Wei Chien Benny CHIN

Academic portfolio

September 2018

A series of works on using computing techniques to uncover geographical structures.

CONTENTS

INTRODUCTION

This portfolio contains most of my works during my MSc and PhD studies, which includes some of my research outputs that were published and some project outcomes. My research consists of two major parts. First, I focus on using the network concept to understand the urban form, specifically on the human movement. Second, I also concentrated on using space-time analysis methods to reveal the infectious disease spreading structure. The following text is organized in three major parts, including (1) **Network analysis in human movement applications**, (2) **Space-time analysis in disease diffusion applications**, and (3) **Other work products**.

(1) NETWORK ANALYSIS IN HUMAN MOVEMENT APPLICATIONS

Published works

A CITY NETWORK STUDY

Geographical distance is a key in spatial network analysis.

Geographically modified PageRank

A network approach, which simplifies geographic settings as a form of nodes and links, emphasizes the connectivity and relationships of spatial features. The PageRank algorithm, a network metric, is often used to help identify important nodes where people concentrate.

In this study, we intended to integrate the spatial perspective, i.e. the distance decay effect, into the network analysis algorithm for identifying the major cities where people tended to go from a transportation network (Fig. 1). The two proposing algorithms are Distance-decay PageRank (DDPR) & Geographical PageRank (GPR).
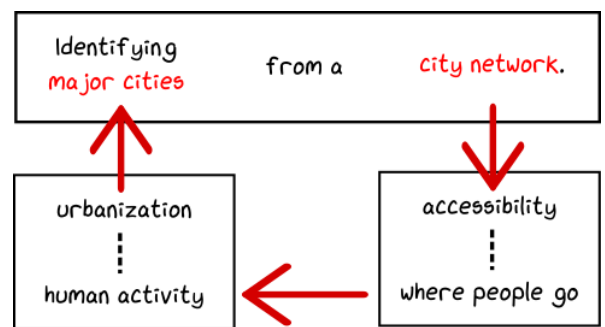

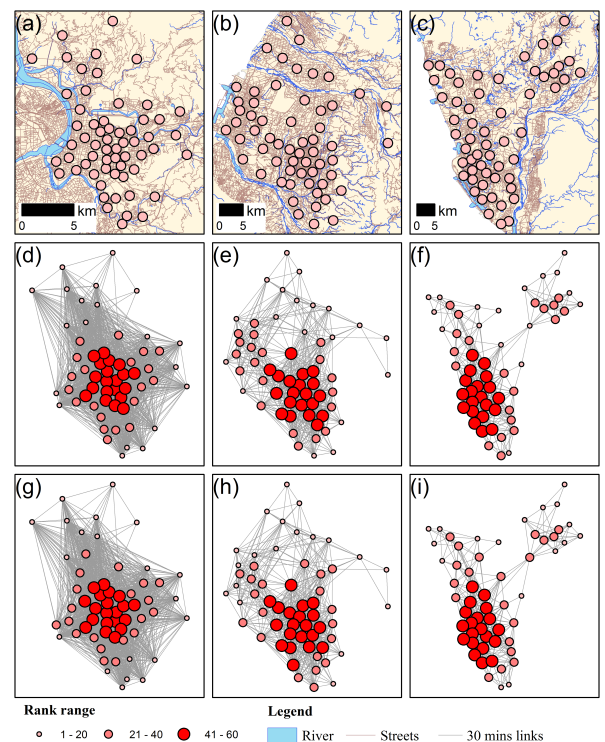Figure 1: The conceptual framework of the city network study.

Figure 2: The DDPR & GPR results for the within-city analysis.

Our findings indicate that in both inter-city (Fig. 3) and within-city (Fig. 2) settings the
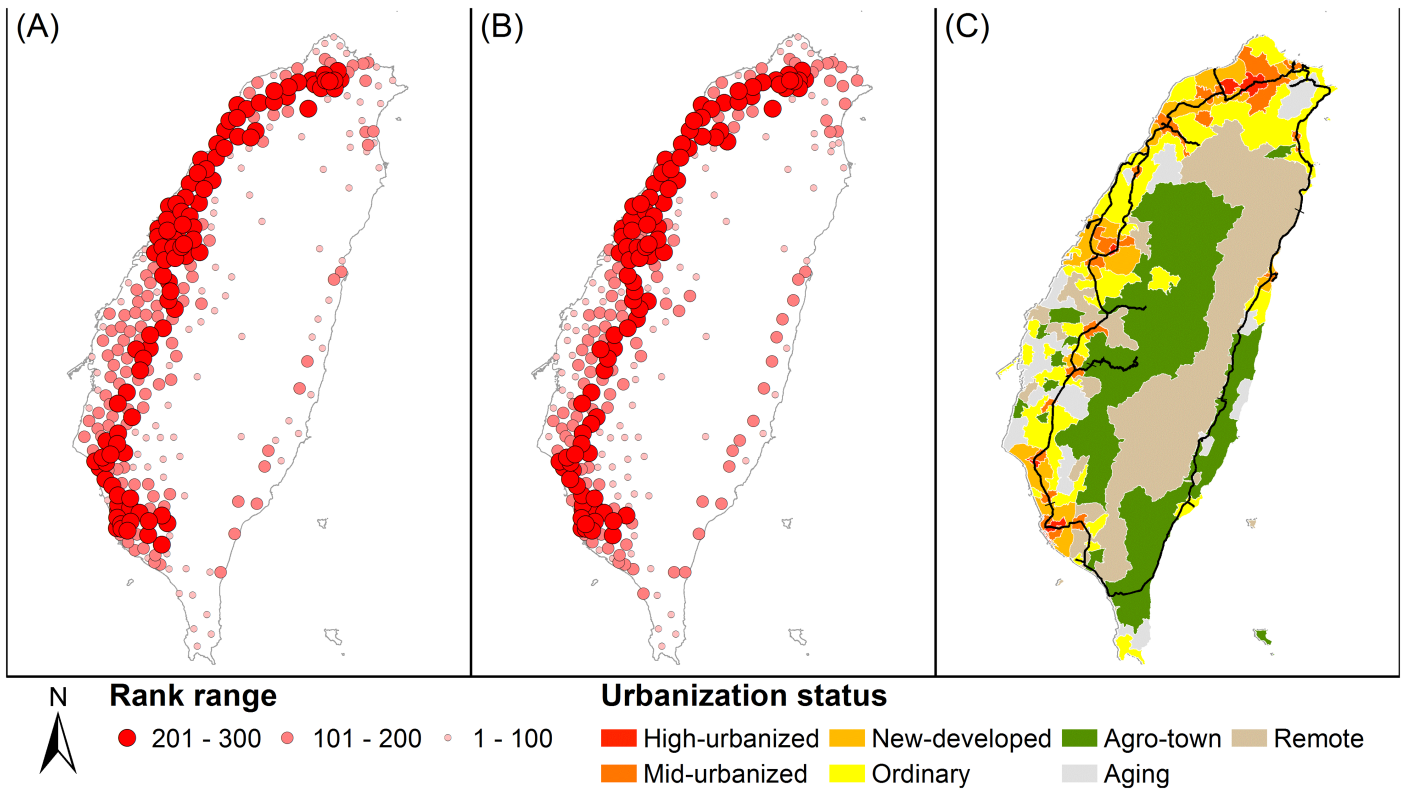
1

Figure 3: The DDPR & GPR results for the inter-city analysis.

Rank range
● 201 - 300 ● 101 - 200 · 1 - 100

Urbanization status
■ High-urbanized ■ New-developed ■ Agro-town ■ Remote
■ Mid-urbanized ■ Ordinary ■ Aging

proposed algorithms more effectively capture the spatial locations where people reside than traditional commonly-used network metrics. In comparing location attractiveness and distance decay, we conclude that the concentration of human movement is largely determined by the distance decay. This implies that geographic proximity remains a key factor in human mobility.

full article info
Geographically Modified PageRank Algorithms: Identifying the spatial concentration of human movement in a geospatial network. `https://doi.org/10.1371/journal.pone.0139509`

STREET CONGESTION STUDIES

The location attractiveness may affect the moving pattern of cars.

Flow-based PageRank

The aim of these studies are to reveal the underlying attractiveness along the streets that may influence the mobility of automobiles (Fig. 4). An algorithm named Flow-based PageRank (FBPR) was proposed, which used a similar idea as the DDPR & GPR, but in a reverse way. While the DDPR & GPR assigned the attractiveness and distance decay effect according to some concepts, FBPR identify the attractiveness based on a heuristic optimization process using a partial collection of the street segments with known traffic flow.

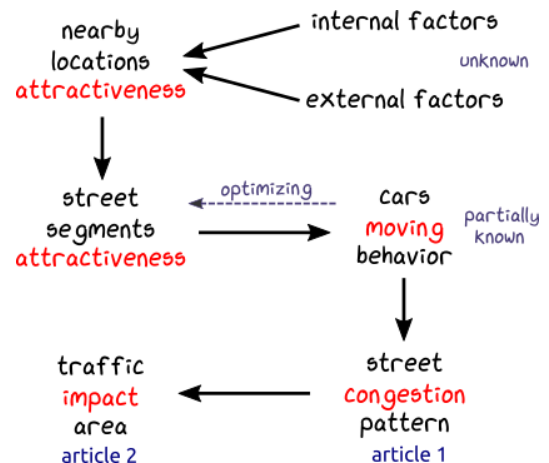The idea is simple: we can count the number of cars moving along some streets in the city through

Figure 4: The conceptual framework of the street congestion studies.

some sensors; only partial of the streets have sensors; the streets are interconnected as a whole system; using the partial streets with flow data for the evaluation procedure, we model the attractiveness of street segments, which affect the flow of cars within the street system.

One of the studies focused on the cars' moving behavior that resulting in street congestion. Another study delineated the traffic impact area of the places with high internal connectivity.

Identifying the topological characteristics and flow complexity of traffic congestion provides network topological insights for sustainable urban planning, and these characteristics can be used to further understand congestion propagation.
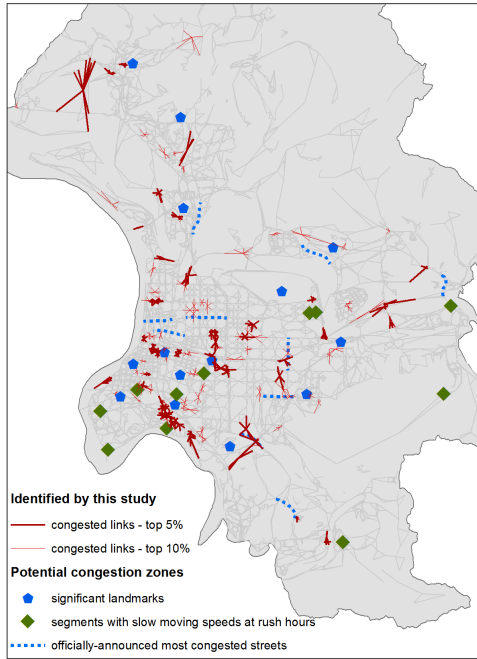
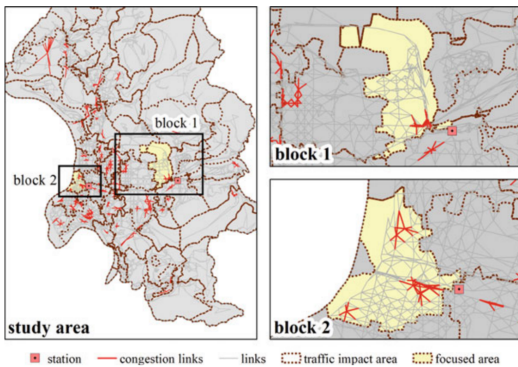Figure 5: The identified street segments that is prone to congestion.

Figure 6: The delineated traffic impact area.

ALONG THE WAY BACK HOME

Modelling the influences of human movement to disease spreading

The two-ways movement

While the basic PageRank (PR), DDPR, GPR, and FBPR were all considering the forwarding movement, some network also implied the backward movement, such as the commuting network that draw a directed link from where people live to where they work; these people will be going back home after work. This may affect the diffusion pattern of infectious disease. Therefore, we proposed an algorithm named EpiRank that considered both forward and backward spreading possibility, with a

parameter to control the effect proportion of the two directions. Fig. 7 showns the concept of the EpiRank, that the same number of people will be commuting in the opposite directions during the morning and evening.
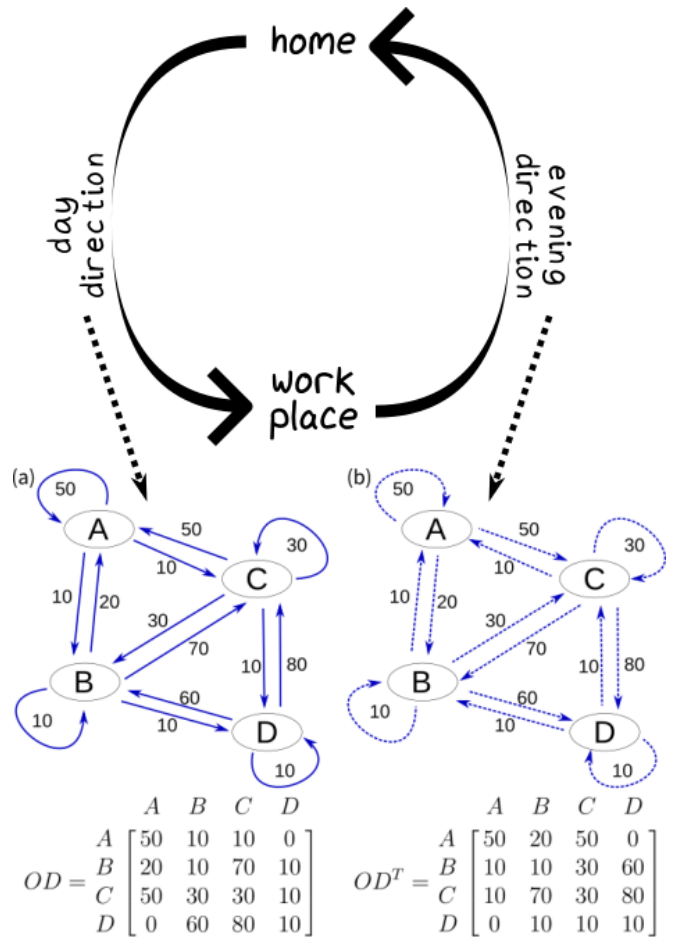

Figure 7: The concept of EpiRank, a modified PageRank considering both forward and backward movement.

The result suggested that the EpiRank was better capturing the disease diffusion pattern than other network metrics; the backward movement of commuters contributed more effects on the spreading than the forward movements.

BRIDGE OR BOND

Differentiating edge type in networks

The HETA algorithm

Many network researchers use intuitive or basic definitions when discussing the importance of strong and weak links and their roles. Others use an approach best described as if not strong, then weak to determine the strengths and weaknesses of individual links, thus deemphasizing hierarchical network structures that allow links to express different strength levels. Here we describe our proposal

3

for a hierarchical edge type analysis (HETA) algorithm for determining link types at multiple network hierarchy levels based on the common neighbor concept plus statistical factors such as bond links, kth-layer local bridges, global bridges, and silk linksall generated during long-term network development and evolution processes. Two sets of networks were used to validate our proposed algorithm, one consisting of 16 networks employed in multiple past studies, and one consisting of two types of one-dimensional small-world networks expressing different random rewiring or shortcut addition probabilities. Two applications with potential for developmental contributions are demonstrated: a network fingerprint analysis framework, and a hierarchical network community partition method.

(a) Network ($char.\ PathLength = 2.282$)
(b) Silk link
(c) Bond link ($T_E^1 = 0.8506 \leq R_e^1$)
(d) Local bridge ($T_l^1 = 0.4471 \leq R_e^1 < T_E^1$)
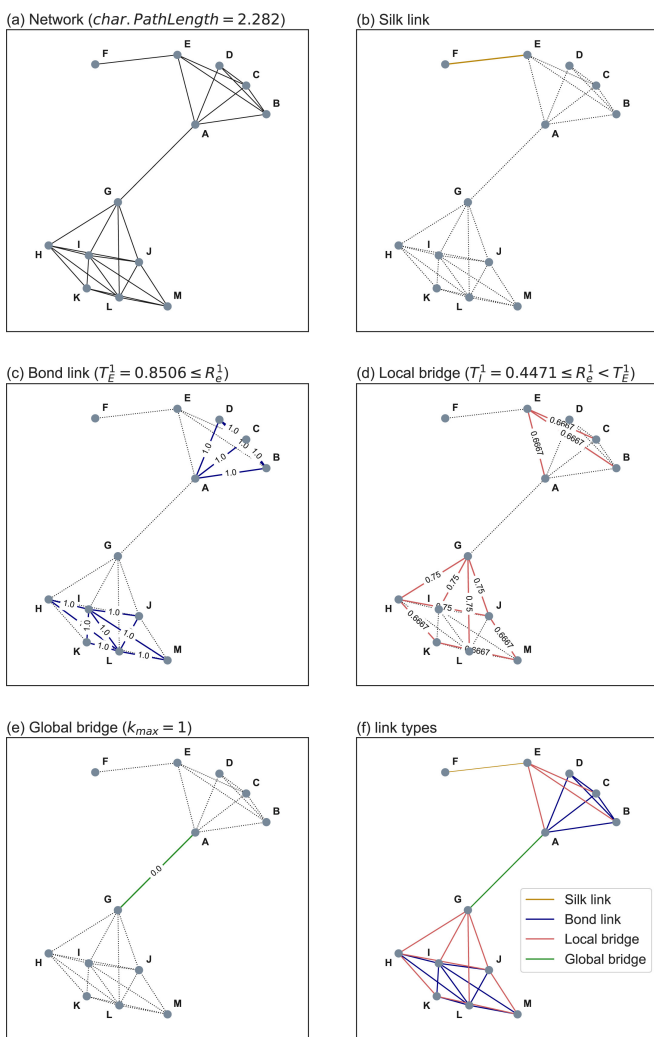(e) Global bridge ($k_{max} = 1$)
(f) link types

Figure 8: A demonstration of HETA, differentiating edge types into the silk link, bond link, local bridge, and global bridge.

Fig. 8 shows the calculation framework and the four identified link types. The example network presented in (a) consists of 13 nodes and 28 links, with a diameter of 5 units and average shortest path length of 2.282 units. Identification procedures for the four link types are shown in (b)-(e), with the final results displayed in (f). The result

of HETA is then used to partition the network into different communities.

A NETWORKED BUILDING STUDY
Controlling disease in campus by isolating building

Class attending network

Students attend classes in different buildings and thus move around the university campus. Therefore, if a contagious disease is introduced into the campus, it can be spreading throughout the campus by the students. In order to stop the disease diffusion, it is important to stop the movements within the campus. In this study, we tried to identify the most important weak links (the less connected pair of buildings), that can effectively separate the campus into different isolated parts if the movement between these buildings is stopped.
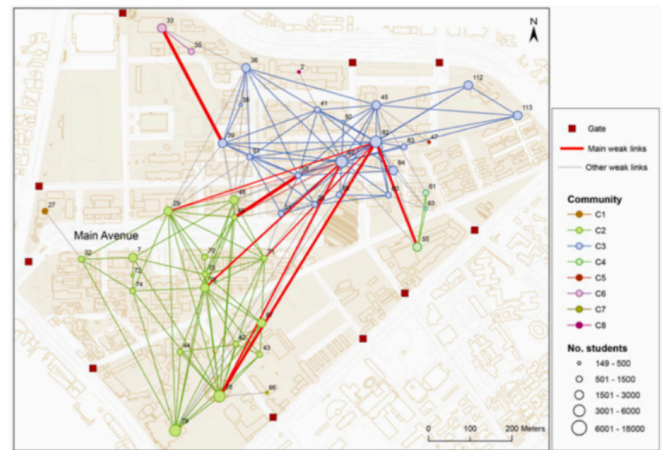

Figure 9: The building community in the campus network, with the red links showing the major weak links. Major weak links indicated the critical connections between community that had a higher edge betweenness, and which should be remove to isolates the communities.

A PATIENT FLOW STUDY
Seeking health care service

Needs vs. Distance

While there are several health care services (HCS) available within an affordable distance, patients may travel to one of these HCS. In other words, while the accessibility to HCS is high within the
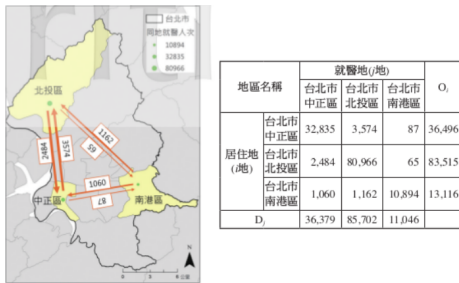
Figure 10: An illustration of the patient flow from the National Health Insurance Database and the Origin-Destination matrix of patient flow.

area (i.e. distance is not a problem), other factors such as the HCS type, resource, etc. will change the patients preferences and the HCS catchment area. This study aimed to analyze the effects of spatial distance between patients and HCS, and the resources provided by the HCS to the preferences of HCS seeking behavior. Using Taiwan as a case study, the result showed that the gravity model with weighting by resources is better to present the HCS accessibility.

full article info

Profiling the patient flow for seeking health care in Taiwan: using gravity modeling to investigate the influences of travel distance and health care resources. (abstract in English) https://doi.org/10.6288/TJPH201635104086

Recent development

Lifestyle pattern of a city

In this study I collected geocoded tweets data from Twitter API for several cities, including Tokyo, Jakarta, Singapore , etc., to delineate activity spaces and time for understanding the lifestyle of the cities. Each city has different activity space and time patterns. The Tokyo cases will be used as example in this portfolio, which results were presented in Tsukuba Global Science Week, CiC forum.

Fig. 11 shows the heatmap distribution of tweets in Tokyo. The tweets is aggregated into cells in Tokyo 23 special wards area, and which is used to delineate the activity spaces based on the co-occurrence relationship and the Louvain community detection method. The resulting activity spaces is shown in Fig. 12.

Using the similar framework, the activity time is also identified, which is shown in Fig. 13, along with the number of users and tweets by hours in days. It is worth to note that the number of tweets and active users during the night hours of Friday and Sunday, that people who stop tweeting at midnight on Monday to Thursday will stay awake later on Friday; and they will stop tweeting at midnight of Sunday and which is similar to the other weekdays. Based on these observation, many tiny information about human behaviors that could represent the lifestyle pattern of people within the city was recorded in the tweets data. In other words, I think this dataset is valuable to understand the everyday life of people in cities, and this is the reason I started and continue to work on this project.

The global communities

This project is still in a beginning stage. I tried to detect the international community through the migration relationship network. The data is downloaded from The World Bank Global Bilateral Migration database. The database recorded the migration between the 231 countries, which can be aggregated by gender (male, female, and total), years (1960 to 2000, with 10 years interval).

Fig. 14 shows the communities from the year 2000 migration network. Some pattern could be observed: while the Europe formed a community, the United Kingdom and Ireland was in another community that was mainly located at the North America and South America; three major communities were found in Africa, that two communities were separated with a line between Republic of the Congo to the northwestern part and Democratic Republic of the Congo to the southern part; another community was located from Sudan to the northeastern part; a community was formed somewhere between Central Asia and Eastern Europe, which location was overlapped with the countries that Russian is their official language or one of the languages in the country.

The structure and framework behind the global migration pattern is a complex combination of different factors, such as political status, economic relationships, socio-demographic pattern, etc. I think the network structure of global migration and the dynamic of this structure requires more investigation.

Figure 11: The tweets distribution in Tokyo City, an example near Shibuya Station. (interactive map is available at http://bit.ly/tokyotweetheatmap)
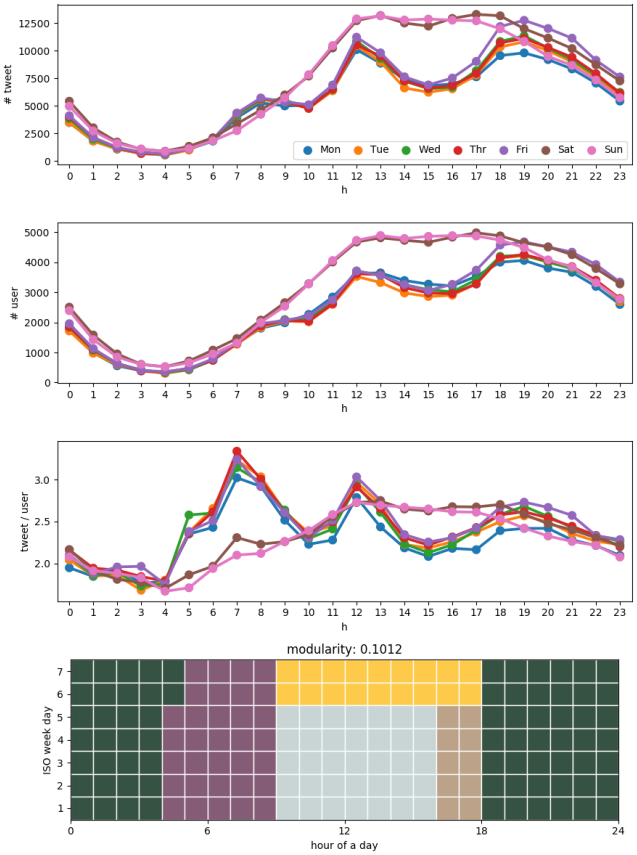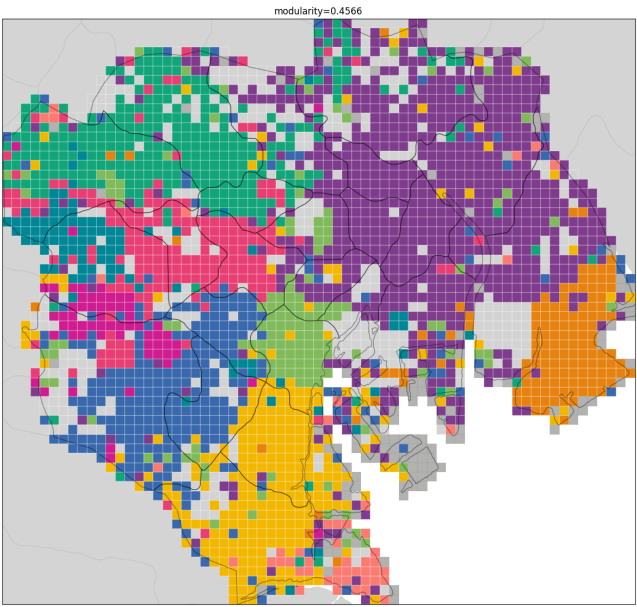

Figure 12: The activity space of Tokyo City.

Figure 13: The activity time of Tokyo City with the number of tweets (a) and number of active users (b).
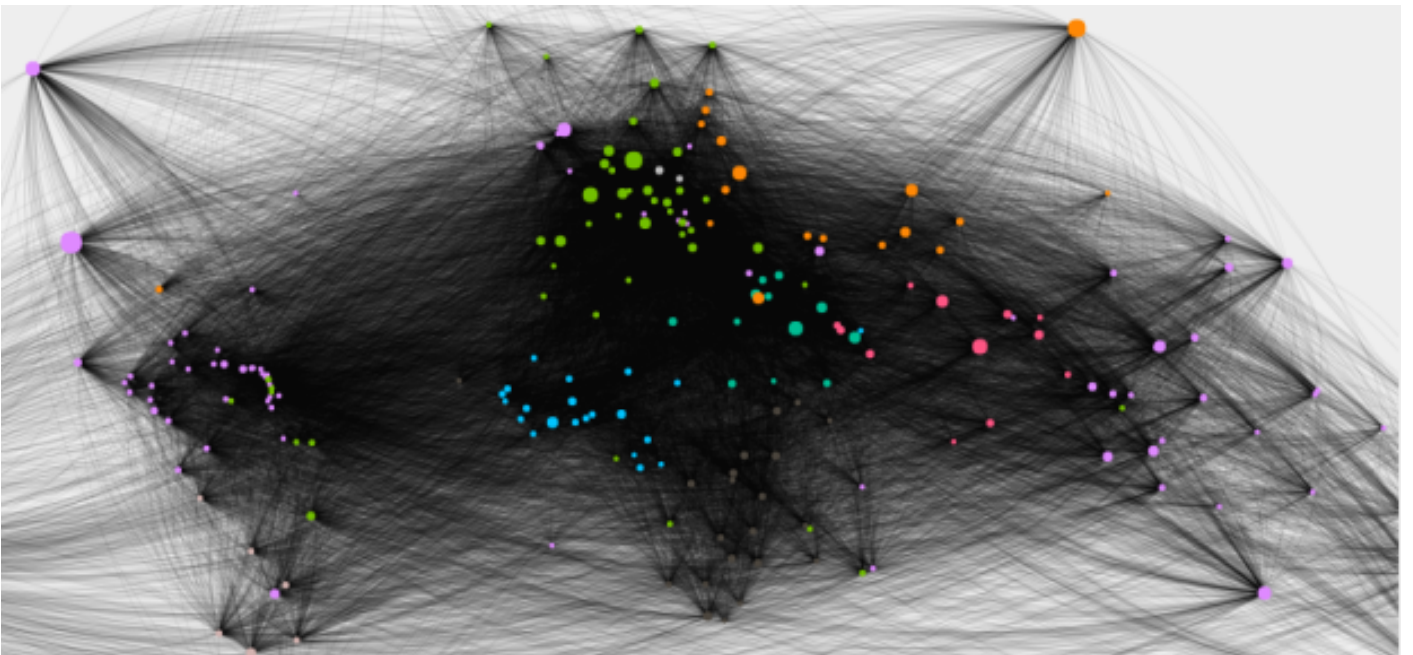
Figure 14: The country communities detected using migration network (2000) (interactive map is available at http://bit.ly/migration_network_2000).

Related projects

Real-time Traffic API Taipei City Gov-

Figure 15: A demonstration of using TGOD package to get the latest bus locations in Taipei City.

ernment released the realtime traffic data through the open data policy. To get the open data, which may be released in xml, json, or the gzipped version of xml and json format, I wrote

a python package to download and process the data into a spreadsheet-like dataframe format, and also clean the time format from varying pattern. Beside these, the package also provided some static geographic dataset such as the county and township polygons, the street and rail transportation polyline and some other datasets. The pacakge is named Taiwan Geographical Open Data (TGOD), which is an open source package.
Manual page: https://wcchin.github.io/tgod/

Taipei City Realtime Traffic

Using the aforementioned TGOD package, I created a simple web-app based on a python server dashboard application–plot.ly dash (Fig. 16). In this web-app, three types of open data is used, including the location of every bus, vehicle detection (VD), and youbike station. The bus data records the latitude, longitude, moving speed, heading direction, etc. The VD dataset records the number of car, trucks passing the sensors, and the average moving speed in each lane. Youbike is a bike-sharing system in Taipei City, which station data records the number of available bike and available port for returning the bike.

OD flows of MRT & Youbike

To visualize the dynamic patterns of human movement in the Taipei City, I downloaded the MRT (Massive Rapid Transit) station to station OD flow data, and the Youbike station to station records. MRT is one of the major commuting transportation for the people in Taipei City, whereas Youbike, a bike sharing service, is the public transportation for the last mile of commuting. The MRT data recorded the number of person who entered into the MRT from one station and leave from another station; the Youbike data recorded each trip information, including the time and station of renting a bike, and the time and station each bike was returned to.

Using processing, a open source graphical library and playground for building electronic arts, I mapped both the MRT and Youbike movements by minutes of a week (Fig. 17). The dynamic pattern of the movements showed the movements of people from where they live to the city center in the morning, i.e. where they work or study, and the movements from city center to the satelite townships in the evening.

A day of Taipei - bus

As part of the project, we had accessed the traffic open data, including bus location, Youbike station status, VD, and MRT station information every five minutes, to store the data locally in our database. Therefore, a lot of bus location data was recorded. In this part, I visualized the bus movement of a day
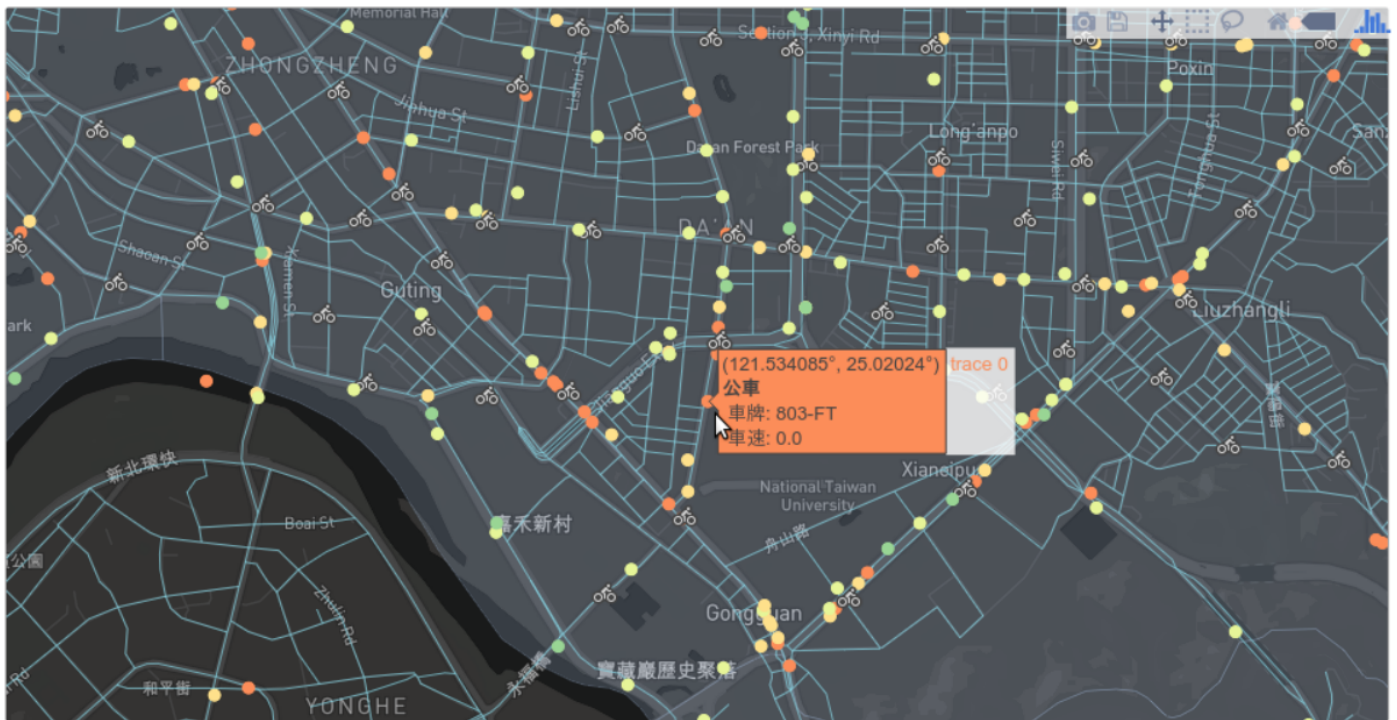
Figure 16: A simple web-app for showing the realtime traffic in Taipei City using TGOD and plotly-dash. The basemap was customized network using Mapbox.
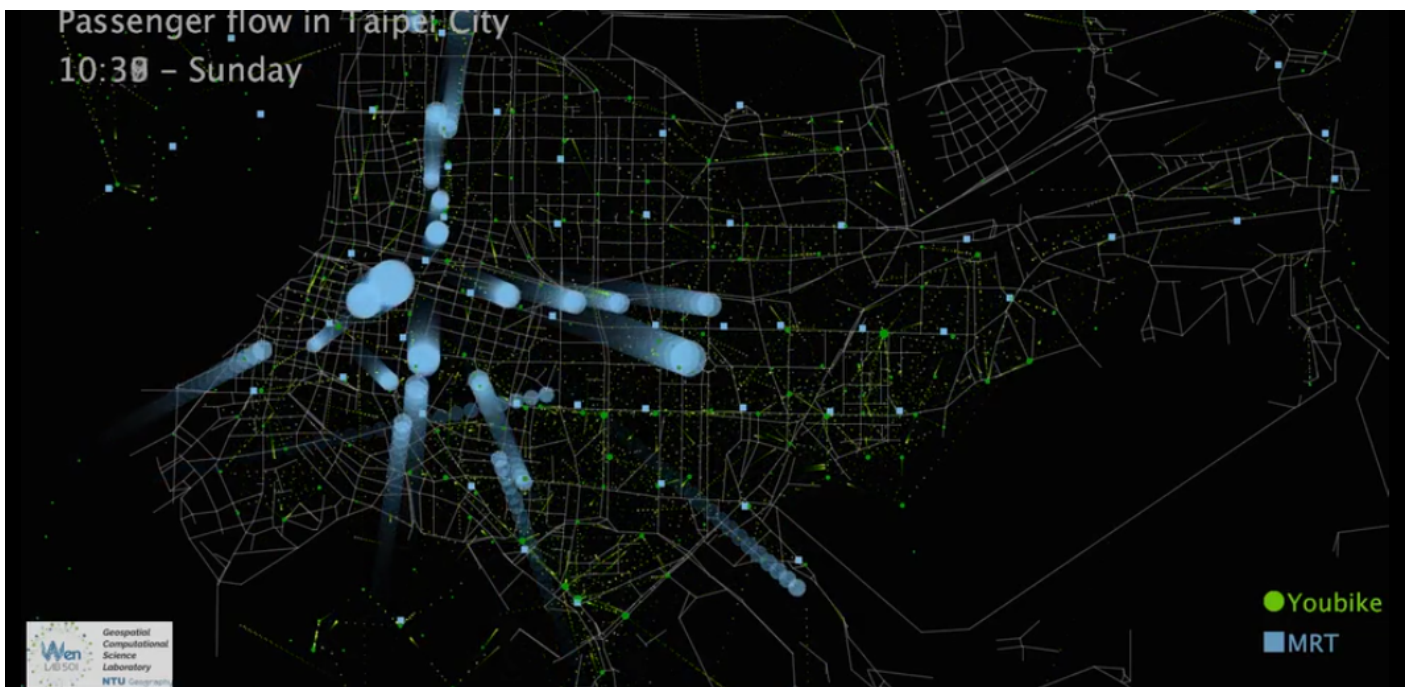

Figure 17: An animation showing the daily dynamic of the human movement using MRT and youbike. **Animation:** https://vimeo.com/210724895

in Taipei City (Fig. 18). The dynamic pattern shows that the buses starting to move into Taipei City center in the morning, and leaving the city center at midnight. While this pattern is intuitive and ex-pected, the dynamic visualization also showed the bus sending the key resources – human resources – into the city to jump start the city metabolism.
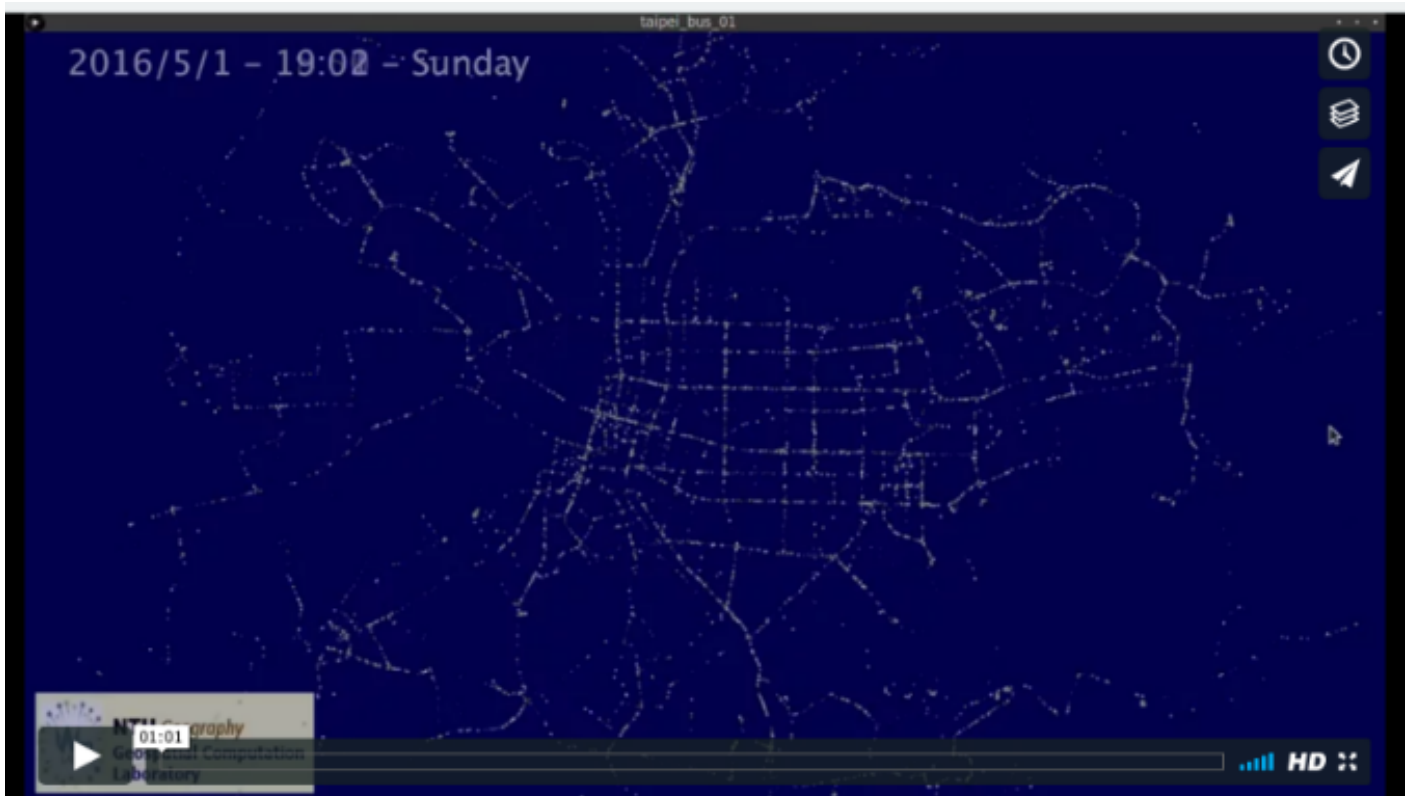
Figure 18: An animation showing a day of Taipei based on bus location data. **Animation:** https://vimeo.com/174331241

(2) SPACE-TIME ANALYSIS IN DISEASE DIFFUSION

Published works

UNCOVERING SPACE-TIME MOVEMENTS

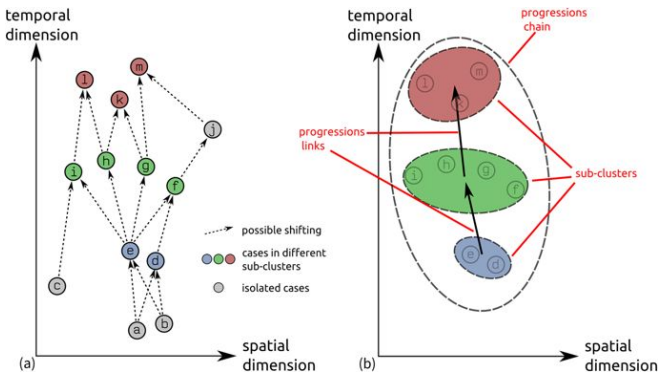Tracking progression of disease diffusion

TaPiTaS algorithm

Figure 19: The objectives of the TaPiTaS algorithm.

A diffusion process can be considered as the movement of linked events through space and time. Therefore, space-time locations of events are key to identify any diffusion process. However, previous clustering analysis methods focused only on space-time proximity characteristics, neglecting the temporal lag of the movement of events. We argue that the temporal lag between events is a key to understand the process of diffusion movement. Using the temporal lag could help to clarify the types of close relationships.

This study aims to develop a data exploration algorithm, namely the TrAcking Progression In Time And Space (TaPiTaS) algorithm, for understanding diffusion processes. Based on the spatial distance and temporal interval between cases, TaPiTaS detects sub-clusters, a group of events that have high probability of having common sources, identifies progression links, the relationships between sub-clusters, and tracks progression chains, the connected components of sub-clusters (Fig. 19).

Dengue Fever cases data was used as an illustrative case study. The temporal lag for extrinsic and intrinsic incubation periods (EIP and IIP) for Dengue Fever were shown in Fig. 20). EIP and IIP are the time for Dengue Fever to develop before the human host become infective. In other words, someone who has been infected will not be infective during these periods, and if some other person is infected on this period, this second person is not infected by the first person, but they could be infected by the same source. Thus, the concept of timelag, is utilized to differentiate the relationships between patients into neighboring pair (the cases are not infecting each other), and shifting pair (the earlier case is possible to infect the latter case). The core concept of TaPiTaS algorithm is then to track the progression of disease transmission from one case to another, and to identify the sub-cluster who may be infected from the same sources.
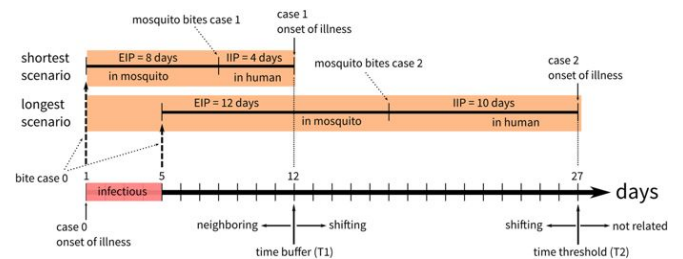

Figure 20: The extrinsic and intrinsic incubation period that cause the mandatory timelag between the the two case that exist infection relationship.

The location and temporal range of sub-clusters are presented, along with the progression links

(Fig. 21). TaPiTaS algorithm contributes a more detailed and in-depth understanding of the development of progression chains, namely the geographic diffusion process. The results were published in animation format, which screen shot is shown in (Fig. 22), with the links for year 2002, 2011, 2014, and 2015, which is the large scale epidemic years in the study area (Kaohsiung City, Taiwan).

full article info

A geo-computational algorithm for exploring the structure of diffusion progression in time and space.

FACTORS AFFECTING TIME-LAG

Time-lag of disease spreading

Risk and duration

Taiwan located on the Tropic of Cancer, causing the local weather to have hot (from about April to September) and cold (from about October to March) seasons. This weather condition is causing the local Dengue Fever epidemic to stop developing during the cold season, and which leads to an annual cycle pattern of Dengue Fever epidemic in Taiwan. The Dengue Fever epidemic of each year is causing by the international travelers who acquire Dengue Fever infection when they are traveling in Southeast Asia, and imported the Dengue virus into Taiwan. In other words, the Taiwan indigenous (local) Dengue Fever epidemic is initiated by the imported cases. Previous studies have shown that imported Dengue Fever cases are able to initiate indigenous epidemics when appropriate weather conditions are present. However, the spatial-temporal associations between imported dengue cases and indigenous epidemics in areas with different social-demographic conditions remain unclear.

This study investigated the determinants of spatial-temporal lags of imported dengue cases that initiated indigenous epidemics from 2003 to 2012 in Taiwan. We constructed Origin-Destination (OD) pairs of transmission from imported townships to local emergence (Fig. 23) to capture two important epidemiological characteristics: 1) the risk of diffusion and 2) the speed of transmission. We further explored the spatial-temporal associations between disease importation and social-demographic indicators with geographic heterogeneity.

Using Poisson regression analysis, the results indicated that there were significant relationships between the time lags from imported cases to local emergence in late spring and early summer. Moreover, urbanization levels influence the speed of transmission. Our findings also showed that the risk of diffusion weighted by distance-decay characteristics increases the explanatory power of socioeconomic variables in imported townships. These findings imply that disease importation in late spring and early summer may be an early warning indicator of indigenous dengue epidemics and that the risk of transmission may be significantly influenced by the effects of inter-township travel.

full article info

Evaluating the role of disease importation in the spatiotemporal transmission of indigenous dengue outbreak.

AN AGENT BASED MODEL

Modeling different diffusion

Nearest neighbor analysis

Spatial diffusion processes can be seen in many geographic phenomena that spread or migrate across space and over time. Studies of these processes were mostly done with verbal description until Hgerstrand (1966) started to approach it with quantitative models. A variety of attempts were made to continue this effort, but only with various degrees of success. Recognizing the critical role that distances between geographic objects or events play in a spatial diffusion process, we experimented with a new approach that uses these distances to detect and distinguish different types of spatial diffusion processes.

My contribution to this study is using the Netlogo to construct a series of experiments that simulate the expansion and relocation diffusion processes (Fig. 24). The outputs of these experiments were then used to study the changes of spatial relationships (using nearest neighbor ratio) over time, and to recognize the space-time development of point pattern.

full article info

An experiment to model spatial diffusion process with nearest neighbor analysis and regression estimation.

Related projects

DENGUE FEVER IN 2018

A fast & auto mapping platform

During the end of August and early September of 2018, Dengue Fever clusters occur in some of the counties in Taiwan, including New Taipei City and Taichung City. Our team, leads by Prof. Tzai-Hung Wen, is helping the local health agency to map the events and to analyze the clustering pattern. To effectively process the raw data (with address of onset cases) into map data, and to analyze the epidemic status, we build an automatic mapping framework to run, publish and map the epidemic patterns. The framework is shown in Fig. 25, which is mainly done in Python environment.

After the local agency provides the raw data, it is put into the framework on localhost, and the procedures is activated by us, which will
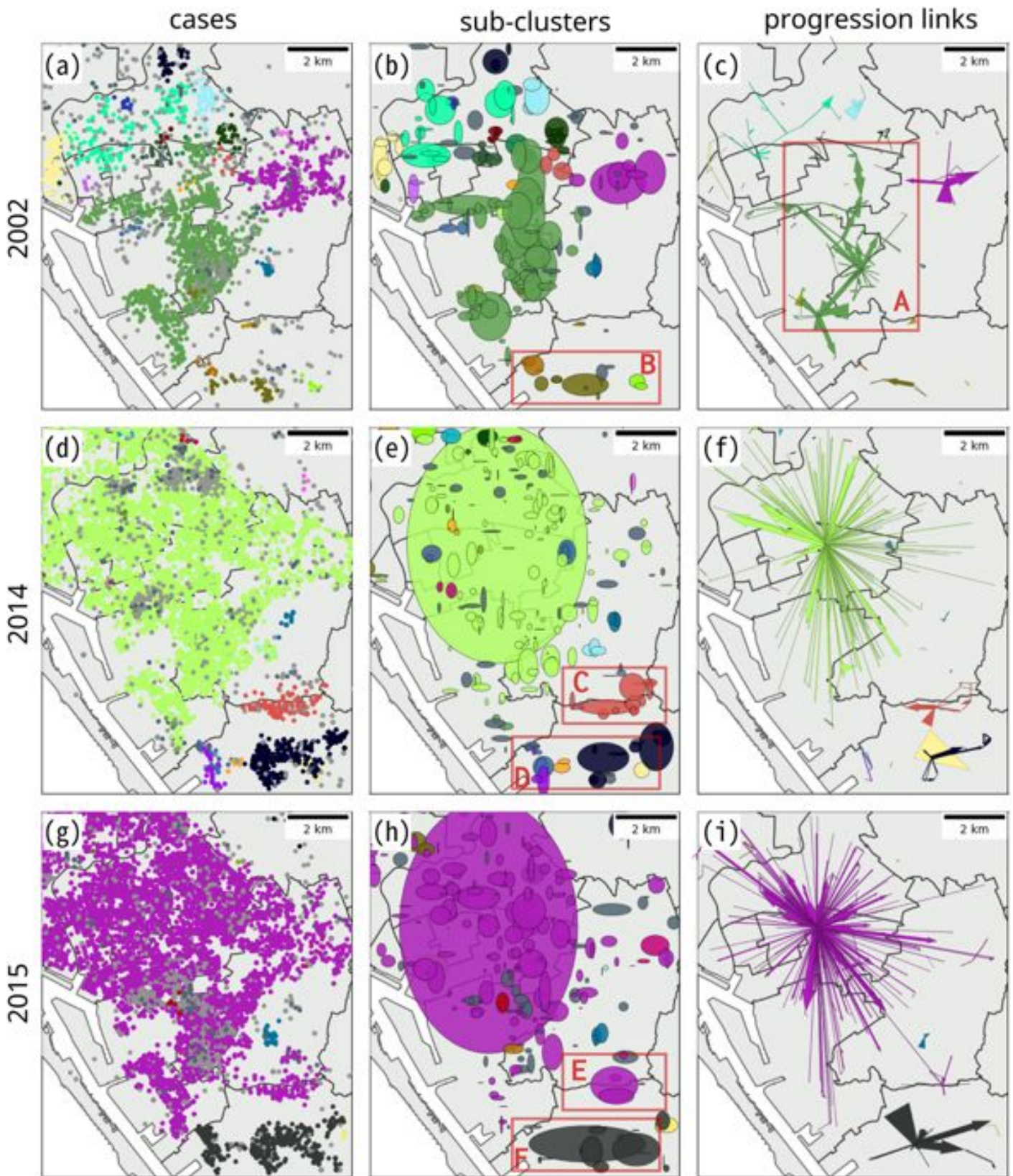
Figure 21: The cases and two outputs of the TaPiTaS algorithm for the three years (2002, 2014, 2015): (a) cases, (b) sub-cluster, (c) progression links (links between sub-clusters).

automatically geocode the address into coordinates (in EPSG:4326 for visualization and is projected to EPSG:3826 for calculation). The data is used to run the TaPiTaS algorithm and MST-DBSCAN (another space-time clustering algorithm developed by another teammate); the results are exported into a published folder, and overwriting the previously existed files. The urls will be automatically sync to the Carto platform in every hour, which will then automatically change the data for the map on Carto (Fig. 26, interactive map is available on
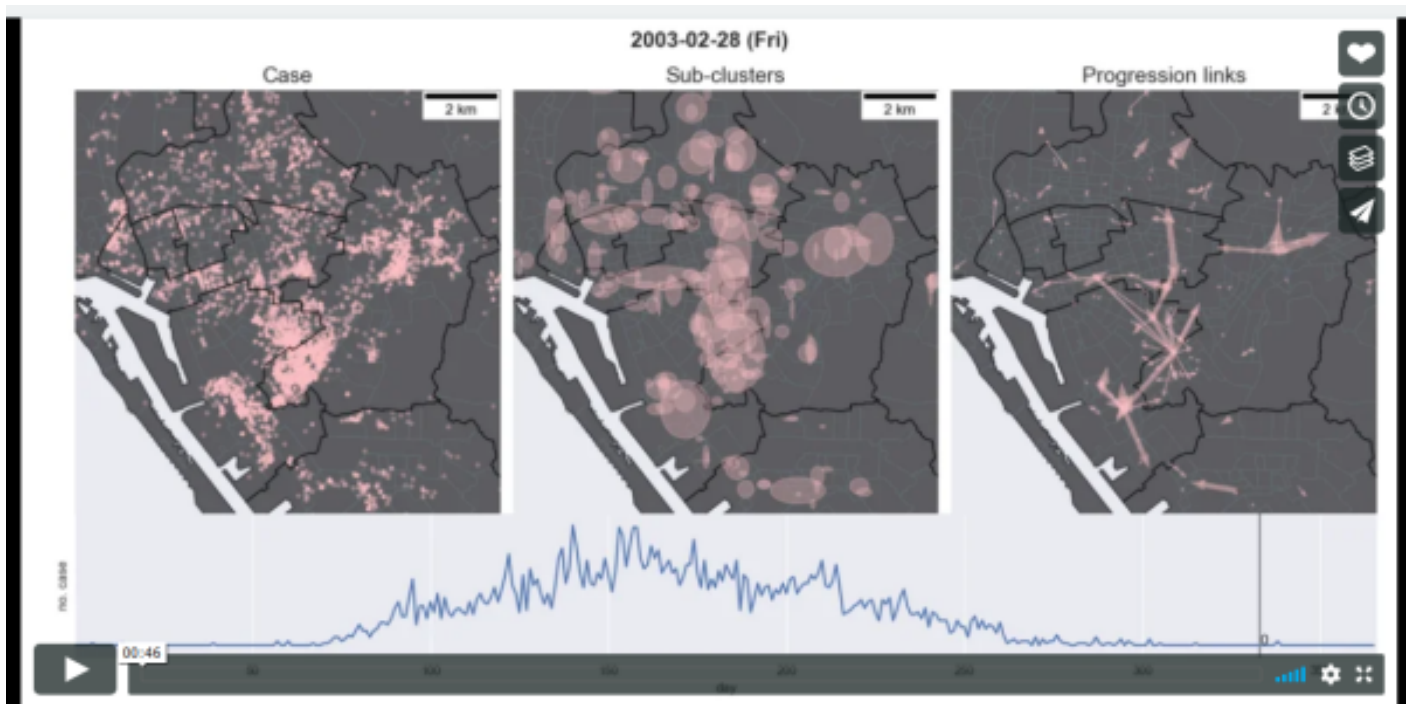
Figure 22: The animated version of the previous output maps. 2002: `https://vimeo.com/210724781`, 2011: `https://vimeo.com/210724791`, 2014: `https://vimeo.com/210724798`, 2015: `https://vimeo.com/210724814`.
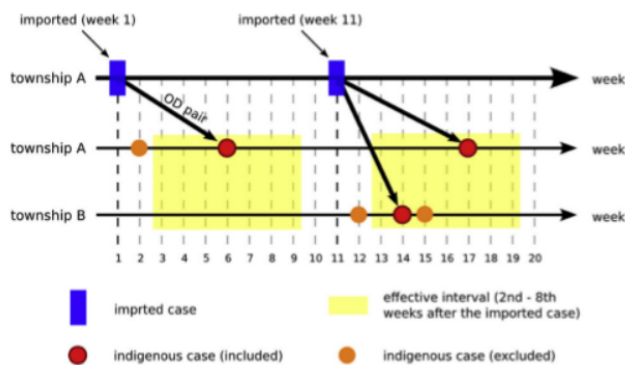

Figure 23: The timelag between imported townships (source) and local townships (target).

`http://bit.ly/wenlab_automapdengue_2018`).

ACUTE DIARRHEA
A foodborne disease monitoring system

Acute diarrhea can be a result of some food-borne disease, which is commonly caused by pathogens including Rotavirus, Salmonella spp., Norovirus, etc. in Taiwan. Taiwan CDC and hospitals have a system named Realtime Outbreak and Disease Surveillance system (RODS), which records and reports the realtime observation of patients with certain symptoms, and acute diarrhea is one of the symptoms. The aim of the RODS system is to monitor the country-wide status for rapid reaction to control disease outbreak. Therefore, it monitor the symptoms instead of the laboratory-confirmed cases, which probably takes several days to weeks for experiments. Our team was assigned to identify possible clusters from the RODS database, and to validate the detected clusters based on another database (Laboratory Automated Reporting System, LARS) after a temporal period. The framework is shown in Fig. 27. The prototype system is shown in Fig. 28, and one of the page for showing recent clusters is shown in Fig. 29.

(3) OTHER WORK PRODUCT

TUTORIAL
Cartography
QGIS learning website

QGIS is an open source software for cartography and spatial analysis. The tutorial aimed to provides people who want to learn making maps in QGIS a self-learning resource. The website is formerly created as a reference for a QGIS cartography class in Department of Geography, National Taiwan University, in which I was a teaching assistant.
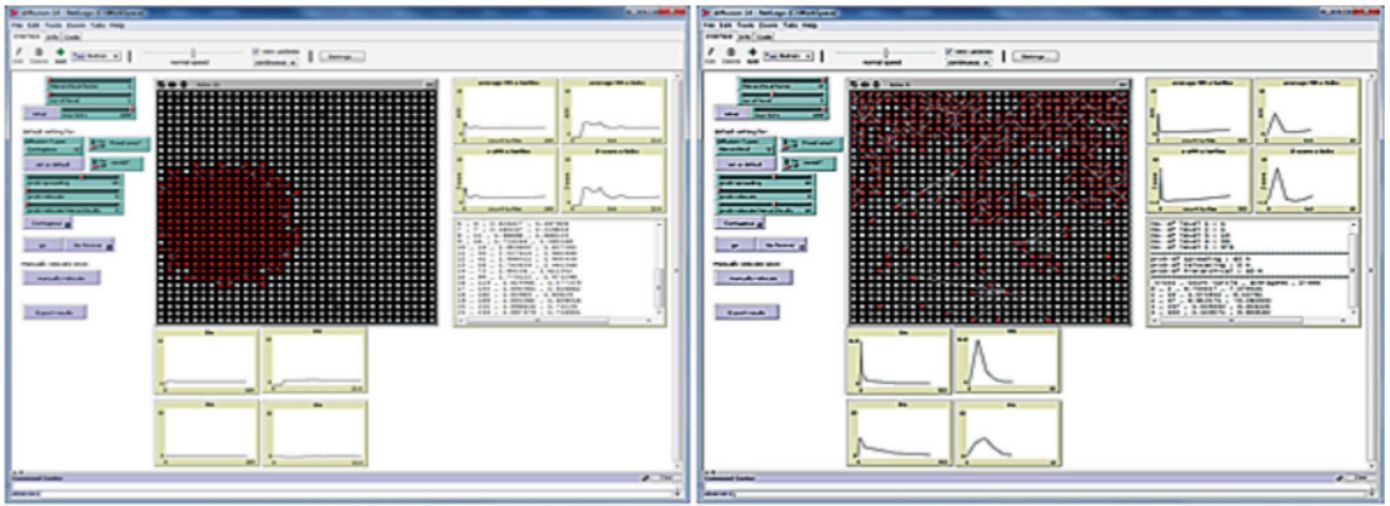
a. Expansion diffusion

b. Relocation diffusion

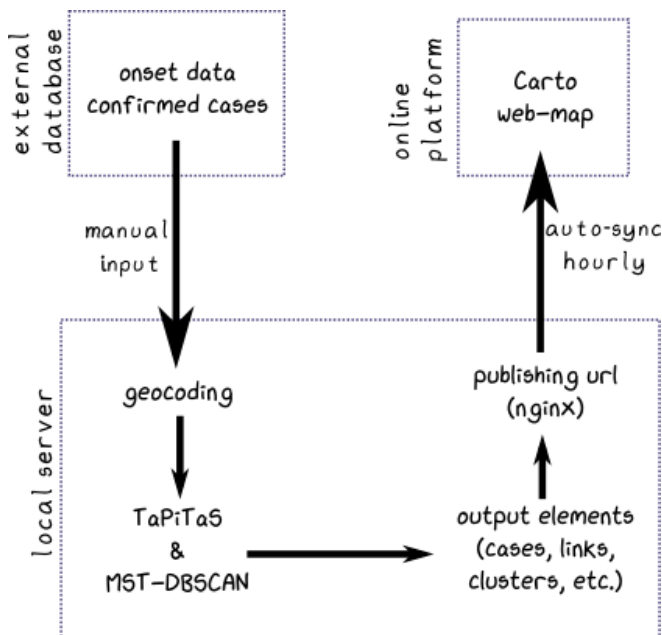Figure 24: The Netlogo app of the diffusion modeling.

Figure 25: The auto-mapping framework.

Website (in mandarin with step by step screen-shots):

https://wcchin_tutorial.gitlab.io/qgis_tutor/

Network analysis

Gephi learning website

Gephi is an open source software for network analysis and visualization. I had an opportunity to teach about using Gephi software in the department in year 2016. So I created this website for the class as a learning note, and for the students to check for the relative steps in the future when they needs the software. It is designed to be a self-learning resource for Gephi, so anyone who wants to learn the software by oneself can use the website directly.

Website (in mandarin with step by step screen-shots):

https://wcchin_tutorial.gitlab.io/gephi_tutor/

Python, open source

making map in python

colouringmap

Making maps in python is part of my life. I use python for daily works, including data visualizing, results of the algorithms or frameworks. So, for spatial data, it is natural to draw map for observation. But the visualization of maps in python is not convenient enough. Every time I map something, some common settings has to be done again for map, (e.g. equal aspect for the two dimensions, hiding the XY ticks, scale bar). In addition, most of the time I need to categorize a column of numbers into groups (based on equal interval, natural break, head-tail break, etc.). Therefore, I created the *colouringmap* package, that is simple to make map in python. I used this package to make almost all of the maps I produced in python, including some of the figures as shown above and in the journal articles. A demonstration of using *colouringmap* is shown in Fig. 32

Project page:

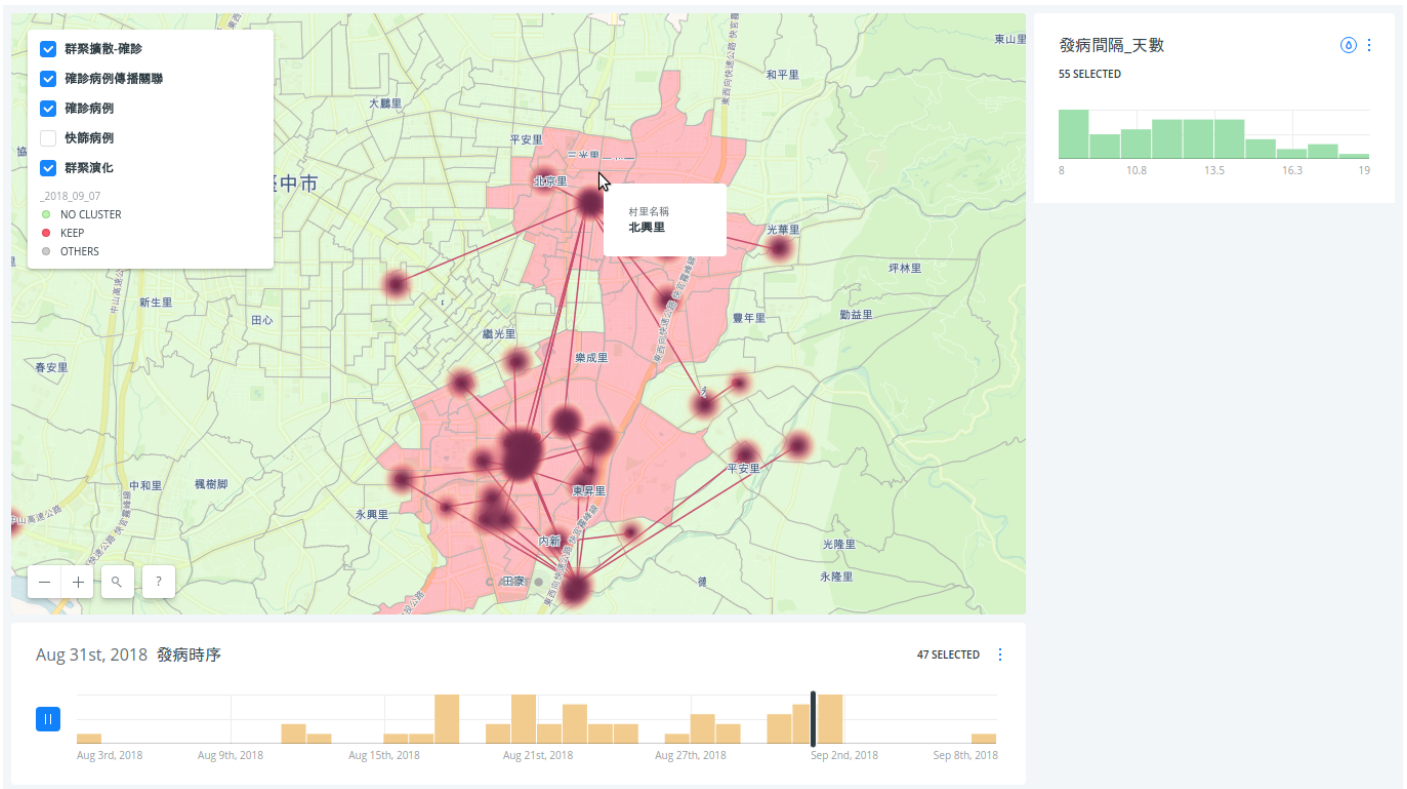https://wcchin.github.io/colouringmap.html

15

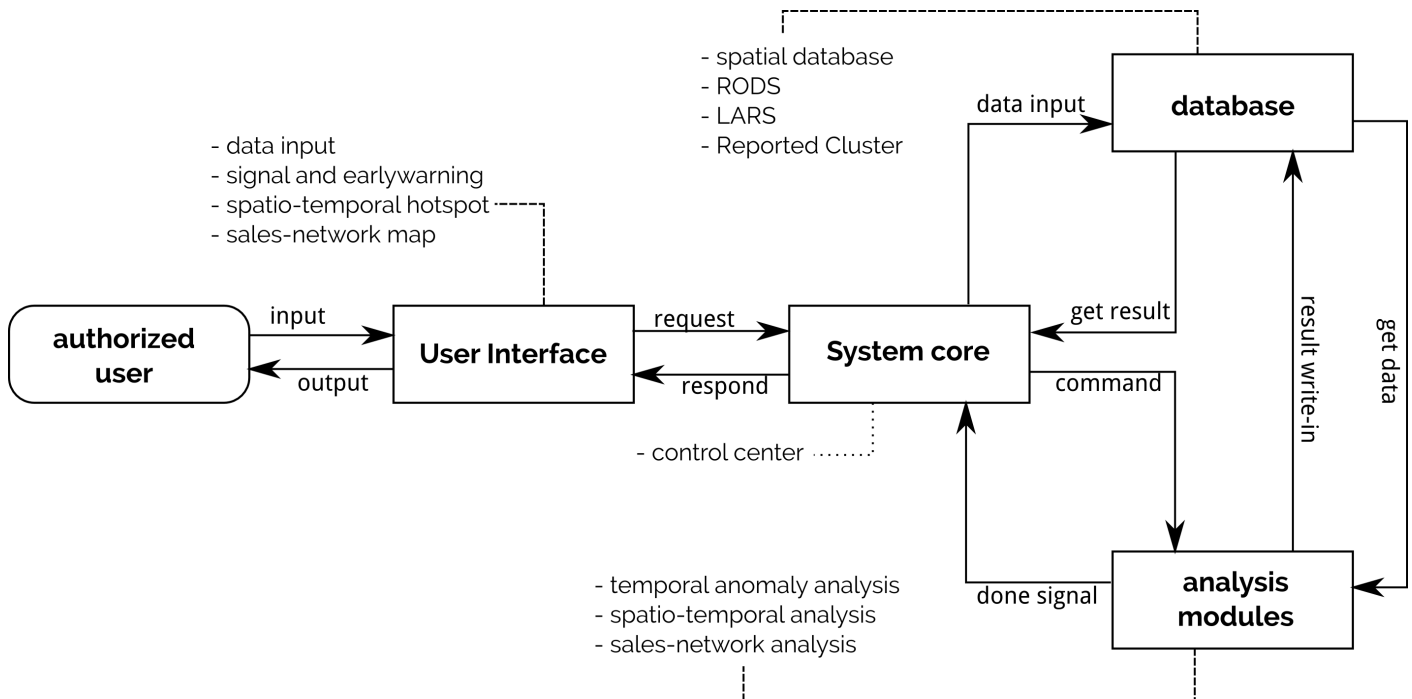Figure 26: A demonstration of the Carto platform of the map.

Figure 27: The framework of detecting possible acute diarrhea clusters from RODS and validate by LARS.

PYTHON, OPEN SOURCE

making SVG map

vmapper

Scalable Vector Grpahics (SVG) is a vector-based graphic written in XML format. SVG can be open in most modern web-browser. So it is good to be used for showing maps on web pages with the abil-

ity to zoom in and out, with no distortion that the raster image does. In addition, a SVG file is also portable and can be published as a static file on a static web-server (such as github pages).

These are the reason why I started to think that it is a good idea to make map with SVG. So, in python environment, I read the common shapefile
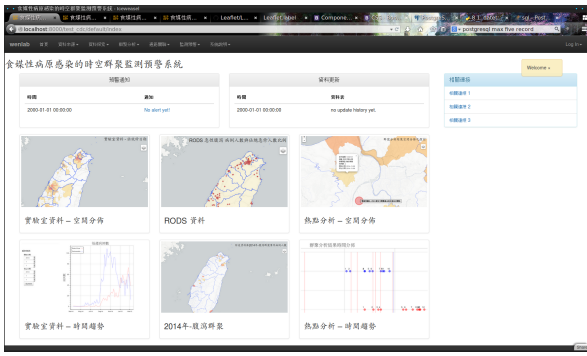
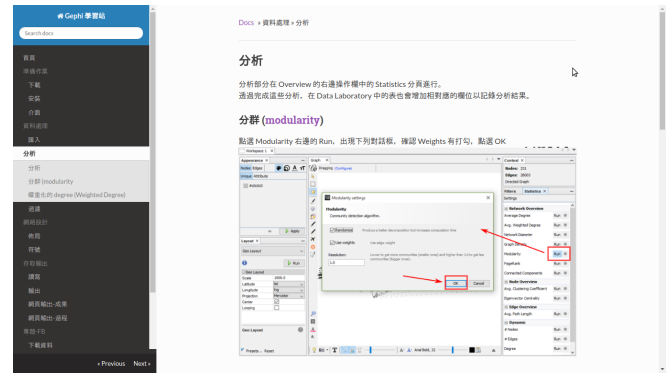Figure 28: The dashboard of the monitoring system prototype.

Figure 29: The map showing current possible clusters based on RODS database.
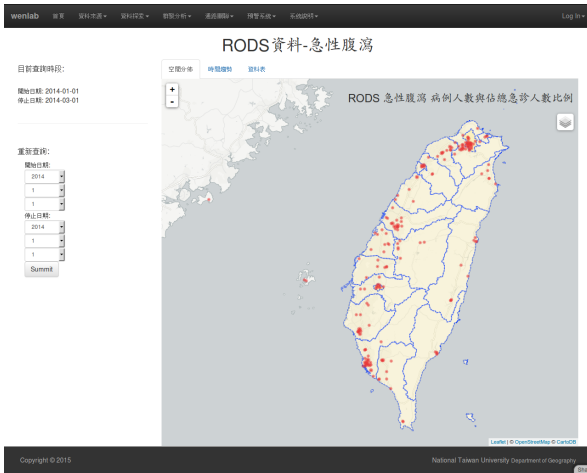

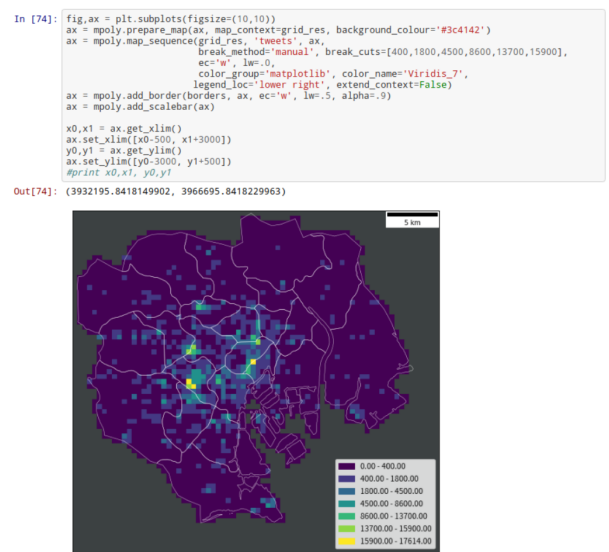Figure 30: A page of the tutorial website for QGIS.

Figure 31: A page of the tutorial website for Gephi.

Figure 32: A demonstration of using *colouringmap* to make a map of the distribution of tweets in Tokyo City.

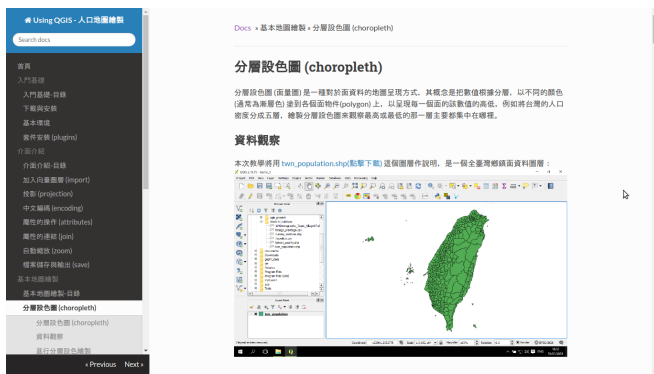with geopandas package and convert the geometry into SVG format, with some controls on the colors and cartography elements. Figure 33 shows a demonstration of a map created with *vmapper*, that is about the distribution of crime density in Taipei City. The colors were determined using the aforementioned colouringmap pacakge.

Project page:
https://wcchin.github.io/vmapper.html

PYTHON, OPEN SOURCE

making slides in markdown

pyreveal

pyreveal is a processor for converting markdown files into html using the Reveal.js html presentation framework (https://revealjs.com/#/). Reveal.js can produce very beautiful slides that is intuitive to be used in presentations and talks.

Although the Jupyter/nbconvert tools can convert the jupyter notebook into slides using the Reveal.js, but the converted slides cannot use a lot functions provided in Reveal.js. So, I jumped out of the Jupyter/nbconvert solution, and tried to create slides directly from markdown files, Reveal.js, and some other dependency library. And then *pyreveal* is created. The current status of the *pyreveal* has too many dependencies, so in the future I intended to reduce some of the dependencies. A demonstration is shown in Fig. 34.

Project page:
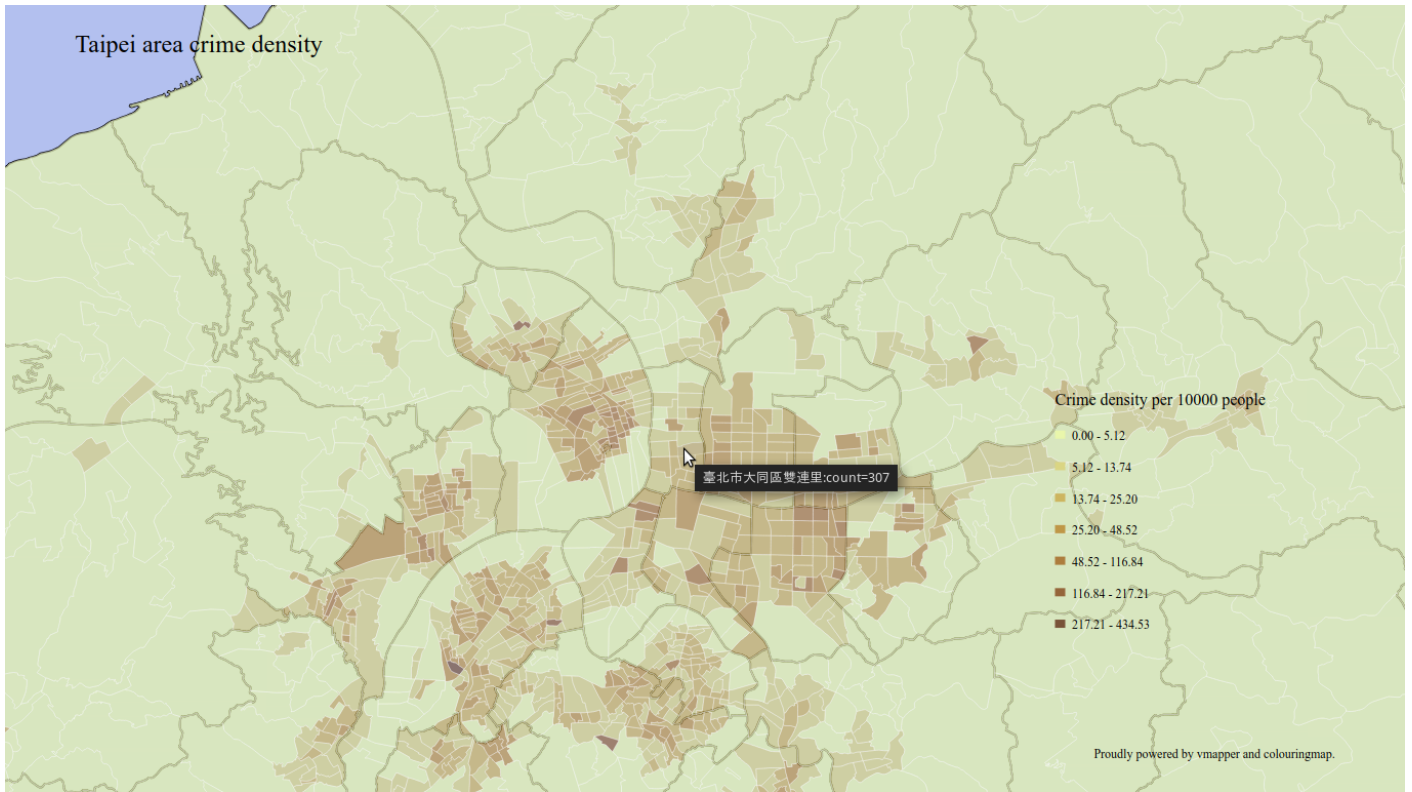https://wcchin.github.io/pyreveal.html

Figure 33: A SVG map that is created using *vmapper*, showing the distribution of crime density in Taipei City (the interactive map is available in `https://wcchin.github.io/svgs/testing_taipei_crime_map2.svg`).
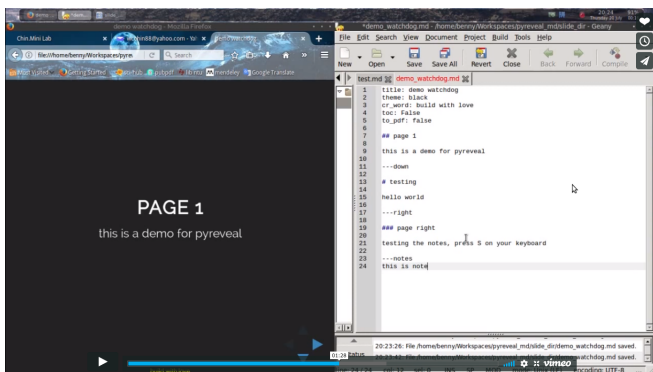

Figure 34: A demonstration of using *pyreveal* to make slides, and checking the output side-by-side (the video version is available on `https://vimeo.com/226295024`).