Faculty of Computer Science and Information Technology

VISUALISATION TOOL TO STUDY MALARIA TRANSMISSION USING NETWORK MODELLING

Wong Ji Yeh

Bachelor of Computer Science with Honours
(Computational Science)
2014

VISUALISATION TOOL TO STUDY MALARIA TRANSMISSION USING NETWORK MODELLING

WONG JI YEH

This project is submitted in partial fulfilment of the requirements for the Degree of Bachelor of Computer Science

Faculty of Computer Science and Information Technology
UNIVERSITI MALAYSIA SARAWAK
2014

UNIVERSITI MALAYSIA SARAWAK

THESIS STATUS ENDORSEMENT FORM

TITLE VISUALISATION TOOL TO STUDY MALARIA TRANSMISSION USING NETWORK MODELING

ACADEMIC SESSION: 2013/2014

(CAPITAL LETTERS)

hereby agree that this Thesis* shall be kept at the Centre for Academic Information Services, Universiti Malaysia Sarawak, subject to the following terms and conditions:

1. The Thesis is solely owned by Universiti Malaysia Sarawak
2. The Centre for Academic Information Services is given full rights to produce copies for educational purposes only
3. The Centre for Academic Information Services is given full rights to do digitization in order to develop local content database
4. The Centre for Academic Information Services is given full rights to produce copies of this Thesis as part of its exchange item program between Higher Learning Institutions [or for the purpose of interlibrary loan between HLI]
5. ** Please tick (√)

☐	CONFIDENTIAL	(Contains classified information bounded by the OFFICIAL SECRETS ACT 1972)
☐	RESTRICTED	(Contains restricted information as dictated by the body or organization where the research was conducted)
☑	UNRESTRICTED	

Validated by

(AUTHOR'S SIGNATURE)

(SUPERVISOR'S SIGNATURE)

Terrin Lim
Lecturer
Faculty of Computer Science and Information Technology
UNIVERSITI MALAYSIA SARAWAK

Permanent Address

No. 6, Jln Nagasari 18,
Segamat Baru, 85000 Segamat,
Johor.

Date: 18/6/2014

Date: 18 June 2014

Note * Thesis refers to PhD, Master, and Bachelor Degree
** For Confidential or Restricted materials, please attach relevant documents from relevant organizations / authorities

ACKNOWLEDGEMENT

First and foremost, I praise God, the Almighty, for His showers of blessings throughout the process to complete my final year project successfully.

I would like to express my deep gratitude to my project supervisor, Mr. Terrin Lim, for giving me the opportunity to do this project and providing invaluable guidance, motivation and inspiration. It was a great pleasure to do this project under his supervision. I would like to thank him for his friendship, empathy and sense of humour.

I would like to extend my gratitude to Assoc. Prof. Dr. Jane Labadin who has given me priceless advises and courage in this project. To my friends that always there to help and encourage whenever I was in need.

I am grateful to my parents for their love, prayers, care and sacrifices for educating and preparing me for my future.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF EQUATIONS

LIST OF NOMENCLATURES

API :	Application Programmer Interface
IDE :	Integrated Development Environment
mean(T) :	Mean of value T
std(T) :	Standard deviation of value T
INT(y) :	Integer part of any real number y
CSV :	Comma-Separated Values file
PDF :	Portable Document Format
Set :	Java data structure which does not allow duplicate elements
$Matrix^T$:	T - Transpose operator for matrix
λ :	Eigenvalue
v :	Eigenvector
epsilon :	Stopping criteria in a loop
maxIteration :	Maximum number of iteration
PNG :	Portable Network Graphics, an image file format

ABSTRACT

Malaria has been described as one of the major killer diseases. There were about 219 million cases of malaria in 2010 and an estimated 660,000 deaths. This deserves urgent scientific investigations and studies on malaria transmission. Since malaria is a vector borne disease, this project aims to produce a visualisation tool to model the disease transmission through formulating a heterogeneous bipartite contact network of two node types (public places and human beings). In addition, the Hypertext Induced Topical Search (HITS) web search algorithm was adapted to implement a search engine, which uses the bipartite contact network as the input. Java was used to implement the visualisation tool. The output from this visualisation tool shows predicted hotspots which harbour the infected malaria vectors. This output was validated with UCINET 6.0 as the benchmark system. A similar dominant node ranking output was obtained when the output from the benchmark system is compared with that of the modelling tool. The resulting information is believed helpful to tackle the issue of malaria transmission from the perspective of vectors detection.

ABSTRAK

Malaria telah digambarkan sebagai salah satu penyakit pembunuh yang utama. Terdapat kira-kira 219 juta kes-kes malaria pada tahun 2010 dengan anggaran 660,000 kematian. Hal ini memerlukan penyiasatan saintifik segera dan kajian mengenai transmisi malaria. Oleh kerana malaria merupakan penyakit bawaan vektor, projek ini bertujuan untuk menghasilkan alat visualisasi untuk memodelkan transmisi penyakit melalui merumuskan jaringan hubungan heterogen dwibahagian daripada dua jenis nod (tempat-tempat awam dan manusia). Di samping itu, algoritma carian web *Hypertext Induced Topical Search (HITS)* telah diguna pakai untuk dilaksanakan dalam enjin carian yang menggunakan jaringan hubungan dwibahagian sebagai input. *Java* digunakan untuk melaksanakan alat visualisasi. Output dari alat visualisasi ini menunjukkan *hotspot* yang diramal meliputi vektor malaria yang dijangkiti. Output itu telah disahkan dengan menggunakan UCINET 6.0 sebagai system penanda aras. Output kedudukan nod dominan yang serupa telah diperolehi apabila output system penanda aras dibandingkan dengan output alat peragaan itu. Maklumat terhasil dipercayai berguna untuk menangani masalah transmisi malaria dari segi pengesanan vektor.

CHAPTER ONE

INTRODUCTION

1.0 Introduction

Malaria is one of the major killers and dangerous diseases. It is a mosquito-borne infectious disease of humans that caused by a parasite called *Plasmodium*. This disease is transmitted via bites from an infected female *Anopheles* mosquito. Every year, there are 350 to 500 million people infected with this disease. In 2010 malaria caused an estimated 655,000 deaths worldwide (WHO, 2013). Thus, it deserves attentions and further scientific investigations on malaria transmission.

The purpose of this project is to develop a Java-based visualisation tool to represent and communicate malaria transmission information clearly and effectively through graphical means. This tool aims to allow users to see, explore and understand large amount of malaria transmission information more easily by providing visual representations and interactive interfaces. The user or analyst does not have to learn the sophisticated network modelling method to study the malaria transmission. Thus, the focus of this project will be on building a visualisation tool that models humans and their visits to public places which might expose them to malaria disease.

1.1 Problem Statement

Malaria transmission in public places is a problem in terms of vector detection that needs scientific investigation. A thesis by Eze (2012) gave a practical scenario observed in 2010 when a team of vector control experts visited UNIMAS to carry out fogging work against malaria and dengue vectors, the team mentioned that they lacked of vector detection tools which resulted in the team possibly does fogging in the wrong places.

There are quite a number of disease modelling and vector detection methods in existence that have some associated disadvantages. One of the examples is the disease modelling methods that assume homogeneity in population where it refers to the assumption that every individual in the disease transmission environment has same probability of mixing with one another and hence get infected. The results based on this assumption may lead to an inaccurate and unrealistic modelling of disease. An improvement over this fault strategy is to employ the proposed network modelling method that takes into account the variation of contacts which leads to disease transmission.

Humans move around in public places that might expose them to risks of being bitten by mosquitoes carrying the malaria disease. Up until now, no system has been developed to study and predict the exposure risks of humans in public places such as markets, schools, food courts and fields. Thus, the purpose of this malaria disease transmission study is to predict transmission of this disease in public places and develop a visualisation tool for such transmission model.

1.2 Objectives

The objectives of project can be listed as follows:

- To apply and implement a network-based model of malaria transmission using HITS web search algorithm.
- To develop a Java-based visualisation tool for such transmission model.
- To predict public places that are hotspots for malaria infections using the visualisation tool.

1.3 Methodology

The purpose of this project is to study and model malaria transmission using contact network model. The model produced is then passed as an input to the Hypertext Induced Topic Search (HITS) web search algorithm to predict and rank the public places that are likely to expose people to risk of malaria infection. To provide better presentation of malaria transmission modelling, a Java-based visualisation tool will be developed based on the network model and HITS algorithm. This project is based mainly on the thesis by Eze (2012) published in Faculty of Computer Science and Information Technology in UNIMAS with the title of "Web Algorithm Search Engine Based Network Modeling of Malaria Transmission". The visualisation tool will be built by implementing the network modelling technique and web search engine algorithm introduced in the thesis.

The methodology for carrying out the project involves a number of procedures as follows:

- Literature Review

  Research on malaria disease transmission, contact network modelling, HITS search algorithm and existing software or system in disease transmission modelling.

- Feasibility Analysis

  Assess the feasibility of different methods to model the disease transmission. The contact network modelling is preferred since it provides better data modelling result.

- Project Plan

  This step includes specifying the scope of this project based on the time available. The project scheduling and activities planning are also be defined.

- System Development

  After the project plan, the Java-based visualisation system is implemented. Java provides a framework that supports social network analysis for modelling and visualisation of relational data. This available Java API is adopted to develop the visualisation tool that helps in study of malaria disease transmission in public places.

- Benchmark Validation

  UCINET 6 is chosen as a benchmark validation system to determine the appropriateness and accuracy of the resulting system.

- Project Evaluation

  Project review should be conducted to determine whether or not the project objectives are achieved.

1.4 Scope

The scope of the project covers how to implement contact network in modelling of disease transmission such as malaria. This work includes understand and implement this modelling method into building a visualisation system. Besides, this project will be looking at a web search algorithm, Hypertext Induced Topic Search (HITS), which will be implemented into the visualisation system to detect and rank public places (such as markets, schools and fields in Kuching area) in terms of their vector density. The factors that are taken into consideration for model calculation in this project are only a few, such as instance human movements, but not all (e.g. weather) of them. The effects of such unused factors are assumed negligible to the study.

The main intention of this project is to develop a visualisation tool to model malaria transmission. The implementation of this visualisation tool is built based mainly upon the published thesis in malaria disease transmission modelling using contact network model by Eze (2012). Thus, it is of assumption that a contact network model is appropriate to study and predict malaria transmission. It is beyond the scope of this work to include further enhancement and study out of the referred thesis, and may indeed be considered as future work.

1.5 Significance of Project

Since malaria disease is transmitted via bites of infected mosquitoes to humans and human beings have different attributes and behaviours in mixing with others, therefore a heterogeneous network rather than a single-node network modelling method would be more suitable in studying malaria disease transmission in public places. With a contact network model given, a search engine algorithm would be appropriate to be used to detect the public places of interest. The HITS web search algorithm is chosen to be implemented in this system for vector density detection.

As no existing tool that implements contact network method in the study of malaria transmission, the significant contribution of this project is to offer a disease modelling system for researchers to visualise and study malaria transmission. This system also aims to provide disease modelling experts an interactive tool to predict and rank public places in terms of their vector population or density, so that appropriate action can be taken to control the disease transmission in those public places. This in turn is believed and expected that this visualisation tool will have a huge impact on malaria transmission detection and control in public places.

1.6 Project Schedule

The project schedule is attached in Appendix A.

1.7 Project Report Outline

This report consists of six chapters. Following this introductory chapter, reviews on some existing disease transmission modelling systems and papers related to the field are presented in Chapter 2. In Chapter 3, the methodology and requirement analysis and design of the proposed system are discussed. Output of the system and their functionalities are also described and explained. Chapter 4 describes the details of the implementation of this project. System specification and requirements are given as well. The programming codes and library functions used will be discussed in detail. In Chapter 5, the system is tested and validated. Besides that, the results of the proposed system are presented. This project's conclusion and future works are in the final chapter, Chapter 6.

CHAPTER TWO

LITERATURE REVIEW

2.0 Introduction

The aim of this chapter is to present the study and research conducted by other researchers on disease transmission modelling, and to review existing software and system that are related to this work. In particular, this chapter shows what malaria is and how its transmission can be modelled using a contact network. Then, a review and study on existing software systems that used to visualise and model disease transmission will be presented. This chapter ends with a brief conclusion.

2.1 Malaria

Malaria is a dangerous disease caused by a parasite called *Plasmodium*, which is transmitted through the bites of infected mosquito (*Anopheles)* to human beings. Without proper treatment, malaria can quickly become life-threatening by disrupting the blood supply to vital organs in human body (W.H.O, 2013). Several efforts towards reducing the spread of malaria have failed since the parasites have developed resistance to a number of malaria medicine and drugs (Erah et al., 2003). World Malaria Day held recently by World Health Organization has urged for research and academic institutions to join and invest in malaria prevention and control (W.H.O, 2013).

2.2 Contact Network

In disease transmission model, a contact network is a graph structure that is used to represent interactions that lead to disease infections (Eze, 2012). Meyers (2007) describes that a contact network is a graph structure where each vertex represents a person (or location) and the edges represent contacts among people (or locations) in the network (Eze, 2012). Hence for this work, the contact network for modelling malaria transmission is a graph with two types of node which are public places that contain the infected malaria vectors and human beings. The edges of the networks depict the relationship between the human beings and the public places, and this relationship is formed as a result of the human visiting those public places such as markets and fields (Eze, 2012).

2.3 Contact Strength Algorithm

The objective of this section is to give a brief review of the algorithm to generate contact strength in malaria transmission network model. Before the algorithm discussion, it is necessary to give a review on the malaria transmission network model.

2.3.1 Malaria Transmission Network

Since malaria transmitted through mosquito bites, there must be points of contact between human beings and mosquitoes. This implies that there is an interaction between these two nodes (human beings and mosquitoes), which can be utilised by researchers to build network models to study malaria transmission (Eze et al., 2011). This contact strength generating algorithm is used to determine the strength of relationship between these interacting nodes of the network.

2.3.2 Contact Strength Generating Algorithm

As discussed by Eze et al. (2011), the major aim of this algorithm is to quantify in numerical terms the strength of relationship between the interacting nodes. Research in the medical field has shown that the higher the exposure to disease-causing agents (such as vectors), the greater the likelihood to contact with such disease. A professor of medicine, Agius (2011) defined exposure as the product of duration and frequency of contact with the disease-causing agents network (Eze et al., 2011). Thus, duration and frequency of visits among other parameters is used to build the strength of relationship between human beings and public places nodes in this network model.

The contact strength generating algorithm assumed that each human being has predictable and scheduled contact time with public places where, for instance, a student goes to school from Monday to Friday. This algorithm gives rise to a normalized contact strength matrix and it is probability theory compliant. The results show that any two heterogeneous nodes can be generated using this algorithm, implying that probabilistic techniques for further analysis and study will be possible (Eze et al., 2011).

The research on network modelling of malaria transmission has contributed in understanding of malaria transmission in a network of interactions comprised of human beings and mosquitoes. Contact strength generating algorithm is proved to be an important component of network modelling research.