

MASTER'S THESIS

Inferring disease transmission networks

Yang, Xiaofei

Date of Award:
2014

[Link to publication](#)

General rights

Copyright and intellectual property rights for the publications made accessible in HKBU Scholars are retained by the authors and/or other copyright owners. In addition to the restrictions prescribed by the Copyright Ordinance of Hong Kong, all users and readers must also observe the following terms of use:

- Users may download and print one copy of any publication from HKBU Scholars for the purpose of private study or research
- Users cannot further distribute the material or use it for any profit-making activity or commercial gain
- To share publications in HKBU Scholars with others, users are welcome to freely distribute the permanent URL assigned to the publication

Abstract

To investigate how an infectious disease spreads, it is desirable to use the observed surveillance data to discover the underlying (often hidden) disease transmission networks. Previous studies have provided methods for inferring information diffusion networks in which each node corresponds to an individual person within the diffusion network. However, in the case of disease transmission, to effectively propose and implement intervention strategies, it is more realistic and reasonable for policy makers to study the diffusion patterns at a metapopulation level, that is, to consider disease transmission networks in which nodes represent subpopulations, and links indicate their interrelationships. Such networks can be useful in several ways: (i) to investigate hidden impact factors that influence epidemic dynamics, (ii) to reveal possible sources of epidemic outbreaks, and (iii) to practically develop and/or improve strategies for controlling the spread of infectious diseases. Therefore, this thesis addresses the problem of inferring disease transmission networks at a metapopulation level. A network inference method called NetEpi (Network Epidemic) is developed and evaluated using both synthetic and real-world datasets. The experimental results show that NetEpi can recover most of the ground-truth disease transmission networks using only surveillance data.

Acknowledgements

I would like to thank all the professors and colleagues who have helped me during my 6-year's study in Hong Kong Baptist University.

In particular, I would like to thank my supervisor Prof. Jiming LIU. My research interest was initiated by his excellent teaching in my undergraduate courses. His inspiring guidance and constructive suggestions in my master's studies had laid a solid foundation for my future work.

I would also like to thank Prof. ZHOU Xiao Nong, Dr. William Kwok Wai CHEUNG, Prof. YUEN, Pong Chi for their insightful advice and encouragement. Thanks also go to the departmental secretaries and technicians for their kind assistance.

I give my special thanks to my parents and my girlfriend He Jing for their persistent support throughout my studies.

Table of Contents

Declaration	i
Abstract	ii
Acknowledgements	iii
Table of Contents	iv
List of Tables	vii
List of Figures	viii
1 Introduction	1
1.1 Overview and Objectives	1
1.2 Contributions and Significance	5
1.3 Outline of the Thesis	7
2 Related Work	9
2.1 Information Diffusion Network Inference	11
2.1.1 Information Diffusion Models	11

2.1.2	Link Prediction Problem	12
2.1.3	Diffusion Network Inference	13
2.2	Disease Transmission Networks	16
2.2.1	Disease Transmission Models	16
2.2.2	Disease Transmission over Networks	22
2.2.3	Inference Based on Surveillance Data	28
2.3	Summary	34
3	Problem Statement	37
3.1	Definitions	38
3.2	Linear Transmission Model	42
3.3	Network Inference Problem	44
3.4	Summary	45
4	The Proposed Network Inference Method	46
4.1	Basic Ideas	47
4.2	Partial Correlation Network Construction	47
4.2.1	Pearson Correlation Analysis	50
4.2.2	First-Order Partial Correlation Analysis	51
4.3	Back-Tracking Bayesian Learning	53
4.3.1	Marginal Likelihood Function	53
4.3.2	Expectation-Maximization Computation	55
4.3.3	Back-Tracking Technique	56
4.4	Discussions	56

4.4.1	Stopping Criteria	56
4.4.2	Computational Complexity of the Algorithm	58
4.5	Summary	59
5	Experiments	60
5.1	Experiments Based on Synthetic Data	61
5.1.1	Experimental Setting	61
5.1.2	Baseline Method	65
5.1.3	Results	66
5.1.4	Sensitivity Analysis	76
5.2	Experiments Based on a Real-world Dataset	93
5.2.1	Dataset Description	93
5.2.2	Experimental Setting	95
5.2.3	Results	97
5.3	Summary	112
6	Conclusion and Future Work	114
6.1	Main Contributions	114
6.2	Future Work	115
	Curriculum Vitae	129