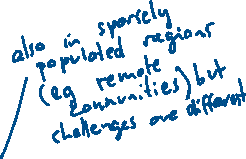
Proposal



1. Background



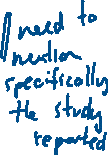
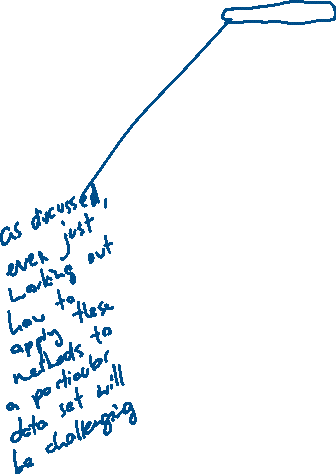
Infection happens frequently recent years, especially at high densely-populated regions. Many researchers try to reveal about the spatial distribution of infection based on the mobility data. Mobility data is a dataset, what describes the mobility behaviour of people in a particular area, such as the journey-to-work data in a country. Many researches have done research at national or international scale, and also others focus on infection that may occur within individual buildings such as school or hospitals. Mobility data used by these kinds of researches is comparably easy to analyse. There are just a few at metropolitan scales. Naturally, how to gain this kind of data is the most important and difficult part in this kind of researches. Yu Zheng et al. proposed the whole process of trajectory data mining and pointed out the possible algorithms used in each section in this process. But there are still some problems cannot be solved: (1) the selection of research scale (e.g. at national scale), if the scale is too small, the possibility that the analysis can be affected by report bias is increased. (2) some important data is hard to collect, such as those characterise private features. (3) it is difficult to determine whether a person is infected or not, and it also involves privacy problems.



1. Outline the question



This research will focus on using existing methodologies and algorithms to mine new trajectory dataset at urban scale that consider the factors that may affect the chance of infecting, such as the length for a person staying at one place and so on.



1. The step to solve the problem

Firstly, I will read literature involves methodologies (such as Mean (or Median) Filter which is used to remove outlier, or how to convert a sequence of raw latitude/longitude coordinates to a sequence of road segments) and algorithms used to mine trajectory dataset (such as “Trajectory Data Mining: An Overview” by Yu Zheng et al. and what it refers), and some papers reveal about the distribution of infection at global, urban and smaller scale. Then, I will learn how to apply the methodologies and algorithms to raw data. Some knowledge of building model based on the dataset is also needed.



Data collection

Social encounter profiles of greater Melbourne residents, by location – a telephone survey

Outline: analyze the data from a survey company based on age, gender, places they stay, encounters with known or unknown people etc. And give conclusions like

This project seeks to define in more detail the attributes of social encounters, beyond degree distribution, that might quantifiably capture such differences.

被调查人平时都跟什么人接触，接触的年龄阶段，以及接触的时长，平时在什么地点停留多长时间

调查了两个区域的数据，先总结了一下两个区域居民大致的整体情况，例如居民的国籍，语言，收入等等。然后是调查方法，本文中采取电话调查的方式，并对受访者根据年龄，性别，household type等进行分类。

location, age, gender and area of residence

得到数据后对数据进行了处理，

按照地点分

This last figure strongly reasserts the importance of household settings in providing opportunities for close-contact transmission of infection.

按照年领分

The highest reported number of median contacts was among individuals aged between 30 and 49 years.

按照是否与认识的人contact

Young men, and women aged 30–49 years, reported many more known contacts than other respondents

不同区域

Conclusion

Assumptions regarding the number, dur- ation and clustering of contacts in model frameworks all have significant implications for the simulated spread of infection.

The incorporation of geographical space into repre- sentations of social networks is a recognised challenge in the field of infectious disease modelling

**Trajectory Data Mining: An Overview**

In this article, we conduct a systematic survey on the major research into trajectory data mining, providing a panorama of the field as well as the scope of its research topics. （总体主题）

Following a road map from the derivation of trajectory data, to trajectory data preprocessing, to trajectory data management, and to a variety of mining tasks, the survey explores the connections, correlations, and differences among these existing techniques. This survey also introduces the methods that transform trajectories into other data formats, such as graphs, matrices, and tensors, to which more data mining and machine learning techniques can be applied. Finally, some public trajectory datasets are presented.

classify the sources generating trajectory data into four groups-> trajectory preprocessing (noise filtering, segmentation, and map matching)-> trajectory data management algorithms -> mining data

1. Trajectory Uncertainty

2. Trajectory Pattern Mining

3. Trajectory Classification

4. Trajectory Outlier Detection

Build a better social network and travel recommendation

Mobility of people

Active Recording

Passive Recording

**Map Matching**

Map matching is a process to convert a sequence of raw latitude/longitude coordinates to a sequence of road segments.