



AI for Global Disease Surveillance

Hsinchun Chen and Daniel Zeng, *University of Arizona*

In this time of increasing concern over the deadly and costly threats of infectious diseases, preparation for, early detection of, and timely response to emerging infectious diseases and epidemic outbreaks are a key public-health priority and are driving an emerging field of multidisciplinary research. Recent health threats to large populations around the world include the Severe Acute Respiratory Syndrome (SARS) epidemic in Asia; the outbreak of avian flu in East Asian countries; this year's global swine flu (H1N1); the catastrophic aftereffects of Hurricane Katrina in New Orleans; and, since the anthrax attacks in the US in October 2001, the ever-pending threat of bioterrorism.

Disease surveillance has been practiced for decades and continues to be an indispensable approach for detecting emerging disease outbreaks and epidemics. Early knowledge of a disease outbreak plays an important role in improving response effectiveness. Although traditional disease surveillance often relies on time-consuming laboratory diagnosis, and the reporting of notifiable diseases is often slow and incomplete, a new breed of global disease surveillance systems has the potential to significantly speed up detection of disease outbreaks. These new, computer-based surveillance systems offer valuable and timely information to hospitals as well as to state, local, and federal health officials. They often rely on intelligent systems and databases, infectious-disease informatics, and advanced analytic techniques such as time-series analysis, text mining, agent-based modeling, social-network analysis, and disease modeling, visualization, and mapping.

With these new advances, public-health surveillance systems are capable of real-time or near real-time detection of serious illnesses and potential exposure to agents of bioterrorism, allowing for a rapid public-health response.

The basis for modern global disease surveillance is the fact that specific diseases of interest can be monitored through various syndromic presentations that can be detected in a timely manner—for example, through nurse calls, medication purchases, and school or work absenteeism. In addition to enabling early detection and reporting of monitored diseases, syndromic surveillance also provides a rich data repository and highly active communication system for situation awareness and event characterization. Multiple participants provide interconnectivity among disparate and geographically separated sources of information to facilitate a clear understanding of the evolving situation. This is important for event reporting, strategic response planning, and disaster victim tracking. Information gained from syndromic surveillance data can also guide the planning, implementation, and evaluation of long-term programs to prevent and control diseases, including distribution of medication, vaccination plans, and allocation of resources.

In recent years, researchers have proposed several syndromic surveillance approaches. According to a study conducted by the US Centers for Disease Control and Prevention (CDC) in 2003, roughly 100 sites throughout the US have implemented and deployed syndromic surveillance systems.¹ These systems, although sharing similar objectives, vary in system architecture, information

processing and management techniques, and algorithms for anomaly detection, and they have different geographic coverage and disease focuses. A recent article by Yan, Chen, and Zeng provides an in-depth survey that analyzes and evaluates existing syndromic surveillance systems and related outbreak modeling and detection work under a unified framework.² Many intelligent global disease surveillance systems have begun to emerge. To illustrate, we present the following case study.

Case Study: BioPortal

In 2003, the University of Arizona Artificial Intelligence Lab along with collaborators in the New York State Department of Health and the California Department of Health Services initiated the BioPortal project to develop an infectious-disease surveillance system. The project has been sponsored by the US National Science Foundation, the Department of Homeland Security, the Department of Defense, the Arizona Department of Health Services, and Kansas State University's BioSecurity Center, under the guidance of a federal interagency working group called the Infectious Disease Informatics Working Committee. Partners in the BioPortal project now include all the original collaborators as well as the US Geological Survey (USGS); University of California, Davis; University of Utah; the Arizona Department of Health Services; Kansas State University; and the National Taiwan University.

The BioPortal research prototype has provided distributed, cross-jurisdictional access to data sets concerning several major infectious diseases, including botulism, West Nile virus, foot-and-mouth disease, livestock syndromes, and chief complaints (both in English and Chinese).

It features advanced spatial-temporal data-analysis methods and visualization capabilities. BioPortal supports syndromic surveillance of epidemiological data and free-text chief complaints. It also supports analysis and visualization of lab-generated gene sequence information. Figure 1 shows BioPortal system architecture.

BioPortal provides automatic syndrome classification capabilities based on free-text chief complaints (CCs). One method recently developed uses a concept ontology derived from the Unified Medical Language System (UMLS).³ For each chief CC, the method first standardizes the CC into one or more medical concepts in the UMLS. The system then maps these concepts into existing symptom groups using a set of rules constructed from a symptom-grouping table. For symptoms not in the table, the system uses a weighted semantic similarity score algorithm, which measures the semantic similarity between the target symptoms and existing symptom groups, to determine the best symptom group for the target symptom. The ontology-enhanced CC classification method has also been extended to handle CCs in Chinese.

BioPortal supports hotspot analysis using various methods for detecting unusual spatial and temporal clusters of events. Hotspot analysis facilitates disease outbreak detection and predictive modeling. BioPortal supports various scan statistics using SaTScan, the nearest-neighbor hierarchical clustering method, and two new machine-learning methods (risk-adjusted support vector clustering, and prospective support vector clustering).⁴ BioPortal makes available a visualization environment called the Spatial-Temporal Visualizer (STV), which allows users to interactively explore spatial and temporal patterns, based on an

integrated toolset consisting of a geographic information systems (GIS) view, a timeline tool, and a periodic pattern tool.⁵

Trends and Controversies

Although syndromic surveillance has gained wide acceptance as a response to disease outbreaks and bioterrorism attacks, many research challenges remain; some reflect future trends, while others present controversies.

False Alarms

Syndromic surveillance systems often generate false alarms because it is difficult to distinguish natural data variations from real outbreaks. Human reviews and follow-up investigations are necessary for verification, which is costly in time and labor. A typical investigation could require a group of epidemiologists, public-health officials, healthcare providers, and their support staff to go through a multistep procedure for alert review and event evaluation. Significant reduction of false alarms is required before any computer-based disease-surveillance system can be used in the field.

Effective Use of Surveillance Systems

There are circumstances in which automated disease surveillance might not be effective or necessary. For example, the potential benefit of disease surveillance can't be realized if hundreds or thousands of people are infected simultaneously. In another scenario, disease surveillance would be ineffective in a case that involved only a few people and thus went undetected.

Data Quality

Real-world disease data tend to be noisy and incomplete. Although reporting of most notifiable diseases

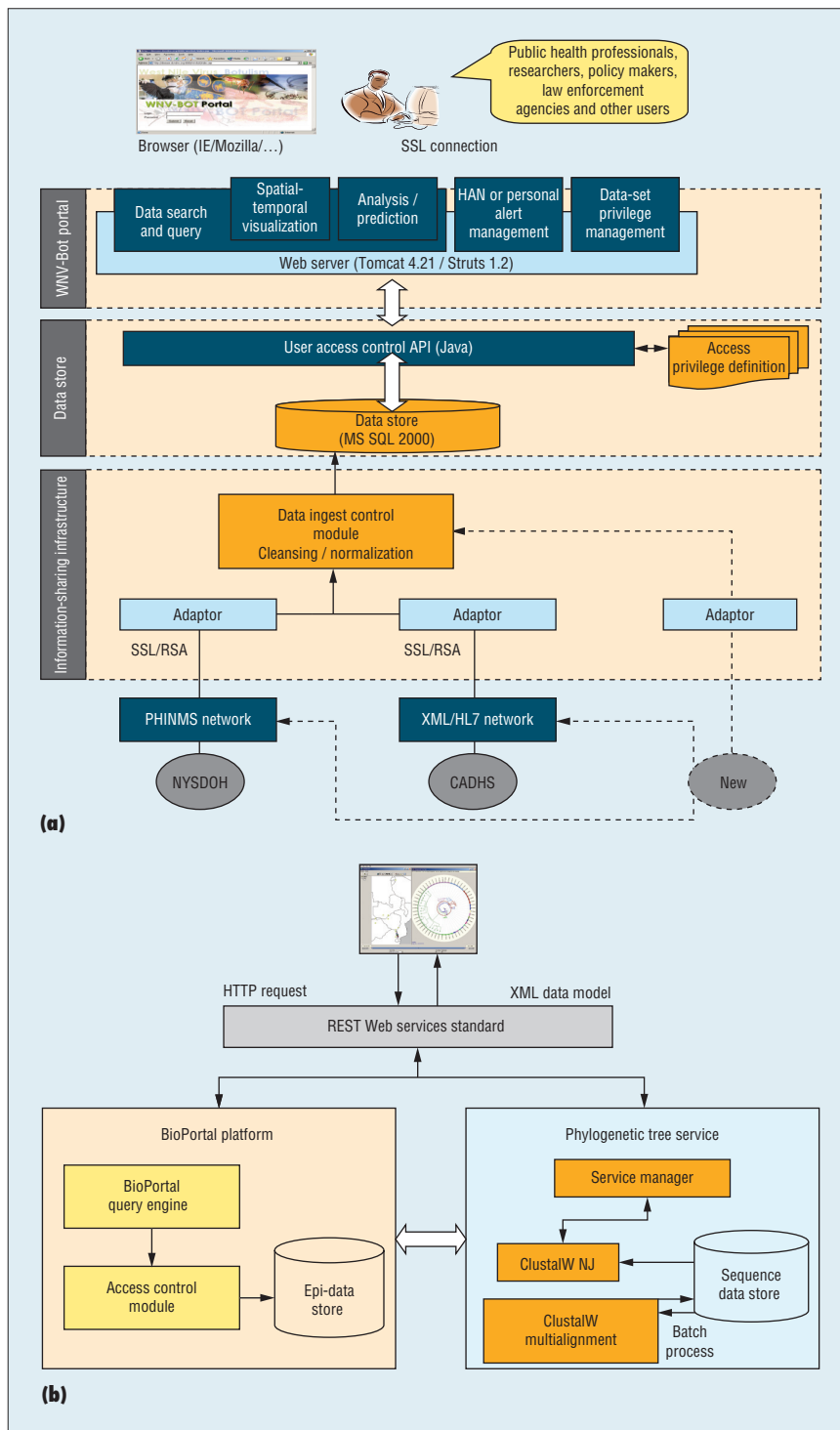


Figure 1. BioPortal system architecture: (a) BioPortal information-sharing and data access infrastructure; (b) BioPortal's enhanced system architecture with epidemiological data and gene sequence data surveillance. PHINMS: Public-Health Information Network Messaging System; HL7: Health Level 7, a medical data-exchange standard; NYSDOH: New York State Department of Health; CADHS: California Department of Health Services; ClustalW: a general-purpose multiple sequence alignment program for DNA or proteins; NJ: the neighbor-joining method used within ClustalW.

through various public-health agencies is required by law, hospitals, laboratories, and clinicians participate for the most part voluntarily. Patients making emergency room visits may not be representative of the population in the neighboring community. This reinforces the need for careful evaluation of data sources and collection procedures.

System Interoperability

Existing systems differ significantly in scope and purpose—for example, in geographical coverage, types of data, and diseases monitored. For instance, some systems might focus on biodefense, whereas other systems might target influenza. The absence of standard vocabularies and messaging protocols leads to interoperability problems among syndromic surveillance systems and the underlying data sources. Health Level Seven (HL7) standards and XML-based messaging protocols represent a potential solution for addressing these problems.

Algorithm Benchmarking and Comparison

Each syndromic surveillance system implements a unique set of outbreak detection algorithms. We urgently need a better understanding of the strengths and limitations of various detection techniques and their applicability. Data visualization techniques—especially interactive visual data exploration techniques—must be further developed to meet the specific analysis needs of syndromic surveillance. Outbreak detection algorithms must be improved in terms of sensitivity, specificity, and timeliness. In particular, how to deal with incomplete data records, how to perform privacy-conscious data mining, and how to leverage multiple data streams, are all interesting research questions.

Assessment, Evaluation, and Deployment

Surveillance system evaluation and comparison are confounded by several practical issues. Systematic, field-based, objective comparative studies among systems are critically needed. Lessons learned from successful and failed implementations must be carefully developed to help with future implementations.

In this Issue

In this issue we include three papers on AI for global disease surveillance from distinguished experts in disease informatics, linguistic analysis, and visual analytics. Each essay presents a unique innovative research framework, computational methods, and selected results and examples.

In “Broadening the Perspective of AI in Public-Health Surveillance: From Local Aberration Detection to Global Epidemic Monitoring and Control,” Buckeridge, Izadi, Verma, and Okhmatovskaia provide an overview of aberrancy detection and present many exciting opportunities for AI research in epidemic monitoring and control on regional and global scales.

In “Mining Biomedical Literature to Identify Viruses and Bacteria as Potential Bioterrorism Weapons,” Hu and Shen present selected computational analysis algorithms for automated identification of viruses and bacteria in biomedical literature. The approach may provide a viable option for mining large-scale biomedical or disease-related knowledge.

In “Disease Surveillance Based on Spatial Contact Networks: A Case Study of the Beijing 2003 SARS Epidemic,” Zeng, Cao, Wang, Chen, Zheng, and Wang present a case study analyzing spatial contact networks to detect the spatial risk of epidemics in large urban areas. Our study uses the survey data concerning spatial

elements of the contact network from the Beijing 2003 SARS epidemic. Research results show that spatial contact networks could be a powerful tool to help detect spatial transmission risk and understand the inherent mechanisms of infectious disease transmission.

Although global surveillance systems have been developed and deployed in many cities, states, and countries, there is a critical need to create a cross-jurisdictional, cross-regional data-sharing infrastructure to maximize the potential benefit and practical impact of global disease surveillance. In a broader context, syndromic surveillance should be turned into a truly global effort for pandemic diseases such as SARS, influenza, foot-and-mouth disease, and so on. We must address issues concerning global data sharing and the development of models that work internationally. In addition to intelligent systems and advanced analytics, international politics, global commerce interests, and cultural and regional considerations are some of the issues that must be considered for effective global disease surveillance.

Acknowledgments

The research is supported in part by the National Science Foundation through Digital Government Grant #EIA-9983304, and Information Technology Research Grant #IIS-0428241.

References

1. J. Buehler et al., “Syndromic Surveillance and Bioterrorism-Related Epidemics,” *Emerging Infectious Disease*, vol. 9, no. 1, 2003, pp. 197–204.
2. P. Yan, H. Chen, and D. Zeng, “Syndromic Surveillance Systems,”

Ann. Rev. Information Science and Technology (ARIST), vol. 41, 2007, pp. 425–495.

3. H. Lu et al., “Ontology-Enhanced Automatic Chief Complaint Classification for Syndromic Surveillance,” *J. Biomedical Informatics*, vol. 4, no. 2, Apr. 2008, pp. 340–356.
4. W. Chang, D. Zeng, and H. Chen, “A Stack-Based Spatio-Temporal Data Analysis Approach Based on Prospective Support Vector Clustering,” *Decision Support Systems*, vol. 45, no. 4, 2008, pp. 697–713.
5. P. Hu et al., “System for Infectious Disease Information Sharing and Analysis: Design and Evaluation,” *IEEE Trans. Information Technology in Biomedicine*, vol. 11, no. 4, 2007, pp. 483–492.

Hsinchun Chen is the McClelland Professor of Management Information Systems at the University of Arizona and director of the Artificial Intelligence Lab. He received his PhD in information systems from New York University. Contact him at hchen@eller.arizona.edu.

Daniel Zeng is an associate professor in the Management Information Systems Department at the University of Arizona and a research professor at the Chinese Academy of Sciences Institute of Automation. Contact him at zeng@email.arizona.edu.

Broadening the Perspective of AI in Public-Health Surveillance: From Local Detection to Global Epidemic Monitoring and Control

David L. Buckeridge, Masoumeh Izadi, Aman Verma, and Anya Okhmatovskaia, *McGill University*

As recently as the 20th century, public-health experts in industrialized countries were confident that infectious

diseases posed little threat to human health. Today, however, infectious diseases remain the leading cause of death among children and young adults, resulting in more than 13 million deaths each year. Moreover, with 335 new infectious diseases identified in the last 60 years and many known organisms, such as influenza, evolving continuously, the world lives under the constant threat of the next pandemic.

Public-health surveillance is critical for detecting and controlling epidemics, and the contribution of AI to surveillance has been considerable over the last decade. In those 10 years, nearly 90 percent of US state public-health departments adopted some form of automated, real-time surveillance.¹ The introduction of these novel, automated systems has presented exciting new opportunities for AI researchers. The systems, usually operated at a city or regional level, acquire large volumes of heterogeneous data in real time from sites such as emergency departments and pharmacies and group the records into syndromes of public-health interest.²

From an applied public-health perspective, the surveillance process occurs in two phases: before and after an epidemic is recognized to be occurring. In this essay, we begin by outlining some contributions of AI to the first phase of surveillance, when the predominant objective is epidemic detection. We then propose two areas of focus for AI in the future: supporting decision making in the phase after an epidemic is detected, and scaling architectures and methods used in local and regional systems to conduct automated surveillance on a global scale.

Contributions of AI to Epidemic Detection

Research on epidemic detection has targeted mainly two problems: identifying

individual cases of a disease or syndrome, and identifying unusual clusters of cases. In terms of case detection, researchers have used ontology to annotate and integrate disparate data and have developed statistical, probabilistic, and rule-based algorithms to process medical text and classify records. At the population level, researchers have developed and validated a wide range of statistical and machine-learning methods to detect unusual patterns. While recent advances in automated epidemic detection have been profound, automated surveillance systems are still perceived as having high false-alarm rates, and there is no consensus on which detection algorithms to use in practice.

Researchers have, not surprisingly, found it difficult to develop detection algorithms that identify weak signals with high accuracy. Perhaps due to the paucity of example epidemic signals, the predominant approach to detection has been to build a model of “normal” data and then monitor for changes from this baseline. Although conceptually straightforward and generically applicable to surveillance for any sort of epidemic, this approach encounters problems as the dimensionality of the data increases. As the number of dimensions increases, so does the likelihood of detecting statistically significant changes from baseline that have no known epidemiological significance. Devising approaches to address this problem, possibly by combining epidemiological knowledge with statistical pattern detection methods, is an important area for future AI research.

Despite the many epidemic detection algorithms that researchers have proposed and evaluated, it is still not clear to public-health practitioners which algorithms to use within their surveillance system. In response to this

need, some research groups have begun to systematically analyze the performance of a wide range of detection algorithms and provide evidence to guide, and possibly even automate, the selection of epidemic detection methods. In the BioSTORM project, for example, researchers are modeling the detection process using a task-analytic approach (Figure 2), systematically evaluating the performance of detection algorithms in terms of this model, and then using machine learning to identify important determinants of epidemic detection.³

Although recent advances in using AI methods to automate detection of individual cases and epidemics have been impressive in many respects, there remains a need for further research. In particular, there is a potential role for AI to make epidemic detection more “intelligent” by highlighting epidemiologically significant aberrations and by guiding the selection of algorithms to suit a particular surveillance situation.

The Limits of Automating Epidemic Detection

Although the methodological advances in epidemic detection over the last decade have been profound, the impact of these advances on public-health practice has been limited. The reasons for this limited impact relate to the role of epidemic detection within the broader context of disease control and the fit of these novel systems within public-health workflows.

To place these issues in perspective, it is helpful to review briefly the evolution of AI in clinical medicine. Initial AI research in clinical decision support focused on diagnosis, and although very accurate algorithms and systems were developed, diagnostic decision support systems are still not used routinely in clinical practice.

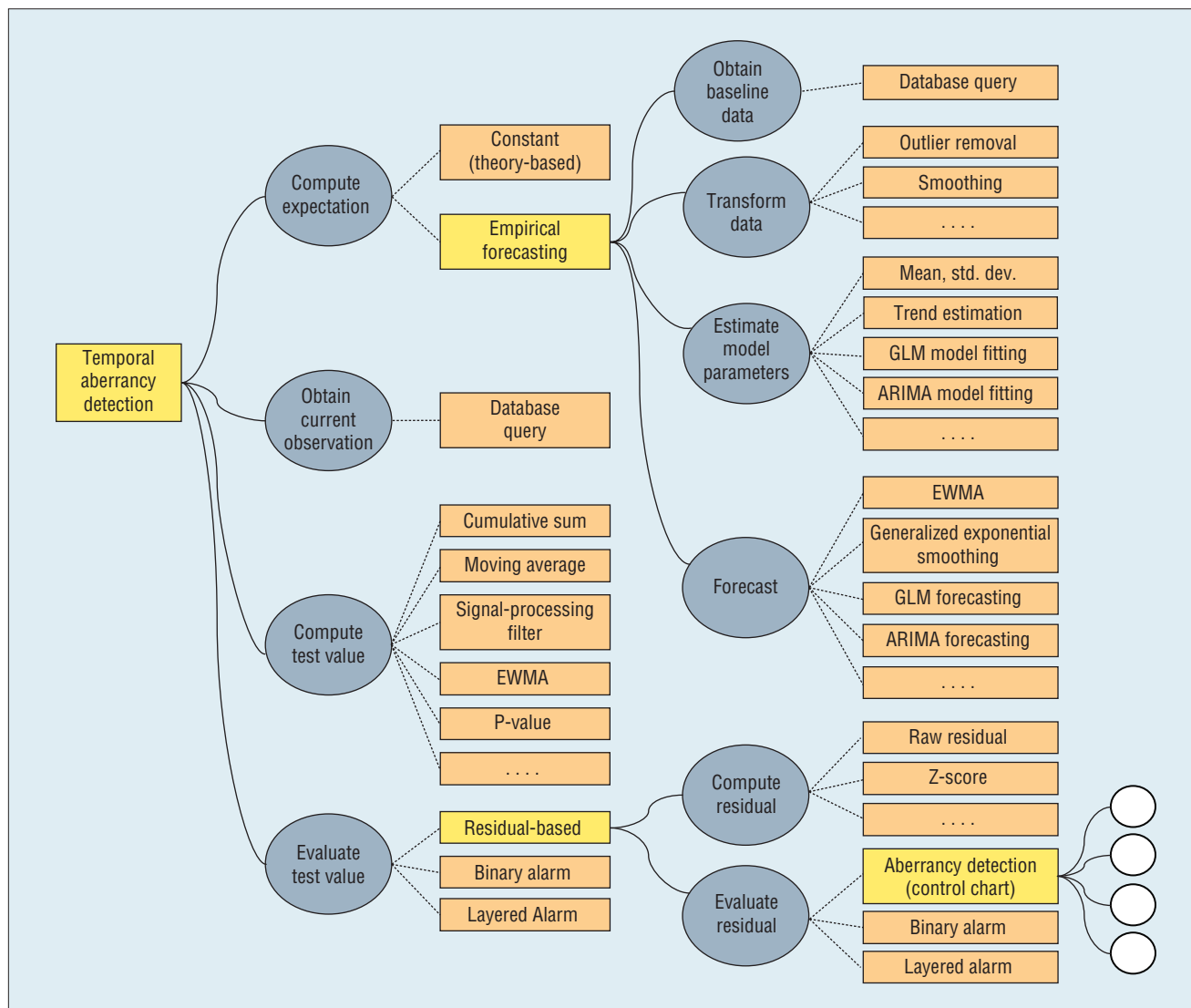


Figure 2. General task structure of temporal detection algorithms. Temporal algorithms are represented as instances of a task-decomposition method (“temporal aberrancy detection” in the figure) that performs the task of detecting aberrations in the surveillance data by decomposing this task into four subtasks (ellipses). Each subtask can be accomplished by different methods (rectangles), some of which perform the task directly (primitive methods, shown as orange rectangles), and some further decompose the task into subtasks (task-decomposition methods, shown as yellow rectangles). For instance, the “compute expectation” task, which constitutes one of the steps (subtasks) of aberrancy detection, can in turn be decomposed into four subtasks, if the “empirical forecasting” method is used. Alternatively, this task can be accomplished directly by a primitive method—“theory-based forecasting.” Similar alternatives exist for the “evaluate test value” task. See our earlier publication for a full description.³

Two critical reasons for this failed adoption are that these systems neither addressed a compelling need nor were integrated with clinician workflow.⁴ In particular, although diagnosis is an interesting problem, clinical practice directs far more time and effort toward disease management. Accordingly, work in clinical decision support has shifted to develop

systems that guide therapy and management decisions.

Analogously, although epidemic detection is an interesting and easily studied problem, most public health departments spend little time attempting to detect epidemics through surveillance; instead they have come to rely on these automated systems to monitor the progression of identified

outbreaks, such as the seasonal influenza epidemic.¹ Recent experiences with using these systems for H1N1 pandemic influenza surveillance indicate that automated systems can be adapted flexibly to monitor novel diseases and can provide critical insights in real time into the epidemiology of previously unknown infectious agents. Understanding how these

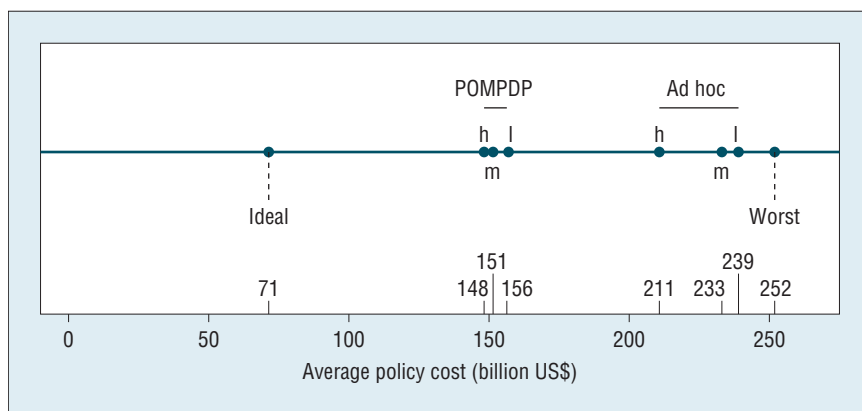


Figure 3. Average cost of an inhalational anthrax outbreak under different policies for responding to alarms from an aberrancy detection algorithm. The ideal policy is one where true alerts are always acted upon immediately and false alerts are always ignored. The worst policy is one where true alerts are always ignored. The ad hoc policies all follow the same series of steps after an alert, reflecting current public-health practice. The POMDP policies respond to each alert in a manner that minimizes costs as estimated using a Partially Observed Markov Decision Process. The letters l, m, and h refer to the sensitivity of the aberrancy detection algorithm as low, medium, and high. See our earlier publication for a full description.⁵

systems are best integrated into public-health workflows, however, remains an open research challenge.

Using AI to Guide Epidemic Monitoring and Response

Although aberration detection is important, the ability to monitor and control an epidemic is as pressing and in many respects a more important problem for public health. Epidemic monitoring and control also pose many challenges for AI researchers, including

- describing epidemiologically relevant patterns of disease,
- forecasting and visualizing an epidemic's likely evolution, and
- guiding the selection and evaluation of epidemic control measures.

The ability to describe and, ideally, explain epidemiologically relevant disease patterns is critical for developing an epidemiological profile of the disease, and in particular for identifying subpopulations with an elevated risk of infection or complications. Factors to be considered include age, symptom profile, severity of

illness, and the spatial extent of cases. Given the high-dimensional nature of the data, approaches to pattern detection and classification that do not incorporate epidemiological knowledge are unlikely to succeed. Hybrid knowledge-based and statistical approaches, however, are more promising. For example, it may be possible to define certain types of patterns that have epidemiological relevance and then implement statistical methods to identify these types of pattern. This type of approach has been applied with some success in other medical settings such as intensive-care unit monitoring.

In order to assess and organize the resources needed to control an epidemic, public-health personnel must be able to forecast and visualize the likely evolution of an epidemic. With infectious diseases, this type of forecasting requires knowledge of how the organism is likely to spread through a population, including topics such as transmissibility and the contact network. The last decade has seen an increase in research on epidemic modeling, resulting in advances in the use of network theory, agent-based

modeling, and other strategies to model epidemics in human populations. However, researchers typically use this type of modeling to draw generic inferences about epidemic control, as opposed to guiding specific day-to-day decisions about managing an ongoing epidemic. The direct linkage of surveillance systems with epidemic simulation models to allow real-time forecasting is a particularly promising area for future AI research.

Guiding the selection and evaluation and epidemic control measures is necessary to ensure that available data and knowledge are brought to bear in making evidence-based decisions during a potentially politically charged and chaotic situation. The problem here is, given likely scenarios for the natural evolution of an epidemic, to identify the optimal public-health control measures to limit morbidity and mortality among the population. The capacity to simulate the progression of an epidemic remains essential, but now this forecasting must be combined with knowledge about potential interventions and their effectiveness. Some researchers have begun to explore this issue by using methods from reinforcement learning to improve decision making following an alert.⁵ Although initial results of this research suggest that optimizing decision making during this phase of surveillance has the potential to improve outcomes (Figure 3), there remains considerable room for innovative research to address this problem.

Moving Automated Surveillance to a Global Scale

So far in this essay, we have considered the historical and potential contributions of AI towards automating local and regional public-health surveillance systems. Infectious diseases, however, are increasingly global in

terms of their emergence and the rate at which they spread through a highly connected world. The need to scale the methods behind local and regional automated surveillance to allow systems to operate at the global scale is an exciting challenge for AI researchers.

There currently exists no automated surveillance of individual cases of disease at the global scale. In the absence of such systems, a handful of automated systems were developed over the last decade to perform global surveillance of Internet media. These systems scan a wide range of information available over the Internet and attempt to integrate findings from disparate sources using rule-based and statistical text-processing methods. Some of these systems have been successful in identifying epidemics rapidly, especially in situations when official reporting has been slow. Although they play an important role in global surveillance, these Internet media surveillance systems lack the clinical detail found in surveillance of cases or syndromes of diseases.

Privacy and legal constraints make global surveillance of individual data practically impossible to achieve, but distributed surveillance presents an alternative approach that is quickly coming to the fore. In this approach, existing automated surveillance systems operating at the local or regional level forward an aggregated version of their data to another system that integrates then analyzes these data to give a wider geographical perspective on the epidemic. For example, the Distribute project of the International Society for Disease Surveillance (<http://distribute.syndromic.org>) operates in this manner to integrate surveillance data about influenza-like illness and gastrointestinal disease from systems in the United States and Canada. Initial results from this project suggest

that the distributed approach is informative, flexible, and potentially expandable to a continental or even global scale. This scale of automated surveillance presents a wide range of exciting AI research opportunities in terms of aberrancy detection and epidemic monitoring and control using global data.

Infectious diseases continue to pose a threat to global health, and public-health surveillance has advanced considerably in the last decade in response to this threat. AI researchers have played an important role in these advances, mainly through the development of epidemic detection methods. Epidemic detection remains an important focus for research, but more broadly, there are many exciting opportunities for AI research on epidemic monitoring and control at regional and global scales.

References

1. J.W. Buehler et al., "Syndromic Surveillance Practice in the United States: Findings from a Survey of State, Territorial, and Selected Local Health Departments," *Advances in Disease Surveillance*, vol. 6, no. 3, 2008.
2. J.A. Lombardo and D.L. Buckeridge, *Disease Surveillance: A Public Health Informatics Approach*, John Wiley & Sons, 2007.
3. D.L. Buckeridge et al., "Understanding Detection Performance in Public Health Surveillance: Modeling Aberrancy-Detection Algorithms," *J. Am. Med. Informatics Assoc.*, vol. 15, no. 6, 2008, pp. 760–769.
4. R.A. Miller, "Medical Diagnostic Decision Support Systems—Past, Present, and Future: A Threaded Bibliography and Brief Commentary," *J. Am. Med. Informatics Assoc.*, vol. 1, no. 1, 1994, pp. 8–27.
5. M.T. Izadi and D.L. Buckeridge, "Optimizing Anthrax Outbreak Detection Methods Using Reinforcement Learning," *Proc. AAAI Conf. Artificial Intelligence*, AAAI, 2007, pp. 1781–1786.

David L. Buckeridge is an assistant professor in the Department of Epidemiology, Biostatistics, and Occupational Health at McGill University. Contact him at david.buckeridge@mcgill.ca.

Masoumeh Izadi is an adjunct professor in the School of Computer Science at McGill University. Contact her at mtabae@cs.mcgill.ca.

Aman Verma is a PhD student in the Department of Epidemiology, Biostatistics, and Occupational Health at McGill University. Contact him at aman.verma@mcgill.ca.

Anya Okhmatovskaia is a research associate in the Department of Epidemiology, Biostatistics, and Occupational Health at McGill University. Contact her at anya.okhmatovskaia@mcgill.ca.

Mining Biomedical Literature to Identify Viruses and Bacteria as Potential Bioterrorism Weapons

Xiaohua Hu, Henan University and Drexel University
Xiaijong Shen, Henan University

Biomedical-literature mining is a procedure to identify important knowledge and information from large biomedical document set. It has been applied to a wide range of applications, such as recognizing biomedical terms, extracting protein interactions, discovering protein functional regions, and generating hypotheses.¹ The problem of mining implicit

Table 1. Geissler's 13 criteria for viruses.⁴

No.	Criterion
1	The agent should consistently produce a given effect: death or disease.
2	The concentration of the agent needed to cause death or disease—the infective dose—should be low.
3	The agent should be highly contagious.
4	The agent should have a short and predictable incubation time from exposure to onset of the disease symptoms.
5	The target population should have little or no natural or acquired immunity or resistance to the agent.
6	Prophylaxis against the agent should not be available to the target population.
7	The agent should be difficult to identify in the target population, and little or no treatment for the disease caused by the agent should be available.
8	The aggressor should have means to protect his own forces and population against the agent clandestinely.
9	The agent should be amenable to economical mass production.
10	The agent should be reasonably robust and stable under production and storage conditions, in munitions, and during transportation. Storage methods should be available that prevent gross decline of the agent's activity.
11	The agent should be capable of efficient dissemination. If it cannot be delivered via an aerosol, living vectors (such as fleas, mosquitoes, and ticks) should be available for dispersal in some form of infected substrate.
12	The agent should be stable during dissemination. If it is to be delivered via an aerosol, it must survive and remain stable in air until it reaches the target population.
13	After delivery, the agent should have low persistence, surviving only for a short time, thereby allowing a prompt occupation of the attacked area by the aggressor's troops.

knowledge or information from biomedical literature was exemplified by D.R. Swanson's pioneering work on Raynaud's disease and fish oil in 1986.² He found from biomedical literature that Raynaud's disease is a peripheral circulatory disorder aggravated by high platelet aggregation, high blood viscosity, and vasoconstriction. In a separate set of literature on fish oils, he discovered that the ingestion of fish oil can reduce these phenomena. But no single article from either set in the biomedical literature in 1986 mentioned Raynaud's disease and fish oil together. Putting these two separate literature sets together, Swanson hypothesized that fish oil might be beneficial to people suffering from Raynaud's disease. Later, Swanson extended his methods to search medical literature for potential viruses.³

Background: Viruses and Bacteria

Some viruses and bacteria have been identified as bioterrorism weapons. However, many other viruses and bacteria have the potential to become bioterrorism weapons. Because biological agents such as viruses evolve through mutations and biological or chemical changes, and because some biological substances have the potential to turn into deadly viruses through chemical, genetic, and biological reactions, there should be an automatic way to keep track of existing suspicious viruses and to discover new viruses that have the potential to become weapons. It would be very useful to identify those biological substances and take precautionary actions or measurements.

Geissler identified and summarized 13 criteria (listed in Table 1) for

identifying biological warfare agents as viruses.⁴ He then compiled 21 viruses that meet some of these criteria. Following are the 21 viruses in MeSH terms:

- Hemorrhagic fever virus, Crimean-Congo
- Lymphocytic choriomeningitis virus
- Encephalitis virus, Venezuelan equine
- Encephalitis virus, Western Equine
- Encephalitis virus, Eastern equine
- Encephalitis virus, Japanese
- Encephalitis viruses, tick-borne
- Encephalitis virus, St. Louis
- Arenaviruses, New World
- Marburg-like viruses
- Rift Valley fever virus
- Yellow fever virus
- Chikungunya virus
- Dengue virus
- Ebola-like viruses
- Hantaan virus
- Hepatitis-A virus
- Orthomyxoviridae
- Junin virus
- Lassa virus
- Variola virus

There are 13 known bacteria that can cause deadly disease. For example, anthrax is an acute infectious disease caused by the spore-forming bacterium *Bacillus anthracis*. Anthrax most commonly occurs in wild and domestic lower vertebrates (cattle, sheep, goats, camels, antelopes, and other herbivores), but it can also occur in humans when they are exposed to infected animals or tissue from infected animals, or when anthrax spores are used as a bioterrorist weapon. *Coxiella burnetii* is a highly infectious agent that is rather resistant to heat and drying. It can become airborne and inhaled by humans. A single *C. burnetii* organism can cause disease in a susceptible person. This agent could be developed for use in biological warfare and is considered a potential terrorist threat.

Swanson developed a text-mining method to search biomedical literature for potential viruses.³ His method's greatest limitation is that it uses only three properties or criteria of a virus as search keywords, and it ignores semantic information in the searching procedure. In this essay, we present a novel biomedical-literature-mining algorithm based on the Swanson philosophy but extending it significantly. Our objective is to extend the existing known virus list compiled by the CDC and the bacteria recognized by domain experts as bioterrorism weapons to include other viruses and bacteria that might have similar characteristics. We thus hypothesize that viruses and bacteria that have been researched with respect to the characteristics possessed by existing viruses are leading candidates for extending the virus and bacterium list.

Design Method

MedMeSH Summarizer summarizes a group of genes by filtering the biomedical literature and assigning relevant keywords describing the functionality of a group of genes.¹ Each gene cluster contains N genes, and each gene has a set of terms associated with it. MedMeSH Summarizer thus builds a co-occurrence matrix, with the number of citations associated with the gene and containing the MeSH (Medical Subject Headings) term as the cell values. On the basis of this matrix and statistical information, our algorithm calculated overall relevance ranking for all the terms describing the topic of a certain cluster of genes. The PubMed database defines 630 bacteria. We found it quite reasonable to extract topical terms for the 13 known bacteria in Table 2; we then used these terms to look for other suspicious bacteria.

Table 2. Bacteria used in biological warfare.

Bacteria name	Disease caused
<i>Bacillus anthracis</i>	Anthrax
<i>Clostridium botulinum</i>	Botulism
<i>Brucella melitensis</i> <i>Brucella abortus</i> <i>Brucella suis</i>	Brucellosis
<i>Vibrio cholerae</i>	Cholera
<i>Francisella tularensis</i>	Tularemia
<i>Burkholderia mallei</i> <i>Burkholderia pseudomallei</i>	Glanders
<i>Coxiella burnetii</i>	Q fever
<i>Salmonella</i>	Salmonellosis, typhoid fever
<i>Yersinia pestis</i>	Plague
<i>Shigella dysenteriae</i>	Shigellosis

Definition of Terms

The *normalized term bacterium matrix* is

$$\tilde{f}_{ij} = F_{ij} / \left(\sum_{i=1}^M F_{ij} \right)^\alpha \quad (0 \leq \alpha \leq 1) \quad (1)$$

where F_{ij} is the number of PubMed documents retrieved by bacterium j that contain term i .

Relevance ranking uses the following three criteria:

1. Cluster topics (major): Terms occur in most bacteria with high frequency. (Criterion R1: Rank the MeSH terms by decreasing order of the means μ_j .)
2. Cluster topics (minor): Terms occur in most bacteria with low frequency. Compute

$$\sigma_i = \sqrt{\left(\sum_{j=1}^N (\tilde{f}_{ij} - \mu_i)^2 \right) / N}.$$

(Criterion R2: Rank the MeSH terms by decreasing order of the ratios μ_j/s_j .)

3. Particular topics: Terms occur in a few bacteria with high frequency. (Criterion R3: Rank the MeSH terms by decreasing order of the ratios s_i^2/μ_j .)

Each MeSH term in Ω is ranked on the basis of criteria R1–R3. We then

give the terms an overall relevance rank R , where

$$R = wR_1 + [(1-w)/2]R_2 + [(1-w)/2]R_3 \quad (2)$$

The weight parameter in Equation 2 is set so that the major topics have weight w , being the most important set of terms in providing a summary of the cluster. The remaining weight $(1-w)$ is divided equally between the minor topics and the particular topics. In our system, we set w to 0.5 because we look for more topical terms of the entire bacteria cluster.

Algorithm Procedure

The algorithm we used to look for suspicious bacteria and viruses consists of the following steps:⁵

1. Submit query “bacteria name [major]” to PubMed and download MeSH terms after applying a stopword list for each biological agent. We downloaded documents for the 13 existing bacteria. Our stopword list is composed of MeSH terms extracted from PubMed documents (1994–2004) by their overall usage. (For example, one very frequently used MeSH term without biomedical meaning is “government supported.”)

Table 3. Ranked bacteria.

Rank	Bacterium	Weight
1	<i>Clostridium tetani</i>	38.8
2	<i>Erysipelothrix</i>	36.96
3	<i>Coxiellaceae</i>	31.57
4	<i>Sarcina</i>	31.27
5	<i>Yersinia pseudotuberculosis</i>	28.16
6	Atypical bacterial	26.41
7	<i>Corynebacterium diphtheriae</i>	26.22
8	<i>Photobacterium</i>	26.13
9	<i>Brucella</i>	24.9
10	<i>Haemophilus ducreyi</i>	24.69
11	<i>Brucellaceae</i>	23.68
12	<i>Campylobacter fetus</i>	22.74
13	<i>Yersinia enterocolitica</i>	21.95
14	<i>Bacillus thuringiensis</i>	21.24
15	<i>Pediococcus</i>	21.2
16	<i>Mycobacterium bovis</i>	20.36
17	<i>Proteus vulgaris</i>	20.23
18	<i>Haemophilus influenzae-b</i>	19.89
19	<i>Nocardia asteroides</i>	19.88
20	<i>Bacillus megaterium</i>	19.69

- Build a normalized matrix (Equation 1) of terms by bacterium (for the already identified 13 bacteria).
- Rank all the terms according to Equation 2 and pick top k terms.
- Download documents for the remaining 617 bacteria, build a matrix of terms by bacteria (Equation 1).
- Let R_i be the rank value of bacterium i ; we use the following formula:

$$R^B = \sum_{i=1}^M \tilde{f}_{ij} \times R_i$$

Results

We applied our method to two data sets: viruses and bacteria. Because of this essay's space limitations, we list only the result for bacteria. Table 3 displays the top-ranked bacteria by criteria.

Table 3 demonstrates that many potential bacterial threats could affect public health on a wide scale if disseminated effectively. We intend our findings to be used as a guide to the virus literature to support further studies that could lead to appropriate defense and public-health measures.

In 1999, the US Department of Defense was involved in building a directory of known emerging infectious diseases and laboratory tests worldwide. It identified approximately 40 high-threat agents for bioterrorism, including many of the hemorrhagic viruses.⁶ However, since that time, we have seen the emergence of SARS, avian flu, the H1N1 and H1N5 viruses, and many other threats to the public health. This situation is

worrisome to public-health officials, who are concerned that the public-health system might not yet be prepared fully for such a crisis as the release of a viral agent in the US population. Certainly, steps have been made in laboratory and public-health preparedness to identify such threats, but potential gaps remain.

Viruses vary in their biological capability to survive, replicate, and infect. Therefore, lists such as those we have produced at least remind us that other viral agents besides the commonly known groups can potentially cause disease or terror in populations. We must be prepared; without continued work such as this to identify additional threats, preparedness efforts may fall short.

Acknowledgments

Xiaohua Hu's research work presented here is supported in part by a US National Science Foundation Career grant (NSF IIS 0448023), NSF grant CCF 0905291, NSF grant IIP-0934197, and a Pennsylvania Department of Health grant.

References

- P. Kankar et al., "MedMeSH Summarizer: Text Mining for Gene Clusters," *Proc. 2nd SIAM Int'l Conf. Data Mining*, SIAM, 2002, pp. 548–565.
- D.R. Swanson, "Fish-Oil, Raynaud's Syndrome, and Undiscovered Public Knowledge," *Perspectives in Biology and Medicine*, vol. 30, no. 1, 1986, pp. 7–18.
- D.R. Swanson, N.R. Smalheiser, and A. Bookstein, "Information Discovery from Complementary Literatures: Categorizing Viruses as Potential Weapons," *J. Am Soc. Information Science and Tech.*, vol. 52, no. 10, 2001, pp. 797–812.
- E. Geissler, ed., *Biological and Toxin Weapons Today*, SIPRI Publications, 1986.

5. X. Hu et al., "Integration of Instance-Based Learning and Text Mining for Identification of Potential Virus/Bacterium as Bio-terrorism Weapons," *Intelligence and Security Informatics*, LNCS 3975, Springer, 2006, pp. 548–553.
6. P. Rumm et al., "A Department of Defense (DOD) Virtual Public Health Laboratory Directory," *Military Medicine*, no. 165, supp. 2, Jul. 2000, p. 73.

Xiaohua Hu is a Yellow River Scholar at Henan University, China. He is an associate professor of the College of Information Science and Technology at Drexel University. Contact him at thu@cis.drexel.edu

Xiaojong Shen is a professor in the College of Computer and Information Engineering at Henan University, China. Contact him at shenxj@henu.edu.cn

Disease Surveillance Based On Spatial Contact Networks: A Case Study of Beijing 2003 SARS Epidemic

Daniel Zeng and Hsinchun Chen,
University of Arizona
Zhidong Cao, Fei-Yue Wang, and
Xiaolong Zheng, *Chinese Academy of Sciences*
Quanyi Wang, *Beijing Center for Disease Control and Prevention*

The explosive growth of urban population and accelerated urbanization in many parts of the world have led to a variety of public-health challenges, as the high population density and the high intensity and frequency of contacts and interactions among the population can be a significant epidemic risk factor.¹ In particular, socioeconomic status and social-network structures are major determinants of the intensity and frequency of direct contacts in the

same way that geographic locations determine the spatial diffusion patterns of infectious diseases. Various properties of people's spatial contact networks must be studied to facilitate a better understanding of disease transmission mechanisms and ultimately help officials make more informed public-health response decisions.

We can construct two types of contact networks—social-contact networks and spatial-contact networks—on the basis of social and geographical relationships among infected and susceptible people. Lately, epidemiological research has focused on the social-contact network-based survey methodology.^{1,2} This approach allows researchers to study several interesting network properties concerning degree distributions, including the well-known small-world and scale-free properties. In the recent literature, researchers have successfully applied social-contact network analysis to analyze AIDS, tuberculosis, and SARS data; the approach is proven to be a useful tool to help predict and respond to infectious-disease outbreaks, complementing existing epidemiological toolsets.

However, spatial diffusion patterns, which are essential for designing response and prevention strategies during outbreaks, have been largely ignored in this line of research. "Everything is related to everything else, but near things are more related than distant things."³ Geographical locations provide contexts for social, economic, and cultural aggregation. Because people are connected via social and geographical contexts, geographical contacts are critical for understanding the outbreaks, finding potential connections among patients, and evaluating the role that those geographical locations play in disease outbreaks. We argue that a better understanding of the characteristics of the spatial-contact network can provide critical insights into disease

dynamics in space, and can serve as a disease surveillance method from the point of view of spatial risk detection. The origins of this idea are as early as 1854, when John Snow mapped all patients and analyzed the spatial relationships in patient-contact networks in urban regions to effectively contain the spread of cholera in London.⁴

Although a substantial amount of work has gone into developing mathematical models of social-contact networks, work on incorporating important geographical contexts into contact networks is just starting, calling for research in both empirical and computational analysis, and theoretical modeling building. A major reason for lack of progress in this area seems to be twofold. First, research in this area must be highly interdisciplinary, cutting across computational sciences, epidemiology and public health, and geographical information science, an area of study that started not long ago. Second, the lack of detailed data sets with spatiotemporal-network data elements capturing both social and geographical contexts has also slowed the research effort in this area. These barriers are disappearing due to a number of intellectual and practical reasons. We believe that the study of spatial-contact networks integrating both social and geographical contexts will become a hotspot for infectious-disease informatics research with both academic and practical significance and relevance.

In this essay, we present a case study analyzing spatial-contact networks to detect the spatial risk of epidemics in large urban areas. Our study uses the survey data concerning spatial elements of the contact network from the Beijing 2003 SARS epidemic. For model validation purposes, we compare the spatial risk detected through spatial-contact networks with that detected through widely used local

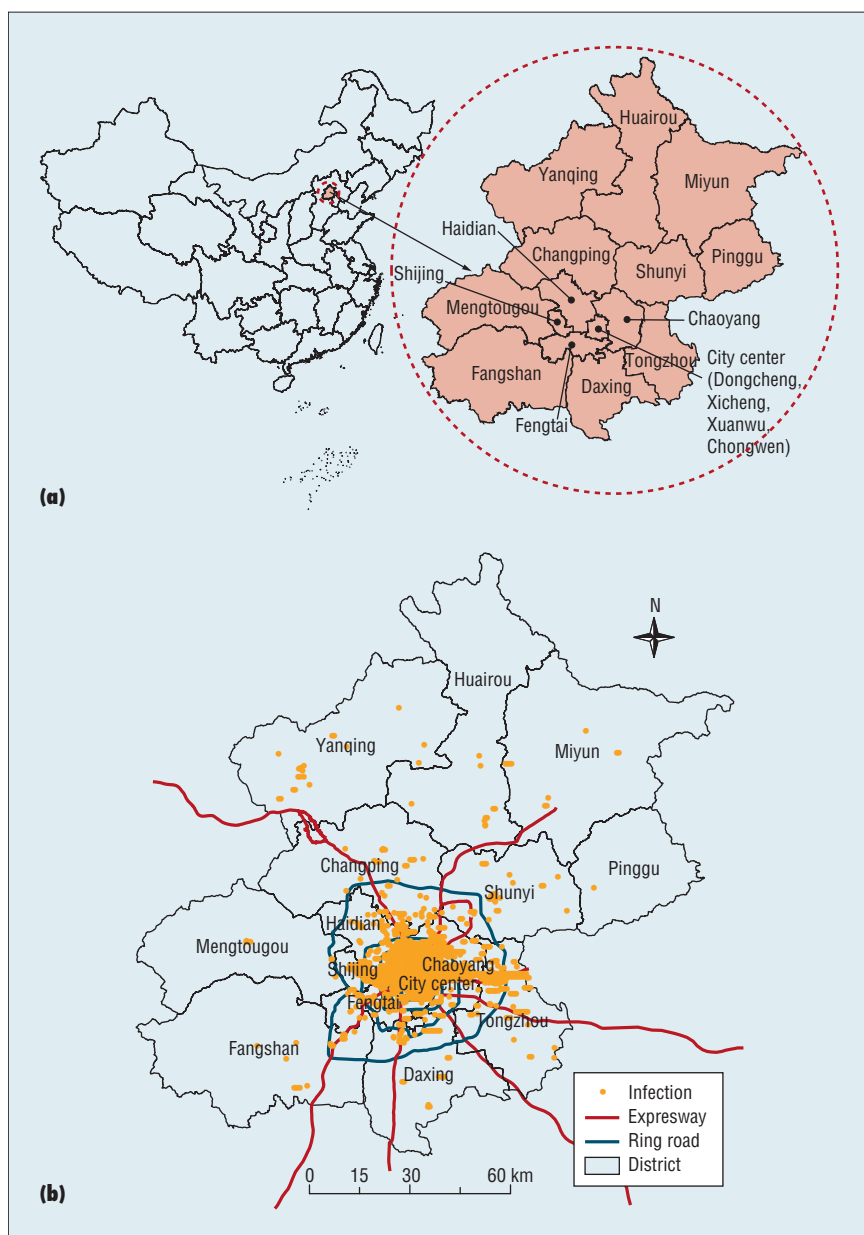


Figure 4. Spatial distribution of 2,321 SARS cases in Beijing in 2003. (a) The city of Beijing covers 14 districts and two counties, of which eight districts (Dongcheng, Xicheng, Xuanwu, Chongwen, Haidian, Chaoyang, Shijing, and Fengtai) are core urban regions. (b) We mapped all 2,321 SARS cases by the patient's home address.

spatial statistics, LISA (local indicators of spatial association).

Case Study: Beijing 2003 SARS Epidemic

We extracted the data used in our study from a survey of 2,444 SARS patients in Beijing, conducted by the Beijing Centers for Disease Control and Prevention and covering the

period from 1 March 2003 to 7 June 2003. Patients or their families were asked to provide information about gender, age, home address, work address, onset of symptoms, and a description of who had contact with whom. To analyze the spatial patterns and understand the complexity of spatial transmission risks, we have adopted home addresses as the spatial

locations for cases, because homes have been widely recognized as one of the highest-risk locations for SARS transmission. The map in Figure 4a shows the districts of Beijing; Figure 4b maps the spatial locations of 2,321 SARS cases studied. (We abandoned the remaining 123 SARS cases because of incomplete, wrong, or unavailable spatial information.)

From our data set, 624 confirmed SARS patients have records of detailed patient contact relationships. Their contact relationships form 482 infectious links; each link is a directed arc pointing from patient A to patient B, if A transmitted the SARS virus to B. We focus on a computational analysis for spatial-risk detection of infectious-disease transmission based on 408 available links whose corresponding nodes are in Beijing. (Our study omits links concerning patients who weren't local residents and weren't infected in Beijing but were treated in Beijing hospitals.) Although the study sample size is small, this data set is representative of the population and is the best one can do given the data collection challenges during a major public-health crisis. Figure 5 shows the spatial-contact network of 408 links.

Figure 6 shows the patterns of the SARS spatial-contact network with respect to distance and direction. The results reveal that SARS transmission is anisotropic in Beijing. The city center has the highest risk of SARS epidemic, and the risk in the east is significantly higher than that in other directions. Transmission distances of links appear to follow approximately log-normal distribution, with the average transmission distance between different districts being around 18 km and the standard deviation 23.4 km. The coefficient of variation ($23.4/18.1 = 1.3$) is fairly large, indicating that transmission distance is scattered in space and that it

is hard to control the spatial diffusion of SARS transmission. We also note an abnormally high-risk region in the east part of Beijing. This could be a result of more frequent interaction between the people working in the city center and those living in the eastern areas of Beijing.

To further analyze transmission risks between different districts, we aggregated all the nodes of links to the geometrical centers of Beijing's districts and constructed a new spatial-contact network among all districts, which we show in Figure 7. The transmission risks among four districts (city center, Chaoyang, Haidian, and Tongzhou) are significantly higher than those among other districts. During the SARS epidemic there were four strong transmission channels: city center–Haidian, city center–Chaoyang, city center–Tongzhou, and Chaoyang–Tongzhou. Intuitively, we expect the high risk of SARS transmission in the city center, Chaoyang, and Haidian because of high-density population in these areas. Surprisingly, the Tongzhou district, located in the far east of Beijing with a low population density and few hospitals receiving SARS patients, also appears to possess high risk. A partial explanation could be economic-social: the cost of living in the Central Business District (CBD) located in the Chaoyang district is high, so many people working in the CBD choose to live in the Tongzhou district, a far more affordable area.

To evaluate the effectiveness and usefulness of an approach to disease surveillance based on spatial-contact networks, we have conducted a comparative study using the spatial statistical methods as a benchmark. Spatial statistics can be applied to detect spatial risks of disease transmission by identifying spatial dependencies among observations. In our study, we

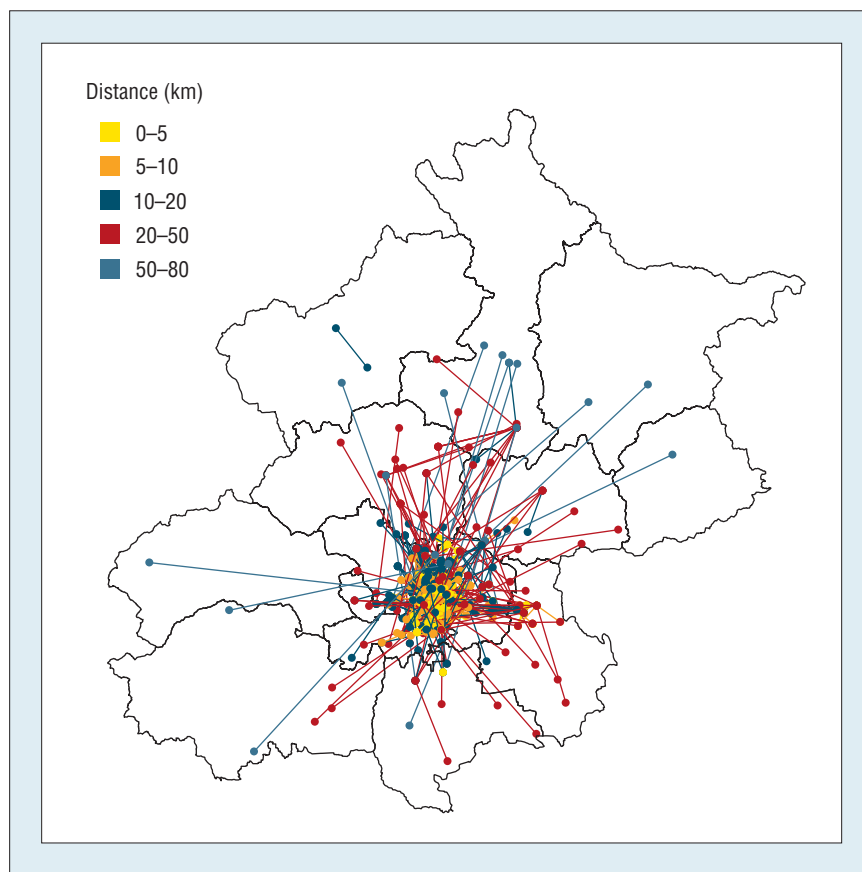


Figure 5. Spatial-contact networks of SARS transmission in Beijing, 2003. The map shows the spatial relationship of the patient contact network with 624 nodes and 408 links, where nodes are SARS-infected persons within the networks, and links indicate infection relationships.

used the widely applied LISA statistic, an indicator of spatial association, to analyze the morbidity rate of Beijing SARS cases using local clusters (regions where adjacent areas have similar values) or spatial outliers (areas distinct from their neighbors),⁴ shown in Figure 8b. A LISA map uses four categories of spatial patterns. The high–high and low–low locations (positive local spatial autocorrelation) are typically called *spatial clusters*, while the high–low and low–high locations (negative local spatial autocorrelation) are termed *spatial outliers*. A cluster is computed as such when the value at a location (either high or low) is more similar to its neighbors than would be the case under spatial randomness. The high–high locations refers to hotspot areas, where the risk

of SARS spreading is higher than average; the low–low locations refers to coolspot areas, where the risk of SARS spread is lower than average.

We found that the spatial-risk detection results of spatial-contact networks were largely consistent with results based on local spatial association.

Discussion

We report in this essay a computational exploration of spatial-contact networks in the context of infectious-disease surveillance. Research results show that spatial-contact networks could be a powerful tool to help detect spatial transmission risk and understand the inherent mechanisms of infectious-disease transmission.

When an epidemic outbreak occurs, especially for new and emerging

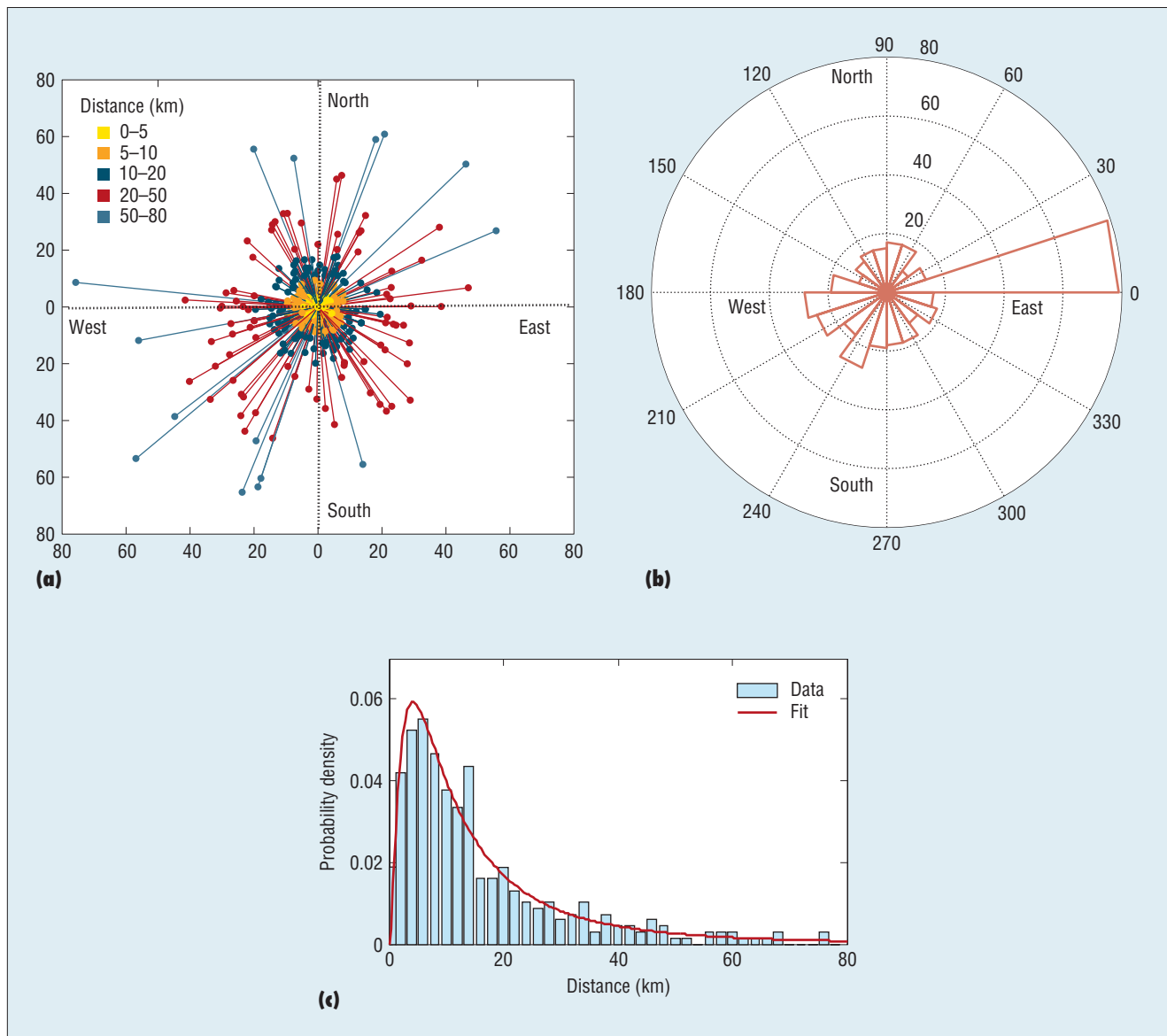


Figure 6. Patterns of the SARS spatial-contact network with 408 links. (a) This graph shows the directions and distances of 408 links, with the start node of each link placed at the origin of the coordinate system. (b) This rose map plots the same data. (c) This graph shows the probability density of all distances of the links with starting and ending nodes in different districts, and a log-normal function fitting the density.

diseases, action must be taken on the basis of sparse knowledge about the pathogen and disease transmission. Public-health response mechanisms such as quarantine are often implemented without a proper understanding of their impact. Geographical locations, as well as social relationships, are crucial for understanding spatial-temporal pattern of infectious diseases and could lead to a better understanding of the complex dynamic of transmission

and become part of informatics disease analytics for actionable use. Our case study presents preliminary empirical evidence about the importance of spatial-contact networks.

Traditional social-contact network analysis models transmission of diseases primarily based on direct or indirect person-to-person contacts. Traditional epidemiological work has

been focusing mainly on time-series analysis and dynamic systems, although researchers are increasingly incorporating spatial elements into epidemiological analyses through spatial statistics modeling techniques and spatial surveillance algorithms.

Any infectious-disease transmission that is based on personal contact is bound to have strong spatial features as people aggregate and move around. As such, analysis based on

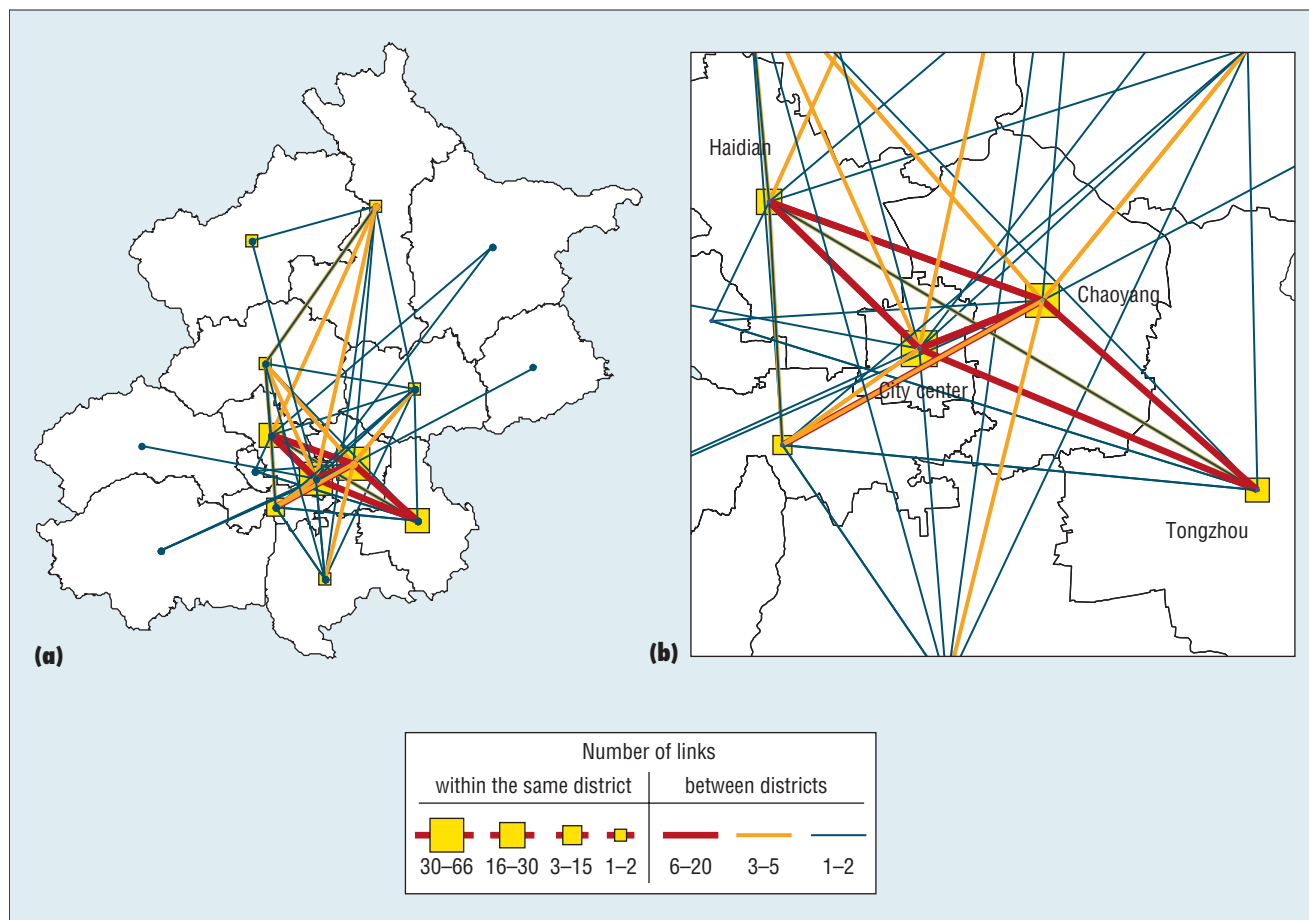


Figure 7. Patterns of the spatial-contact network among districts. Lines indicate links between different districts, and blocks refer to the links within the same district. Heavier lines and larger blocks indicate higher risk of SARS transmission.

social-spatial contact networks has wide applicability. Spatial-contact networks can reveal potential transmission risks among patients in space, complementing a pure spatial epidemiological analysis that ignores patient interaction information. At the same time, spatial-contact networks complement social-contact networks by enriching the context of patient contacts with spatial information and enabling analysis of interactions between spatial and social factors.

From both domain science and methodological viewpoints, however, research on spatial-contact networks is just beginning, and many challenges remain.

First, we need empirical studies to assess the usefulness of spatial-contact networks in practice. The

case study we have presented here is just one such example.

Second, many options are possible to represent the key elements of data of interest: spatial, temporal, and contact networks. In our case study, we imposed contact information on the spatial structure. Researchers have also explored the opposite: adding spatial information on top of the social networks. The various ways of encoding and representing temporal information lead to further complications with representation selection. Insights and computational experience with these different representations, especially in a comparative context, are yet to be reported. Such comparisons could have interesting modeling and practical implications.

Third, principle-based models of disease transmission on spatial-contact networks are lacking. The standard epidemiological models such as SIR (susceptible, infectious, recovered) provide the references to guide empirical models and can produce predictions and decision support. Such models are typically confined to the temporal domain. What are the counterparts or extensions of these temporal models in the spatial domain and in the spatial-network domain? We believe that these models—yet to be worked out—may lead to exciting new developments in public-health informatics and epidemiology.

Fourth, given empirical data and possibly spatial-temporal-network models of infectious-disease transmission, how can actionable public-health

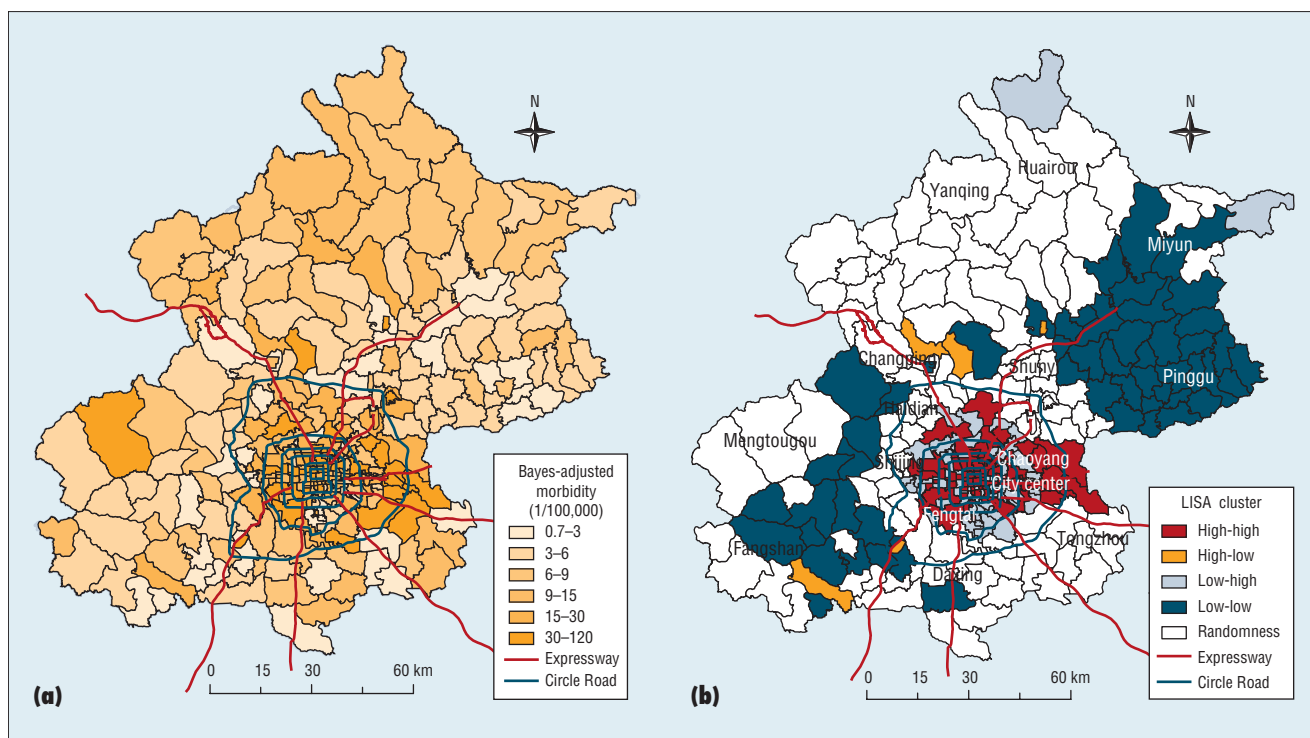


Figure 8. Morbidity and spatial clustering based the LISA statistic. (a) The first map shows the spatial distribution of Bayes-adjusted morbidities according to population. (b) The second map gives the outlines of four types of spatial clustering. We adopted spatial adjacent matrix-based distance; the threshold is the expected distance of 408 available links, 18.1 km.

response advice be generated? We suspect that analytical solutions could only help make such decisions in simple scenarios. A data-driven, real-time response framework may necessarily need AI-enabled human and group behavior simulation and experimental methods such as those based on agents. ■

Acknowledgments

Work reported in this article is supported by the US National Science Foundation (grants IIS-0839990 and IIS-0428241); the US Department of Homeland Security (grant 2008-ST-061-BS0002); the Chinese Ministry of Health (grant 2009ZX10004-315100); the Chinese Academy of Sciences (grants 2F07C01 and 2F08N03); the National Natural Science Fund of China (grant 60621001); and the China Postdoctoral Science Fund (grant 20080440559).

References

1. L.A. Meyers, M.E.J. Newman, and B. Pourbohloul, "Predicting Epidemics

- on Directed Contact Networks," *J. Theoretical Biology*, vol. 240, no. 3, 2006, pp. 400–418.
2. L.A. Meyers et al., "Network Theory and SARS: Predicting Outbreak Diversity," *J. Theoretical Biology*, vol. 232, no. 1, 2005, pp. 71–81.
3. W. Tobler, "A Computer Movie Simulating Urban Growth in the Detroit Region," *Economic Geography*, vol. 46, no. 2, 1970, pp. 234–240.
4. S.W.B. Newsom, "Pioneers in Infection Control: John Snow, Henry Whitehead, the Broad Street Pump, and the Beginnings of Geographical Epidemiology," *J. Hospital Infection*, vol. 64, no. 3, 2006, pp. 210–216.
5. L. Anselin, "Local Indicators of Spatial Association—LISA," *Geographical Analysis*, vol. 27, no. 2, 1995, pp. 93–115.

Daniel Zeng's and **Hsinchun Chen's** biographies appear at the end of the introductory essay.

Zhidong Cao is an assistant professor at the Chinese Academy of Sciences. Contact him at caozdong@gmail.com.

Fei-Yue Wang is the vice president of the Institute of Automation and a director and professor at the Key Laboratory of Complex Systems and Intelligence Science, Chinese Academy of Sciences. He is also affiliated with the University of Arizona. He is the Editor in Chief of *IEEE Intelligent Systems*. Contact him at feiyue@ieee.org.

Xiaolong Zheng is an assistant professor at the Chinese Academy of Sciences. Contact him at xiaolongzhengemail@gmail.com.

Quanyi Wang is a professor at the Beijing Center for Disease Control and Prevention. Contact him at wangquanyi@yahoo.com.

cn Selected CS articles and columns are also available for free at <http://ComputingNow.computer.org>.