
Learning to build network-oriented epidemic simulation models in epidemiology education

Ji-Lung Hsieh

Department of Computer Science,
National Chiao Tung University,
1001 Ta Hsueh Road,
Hsinchu 300, Taiwan, Republic of China
E-mail: jirlong@gmail.com

Chung-Yuan Huang*

Department of Computer Science and Information Engineering
and Research Center for Emerging Viral Infections,
Chang Gung University,
259 Wen Hwa 1st Road,
Taoyuan 333, Taiwan, Republic of China
Fax: +886-3-2118700
E-mail: gscott@mail.cgu.edu.tw
*Corresponding author

Chuen-Tsai Sun and Yu-Shiuan Tsai

Department of Computer Science,
National Chiao Tung University,
1001 Ta Hsueh Road,
Hsinchu 300, Taiwan, Republic of China
E-mail: ctsun@cs.nctu.edu.tw
E-mail: mathoowind@gmail.com

Gloria Yi-Ming Kao

Department of Information Communication,
Yuan Ze University,
135 Yuan-Tung Road,
Chung-Li 32003, Taiwan, Republic of China
E-mail: gloriakao@gmail.com

Abstract: Epidemic simulations and intervention strategy assessments are attracting interest in light of recent and potential outbreaks of infectious diseases such as SARS and avian flu. Universities are using computational modelling and simulation tools to teach epidemiology concepts to students, but integrating domain-specific knowledge and building network-based simulation models are difficult tasks in terms of teacher preparation and learner evaluation. To illustrate challenges to creating network-oriented models in epidemiology education, we introduce an architecture based on demographic and geographic data for building network-oriented epidemic simulation models, and describe our experiences simulating the transmission dynamics of three infectious diseases in Taiwan.

Keywords: computational modelling and simulation; network-oriented simulation models; epidemic dynamics; bipartite networks; social mirror identities; cellular automata; SARS; HIV; seasonal influenza.

Reference to this paper should be made as follows: Hsieh, J-L., Huang, C-Y., Sun, C-T., Tsai, Y-S. and Kao, G.Y-M. (2009) 'Learning to build network-oriented epidemic simulation models in epidemiology education', *Int. J. Simulation and Process Modelling*, Vol. 5, No. 1, pp.31–41.

Biographical notes: Ji-Lung Hsieh is a PhD candidate in Computer Science at National Chiao-Tung University, Taiwan. His research interests include complex network systems, agent-based modelling and network-oriented simulation approaches for social science research, computational epidemiology, information engineering, and digital game cultures.

Chung-Yuan Huang received his PhD in Computer Science (2005) from National Chiao Tung University, Taiwan. He is currently an Assistant Professor in the Department of Computer Science and Information Engineering and a member of the Research Center for Emerging Viral Infections at Chang Gung University, Taiwan. His research interests include complex adaptive networks and systems, agent-based modelling and network-oriented simulation approaches for social science research, computational epidemiology, computational systems biology, and evolutionary computations.

Chuen-Tsai Sun received his BS in Electrical Engineering (1979) and MA in History (1984) from National Taiwan University, and his PhD in Computer Science from the University of California at Berkeley (1992). He is currently a Professor in the Department of Computer Science, National Chiao Tung University. His research interests include creative evolutionary systems, web-based collaborative learning, and computer simulations for social science research.

Yu-Shiuan Tsai is a PhD student in the Department of Computer Science, National Chiao Tung University. His research interests include complex networks and systems, social science models, and mathematical analyses.

Gloria Yi-Ming Kao received her PhD in Computer Science (2008) from National Chiao Tung University, Taiwan. She is currently an assistant professor in the Department of Information Communication, Yuan Ze University. Her research interests include e-learning, creativity and mind-maps.

1 Introduction

Computational modelling and simulation approaches are widely used for researching emerging infectious diseases, social phenomena, and processes involving human interactions and daily contacts. These approaches have the advantages of speed, parallel processing, and the ability to work with large amounts of data over many repetitions. They are currently used in such domains as sociology (Newman, 2001, 2004; Axelrod, 1997; Gilbert and Troitzsch, 1999), economics (Born, 2003), bioinformatics (Barabasi and Oltvai, 2004; Ravasz et al., 2002), ecology (Grimm, 2005) and epidemiology (Bagni et al., 2002). Specific examples include the use of cellular automata by transportation management researchers to test traffic control strategies (Nagel, 2004; Koelle and Johenning, 1997); agent-based models constructed by economists to analyse price fluctuations, cash flows, and income and wealth distributions to identify instances of market manipulation (Siallagan et al., 2006); simulations run by sociologists to study the diffusion effects of emergent social institutions, innovations, norms, social moods, and collective behaviours (Gilbert and Troitzsch, 1999).

Instructors from a broad range of disciplines are collaborating with computer scientists to explore complex epidemic dynamics for purposes of training novice researchers. Until recently, epidemiology students have been required to master the use of questionnaires or field investigation techniques when studying epidemic outbreaks, a process that prevents many from gaining a macro view of epidemic dynamics or from quickly evaluating the efficacies of intervention strategies for prevention and control.

Today's students have access to computational modelling and simulation courses and applications. In addition, research centres are taking advantage of high-level object-oriented programming languages and agent-based modelling toolkits (e.g., *NetLogo*, *StarLogo*, *Repast* and *Swam*) to design and run epidemiological 'what-if' simulation experiments (Bagni et al., 2002).

Our focus in this paper is on teaching skills for building network-oriented epidemic simulation models that epidemiologists can use for such tasks as analysing spreading situations and outbreak patterns; predicting future transmission dynamics; assessing the efficacies of intervention strategies for disease prevention and control, vaccine development, and other efforts to fight the effects of viruses. We acknowledge that researchers and instructors in computational epidemiology are meeting unique roadblocks to solve complex problems involving movement and infection patterns among millions of people who span wide ranges of age, disease resistance status, and types and amounts of interaction with other individuals. Epidemic outbreaks can develop randomly and unexpectedly according to such factors as breadth of early stage outbreaks, number and type of random imported cases, an infected individual's current health status, and individual disease progress. Intervention strategies executed by health authorities also directly and indirectly influence transmission dynamics and spreading situations.

Network-oriented simulation models are proving successful for investigating epidemic dynamics and the efficacies of intervention strategies for prevention and control (Barrett et al., 2005; Hsieh et al., 2005; Schneeberger et al., 2004; Ferguson et al., 2005;

Huang et al., 2004, 2005b; Sumodhee et al., 2005; Stroud et al., 2007). By approaching computational epidemiology from a network science perspective, static domain information (e.g., the properties of individuals) can be represented in attribute form using network nodes. Dynamic domain information regarding individual factors such as movement, contact and interaction requires support in the form of statistical or demographic data plus well-constructed assumptions. In previous studies, we have examined differences and similarities among network-oriented epidemic simulation models that focus on local interaction patterns, and have applied our epidemiological modelling experiences to simulating the transmission dynamics of SARS, influenza, and HIV in Taiwan. Our goals in this paper are to propose an architecture based on demographic and geographic properties, to discuss ways of teaching computational modelling and simulation approaches to undergraduate students and novice researchers, and to describe three scenarios in which we either taught or helped construct network-oriented epidemic simulation models.

2 Background

Three practice categories reflect different combinations of instructional goals, available hardware and software, and human-computer interaction. Simulations on computer systems entail the use of expert system-based simulation software for practicing specific skills; examples include flight and vehicle simulators (Colpitts, 2002; Rolfe and Staples, 1988). Computer-mediated simulations provide operational environments in which human participants solve problems, play roles and negotiate with or compete against other participants; examples include market simulation environments that allow users to practice investing strategies without using cash resources (Sato et al., 2001). All-in-computer (also called agent-based) simulations have system components (object or agents) that represent computer entities; human participation is limited to setting initial parameters and controlling simulation processes via automatic or manual sequential instruction. Most of the examples in this paper belong to the third category. All-in-computer simulations can be further classified according to simulated target type:

- modelling that simulates the flow paths and patterns of complex issues and verifies simulation models using actual data (e.g., economic decision modelling)
- agent-based modelling that focuses on heterogeneous individuals, changes in individual status, and interactions between individuals in simulated environments (e.g., SARS simulation (Huang et al., 2004, 2005b))
- adaptive agent-based modelling, in which automatic agents develop their own intelligent rules (e.g., adaptive agents that use genetic algorithms to adjust economic strategies (Holland, 1992)).

2.1 Network-oriented simulations

Compared with heterogeneous agent-based models, network-oriented simulation models place greater emphasis on relationships between individuals. In complex networks, nodes represent individuals and links represent their various relationship types. Results from mathematical analyses and computer simulations show that the topological features of social networks exert considerable influence on both transmission dynamics and spreading situations associated with epidemics, which allows for analyses of subtle details that non-network-oriented approaches (e.g., system dynamics) are incapable of performing (Barrett et al., 2005; Huang et al., 2004, 2005a, 2005b; Moore and Newman, 2000; Sumodhee et al., 2005; Kao et al., 2006). Furthermore, the need to identify targeted and more efficient intervention strategies requires accurate model representations of public health policies with geographic properties (e.g., home quarantines and hospital visitation bans).

Complex network topologies – including scale-free networks, small-world networks, random networks, regular networks, and bipartite networks – have been observed in gene transcriptions (Eisenberg and Levanon, 2003), food webs (Dunne et al., 2002), research paper citation networks (Davenport and Cronin, 2000; Small, 1995), opinion exchanges (Grabowski and Kosinski, 2006) and human daily contact networks (Miramontes and Luque, 2002; Huang et al., 2005b; Sumodhee et al., 2005). Such complex networks frequently reveal statistical topology features such as high degree of clustering, low degree of separation, and power-law connectivity distributions.

Network-oriented simulation models entail computer entities that imitate patterns of social interactions and daily contacts between individuals and that use parallel or sequential processes for the movement of heterogeneous individuals. Regular networks have been created for discussing human relationships. For example, the von Neumann neighbourhood concept (used with two-dimensional regular networks) perceives any node's four adjacent nodes as its neighbours. Using a regular network for an underlying network topology allows for the easy measurement of abstract geographic distances between individuals. In contrast, random networks support the features of casual interactions between individuals and the low degree of separation found in complex networks. However, random networks have an important drawback in that they cannot ensure the topological feature of local interactions.

Network-oriented simulations have been created to capture the properties of geographical distance and randomness. Many small-world approaches to building complex simulation models have been proposed – for example, the short-cut model (Newman and Watts, 1999) and the Cellular Automata with Social Mirror Identities Model (CASMIM) (Huang et al., 2004, 2005b). The mirror identity concept used to create CASMIM allows for simulations of individuals who have regular contact

locally and intermittent contact with others over far distances; this is representative of long-distance air travel or (in non-disease scenarios) modern communication methods such as e-mail, text-messaging, express mail, and cell phones. For purposes of model validation, the building and modification of these concepts should correspond to statistical topology features such as small-world phenomena.

However, different complex issues require different network topologies for building simulation models based on social interaction or daily contact type. For example, heterosexual contact, homosexual contact, and illegal drug use all cause HIV infections, but the network topologies formed by these approaches are very different. The heterosexual network should be scale-free to reflect the lower average number of sex partners, whereas bipartite networks are more likely to accurately reflect the sharing of syringes among drug users. Data granularity and detail also influence topological structure: if a disease control agency can trace drug user contacts, it is possible to use identified matches to build a relatively precise bipartite contact network. In contrast, for researchers who are limited to knowledge of locations where drug users congregate, simulations require more assumptions regarding interactions. Accordingly, demographic data and information on individual interactions tend to be significant challenges to novice researchers interested in complex network computer simulations.

2.2 Learning through simulations

Using computer simulations as a pedagogical tool is now common in scientific technology training (Colpitts, 2002) and the teaching of science concepts (Liao and Sun, 2001). Computer simulations are also being used in other disciplines to support educational and training efforts based on constructivist learning principles. In addition to mitigating learner obsession with the minutiae of complex procedures described in textbooks (Wenglinsky, 1998), they provide multiple opportunities for “learning by doing” (Oehme, 2000). Constructivists believe that learners draw upon prior knowledge when forming new schema via discovery learning (Bruner and Lufburrow, 1963). When confronted with a new stimulus, learners apply their own knowledge bases to accommodate new information and to alter their existing schema (Piaget, 1978). When constructive learning processes are embedded in simulation tools, students can learn by doing, have more and better opportunities for discovering interesting primary and secondary issues, and gain hands-on experience for dealing with real-world issues.

Learning via simulation is an example of Problem-Based Learning (PBL) in that it confronts learners with authentic problems that serve as contexts for practice. Originally developed for medical education in the early 1970s, PBL is now considered a core teaching tool in over 60 medical schools (Savery and Duffy, 1995). Two characteristics make PBL compatible with the

theoretical foundations of learning and teaching via simulations:

- *Engagement.* Students often ask for simulations to assist with learning and to gain a sense of engagement with real-world problems. This allows for the introduction of related concepts to the learning process. There is no ‘perfect’ educational simulation, but simulations can still support meaningful learning experiences as long as scenario limitations are taken into account (Aldrich, 2004).
- *Interaction flexibility.* Simulation tools can be used with interaction and feedback methods to illustrate how complex systems work under different circumstances (Aldrich, 2004). Simulated problems are usually complex and rarely have single ‘correct’ answers, which encourages learners to repeatedly manipulate parameters. With sufficient practice, learners or novice researchers can learn how to transfer their new knowledge to real-world issues.

Learning via computer simulations has at least four potential advantages:

- *Operational.* Complex problems often require examinations of the influences of specified variables on an entire system. Using an epidemic outbreak as an example, epidemiologists may want to measure the potential impacts of individual health policies, but it is impossible to do so when running real-world experiments. Researchers and students can examine the influences of different variables on complex systems and execute ‘what-if’ experiments to study the emerging behaviours of such systems while temporarily neglecting irrelevant variables. In short, simulations can be optimised for learning (Bertsche et al., 1996).
- *Observational.* Users can take simulation processes and slow them down (e.g., to study traffic jams) (Burmeister et al., 1997), speed them up (e.g., to study the propagation of diseases such as HIV that have long incubation periods) (Sumodhee et al., 2005), or simply adjust their scales for observation purposes. Computer simulations are recognised as an efficient approach to review or prove textbook concepts. Post-simulation reports allow teachers to determine which concepts their students have mastered (Hargrave and Kenton, 2000).
- *Bridging reality and theory.* Simulations allow novices to practice professional skills (e.g., aircraft piloting) and to practice or prove textbook concepts (Hargrave and Kenton, 2000); this protects them from having to jump into high-risk situations for learning purposes. Investment students can use simulations to practice theoretical concepts without having to invest large amounts of capital (Levy et al., 1995; Klein et al., 2004).

- **Construction.** Computer simulation tools can be used to create, explore, or construct environments. Using health policy assessments as an example, learners can practice predicting potential developments that may result from different combinations of strategies or interaction rules. When combined with instruction-based learning, teachers can exert relatively precise control over knowledge construction and accumulation (Hargrave and Kenton, 2000).

The learning processes and goals associated with learning via computer simulations differ from those of traditional classroom and textbook-centred learning. Scenarios used for computer simulations are often ill-defined and open-ended (Hsieh et al., 2005), and problems frequently arise after a simulation starts. Novices are therefore required to use instruction-based manuals to build or run simulations and to create professional-quality reports or presentations of their results. Teacher preparation time varies depending on the amount of required background, scenario construction, and necessary instruction to help learners formulate problem statements, collect data, run simulations, and create reports. Evaluative techniques for learning results also differ from those used in traditional classroom settings and require some training on the part of instructors. On the basis of our past experience in teaching epidemic outbreak principles through simulations, we have written comprehensive teaching instructions that we will discuss in a later section.

3 Framework for analysing network topologies in computational modelling and simulations

Instead of using mathematical equations to generate an analytical network model, complex issue simulations rely more on imitating real-world interaction patterns (Barrett et al., 2005; Huang et al., 2004, 2005b; Stroud et al., 2007). Two issues associated with building network-oriented simulation models must be considered: choosing the appropriate complex network model type and integrating professional knowledge of complex issues into the system. Several types of network-oriented simulation models can be implemented, including real contact networks that trace individual contacts (Barrett et al., 2005; Stroud et al., 2007), bipartite networks composed of media and individual layers (Kao et al., 2006), regular lattices with shortcuts that reflect small-world phenomena and that are suitable for modelling opinion transmission (Newman and Watts, 1999) and regular lattices with mirror identities for simulating regular long-distance movement (Hsieh et al., 2005, 2006; Huang et al., 2004, 2005b). However, novice learners tend to feel challenged in terms of domain knowledge and complex network theorems, including the need to identify real contact interaction patterns, choosing or adopting existing network simulation models, building new models and validating simulation models.

Individual characteristics and behaviours determine domain knowledge and properties embedded in complex network simulation models; here we will offer

several examples. In Table 1, we present a list of individual attributes used in epidemic disease simulation. Opinion communication and degrees of propagation acceptance and active transmission are determined via the persistence shown by individuals for holding onto their opinions. Sexual and injecting drug use behaviours lead to different types of HIV propagation networks. Finally, the use of different media (e.g., telephone, e-mail, mass media, instant messenger software) also results in differences in contact or communication network type.

Table 1 Common individual attributes in epidemic simulations

Attribute	Description	Default value
ID	Unique serial number identifying 1 ~ P each CASMIM agent	
E	Epidemiological progress of each individual	Susceptible
Mobility	Denotes individual activity. Default value is 'free'—i.e., no restrictions on interacting with the mirror identities of neighbouring agents	Free
Age	Agents categorised as young (1 to 20), prime (21 to 60), or old (61 and above)	Young, Prime, Old
Super	Denotes whether an agent is a super-spreader	True, False
Immunity _{Permanent}	Denotes whether an agent is permanently immune	True, False
Day	Number of days of a patient's recorded disease status	
Rate _{Contact}	Rate of contact with other agents	0 ~ 1

The three properties considered most important for simulation network construction are:

- **Time scale.** In the case of HIV propagation via heterosexual behaviour, frequency distributions of sexual behaviour over one month or one year will show power-law characteristics (Schneeberger et al., 2004), but this is not true when the time scale is shortened to 1 day. It is also important to remember that different diseases have different incubation periods – five days for SARS vs. six months to 20 years for HIV.
- **Geographical properties.** Flu epidemics tend to be large-scale, therefore models for countries that have multiple regions require the consideration of inter-city transportation networks. Building a contagion model for any modern city with an established mass transportation system must assume a strong and varied mix of human movement, which can affect considerations of inter-regional transportation. In previous studies, we used regular lattices with mirror identities for simulations of SARS outbreaks in Singapore, Taipei and Toronto (Huang et al., 2004, 2005b).

- **Robustness and ability to expand.** Owing to the diversity of data collected for complex issues, simulations of specific cases often require modifications to selected simulation networks. For example, the CASMIM model is suitable for simulating a well-mixed underlying environment with randomly distributed mirror identities, but assumptions of well-mixed distributions of locations for sexual contact or needle sharing cannot be supported. Therefore, when applying an existing simulation network to other complex issues, model robustness is a concern due to data diversity and changes in demographic structure.

We built a four-layer architecture based on geographic and demographic properties, and found that simulation network model choices are highly data-dependent. The Lv1 model in Table 2 reflects the use of real contact tracing for interaction network construction. For example, during the Asian SARS outbreak, health authorities in Taiwan and Singapore attempted to construct contact histories for all infected individuals so as to quarantine anyone who had been in contact with a carrier. The Lv2 model consists of bipartite matching networks of individuals and interacting media. In this case, saunas and bars frequented by homosexuals can be viewed as media that bridge individuals; for illegal drug users, media include syringes and chemicals used for drug dilution. To construct bipartite networks, researchers must determine how many times a user shares a syringe with others during one month or how many users share the same diluting agent in a single session.

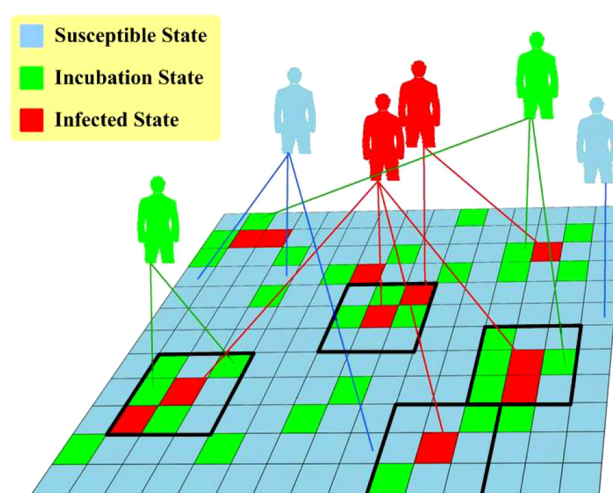
Table 2 Network-oriented simulation model architecture

Layer	Description	Example
Lv1	Individual-to-individual real connection	E-mail networks, research paper citations, patent citations
Lv2	Individuals passively connected by content	BBS reply networks, research paper co-authoring
Lv3	Statistical topology features (e.g., low degrees of separation or power-law connectivity distribution) and abstract geographical properties (e.g., neighbourhood concept)	CASMIM (Huang et al., 2004, 2005b), BBS, or game community
Lv4	Physical geographic properties; often applied to demographic statistics	South Asian flu simulation (Ferguson et al., 2005); Portland smallpox simulation (Barrett et al., 2005); EpiSimS simulation (Stroud et al., 2007)

The Lv3 model reflects individual relationship networks with abstract geographical properties. In the absence of real contact information, researchers may need to build well-mixed modern societies using abstract distances. In a previous study (Huang et al., 2004, 2005b), we proposed a social mirror identities model (CASMIM) consisting of

a 2-D lattice with cellular automata as a basic underlying network for retaining an abstract individual's geographical mobility (Figure 1). The mirror identity concept utilises simple social networks to preserve the properties of agents that interact with their neighbours within two-dimensional lattices, and reflect such activities as long-distance movement and daily visits to fixed locations. The main characteristic of models in this layer is their abstract geographical property, which allows for observations of emerging complex issues as well as for simulating geographical policies or strategies.

Figure 1 Cellular automata and social mirror identity model (CASMIM) (see online version for colours)



The Lv4 model consists of individual relationship network models based on demographic data with geographical properties. Building network-oriented simulation models at this layer often requires significant support in the form of demographic data. For example, Ferguson's Southeast Asian flu simulation (Ferguson et al., 2005) uses data for population density, household size, age distribution, school and workplace size, and individual travel data. In contrast, the spread of HIV among homosexuals serves as a negative example: movement, location, and means of sexual contact are less obvious, making it more difficult to build a fourth layer simulation model (Sumodhee et al., 2005). We will describe a similar flu simulation case in Section 4.

In conclusion, three points must be considered when selecting a network-oriented simulation model according to our proposed architecture:

- **Issue diversity.** The Lv1 model can be used for a patent citation network simulation because each patent belongs to only one agency (in other words, there is an absence of multiple authorities), and each citation can be placed along a specific timeline. This reflects the patent citation research emphasis on how patent holders cite each other without intervening media. For academic paper citation networks, if the focus is on how often researchers cite each other, a simulation network model can be chosen from the first layer; if the focus is on co-authorship, research papers must be treated as media in a bipartite (Lv2) model.

- *Data-dependency.* We mentioned earlier that data granularity or type determines the method for building a network simulation model. Using homosexual HIV transmission as an example, a situation in which data are limited to sexual contact frequency requires a Lv3 model utilising the neighbourhood concept and abstract geographical distances (Sumodhee et al., 2005). However, if movement within a high-risk contact population can be traced, a Lv1 model can be used to create a simulation and to predict further development.
- *Scale.* Care must be taken in selecting a proper geographic scale. Whereas CASMIM can be applied to simulate epidemic outbreaks in modern societies, simulating multi-city transmission dynamics requires additional demographic data. One possible solution is building a separate CASMIM for each city and measuring transportation flow between paired cities.

4 Three computational epidemic simulation cases

In this section, we will share our experiences building epidemic simulation models for three of the four layers described earlier.

4.1 Bipartite simulation model: drug users

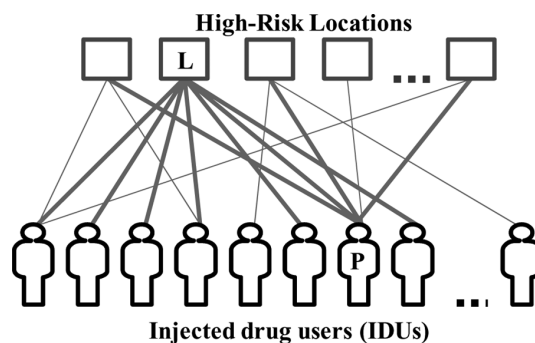
Introduction. According to the AIDS Prevention and Research Centre of National Yang-Ming University, Injecting Drug Users (IDUs) now constitute the largest population of HIV-infected individuals in Taiwan (Sumodhee et al., 2005). Owing to our success in simulating the SARS outbreak of 2004, we were asked to build an HIV simulation model for IDU inmates in three jails located in central and northern Taiwan.

Data collection. According to responses to a participant survey, each IDU shared a chemical for drug dilution an average of 2–3 times per month and shared it with 2–3 persons on each occasion. The Taiwan Centers for Disease Control (CDC) have collected data on the country's HIV-positive population between 1984 and September 2006. In terms of age group, teenagers constitute 3% of the HIV-positive population, adults 73%, and elders (defined as 60 years or older) 24%. The mortality rate for HIV in Taiwan is 13%. Less than one-quarter of all identified HIV carriers during this period eventually became AIDS patients (2,398/10,158, or 23.6%); the mortality rate for these patients was 43%. We used this data to build a two-layer bipartite network for modelling individual disease status.

Model building. Since we only had access to data on the average number of syringe-sharing events per month and average number of persons sharing either syringes or a chemical dilutant during each occasion, we treated users who shared drug paraphernalia as neighbours when creating a model based on CASMIM's mirror identity concept. As part of our one-month simulation model, we used the

average number of syringe-sharing experiences per month as our mirror identity value. However, actual IDU gathering places are spread throughout the country as well as throughout individual cities. Therefore, the spread of HIV among different locations has no effect on local spreading. Note that the original CASMIM incorporates the effect of local spreading, meaning that the disease can still spread to any other location via transmission between local neighbours in the absence of mirror identities or shortcuts. Since these high-risk locations are not adjacent, it is inappropriate to use CASMIM or the mirror identity concept to simulate HIV transmission among IDUs. By collecting data regarding the distribution of the number of syringe and dilutant-sharing events plus the distribution of gathering locations, we were able to build a bipartite model without geographical properties. The bipartite relationship between high-risk locations and IDUs is shown in Figure 2. The individual labelled 'P' visited more places than the other IDUs; location 'L' was a frequently visited gathering place.

Figure 2 Bipartite relations among injecting drug users (IDUs) and their meeting locations



Discussion. Whereas we were able to obtain data on the distribution of shared syringe and dilutant event numbers, we could not collect data on the distribution of numbers of IDUs visiting each location. We assumed that this reflects a power-law distribution, but we had insufficient data for model validation – especially since the reported number of infected individuals may not have contained concealed cases. This example shows how difficult it is to build a simulation model without sufficient support in the form of demographic data for building local interaction rules.

4.2 Abstract geographical model: SARS

Introduction. Students used CASMIM to create simulations of the 2004 SARS outbreak in Taipei, Singapore and Toronto (Huang et al., 2004, 2005b). They were also asked to assess the efficacies of different combinations of public health policies to control the spread of SARS.

Data collection. Our epidemiological data focused on incubation period, infectious period, mortality rate, and infection rate. In terms of building an underlying simulation network, these parameters address the age distribution and number of mirror identities of individuals as well as the neighbourhood concept. Input data also addressed the

ratios and numbers of normal and super-spreaders as well as the numbers, entry dates and disease statuses of imported cases.

Model building. We used CASMIM (Huang et al., 2004, 2005b) to conceptualise individual daily contact networks (Figure 1). Since SARS is a short-distance contagious disease, we defined one time step as equivalent to one day in the real world. The model consists of two layers: the lower layer uses 2D cellular automata to represent the activity space as a social mirror identity, and the upper layer represents individuals with several heterogeneous identities in the real world. Each individual in the upper layer can be abstracted as a single agent in a multi-agent system with different social mirror identities; those identities reflect geographic locations that the agent visits on a regular basis (e.g., homes, workplaces, and transportation stations). Interactions between individuals represent infection routes.

Discussion. As a model for individuals with different social mirror identities, CASMIM has an important flaw when discussing epidemic dynamics: individuals cannot exist in all mirror identities at the same time. A possible response to this flaw is partitioning days into time segments based on the length of time spent in each identity. Accordingly, the large majority of time will be partitioned for home and workplaces for adults, and home and school for children.

4.3 Physical geographic model: flu outbreak

Introduction. Since 2006, we have been building a multi-city influenza simulation for all of Taiwan. In the absence of data on past outbreaks, we have placed greater emphasis on predicting the efficacies of public health policies and potential vaccines.

Data collection. We used inter-city transportation network data to model inter- and intra-city contact between individuals – specifically, statistics on vehicle flow between pairs of cities along the country's National Highway 1. A vehicle flow matrix was created to estimate the movement of Taiwanese along the highway. Demographic data published by the Taiwanese government were used to assign individuals to various locations. The data included the distributions of numbers of individuals per household in each county, numbers of employees in workplaces and numbers of students in classrooms. The data were combined to achieve an approximate understanding of the overall distribution of numbers of persons in each location. As shown in Figure 3, most locations had fewer than ten; exceptions included movie theatres and classrooms, with 40 or more.

Model building. On the basis of our SARS experience, we knew that CASMIM is suitable for simulating contact problems in well-mixed but not in poorly mixed cities. We therefore assigned a separate CASMIM to each city, with model size determined by city population. Each CASMIM cell represented a real-world household, workplace, or classroom that contained several mirror

identities. The number of individuals in each cell was assigned according to the distribution of numbers of persons in a location. Transportation flow statistics were used to represent inter-city movement by individuals, with each movement represented by a pair of mirror identities in different cities but belonging to the same individual. For the majority, mirror identities were allocated to the same city (Figure 4). Since the incubation period for influenza is only 1–3 days, the simulation time parameter was set to one day.

Figure 3 Statistical distribution of number of persons in regularly visited locations such as households, workplaces, and classrooms

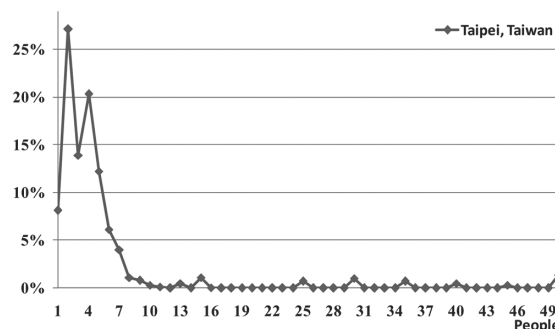
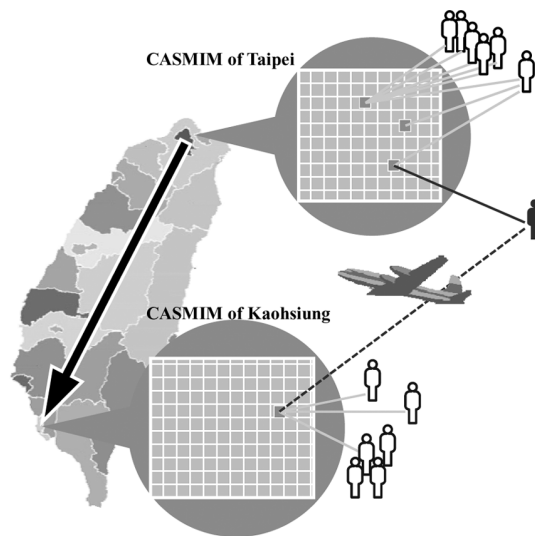


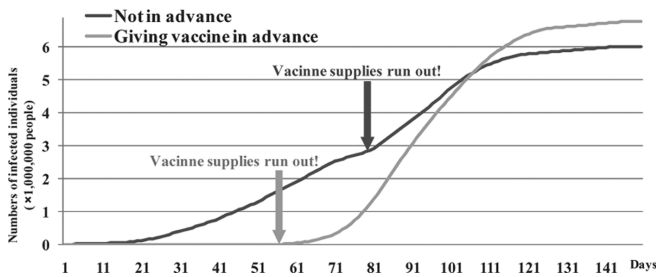
Figure 4 Multiple cellular automata with intra- and inter-city social mirror identities. Each mirror identity represents inter-city movement



Discussion. One weak point of this multi-city CASMIM was its lack of complex network statistical properties for a model consisting of multiple poorly mixed cities. Unlike a well-mixed modern city, one cannot simply assume that the contact network of individuals distributed among multiple cities has small-world properties. Furthermore, there is a lack of data on recent outbreaks for empirical validation. Some researchers have constructed simulations of the 1918–1919 influenza outbreak (Sattenspiel and Herring, 2003), but we believe that the topological structure of modern cities is far different from that observed in 1918. Despite these weaknesses, we used the model to predict the efficacy of a

vaccine and related prevention strategies – for instance, giving the vaccine to all residents living in high-risk areas in advance of an initial outbreak. Our simulation data indicate that vaccine supplies might run out very early, resulting in a more serious outbreak (Figure 5).

Figure 5 Results from using a multi-city CASMIM to assess a prevention strategy in which all individuals thought to be susceptible to an epidemic disease are vaccinated. When the vaccine was given to all residents living in high-risk areas, supplies ran out on day 57. When given to susceptible patients only, supplies ran out on day 78



4.4 Teaching computational modelling and simulation

Owing to the properties of complex issues (occasionally non-linear, random, and ill-conditioned) and the large amount of required domain knowledge from multiple disciplines, network-oriented training simulations present challenges in terms of problem design, preparation and introduction of course material, student participation, and evaluation. Since a large amount of background knowledge is required (Hargrave and Kenton, 2000), we suggest using

- pre-instructional roles to teach public health policy assessment and epidemic outbreak trend prediction skills
- post-instructional roles to teach simulation model building and analysis. Both are appropriate for learning-by-doing experiences.

We designed a teaching schedule consisting of six steps:

- introducing background events and complex issue knowledge
- introducing professional domain knowledge of a complex issue and computer science techniques
- preparing a pre-test for guiding students to key points of an issue
- creating step-by-step instruction-based simulations with appropriate test cases using sample data, user manuals for operating simulation models, and experiment design examples

- unrestrained operating time, which allows students to construct and develop their own experiments
- post-tests or final presentations to evaluate student understanding of an issue.

Since complex issues often have no single or absolute approach, it is difficult to evaluate how well novice learners understand the operational aspects of simulating them. One potential solution is to design constructive pre-tests and post-tests. Using public health policy simulations as an example, novice learners may be asked to compare the efficacies of different combinations of health policies before and after a simulation is performed. In addition, we have observed that novice learners exhibit wide differences in terms of controlling simulation parameters (Hsieh et al., 2006), and therefore suggest that parameters be used as an evaluation criterion.

When using the SARS scenario as a teaching example, our participants were 34 college students recruited from a private university in northern Taiwan. They were assigned to working pairs for collaboration and for discussing simulation results. The 17 participant pairs were given three assignments. For each assignment, they were given a pre-test to examine their understanding of epidemic dynamics, verbal and written information on simulation goals, tools for using simulation results to answer core questions, a user manual and three sets of sample data for running SARS simulations using CASMIM, and a post-test to determine the effects of the simulation tool on learning. Copies of the pre-test, user manual, sample data, and post-test can be found in Hsieh et al. (2006).

All of the participants were undergraduate students in the school's computer science department; most had no background knowledge of epidemiology principles. Each pair was given a user manual and dataset for running the simulation. After the simulations were run, we evaluated epidemiological knowledge accumulation. For example, we presented students with a list of 12 parameters (half of them required for running an epidemic simulation and the other half not required) and asked them to categorise the parameters. Results from a paired-sample *t*-test indicate statistically significant improvement in the students' overall understanding of epidemic concepts (Table 3). A second question for which results were statistically significant was, "Assess the efficacies of different combinations of public health policies". Significant differences were not found for the other evaluation questions. The results indicate that running a simulation can help students learn background knowledge and compare different public health policies. However, the participants had difficulty verifying parameter sensitivity.

Table 3 Statistical results for pre-tests and post-tests when teaching CASMIM Principles to 34 undergraduate students

Question set	Pre-test score		Post-test score		t-test	p-test
	M	SD	M	SD		
Set 1: Understanding of epidemic concepts	31.71	3.11	35.15	2.77	-5.36	$p < 0.001$
Set 2: Comparing simulation and actual case results for outbreaks in Singapore and Taiwan	6.12	1.01	6.91	1.44	-2.17	$p < 0.05$
Set 3: Assessing efficacies of different combinations of public health policies	6.59	1.37	8.41	0.71	-5.12	$p < 0.001$
Set 4: Analysing effects of protective mask wearing by the general public	6.24	1.15	6.53	1.23	-1.32	
Set 5: Analysing effects of mask wearing by healthcare workers	6.76	1.09	7.06	0.66	-1.32	
Set 6: Estimating social costs of different combinations of public health policies	6.00	1.37	6.24	1.15	-1.00	

5 Conclusion

In this paper, we proposed a pre-analysis architecture of network-oriented simulation models for the benefit of teachers and novice learners, and illustrated architecture reduction and extension in terms of collected geographic and demographic data. In addition, descriptions of three sample cases were given to illustrate the process of building a network-oriented simulation model. Our experiences of teaching epidemiology concepts via network-oriented simulations have allowed us to identify three challenges for instructors: choice of a suitable simulation model, preparation for instruction-based teaching and evaluating student understanding.

Unlike traditional computer systems that simulate scientific tools, special skills, or natural phenomena, network-oriented simulation models for solving complex issues require more demographic data support and greater amounts of initial domain knowledge. In other words, most complex issues require collaborations among scientists from multiple disciplines. Regarding student achievement evaluation, we acknowledge that the traditional method of pre-tests and post-tests is insufficient for computational modelling and simulation courses; however, in one situation we found that students did learn basic epidemiology concepts even though they had difficulty using a complex network model to perform high-level assessments of public health policies.

Acknowledgements

This work was supported in part by the Republic of China National Science Council (grant No. NSC 97-2221-E-182-046), Chang Gung University (No. UERPD270281) and Chang Gung Memorial Hospital (No. CMRPD260022).

References

- Aldrich, C. (2004) *Simulations and the Future of Learning: An Innovative (and Perhaps Revolutionary) Approach to e-Learning*, Pfeiffer, San Francisco, CA.
- Axelrod, R. (1997) 'Advancing the art of simulation in the social sciences', *Complexity*, Vol. 3, No. 2, pp.16–22.
- Bagni, R., Berchi, R. and Cariello, P. (2002) 'A comparison of simulation models applied to epidemics', *Journal of Artificial Societies and Social Simulation*, Vol. 5, No. 3, <<http://jasss.soc.surrey.ac.uk/5/3/5.html>>
- Barabasi, A.L. and Oltvai, Z.N. (2004) 'Network biology: understanding the cell's functional organization', *Nature Reviews Genetics*, Vol. 5, No. 2, pp.101–113.
- Barrett, C.L., Eubank, S.G. and Smith, J.P. (2005) 'If smallpox strikes Portland', *Scientific American*, Vol. 292, No. 3, pp.42–49.
- Bertsche, D., Crawford, C. and Macadam, S.E. (1996) 'Is simulation better than experience', *The McKinsey Quarterly*, Vol. 1, No. 1, pp.50–58.
- Born, R.G. (2003) 'Teaching discrete event simulation to business students: the alpha and omega', *Simulation Conference, Proceedings of the 2003*, Winter, Piscataway, New Jersey, Vol. 2, pp.1964–1972.
- Bruner, J.S. and Lufburrow, R.A. (1963) 'The process of education', *American Journal of Physics*, Vol. 31, pp.468.
- Burmeister, B., Haddadi, A. and Matylis, G. (1997) 'Application of multi-agent systems in traffic and transportation', *Software Engineering, IEE Proceedings*, Vol. 144, No. 1, pp.51–60.
- Colpitts, B.G. (2002) 'Teaching transmission lines: a project of measurement and simulation', *Education, IEEE Transactions on*, Vol. 45, No. 3, pp.245–252.
- Davenport, E. and Cronin, B. (2000) 'The citation network as a prototype for representing trust in virtual environments', *The Web of Knowledge: A Festschrift in Honor of Eugene Garfield*, Metford, Information Today Inc., ASIS Monograph Series, NJ, pp.517–534.
- Dunne, J.A., Williams, R.J. and Martinez, N.D. (2002) 'Food-web structure and network theory: the role of connectance and size', *Proceedings of the National Academy of Sciences*, Vol. 99, No. 20, pp.12917–12922.
- Eisenberg, E. and Levanon, E.Y. (2003) 'Preferential attachment in the protein network evolution', *Physical Review Letters*, Vol. 91, No. 13, p.138701.
- Ferguson, N.M., Cummings, D.A.T., Cauchemez, S., Fraser, C., Riley, S., Meeyai, A., Iamsirithaworn, S. and Burke, D.S. (2005) 'Strategies for containing an emerging influenza pandemic in Southeast Asia', *Nature*, Vol. 437, No. 7056, pp.209–214.

- Gilbert, G.N. and Troitzsch, K.G. (1999) *Simulation for the Social Scientist*, Open University Press, Philadelphia, PA.
- Grabowski, A. and Kosinski, R.A. (2006) 'Ising-based model of opinion formation in a complex network of interpersonal interactions', *Physica A: Statistical Mechanics and Its Applications*, Vol. 361, No. 2, pp.651–664.
- Grimm, V. (2005) *Individual-Based Modelling and Ecology*, Princeton University Press, Princeton, NJ.
- Hargrave, C.P. and Kenton, J.M. (2000) 'Preinstructional simulations: implications for science classroom teaching', *Journal of Computers in Mathematics and Science Teaching*, Vol. 19, No. 1, pp.47–58.
- Holland, J.H. (1992) *Adaptation in Natural and Artificial Systems*, MIT Press Cambridge, MA, USA.
- Hsieh, J.L., Huang, C.Y., Sun, C.T. and Chen, Y.M.A. (2005) 'Using the CAMIM small-world epidemic model to analyze public health policies', *Western Multiconference on Health Sciences Simulation*, New Orleans, LA, USA, pp.63–69.
- Hsieh, J.L., Sun, C.T., Kao, G.Y.M. and Huang, C.Y. (2006) 'Teaching through simulation: epidemic dynamics and public health policies', *Simulation*, Vol. 82, No. 11, pp.731–759.
- Huang, C.Y., Sun, C.T., Hsieh, J.L. and Lin, H. (2004) 'Simulating SARS: small-world epidemiological modelling and public health policy assessments', *Journal of Artificial Societies and Social Simulation*, Vol. 7, No. 4, <<http://jasss.soc.surrey.ac.uk/7/4/2.html>>
- Huang, C.Y., Sun, C.T. and Lin, H.C. (2005a) 'Influence of local information on social simulations in small-world network models', *Journal of Artificial Societies and Social Simulation*, Vol. 8, No. 4, <<http://jasss.soc.surrey.ac.uk/8/4/8.html>>
- Huang, C.Y., Sun, C.T., Hsieh, J.L., Chen, Y.M.A. and Lin, H. (2005b) 'A novel small-world model: using social mirror identities for epidemic simulations', *Simulation*, Vol. 81, No. 10, pp.671–699.
- Kao, R.R., Danon, L., Green, D.M. and Kiss, I.Z. (2006) 'Demographic structure and pathogen dynamics on the network of livestock movements in Great Britain', *Proceedings of the Royal Society B: Biological Sciences*, Vol. 273, No. 1597, pp.1999–2007.
- Klein, C.A., Berlin, L.S., Kostolansky, T.J. and Del Palacio, J.R. (2004) *Stock Simulation Engine for an Options Trading Game*, United States Patent 6709330.
- Koelle, H.H. and Johanning, B. (1997) *Space Transportation Simulation Model (TRASIM 2.0)*, ILR Mitt. 319, Technische Universität Berlin, Germany.
- Levy, M., Levy, H. and Solomon, S. (1995) 'Microscopic simulation of the stock market: the effect of microscopic diversity', *Journal de Physique I*, Vol. 5, pp.1087–1107.
- Liao, Y.H. and Sun, C.T. (2001) 'An educational genetic algorithms learning tool', *IEEE Transactions on Education*, Vol. 44, No. 2, p.20.
- Miramontes, O. and Luque, B. (2002) 'Dynamical small-world behavior in an epidemical model of mobile individuals', *Physica D*, Vol. 168, No. 169, pp.379–385.
- Moore, C. and Newman, M.E.J. (2000) 'Epidemics and percolation in small-world networks', *Physical Review E*, Vol. 61, No. 5, pp.5678–5682.
- Nagel, K. (2004) 'Multi-agent transportation simulation', *Traffic*, Vol. 2, p.2.
- Newman, M.E.J. (2001) 'Who is the best connected scientist? A study of scientific coauthorship networks', *Physics Review*, Vol. 64, No. 4, pp.046131–046132.
- Newman, M.E.J. (2004) 'Coauthorship networks and patterns of scientific collaboration', *Proceedings of the National Academy of Sciences of the United States of America*, Vol. 101, pp.5200–5205.
- Newman, M.E.J. and Watts, D.J. (1999) 'Renormalization group analysis of the small-world network model', *Phys. Rev. E*, Vol. 60, pp.7332–7342.
- Oehme, F. (2000) 'Learn by doing: how to include new requirements of research in engineering education', *European Journal of Engineering Education*, Vol. 25, No. 2, pp.131–137.
- Piaget, J. (1978) *The Development of Thought: Equilibration of Cognitive Structures*, Viking Press, New York, NY.
- Ravasz, E., Somera, A.L., Mongru, D.A., Oltvai, Z.N. and Barabási, A.L. (2002) 'Hierarchical organization of modularity in metabolic networks', *Science*, Vol. 297, No. 5586, pp.1551–1555.
- Rolfe, J.M. and Staples, K.J. (1988) *Flight Simulation*, Cambridge University Press, New York, NY, USA.
- Sato, H., Koyama, Y., Kurumatani, K., Shiozawa, Y. and Deguchi, H. (2001) 'U-Mart: a test bed for interdisciplinary research in agent based artificial market', *Evolutionary Controversies in Economics*, pp.179–190.
- Sattenspiel, L. and Herring, D.A. (2003) 'Simulating the effect of quarantine on the spread of the 1918-19 flu in Central Canada', *Bulletin of Mathematical Biology*, Vol. 65, No. 1, pp.1–26.
- Savery, J.R. and Duffy, T.M. (1995) 'Problem based learning: an instructional model and its constructivist framework', *Educational Technology*, Vol. 35, No. 5, pp.31–38.
- Schneeberger, A., Mercer, C.H., Gregson, S.A.J., Ferguson, N.M., Nyamukapa, C.A., Anderson, R.M., Johnson, A.M. and Garnett, G.P. (2004) 'Scale-free networks and sexually transmitted diseases: a description of observed patterns of sexual contacts in Britain and Zimbabwe', *Sexually Transmitted Diseases*, Vol. 31, No. 6, pp.380–387.
- Siallagan, M., Juhana, N., Novani, S. and Putro, U.S. (2006) 'Applications of agent-based modelling and simulation of fuel distribution in indonesia', *World Congress of Social Simulation*, Kyoto, Japan.
- Small, H. (1995) 'Navigating the citation network', *Proc. 58th Annual Meeting of the American Society for Information Science*, Vol. 32, pp.118–126.
- Stroud, P., Valle, S.D., Sydoriak, S., Riese, J. and Mniszewski, S. (2007) 'Spatial dynamics of pandemic influenza in a massive artificial society', *Journal of Artificial Societies and Social Simulation*, Vol. 10, No. 4, pp.1–18, <http://jasss.soc.surrey.ac.uk/10/4/9.html>
- Sumodhee, C., Hsieh, J.L., Sun, C.T., Huang, C.Y. and Chen, Y.M.A. (2005) 'Impact of social behaviors on HIV epidemic: a computer simulation view', *Computational Intelligence for Modelling, Control and Automation, 2005 and International Conference on Intelligent Agents, Web Technologies and Internet Commerce, International Conference on*, Vol. 2, pp.550–556.
- Wenglinsky, H. (1998) *Does it Compute? The Relationship Between Educational Technology and Student Achievement in Mathematics*, Educational Testing Service, Princeton, NJ.