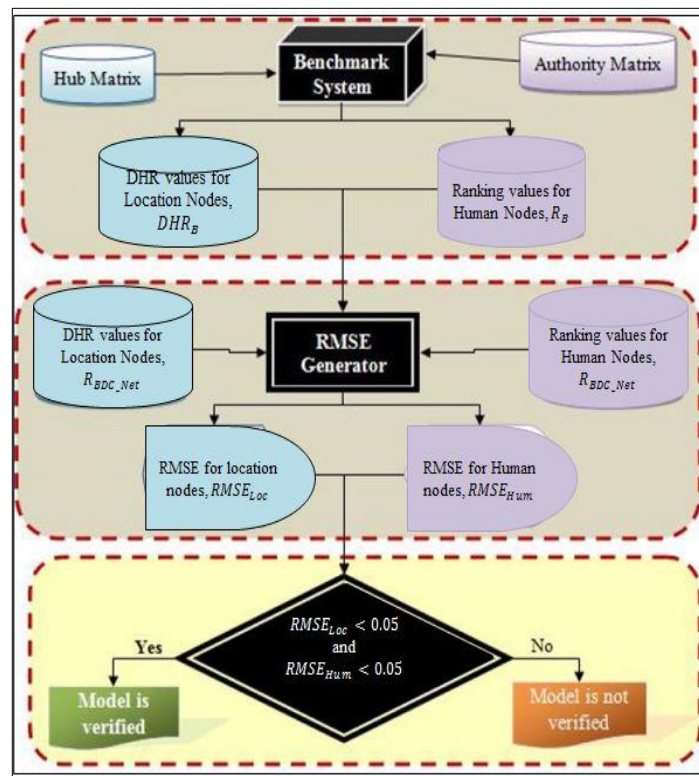


Figure 5.1: Benchmark Verification Process (Liew, 2016)

UCINET 6 for Windows is a powerful network analysis software and able to handle 2 million nodes and also two-type-node networks (Borgatti, Everett & Freeman, 2002). UCINET

is one of the most trusted network analysis, comprehensive and wide usability software (Boyd, Fitzgerald, & Beck, 2007). It supports many network analysis method, which includes centrality measure, subgroup identification and elementary graph theory. UCINET has strong matrix analysis routines which fit in this research as the input of the data is hub and authority matrix. The centrality measure analysis feature available in UCINET can determine the node centrality of a network. The node centrality obtained by measuring the relative importance of a node within the network. Bonacich eigenvector centrality provided in UCINET is one of the centrality measures. This Bonacich eigenvector centrality can identify the node which is more central if they are connected to other nodes that are themselves central. Hence, it relates both the number of connections to other nodes and also measure how well the neighbors are connected.

Firstly, the hub matrix in Excel format is loaded into UCINET as shown in Figure 5.1. It is then transformed into UCINET system files, header file (in `##H` file) and data file (in `##D` file). Next, the header file is used for centrality analysis in UCINET. UCINET generate the principal eigenvector which the elements are taken as the ranking values for location nodes and presented in a log file. These ranking values produced by the benchmark system are termed as benchmark values of DHR , DHR_B . The location nodes are ranked based on the DHR_B and this ranking is the benchmark ranking of location nodes.

Figure 5.2 shows the benchmark ranking result of the location nodes and the result obtained in this research which is termed as BDC network model ranking. It showed that the benchmark ranking is the same as the BDC network model ranking. However, the range of DHR_B is 0 to 0.531 which is different from the range of DHR_{BDC_Net} which is 0 to 1. Calibration need to be carried out to bring two different rates or units into equivalent units, else the values cannot be used in RMSE analysis (A2LA, 2014). Calibration is carried out with an

equation (5.1). The calibrated DHR values are produced and termed as *Calibrated DHR_B* as shown in the first step in Figure 5.1.

$$\text{Calibrated } DHR_B = \frac{DHR_{Each\ Location}}{\text{maximum } DHR_B} \quad (5.1)$$

Benchmark Ranking		BDC Network Model Ranking	
Location node	DHR _B	Location node	DHR _{BDC Net}
L4	0.531	L4	1
L12	0.503	L12	0.947297
L17	0.497	L17	0.937019
L19	0.468	L19	0.881315
L3	0.000	L3	0.001617
L11	0.000	L11	0.00144
L16	0.000	L16	0.001427
L7	0.000	L7	7.49E-07
L6	0.000	L6	4.64E-07
L18	0.000	L18	4.19E-07
L13	0.000	L13	2.89E-07
L10	0.000	L10	3.3E-116
L2	0.000	L2	3.1E-116
L15	0.000	L15	2.9E-116
L1	0.000	L1	0
L5	0.000	L5	0
L8	0.000	L8	0
L9	0.000	L9	0
L14	0.000	L14	0

Figure 5.2: Benchmark Ranking and BDC Network Model Ranking

The second step is to execute RMSE analysis. Both DHR_B and DHR_{BDC_Net} are the inputs to the RMSE Generator as shown in Figure 5.1. RMSE for location nodes is defined in equation (5.2) where N_{Loc} = the number of location node = 19 in this research and i is a natural number starting from 1. The calculations of RMSE on location nodes are tabulated in Table 5.1. The sum of squares difference obtained from Table 5.1 is substituted into equation (5.2). As a

result, RMSE for location node is 0.0006419 is shown in equation (5.3), correct to four significant figures. The RMSE is much lesser than the threshold value of 0.05.

$$RMSE_{Loc}(DHR_B, DHR_{BDC_Net}) = \sqrt{\frac{1}{N_{Loc}} \left[\sum_{i=1}^{N_{Loc}} (\{DHR_B\}_i - \{DHR_{BDC_Net}\}_i)^2 \right]} \quad (5.2)$$

Table 5.1: RMSE Analysis of Location Nodes

Location Node	DHR_B	DHR_{BDC_Net}	$DHR_B - DHR_{BDC_Net}$	$(DHR_B - DHR_{BDC_Net})^2$
L1	0	0	0	0
L2	0	3.1E-116	-3.1E-116	9.61E-232
L3	0	0.001617	-0.001617	2.61469E-06
L4	1	1	0	0
L5	0	0	0	0
L6	0	4.64E-07	-0.000000464	2.15296E-13
L7	0	7.49E-07	-0.000000749	5.61001E-13
L8	0	0	0	0
L9	0	0	0	0
L10	0	3.3E-116	-3.3E-116	1.089E-231
L11	0	0.00144	-0.00144	2.0736E-06
L12	0.947269	0.947297	-2.76968E-05	7.67113E-10
L13	0	2.89E-07	-0.000000289	8.3521E-14
L14	0	0	0	0
L15	0	2.9E-116	-2.9E-116	8.41E-232
L16	0	0.001427	-0.001427	2.03633E-06
L17	0.93597	0.937019	-0.001049132	1.10068E-06
L18	0	4.19E-07	-0.000000419	1.75561E-13
L19	0.881356	0.881315	4.09322E-05	1.67545E-09
Sum of $(DHR_B - DHR_{BDC_Net})^2$:				7.82774E-06

$$RMSE_{Loc}(DHR_B, DHR_{BDC_Net}) = \sqrt{\frac{1}{19} [7.82774E - 06]} = 0.0006419 \quad (5.3)$$

The RMSE for the human node has to be evaluated and this can be achieved by having the authority matrix as the input to the benchmark system as shown in the first step in Figure 5.1. The ranking value of each human that determined through the UCINET system is termed as R_B while the principal eigenvector (Table 4.18 (b)) of human yielded from the BDC network model is termed as R_{BDC_Net} . RMSE for human, $RMSE_{Hum}$ is defined in equation (5.4) where

N_{Hum} = number of human node = 8 and j is a natural number starting from 1. Table 5.2 shows the calculations of RMSE analysis onto the human nodes using equation (5.4).

$$RMSE_{Hum}(R_B, R_{BDC_Net}) = \sqrt{\frac{1}{N_{Hum}} \left[\sum_{j=1}^{N_{Hum}} (\{R_B\}_j - \{R_{BDC_Net}\}_j)^2 \right]} \quad (5.4)$$

Table 5.2: RMSE Analysis of Human Nodes

Human Node	R_B	R_{BDC_Net}	$R_B - R_{BDC_Net}$	$(R_B - R_{BDC_Net})^2$
H1	0	0	0	0
H2	0	2.1E-107	-2E-107	4.4429E-214
H3	0	0.002599	-0.0026	6.75281E-06
H4	1	1	0	0
H5	0	0	0	0
H6	0	1.96E-06	-2E-06	3.85387E-12
H7	0	1.34E-06	-1.3E-06	1.791E-12
H8	0	0	0	0
Sum of $(R_B - R_{BDC_Net})^2$:				6.75282E-06

From the sum of square of difference in Table 5.2, RMSE is calculated as in the expression (5.5). The resulted value of RMSE obtained for human node is 0.0009188, correct to four significant figures. This RMSE is much lesser than the threshold value of 0.05.

$$RMSE_{Hum}(R_B, R_{BDC_Net}) = \sqrt{\frac{1}{8} [6.75282E - 06]} = 0.0009188 \quad (5.5)$$

The value of RMSE of location and human nodes is passed to the last step of the verification process (Figure 5.1). It is observed that the RMSE of location (0.0006419) and RMSE of human (0.0009188) are much lesser than the threshold RMSE value of 0.05. Therefore, it can be concluded that the BDC network model formulated in this study is a verified model.

5.2.2 Analytical Verification

Spearman coefficient is used to measure the degree of relationship between a pair of rankings. Spearman's Rank Correlation Coefficient (SRCC) used in analytical verification in this study to measure the closeness of the ranking values for each location nodes and human nodes. Due to the small data size, SRCC or ρ is used in verifying the network model (Bouzerdoun, Havstad, & Beghdadi, 2004; Lim, Wang, Lefebvre, & Califano, 2007; Tetko & Tanchuk, 2002). The values of SRCC can be interpreted as Table 5.3. The SRCC values range from -1 until +1 where +1 indicates very strong positive association of ranks; 0 indicates no associations between ranks and -1 indicates very strong negative association of ranks. SRCC with value 1 indicates perfect similarity which is exactly the same ranking while value 0 indicates the absence of similarity which is exactly different ranking. This coefficient is used by other researchers to evaluates the monotonic relationship between ordinal variables in different network approach studies (Chiu & So, 2004; Liew, 2016).

Table 5.3: 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

In order to verify the model analytically, a positive and high correlation with a coefficient of not less than 0.7 between DHR and hub matrix of each location node should be obtained (Liew, 2016). Thus, SRCC for location nodes, $\rho_{DHR_HubMatrix}$ is defined in equation (5.6) where $N_{Loc}=19$ and i is a natural number starting from 1. Calculation for $d_{DHR_HubMatrix}$

in equation (5.6) is shown in Table A.2 (in Appendix A). Using the sum obtained from Table A.2, calculation of $\rho_{DHR_HubMatrix}$ is shown in expression (5.7).

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

where $\{d_{DHR_HubMatrix}\}_i = \{RankDHR\}_i - \{RankHubMatrix_{sum}\}_i$ (5.6)

$$\rho_{DHI_HubMatrix} = 1 - \frac{6 \times 196}{19(19^2 - 1)} = 0.828 \quad (5.7)$$

The value of $\rho_{DHR_HubMatrix} = 0.828$ shows that DHR of each location node is positively and highly correlated with the sum of a hub matrix of the edges that are linked to it. The ρ value is far above the threshold value of 0.70. Therefore, this has further strengthened the conclusion made in benchmark verification about the processes involved in developing BDC network model and hence this model is verified. The model validation will be discussed next.

5.3 Model Validation

Model validation aims to answer the question of “are you building the right thing?”. Validation here is defined as a process to check the accuracy of the results obtained from the model. This section discusses the process 3.2 of BNM-RMF in order to validate the BDC network model. The model validation is performed through analysis using validated model as tabulated in Table 3.3. The result yielded by the first targeted model should agree with the result obtained from the validated model (Epi Week 30-31).

From the Table 3.3, the first BDC network (first targeted model) is the network model for Epi Week 28 and 29, this network consists of dengue positive and possible positive patients and the location nodes were visited by the dengue positive and possible positive patients. This BDC network model for Epi Week 28 and 29 is the targeted model. The subsequent Epi Week

from the targeted model is used to build the validated model in order to validate the model. In the objectives mentioned in section 1.3, bipartite network model is able to detect the possible dengue breeding location from a pool of suspected location nodes. Thus, a new BDC network model for Epi Week 30 and 31 which consists of dengue positive patients only and the location nodes were visited by the dengue positive patients is formulated. This new BDC network model for Epi Week 30 and 31 is the validated model. The comparison of targeted and validated model is performed to identify the ability of the prediction by the targeted model.

In validation analysis, SRCC (ρ) is calculated to compare the location rank generated from the targeted model and the validated model which utilised the real data. As mentioned in section 5.2.2, this research resolves to adopt a threshold value, at least 0.70 which implies there exists a high correlation between the targeted model ranking and the ranking produced by the validated model for the SRCC reported. This also implies if the correlation coefficient is higher than 0.70 and thus the BDC network model is validated.

One of the approaches to validate the model is comparing the simulation results with real data. It is undeniable that the ability to reproduce the reality with sufficient exactness is the ultimate objective in validation. In this study, the validation analysis calculates three groups of BDC network models as tabulated in Table 5.4.

Table 5.4: BDC Network Model Validation

Group	BDC Network	Model	Epi Week	SRCC Values
1	1	Targeted Model	28-29	1.0000
	2	Validated Model	30-31	
2	3	Targeted Model	32-33	0.8000
	4	Validated Model	34-35	
3	5	Targeted Model	36-37	0.8424
	6	Validated Model	38-39	

From Table 5.4, the first group consists of 2 models, the targeted model and validated model. First targeted model utilises the data on Epi Week 28 and 29, consist of 8 human nodes which only 2 dengue positive patients and there are 6 possible positive patients. The targeted model in the first BDC network model has been quantified formulated and the location nodes are ranked accordingly to detect any possible dengue hotspots in section 4.7.4. This first BDC network model is validated if the closeness of the location ranking is more than 0.7 (of SRCC values) between the comparison of the subsequent Epi Week 30 and 31. The network model formulated for Epi Week 30 and 31 is termed as a validated model to test whether the location in the targeted model correlates to the location in validated model.

The validated model for Epi Week 30 and 31 consists of three patient nodes and eleven location nodes. There are eight location nodes in this validated dataset are different from the targeted model and thus the new incoming location nodes are labelled as L20, L21,..., L26 and L27. Only three location nodes, L2, L4 and L5 visited by the patients in dataset Epi Week 30 and 31. Nevertheless, L1, L3 - L19 from the targeted model are still included in the validated model to give a complete picture. There are 3 patients with a positive dengue serological test only in this validated model and thus the human nodes are H1, H2 and H3. The human nodes in validated model are not identical with the human nodes in targeted model based on observation of the dataset. Since the human nodes in targeted model are different to the one in validated model, therefore, the validation only focusses in the location node ranking.

Next, the location node parameters – Lc , S , B , Pre , Al , K , and Fl need to be quantified. The quantification of Lc , S , B and Fl are determined using the method described in section 4.3.1 until 4.3.4. All attributes required by the parameters are extracted from the eight location nodes of the validated dataset, the attributes have been formatted and scaled accordingly. Furthermore, values for Pre , Al , K are taken from the dataset. Table A.3 (in Appendix A)

display all the parameter values involved in each of the location nodes of the first validated model. After that, the parameter values are normalised and used for the computation of DCS using equation (4.13) as introduced in Section 4.4. Table 5.5 presents the DCS for all links in validated model.

Table 5.5: Dengue Contact Strength (DCS) Matrix of the First Validated Dataset

	H1	H2	H3
L1	0	0	0
L2	0	0	5.683891
L3	0	0	0
L4	0	0	6.647323
L5	5.481313	0	0
L6	0	0	0
L7	0	0	0
L8	0	0	0
L9	0	0	0
L10	0	0	0
L11	0	0	0
L12	0	0	0
L13	0	0	0
L14	0	0	0
L15	0	0	0
L16	0	0	0
L17	0	0	0
L18	0	0	0
L19	0	0	0
L20	6.528254	0	0
L21	0	6.237059	0
L22	0	0	5.636161
L23	5.354749	0	0
L24	0	0	5.092558
L25	5.117741	0	0
L26	0	0	5.270074
L27	0	0	5.343166

DCS matrix in Table 5.5 is the input to the HITS search algorithm. The same implementation design as depicted in Figure 4.8 is followed. Table 5.6 shows the ranking of

location nodes in the validated model with the DHR. As mentioned earlier in this section, there are three location nodes that are identical to the targeted and validated model. Thus, these three identical location nodes will be utilised in the model validation by calculating the SRCC between the ranking of targeted and validated model. Equation (5.8) gives the formula used for the calculation of ρ_{TarVar} where $N_{Loc} = 3$ and a is a natural number start from 1. Table 5.7 presents the calculation of various terms in equation (5.8). Expression (5.9) shows the calculation of ρ_{TarVar} according to the equation defined in (5.8).

$$\rho_{TarVar} = 1 - \frac{6 \sum_{a=1}^{N_{Loc}} [d_{TarVal}]_a^2}{N_{Loc}(N_{Loc}^2 - 1)}$$

$$\text{where } \{d_{TarVal}\}_a = \{RankDHR_{Targeted}\}_a - \{RankDHR_{Validated}\}_a \quad (5.8)$$

$$\rho_{TarVar} = 1 - \frac{6 \times 0}{3(3^2 - 1)} = 1.0000 \quad (5.9)$$

Table 5.6: The Ranking of Location Nodes (First Validated Model)

Rank	Location Node	DHR _{Validated}	Rank	Location Node	DHR _{Validated}
1	L4	1	14	L6	0
2	L2	0.855065	15	L7	0
3	L22	0.847884	16	L8	0
4	L27	0.803807	17	L9	0
5	L26	0.792811	18	L10	0
6	L24	0.766107	19	L11	0
7	L20	8.62E-06	20	L12	0
8	L5	7.24E-06	21	L13	0
9	L23	7.07E-06	22	L14	0
10	L25	6.76E-06	23	L15	0
11	L21	8.33E-21	24	L16	0
12	L1	0	25	L17	0
13	L3	0	26	L18	0
			27	L19	0

Table 5.7: SRCC Calculation of Location Nodes between the Models in Group 1

Location Node	DHR _{Targeted}	DHR _{Validated}	Rank DHR _{Targeted}	Rank DHR _{Validated}	d_{TarVal}	$[d_{TarVal}]^2$
L2	3.1E-116	0.855065	2	2	0	0
L4	1	1	1	1	0	0
L5	0	7.24E-06	3	3	0	0
Sum of $[d_{TarVal}]^2$:						0

The result of this validation, $\rho_{TarVar} = 1.0000$ which is much higher than the threshold value of 0.70. Therefore, the BDC network models in group 1 are validated using real-world data.

Next, the BDC network models in group 2 are constructed to further validate the network model. For group 2, the targeted model built by using the data on Epi Week 32 and 33 which consists of 12 human nodes and 78 location nodes. From the 78 locations, there are 51 new incoming location nodes found and added into the network model. The database of location is extended from 19 location nodes (during the implementation of the first BDC network) to 27 location nodes (during the implementation of the second BDC network) and now there are 78 location nodes (during the implementation of the third BDC network). On the other hand, the validated model in group 2 utilised the data on Epi Week 34 and 35 that consist of 2 human nodes and 81 location nodes. There are 3 new incoming location nodes added.

Similarly, the DCS values for targeted and validated model in group 2 are calculated and tabulated in Table A.4 (in Appendix A). The locations' ranking results for targeted and validated model are tabulated in Table A.5 (in Appendix A). From the Table A.5, there are four identical location nodes in targeted and validated model. Thus, these four location nodes are used to calculate the SRCC value as depicted in Table 5.8. Equation (5.8) gives the formula used for the calculation of ρ_{TarVar} where $N_{Loc} = 3$ and a is a natural number start from 1. Table

5.8 presents the calculation of various terms in equation (5.8). Expression (5.10) shows the calculation of ρ_{TarVar} according to the equation defined in (5.8).

Table 5.8: SRCC Calculation of Location Nodes between the Models in Group 2

Location Node	DHR _{Targete}	DHR _{Validate}	Rank DHR _{Targete}	Rank DHR _{Validate}	d_{TarVal}	$[d_{TarVal}]^2$
L53	1.1E-124	3.72E-06	3	4	-1	1
L64	1.4E-299	0.774448	4	3	1	1
L69	4E-114	0.859498	2	2	0	0
L77	0.000259	0.978416	1	1	0	0
Sum of $[d_{TarVal}]^2$:						2

$$\rho_{TarVar} = 1 - \frac{6 \times 2}{4(4^2 - 1)} = 0.8 \quad (5.10)$$

The validation result, $\rho_{TarVar} = 0.8$ which is higher than the threshold value of 0.70. Therefore, the BDC network models in group 2 are validated.

From Table 5.4, the BDC network models in group 3 are constructed for validation. For group 3, the targeted model built by using the data on Epi Week 36 and 37 which consists of 10 human nodes and 98 location nodes. From the 98 locations, there are 17 new incoming location nodes found and added into the network model. On the other hand, the validated model in group 3 utilised the data on Epi Week 38 and 39 that consist of 7 human nodes and 100 location nodes. There are 2 new incoming location nodes added.

The quantification processes for nodes and link weights for group 3 are the same with the previous network model as illustrated in Figure 4.6 and Figure 4.8. The locations' ranking results for the networks are tabulated in Table A.6 (in Appendix). There are ten identical location nodes in both of the targeted and validated models. Thus, the SRCC between these

model are calculated in Table 5.9. Equation (5.8) is used and the expression (5.11) shows the calculation of $\rho_{testLoc}$.

Table 5.9: SRCC Calculation of Location Nodes between the Models in Group 3

Location Node	Rank DHR _{Targete}	Rank DHR _{Validate}	d_{TarVal}	$[d_{TarVal}]^2$
L89	3	1	2	4
L90	8	10	-2	4
L91	10	7	3	9
L92	2	3	-1	1
L93	9	8	1	1
L94	7	9	-2	4
L95	5	4	1	1
L96	6	6	0	0
L97	1	2	-1	1
L98	4	5	-1	1
Sum of $[d_{TarVal}]^2$:				26

$$\rho_{TarVar} = 1 - \frac{6 \times 26}{10(10^2 - 1)} = 0.8424 \quad (5.11)$$

The SRCC, $\rho_{TarVar} = 0.8424$ which is higher than the threshold value of 0.70. Therefore, the BDC network models in group 3 are validated. The SRCC values of the network models are tabulated in Table 5.4. From the observation, the SRCC of these three groups are higher than the threshold value 0.70. Thus, the BDC network models are validated.

Combining the results produced from the verification and validation analysis of the research results, it can be concluded that the BDC network model is now verified and validated.

5.4 Parameter Significance Analysis

This analysis aims to determine the relative importance of individual parameter included in the BDC network model. One of the objectives is to identify the critical parameters which have greater influences on the performance of the algorithm. These parameters include seven

location node parameters - Lc, S, B, Pre, Al, H and Fl – and two human parameters - Du and Fh . The analyses performed are (i) “leave-one-out” analysis for each of the parameters; (ii) “leave-multiple-out” analysis for more than one parameter.

In “leave-one-out” analysis, parameter will be excluded one at a time from the BDC network model for computation of DHR and the subsequent processes of the research methodology outlined in the BNM-RMF. The result is tabulated in Table 5.10. For the “leave-multiple-out” analysis, at least two parameters will be excluded from the BDC network model. The result is tabulated in Table 5.11. The new ranking results computed will be compared with the original ranking through calculation of SRCC. The SRCC value will be interpreted based on the general rule of thumb (Hinkle et al, 2003; Mukaka, 2012). These results are used to suggest further analyses to determine the relationship between the parameter(s) that cause the disturbance to the research result and the performance of BDC network model.

General equations used to calculate the SRCC related to location node and human node ranking is given in equations (5.11) and (5.12) respectively. $RankDHR_{BDC}$ and $RankR_{BDC}$ represent to the original location and human node ranking respectively, and $RankDHR_{New}$ and $RankR_{New}$ represent to the new location and human node ranking. $N = 8$ and a is a natural number starts from 1 for both equations.

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

where $\{d_{ParaLoc}\}_a = \{RankDHR_{BDC}\}_a - \{RankDHR_{New}\}_a$ (5.11)

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

where $\{d_{ParaHum}\}_a = \{RankR_{BDC}\}_a - \{RankR_{New}\}_a$ (5.12)

Table 5.10: SRCC for “Leave-one-out” Parameter Analysis

Analysis	Parameter Excluded	ρ_{ParaLoc}	ρ_{ParaHum}
1	<i>Lc</i>	1	0.929
2	<i>S</i>	0.984	0.976
3	<i>B</i>	0.977	0.929
4	<i>Pre</i>	0.802	0.833
5	<i>Al</i>	0.975	0.929
6	<i>K</i>	0.977	0.929
7	<i>Fl</i>	0.977	0.929
8	<i>Du</i>	0.967	0.929
9	<i>Fh</i>	0.707	0.786

Table 5.11: SRCC for “Leave-multiple-out” Parameter Analysis

Analysis	Parameters Excluded	ρ_{ParaLoc}	ρ_{ParaHum}
1	<i>Lc</i> and <i>S</i>	0.930	0.952
2	<i>Lc</i> , <i>S</i> and <i>B</i>	0.688	0.905
3	<i>Lc</i> , <i>S</i> and <i>Pre</i>	0.881	0.905
4	<i>Lc</i> , <i>S</i> , <i>Pre</i> and <i>Al</i>	0.881	0.857
5	<i>Lc</i> , <i>S</i> , <i>Pre</i> , <i>Al</i> and <i>K</i>	0.849	0.857
6	<i>Lc</i> , <i>S</i> , <i>Pre</i> , <i>Al</i> , <i>K</i> and <i>Fl</i>	0.763	0.762
7	<i>Lc</i> , <i>S</i> , <i>Pre</i> , <i>Al</i> , <i>K</i> , <i>Fl</i> and <i>Du</i>	0.643	0.905
8	<i>Lc</i> , <i>S</i> , <i>Pre</i> , <i>Al</i> , <i>K</i> , <i>Fl</i> and <i>Fh</i>	0.361	0.619
9	<i>Lc</i> and <i>B</i>	0.911	0.905
10	<i>B</i> and <i>Du</i>	0.688	0.905
11	<i>B</i> and <i>Fh</i>	0.707	0.786
12	<i>Du</i> and <i>Fh</i>	NA	NA
13	<i>Lc</i> , <i>S</i> , <i>B</i> , <i>Pre</i> , <i>Al</i> , <i>K</i> , <i>Fl</i> and <i>Fh</i>	0.409	0.619

The analysis 1 in Table 5.11 showed that the SRCC value still entails a strong and positive correlation. However, when the life cycle (*Lc*), survival (*S*) and biting rate (*B*) are excluded in analysis 2, the SRCC value drops dramatically to 0.688 which is less than the threshold value 0.70. It suggested that *B* might be the significant parameter. It might be a joint effect of the parameter and *B* is not certainty a significant one. This is proven by the analysis 9 after excluded the combination of *Lc* and *B*, the SRCC remained high correlation. Nevertheless, *B* is maintained and other parameters such as *Pre*, *Al* and *Fl* are excluded from the network as

in analyses 3 until 6, the SRCC remains high correlation. In analysis 7 of the Table 5.11, when the *Du* is excluded with the other 6 parameters, the SRCC drops again to 0.643. Similarly, it is observed that *Du* might be the significant parameter. Thus, *Du* is keep in the network model and to be examined later. In the analysis 8, the SRCC drops to 0.361 which indicates the weakest correlation when the *Fh* is excluded together with the other 6 parameters. Next, *B*, *Du* and *Fh* are further examined to identify the relative importance of the parameters.

In analysis 10, the exclusion of both *B* and *Du* reports a low correlation with the original location node ranking but a high correlation with the human node ranking. However, when *B* and *Fh* are excluded in analysis 10, the SRCC value exceed the threshold value 0.7. These two exclusion analyses showed that the *B*, *Du* and *Fh* are not the significant parameters.

After that, *Du* and *Fh* are excluded in the analysis 12. Since there are only two parameter, *Du* and *Fh* for human nodes, there is no link in the network model if these two parameters are excluded from the network model. Since network model cannot be formulated when *Du* and *Fh* are excluded at the same time, hence the SRCC value is not applicable in this analysis. The exclusion of the eight from nine parameters reports a very low correlation and the exclusion of both *Du* and *Fh* is not applicable. Therefore, it can be concluded that all nine parameters are relatively significant.

In conclusion, a statistical study of the different parameter involved has been carried out to assess the interrelations between the model parameters. All nine parameters are relative importance towards the performance of BDC network model. These parameters need to be taken into account in order to produce an optimal performance of the BDC network model.

5.5 Predictive Power Analysis

This section discusses the potential predictive ability of BNM. According to Shmueli (2010), “a predictive modeling is a process of applying the statistical model or algorithm onto the data for the purpose of predicting the new and future observations.” In other word, prediction includes the observation until time t for the input to forecast new observations at time $t + k, k > 0$ for the output (Geisser, 1993). The predictions include predictive distribution, prediction interval or points, and the new rankings of the new observations. It is important to undergo prediction as a necessary scientific endeavour beyond utility for the purpose of developing and validated theories. Predictive modeling can capture the complex patterns and relationships from the large dataset which are hard to hypothesise and analysis. In short, prediction can be produced regardless of the underlying methods.

From Table 5.4, it is reported that the targeted model possesses the potential to predict the hotspot in the subsequent week that constructed in the validated model due to the high SRCC values (exceed the threshold value, 0.70).

Therefore, the section further investigates the potential of bipartite network model to predict the possible hotspot. To examine the predictive power, the ranking of the targeted model need to be compared with the ranking of the validated model to check whether the top-ranked location of the targeted model remained in the top position in the validated model.

In order to compare these top-ranked locations, all link weights of the location nodes need to be considered in the validated model. In other words, the location nodes where do not visited by the patients in the validated model will be taken into the calculation. In the first validated model shown in Table 5.5, only location nodes L2, L4, L5, and L20 until L27 are

visited by the patients during Epi Week 30 and Epi Week 31. In this section, the link weight of the L1, L3, L6 until L19 need to be calculated.

The location parameters include Lc, S, B, Pre, Al, K and Fl and human parameters include Du and Fh of all 27 location nodes are tabulated in Table A.7 (Appendix A). Even though there is no patient visited some of the location nodes during EW 30 and 31, the environmental effects of these locations still remained in the model. The link weight of the locations in this validated model is the summation effect between the current model and the previous model as shown in Table A.7. These link weights are DCS that acts as an input to the HITS search algorithm. The resulted ranking is tabulated in Table 5.12.

From the Table 5.12, it is observed that 4 of the top-5 ranked locations in network 1 remained on the top-5 ranked locations in network 2. The accuracy of the network model to predict the hotspot is given in equation (5.13) which is 80%. A similar research revealed that 60% accuracy in predicting the dengue hotspot in Subang Jaya, Malaysia (Dom, Ahmad, Latif & Ismail, 2017).

$$\begin{aligned}
 Accuracy &= \frac{\text{Number of the Remained Location}}{\text{Top} - N \text{ Ranked Location}} \times 100\% \\
 &= \frac{4}{5} \times 100\% \\
 &= 80\%
 \end{aligned} \tag{5.13}$$

There is a list of hotspot location identified based on the dengue cases in overall obtained from the public health authority. The GPS coordinates of these 9 hotspots are identified and labelled as P1, P2, ..., P8 and P9. The distance matrix between the location nodes of network model and the real data is computed and tabulated in Table 5.13. From Table 5.13, the distance between L3 and P6 is only 140 meters. In other words, the network model in Table

5.12 managed to predict the hotspot where only 140 meter away from one of the actual hotspot released by public health. This further validated that the BNM approach possesses the potential to predict the future hotspot.

Table 5.12: Result of Prediction

Ranking	Network 1	Network 2
1	L4	L4
2	L12	L3
3	L17	L17
4	L19	L19
5	L3	L5
6	L11	L7
7	L16	L11
8	L7	L1
9	L6	L16
10	L18	L13
11	L13	L9
12	L10	L15
13	L2	L18
14	L15	L6
15	L1	L8
16	L5	L14
17	L8	L10
18	L9	L12
19	L14	L2
20	-	L20
21	-	L21
22	-	L22
23	-	L23
24	-	L27
25	-	L26
26	-	L25
27	-	L24

The targeted model includes the possible positive patients, this promotes the predictive power since the new location where never visited by the dengue positive patient will be included.

The environmental characteristics and the contact frequency between the dengue positive and possible positive patients will be quantified into the network.

Table 5.13: Distance Matrix of Location Node versus the Real Data

	P1	P2	P3	P4	P5	P6	P7	P8	P9
L1	4.52	18.94	20.04	33.73	14.83	34.14	35.09	36.76	36.1
L2	30.83	45.12	11.91	4.58	16.85	4.14	5.33	6.54	6.27
L3	35.03	48.56	15.25	4.59	20.46	0.14	1.2	2.68	2.03
L4	35.31	49.84	16.58	2.45	21.59	2	3.38	2.4	3.61
L5	24.59	36.78	3.47	13	8.73	11.65	12.37	14.45	13.37
L6	45.72	59.35	26.12	12.15	31.38	11.08	10.3	8.41	9.32
L7	41.31	50.73	19.54	14.78	24.76	10.31	9.14	10.6	8.81
L8	36.95	51.93	18.71	3.12	23.66	4.56	4.99	2.84	4.81
L9	41.07	54.6	21.35	8.03	26.6	6.31	5.54	3.71	4.58
L10	34.03	49.81	16.96	0.96	21.58	5.51	6.53	5.43	6.85
L11	7.36	26.14	14.89	26.62	10.63	27.53	28.58	30.01	29.58
L12	13.21	32.13	12.24	20.8	10.19	22.27	23.41	24.55	24.38
L13	41.44	52.44	20.13	12.07	25.48	8.02	6.74	7.48	6.08
L14	17.98	15.22	18.75	35.06	14.28	33.67	34.21	36.49	35.16
L15	38.74	51.4	18.3	7.67	23.63	4.04	2.83	3.05	1.9
L16	24.61	41.83	11.59	9.57	14.51	12.01	13.27	13.8	14.13
L17	16.62	32.83	37.05	45.49	32.91	47.7	48.91	49.74	49.84
L18	18.64	27.38	6.02	22.33	2.78	21.05	21.68	23.87	22.65
L19	35.65	50.28	17.03	2.45	22.03	2.97	3.67	2.24	3.79

5.6 Data Size Analysis

The size of data used in the study is associated with the performance of a network model. This section discusses the effect to the model accuracy when data size used to formulate the BDC network is varied by increasing the size gradually. Implementation design of this analysis is shown in Figure 5.3 where the three main processes are shown on the left, inputs and outputs are shown on the right.

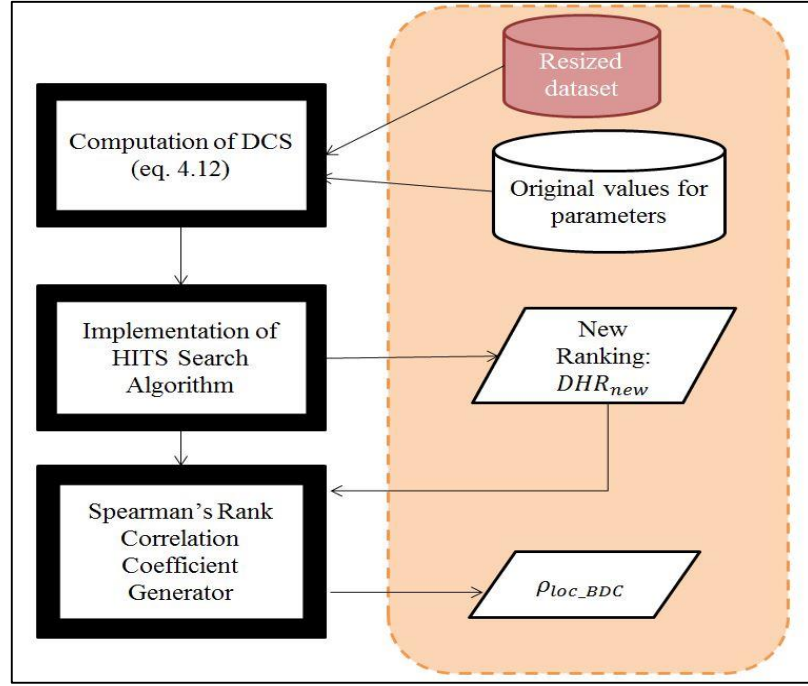


Figure 5.3: Implementation Design of Data Size Analysis

The data used to formulate one BDC network model is the human movement of two Epi Week. As shown in Figure 5.3, the resized data is needed to formulate new BDC network models, targeted and validated models. Together with the original values of parameters for both location and human nodes, new DCS matrix is generated based on the equation 4.12. The new DCS matrix computed is then passed to the implementation of HITS search algorithm. After that, new ranking values for location node represented as DHR_{new} of the targeted and validated models are produced. The degree of similarity between the targeted and validated models is measured using SRCC as indicated in the final process of Figure 5.3. Interpretation of the coefficient obtained ρ_{loc_new} will follow the general rule of thumb for correlation coefficient (Mukaka, 2012).

Table 5.14 shows the summary of the characteristics for the new BDC network models constructed respectively. These three new models are represented as BDC, BDC-4 and BDC-6

where the BDC model is the original network models which utilise 2 weeks in formulating the BDC network model, BDC-4 model utilise 4 weeks and BDC-6 used 6 weeks in formulating the BDC network models.

Table 5.14: Summary of Characteristics of the New BDC Network

Network Model	Number of Epi Week per Model	Model	Epi Week
BDC	2	Targeted model	28-29
		Validated model	30-31
BDC-4	4	Targeted model	28-31
		Validated model	32-35
BDC-6	6	Targeted model	28-33
		Validated model	34-39

The SRCC calculation of locations' ranking for BDC-4 are tabulated in Table A.8 (in Appendix) and for BDC-6 are tabulated in Table A.9 (in Appendix). The SRCC of the BDC-4 and BDC-6 can be obtained by using the equation 5.14 and 5.15.

$$\rho_{loc_{BDC4}} = \left| 1 - \frac{6 \times 98}{8(8^2 - 1)} \right| = 0.1667 \quad (5.14)$$

$$\rho_{loc_{BDC6}} = \left| 1 - \frac{6 \times 34}{5(5^2 - 1)} \right| = 0.7 \quad (5.15)$$

Table 5.15: SRCC for Comparison of BDC Networks' Location Ranking

Network Model	SRCC, ρ_{loc_BDC}
BDC	1.000
BDC-4	0.1667
BDC-6	0.7000

Today big data is expensive to maintain, manage and process. The larger the data we use, the higher the cost to process and analyse it. This analysis aimed to speculate whether the bigger data size would improve the model performance. According to the result presented in Table 5.15, the new rankings of location nodes for BDC-4 and BDC-6 models are lower than

the original BDC network model. The SRCC of BDC-4 is 0.1667 implies a very weak correlation between the targeted and validated model. This value is also far lower than the threshold value of 0.70. On the other hand, the SRCC of BDC-6 is equal to the threshold value of 0.70 which implies a strong correlation. However, the SRCC value of BDC-6 is much lower compared with the original BDC network model which utilise only two epi weeks. Based on Table 5.15, a network model is able to generate effective result with small data set. This is due to the fact that the patient who is already infected with the dengue virus can only transmit the infection for maximum 12 to 14 days via *Aedes* mosquitoes after the first symptoms appear (WHO, 2012). Therefore, two weeks is appropriate to use in formulate a BDC network. BDC-4 and BDC-6 used four and six weeks exceed the infectious period and therefore, this network cannot detect the dengue hotspot effectively even though the data size is bigger. The temporal aspect cannot be neglected when formulating the network model for location detection. Vary the number of weeks per one network does affect the model accuracy. In short, a network model is able to generate effective result with small data set.

This study collected the patient movement data at district level. However, there is a possibility that the patient will visit other states than Kuching. The environmental properties of other states are different from Kuching. Vary the number of patients per one network when the number of Epi Week remained constant as 2, I hypothesise that by increasing the data size from district level to national level does not affect the model accuracy. This can be included as future work as the data collection at national level is out of this research scope.

5.7 Summary

This study has managed to verify and validate this ranking result. Verification analysis is achieved through comparing the result with the result produced from UCINET 6.0 as

benchmark system. An RMSE of 0.0006419 is reported for the ranking of location nodes and human nodes respectively. These RMSE values are lower than the threshold value of 0.05 signifies that the BDC network model is verified. Besides that, analytical verification is carried out, then shows the DHR and a hub matrix of location nodes are highly correlated as concluded from the SRCC 0.868 reported. The analytical verification further confirms the result of benchmark verification and both verification analyses have verified the model.

Validation analyses are accomplished via comparison of research result between the targeted and validated model. SRCC of 1 is reported for the ranking of location nodes in group 1. The SRCC of group 2 is 0.8 while the SRCC of group 3 is 0.8424. These coefficients are much higher than the threshold value of 0.70 in this research, which signifies that the research result is positive and very highly correlated with the result obtained from the real data. This validation analysis has validated the model formulated in this research. Consequently, it can be concluded that the BDC network model is verified and validated.

Parameter significance analysis was conducted to identify the important parameters in the network. It is observed that there is no single parameter has higher influence to affect the ranking. This is validated by the low SRCC value when multiple parameters are leave-out. The BNM approach possess the potential to predict the possible hotspot and the accuracy is 80%. Data size analysis is conducted and BNM approach can generate effective result with small data set. In the next chapter, the discussion of the summary and the result and the contribution of the research with recommendations for future research will be presented.

CHAPTER 6

CONCLUSION

6.1 Introduction

This chapter concludes the overall implementation of this research project and it summarises the contributions of the present study to the related body of knowledge. Possible future work that can be done for this research is also discussed.

6.2 Conclusion of Research

A case study of modeling the location detection of dengue has been investigated in this study. Current control measures eradicate the spread of dengue by fogging the reported residential neighbourhood. It was reported that the proactive approach, for instance identifies the possible breeding sites of the vector before disease outbreak can eradicate dengue. However, the typical studies of dengue transmission, for example compartmental modelling, are lacking in support to detect risk location. The bipartite network modelling approach was able to integrate the human mobility to detect the breeding site.

This study aimed to formulate a network model to describe the contact between the human and the location that human visited. In order to achieve the main research objective, three main components - host, location and environmental properties are identified in typical dengue transmission studies. In stage 1, ET was formulated and served as the basic building block for bipartite graphs. This basic building block was then used to construct the bipartite graph.

In stage 2, the data pre-processing algorithm was developed to identify the GPS coordinates and generate the distance matrix to calculate the distance between the location nodes. The human data was pre-processed and the human nodes were replaced by the algorithm-generated ID. The pre-processing of the location parameters included life cycle (*Lc*), survival (*S*), biting (*B*), precipitation (*Pre*), altitude (*Al*), humidity (*K*) and number of times a location is visited by human (*Fl*). Polynomial curve fitting was used to generate life cycle and survival from the experimental data obtained from literatures. *Fl* was quantified through the sum of the link matrix. For the human parameter, the duration time on the human stay at one location (*Du*) and number of times a human visited a location (*Fh*) were identified. The summation rule was used to quantify the link weight (DCS) in Table 4.14. As a case study, the first bipartite dengue contact (BDC) network of Epi Week 28 and 29 was constructed, the network consists of eight human nodes, nineteen location nodes and twenty weighted links as presented.

The BDC network, then served as an input to the adapted HITS search algorithm to generate DHR value for each location node and ranking values for each human node. The location nodes were ranked according to the DHR values which also a measurement of vector density. At the end of this stage, the bipartite dengue contact network was formulated.

UCINET was used to verify the model in stage 3. The DHR values obtained in the study were compared with the estimated DHR values produced through RMSE analysis. The RMSE of location node is 0.0006419 and RMSE of human node is 0.0009188. Both values are well below the threshold value of 0.05. Besides, analytical verification was performed by calculating the correlation value (SRCC) between DHR values and the sum of the DCS for location nodes. The analysis reported a value of 0.868 which are exceeding the threshold value of 0.70. The analysis showed the model is high correlation and well above the hypothesised value. For model validation, the SRCC value between the targeted and validated model was calculated. The

SRCC for the first targeted and the first validated model of group 1 is 1.00 which imply the perfect positive correlation.

Next, parameter significance analysis was performed to study the relative importance of each parameter. The output of the analysis supports the effectiveness of the BDC network model. There is a link association between the parameters life cycle (Lc), survival (S), biting (B), precipitation (Pre), altitude (Al), humidity (K), frequency of a location is visited by human (Fl), duration time on the human stay at one location (Du) and the frequency of a human visited location (Fh).

In addition, the predictive power analysis was conducted and the BNM approach possess the potential to predict the dengue hotspot in the subsequent week and the accuracy obtained is 80%.

Last but not least, data size analysis was conducted and this analysis revealed that a network model is able to generate effective result with small data set.

6.3 Research Contributions

There are several contributions that this research has achieved upon completion:

- (a) A contact network model of dengue transmission where “human heterogeneity” and “vector characteristics” are put into consideration was developed.
- (b) This research successfully demonstrated the BNM-RMF (Figure 1.1) to formulate a network model by following the processes involved in the BNB-RMF.
- (c) The application of the Epidemiological Triangle (ET) in this research can be applied in any other vector-borne disease problem. ET and the basic building block are important to understand in order to employ the bipartite network approach.

- (d) Algorithm to identify the distance matrix to declare the location node was developed in the data pre-processing step. This algorithm can improve the computing time and the solution quality, it is time-saving for the researcher in preparing the data ready for the quantification process.
- (e) This research developed a technique in quantifying the life cycle and survival rate parameters. There is no direct life cycle and survival rate data have been recorded from any data source. Thus, polynomial curve fitting is used to imply the life cycle and survival parameter.
- (f) This research developed the dengue hotspot ranking (DHR) value for location to capture the interaction between the behaviour of human, the environmental properties of the location and the vector characteristic.
- (g) This research is able to rank the location nodes according to the vector density and the risk level of the location. The human node's ranking can help identify the most significant individual which can contribute in recognising the possible breeding locations.
- (h) This research identified significant parameters to detect possible breeding site for mosquito, which includes mosquito's life cycle, survival and biting rate; precipitation; humidity; altitude; the number of times human visited a location; duration of human stay one location and the number of times a human visited a location.
- (i) The spatial and temporal effects are accounted for the formulation of the contact network model where the subsequent week data is used to identify the usefulness of the network model. The spatial component is integrated and projected onto the network modeling approach to support location detection.

6.4 Future Work

This section outlines the recommendations for future work with the objective of improve and expand the research.

- (a) This research reported that by using the basic building block in ET (Figure 3.3) can detect the possible breeding location of dengue in Kuching by ranking the additional location nodes, however there is no temporal element in the framework. Based on the data size analysis in section 5.6, latent period and the timeline of a patient first symptom emerge is crucial in determining a dengue hotspot. A new layer, time can be incorporated into the modified ET to serve as a new basic building block especially in determine the hotspot for mosquito-borne diseases.
- (b) It is time consuming to clean and prepare the data input for instance transforming the address into a geolocation coordinate, check the distance of each location node in the database, quantify the parameter values. The process is done manually, thus, a automated system should be developed to reduce the processing time of the data cleaning.
- (c) The link association between the parameters life cycle (Lc), survival (S), precipitation (Pre), altitude (Al), humidity (K), frequency of a location is visited by human (Fl) and the frequency of a human visited location (Fh) can be further investigate to understand the joint effect of the climate variability and socio-ecological factors on dengue transmission.
- (d) BNM approach can be applied in different domain for future research.

Transportation Modeling

Transportation management is important, especially in the countries with high concentration of economic activities such as Malaysia. Congestion is one of the most prevalent transport problems in large urban cities such as Kuala Lumpur, Bangkok, Beijing and other large cities. A brief reviewing on a few literature suggest that reduction in travel time can overcome the congestion problems. This can be done by identifying the potential congestion location and determine the shortest route to reach the destination. Network-based approach can be applied to solve different transportation networks for instance airport, urban road and railway networks (Mukherjee, 2012; Seaton & Hackett, 2004; Towles & Dally, 2001; Wang, Zhou, Shi, Wang, & He, 2009; Wu et al., 2006). These surveys suggest that the transportation modelling could be represented by two distinct components: location and vehicle. For illustration, a railway bipartite network can be constructed with station and train. A train connects to the stations where it stops. The shortest path and the most efficient route can be identified to avoid the train idle time. Thus, this can optimise the resources allocation since there is no congestion in one particular busy station where more than one train need to queue and idle when the train is waiting. The busiest station can be identified so that the train frequency and the number of train will be increase and allocate to the particular congestion station to reduce the railway traffic problem. The traffic characteristic includes number of connecting trains of one route, train speed, duration required to reach different stations, and environmental factors can be incorporated into the network.

REFERENCES

- Akter, R., Hu, W., Naish, S., Banu, S., & Tong, S. (2017). Joint effects of climate variability and socio-ecological factors on dengue transmission: Epidemiological evidence. *Tropical Medicine & International Health*, 22(6), 656-669.
- Albright, J. J., & Park, H. M. (2009). Confirmatory factor analysis using amos, LISREL, Mplus, SAS/STAT CALIS. Retrieved from <http://www.iub.edu/~statmath/stat/all/cfa/cfa2008.pdf>.
- Arias, J. R. (2016). *What is an "Epi Week?"* Retrieved from <http://www.cmmcp.org/epiweek.htm>.
- Arifin, S. M., Arifin, R. R., Pitts, D. D. A., Rahman, M. S., Nowreen, S., Madey, G. R., & Collins, F. H. (2015). Landscape epidemiology modeling using an agent-based model and a geographic information system. *Land*, 4(2), 378-412.
- Aziz, S., Aidil, R. M., Nisfariza, M. N., Ngui, R., Lim, Y. A. L., Yusoff, W. W., & Ruslan, R. (2014). Spatial density of Aedes distribution in urban areas: A case study of breteau index in Kuala Lumpur, Malaysia. *Journal of Vector Borne Diseases*, 51(2), 91.
- 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.
- Barnes, J. A., & Harary, F. (1983). Graph theory in network analysis. *Social Networks*, 5(2), 235-244.
- Belik, V. V., Geisel, T., & Brockmann, D. (2009, August). The impact of human mobility on spatial disease dynamics. In *Computational Science and Engineering, 2009. CSE'09. International Conference on* (Vol. 4, pp. 932-935). IEEE.
- 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.

- Benedict, M. Q., Levine, R. S., Hawley, W. A., & Lounibos, L. P. (2007). Spread of the tiger: global risk of invasion by the mosquito *Aedes albopictus*. *Vector-Borne and Zoonotic Diseases*, 7(1), 76-85.
- Bian, L., & Liebner, D. (2007). A network model for dispersion of communicable diseases. *Transactions in GIS*, 11(2), 155-173.
- Biggerstaff, B. J., Gallagher, N., Marano, N., Staples, J. E., & Johansson, M. A. (2012). Assessing the risk of international spread of yellow fever virus: a mathematical analysis of an urban outbreak in Asuncion, 2008. *The American Journal of Tropical Medicine and Hygiene*, 86(2), 349-358.
- Bishop, C. M. (2013). *Pattern recognition and machine learning*. Springer, New York.
- Borgatti, S. P., Everett, M. G., & Freeman, L. C. (2002). Ucinet for Windows: Software for social network analysis. Harvard, MA Analytic Technologies.
- Bouzerdoun, A., Havstad, A., & Beghdadi, A. (2004). Image quality assessment using a neural network approach. In *Signal Processing and Information Technology, 2004. Proceedings of the Fourth IEEE International Symposium on* (pp. 330-333). IEEE.
- Boyd, J. P., Fitzgerald, W. J., & Beck, R. J. (2007). Computing core/periphery structures and permutation tests for social relations data. *Social Networks*, 28(2), 165-178.
- Browne, M. W., & Cudeck, R. (1992). Alternative ways of assessing model fit. *Sociological Methods & Research*, 21(2), 230-258.
- Bruch, E., & Atwell, J. (2015). Agent-based models in empirical social research. *Sociological Methods & Research*, 44(2), 186-221.
- Carrington, L. B., Armijos, M. V., Lambrechts, L., Barker, C. M., & Scott, T. W. (2013). Effects of fluctuating daily temperatures at critical thermal extremes on *Aedes aegypti* life-history traits. *PloS One*, 8(3), e58824.

- Carrington, L. B., Armijos, M. V., Lambrechts, L., & Scott, T. W. (2013). Fluctuations at a low mean temperature accelerate dengue virus transmission by *Aedes aegypti*. *PLoS Neglected Tropical Diseases*, 7(4), e2190.
- CDC. (2014). *Dengue*. Retrieved from <https://www.cdc.gov/dengue/index.html>
- CDC. (2014). *Laboratory Guidance and Diagnostic Testing*. Retrieved from <https://www.cdc.gov/dengue/clinicallab/laboratory.html>
- Centers for Disease Control and Prevention. (2010). *Epidemiology*. Retrieved from <https://www.cdc.gov/dengue/epidemiology/index.html>
- Chan, M., & Johansson, M. A. (2012). The incubation periods of dengue viruses. *PloS One*, 7(11), e50972.
- Cheng, A. D., Golberg, M. A., Kansa, E. J., & Zangwill, G. (2003). Exponential convergence and H-c multiquadric collocation method for partial differential equations. *Numerical Methods for Partial Differential Equations*, 19(5), 571-594.
- Chiu, T. L., & So, S. S. (2004). Development of neural network QSPR models for Hansch substituent constants. 2. Applications in QSAR studies of HIV-1 reverse transcriptase and dihydrofolate reductase inhibitors. *Journal of Chemical Information and Computer Sciences*, 44(1), 154-160.
- Choi, S. (2005, September). On variations of power iteration. In *International Conference on Artificial Neural Networks* (pp. 145-150). Springer, Berlin, Heidelberg.
- 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), 20-23.
- Craft, M. E., & Caillaud, D. (2011). Network models: an underutilized tool in wildlife epidemiology? *Interdisciplinary Perspectives on Infectious Diseases*, 2011, 1-12.

- 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.
- Davis, P. K. (1992). *Generalizing concepts and methods of verification, validation, and accreditation (VV&A) for military simulations*. RAND CORP SANTA MONICA CA.
- 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.
- Dom, N. C., Ahmad, A. H., Latif, Z. A., & Ismail, R. (2017). Integration of GIS-Based Model with Epidemiological data as a Tool for Dengue Surveillance. *Environment Asia*, 10(2), 135-146.
- Dommar, C. J., Lowe, R., Robinson, M., & Rodó, X. (2014). An agent-based model driven by tropical rainfall to understand the spatio-temporal heterogeneity of a chikungunya outbreak. *Acta Tropica*, 129, 61-73.
- Esteva, L., & Yang, H. M. (2005). Mathematical model to assess the control of *Aedes aegypti* mosquitoes by the sterile insect technique. *Mathematical Biosciences*, 198(2), 132-147.
- Esteva, L., & Vargas, C. (2003). Coexistence of different serotypes of dengue virus. *Journal of Mathematical Biology*, 46(1), 31-47.
- Eze, M. O. (2013). *Web Algorithm search engine based network modelling of Malaria Transmission*. Doctoral dissertation, Universiti Malaysia Sarawak (UNIMAS), Sarawak, Malaysia.
- Eze, M., Labadin, J., & Lim, T. (2011a). Emerging computational strategy for eradication of malaria. In *Computers & Informatics (ISCI), 2011 IEEE Symposium on* (pp. 715-720). IEEE.
- Eze, M., Labadin, J., & Lim, T. (2011b). Contact strength generating algorithm for application in malaria transmission network. In *Information Technology in Asia (CITA 11), 2011 7th International Conference on* (pp. 1-6). IEEE.

- Fehlberg, E. (1969). Low-order classical Runge-Kutta formulas with stepsize control and their application to some heat transfer problems. *NASA Technical Report*, 315.
- Focks, D. A., Haile, D. G., Daniels, E., & Mount, G. A. (1993). Dynamic life table model for *Aedes aegypti* (Diptera: Culicidae): analysis of the literature and model development. *Journal of Medical Entomology*, 30(6), 1003-1017.
- Focks, D. A., Patz, J. A., Martens, W. J., & Jetten, T. H. (1998). Dengue fever epidemic potential as projected by general circulation models of global climate change. *Environmental Health Perspectives*, 106(3), 147.
- Fry, S. R., Meyer, M., Semple, M. G., Simmons, C. P., Sekaran, S. D., Huang, J. X., McElnea, C., Huangm C. Y., & Cooper, M. A. (2011). The diagnostic sensitivity of dengue rapid test assays is significantly enhanced by using a combined antigen and antibody testing approach. *PLoS Neglected Tropical Diseases*, 5(6), e1199.
- Fujimura, K., Sawaki, H., Sakai, T., Hiruma, T., Nakanishi, N., Sato, T., & Narimatsu, H. (2005). LARGE2 facilitates the maturation of α -dystroglycan more effectively than LARGE. *Biochemical and biophysical research communications*, 329(3), 1162-1171.
- Galea, S., Riddle, M., & Kaplan, G. A. (2010). Causal thinking and complex system approaches in epidemiology. *International Journal of Epidemiology*, 39(1), 97-106.
- Gardner, L. M., & Sarkar, S. (2015). Risk of dengue spread from the Philippines through international air travel. *Transportation Research Record. Journal of the Transportation Research Board*, 2501, 25-30.
- Garnett, G. P. (2002). An introduction to mathematical models in sexually transmitted disease epidemiology. *Sexually Transmitted Infections*, 78(1), 7-12.

- Garnett, G. P., Cousens, S., Hallett, T. B., Steketee, R., & Walker, N. (2011). Mathematical models in the evaluation of health programmes. *The Lancet*, 378(9790), 515-525.
- Graham, M., & House, T. (2014). Dynamics of stochastic epidemics on heterogeneous networks. *Journal of Mathematical Biology*, 68(7), 1583-1605.
- Handcock, M. S., & Jones, J. H. (2004). Likelihood-based inference for stochastic models of sexual network formation. *Theoretical Population Biology*, 65(4), 413-422.
- Harary, F., Hedetniemi, S. T., & Robinson, R. W. (1969). Uniquely colorable graphs. *Journal of Combinatorial Theory*, 6(3), 264-270.
- Hassarangsee, S., Tripathi, N. K., & Souris, M. (2015). Spatial pattern detection of tuberculosis: a case study of Si Sa Ket Province, Thailand. *International Journal of Environmental Research and Public Health*, 12(12), 16005-16018.
- Heath, B., Hill, R., & Ciarallo, F. (2009). A survey of agent-based modeling practices (January 1998 to July 2008). *Journal of Artificial Societies and Social Simulation*, 12(4), 9.
- Hethcote, H. W. (2000). The mathematics of infectious diseases. *SIAM Review*, 42(4), 599-653.
- Hinkle, D. E., Wiersma, W., & Jurs, S. G. (2003). *Applied statistics for the behavioral sciences* (Fifth edition). Boston, MA: Houghton Mifflin Company.
- Jang, M. J., Chen, C. L., & Liu, Y. C. (2001). Two-dimensional differential transform for partial differential equations. *Applied Mathematics and Computation*, 121(2), 261-270.
- Jeefoo, P., Tripathi, N. K., & Souris, M. (2010). Spatio-temporal diffusion pattern and hotspot detection of dengue in Chachoengsao province, Thailand. *International Journal of Environmental Research and Public Health*, 8(1), 51-74.

- Jie, S., Chen, C., Hui, Z., Rong-Shuang, S., Yan, Z., & Kun, H. (2008). Tagrank: a new rank algorithm for webpage based on social web. *Computer Science and Information Technology*, 3(2), 254-258, IEEE.
- Johansson, M. A., Dominici, F., & Glass, G. E. (2009). Local and global effects of climate on dengue transmission in Puerto Rico. *PLoS Neglected Tropical Diseases*, 3(2), e382.
- Jong, S. L., Lam, H. Y., D'Amico, M., & Din, J. (2016, August). Analysis of precipitation characteristics over Southern peninsular Malaysia for satellite propagation application. In *URSI Asia-Pacific Radio Science Conference (URSI AP-RASC)* (pp. 1729-1732). IEEE.
- Kesorn, K., Ongruk, P., Chompoosri, J., Phumee, A., Thavara, U., Tawatsin, A., & Siriyasatien, P. (2015). Morbidity rate prediction of dengue hemorrhagic fever (DHF) using the support vector machine and the *Aedes aegypti* infection rate in similar climates and geographical areas. *PloS One*, 10(5), e0125049.
- Kon, C., & Labadin, J. (2013, July). Reaction-diffusion generic model for mosquito-borne diseases. In *Information Technology in Asia (CITA), 2013 8th International Conference on* (pp. 1-4). IEEE.
- Kramer, L. D., & Ebel, G. D. (2003). Dynamics of flavivirus infection in mosquitoes. *Advances in Virus Research*, 60, 187-232.
- Kyle, J. L., & Harris, E. (2008). Global spread and persistence of dengue. *Annual Review of Microbiology*, 62, 71-92.
- Lam, S. K. (1994). Strategies for dengue control in Malaysia. *Tropical Medicine*, 35(4), 303-307.
- Lambrechts, L., Paaijmans, K. P., Fansiri, T., Carrington, L. B., Kramer, L. D., Thomas, M. B., & Scott, T. W. (2011). Impact of daily temperature fluctuations on dengue virus transmission by *Aedes aegypti*. *Proceedings of the National Academy of Sciences*, 108(18), 7460-7465.

- Liew, C. Y. (2016). *Bipartite Network Modeling of Habitat Suitability*. Unpublished doctoral dissertation, Universiti Malaysia Sarawak, (UNIMAS), Sarawak, Malaysia.
- Liew, C. Y., Labadin, J., Chai, W. Y., Tuen, A. A., & Peter, C. (2015, May). Modeling using support vector machines on imbalanced data: A case study on the prediction of the sightings of Irrawaddy dolphins. In *AIP Conference Proceedings* (Vol. 1660, No. 1, p. 050011). AIP Publishing.
- 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.
- Luz, P. M., Lima-Camara, T. N., Bruno, R. V., Castro, M. G. D., Sorgine, M. H. F., Lourenço-de-Oliveira, R., & Peixoto, A. A. (2011). Potential impact of a presumed increase in the biting activity of dengue-virus-infected *Aedes aegypti* (Diptera: Culicidae) females on virus transmission dynamics. *Memórias do Instituto Oswaldo Cruz*, 106(6), 755-758.
- Maciel-de-Freitas, R., Eiras, Á. E., & Lourenço-de-Oliveira, R. (2008). Calculating the survival rate and estimated population density of gravid *Aedes aegypti* (Diptera, Culicidae) in Rio de Janeiro, Brazil. *Cadernos de saude publica*, 24(12), 2747-2754.
- Maidana, N. A., & Yang, H. M. (2008). Describing the geographic spread of dengue disease by traveling waves. *Mathematical Biosciences*, 215(1), 64-77.
- Malik, H. A. M., Mahesar, A. W., Abid, F., & Wahiddin, M. R. (2014, November). Two-mode complex network modeling of dengue epidemic in Selangor, Malaysia. In *Information and Communication Technology for The Muslim World (ICT4M), 2014 The 5th International Conference on* (pp. 1-6). IEEE.
- Mason, O., & Verwoerd, M. (2007). Graph theory and networks in biology. *IET Systems Biology*, 1(2), 89-119.

- Meyers, L. A., Newman, M. E. J., & Pourbohloul, B. (2006). Predicting epidemics on directed contact networks. *Journal of Theoretical Biology*, 240(3), 400-418.
- Miksch, F., Urach, C., Einzinger, P., & Zauner, G. (2014, April). A Flexible Agent-Based Framework for Infectious Disease Modeling. In *Information and Communication Technology-EurAsia Conference* (pp. 36-45). Springer, Berlin, Heidelberg.
- Mukaka, M. M. (2012). A guide to appropriate use of correlation coefficient in medical research. *Malawi Medical Journal*, 24(3), 69-71.
- Mukherjee, S. (2012). Statistical analysis of the road network of India. *Pramana-Journal of Physics*, 79(3), 483-491.
- Nagao, Y., Thavara, U., Chitnumsup, P., Tawatsin, A., Chansang, C., & Campbell-Lendrum, D. (2003). Climatic and social risk factors for Aedes infestation in rural Thailand. *Tropical Medicine & International Health*, 8(7), 650-659.
- Najid, S. K., Zaharim, A., Razali, A. M., Zainol, M. S., Ibrahim, K., & Sopian, K. (2009). Analyzing the east coast Malaysia wind speed data. *International Journal of Energy and Environment*, 3(2), 53-60.
- 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.
- Newton, E. A., & Reiter, P. (1992). A model of the transmission of dengue fever with an evaluation of the impact of ultra-low volume (ULV) insecticide applications on dengue epidemics. *The American Journal of Tropical Medicine and Hygiene*, 47(6), 709-720.
- Normalisation Overview. (2004). Retrieved February 11, 2018, from

http://www.improvedoutcomes.com/docs/WebSiteDocs/PreProcessing/Normalization/Normalization_Overview.htm

- Packierisamy, P. R., Ng, C. W., Dahlui, M., Inbaraj, J., Balan, V. K., Halasa, Y. A., & Shepard, D. S. (2015). Cost of dengue vector control activities in Malaysia. *The American Journal of Tropical Medicine and Hygiene*, 93(5), 1020-1027.
- Parikh, N., Youssef, M., Swarup, S., & Eubank, S. (2013). Modeling the effect of transient populations on epidemics in Washington DC. *Scientific Reports*, 3, 3152.
- Parkash, O., & Shueb, R. H. (2015). Diagnosis of dengue infection using conventional and biosensor based techniques. *Viruses*, 7(10), 5410-5427.
- 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.
- Perez, L., & Dragicevic, S. (2009). An agent-based approach for modeling dynamics of contagious disease spread. *International Journal of Health Geographics*, 8(1), 50.
- Phaijoo, G. R., & Gurung, D. B. (2015). Mathematical Study of Biting Rates of Mosquitoes in Transmission of Dengue Disease. *Journal of Science, Engineering and Technology*, 11, 25-33.
- Poletto, C., Tizzoni, M., & Colizza, V. (2013). Human mobility and time spent at destination: impact on spatial epidemic spreading. *Journal of Theoretical Biology*, 338, 41-58.
- Potterat, J. J., Rothenberg, R. B., & Muth, S. Q. (1999). Network structural dynamics and infectious disease propagation. *International Journal of STD & AIDS*, 10(3), 182-185.
- Reiner, R. C., Stoddard, S. T., & Scott, T. W. (2014). Socially structured human movement shapes dengue transmission despite the diffusive effect of mosquito dispersal. *Epidemics*, 6, 30-36.

- Ritchie, S. A., & Johnson, B. J. (2017). Advances in Vector Control Science: Rear-and-Release Strategies Show Promise... but Don't Forget the Basics. *The Journal of Infectious Diseases*, 215(suppl_2), S103-S108.
- Rohani, A., Wong, Y. C., Zamre, I., Lee, H. L., & Zurainee, M. N. (2009). The effect of extrinsic incubation temperature on development of dengue serotype 2 and 4 viruses in *Aedes aegypti* (L.). *Southeast Asian Journal of Tropical Medicine and Public Health*, 40(5), 942.
- Rueda, L. M., Patel, K. J., Axtell, R. C., & Stinner, R. E. (1990). Temperature-dependent development and survival rates of *Culex quinquefasciatus* and *Aedes aegypti* (Diptera: Culicidae). *Journal of Medical Entomology*, 27(5), 892-898.
- 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.
- Sarfraz, M. S., Tripathi, N. K., Tipdecho, T., Thongbu, T., Kerdthong, P., & Souris, M. (2012). Analyzing the spatio-temporal relationship between dengue vector larval density and land-use using factor analysis and spatial ring mapping. *BMC Public Health*, 12(1), 853.
- Scott, T. W., Amerasinghe, P. H., Morrison, A. C., Lorenz, L. H., Clark, G. G., Strickman, D., Kittayapong, P., & Edman, J. D. (2000). Longitudinal studies of *Aedes aegypti* (Diptera: Culicidae) in Thailand and Puerto Rico: blood feeding frequency. *Journal of Medical Entomology*, 37(1), 89-101.
- Seaton, K. A., & Hackett, L. M. (2004). Stations, trains and small-world networks. *Physica A: Statistical Mechanics and its Applications*, 339(3), 635-644.
- Selvaretnam, A. A., Sahu, P. S., Sahu, M., & Ambu, S. (2016). A review of concurrent infections of malaria and dengue in Asia. *Asian Pacific Journal of Tropical Biomedicine*, 6(7), 633-638.

- Shirinivas, S. G., Vetrivel, S., & Elango, N. M. (2011, July). Applications of graph theory in computer science. In *Computational Intelligence, Communication Systems and Networks (CICSyN), 2011 Third International Conference on* (pp. 142-145). IEEE.
- Sincovec, R. F., & Madsen, N. K. (1975). Software for nonlinear partial differential equations. *ACM Transactions on Mathematical Software (TOMS)*, 1(3), 232-260.
- Suaya, J. A., Shepard, D. S., Siqueira, J. B., Martelli, C. T., Lum, L. C., Tan, L. H., Kongsin, S., Jiamton, S., Garrido, F., Montoya, R., & Armien, B. (2009). Cost of dengue cases in eight countries in the Americas and Asia: a prospective study. *The American Journal of Tropical Medicine and Hygiene*, 80(5), 846-855.
- Sutherst, R. W. (2004). Global change and human vulnerability to vector-borne diseases. *Clinical Microbiology Reviews*, 17(1), 136-173.
- Syafruddin, S., & Noorani, M. S. M. (2012). SEIR model for transmission of dengue fever in Selangor Malaysia. In *International Journal of Modern Physics: Conference Series* (Vol. 9, pp. 380-389). World Scientific Publishing Company.
- Sylvestre, G., Gandini, M., & Maciel-de-Freitas, R. (2013). Age-dependent effects of oral infection with dengue virus on *Aedes aegypti* (Diptera: Culicidae) feeding behavior, survival, oviposition success and fecundity. *PloS One*, 8(3), e59933.
- Tao, H.Y., Pan, Z. Z., Pan, M. L., Zhuo, L. Xu, Y. & Lu, M. (2016). Mixing spatial-temporal transmission patterns of metropolis dengue fever- A case study of Guangzhou, China. *Acta Geographica Sinica*, 71(9), 1653–1662.
- Tetko, I. V., & Tanchuk, V. Y. (2002). Application of associative neural networks for prediction of lipophilicity in ALOGPS 2.1 program. *Journal of Chemical Information and Computer Sciences*, 42(5), 1136-1145.

- Tran, A., L'Ambert, G., Lacour, G., Benoît, R., Demarchi, M., Cros, M., Cailly, P., Aubry-Kientz, M., Balenghien, T., & Ezanno, P. (2013). A rainfall-and temperature-driven abundance model for *Aedes albopictus* populations. *International Journal of Environmental Research and Public Health*, 10(5), 1698-1719.
- Towles, B., & Dally, W. J. (2002, August). Worst-case traffic for oblivious routing functions. In *Proceedings of the Fourteenth Annual ACM Symposium on Parallel Algorithms and Architectures* (pp. 1-8). ACM.
- Tun-Lin, W., Burkot, T. R., & Kay, B. H. (2000). Effects of temperature and larval diet on development rates and survival of the dengue vector *Aedes Aegypti* in north Queensland, Australia. *Medical and Veterinary Entomology*, 14(1), 31-37.
- Viennet, E., Ritchie, S. A., Williams, C. R., Faddy, H. M., & Harley, D. (2016). Public health responses to and challenges for the control of dengue transmission in high-income countries: four case studies. *PLoS Neglected Tropical Diseases*, 10(9), e0004943.
- Wahid, J., & Hussin, N. M. (2016, August). Combination of graph heuristics in producing initial solution of curriculum based course timetabling problem. *AIP Conference Proceedings*, 1761(1), 020104.
- Wang, S. M., & Sekaran, S. D. (2010). Early diagnosis of Dengue infection using a commercial Dengue Duo rapid test kit for the detection of NS1, IGM, and IGG. *The American Journal of Tropical Medicine and Hygiene*, 83(3), 690-695.
- Wang, X., Tang, S., & Cheke, R. A. (2016). A stage structured mosquito model incorporating effects of precipitation and daily temperature fluctuations. *Journal of Theoretical Biology*, volume 411, 27-36.

- Wang, Y. L., Zhou, T., Shi, J. J., Wang, J., & He, D. R. (2009). Empirical analysis of dependence between stations in Chinese railway network. *Physica A: Statistical Mechanics and its Applications*, 388(14), 2949-2955.
- Wesolowski, A., Qureshi, T., Boni, M. F., Sundsøy, P. R., Johansson, M. A., Rashed, S. B., & Buckee, C. O. (2015). Impact of human mobility on the emergence of dengue epidemics in Pakistan. *Proceedings of the National Academy of Sciences*, 112(38), 11887-11892.
- WHO. (2012). *Rapid Risk Assessment of Acute Public Health Events*. WHO, Geneva, Switzerland.
- Wijayanti, S. P., Porphyre, T., Chase-Topping, M., Rainey, S. M., McFarlane, M., Schnettler, E., & Kohl, A. (2016). The Importance of Socio-Economic Versus Environmental Risk Factors for Reported Dengue Cases in Java, Indonesia. *PLoS Neglected Tropical Diseases*, 10(9), 1-15.
- Wilensky, U. (1999). *NetLogo*. Retrieved from <http://ccl.northwestern.edu/netlogo/>.
- Willem, L., Stijven, S., Vladislavleva, E., Broeckhove, J., Beutels, P., & Hens, N. (2014). Active learning to understand infectious disease models and improve policy making. *PLoS Computational Biology*, 10(4), 3563-3574.
- Wisn, M. S., Hijmans, R. J., Li, J., Peterson, A. T., Graham, C. H., & Guisan, A. (2008). Effects of sample size on the performance of species distribution models. *Diversity and Distributions*, 14(5), 763-773.
- Wiwatitkit, V. (2011). Concurrent malaria and dengue infection: a brief summary and comment. *Asian Pacific Journal of Tropical Biomedicine*, 1(4), 326-327.
- World Health Organisation. (2017). *Dengue and severe dengue* (Fact sheet). Retrieved from <http://www.who.int/mediacentre/factsheets/fs117/en/>

- Wu, P. C., Lay, J. G., Guo, H. R., Lin, C. Y., Lung, S. C., & Su, H. J. (2009). Higher temperature and urbanization affect the spatial patterns of dengue fever transmission in subtropical Taiwan. *Science of the Total Environment*, 407(7), 2224-2233.
- Wu, Z., Braunstein, L. A., Havlin, S., & Stanley, H. E. (2006). Transport in weighted networks: partition into superhighways and roads. *Physical Review Letters*, 96(14), 148702.
- Xiao, F. Z., Zhang, Y., Deng, Y. Q., He, S., Xie, H. G., Zhou, X. N., & Yan, Y. S. (2014). The effect of temperature on the extrinsic incubation period and infection rate of dengue virus serotype 2 infection in *Aedes albopictus*. *Archives of Virology*, 159(11), 3053-3057.
- Yong, L. S., & Koh, K. C. (2013). A case of mixed infections in a patient presenting with acute febrile illness in the tropics. *Case reports in Infectious Diseases*, 2013, 1-3.
- Yu, H., Yang, J., & Han, J. (2003, August). Classifying large data sets using SVMs with hierarchical clusters. In *Proceedings of the ninth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (pp. 306-315). ACM.
- Zhang, Z., Geng, J., Tang, X., Fan, H., Xu, J., Wen, X., Zhanshan, M., & Shi, P. (2014). Spatial heterogeneity and co-occurrence patterns of human mucosal-associated intestinal microbiota. *The ISME Journal*, 8(4), 881.
- Zhou, T., Ren, J., Medo, M., & Zhang, Y. C. (2007). Bipartite network projection and personal recommendation. *Physical Review E*, 76(4), 046115.

APPENDICES

Appendix A

Table A.1: Epi Week Calendar From July until October in Year 2015 (Arias, 2016)

(a) July

July							
Epi Week #	Sun	Mon	Tue	Wed	Thu	Fri	Sat
26				1	2	3	4
27	5	6	7	8	9	10	11
28	12	13	14	15	16	17	18
29	19	20	21	22	23	24	25
30	26	27	28	29	30	31	

(b) August

August							
Epi Week #	Sun	Mon	Tue	Wed	Thu	Fri	Sat
30							1
31	2	3	4	5	6	7	8
32	9	10	11	12	13	14	15
33	16	17	18	19	20	21	22
34	23	24	25	26	27	28	29
35	30	31					

(c) September

September							
Epi Week #	Sun	Mon	Tue	Wed	Thu	Fri	Sat
35			1	2	3	4	5
36	6	7	8	9	10	11	12
37	13	14	15	16	17	18	19
38	20	21	22	23	24	25	26
39	27	28	29	30			

(d) October

October							
Epi Week #	Sun	Mon	Tue	Wed	Thu	Fri	Sat
39					1	2	3
40	4	5	6	7	8	9	10
41	11	12	13	14	15	16	17
42	18	19	20	21	22	23	24
43	25	26	27	28	29	30	31

Table A.2: Calculation of SRCC for Location Nodes (Analytical Verification)

L Node	DHR	Hub Mat	RankDHR	RankHub	$d_{DHR_HubMatrix}$	$d_{DHR_HubMatrix}]^2$
L1	0	66.88	15	15	0	0
L2	3.1E-116	104.08	13	10	3	9
L3	0.001617	130.24	5	2	3	9
L4	1	128.52	1	3	-2	4
L5	0	36.14	16	19	-3	9
L6	4.64E-07	102.97	9	11	-2	4
L7	7.49E-07	167.36	8	1	7	49
L8	0	68.65	17	14	3	9
L9	0	65.74	18	16	2	4
L10	3.3E-116	109.04	12	9	3	9
L11	0.00144	115.94	6	6	0	0
L12	0.947297	121.74	2	4	-2	4
L13	2.89E-07	65.1	11	18	-7	49
L14	0	65.62	19	17	2	4
L15	2.9E-116	96.54	14	12	2	4
L16	0.001427	114.89	7	7	0	0
L17	0.937019	120.41	3	5	-2	4
L18	4.19E-07	92.88	10	13	-3	9
L19	0.881315	113.26	4	8	-4	16
Sum of $[d_{TrainTest}]^2$:						196

Table A.3: Parameter Values of Location Nodes in Validated Model in Group 1(a) Parameter L_c , S , B , Al , H and Pre

Location Node	L_c	S	B	Al (m)	K (%)	Pre (mm)
L2	0.9068	0.7286	1	0.8269	0.9581	0.2621
L4	0.9918	0.7286	1	0.3654	0.9892	0.5971
L5	0.9682	0.7286	1	0.1346	0.9892	0.5631
L20	0.9918	0.7286	1	0.8269	0.9892	0.5631
L21	0.9918	0.7286	1	0.25	0.9892	0.5631
L22	0.9068	0.7129	1	0.1538	0.9437	0.4902
L23	0.9682	0.8037	1	0.07692	0.9581	0.2621
L24	0.9405	0.7474	1	0.1923	0.9473	0.2621
L25	0.9405	0.7474	1	0.2115	0.9473	0.2621
L26	0.9682	0.7129	1	0.1346	0.9581	0.4903
L27	0.9682	0.7286	1	0.1346	0.9473	0.5631

(b) Parameter Du

L node	H Node	Du (s)
L2	H1	0.001290
L4	H3	0.9752
L5	H1	0.09752
L20	H1	0.4286
L21	H2	0.7143
L22	H3	0.4286
L23	H1	0.2857
L24	H3	0.002976
L25	H1	0.008929
L26	H3	0.005952
L27	H3	0.001290

Table A.4: DCS Matrix of the Targeted and Validated Model in Group 2

(a) Targeted Model (Column 1-6)

	H1	H2	H3	H4	H5	H6
L19	0	0	4.4888	0	0	0
L23	0	5.3718	0	0	0	0
L47	0	4.5275	0	0	0	0
L52	4.7133	0	0	0	0	0
L53	0	4.1133	0	0	0	0
L54	0	0	4.7548	0	0	0
L55	0	0	0	4.8064	0	0
L62	5.2323	0	0	0	0	0
L63	0	0	0	5.7359	0	0
L64	0	0	0	0	5.8979	0
L65	0	0	0	0	0	5.1045
L71	3.8893	0	0	0	0	0
L72	0	0	4.5817	0		0
L73	0	0	0	4.7583	0	0
L77	4.3789	0	0	0	0	0
L78	0	0	0	5.0940	0	0

(b) Targeted Model (Column 7-12)

	H7	H8	H9	H10	H11	H12
L33	0	4.1390	0	4.1731	0	0
L56	4.6411	0	0	0	0	0
L57	0	4.4594	0	0	0	0
L58	0	0	5.4041	0	0	0
L59	0	0	0	4.6307	0	0
L60	0	0	0	0	4.9134	0
L61	0	0	0	0	0	5.0886
L62	0	0	0	4.8668	0	0
L66	5.1674	0	0	0	0	0
L67	0	5.0026	0	0	0	0
L68	0	0	4.6137	0	0	0
L69	0	0	0	0	5.2174	0
L70	0	0	0	0	0	4.8682

L74	4.6482	0	0	0	0	0
L75	0	0	4.0197	0	0	0
L76	0	0	0	0	4.1852	0

(c) Validated Model

	H1	H2
L53	5.6421	0
L64	0	4.0252
L69	0	5.0221
L77	0	5.7170
L79	4.5602	0
L80	0	5.8431
L81	0	4.9992

Table A.5: Location Node Ranking of the Targeted and Validated Model in Group 2

(a) Ranking result of Targeted Model

Rank	Location No	DHR_Train
1	L63	1
2	L78	0.888077
3	L55	0.837945
4	L73	0.829556
5	L62	0.000528
6	L52	0.000279
7	L33	0.000266
8	L77	0.000259
9	L71	0.00023
10	L59	0.000208
11	L67	9.43E-05
12	L57	8.41E-05
13	L66	2.5E-110
14	L74	2.2E-110
15	L56	2.2E-110

Rank	L Node	DHR_Train
16	L69	4E-114
17	L60	3.8E-114
18	L76	3.2E-114
19	L58	4.9E-123
20	L68	4.2E-123
21	L75	3.6E-123
22	L23	1.4E-124
23	L47	1.2E-124
24	L53	1.1E-124
25	L54	3.7E-135
26	L72	3.6E-135
27	L19	3.5E-135
28	L61	2.6E-203
29	L70	2.5E-203
30	L64	1.4E-299

(b) Ranking result of Validated Model

Rank	Location No	DHR _{Test}
1	L80	1
2	L77	0.978416
3	L69	0.859498
4	L81	0.855578
5	L64	0.688878
6	L53	2.19E-06
7	L79	1.77E-06

Table A.6: Location Node Ranking of the Targeted and Validated Model in Group 3

(a) Ranking result of Targeted Model

Rank	Location Node	Rank	L Node
1	L80	13	L96
2	L83	14	L94
3	L97	15	L90
4	L43	16	L85
5	L92	17	L88
6	L89	18	L86
7	L84	19	L97
8	L5	20	L11
9	L77	21	L93
10	L82	22	L64
11	L98	23	L91
12	L95		

(b) Ranking result of Validated Model

Rank	Location No
1	L89
2	L97
3	L92
4	L95
5	L99
6	L98
7	L100
8	L96
9	L20

10	L97
11	L41
12	L93
13	L94
14	L90
15	L14

Table A.7: Location Node Ranking of the Targeted and Validated Model in Group 3

(a) Parameter L_c , S , B , Al , H and Pre

Location Node	L_c	S	B	Al	H	Pre
L1	0.9762	0.8044	1	0.8032	0.9685	0.8211
L2	0.9762	0.8044	1	1	0.9685	0.8211
L3	0.9762	0.8044	1	0.4325	0.9685	0.8211
L4	1	0.8221	1	0.4943	1	0.9431
L5	0.9762	0.8044	1	1	0.9685	0.8211
L6	0.9762	0.8044	1	0.2471	0.9685	0.8211
L7	0.9762	0.8044	1	0.5561	0.9685	0.8211
L8	0.9762	0.8044	1	0.3707	0.9685	0.8211
L9	0.9762	0.8044	1	0.8032	0.9685	0.8211
L10	0.9762	0.8044	1	1	0.9685	0.8211
L11	0.9762	0.8044	1	0.5561	0.9685	0.8211
L12	0.9762	0.8044	1	0.2471	0.9685	0.8211
L13	0.9762	0.8044	1	1	0.9685	0.8211
L14	0.9762	0.8044	1	0.4943	0.9685	0.8211
L15	0.9762	0.8044	1	1	0.9685	0.8211
L16	0.9762	0.8044	1	0.4943	0.9685	0.8211
L17	0.9762	0.8044	1	0.6178	0.9685	0.8211
L18	0.9762	0.8044	1	1	0.9685	0.8211
L19	0.9762	0.8044	1	0.5561	0.9685	0.8211
L20	1	0.8221	1	1	1	0.9431
L21	1	0.8221	1	0.8032	1	0.9431
L21	1	0.8221	1	0.8032	1	1
L21	0.9143	0.9068	1	0.8032	0.9540	0.4390
L22	0.8764	1	1	0.8032	0.9031	0.0163
L23	0.9482	0.8433	1	0.6178	0.9576	0.4390
L24	0.9482	0.8433	1	0.6796	0.9576	0.4390
L25	0.8764	1	1	0.8032	0.9031	0.0163
L26	0.9762	0.8044	1	0.4943	0.9685	0.8211
L27	0.9762	0.8044	1	1	0.9685	0.8293

(b) Parameter Fl , Du , DCS , DCS_Pre and DCS_T

Location Node	Fl	Du	DCS	DCS_Pre	DCS_T
L1	1	0.00129	5.3748	5.8076	11.1824
L2	1	0.00129	5.6839	5.914	11.5979
L3	1	0.00129	5.0041	6.8536	11.8577
L4	1	0.00129	6.6473	5.8419	12.4892
L5	1	0.005952	5.4813	6.0119	11.4932
L6	1	0.00129	4.8187	6.0518	10.8705
L7	1	0.00129	5.1276	6.2569	11.3845
L8	1	0.00129	4.9423	5.9245	10.8668
L9	1	0.00129	5.3748	5.7087	11.0835
L10	1	0.00129	5.5716	6.1964	11.7680
L11	1	0.00129	5.1276	6.1014	11.2290
L12	1	0.00129	4.8187	5.534	10.3527
L13	1	0.00129	5.5716	5.5251	11.0967
L14	1	0.00129	5.0659	5.6635	10.7294
L15	1	0.00129	5.5716	5.4862	11.0578
L16	1	0.00129	5.0659	6.0466	11.1125
L17	1	0.00129	5.1894	5.474	10.6634
L18	1	0.00129	5.5716	5.4588	11.0304
L19	1	0.00129	5.1276	5.1485	10.2761
L20	1	0.4286	6.5283	0	6.5283
L21	1	0.9494	6.5178	0	6.5178
L21	1	0.09752	5.7229	0	5.7229
L21	1	0.2857	5.3031	0	5.3031
L22	1	0.4167	5.6362	0	5.6362
L23	1	0.09752	5.3547	0	5.3547
L24	1	0.09752	5.0926	0	5.0926
L25	1	0.4167	5.1177	0	5.1177
L26	1	0.09752	5.2701	0	5.2700
L27	1	0.008929	5.3432	0	5.3432

Table A.8: SRCC Calculation of Location Nodes in BDC-4 Network

Location Node	Rank DHR _{Target}	Rank DHR _{Val}	$d_{TargetVal}$	$[d_{TargetVal}]^2$
L7	6	5	1	1
L8	2	7	-5	25
L23	4	2	2	4
L26	7	8	-1	1
L32	5	1	4	16
L41	3	4	-1	1
L43	8	3	5	25
L77	1	6	-5	25
Sum of $[d_{TargetVal}]^2$:				98

Table A.9: SRCC Calculation of Location Nodes in BDC-6 Network

Location Node	Rank DHR _{Target}	Rank DHR _{Val}	$d_{TargetVal}$	$[d_{TargetVal}]^2$
L10	5	3	2	4
L23	1	5	-4	16
L29	4	1	3	9
L62	2	4	-2	4
L68	3	2	1	1
Sum of $[d_{TargetVal}]^2$:				34

Appendix B

NO. KES eDENGUE :		PBV (DD/DB/IE/CHIKU) 301 Pindaan 4/13					
NOTA PEMBERITAHUAN KES DF / DHF / DSS / SEVERE DENGUE							
DAERAH :	Kota Samarahan	Bil. Minggu Epid :	30				
		1) Kedudukan Kes <table border="1" style="margin: auto;"> <tr> <td>Kes Sporadik</td> <td>WB</td> <td>WT</td> <td>WTK</td> </tr> </table> (Pilih yang Berkenaan)		Kes Sporadik	WB	WT	WTK
Kes Sporadik	WB	WT	WTK				
2) Jika kes berada dalam Lokaliti Wabak , nyatakan nama lokaliti wabak		: TB					
3) Nyatakan juga Kes Pertama dalam lokaliti wabak yang tersebut di atas		: TB					
4) Nama Pesakit : Ally Lim							
5) Bangsa : Cina		6) Warganegara / ~~Bukan warganegara~~ : MALAYSIA					
7) Umur : 33 tahun		8) Jantina : Perempuan					
9) No. K/P :		10) Pekerjaan : Pensyarah					
11a) Jenis K/P : / Sendiri		11b) Nama Pengiring :					
12a) Alamat Tempat Tinggal Semasa : One Residency Courtyard, Batu Kawa New Township, 93250		12b) Bandar Luar Bandar					
Koordinat GPS Latitude: ---		Longitude: ---					
12c) No. Telefon :		12d) No. H/P : 010-1234567					
13a) Alamat Lain / Tempat Kerja / Tadika / Sekolah / Kolej / Universiti : Kolej IKMAS		13b) Bandar Luar Bandar					
Koordinat GPS Latitude: ---		Longitude: ---					
13c) No. Telefon : ---							
14) Nama Klinik / Hospital yang melaporkan notifikasi ini :		Sarawak General Hospital					
15) Nama wad : ---		16) No. Pendaftaran :					
17) Diagnosa Dalam Notifikasi Kes :		DF / DHF / DSS / SEVERE DENGUE					

(a)

MAKlumat Klinik Primer	18) Tarikh Onset	:	29/7/2016	19) Tarikh Masuk	:	1/8/2016
	20) Tarikh Diagnosa	:	1/8/2016	21) Tarikh Notifikasi	:	1/8/2016
	22) Tarikh Ujian Serologi	:	1/8/2016	23) Keputusan serologi	:	NS1: VE IGM: VE IGG: VE
	24) Tanda-tanda Klinikal	(☒)	pada yang berkenaan			
	☒ Fever		☐ Headache		☐ Retroorbital Pain	
	☒ Joint Pain		☐ Myalgia/Muscle Ache		☐ Backache	
	☐ Vomiting		☐ Gum Bleeding		☐ Rash	
	☐ Nose Bleeding		☐ Ecchymosis		☐ Petechiae	
	☐ Altered Consciousness		☐ Purpura		☐ Leukopenia	
	☐ Haemetesis		☐ Nauseated		☐ Malaena	
		Hess's Test :				
25) Warning Signs	(☒)	pada yang berkenaan				
☐ Mucosal Bleed		☐ Abdominal Pain/Tenderness		PCV/Hematocrit	:	
☐ Enlarged Liver (>2cm)		☐ Clinical Fluid Accumulation		WBC	:	
☐ Persistent Vomiting		☐ Lethargy/Restless		Platelet Count	:	
26) Tanda-tanda Klinikal Lain (☒ pada yang berkenaan)				H	:	
☐ Epidemiological Link				B	:	
27) Adakah kes ini memenuhi kriteria definisi Kes Denggi?		Ya / Tidak				
		Jika TIDAK, nyatakan sebab kes ini dinotifikasikan sebagai kes denggi :				
28) Nama Pegawai yang mendiagnos kes ini	:	Encik Abu bin Ali				
28) Nama Pegawai yang melaporkan notifikasi ini	:	Cik Malanie Anak Juli				
29a) Nama Klinik Primer yang dilawati oleh pesakit sebelum notifikasi ini (jika ada)	:					
		(definisi Klinik Primer - OPD, KK, Poliklinik, Klinik Swasta, A&E Hospital, tempat pesakit menerima rawatan OPD sahaja)				
29b) Tarikh melawat Klinik Primer oleh pesakit	:					
29c) Diagnosa oleh Klinik Primer yang dilawati oleh pesakit	:	DF/DHS / DSS / Severe Dengue / TRO Dengue / Lain				

(b)

RUJUKAN	29d) Tanda klinikal semasa pesakit melawat Klinik Primer	:	
	29e) Jika diagnosa ialah Denggi, adakah Rekod Pemantauan Pesakit Denggi diberi?	:	
	29f) Adakah notifikasi dibuat oleh Klinik Primer yang dilawati oleh pesakit?	:	
	30a) Adakah kes ini merupakan kes yang dirujuk oleh mana-mana pihak?	:	
	30b) Jika dirujuk, namakan institusi yang merujuk kes ini	:	TB
	30c) Tarikh kes dirujuk	:	TB
	30d) Diagnosa semasa dirujuk	:	TB
	31) Pergerakan Pesakit - Sila lampirkan Carta Pergerakan Pesakit secara berasingan.		
	32a) Nama Pegawai yang menerima notifikasi	:	
	32b) Tarikh Menerima	:	
33a) Nama Pegawai yang menyiasat	:		
33b) Tarikh Menyiasat	:		
34a) Nama pegawai yang melapor	:		
34b) Tarikh Melapor	:		

(c)

CARTA PERGERAKAN PESAKIT (Untuk dikepilkan bersama dengan Nota Pemberitahuan Kes DF / DHF / DSS / SEVERE DF)				
Nama Pesakit: Ally Lim				
TEMPAT	HARI	TARIKH	AKTIVITI DIJALANKAN	TEMPOH HARI DARI ONSET
Tadika Eduland, IKMAS, Rumah	Jumaat	15/7	Betolak dari rumah hantar anak ke Tadika Eduland pada 7:30 pagi. Pergi IKMAS dari jam 9 hingga 12.30 tengahari dan dari 7:30 malam hingga 9:30 malam, makan tengahari di rumah	-14
Emart Batu Kawa, rumah	Sabtu	16/7	Pergi Emart Batu Kawa jam 4pm hingga 6pm, masa lain berada dalam rumah	-13
Samariang	Ahad	17/7	Pergi Samariang dari jam 5 hingga 6:30pm	-12
Tadika Eduland, IKMAS, Rumah	Isnin	18/7	Betolak dari rumah hantar anak ke Tadika Eduland, Jalan Green 7:30 pagi. Pergi IKMAS dari jam 9 hingga 12.30 tengahari dan dari 7:30 malam hingga 9:30 malam, makan tengahari di rumah	-11
Tadika Eduland, IKMAS, Rumah, Plaza Merdeka	Selasa	19/7	Betolak dari rumah hantar anak ke Tadika Eduland, Jalan Green 7:30 pagi. Pergi IKMAS dari jam 9 hingga 12.30 tengahari dan dari 7:30 malam hingga 9:30 malam, makan tengahari di Plaza Merdeka dari 2pm hingga 4pm.	-10
Tadika Eduland, IKMAS, Rumah	Rabu	20/7	Betolak dari rumah hantar anak ke Tadika Eduland, Jalan Green 7:30 pagi. Pergi IKMAS dari jam 9 hingga 12.30 tengahari dan dari 7:30 malam hingga 9:30 malam, makan tengahari di rumah	-9
Tadika Eduland, IKMAS, Rumah	Khamis	21/7	Betolak dari rumah hantar anak ke Tadika Eduland, Jalan Green 7:30 pagi. Pergi IKMAS dari jam 9 hingga 12.30 tengahari dan dari 7:30 malam hingga 9:30 malam, makan tengahari di rumah	-8
Tadika Eduland, IKMAS, Rumah	Jumaat	22/7	Betolak dari rumah hantar anak ke Tadika Eduland, Jalan Green 7:30 pagi. Pergi IKMAS dari jam 9 hingga 12.30 tengahari dan dari 7:30 malam hingga 9:30 malam, makan tengahari di rumah	-7
Rumah	Sabtu	23/7	Berada dalam rumah	-6
Pasar Batu 3, rumah	Ahad	24/7	Pergi pasar batu 3 masa 8 pagi hingga 10pagi. Masa lain berada dalam rumah sahaja.	-5
Tadika Eduland, IKMAS, Rumah	Isnin	25/7	Betolak dari rumah hantar anak ke Tadika Eduland, Jalan Green 7:30 pagi. Pergi IKMAS dari jam 9 hingga 12.30 tengahari dan dari 7:30 malam hingga 9:30 malam, makan tengahari di rumah	-4

(d)

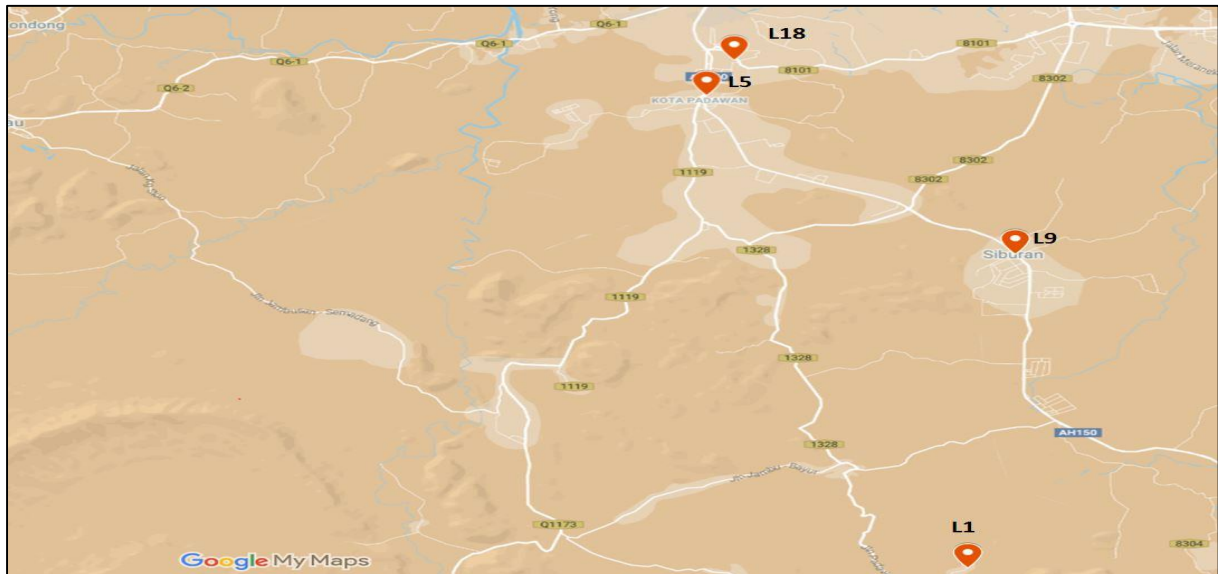
Tadika Eduland, IKMAS, Rumah	Selasa	26/7	Betolak dari rumah hantar anak ke Tadika Eduland, Jalan Green 7:30 pagi. Pergi IKMAS dari jam 9 hingga 12.30 tengahari dan dari 7:30 malam hingga 9:30 malam, makan tengahari di rumah	-3
Tadika Eduland, IKMAS, Rumah	Rabu	27/7	Betolak dari rumah hantar anak ke Tadika Eduland, Jalan Green 7:30 pagi. Pergi IKMAS dari jam 9 hingga 12.30 tengahari dan dari 7:30 malam hingga 9:30 malam, makan tengahari di rumah	-2
Tadika Eduland, IKMAS, Rumah	Khamis	28/7	Betolak dari rumah hantar anak ke Tadika Eduland, Jalan Green 7:30 pagi. Pergi IKMAS dari jam 9 hingga 12.30 tengahari dan dari 7:30 malam hingga 9:30 malam, makan tengahari di rumah	-1
Rumah	Jumaat	29/7	Ada demam dan pening kepala, ambil cuti dan berada di rumah sahaja	ONSET
			Jumpa doctor dalam hospital	1
				2
				3
				4
				Lawatan ke Klinik Primer(jika ada)
				Diagnosas
				Notifikasi (jika ada)
				Kemasukan Hospital (Jika ada)
				Diagnosa
				Notifikasi (Jika ada)
				Dirujuk (Jika Ada)
				Siasatan

(e)

Figure B.1: Investigation Form (a) - (e)



(a)



(b)

Figure B.2: Actual Location of the First BDC Network Model

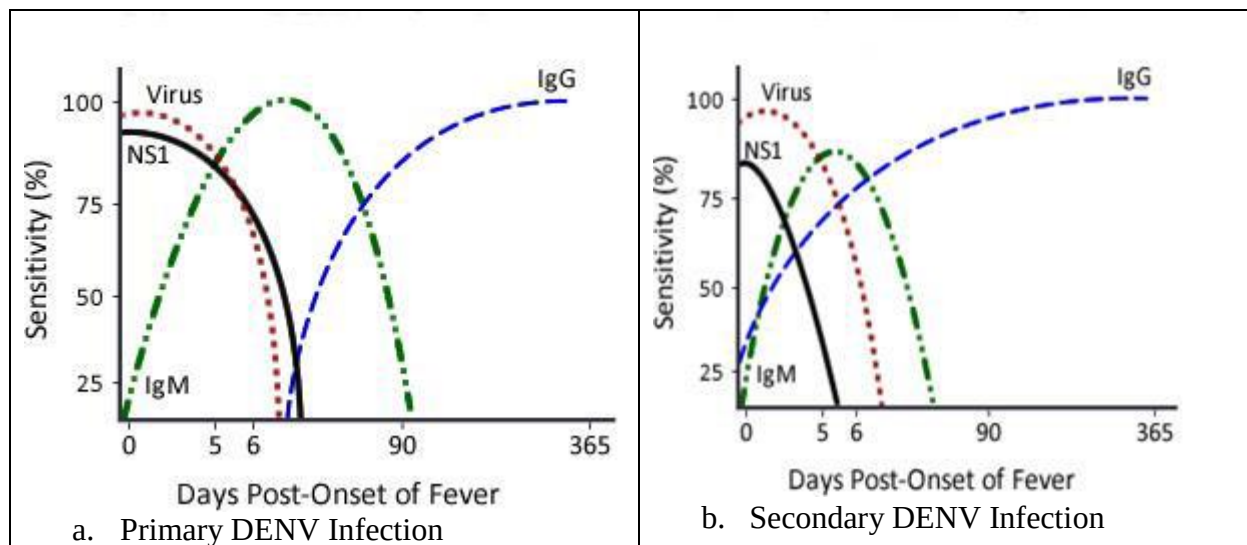


Figure B.3: Sensitivity of dengue tests when dengue occurs (CDC, 2017)