

Visual Analysis of Correlation Between Diseases Evolution and Human Dynamics

Lanyun Zhang[#], Hongyu Jiang^{*}, Weixin Zhao^{**}

[#] School of Information Engineering, Southwest University of Science and Technology, Mianyang, China

^{*} School of Computer and Information Sciences, Hosei University, Koganei Kajinocho, Tokyo, Japan

^{**} School of Computer Science and Technology, Southwest University of Science and Technology, Mianyang, China
E-mail: lanyun.zhang@outlook.com, hongyuj@outlook.com, weixin.zhao@outlook.com

Abstract— With the urbanization and the increasing pervasiveness of medical system, the systems utilizing digital technologies for their operation generates enormous amounts of digital traces capable of reflecting in real-time human behaviors and health situation in the city. It is not only transforming how we study the urbanization effect on disease's emergence and spread in cities but opens up new possibilities for tools that give people access to up-to-date information about urban dynamics the situation of diseases. Moreover, it was allowing us to make decisions that are more in sync with their environment. This paper introduces a prototype for exploring the dynamic of diseases-urbanization and supports urban planner meaningful access to large amounts of data capable of informing their decisions. We describe the technology context in this project, illustrate the requirements and the architecture of the platform to serve as a base for monitoring the health situation of the city. Finally, we shows the validity and practicability of the system by using real data in M city, China, which including electronic medical record, cellular network data, public transport, Census data.

Keywords— Visual Analysis, Healthcare, Urbanization, Correlation Analysis.

I. INTRODUCTION

Urbanization is a key driver of land-use change that is likely to increase at an unprecedented rate in the coming decades, particularly in developing countries, where as much of 90% of population growth is projected to occur in cities [1, 2]. Human population density and growth are significant predictors of historical events, and thus urbanization is likely to have a profound effect on public health as rural pathogens adapt to urban conditions, and other pathogens emerge (or re-emerge) in urban areas [3]. The human factors, such as population density, migration, trade, sanitation, and access to clean water can promote the transmission of pathogens and alter vector dynamics. While social factors that drive health inequality (socioeconomic status, housing, race, ethnicity, gender, and education) also influence the epidemiology of infectious disease in urban areas [4, 5]. For cities in developing countries, the epidemiological effects of these factors are often concentrated in informal settlements, where population growth and density is highest [5]. The urban computing emerge as the times require, urban computing is the process of collecting, integrating and analyzing massive heterogeneous data generated by various sources (such as sensors, equipment, vehicles, buildings, and humans) in urban space [6]. It is designed to address the main

problems in the city, such as air pollution, increased energy consumption, and traffic congestion. Urban computing can connect unobtrusive ubiquitous perceptual technologies, advanced data management, and analysis models, and cutting-edge visualization methods to create a win-win solution that improves the urban environment, human quality of life, and city operations. And it also helps us to understand the nature of urban phenomena and even predict the future of the city. From health-care aspect, hospital clinics and related institutions have a wealth of electronic medical records and diseases, and their diagnostic records can be used to describe their health status, the spread of specific diseases and the location of outbreaks. If the urban data combined with health-care data, we can even use the data to study how the environment influences the humans' health status and how the human behavior affects the spread of disease. For example, whether bad air quality makes more people's lung disease, or high frequency and extent of humans' movement makes more infection range. Besides, monitoring health situations and trends are the core mandates and underpin evidence-based decision making in every facet of health systems strengthening. It improve the quality, analysis, and use of country and regional health information for better measurement and accountability for health. This information is vital for monitoring and

evaluation of progress towards universal health coverage and achieving the health-related Sustainable Development Goals, national public health decision making, annual health sector reviews, planning and resource allocation and even individual health-care and service delivery. In this paper, though the urban and health-care data and urban computing technique, we focus on human activities as a driver of disease transmission and use it to explore how anthropogenic changes are driving interactions and the latent factor between disease emergence between people and urban.

Our work addresses the value of urban data for urban health situation assessment as well as to give urban planner meaningful access to large amounts of data capable of informing their decisions. We describe the technology context within which this project is framed, illustrate the requirements and the architecture of the multi-source data platform to serve as a base for monitoring the dynamics of city and health. And finally, we introduce and discuss the prototype of exploration platform by using real world data including the electronic medical record, cellular network data, public transport, Census data. This paper is structured as follows. Section 2 reviews related work. And we describe design considerations in section 3. After introducing the urban and health-care data and illustrate the spatiotemporal analysis for multi-source data in sections 5. We demonstrate the effect of human dynamic on the urban health situation in section 6. Finally, we evaluate our system in section 7 and discuss and the limitations and future of this work in section 8,9.

II. PAGE LAYOUT

An easy way to comply with the conference paper formatting requirements is to use this document as a template and simply type your text into it.

A. Urban data Visualization

LIVE Singapore! [7] project explores the development of an open platform for the collection, elaboration and distribution of a large and growing number of different kinds of real-time data that originate in a city. The platform is structured to become itself a tool for developer communities, allowing them to analyze data and write applications that create links between cities different real-time data streams, offering new insights and services to citizens. Urban Pulse [8] define the concept of an urban pulse which captures the spatiotemporal activity in a city across multiple temporal resolutions. The prominent pulses in a city are obtained using the topology of the data sets and are characterized as a set of beats. The beats are then used to analyze and compare different pulses. [9] focus on rapid urbanization (predominantly a feature of developing countries) as a driver of disease emergence and use it to explore how anthropogenic changes are driving interactions and the potential for disease emergence between humans and urban.

B. Human Mobility Visualization

Human activities have a great influence on disease emergence and transmission, the various data which related

movement of human can be used to obtain the human mobility, such as the data collected by wearable sensors, mobile phones can be used to describe the long-term movement of humans. Candia et al[10] study the human dynamic in both time and space two aspects from the mobile phone data. TelCoVis[11], an interactive visual analytics system, which helps analysts leverage their domain knowledge to gain insight into the co-occurrence in urban human mobility based on mobile data. Present a visual analysis system to predict non-visual city attributes. With the same data[12] combined call and geographical relationship to help the users to detect the real relationship of persons.

C. Healthcare Visualization

There are various works toward on the visualization of medical patient records, and they have been proposed, and new systems are likely to emerge as Electronic Health Record (EHR) is adopted widely. e.g., VISCARETRAILS [13] is a system based on timed word trees visualization that summarizes event paths relative to a given root event and is obtained through a simple drag-and-drop user interface. CareCruiser [14] is a system designed to visualize the effects of applying clinical treatment plans and to support the exploration of the effects that the treatments have on the patients condition. Zhang et al [15] propose a framework composed of a suite of cooperating visual information displays to represent the Five Ws and demonstrate its use within a health-care informatics application. TimeSpan [16], an exploratory visualization tool designed to gain a better understanding of the temporal aspects of the stroke treatment process. More recent work in this direction like the CDDV (Critical Care Patient Data Visualization) [17] in which health records are distinguished by their inherent aspects, such as problems, symptoms, tests/results, diagnosis, treatments and medications, etc. VisuExplore [18] and Miksch[19] both used color to indicate severity or type, and a level of detail mechanism allows one to zoom into patient records. Particularly interesting in this context is the Midgaard system [20] which makes practical use of illustrative abstractions to gradually transition between broad qualitative overviews of temporal data (for example, blood pressure) to complete, quantitative time signals. These techniques are further elaborated on by Aigner [21]. At present, there are little works focusing the correlation analysis between disease and urbanization, but there are many works on urban data or health data analysis individually, those works inspired us to design the visualization model and analysis method. In contrast, we are not focusing the visualization of the time-varying process of individual event. We aim to explore how effect can cause on disease's emergence and transmission by human activities in urban. Also, measuring the correlation among attributions of the urban and various disease condition. Some cases were used to describe how visual analytics techniques were used to help users explore the relationship between human behavior and emergence, the spread of disease in the complex urban environment.

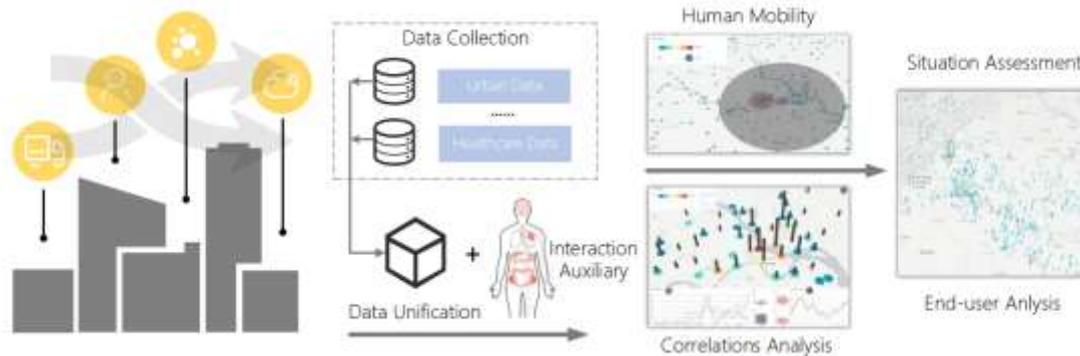


Fig .1. The analysis flow of this work, first we collect the data which produced in urban, various data was been unification for further processing. Plus the visual model, human dynamic and correlation analysis technologies, the users can obtain the health situation of the whole city.

III. DESIGN TASKS

The emergence and spread of disease in urban was being affected by human dynamic, there are a lot of potential relations linked the various attribution of urban to health conditions (such as disease emergence and transmission), at meantime, the operation of medical institutions becoming more and more modernized, EHRs are increasingly being used and more generated, collect and manage by new health information technology systems. Hence, the amount of clinical data which produced by the health-care system, Increased data can be used for managers and researchers which are never happened before. Besides, urban as a complex system running with large-scale residents and their activities, events and data are producing continuously. Researchers and analyzers are hard to recognize the relationship in urban.

While many people hope to make a better decision with more data scale and type, but the large-scale, high-dimension clinical data caused by multi-sources make us harder to analyze with the traditional method. Visual analysis is an emerging discipline that shows a significant ability to solving the challenge which caused by information overload. Also, it is a science to promote the detection and explanation process which combines the data mining, machine learning, human-computer interaction and the human cognition. Because health-related data continues growing with unprecedented rate, complicated relationships among heterogeneous and large data which collected by various information systems. The visual analytic technology should be used to avoid information overload and recognize deeper connection and pattern.

In this work, we build a framework of health-care data and urban data. With three cases at disease monitoring, disease tracking, and evolution presenting, we will demonstrate the effectiveness of our frame detailed. However, the high-dimension and complex relation features of multi-source data set make a hard analysis work for us, and we will solve the following issues:

- T1** Multi-source and high-dimension data unification
- T2** Correlation analysis between health-care and urban data
- T3** Support diseases prevention and control for related departments and public

In order to deal with the above issues, data fusion and spatiotemporal analysis method are introduced (T1). The dynamic time warping and geo-hash algorithm and some visualization models were designed to support correlation analysis of diseases and urbanization, discuss the effect of human dynamic on diseases spread and happening (T2). Final, we use map and chart to present the effeteness of human dynamic and the time/geo- distribution of various data (T3). Besides, in this work, the 3D visualization technology is introduced, it provides more display space for graph element.

IV. URBAN DATA MEETS DISEASES DATA

When urban computing meets the disease correlation analysis, more potential information can be detected. The human mobility data, vehicles' trajectories and POI data obtain a lot of valuable information about the spread of disease modeling if combined with the health data. When model the diseases spread, it should consider the factor of human behavior. The health data combined with vehicle data we will obtain the mobility pattern and propensity of urban hence the transportation is often used to support long-distance travel for the human. Movement pattern of bacterial carrier's study will make us realize the produce and spread process of diseases more globally. Besides, those various data can be used to obtain the global situation of the urban health situation.

A. Urban Data

Urban data can be classified as four class, 1.subject data produced by human(such as the geo-tag social media data, public security and service data.) 2. object data produced by human(such as traffic data, mobile data and health data.) 3. data generalized by human(such as the social network data) and 4.environment and POI data(such as the road network and building data). In this work, we are only focusing the data produced by human and environment, POI data. Those data can be categorized as fixed attributions of urban, data provided by fixed sensors and data generated by motion sensors(random location and timestamps).

Urban Attributions: Including the point of interest(POI), Census data which contains the buildings area, how many people living in there and so on, but without time information.

Fixed Sensors Data : The vehicles and peoples which collected by cameras, the call records which collected by communication base stations, long span and irregular capture cycle and fixed location make the steady characteristics of this data.

Sensors with irregular location : The source of data do not have a fixed location, and irregular capture cycle makes this data is difficult to process mostly, such as GPS data track the persons and GPS track the vehicle.

Different data contain various collection frequency makes our analysis task more hard to achieve. Hence the first step is to unite the time window to fusing the various type of data.

B. Healthcare Data

Electronic Health Records : Electronic Health Records (EHR) have become an effective instrument not only for medical diagnosis because it assists as support tool presenting and storing patient information, but also for the patient who benefits directly having the proper management of the disease. Benefits of the EHR over paper-based records including the creation of new medical knowledge. Due to structured EHR content, statistics can be obtained, and effective research can be done contribute to the improvement of international standards.

ICU Data : The intensive care unit (ICU) is a data-rich environment. By this definition, ICU patients have significant organ system disorder and have a little physiologic reserve. Optimal medical care requires intensive monitoring of organ function, frequent multimodal diagnostic testing, and many consultations with sub-speciality physicians and other health-care providers. In this kind of a critical care environment, it is often difficult to make rapid evaluations of a patient's condition because of them the overwhelming volume of data that is continuously generated [22]. Sources of patient-generated data include continuous automatic physiologic monitoring and intermittently determined data that is gathered by bedside care providers and from various diagnostic testing sources. In addition to patient-generated data, there are vast arrays of clinical data generated that document the treatment that is received by the patient. This data includes drug therapy, respiratory therapy, physical therapy and all other clinical interventions.

In this work, we interested in EHR data, and the following work will be discussed with it individually

C. Time Series Analysis

Timing data produced most frequently in the urban, such as sensor collection data, the human movement trajectory data. Moreover, an evolution processing of the disease can also be recorded via timing data. Hence, timing analysis technology is a powerful tool to obtain the potential relationship among various data. Also, those data produced by multi-source caused the different format and mismatch dimension. There is possible timing relationship among multi-source data. For example, we can obtain the relevance between pulmonary disease emergence and the volume of car's exhausted gas. Moreover, how does the water quality affect the infectious disease transmission? However, that relevance and effect may not be revealed immediately because of it is a cumulative process. These

characteristics make the timing data association analysis among urban data and health situation more difficult. For example, there is a new chemical plant was built in the 2016 year making the air quality decreased day by day. When it comes 2017 year, many residents were diagnosed with the pulmonary disease by the doctor. That is what we want to identify and analyze.

1) *Time Window Integration:* Difference source of data with different collection cycle, which caused different time windows in the time dimension. For example, to unite the time window of records at air quality monitor station and checkpoint records at vehicle checkpoint which captures data when vehicles passed, former has regular collection cycle while later has irregular collection cycle. For conduct associated analysis, different time window must be unified. Hence the data with smaller time window should be grouped to rise time window.

2) *Correlation Analysis:* Typical time-series techniques usually apply algorithms, such as SVM, logistic regressions, decision trees, after transforming away temporal dynamics through feature engineering. By creating features, we often remove any underlying temporal information, resulting in a loss of predictive power.

This work, we use the dynamic time warping (DTW) [23] to measure the correlation between different data, which is a useful distance-like similarity measure that allows comparisons of two time-series sequences with varying lengths and speeds. Simple examples include detection of people 'walking' via wearable devices, arrhythmia in ECG, and speech recognition. Notice that while DTW is a distance-like quantity since the triangle inequality doesn't hold, it is not a metric.

This measure distinguishes the underlying pattern rather than looking for an exact match in the raw time-series. As its name suggestions, the usual Euclidean distance in problems is replaced with a dynamically adjusted metric. DTW allows us to retain the temporal dynamics by directly modeling the time-series.

Let X and Y be two series with the length of n and m , respectively, and an $n \times m$ matrix M can be defined to represent the point-to-point correspondence relationship between X and Y , where the element M_{ij} indicates the distance between x_i and y_j . Then the point-to-point alignment and matching relationship between X and Y can be represented by a time warping path W , where the element w_k indicates the alignment and matching relationship between x_i and y_j . If a path is the lowest cost path between two series, the corresponding dynamic time warping distance is required to meet:

$$DTW(X, Y) = \min_W \left\{ \sum_{k=1}^K d_k, W = \langle w_1, w_2, \dots, w_k \rangle \right\} \quad (1)$$

where $d_k = d(x_i, y_j)$ indicates the distance represented as $w_k = (i, j)$ on the path W .

D. Spatial Analysis

According to the spatial distribution of the infectious

disease emergency locations, combined with the characteristics of the residents and the attributes of each field. The rough number of infected persons and the disease spread range can be predicted. Besides, with the characteristics of the residents and disease's pathological features, the reason of disease's emergence can be surmised.

The spatial dimension of various data may have different encoding methods. e.g., latitude and longitude encoding method often used in sensors data while census data tend to use area code to encode geographic information. But, different geographic encoding method caused hard to link the geographic information among various data. Hence, unifying the geographic dimension of varying data is essential, the latitude and longitude encode method is grouped by point, the data which encoded by area code and address is group by region.

Geo-hash[24] is a geo-coding system invented by Gustavo Niemeyer and placed into the public domain. It is a hierarchical spatial data structure which subdivides space into buckets of grid shape with space-filling curves. The scale of the grid can be defined by the user. For point grouped data, we can use it to allocate the points to the grids directly. As to the region grouped data, we use the buckets of grid shape to split area to grid. If the area of the grid is less than one grid, it will be removed, otherwise, extended to a complete grid which illustrated in figure 2. Final, assign the result to the processed grid with the average weight (shown in figure 3). By this way, the population data with uneven distribution can be uniformly distributed in the whole city, and each grid will be allocated the same number of residents. Depending on the location and density of the patient at different places of the city, the disease emergence's centre can be estimated. Moreover, combined with various urban data, to recognize the relationship between urban and epidemic events. For example, with air quality data to determine whether the air quality affected the lung disease of person, and find the relationship between infectious disease and population density.

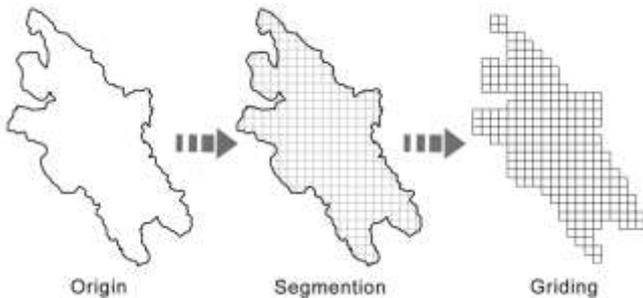


Fig .2. The gridding process of a region of a city, it is divided into two steps.

1) *Modeling data as scalar functions over time*: A scalar function, $f : \mathbf{D} \rightarrow \mathbf{R}$, maps points in a spatial domain \mathbf{D} to real values. In this work, we are interested in the spatial region corresponding to a city, which is represented by a planar domain. A function value is defined over each point in this planar domain with the goal to capture the activity at

that location corresponding to a given data set. An example of a scalar function that is used in this work is the density function. Assume that the input data is provided as a set of points (data points) having location and time. The density function at a given location p is defined as the Gaussian weighted sum:

$$f(p) = \sum_{x_i \in N(p)} e^{-\frac{d(p,x_i)^2}{\varepsilon^2}} \quad (2)$$

The data points x_i are the points in p 's neighborhood $N(p)$. Here, d is the Euclidean distance between two points, and ε is the extent of the influence region for a given data point. The neighborhood $N(p)$ is defined as a circular region centered at p . Intuitively, the density function captures the level of heat of particular or global disease over different locations in a city. For example, consider data corresponding to pulmonary disease. Each data point corresponds to a case and provides the place where the patient was living together with the time when he gets it. High function value at a given location implies a lot of activity (there are many residents get this disease), the result divided by population indicating the prevalence of pulmonary disease in that location.

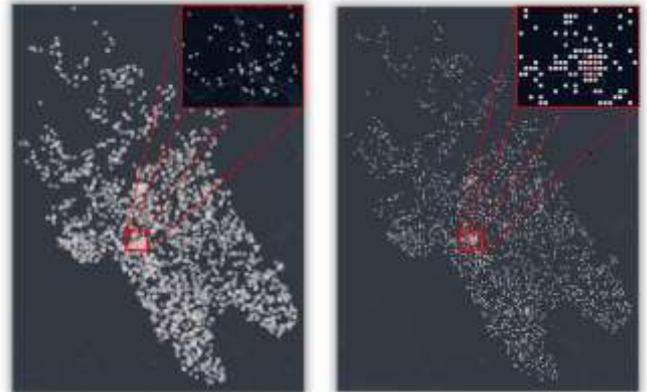


Fig .3. The comparison of original(Left) and geo-hashed results(Right) for population distribution in M city.

In order to efficiently compute the topological features of the scalar function f , it is represented as piece-wise linear (PL) function $f : \mathbf{K} \rightarrow \mathbf{R}$. The planar domain \mathbf{R} of the function is represented by a rectangle mesh \mathbf{K} . It is defined on the vertices of the mesh and linearly interpolated within each triangle.

The set of data points are grouped into a discrete set of time steps corresponding to the temporal resolution if we are taking into account the variation with time, and the scalar function is computed for each of these time steps.

V. HUMAN DYNAMIC EFFECT ON DISEASE DYNAMICS

The mechanisms of disease transmission and spread are usually complex and possibly involve social, economic and psychological factors in addition to the fundamental disease biology and ecology. In particular, the human behavior could have a significant influence on disease

transmission and vice versa[25]. For example, individuals avoid close contact with obviously sick persons to protect themselves, and therefore the frequency and strength of connections between uninfected and infected people are reduced. In case of severe disease outbreaks, people will attempt to change their routine schedules (including, but not limited to, work, recreation, and travel), to minimize their risk of infection. Nowadays, the fast growth of information technology allows timely and up-to-date reports on the details of disease outbreaks from the internet (especially those favorite social networking ~sites), newspaper, television, radio stations, and government announcements. Consequently, these media coverage and health education will, to a wide extent, affect human behavior which can lead to a significant reduction in outbreak morbidity and mortality. It is clear that human behavior could play a significant role in shaping the complex epidemic and endemic pat- tern of a disease. i.e. if we present the disease situation such as dangerous area and the ways infectious diseases spread, we can reduce the infection probability activity. Hence we aim to obtain the high-risk regions and periods, control the infection probability lowest.

A. Human Dynamic

To describe the influence of human behavior on disease dynamics, model the humans' dynamic is necessary. At the large scale, when the behavior is modeled over a period of relatively long duration (e.g. more than one day), human mobility can be described by three major components:

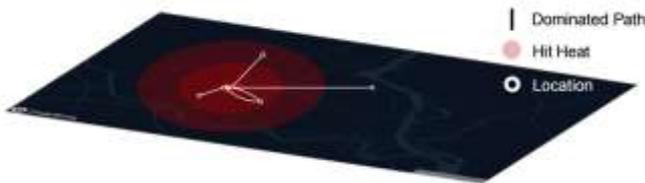


Fig. 4. Presenting the host movement influence range and possible of disease spread, the black circles refers as the radius of gyration, the red lines denote the dominant paths of individuals, and the black points indicate the locations which has high frequency visited.

- Trip distance distribution P_r ;
- Radius of gyration $r_g(t)$;
- Number of visited locations $S(t)$.

High reproducibility and predictability are measured in human movement. Brockmann [26], by analyzing bank notes, found that the probability of travel distance follows a scale-free random walk known as Levy flight of form $P_r \propto r^{-(1+\beta)}$ where $\beta = 0.6$. It was later confirmed by two studies that used cell phone data[27] and GPS data to track users [28]. The implication of this model is that, as opposed to other more traditional forms of random walks such as Brownian motion, personal trips tend to be of mostly short distances with a few long-distance ones. In the Brownian motion, the distribution of trip distances is governed by a bell-shaped curve, which means that the next trip is a roughly predictable size, the average, wherein

Levy flight it might be an order of magnitude larger than the average.

The radius of gyration is used to capture the travel distance of individual, and it indicates the characteristic distance traveled by a person during a period t [27]. Each user, within his radius of gyration $r_g(t)$, will choose his trip distance according to P_r . The $S(t)$ models the fact that humans tend to visit some locations more often than what would have happened under a random scenario. For example, home or workplace or favorite restaurants are visited much more than many other places in a user's radius of gyration. It has been discovered that $S(t) \propto t^\mu$ where $\mu = 0.6$, which indicates a sub-linear growth in a different number of places visited by an individual.

These three measures capture the fact that most trips happen between a limited number of locations, with less frequent travels to places outside of an individual's radius of gyration. By measuring the entropy of each person's movement, it has been shown [29] that there is a 93% potential predictability. It means that although there is a significant variance in the type of users and the distances that each of them travels, the overall characteristic of them is highly predictable. The implication of it is that in principle, it is possible to accurately model the processes that are dependent on human mobility patterns, such as disease or mobile phone virus spreading patterns. Network topology, traffic structure, and individual mobility patterns are all essential for accurate predictions of disease spreading [30]. Some studies have shown that a smaller spatial scale, the regularity of human movement patterns and its temporal structure should be taken into account in models of infectious disease spread [31].

B. Mobility Visualization:

Humans' mobility can be visualized to help the users understanding the effect of individual's mobility to disease's transmission and spread. Hence we use three visual elements to present the infectious disease's active range of carriers, dominant path, and frequently-visited locations, as shown in Fig 4. The radius of circle refers to the field of influence. Also, the darker degree of color denotes the infection probability which indicated by the radius of gyration and trip distance distribution. The white lines refer to the dominant path, the place which passed is the dangerous region. However, the most critical area is the frequency- visited locations, such as home, school and workplace. The three elements were presented in a map at same time aim to describe the transmission ability of the individual.

C. Diseases Monitoring and Pre-warning

Through the understanding of the characteristics of infectious diseases and its influencing factors, establish and improve the pre- diction and early warning techniques and methods. Detect the occurrence and development of infectious disease timely to obtain the abnormal trend, can

help the health department to take scientific measures to prevent and reduce the harm immediately and improve their ability to deal with the epidemic infectious diseases. Moreover, scientific and comprehensive early warning will help the public to deal with the infectious diseases correctly, appropriate to take self-protection measures and consciously co-operate with professional institutions to conduct prevention and control work.

Systemic monitoring for disease began at the end of the last century at the US Centers for Disease Control. After that, monitoring was extensively appeared in many countries, from the observe the dynamic spread of infectious diseases to non-infection diseases. Also, the focus is progressively shifted from a simple bio-medical issue to a biologically-psychologically aspect. Disease monitoring means to continuous and systematic collect the dynamic distribution and influential factor in a long-term, the analysis result will be reported and feedback to take timely action to intervention and evaluate its effectiveness. Disease surveillance is an essential method for prevention and control of disease it includes information collection, analysis, feedback, and information use. Moreover, data obtained from monitoring can be used to understand the situation of disease distribution, predict the epidemic, evaluate the effect of an intervention and confirm the primary health issue, support theoretical foundations for drawing up the strategy and methods for disease prevention and control.

The early-warning system is vital for disease prevention. To build an early-warning system is a complex system engineering. Theoretically, it is mainly composed of risk information system; early warning evaluates index system; early warning detection system; alarm system and early warning preventive measures. However, the basis for an early warning system's building is a complete disease monitoring architecture, and we can identify the early warning indicators which affect the spread of infectious disease and then to monitor these indicators to conduct the early warning. Therefore, the modernized early warning system must be a composed of complete disease monitoring network system, professional team network management, and computer network technology.

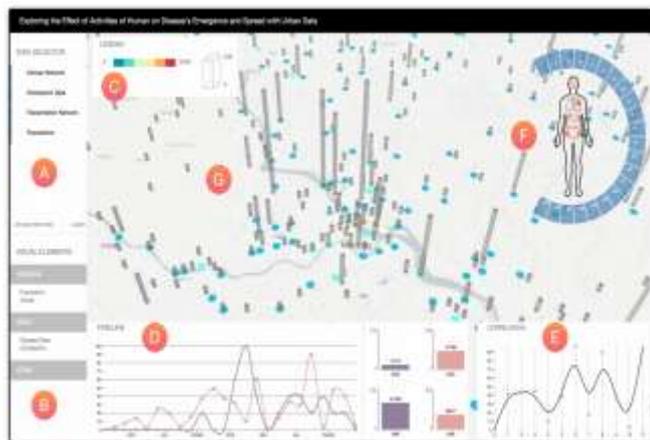


Fig .5. The DMUD system, which mainly comprised of four parts.

VI. CASE STUDY

In this work, we design a visual analytic system to achieve the goal demonstrated in the previous section, this system is an open platform and mainly composed of four parts which shown as figure 5, the A, B component shows the data name and visual channel respectively. The centre of our system is the map part including C, G and F component. The map-view is core part of our system. It presenting mostly visual elements, C is a legend window describe the colors and glyph scheme, and F support a natural and convenient way for interaction with the users, the users can click an organ of human in model to highlight the corresponding diseases' class. In the bottom of the system are D and E components describe the timeline of each selected data and their correlation level during the whole time. It aims to analyze the potential relations between urbanization and disease dynamic. To shows the effectiveness of our method, the Census data, one-year call records, roads checkpoints data and EHRs of M city which have five hundred thousand residents was used.

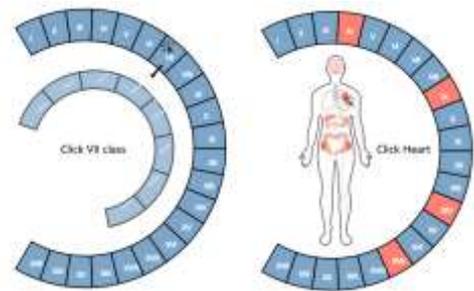


Fig .6. ICD code is organize as a hierarchical structure. Hence, we design a ring to visualize the elements information of each level, the users can click the parents nodes to unfold the children nodes of corresponding class.

A. Diseases Hierarchy Illustration

In the current medical domain, the disease organization architecture is a tree structure. Hence there is an international code ICD-10 [32] which aim to describe the class of each disease. The ICD-10 is the 10th revision of the International Statistical Classification of Diseases and Related Health Problems (ICD), a medical classification list by the World Health Organization (WHO). It contains codes for diseases, signs and symptoms, abnormal findings, complaints, social circumstances, and external causes of injury or diseases. To presenting the structure of disease, we use a tree visualization method to illustrate it. For each symptom or diagnosis the patient has, we always have a corresponding ICD-10 code, the user clicks parent nodes, the children nodes will unfold and replace parent nodes (figure 6). Although it is friendly and convenient to domain expert (doctors and analyzer), it is not suited for people who study other disciplines. Hence, we present an anatomical map in the centre of the radial chart, to enhance universality of the radial chart, the anatomical was designed as the users click the organs will highlight the related class of ICD-10.

B. Disease Monitoring

Individual's trajectories can be collected and recognized

via mobile phone, from the human mobile data we can know a city's daily lifestyle and the distribution of functional area. If the mobile data plus the disease situation, we can extract from the infection path and area range of infectious diseases. Through the pattern recognition for movement and then make a predicted result, we can obtain the high-risk regions with the predicted result. On the other hand, we can trace the process of disease production and source the origin of it. Besides, if we combine the population with the above data, can also estimate the scale of infected populations and enhance our disease spread model.

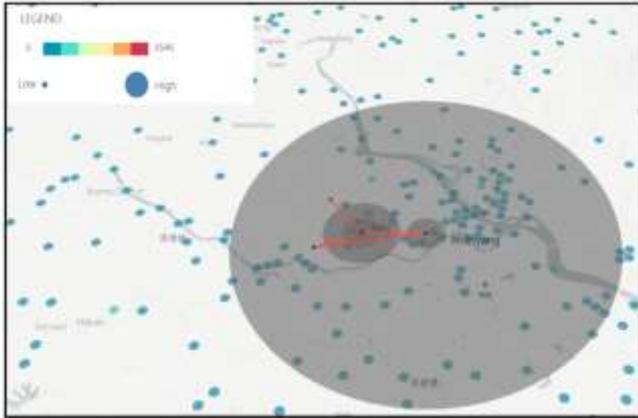


Fig. 7. A case of a host's influence range who carrying the flu virus through mobile phone and EHRs.

From call records we can get the sparse trajectories of individuals, their traces are producing with the low sampling rate, and irregular sampling cycle made the analysis work more difficult. However, the rough location of mobile trajectories of individuals can be extracted. Besides, if the sampling period is long enough, we can detect the dominant path and the frequency region of the individual, dominant paths are the way frequently visited, such as the route from home to the workplace. With above information, the movement pattern of each mobile user can be modelling. Also, with the individual mobility theory, we can estimate the distribution of trip distance P_r ; the radius of gyration $r_g(t)$ and number of visited locations $S(t)$. Those characters can be used to illustrate the influence area range of the infectious disease's carrier, and describe the dangerous trend of the whole city indirectly. As shown in figure 7, we illustrate the influence of the human movement. The white circle refers to the radius of gyration of individual, more opaque area means higher possible to visit. Also, the red lines denote the dominant paths of individuals, while the black point implies the frequency visited locations. Combined with the population density, we can pre-warning the highest dangerous region for prevention, if the samples are large enough, give your GPS trajectory during one week we can compute the safe level of individuals.

C. Disease Tracing

Shang et al. [33] used the GPS trajectory sample of the vehicles (such as taxis and buses) to detect the gas

consumption and pollutant emissions of vehicles travelling on the city's road network. Based on this idea, if we plus the information about the factory which has a large volume of gas emissions to gas consumption, it is possible to obtain the quality of air real-time and detect the health situation of residents and the whole city. These are the case to identify the relationship between the urban condition and health status via visual analysis technology.

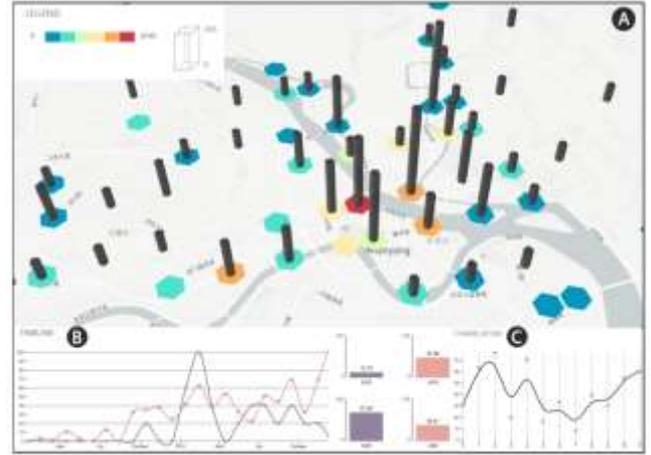


Fig. 8. Illustration the comparison of city attributions and pneumonia emergence. The A presenting the geographic correlation, the hexagon refer as the heat of vehicles passing while the B,C shows the timing line and the Pearson correlation of two timing data

Tracking the disease and explore the relationship between urban status and disease emergence is important for the prevention and control of diseases. As discussed previously, the gas and pollution emission can be estimated by car's stay time and activeness in the road network through the visual analysis technology. Hence, we aim to detect the relationship between gas emission and the heat of lung disease emergence. Those stations are mainly located on the primary road network and seldom located on branch roads.

D. Disease Evolution Backtracking

Present the evolution process of infectious diseases will help researchers and public understand the feature of infectious diseases. Moreover, playing an infectious disease evolution animation will let city's residents aware of the diseases are so close to them, called the public for more self-protection ability and the awareness of the friendly environment, it will produce a positive impact in society. Besides, it will let public learn the mechanism of different infectious disease, reduce the fear of infectious disease after learning it.

In this work we group the same illness and grids which generated by geo-hash, then segment it with the time window, map the count value in the map by the height of bars, then smoothly change the height of bars in chronological order, we can watch the trend of disease emergency. Shown as the figure 9, the infectious disease dysentery was emergence at the centre of city and spreading from west to east, using animation will illustrate the evolution of spread of the infectious disease.

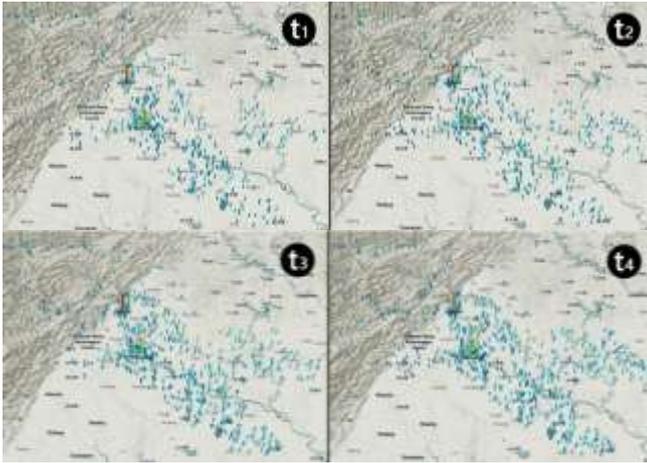


Fig .9. The infectious disease dysentery was emergence at the centre of urban and spreading from west to east, this figure shows the snapshots at specially periods.

VII. EVALUATION

To evaluate the usability and efficacy of our prototype we invited three physicians (all were residents), two infectious disease professionals and three college student to participate in a pilot user study. None had previous experience with our system. After we introduced the guide for them, they had ten minutes to learn how to use it. Then, we explained the idea behind each layout and the basic functionalities, including the data upload, dimension selection, visual style selection and the various interaction modes, especially the interactive method in the anatomical map, etc. We prepared two sets of questions. The first set aimed at finding out whether our system can help physicians to quickly and accurately find information. For the anatomical map evaluation. At the first round, the interactive speed of students is slower than physicians, but when it comes third round, we found that student beyond three physicians. That is because of the professionals tend to use the code to find disease, while the student tends to use the anatomical map to find focused information. The second set was more focused on design details along with some general questions. We support three types of datasets to let the users find the correlation among different data with our system, with the creative thinking of students, they show better results than professionals.

VIII. DISCUSSION

Presently, our system enables doctors and ordinary users to explore the effect of human dynamic to disease emergence and spread, and finally help users with situation assessment. On the one hand, our work aims to help urban planner to improve the city. On the other hand, our system was also designed to call on people to protect the environment and used the graph and animation to shows the relationship among human and nature, made public learning about more knowledge about disease prevention and control.

This work only discusses the combination of urban-produced and health data some cases were reviewed and

experimented by us. However, the application of joint research far more than that, for example, we also can explain the relationship between the city's mobility, how long they stay at home and the town the number of newborns. As a result, the combination of data will create more value for the health-care and urban.

IX. CONCLUSIONS

This work first discusses which effect can cause on disease's emergence and transmission by human activities in urban. Then proposed a method to measure the correlation among attributions of an urban and various disease condition. Final provided a visual analytic framework for exploring the relationship between human dynamic and emergence, the spread of disease in the complex urban environment. Besides, we hope to use the result to call for people's care for environment protection. Final, we illustrate the effectiveness of our frame through several analysis cases.

ACKNOWLEDGMENT

The research of authors is partially supported by National key research and development program (2016QY04W0801), the Fund of Fundamental Sichuan Civil-military Integration Institute (No.300015) and the Fund of Sichuan Provincial Science and Technology Department (2017TJPT0200, 2017KZ0023, 2017GZ0186).

REFERENCES

- [1] S. Knobler, A. A. Mahmoud, S. M. Lemon, et al., The impact of globalization on infectious disease emergence and control: exploring the consequences and opportunities: workshop summary, Joseph Henry Pr, 2006.
- [2] U. Nations, World urbanization prospects: The 2014 revision, highlights. department of economic and social affairs, Population Division, United Nations.
- [3] K. E. Jones, N. G. Patel, M. A. Levy, A. Storeygard, D. Balk, J. L. Gittleman, P. Daszak, Global trends in emerging infectious diseases, *Nature* 451 (7181) (2008) 990–993.
- [4] C. A. Bradley, S. Altizer, Urbanization and the ecology of wildlife diseases, *Trends in ecology & evolution* 22 (2) (2007) 95–102.
- [5] E. Alirol, L. Getaz, B. Stoll, F. Chappuis, L. Loutan, Urbanisation and infectious diseases in a globalised world, *The Lancet infectious diseases* 11 (2) (2011) 131–141.
- [6] Y. Zheng, L. Capra, O. Wolfson, H. Yang, Urban computing: concepts, methodologies, and applications, *ACM Transactions on Intelligent Systems and Technology (TIST)* 5 (3) (2014) 38.
- [7] K. Kloeckl, O. Senn, C. Ratti, Enabling the real-time city: Live singapore!, *Journal of Urban Technology* 19 (2) (2012) 89–112.
- [8] F. Miranda, H. Doraiswamy, M. Lage, K. Zhao, B. Goncalves, L. Wilson, M. Hsieh, C. T. Silva, Urban pulse: Capturing the rhythm of cities, *IEEE transactions on visualization and computer graphics* 23 (1) (2017) 791–800.
- [9] J. M. Hassell, M. Begon, M. J. Ward, E. M. Fe`vre, Urbanization and disease emergence: Dynamics at the wildlife–livestock–human interface, *Trends in ecology & evolution* 32 (1) (2017) 55–67.
- [10] J. Candia, M. C. Gonzalez, P. Wang, T. Schoenharl, G. Madey, A.-L. Barabási, Uncovering individual and collective human dynamics from mobile phone records, *Journal of physics A: mathematical and theoretical* 41 (22) (2008) 224015.
- [11] W. Wu, J. Xu, H. Zeng, Y. Zheng, H. Qu, B. Ni, M. Yuan, L. M. Ni, Telcovis: Visual exploration of co-occurrence in urban human mobility based on telco data, *IEEE transactions on visualization and computer graphics* 22 (1) (2016) 935–944.
- [12] H. Jiang, Y. Wu, Y. Zhang, S. Wang, Y. Zhang, From social community to spatio-temporal information: A new method for

- mobile data exploration, *Journal of Visual Languages & Computing* 41 (2017) 1–11.
- [13] L. Lins, M. Heilbrun, J. Freire, C. Silva, Viscaretrails: Visualizing trails in the electronic health record with timed word trees, a pancreas cancer use case, in: *Proc. IEEE Visual Analytics in Health Care (VAHC) Workshop*, 2011.
- [14] T. Gschwandtner, W. Aigner, K. Kaiser, S. Miksch, A. Seyfang, Carecruiser: exploring and visualizing plans, events, and effects interactively, in: *Pacific Visualization Symposium (PacificVis)*, 2011 IEEE, IEEE, 2011, pp. 43–50.
- [15] Z. Zhang, B. Wang, F. Ahmed, I. Ramakrishnan, R. Zhao, Viccellio, K. Mueller, The five Ws for information visualization with application to healthcare informatics, *IEEE transactions on visualization and computer graphics* 19 (11) (2013) 1895–1910.
- [16] M. H. Loorak, C. Perin, N. Kamal, M. Hill, S. Carpendale, Timespan: Using visualization to explore temporal multi-dimensional data of stroke patients, *IEEE transactions on visualization and computer graphics* 22 (1) (2016) 409–418.
- [17] A. Faiola, S. Hillier, Multivariate relational visualization of complex clinical datasets in a critical care setting: A data visualization interactive prototype, in: *Information Visualization, 2006. IV 2006. Tenth International Conference on*, IEEE, 2006, pp. 460–468.
- [18] A. Rind, S. Miksch, W. Aigner, T. Turic, M. Pohl, Visuexplore: Gaining new medical insights from visual exploration, in: *International Workshop on Interactive Systems in Healthcare*, 2010.
- [19] R. Kosara, S. Miksch, Visualization techniques for time-oriented, skeletal plans in medical therapy planning, *Artificial Intelligence in Medicine* (1999) 291–300.
- [20] R. Bade, S. Schlechtweg, S. Miksch, Connecting time-oriented data and information to a coherent interactive visualization, in: *Proceedings of the SIGCHI conference on Human factors in computing systems*, ACM, 2004, pp. 105–112.
- [21] W. Aigner, S. Miksch, W. Müller, H. Schumann, C. Tominski, Visual methods for analyzing time-oriented data, *IEEE transactions on visualization and computer graphics* 14 (1) (2008) 47–60.
- [22] W. Horn, C. Popow, L. Unterasinger, et al., Support for fast comprehension of ICU data: Visualization using metaphor graphics, *Methods Archive* 40 (5) (2001) 421–424.
- [23] S. Salvador, P. Chan, Toward accurate dynamic time warping in linear time and space, *Intelligent Data Analysis* 11 (5) (2007) 561–580.
- [24] G. Niemeyer, Geohash, <http://geohash.org>, accessed: 2017-07-22 (2008).
- [25] X. Wang, D. Gao, J. Wang, Influence of human behavior on cholera dynamics, *Mathematical biosciences* 267 (2015) 41–52.
- [26] D. Brockmann, L. Hufnagel, T. Geisel, The scaling laws of human travel, *Nature* 439 (7075) (2006) 462.
- [27] M. C. Gonzalez, C. A. Hidalgo, A.-L. Barabasi, Understanding individual human mobility patterns, *arXiv preprint arXiv:0806.1256*.
- [28] L. Pappalardo, S. Rinzivillo, Z. Qu, D. Pedreschi, F. Giannotti, Understanding the patterns of car travel, *The European Physical Journal Special Topics* 215 (1) (2013) 61–73.
- [29] C. Song, Z. Qu, N. Blumm, A.-L. Barabási, Limits of predictability in human mobility, *Science* 327 (5968) (2010) 1018–1021.
- [30] C. Nicolaides, L. Cueto-Felgueroso, M. C. Gonzalez, R. Juanes, A metric of influential spreading during contagion dynamics through the air transportation network, *PloS one* 7 (7) (2012) e40961.
- [31] V. Belik, T. Geisel, D. Brockmann, Natural human mobility patterns and spatial spread of infectious diseases, *Physical Review X* 1 (1) (2011) 011001.
- [32] W. H. Organization, The ICD-10 classification of mental and behavioural disorders: clinical descriptions and diagnostic guidelines, Vol. 1, World Health Organization, 1992.
- [33] J. Shang, Y. Zheng, W. Tong, E. Chang, Y. Yu, Inferring gas consumption and pollution emission of vehicles throughout a city, in: *Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining*, ACM, 2014, pp. 1027–1036.