Simulations for epidemiology and public health education

C-Y Huang[1]*, Y-S Tsai[2] and T-H Wen[3]

[1]Chang Gung University, Taoyuan, Taiwan; [2]National Chiao Tung University, Hsinchu, Taiwan; [3]National Taiwan University, Taipei, Taiwan

Recent and potential outbreaks of infectious diseases are triggering interest in predicting epidemic dynamics on a national scale and testing the efficacies of different combinations of public health policies. Network-based simulations are proving their worth as tools for addressing epidemiology and public health issues considered too complex for field investigations and questionnaire analyses. Universities and research centres are therefore using network-based simulations as teaching tools for epidemiology and public health education students, but instructors are discovering that constructing appropriate network models and epidemic simulations are difficult tasks in terms of individual movement and contact patterns. In this paper we will describe (a) a four-category framework (based on demographic and geographic properties) to discuss ways of applying network-based simulation approaches to undergraduate students and novice researchers; (b) our experiences simulating the transmission dynamics of two infectious disease scenarios in Taiwan (HIV and influenza); (c) evaluation results indicating significant improvement in student knowledge of epidemic transmission dynamics and the efficacies of various public health policy suites; and (d) a geospatial modelling approach that integrates a national commuting network as well as multi-scale contact structures.
Journal of Simulation (2010) 4, 68–80. doi:10.1057/jos.2009.13

1. Introduction

Network-based simulation approaches have gained acceptance as trustworthy means of investigating a wide variety of epidemics (Boccara and Cheong, 1993; Axelrod, 1997; Gilbert and Troitzsch, 1999; Alfonseca *et al*, 2000; Barrett *et al*, 2003; Huang *et al*, 2004, 2005b; Schneeberger *et al*, 2004; Ferguson *et al*, 2005; Hsieh *et al*, 2005, 2006; Sumodhee *et al*, 2005; Stroud *et al*, 2007). For the purposes of training students and novice researchers, epidemiology instructors from many disciplines are collaborating with simulation researchers to recreate the transmission dynamics of infectious diseases and to improve our understanding of the efficacies of public health policies (Huang *et al*, 2005a; Hsieh *et al*, 2006). However, computational epidemiology researchers and instructors are still addressing individual problems involving movement and contact patterns among millions of people of different ages and with different professions, educational levels, marital/partner statuses, and levels of epidemiological resistance (Boccara and Cheong, 1993; Barrett *et al*, 2003; Huang *et al*, 2004, 2005b).

In addition, emerging and re-emerging infectious disease outbreaks can develop randomly and unexpectedly depending on the breadth of early stage outbreaks, numbers of randomly imported cases, infected individuals' responses, and contacts with other susceptible individuals (Barrett *et al*, 2003; Huang *et al*, 2004, 2005a, b). Public health policies executed by health authorities also directly and indirectly affect epidemic dynamics and spreading situations (Hsieh *et al*, 2005). Furthermore, improper implementation and the inappropriate timing of public health policy activation occasionally produces such secondary impacts as disease concealment and social discrimination against infected patients and the health care employees who work with them (Huang *et al*, 2004). In spite of these factors, most students and novice researchers in public health and related disciplines still use questionnaires or field investigation techniques when studying epidemic outbreaks—a process that prevents many from gaining a macro view of epidemic dynamics or from assessing the potential efficacies of public health policies for prevention and control.

Network-based simulations are proving successful for solving individual movement and contact problems, exploring epidemic dynamics, and assessing the efficacies of public health policies (Barrett *et al*, 2003; Huang *et al*, 2004, 2005b; Schneeberger *et al*, 2004; Ferguson *et al*, 2005; Hsieh *et al*, 2005, 2006; Sumodhee *et al*, 2005; Stroud *et al*, 2007). In previous studies we have applied our social network simulation experiences to the transmission dynamics of HIV, SARS, and influenza in Taiwan (Huang *et al*, 2004, 2005b; Hsieh *et al*, 2005, 2006; Sumodhee *et al*, 2005). In this paper we will explain our proposal for a four-category

*Correspondence: C-Y Huang, Department of Computer Science and Information Engineering, Research Center for Emerging Viral Infections, Chang Gung University, 259 Wen Hwa 1st Road, Taoyuan 333, Taiwan.
E-mail: gscott@mail.cgu.edu.tw

framework based on demographic and geographic properties, discuss ways of applying network-based simulation approaches to undergraduate students and novice researchers, and describe our experiences simulating the transmission dynamics of two infectious disease scenarios in Taiwan (HIV and influenza). In doing so, we hope to clearly illustrate existing challenges to building network-based epidemic simulations. We hope that our pre-analysis framework and two application examples will assist epidemiology students and novice researchers in their efforts to predict the transmission dynamics of emerging and re-emerging infectious diseases, as well as to improve current public health policies and immunization strategies.

2. Simulations for epidemiology and public health education

Using computer simulations as a pedagogical tool is now common in many scientific technology training and the teaching of science concepts (Liao and Sun, 2001; Colpitts, 2002; Hsieh et al, 2006). Computer simulations are also being used in epidemiology disciplines to support educational and training efforts based on constructivist learning principles. In addition to mitigating learner obsession with the minutiae of complicated procedures described in epidemiological textbooks (Wenglinsky, 1998), simulations 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 epidemic simulations, students can learn by doing, have more and better opportunities for discovering interesting primary and secondary epidemic issues, and gain hands-on experience for dealing with real-world public health issues.

Instructional simulations exemplify problem-based learning. Originally developed for medical education in the early 1970s, problem-based learning is now considered a core teaching model in over 60 medical schools (Savery and Duffy, 1995). The use of simulations for learning and teaching has two characteristics that make it compatible with the theoretical foundations of problem-based learning:

1. *Engagement*. Students often request epidemic simulations to assist with learning and to gain a sense of engagement with real-world epidemiology problems. This allows for the introduction of related concepts to the learning process. There is no 'perfect' simulation, but simulations can still support meaningful learning experiences as long as scenario limitations are taken into account (Aldrich, 2004).

2. *Interaction flexibility*. Epidemic simulations can be used with interaction and feedback methods to illustrate how infectious diseases are spread under different conditions and circumstances (Aldrich, 2004). Epidemiology problems are usually complicated 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 infectious diseases.

Learning through epidemic simulations has at least three potential benefits:

1. *Operational*. Epidemiology problems often require examinations of the influences of various public health policies in specific environments. Using the SARS outbreak of 2003 as an example, epidemiologists may want to measure the potential impacts of public health policies, but it is impossible to do so when running real-world experiments. Epidemiology instructors and students can examine the influences of different public health policies in different regions, and execute 'what-if' experiments to study the emerging behaviours of infections when irrelevant health policies are temporarily removed. In short, simulations can be optimized for learning (Bertsche et al, 1996).

2. *Observational*. Users can take epidemic simulation processes and adjust their scales for observation purposes, slow them down, or speed them up (Sumodhee et al, 2005). Epidemic simulations not only allow novice researchers to practice professional skills without having to invest large amounts of resources, but are also recognized as an efficient approach to reviewing or proving epidemiological concepts. This protects them from having to jump into high-risk situations for learning purposes. In classrooms, post-simulation reports allow teachers to determine which concepts their students have mastered (Levy et al, 1995; Hargrave and Kenton, 2000; Klein et al, 2004).

3. *Construction*. Epidemic simulations can be used to create or explore environments. Using public health policy assessments as an example, learners can practice predicting developments that might result from different combinations of public health policies. In classrooms, epidemiology instructors can exert relatively precise control over knowledge construction and accumulation (Hargrave and Kenton, 2000).

Processes and goals associated with learning via epidemic simulations differ from those associated with traditional classroom and textbook-centred learning. Epidemic simulation scenarios are often open-ended and poorly defined (Hsieh et al, 2005), and problems frequently arise after

simulations are started. We therefore suggest that novices be required to use instruction-based manuals to run epidemic simulations and to create professional quality reports or presentations of their learning results. Teacher preparation time will vary depending on the required epidemiology background, scenario construction requirements, and instruction needed 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.

In light of the amount of required background knowledge (Hargrave and Kenton, 2000), we suggest using pre-instructional time to teach public health policy assessment and epidemic outbreak prediction skills, and post-instructional time to teach skills in epidemic simulation construction and analysis. Both are appropriate for learning-by-doing experiences. We designed a five-step epidemiology teaching process: (a) introducing epidemiology knowledge and background scenarios; (b) preparing a pre-test for guiding students to key properties of an epidemiology issue; (c) creating step-by-step instruction-based epidemic simulations with appropriate sample data, user manuals for operating epidemic simulations, and experiment design examples; (d) unrestrained operating time, which allows students to construct and develop their own experiments; and (e) post-tests or final presentations to evaluate student understanding of the issue.

Since disease scenarios often have no single or absolute approach, it is difficult to evaluate how well novice learners understand the operational aspects of simulations. One potential solution is to design constructive pre-tests and post-tests. Using epidemic simulations associated with public health policies as an example, novice learners may be asked to compare the efficacies of different combinations of public health policies before and after an epidemic simulation is run. 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.

3. Network-based epidemic simulations

Compared to agent-based simulations, network-based simulations rely more on relationships among individuals. Recent mathematical studies and experimental simulations indicate that the topological features of social networks exert considerable influence on the transmission dynamics and critical thresholds of infectious diseases, thereby supporting subtle analyses that agent-based simulations are incapable of performing (Moore and Newman, 2000; Barrett *et al*, 2003; Huang *et al*, 2004, 2005a, b; Sumodhee *et al*, 2005; Kao *et al*, 2006). Accordingly, epidemiologists are focusing on the transmission dynamics of specific network models to investigate in the spread of emerging infectious diseases.

Network-based simulations entail computer entities that imitate contact patterns between individuals and apply state-of-the-art computing technology to study the movement of heterogeneous individuals. Lattice graphs have been applied for purposes of determining distance relationships between individuals. For example, von Neumann and Moore neighbourhood concepts are commonly used in two-dimensional lattice graphs in which one node's four or eight adjacent nodes are defined as neighbours, respectively (Gilbert and Troitzsch, 1999). The use of two-dimensional lattice graphs allows for the easy representation and measurement of geographic spatial and distance concepts. In contrast, random graphs support features associated with casual contacts among mobile individuals in addition to the low degree of separation commonly observed in social networks.

Two major modelling issues associated with network-based epidemic simulations must be considered: choosing an appropriate network model, and integrating knowledge and properties of epidemiology issues into that model. Different epidemiology issues require different network topology structures for building epidemic simulations based on specific contact patterns. For example, both sexual contact and illegal drug use cause HIV infections and diffusion, but their network topology structures are very different. Sexual contact networks should be scale-free to reflect the power-law statistical distribution of sex partners, but bipartite networks are more likely to accurately reflect needle sharing among injecting drug users (IDUs). Several network model types can be adopted, including daily contact networks that make use of individual activity records (Barrett *et al*, 2003; Stroud *et al*, 2007) or bipartite networks composed of individuals and most frequently visited places (eg homes, train stations, school buses, workplaces, restaurants) (Kao *et al*, 2006).

Data granularity and detail also affect network topology structures: if a disease control agency can trace all IDU activities, it is possible to use identified matches to build a relatively precise daily contact network. In contrast, simulations performed by researchers limited to knowing specific locations where IDUs congregate require more assumptions regarding contact patterns. Static demographic data (eg age, gender, profession, educational level, marital status) can be represented as the social network attributes of nodes. In addition to well-constructed and appropriate assumptions, individual movement and contact patterns require support in the form of demographic, geographic, and transportation data—for instance, average daily movement statistics for railway passengers between counties and cities, statistical distributions of family members per household in each county, numbers of employees in workplaces, and numbers of students in classrooms. For novice researchers interested in network-based epidemic simulations, these requirements pose significant challenges to the collection and analysis of demographic, geographic, and related statistical data on individual movement and contact patterns.

In response to a wide variety of geographic and demographic restrictions, we divided all network-based epidemic simulations into four categories. The first category reflects the use of real contact tracing for constructing small-scale individual-to-individual contact networks. Using the 2003 SARS outbreak as an example, health authorities in Taiwan and Singapore attempted to construct contact histories for all infected individuals in order to quarantine anyone who had come into contact with a carrier.

The second category consists of individuals and locations, with individuals passively connected by activity locations. For example, saunas and bars frequented by homosexuals can be viewed as activity locations bridging susceptible individuals with HIV carriers; for illegal drug users, infection locations include syringes and chemicals used for drug dilution. To construct social networks for illegal drug users, epidemiologists must determine how many times a user shares a syringe with other users during 1 week/month, or how many users share the same diluting agent in a single session.

The third category reflects individual neighbourhood concepts using statistical geographic properties. In the absence of real contact data, epidemiologists may need to build a specific and customized social network using well-constructed and appropriate interaction and contact assumptions. In previous studies we proposed a Cellular Automata with Social Mirror Identity Model (CASMIM) consisting of two layers: the upper layer is a simplified multi-agent system for simulating heterogeneous cohorts, and the lower layer contains two-dimensional cellular automata for retaining the geographic mobility of individuals and for representing real-world activity spaces (Figure 1) (Huang *et al*, 2004, 2005b). The social mirror identities that connect the two layers establish CASMIM as a small-world social network and preserve the properties of individuals who interact with their neighbours within two-dimensional geographic spaces. Those properties reflect such activities as long-distance movement and daily visits to fixed locations.

The fourth category frequently requires significant support in the form of demographic or geographic data. For example, Ferguson's Southeast Asian influenza simulation (Ferguson *et al*, 2005) uses statistical data for group density, household size, age distribution, school and workplace size, and individual travel information. The spread of HIV among homosexuals serves as a negative example—that is, movement, location, and means of sexual contact are less obvious, making it more difficult to build a network-based HIV epidemic simulation (Sumodhee *et al*, 2005).

The four properties considered most important for building network-based epidemic simulations are:

1. *Time scale*. In the case of HIV diffusion via heterosexual intercourse, frequency distributions of sexual behaviours over 1 month or 1 year show power-law distribution features (Schneeberger *et al*, 2004), but the same is not true when the time scale is reduced to 1 day or 1 week. It is also important to remember that different diseases have different incubation periods (5 days for SARS *versus* 6 months to 20 years for HIV) and immunization periods.

2. *Geographic scale*. Care must be taken when selecting this scale. Whereas CASMIM can be applied to simulate
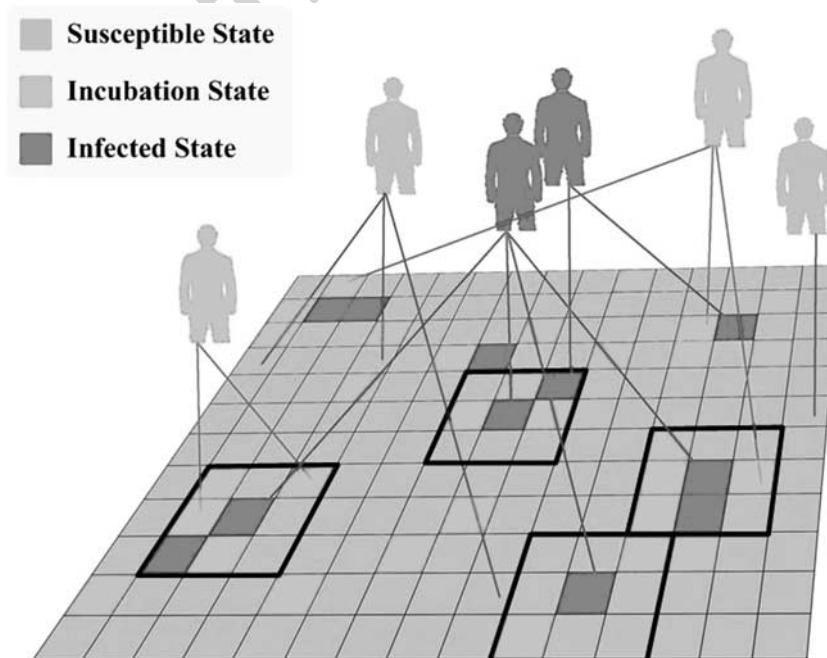

Figure 1   Cellular Automata and Social Mirror Identity Model (CASMIM).

SARS outbreaks in modern cities such as Taipei or Singapore, simulating multi-region epidemic dynamics requires additional demographic data. One possible solution is building a separate CASMIM for each city and measuring transportation flow between paired cities. For example, a new form of influenza tends to be expressed as a large-scale epidemic, therefore models for countries that have multiple regions require the consideration of cross-border transportation networks. Building a social network 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.

3. *Data dependency*. Data granularity determines the best method for building a network model. Using homosexual HIV diffusion as an example, a situation in which data are limited to frequency distributions of sexual contact restricts modellers to using abstract von Neumann and Moore neighbourhood concepts (Huang *et al*, 2005b). However, if movement within a high-risk contact group can be traced, modellers can create simulations capable of predicting further development.

4. *Extendability*. Due to the diversity of data collected for epidemiology issues, simulations of specific infectious diseases often require modifications to existing network models. For example, the homogeneous mixing hypothesis used in random networks assumes that all members of a group are well-mixed (ie equal probabilities exist for contact between any two members), but data on sexual contact or needle sharing do not support this hypothesis. Therefore, extendibility is a major concern when applying an existing network model to new epidemic simulations.

In the next two sections we will share our modelling experiences to construct network-based epidemic simulations. We applied an event-driven programming concept to implement the user and input/output interfaces of epidemic simulations. In addition to providing specific statistical reports and charts presenting experimental results, the two epidemic simulations let learners use browser windows to observe real-time infection situations in an agent society. For a detailed description of our epidemic simulations, please contact the corresponding author.

4. HIV simulation

According to annual statistics presented by Taiwanese health authorities (World Health Organization (WHO), 2003), the number of HIV-1-infected patients increased nationally from 861 in 2003 to 1519 in 2004 to 3386 in 2005 (Table 1). The proportion of IDUs in this population increased from 8.6% in 2003 to 35.8% in 2004 to 71% in 2005 (Sumodhee *et al*, 2005). In light of these sharp increases, government agencies initiated several projects aimed at identifying at-risk

Table 1 Numbers of HIV-1 infections in Taiwan from January 1984 to December 2008

Year	Number of HIV-1 cases per year
1984	9
1985	15
1986	11
1987	12
1988	29
1989	43
1990	36
1996	277
1997	348
1998	401
1999	478
2000	530
2001	654
2002	771
2003	861
2004	1519
2005	3386
2006	2924
2007	1935
2008	1752

populations and controlling the rate of new infections. Due to our success simulating the 2003 SARS outbreak (Huang *et al*, 2004, 2005b), we were asked by the Taiwan Centers for Disease Control (CDC) to collaborate with Professor Yi-Ming A. Chen of the AIDS Prevention and Research Center of National Yang-Ming University to build a network-based epidemic simulation capable of predicting HIV-1 infections among Taiwanese IDUs.

4.1. Data collection

We used data on HIV-positive Taiwanese gathered between 1984 and 2008. In addition, between November 2004 and December 2006 we collected HIV-1-seropositive blood samples and conducted questionnaire interviews with 518 inmates living in four detention centres and two prisons across Taiwan; 3% were teenagers, 73% adults, and 24% adults 60 years of age or older (mean age 32.6 ± 7.7). The large majority (505, or 97.4%) described themselves as IDUs. The women in our sample were three times more likely than their male counterparts to have had sex partners who were also IDUs (65.8% *versus* 19.4%). On average, each IDU shared a drug dilution chemical with two or three other IDUs between two and three times per month; 86.9% stated that they had shared at least a heroin diluent, and 98% said they had shared either diluent or syringes. Results from a logistic regression produced 17.2, 34.0, and 46.7 odds ratios for sharing heroin diluent, sharing syringes, or sharing both diluent and syringes, respectively. In summary, the major causes of HIV-1 infection among the IDUs in our sample were identified as syringe sharing, heroin dilution sharing, low education level, and number of IDUs using the same syringe.
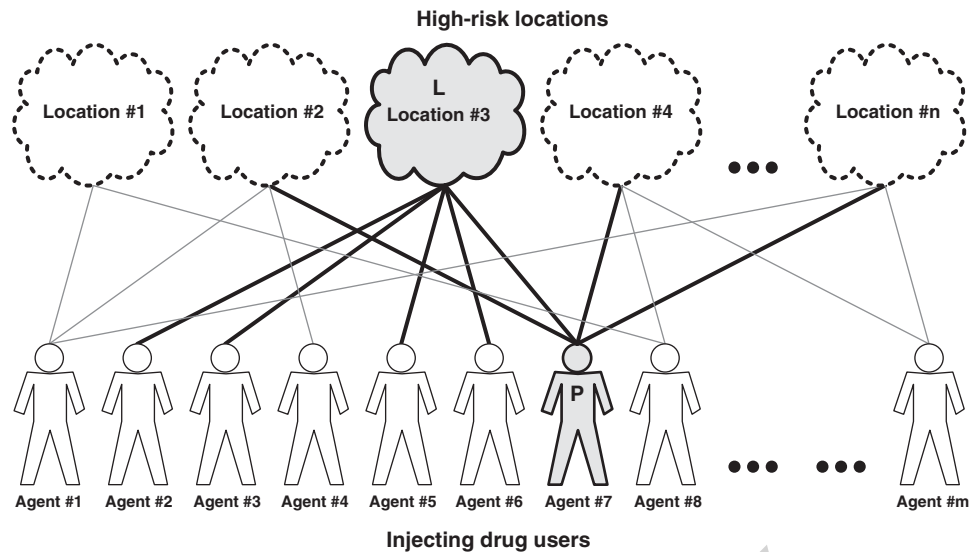
High-risk locations

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

4.2. Simulation model

Since our data were limited to the average number of syringe- and diluent-sharing events per month and average number of persons sharing them per occasion, we treated users as abstract neighbours when developing a bipartite network model based on our social mirror identity concept (Figure 2). The model consisted of two layers connected by social mirror identities: an upper layer representing real-world high-risk locations and a lower layer for simulating IDUs. The individual labeled 'P' in Figure 2 visited more high-risk places than the other IDUs; 'L' marks a frequently visited gathering place.

The time unit used in this simulation was equivalent to 1 month in the real world; results and epidemic curves are reported for each year. We used the average number of syringe and diluent-sharing events per month to represent the number of social mirror identities owned by a lower layer IDU. For the upper layer we used the average number of persons sharing either syringes or a chemical diluent during each occasion to represent the number of IDUs gathered at a high-risk location (Figure 2). Real-world IDU gathering places are located throughout the country and in multiple locations in individual cities, meaning that the spread of HIV among different high-risk locations has no effect on local spreading. In contrast, note that our SARS simulation CASMIM incorporated the effect of local spreading, meaning that SARS could still spread to any other location via transmission between local neighbours in the absence of social mirror identities or shortcuts. Since these high-risk locations are not adjacent, the same CASMIM is inappropriate for simulating HIV transmission among IDUs.

While we were able to obtain data on the statistical distribution of shared syringe and diluent events, we had no

data on the statistical distribution of IDUs visiting each location. According to the most common topological features of social networks, we assumed that this reflects a power-law connectivity distribution. However, we had insufficient empirical data for model validation—especially since the reported number of infected individuals may not have contained concealed cases. This situation shows how difficult it is to build an HIV epidemic simulation without sufficient support in the form of demographic data for building detailed syringe- and diluent-sharing rules for IDUs.

4.3. Discussion

The gray bar in Figure 3 represents the 19 years (1984–2002) of HIV data used for model training and fitting, and the black bar represents 6 years (2003–2008) of data used for model testing and validation.[1] For each simulation we activated a harm reduction policy in December 2005 (simulation time step 264) and increased the policy participation rate from 30% to 80% in June 2006 (time step 270). Prediction results are presented as the curves marked with squares: pre-2006 results were well above actual 2003–2008 HIV epidemic curves (eg 1047 cases predicted for 2003 compared to 861 actual) and post-2007 results were below the actual number (1782 *versus* 1935).

Taiwanese health authorities initiated an HIV harm reduction policy in December 2005. Despite the low policy participation rate (30%), it exerted a strong positive effect in terms of bringing the HIV epidemic under control by reducing the number of new HIV-positive cases from 3386 to 2924 by the end of 2006. This represents the first decrease in the number of new HIV cases in Taiwan since 1986. A stronger harm reduction policy was activated in June 2006,
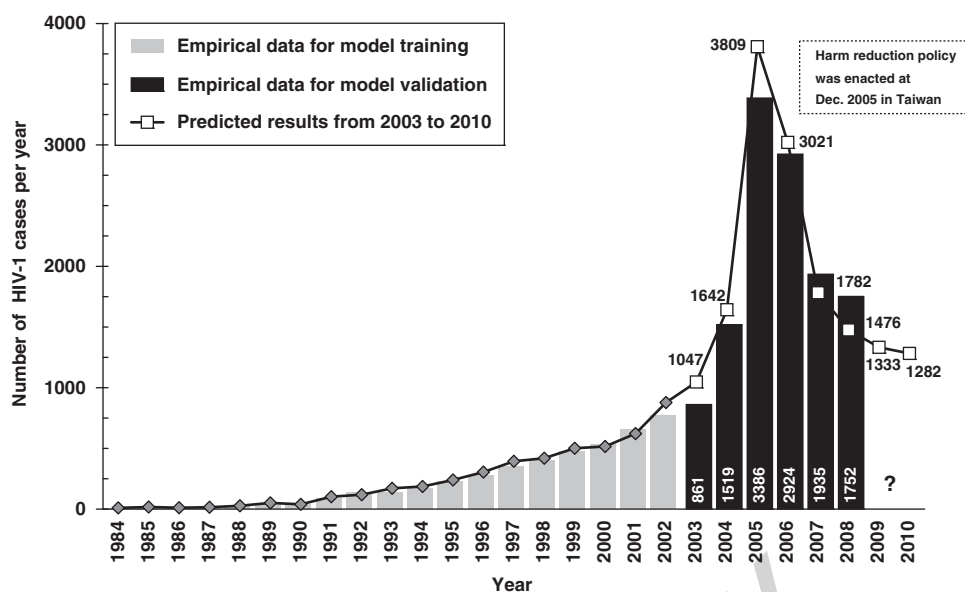
Figure 3 A comparison of actual and predicted[1] HIV epidemic curves from 2003 to 2010 in Taiwan.

resulting in a further decrease in the number of new HIV-positive cases to 1935 by the end of 2007. Our 2002–2010 simulation results also indicate a decreasing trend and suggest that an ongoing harm reduction plan would both lower the number of new HIV-positive individuals and reduce the HIV *reproduction number* from 28 (during 1984–2006) to 1.1 (during 2007–2008).

4.4. Evaluation results

Participants were 14 graduate students recruited from a 2008 infectious disease informatics course taught as part of Chang Gung University's Advanced Biotechnology Education Program. Students worked in pairs to design simulations and to discuss results. For each of two experiments, students were given a pre-test to determine their understanding of (a) HIV transmission dynamics among IDUs, and (b) harm reduction policies activated by Taiwanese health authorities in December 2005. Participants were also given verbal and written information on simulation goals and a post-test to measure the effects of the simulation activity on learning. Each pair was given a user manual for running simulations. After the end of the experiment we evaluated student knowledge on the HIV epidemic, the role of IDUs in HIV transmission, and harm reduction policies. As predicted, results from a paired sample *t*-test indicated statistically significant improvement in the students' overall understanding of HIV epidemic dynamics among IDUs and the efficacies of various harm reduction policies based on different participation rates (Table 2).

5. Influenza simulation

Up to two billion people may be susceptible to the next high pathogenic influenza virus; the predicted mortality rate will approach 65% (WHO, 2008). According to WHO surveillance reports (WHO, 2007), a novel influenza virus is inevitable, yet it is impossible to predict when and in what form the virus will invade individual countries, or how it will specifically threaten the health of individuals. Taiwanese health authorities have already announced three major public health policies: vaccines, antiviral drugs, and rapid containment operations. As part of this programme, starting in 2006 we participated in a 2-year project managed by the Taiwan CDC to work with sociologists to develop a multi-region influenza simulation for the entire country. Lacking epidemiological data for prior outbreaks, we simulated the transmission dynamics of seasonal influenza and assessed the efficacies of prevention and public health policies to determine the optimal application timing of vaccine and antiviral drug responses.

5.1. Data collection

We used transportation data to establish a model of daily inter- and intra-regional contact between individuals—specifically, statistics for the average daily movement of railway passengers between counties and cities (Institute of Transportation, Executive Yuan, Republic of China, 2008). Demographic data from the 2006 Social Indicators Report (Directorate General of Budget, Accounting and Statistics, Executive Yuan, Republic of China, 2006) published by the

---

[1] In this prediction simulation, we averaged 1000 independent experiments to obtain the mean value of error in the epidemic parameter space, then chose parameters and named the optimal parameters with minimal mean error. As shown in the figure, the prediction result corresponds to the best one (ie minimal error) among 30 simulations under the optimal parameters. The parameters did not change with time—that is, predictions are only valid for cases with no additional interventions.

Table 2 Statistical results for (a) HIV and (b) Flu simulations pre-tests and post-tests

Question set		Pre-test score		Post-test score		t-test	p-test
		M	*SD*	*M*	*SD*		
(a) HIV simulation							
Set 1.	Understanding of HIV epidemic concepts and comparisons of actual and predicted HIV epidemic curves from 2003 to 2010 in Taiwan.	6.57	0.73	8.14	0.83	−5.08	*P<0.001*
Set 2.	Understanding of harm reduction policies associated with HIV and assessing efficacies according to different participation rates and activation dates.	6.71	1.03	8.29	0.45	−5.08	*P<0.001*
(b) Flu simulation							
Set 1.	Understand of epidemiology concepts associated with influenza and transmission dynamics of the 1918 influenza pandemic.	5.63	1.15	7.77	1.26	−5.89	*P<0.001*
Set 2.	Assessing and analysing the prevention effects of five public health policies at low and high regional densities and with three policy activation dates (10/1∼10/07, 10/22∼10/28, and 11/19∼11/25).	5.13	0.62	6.68	0.9	−6.86	*P<0.001*
Set 3.	Assessing and analyzing the cost-efficacies of five public health policies at low and high regional densities with three policy activation dates (10/1∼10/07, 10/22∼10/28, and 11/19∼11/25).	5.68	0.47	6.40	0.49	−5.41	*P<0.001*

Taiwanese government were used to assign individuals to various locations. These data include statistical distributions of family members per household in each county, numbers of employees in workplaces, and numbers of students in classrooms. We combined data in an effort to achieve an approximate understanding of the overall distribution of the number of persons in each regularly visited activity location. As shown in Table 3, most activity locations had fewer than 10; exceptions included movie theatres and classrooms, each with 40 or more.
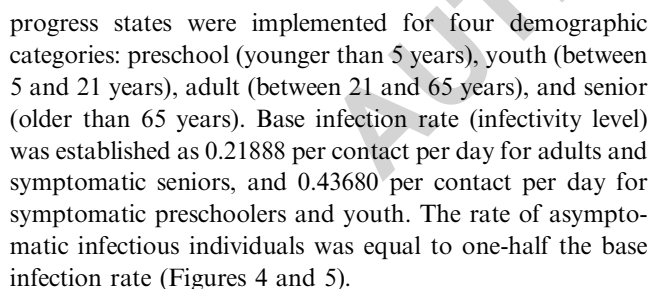
5.2. Simulation model

Based on our SARS modelling and simulation experience, we knew that CASMIM is suitable for simulating the transmission dynamics of contagious diseases in well-mixed but not in poorly mixed modern cities. We therefore assigned a separate CASMIM to each county, with model scale determined by the number of counties included in the overall simulation. Each CASMIM model cell represented one real-world activity location (household, classroom, train station, etc), and the number of individuals in each cell was assigned according to the statistical distribution of numbers of observed persons in regularly visited locations. Railway transportation data were used to represent inter-county movement, with each instance representing a pair of social mirror identities belonging to the same individual but in different counties. Social mirror identities for the majority

Table 3 Statistical of numbers of persons in regularly visited locations such as households, workplaces, and classrooms

Number of persons in regularly visited locations	Percentage of regularly visited locations
1	9
2	28
3	14
4	22
5	13
6	5
7	4
8	1
15	1
25	1
30	1
35	1
40	0.3
45	0.3
50	0.3

of individuals stayed within the same county. Since the incubation period for influenza is only 1–3 days, the simulation time unit in this simulation was equivalent to 1 day in the real world.

The epidemiological progress states for influenza shown in Figure 4 are the same as those described by Longini *et al* (2005) and used in Stroud *et al*'s influenza simulation system (Stroud *et al*, 2007). Separate epidemiological

Figure 4 Epidemiological progress states of epidemic influenza disease manifestations for four age categories with no treatment (Longini *et al*, 2005; Stroud *et al*, 2007).

progress states were implemented for four demographic categories: preschool (younger than 5 years), youth (between 5 and 21 years), adult (between 21 and 65 years), and senior (older than 65 years). Base infection rate (infectivity level) was established as 0.21888 per contact per day for adults and symptomatic seniors, and 0.43680 per contact per day for symptomatic preschoolers and youth. The rate of asymptomatic infectious individuals was equal to one-half the base infection rate (Figures 4 and 5).

A weak point of our multi-region influenza simulation model was its lack of network topology properties for an epidemic simulation model consisting of multiple poorly mixed cities. Unlike a well-mixed modern city, modellers cannot assume that the interaction and contact networks of individuals distributed among multiple counties have small-world properties. Furthermore, there is a lack of epidemiological data for recent novel influenza pandemics to use for empirical validation. Some epidemiologists have constructed simulations of the 1918–1919 influenza pandemic (Stroud *et al*, 2007), but we believe the network topology structure of modern counties is far different from that observed in 1918. Despite these weaknesses, we used our

multi-region influenza model to simulate the transmission dynamics of seasonal influenza and to assess the efficacy of a vaccine policy and related public health policies under different conditions and in different regions.

5.3. Discussion

We used two assessment indicators—prevention effect and cost-efficacy—to replace previously employed epidemic curves to help epidemiologists and public health experts assess the efficacies of public health policies. The first indicator (Equation 1), *prevention effect*, was used to evaluate the effect of specific public health policies; its value was set at > 1 to produce better prevention effects (for values < 1, the policy actually supports the spread of a disease). By comparing different policies at this level, public health experts can identify the best public health policy. The second indicator (Equation 2), *cost-efficacy*, was used to evaluate the prevention effect per unit cost; it was given a positive value to achieve better prevention effects (when the value $= 0$, the policy has no preventive effect). Public health experts can use the same benchmark to make decisions for
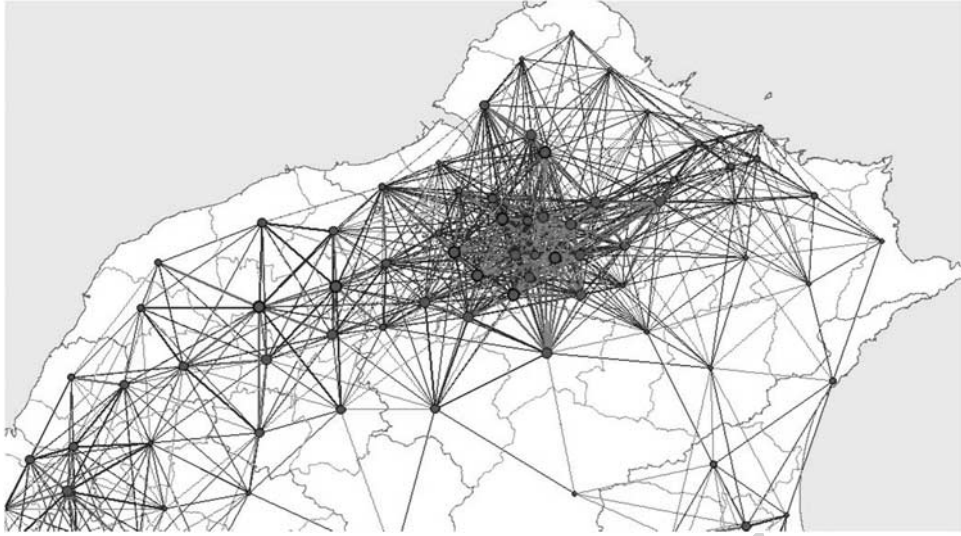
Figure 5 Northern Taiwan commuter network.
Each node represents a city or town and each edge represents a commuter connection. Node size reflects the percentage of persons who work and live in the same city. Edge thickness reflects the number of commuters travelling between cities.

the best timing of public health policies in response to a novel influenza.

$$\text{Prevention\_effect (Policy A)}$$
$$= \frac{\text{Total infected cases without activating any policy}}{\text{Total infected cases with A activated}} \in [0, \infty] \tag{1}$$

$$\text{Cost\_Efficacy (Policy A)}$$
$$= \frac{\begin{array}{c}\text{Total infected cases with activating any policy}-\\ \text{Total infected cases with A activated}\end{array}}{\text{Total consumed resource costs of A}} \tag{2}$$

We used our influenza simulation to compare prevention effects and cost-efficacy for five public health policies, two regional densities (low and high), and three policy application dates (the first at $10/1 \sim 10/07$, the fourth at $10/22 \sim 10/28$, and the eighth at $11/19 \sim 11/25$). The five public health policies and corresponding objects were (a) give the vaccine to randomly chosen individuals, (b) track and inoculate individuals coming into contact with infected individuals, (c) strongly encourage hand washing and mask-wearing by the general public during the influenza season, (d) enforce home quarantines for infected individuals until they recover and for individuals who have come into contact with them for a minimum of 8 days, and (e) give anti-virus medicine in advance to all individuals. As shown in Table 4, the two best public health policies were giving vaccines to randomly chosen individuals and the use of anti-virus medicines; encouraging hand washing and mask-wearing was the third

best. All three policies were more effective when activated as early as possible, with little difference in effect between activating the policies in late October and late November. Our main conclusions derived from the simulation results were (a) hand washing and mask-wearing by the general public during the influenza season is the most cost-effective policy, and (b) using anti-virus medicine in advance is more cost-effective than buying and using a mix of vaccines and anti-virus medicines (Table 4).

5.4. Evaluation results

Participants were 22 graduate students recruited from a spring, 2009 data mining course given by the Department of Computer Science and Information Engineering at Chang Gung University. Most of the participants had no previous knowledge of or experience with influenza epidemiology. Students worked in pairs to construct simulations and to discuss results. Each participant pair was asked to conduct three instructional experiments. For each experiment they were given a pre-test to examine their understanding of transmission dynamics of the 1918 influenza pandemic and five public health policies that were established in response to the pandemic, verbal and written information on simulation goals, and a post-test to determine the effects of the simulation on learning. Each pair was given an instructional manual for running multi-region influenza simulations. As shown in Table 2, results from a paired-sample *t*-test of evaluation scores indicate statistically significant improvement in the students' overall understanding of 1918 influenza transmission dynamics and public health policies.

Table 4 Comparisons of (a) prevention effects and (b) cost-efficacies among five public health policies

Public health policy		Activated at the beginning of October	Activated at the end of October	Activated at the end of November
(a) Prevention effects				
Densely populated region	#1	40.34	10.63	3.00
	#2	2.16	1.98	1.62
	#3	28.68	6.92	2.65
	#4	1.84	1.67	1.45
	#5	41.55	7.38	2.64
Sparsely populated region	#1	14.79	5.60	2.05
	#2	3.85	3.00	2.14
	#3	10.83	5.02	3.26
	#4	1.86	1.80	1.46
	#5	15.11	6.96	3.38
(b) Cost-efficacies				
Densely populated region	#1	0.81	0.75	0.55
	#2	0.45	0.41	0.32
	#3	1.39	1.23	0.90
	#4	0.01	0.01	0.01
	#5	0.81	0.72	0.52
Sparsely populated region	#1	0.28	0.25	0.20
	#2	0.22	0.20	0.16
	#3	0.48	0.42	0.36
	#4	0.01	0.01	0.01
	#5	0.28	0.26	0.21

Note: #1: Inoculate individuals at random; #2: Locate and inoculate those who have come into contact with infected individuals; #3: Encourage hand washing and mask-wearing by the general public during the flu season; #4: Quarantine infected individuals until complete recovery and home quarantine individuals who have come into contact with them for a minimum of 8 days; #5: Give anti-virus medicine in advance for prevention purposes.

6. Multiple-scale epidemiological dynamics for the geospatial diffusion of directly transmitted infectious diseases

Using high quality epidemiologic data to understand the mechanisms underlying the geographic spread of diseases is central to devising spatially targeted prevention and control strategies, a task requiring comparisons of potential outcomes among control measures. Spatial simulation modelling represents a practical approach to helping policy makers decide among various combinations of control strategies in response to real and potential disease epidemics. In this section we will describe a multi-scale simulation process to support the efforts of students, academic researchers, and policy makers to study the spatial dynamics of epidemics and to observe various control measure scenarios.

6.1. Data collection

The diffusion of directly transmitted (droplet and air-borne) diseases results from close human contact and population movement. Accordingly, contact structures and commuting routes between homes and workplaces play important roles in spatial dynamics. To determine the role of commuters in disease transmission, we used data from the 2000 Taiwan

Census Database to create a link-node network structure in order to simulate complex relationships between cities and towns. Figure 5 is an illustration of the commuter network in north Taiwan (the country's economic, cultural, and political centre). It consists of 359 nodes (cities or towns) and more than 120 000 edges (commuter connections between cities). We calculated *contact density* as the number of commuters between cities multiplied by each city's population density.

6.2. Simulation model

To simulate different contact patterns at different levels, we constructed the four-layer multi-scale model shown in Figure 6. To model intra-city contacts and to reflect different contact structures among age groups, we divided the population into children (younger than 15 years), adults (16–64 years), and seniors (older than 65 years). Layer 1 addresses population contact within the same age group, with *Susceptible, Exposed, Infective* and *Removed* (S-E-I-R) statuses used as governing equations to simulate disease transmission. The *Susceptible-to-Exposed* transmission process begins with the initial transmission of a very small number of pathogens. During the early part of the *Exposed* stage, pathogens reproduce but remain below a threshold for active transmission to other *Susceptible* individuals. After
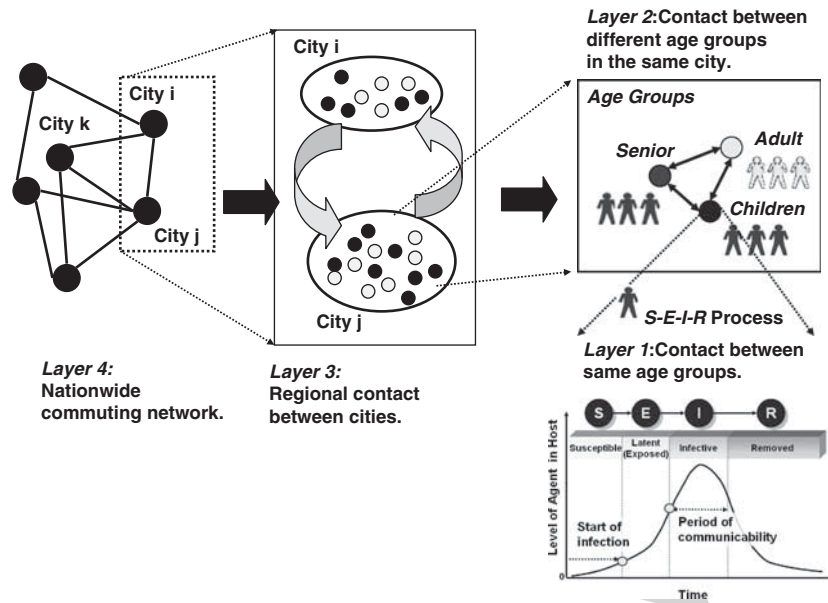
Figure 6   Multi-scale framework for epidemiologic dynamics simulation.

a certain period in which individuals move from *Exposed* to *Infective* status, the number of pathogens becomes sufficiently large for transmission to other *Susceptible* hosts. Eventually the host enters the *Removed* stage in which it achieves a state of immunity or dies.

Layer 2 addresses the difference in contact densities between two age groups. Layer 3 addresses regional commuting between population centres. We assumed that children and seniors are immobile, since the bulk of their activities occur close to their homes. In contrast, adults are more likely to move between fixed geographic locations on a daily basis, increasing their likelihood of carrying pathogens between two or more populations. Layer 4 addresses the national commuting network that can be used to simulate daily population movement throughout Taiwan. The link-node network structure is used to determine disease dynamics among cities.

6.3. Discussion

Developing appropriate prevention and control measures entails making and monitoring the results of multiple rules and decisions made at different points during an epidemic. Conflicts among decision criteria for different strategies are inevitable, and our proposed multi-scale simulation framework can help decision makers test and refine different strategies at different layers. For example, Layer 1 can be used to simulate and evaluate a vaccination policy by changing transmission rates among groups at greater risk of infection (eg children or seniors). This would allow for the testing of social distance measures such as school closures. Layer 2 can be used to evaluate quarantine strategies by changing contact rates among different age groups, Layer 3

can be used to evaluate travel restrictions by changing regional contact rates among cities, and Layer 4 can be used for the same purpose by changing the structure of the commuting network. By analysing multi-scale interactions, decision makers can prepare themselves for making rapid proactive intervention decisions in response to identified outbreak transmission pathways.

Furthermore, our simulation framework can provide additional geospatial insight into epidemiological processes underlying control measures. Spatial orientation and visualization are necessary when monitoring disease progression and generating potential control strategies. We incorporated a geographic information system (GIS) into our multi-scale simulation framework in order to capture spatial variation in disease transmission throughout Taiwan. The GIS supports a visual analysis of the spatial impacts of individual control measures. Combining multi-scale simulations, spatial visualization, and geographic information can clarify spatial and temporal characteristics in support of potential pandemic preparation and control measures.

7. Conclusion

In this paper we proposed a pre-analysis framework for network-based epidemic simulations for purposes of training students and novice researchers, and gave framework reduction and extension examples in terms of collected geographic and demographic data. Epidemiologists can use this information to support such tasks as analysing spreading situations and outbreak patterns, predicting future transmission dynamics, and assessing the efficacies of public health policies for disease prevention and control, vaccine development, and other efforts to fight epidemics. We also described two sample cases to discuss applications of

network-based epidemic simulations. Our experiences in teaching epidemiological modelling and simulations have allowed us to identify three challenges for instructors: choice of a suitable network model, preparation for instruction-based teaching, and evaluating student understanding. Network-based simulations for solving epidemiology issues require more demographic and geographic data support and larger amounts of initial domain knowledge. In other words, most epidemiology issues require collaborations among computer scientists, sociologists, epidemiologists, and policy decision-makers.

References

Aldrich C (2004). *Simulations and the Future of Learning: An Innovative (and Perhaps Revolutionary) Approach to e-Learning*. Pfeiffer: San Francisco, CA.

Alfonseca M, Martinez-Bravo MT and Torrea JL (2000). Mathematical models for the analysis of Hepatitis B and AIDS epidemics. *Simulation* **74**(4): 219–226.

Axelrod R (1997). Advancing the art of simulation in the social sciences. *Complexity* **3**(2): 16–22.

Barrett CL, Eubank SG and Smith JP (2003). If smallpox strikes Portland. *Sci Am* **292**(3): 42–49.

Bertsche D, Crawford C and Macadam SE (1996). Is simulation better than experience. *McKinsey Quart* **1**(1): 50–58.

Bruner JS and Lufburrow RA (1963). The process of education. *Am J Phys* **31**: 468.

Boccara N and Cheong K (1993). Critical-behavior of a probabilistic-automata network SIS model for the spread of an infectious-disease in a group of moving individuals. *J Phys A-Math Gen* **26**: 3707–3717.

Colpitts BG (2002). Teaching transmission lines: A project of measurement and simulation. *IEEE T Educ* **45**(3): 245–252.

Directorate General of Budget, Accounting and Statistics (2006). *Social indicators*, Executive Yuan, Republic of China.

Ferguson NM *et al* (2005). Strategies for containing an emerging influenza pandemic in Southeast Asia. *Nature* **437**(7056): 209–214.

Gilbert GN and Troitzsch KG (1999). *Simulation for the Social Scientist*. Open University Press: Philadelphia, PA.

Hargrave CP and Kenton JM (2000). Preinstructional simulations: Implications for science classroom teaching. *J Comput Math Sci Teach* **19**(1): 47–58.

Hsieh JL, Huang CY, Sun CT and Chen YMA (2005). Using the CAMIM small-world epidemic model to analyze public health policies. In: *Proceedings of Western Simulation Multiconference on Health Sciences Simulation*. New Orleans, Louisiania, USA, pp 63–69. The Society for Modeling and Simulation International: San Diego, California, USA.

Hsieh JL, Sun CT, Kao GYM and Huang CY (2006). Teaching through simulation: Epidemic dynamics and public health policies. *Simulation* **82**(11): 731–759.

Huang CY, Sun CT, Hsieh JL and Lin H (2004). Simulating SARS: Small-world epidemiological modelling and public health

policy assessments. *JASSS* **7**(4), http://jasss.soc.surrey.ac.uk/7/4/2.html.

Huang CY, Sun CT and Lin HC (2005a). Influence of local information on social simulations in small-world network models. *JASSS* **8**(4), http://jasss.soc.surrey.ac.uk/8/4/8.html.

Huang CY, *et al* (2005b). A novel small-world model: Using social mirror identities for epidemic simulations. *Simulation* **81**(10): 671–699.

Institute of Transportation, Executive Yuan, Republic of China (2008). *General profile of respective transportation and communications sectors-RAILWAY*, IOT – Transportation Information-Statistical Trend, http://www.iot.gov.tw/english/ct.asp.

Kao RR, Danon L, Green DM and Kiss IZ (2006). Demographic structure and pathogen dynamics on the network of livestock movements in Great Britain. *P Roy Soc B: Biol Sci* **273**(1597): 1999–2007.

Klein CA, Berlin LS, Kostolansky TJ and Del Palacio JR (2004). *Stock simulation Engine for an Options Trading Game*, Issued on March 23, 2003. United States Patent No. 6709330.

Levy M, Levy H and Solomon S (1995). Microscopic simulation of the stock market: The effect of microscopic diversity. *J Phys I* **5**: 1087–1107.

Liao YH and Sun CT (2001). An educational genetic algorithms learning tool. *IEEE T Educ* **44**(2): 20.

Longini Jr IM *et al* (2005). Containing pandemic influenza at the source. *Science* **309**(5737): 1083–1087.

Moore C and Newman MEJ (2000). Epidemics and percolation in small-world networks. *Phys Rev E* **61**(5): 5678–5682.

Oehme F (2000). Learn by doing: How to include new requirements of research in engineering education. *Eur J Eng Educ* **25**(2): 131–137.

Piaget J (1978). *The Development of Thought: Equilibration of Cognitive Structures*. Viking Press: New York, NY.

Savery JR and Duffy TM 1995. Problem based learning: An instructional model and its constructivist framework. *Educ Technol* **35**(5): 31–38.

Schneeberger A *et al* (2004). Scale-free networks and sexually transmitted diseases: A description of observed patterns of sexual contacts in Britain and Zimbabwe. *Sex Transm Dis* **31**(6): 380–387.

Stroud P *et al* (2007). Spatial dynamics of pandemic influenza in a massive artificial society. *JASSS* **10**(4), http://jasss.soc.surrey.ac.uk/10/4/9.html.

Sumodhee CJ, *et al* (2005). Impact of social behaviors on HIV epidemic: A computer simulation view. In: *Proceedings of International Conference on Computational Intelligence for Modelling, Control and Automation*, Vienna, Austria **2**: 550–556. IEEE Computer Society: Los Alamitos, CA, USA.

Wenglinsky H 1998. *Does it Compute? The Relationship Between Educational Technology and Student Achievement in Mathematics*. Educational Testing Service: Princeton, NJ.

World Health Organization (WHO) (2003). *HIV/AIDS in Asia and the Pacific region*. http://www.wpro.who.int/NR/rdonlyres/11ED3283-9821-43BE-9B73-B3444A3DADE6/0/HIV_AIDS_Asia_Pacific_Region2003.pdf.

WHO (2007). *Ten things you need to know about pandemic influenza*, http://www.who.int/csr/disease/influenza/pandemic10things/en.

WHO (2008). *Confirmed Human Cases of Avian Influenza A (H5N1)*, http://www.who.int/csr/disease/avian_influenza/country/cases_table_2008_09_10/en/index.html.