MODELING INFECTIOUS DISEASES USING GLOBAL STOCHASTIC CELLULAR AUTOMATA

ARMIN R. MIKLER[*], SANGEETA VENKATACHALAM[†] and KAJA ABBAS[‡]

Department of Computer Science and Engineering
University of North Texas, Denton, TX 76203, USA
[*]mikler@cs.unt.edu,
[†]venkatac@cs.unt.edu,
[‡]kaja@cs.unt.edu

Susceptibles-infectives-removals (SIR) and its derivatives are the classic mathematical models for the study of infectious diseases in epidemiology. In order to model and simulate epidemics of an infectious disease, we use cellular automata (CA). The simplifying assumptions of SIR and naive CA limit their applicability to the real world characteristics. A global stochastic cellular automata paradigm (GSCA) is proposed, which incorporates geographic and demographic based interactions. The interaction measure between the cells is a function of population density and Euclidean distance, and has been extended to include geographic, demographic and migratory constraints. The progression of diseases using traditional CA and classic SIR are analyzed, and similar behavior to the SIR model is exhibited by GSCA, using the geographic information systems (GIS) gravity model for interactions. The limitations of the SIR and naive CA models of homogeneous population with uniform mixing are addressed by the GSCA model. The GSCA model is oriented to heterogeneous population, and can incorporate interactions based on geography, demography, environment and migration patterns. The progression of diseases can be modeled at higher levels of fidelity using the GSCA model, and facilitates optimal deployment of public health resources for prevention, control and surveillance of infectious diseases.

Keywords: Global Stochastic Cellular Automata; Infectious Diseases; Computational Epidemiology.

1. Introduction

Globalization and the ever-increasing population diversity accelerates the spread of communicable diseases in the modern society.[1,2] The World Health Organization (WHO)[3] and the Centers for Disease Control and Prevention (CDC)[4] involve in worldwide surveillance of infectious diseases, and prioritize prevention measures at the root cause of epidemics. As the significance of *public health* is being recognized, the role of epidemiologists has become more prominent. *Epidemiology* deals

---

[‡] Corresponding author.

with the study of cause, spread, and control of diseases. The goal for epidemiologists is to implement mechanisms for surveillance, monitoring, prevention and control of diseases. Epidemiological studies may require large data sets of disease outbreaks which are often spatially and temporally distributed. It is in fact ironic that, for epidemiologists to study the dynamics of different diseases, it is imperative for an outbreak to occur. Epidemiologists have been studying and analyzing disease outbreak data by means of statistical tools. In order for the epidemiologists to prepare for a sudden outbreak of an infectious disease or a bio-terror attack, the need for simulation arises. Hence, it is imperative to develop new models that take advantage of today's computational capabilities, and help epidemiologists to analyze and quantify the progression of an epidemic in a given geographic region with specific demographic characteristics. The computational models also enhance the quality of information, accelerate the generation of answers to specific questions and facilitate prediction. To this end, we propose the use of *Global Stochastic Cellular Automata* (GSCA) to simulate outbreaks of infectious diseases,[5] thereby facilitating the optimal allocation of public health resources.

2. Susceptibles-Infectives-Removals Model

Mathematical models of infectious diseases are based on the principles of *susceptibles*, *infectives*, *and removals*, namely the SIR model. *Susceptibles* are those individuals in a population who can be infected by the disease under study. *Infectives* are those individuals who have been infected by the disease and are infectious. *Removals* include all individuals that are incapable of transmitting the infection, and are either recovering, fully recovered, expired from the disease, or immune to the disease. In complex models, the removals who recover may revert to susceptibles. In case of influenza, a recovered individual cannot be infected by the same influenza strain due to acquired immunity during the infection. Nevertheless, he/she may remain susceptible to other influenza strains.

The Kermack-McKendrick Threshold Theorem[6] is the basis for the SIR model. A continuous influx of susceptibles is a requisite for sustained infection in a population. This is the case of endemic diseases, such as tuberculosis, which prevail in a community at all times. The model is based on the presumption of a closed population, assuming that the epidemic spreads rapidly enough that the changes brought in by births, deaths, migration and demographic changes are negligible.[7]

During the start of a disease epidemic, the total population comprises of susceptibles, excluding those that have inherent immunity to the disease. The *index case* is the first infected individual and is the source of the infection. During the infectious period, the infection is passed on to some susceptibles, who interact with the index case close enough to contract the infection. This triggers the cycle of infections spreading through the population. Once the infected individuals become non-infectious, they move over to the removals category. A point of interest is that the total number of susceptibles (S), infectives (I), and removals (R) is a constant

[Eq. (2.2)]. The rising infection on reaching the peak starts to recede due to the decrease in the number of susceptibles, and diminishes eventually.

$$S + I + R = \text{constant} \tag{2.1}$$

$$\frac{dS}{dt} = -\beta SI$$

$$\frac{dI}{dt} = +\beta SI - \gamma I \tag{2.2}$$

$$\frac{dR}{dt} = +\gamma I.$$

The random mixing of susceptibles and infectives[7] is given by the multiplicative product, $S I$. β defines the transmission coefficient[8] based on contact rate between susceptibles (S) and infectives (I), and infectivity of the disease. γ defines the rate of infectives (I) becoming non-infectious. Hence, the average duration of infectivity is given by $1/\gamma$.[7] The set of differential equations used in classic SIR model for a closed population are shown in Eq. (2.2). The transfer rates of individuals from $S \rightarrow I$ and $I \rightarrow R$ are given by dS/dt and dR/dt, respectively. The rate of change of infectives is given by dI/dt.

The SIR/SIRS state diagram (Fig. 1) illustrates the course of a disease in an individual. A susceptible individual may be exposed to a disease pathogen and continue to be in the susceptible state. A susceptible becomes an infective, once the susceptible is able to transmit the pathogen onto others. The recovery state begins once the ability to infect ceases. The individual continues the state of recovery from the disease, or may expire. On full recovery, the individual may acquire full immunity from disease, and hence is no more susceptible to the disease (SIR model). The individual reverts to a susceptible on full recovery when lacking disease immunity (SIRS model).

The SIR model provides a simple framework for understanding the spread of a disease. However, it cannot be used to model a real epidemic for a specific population and region at sufficient fidelity. The SEIR model is an extension of the SIR model, in which the exposed/latent stage of a disease transmission is considered to account for the time period between the onset of the infection in the body and
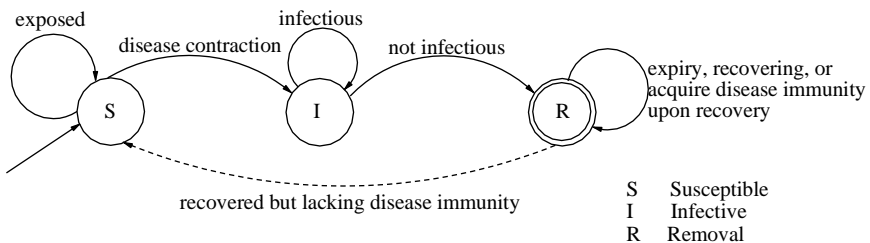

Fig. 1. SIR/SIRS state diagram.

becoming infectious. The SIR and its related models do not take into consideration the geography or the spatial dimensions of a region. In general, interactions among individual is distance-dependent and it is often more likely to interact with individuals at closer proximity. Consequently, the probability of acquiring an infection from an infectious individual is inversely proportional to the interaction proximity. The spread of a disease is dependent on the levels of interaction in the given population of a specific region. The SIR model considers a uniform population with homogeneous mixing and null consideration of specific interaction measures. Also, it is assumed that the epidemic recedes to an end. The model cannot be used effectively for smaller population sizes. The SIR model can be extended to include geography and demographics, but makes it complicated and unwieldy.

3. Cellular Automata

Cellular automata have been used for several decades[9] in the domain of computational models. Nevertheless, in modeling epidemics, this paradigm has rarely been utilized to its full potential.[9−12] *Cellular automata*, as defined by Lyman Hurd, is a discrete dynamic system, where space, time, and the states of the system are distinct.[13] An automaton is best exemplified by representing a point in space as a cell C_i surrounded by other cells, thereby defining the neighborhood H_i of C_i. The cells are most often arranged to constitute a regular spatial lattice (see Fig. 2).

In general, we can define a cellular automaton of any dimension. One-, two-, and three- dimensional automata are often used in science. For a one-dimensional automaton, $H_i = 2$, i.e. cell C_i has a left and a right neighbor (ignoring edge conditions). A two-dimensional automaton is best represented as a regular spatial lattice or grid. Here, cell $C_{i,j}$ is surrounded by cells that form its neighborhood $H_{i,j}$. Traditionally, there are two possible sizes of $C_{i,j}$'s neighborhood in a two-dimensional automaton, namely, $|H_{i,j}| = 4$ in the *von Neumann neighborhood* and $|H_{i,j}| = 8$ in the *Moore neighborhood*[13] (see Fig. 2). Table 1 specifies the neighboring cells for $C_{i,j}$ in both the neighborhoods.

At a particular time t, each cell C of the automaton is said to be in a specific state $s(t)$, which depends on the specific application. $s(t) \in S$ where S is the state space of the cellular automaton. In a simple scenario, cells are assuming binary states
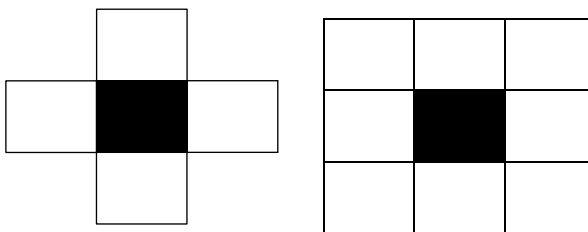

Fig. 2.   von Neumann and Moore neighborhood.

Table 1. Neighborhood specification.

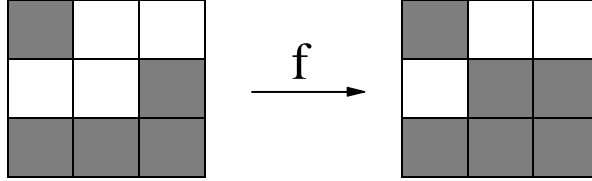| Neighborhood	Neighboring cells for $C_{i,j}$
von Neumann	$C_{i+1,j}$, $C_{i-1,j}$, $C_{i,j+1}$, $C_{i,j-1}$
Moore	$C_{i+1,j}$, $C_{i-1,j}$, $C_{i,j+1}$, $C_{i,j-1}$,
	$C_{i+1,j+1}$, $C_{i-1,j-1}$, $C_{i-1,j+1}$, $C_{i+1,j-1}$

Fig. 3. Cellular automata update from time step $t-1$ to t.

$(0, 1)$. For more complex applications, any size of discrete (and even continuous) state space can be defined. The state of cell $C_{i,j}$ at time t is determined by the state of its neighborhood $H_{i,j}$ at time $t-1$ [see Eq. (3.1)]. The function f can be considered as the rule that dictates how a particular state configuration of $H_{i,j}$ determines the next state of $C_{i,j}$. For a deterministic cellular automaton, the initial states of each cell and the update rule f completely describes the automaton. During a time step t, a new state $s(t)$ is computed for every cell as described above. An initial state configuration will hence evolve, thus representing a dynamic system.

$$s_{i,j}(t) = f(H_{i,j}(t-1)). \tag{3.1}$$

An example of a cellular automata update rule is shown in Fig. 3. Here, the function f is defined by a majority rule. The state of the center cell transitions to a state, which is in majority among the cells in the neighborhood and itself. The update rule determines the deterministic or stochastic behavior of CA. Stochastic behavior is seen by probabilistic update rules in non-deterministic state transitions. For example, in stochastic CA, for every update, a cell can choose probabilistically from a set of update rules, or for a particular update rule, probabilistically choose from a set of states for the stochastic transition.

4. Disease Modeling with Cellular Automata

The traditional cellular automata paradigm forms the basis of our disease model and incorporates the spatial distribution of the population using the Moore neighborhood. The basic unit of cellular automata is a cell. In our model, a cell represents an individual or a sub-population. Each cell can be characterized with state and likelihood risks for exposure and contracting the disease. Unlike the SIR model, every cell comes in contact with the cells in its defined neighborhood. Similar to the SIR model, state S for susceptible is defined as the state in which the cell is
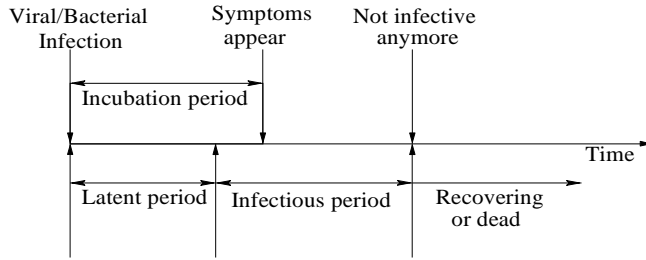
Fig. 4. Infection time-line.

capable of contracting a disease from its neighbors. In the infectious state I, the cell is capable of transmitting the infection to its neighbors. In the recovery state R, the cell is neither capable of passing on the infection nor capable of contracting the infection. On full recovery and acquiry of disease immunity, the cell shall continue in the removal state (R). The time-line for infection is illustrated in Fig. 4.

Infectivity ψ of a disease is defined as the probability of a susceptible cell becoming infectious, when coming in contact with an neighboring infectious cell. Latency λ is defined as the time period between the cell becoming infected and it becoming infectious. Infectious period θ is the time period during which the infected cell is capable of transmitting the disease to neighboring cells. Recovery period ρ is defined as the time period the cell takes to recover, wherein it is neither capable of transmission of the infection nor capable of contracting the infection.

4.1. *Rules for disease spread*

The rules described below determine the state transitions of individual cells in the CA for the SEIR and SEIRS models.

(1) A cell changes its state from *susceptible* to *latent* ($S \rightarrow L$) when it comes in contact with an infected cell in its defined neighborhood. The probability of acquiring the disease from an infected neighbor is a function of infectivity ψ. The cell remains in the latent state for the number of time steps (updates) as defined by the parameter latency λ.

(2) The state of the cell changes from *latent* to *infectious* ($L \rightarrow I$) after being in state L for a given λ. In our model, we assume that every cell exposed to the pathogen will become infectious. In state I, the cells are capable of passing on the infection to neighborhood cells. For example for a disease D, with $\lambda = 2$ units the cell will enter the infectious state I after two time steps of initial exposure.

(3) After the infectious period θ, the cell changes its state from *infectious* to *recovered* or *removed* ($I \rightarrow R$). Once the cell enters the state R, the cell is no more capable of passing on the infection.

(4) From the state R, the cell's state changes back to either susceptible S for the SEIRS model or it remains in state R, for the SEIR signifying complete immunity.

4.2. *Neighborhood saturation*

Figure 5 depicts the cell layers with respect to a central cell in $layer_1$. $Layer_1$ has eight neighboring cells in its outer-line $layer_2$ in a Moore neighborhood model. The outer-line neighborhood of $layer_i$ is $layer_{i+1}$ and the inner-line neighborhood is $layer_{i-1}$. The total neighbors of a layer is defined by a summation of its outer- and inner-line neighborhoods. The ratio of neighboring cells to the cells in the current layer defines the *effective neighbors* per cell of the current layer. L_i is the number of cells in $layer_i$ and is defined in Eq. (4.1). It can be visualized as the area enclosed by layer L_{i-1} subtracted from the area enclosed by layer L_i [see Eq. (4.1)]. The effective outer-line neighbors of $layer_i$ are defined by L_{i+1}/L_i and the inner-line neighbors are L_{i-1}/L_i. Figure 6 illustrates the effective inner- and outer- line neighbors from $layer_1$ up to $layer_{50}$. Even though the effective outer-line neighbors of $layer_1$ is 8, it converges to 1 for higher layers. The effective inner-line neighbors increase from 0 for $layer_1$ to 1 for higher layers.

$$
\begin{aligned}
L_i &= 1 & i &= 1 \\
&= (2*i-1)^2 - (2*i-3)^2 & i &> 1 \\
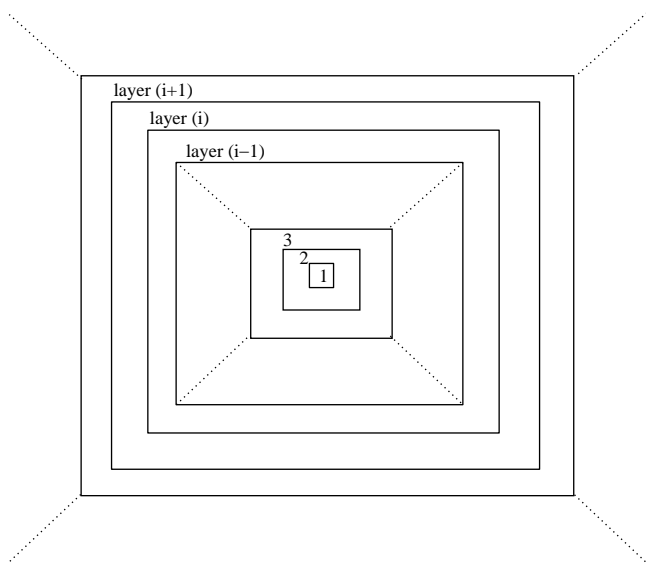L_{i+1}/L_i &\to 1 & i &\to \infty.
\end{aligned}
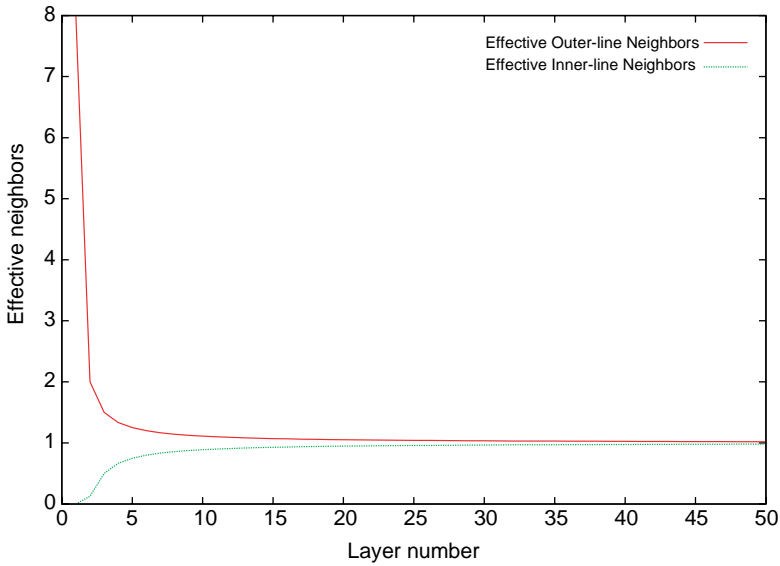\tag{4.1}
$$

Fig. 5. Cell layers.

Fig. 6.   Effective neighborhood.

In the context of epidemiology, we consider a disease progressing at 100% infectivity through neighboring layers. An index case at the central cell in $layer_1$ shall effectively infect eight outer-line neighbors at $layer_2$. However, at higher layers, each cell at $layer_i$ is able to infect effectively only one outer-line cell at $layer_{i+1}$. This resulting *neighborhood saturation* is a primary limitation of naive cellular automata in depicting the spatial progression of a disease.

4.3. *Restrictions of classic cellular automata*

The classic cellular automata methodology suffers from saturation of a limited neighborhood, as described above. A neighborhood of eight cells quickly saturates and thus reduces the number of susceptibles. In such a situation the increase of infectivity parameter plays no role and has the same effect on the spread of the disease. Neighborhood saturation dominates the effects of increasing infectivity and limits the spread of the disease. Further, the need to model a disease where an infective can spread the disease to an extended neighborhood in one time step cannot be modeled. The movement of individuals, migration, or travel is not considered. Some of the models discussed in the literature, deal with movement of individuals from one cell to another in the defined neighborhood. Clearly as discussed above they are deemed to be hampered by early saturation. In order to overcome the limitations posed by naive cellular automata, we introduce the global stochastic model for cellular automata, that shall incorporate the demographics of location and population density.

5. Global Stochastic Cellular Automata

Disease modeling over small regions with local interactions can be implemented using traditional cellular automata. However, its accuracy diminishes for simulating disease spread over large geographic regions because of neighborhood saturation. We propose *global stochastic cellular automata* (GSCA) that includes demographic parameters of a given geographic region.[14] This facilitates understanding of the effects of different demographics, the population density, socio-economics and culture of a region. It can also be used effectively for investigating different vaccination strategies and understanding the effects of travel.

For simulating the spread of diseases in such an environment, contacts need to be established between cells. In this model, every cell may interact with every other cell in the environment. The probability of contact varies based on what is defined to be the interaction coefficient. The interaction coefficient reflects the factors which are important when considering contact between two cells, such as distance, population and other demographics or socio-economic factors. The interaction coefficient in the present model is based on the distance between cells.

The *neighborhood* of cell $C_{i,j}$ in GSCA is defined using a fuzzy set formulation as follows:

$$G_{i,j} := \{(C_{k,l}, \Upsilon_{C_{i,j},C_{k,l}}) \mid \forall C_{k,l} \in C, 0 \leq \Upsilon_{C_{i,j},C_{k,l}} \leq 1\}. \tag{5.1}$$

Here C is the set of all cells in the CA. The above formulation allows for the construction of arbitrary neighborhoods. The membership strength $\Upsilon_{C_{i,j},C_{k,l}}$ represents an interaction coefficient that controls all possible interactions between a cell $C_{i,j}$ and its global neighborhood $G_{i,j}$. Further, it should be noted that $((C_{k,l}, \Upsilon_{C_{i,j},C_{k,l}})$ $\in G_{i,j}) \not\models ((C_{i,j}, \Upsilon_{C_{k,l},C_{i,j}}) \in G_{k,l})$. In what follows, the interaction coefficient $\Upsilon_{C_{i,j},C_{k,l}}$ is a function of *inter-cell distance* and *cell population density* and has been extended to include geographic and demographic constraints.

5.1. *Interaction metrics*

The *interaction coefficient* $\Upsilon_{C_{i,j},C_{k,l}}$ is defined as the strength or likelihood of interaction between two cells, $C_{i,j}$ and $C_{k,l}$. We presently consider the distance between cells as the factor influencing the interaction coefficient. It is calculated as the inverse of the Euclidean distance between the cells [see Eq. (5.2)]. Experiments were conducted on calculating the coefficient based on distance and population as derived from the geographic information systems (GIS) gravity model.[15] Equation (5.3) shows the calculation of interaction coefficient based on distance and population of the two cells, $P_{C_{i,j}}$ and $P_{C_{k,l}}$.

$$\Upsilon_{C_{i,j},C_{k,l}} = \frac{1}{(i-k)^2 + (j-l)^2} \tag{5.2}$$

$$\Upsilon_{C_{i,j},C_{k,l}} = \frac{P_{C_{i,j}} \times P_{C_{k,l}}}{(i-k)^2 + (j-l)^2}. \tag{5.3}$$

The *state of infection* $\delta_{C_{i,j}}$ for a cell $C_{i,j}$ indicates the level of infection present in the cell, and $\delta_{C_{i,j}} \subseteq [0, 1]$. 0 indicates null infection, and 1 indicates full infection. This parameter is used to determine whether the subject or group is capable of transmission of the infection.

The *global interaction coefficient* $\Gamma_{C_{i,j}}$ of cell $C_{i,j}$ is the sum of all the individual n^2_1 interaction coefficients of the cell in a n_*n grid. This coefficient represents the overall interaction of the particular cell. It varies for every cell based on its location. Figure 7 shows the global interaction coefficient based on distance for every cell on a 50_*50 grid. The center cell has the maximal interaction coefficient, since it has a relatively higher number of neighbors at closer proximity. Figure 8 illustrates the
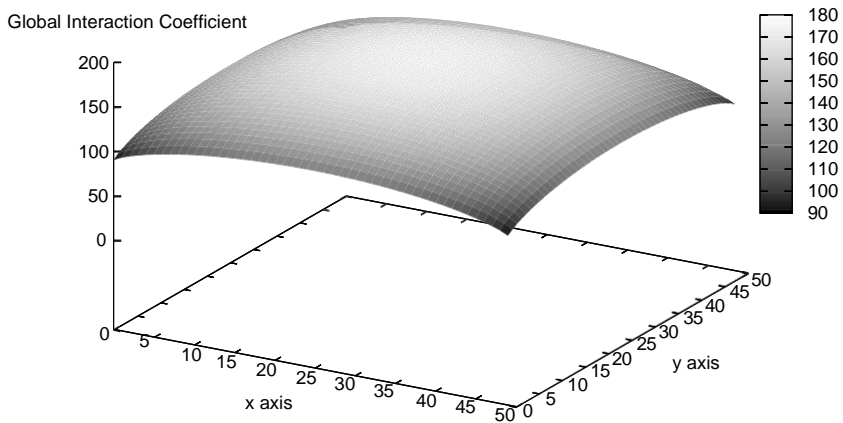

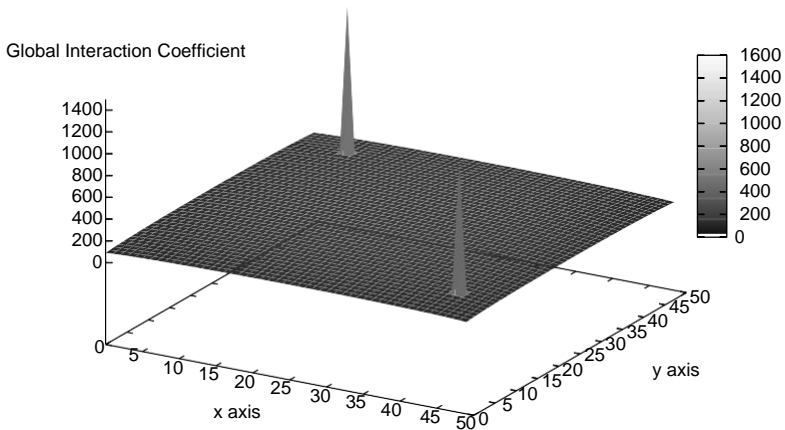Fig. 7. Global interaction coefficient based on distance.

Fig. 8. Global interaction coefficient based on distance and population.

global interaction coefficient based on distance and population for every cell on a 50_*50 grid. Experiments were conducted on two cities with significantly higher population. In the example shown in Fig. 8, population dominates distance. This however may not hold true if the interaction coefficient incorporates measures of population and other demographic values. To normalize $\Upsilon_{C_{i,j},C_{k,l}}$, we calculate the global interaction coefficient Γ [Eq. (5.4)].

$$\Gamma_{C_{i,j}} = \sum_{\forall C_{k,l}=C_{i,j}} \Upsilon_{C_{i,j},C_{k,l}}. \tag{5.4}$$

The *infection factor* $\Phi_{C_{i,j}}$ of cell $C_{i,j}$ with respect to cell $C_{k,l}$ is calculated as a ratio of the interaction coefficient between the two cells to the global interaction coefficient $\Gamma_{C_{i,j}}$. It is also based on the virulence and infectivity of a disease (ψ) and the state of infection (δ) of the infecting agent.

$$\Phi_{C_{i,j}} = \frac{\Upsilon_{C_{i,j},C_{k,l}}}{\Gamma_{C_{i,j}}} \times \delta^{C_{k,l}} \times \psi. \tag{5.5}$$

5.2. *Global stochastic models*

Global stochastic cellular automata (GSCA) models the population with uniform distribution over the grid. Each cell is considered as a sub-population, with certain epidemiological and demographic properties. As derived by the GIS gravity model, the probability of contacts between cells is inversely proportional to the distance between them. This concept is applied in the global model to select contacts for interaction of individual cells.

Although the global model simulates the SIR model, the basic global model considers homogeneous population, and demographics or distances are not included. Figure 9 shows the result of an outbreak simulated in a uniform population with homogeneous mixing, and exhibits similar disease progression behavior of the SIR model. The experiment was conducted on a 50_*50 grid consisting of 2500 cells where every cell constituted one individual. The disease parameters considered were those of influenza. Every individual had an average contact rate of six contacts per day. In a similar experiment, traditional cellular automata restricts the spread of infection due to neighborhood saturation. This is evident in Fig. 10 which compares the infection in traditional CA and the global neighborhood model.

Experiments were also conducted in the global model to investigate the effects of distance based interaction coefficient. Figure 11 depicts the results that illustrate that the rate of disease progression is relatively slower in the global model, when the distance demographic parameter is incorporated.

Using the same metrics of population and grid size, experiments were conducted with the global neighborhood model for three different diseases, namely, common cold, conjunctivitis and influenza, under the assumption of similar virulence/infectivity of disease. The infectious period, latency period and recovery period of
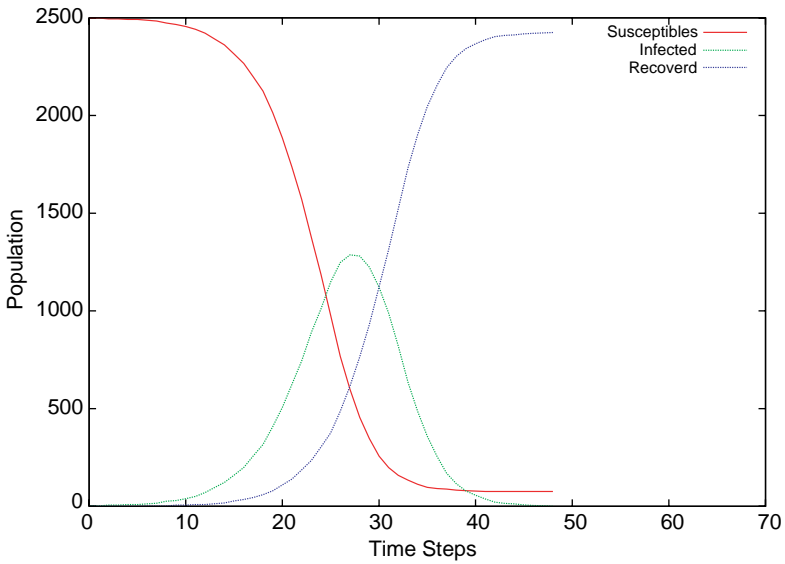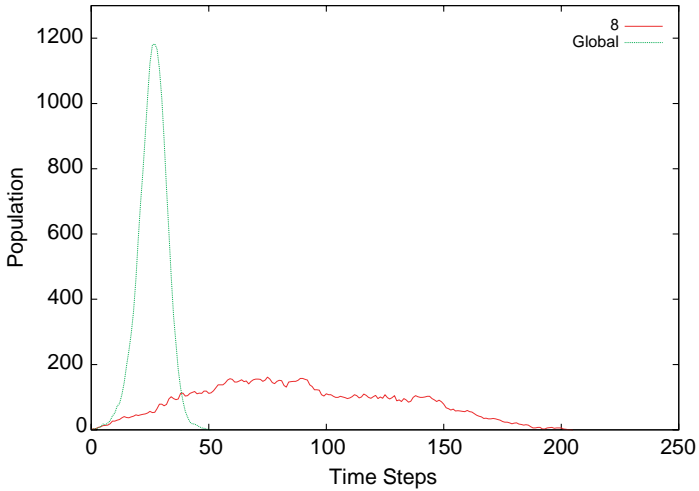
Fig. 9. Global model simulation.

Fig. 10. Comparison of spread of infection in traditional cellular automata with neighborhood saturation and global neighborhood model.

the diseases, shown in Table 2[16,17] were used in the experiments. Due to the relatively smaller incubation period and higher infectious period of conjunctivitis, the rate of spread and the prevalence of conjunctivitis is relatively higher in comparison to common cold and influenza (see Fig. 12).
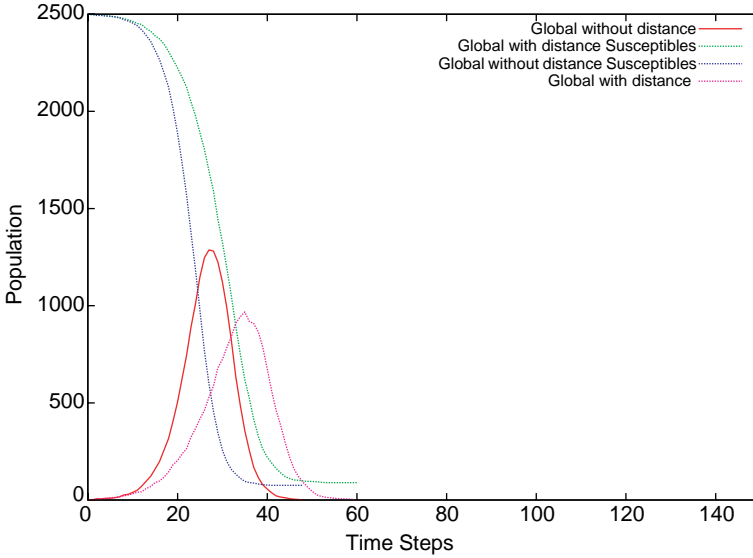
Fig. 11. Disease progression with and without distance demographic parameter.

Table 2. Infection timelines for common cold, conjunctivitis and influenza.

Disease	Incubation period	Latent period	Infectious period
Common cold	3 days	2 days	5 days
Conjunctivitis	3 days	1 day	6 days
Influenza	3 days	3 days	5 days

5.3. *Heterogeneous population models*

The GSCA model is extended to incorporate heterogeneous populations. Rasterized GIS census block data of the area around city of Denton, Texas for the total population of 110,000 is overlaid on a grid of size 50 $*$ 98. Each cell is involved in k contacts, where k is computed based on the cell population and contact rate of individuals per day. Assuming that contacts among individuals are Poisson distributed over time, and individuals make contacts at an average rate of λ, the effective contact rate for a cell is determined by a Poisson random variate. For a cell with population p, $k = p\lambda$. The probability of exposure along with infectivity decides the transmission of infection for a given contact. This leads to heterogeneous interactions in the population, thereby overcoming the presumption of homogeneous mixing in the SIR model. An interaction is a contact between two individuals that may result in successful disease transmission. Figure 13a shows the heterogeneous population distribution of area around Denton city, while Fig. 13b illustrates the disease prevalence of influenza over that region. The total population of the region is 110,000 and the total number of infected people is 48,000.
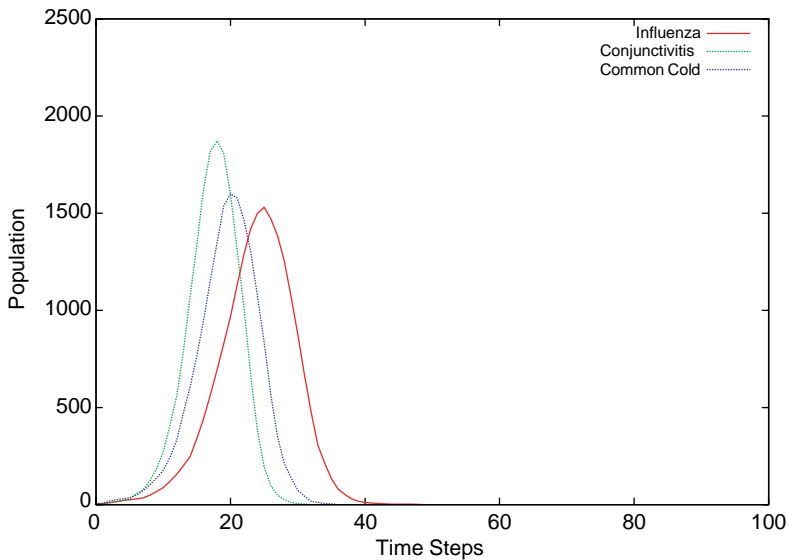
Fig. 12. Comparison of spread of infection for different diseases.

We have implemented a dichotomy of global and local interactions to model distance dependency. For global interactions, contacts are initiated between any two cells in the grid, while for local interactions, the contacts are between neighboring cells. In general, locality can be defined as the set of cells (census blocks) within a specified distance range. The mixing patterns of the population are varied over different proportions of global and local interactions. The prevalence levels of influenza is witnessed to be the same, irrespective of the proportions of local and global mixing. This suggests that influenza prevalence is independent of the spatial domain, and correlates to the results of influenza prevalence in France.[18] The incidence of influenza is further analyzed for varied rates of local and global interactions to generate the corresponding epidemic curves, as shown in Fig. 14. The incidence decreases with higher proportions of local interactions. The results indicate that although influenza prevalence is independent of the spatial domain, the incidence of the epidemic is lowered with higher proportions of local interactions.

The modeling of disease progression through classic SIR and traditional CA are limited by the assumptions of homogeneous population and uniform mixing. These limitations are addressed by the GSCA model, which is oriented towards heterogeneous population. The cell interactions are currently based on population density and Euclidean distance, and can be extended to incorporate geography, demography, environment and migration patterns. The following section summarizes related work in CA epidemiological models as well as the classic SIR models and the newer modes of mathematical reasoning methodologies for epidemiology.
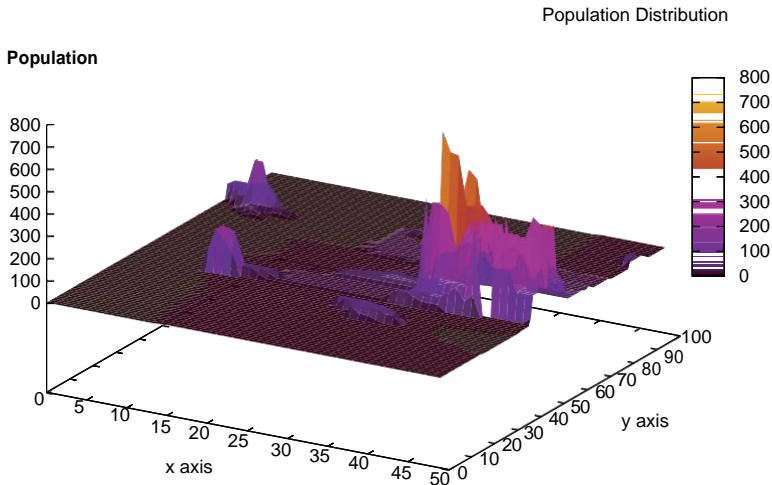
(a) Heterogeneous population distribution.
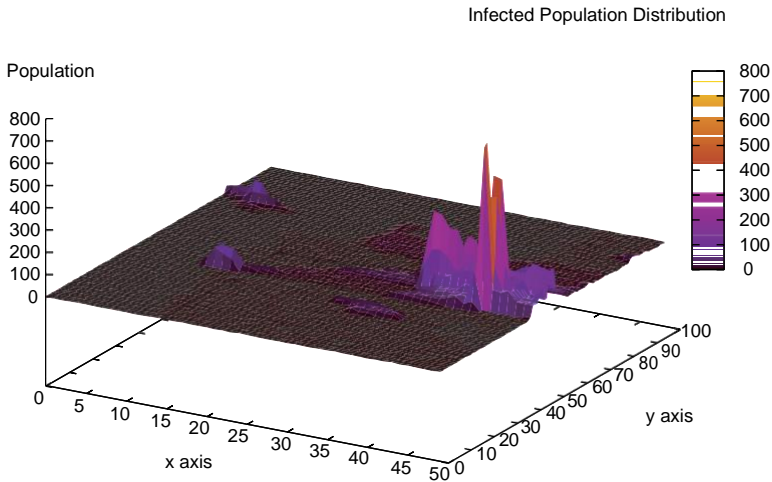
(b) Disease prevalence distribution.

Fig. 13.   Disease prevalence in heterogeneous population.

6. Related Work

Most of the work in modeling infectious disease epidemics is mathematically inspired and based on differential equations and SIR/SEIR model.[7,19] Differential equation SIR modeling rely on the assumption of closed population and neglect the spatial effects.[20,21] They often fail to consider individual contact/interaction process and assume populations are homogeneously mixed and do not include variable
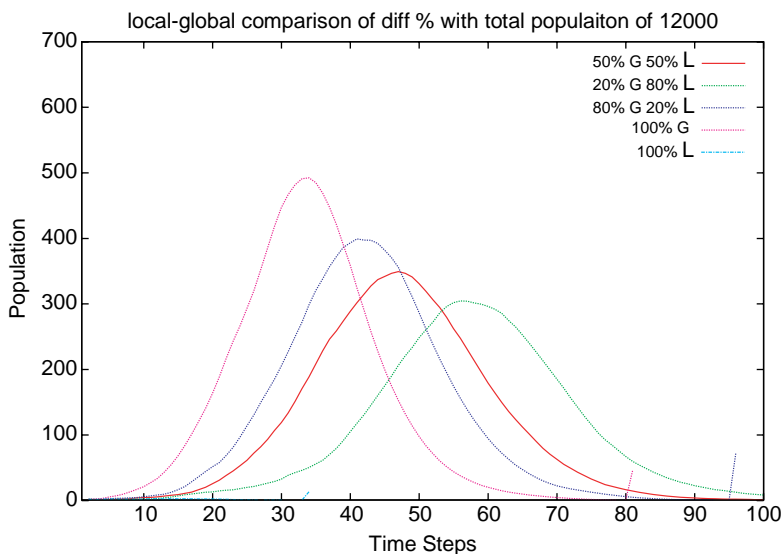
Fig. 14. Epidemic curves for varied rates of global and local interactions in the heterogeneous population of Denton city.

susceptibility. Both partial and ordinary differential equation models are deterministic in nature and neglect stochastic or probabilistic behavior.[12] Nevertheless, these approaches/models have been shown to be effective in regions of small population.[12] Boccara and Cheong[20] study the SIS model for spread of infectious diseases in a population of mobile individuals, thereby introducing non-uniform population density. Ahmed and Elgazzar[22] model variations in population density by allowing cyclic host movement. Ahmed and Agiza[10] introduce incubation and latency time that lends to an accelerating impact on the spread of a disease epidemic. Boccara *et al.*[21] concentrate on SIR epidemic models and take into consideration the fluctuation in the population by births and deaths, exhibiting a cyclic behavior with primary emphasis on moving individuals.

The earliest example of use of cellular automata is Bailey's lattice model[23] for the spread of diseases from micro-level interactions. Schönfisch has analyzed varied cellular automata models to study the dynamics of epidemics.[24] Di Stefano *et al.*[12] have developed a lattice gas cellular automata model to analyze the spread of epidemics of infectious diseases. The model is based on individuals who can change their state independent of others and can move from one cell to other. However, this approach does not consider the critical factor of the infection time-line. Fu has used stochastic cellular automata to model epidemic outbreaks that take into account the heterogeneous spatiality.[25] Situngkir has developed a dynamic model of spatial epidemiology to study avian influenza disease in Indonesia and uses cellular automata for computing analysis.[26] Bonabeau has studied the spatio-temporal characteristics of influenza outbreaks in France. The study infers that the global transportation

systems of the modern world lend to propagation of influenza epidemics dominated by a global mixing process in comparison to local dynamic heterogenities.[18] Duryea has analyzed spatially detailed epidemic models using probabilistic cellular automata for heterogeneous population densities in a region.[27] Benyoussef has used a one-dimensional lattice model and a two-dimensional automata network model to illustrate the spatial spread of rabies among foxes.[28] Fuks describes a SIR epidemic in the spatio-temporal domain via a lattice gas cellular automaton for both human and animal populations. Vaccination strategies are incorporated and dynamics of the disease spread are investigated in relation to the spatial distribution of the vaccinated individuals.[29]

Disease epidemics have been modeled using mean field type (MFT) approximations.[30] Even though the MFT models are similar to the differential equations, they add a probabilistic nature by adding different probabilities for the mixing among individuals. According to Boccara and Cheong,[20] mean field approximations tend to neglect spatial dependencies and correlations and assume that the probability of the state of a cell being susceptible or infective is proportional to the density of the corresponding population. Bayesian analysis of epidemiological data highlights the significance of analyzing demographics to uncover the higher risk spectrums of the population for infectious diseases.[31] A Monte Carlo simulation using a Markov model is implemented to study the infection models that occur naturally, such as influenza, whose viral pathogen spreads through a susceptible community, or induced deliberately, as in the case of bio-terror attacks.[32]

7. Conclusion

Modeling outbreaks of infectious diseases using the traditional cellular automata (CA) model is constrained by neighborhood saturation. The classic susceptibles-infectives-removals (SIR) model is oriented towards a homogeneous population with uniform mixing. The limitations of traditional CA and classic SIR models necessitates the need for new computational models to study the complexity of the spread of diseases in the real world. The global stochastic cellular automata (GSCA) paradigm is used to model outbreaks of infectious diseases. The GSCA model supports modeling and analysis of disease progression in heterogeneous environments, and can incorporate geography, demography, environment, and migration patterns into the interaction measure between cells on a global neighborhood level. The GSCA model includes interactions based on population density and Euclidean distance, and has been implemented to model the progression of three diseases, namely, common cold, conjunctivitis, and influenza. Rasterized GIS population data of Denton city is incorporated to model heterogeneous population through GSCA. The spatial progression of influenza across the heterogeneous population reveals the independence of influenza prevalence for the spatial domain, while influenza incidence decreases with higher rates of local interactions. To facilitate surveillance, monitoring, prevention and control of different diseases, computational models must

be developed. To this end, the GSCA model shall prove to be an valuable asset in the analysis of progression of infectious diseases, thereby leading to optimal utilization of public health resources.

Acknowledgments

References

1. Yaganehdoost A, Graviss E, Ross M, Adams G, Ramaswamy S, Wanger A, *et al.*, Complex transmission dynamics of clonally related virulent mycobacterium tuberculosis associated with barhopping by predominantly human immunodeficiency virus-positive gay men, *J Infect Dis* **180**(4):1245 – 1251, 1999.
2. Youngblut C, *Educational Uses of Virtual Reality Technology*, Institute for Defense Analyses, Alexandria, VA, Tech. Rep., 1998, Technical Report D-2128 IDA Document.
3. World Health Organization (WHO) website. [Online]. Available: http://www.who.org/ (2004).
4. Centers for Disease Control and Prevention (CDC) website. [Online]. Available: http://www.cdc.gov/ (2004).
5. Venkatachalam S, Mikler A, An infectious outbreak simulator based on the cellular automata paradigm, in *Proceedings of the International Conference on Innovative Internet Community Systems*, Guadalajara, Mexico, June 2004.
6. Bailey N, *The Mathematical Theory of Epidemics*, Hafner Publishing Company, NY, USA, 1957.
7. Aron J, *Mathematical Modeling*: *The Dynamics of Infection*, Aspen Publishers, Gaithersburg, MD, 2000, Ch. 6.
8. Allman E, Rhodes J, *Mathematical Models in Biology An Introduction*, Cambridge University Press, 2004.
9. Fu S, Milne G, Epidemic modelling using cellular automata, in *Proceedings of the Australian Conference on Artificial Life*, 2003.
10. Ahmed E, Agiza H, On modeling epidemics, including latency, incubation and variable susceptibility, *Physica A* **253**:347 – 352, 1998.
11. Situngkir H, *Epidemiology through Cellular Automata*, Bandung Fe Institute, Tech. Rep., 2004.
12. Stefano D, Fukś H, Lawniczak A, Object-oriented implementation of CA/LGCA modeling applied to the spread of epidemics, in *Canadian Conference on Electrical and Computer Engineering*, IEEE, Halifax, pp. 26 – 31, 2000.
13. Wolfram S, Statistical mechanics of cellular automata, *Rev Mod Phys* **55**:601 – 644, 1983.
14. Venkatachalam S, Towards computational epidemiology: using stochastic cellular automata in modeling spread of diseases, in *Proceedings of the 4th Annual International Conference on Statistics*, January 2005.
15. Ghosh A, Rushton G (eds.), *Spatial Data Analysis and Location-Allocation Models*, Van Nostrand Reinhold Company, 1987.

16. Benenson A (ed.), *Control of Communicable Diseases Manual*, American Public Health Association, 1995.
17. Timmreck T, *An Introduction to Epidemiology*, Jones and Bartlett, Boston, Ch. 2, pp. 38 – 39, 2002.
18. Bonabeau E, Toubiana L, Flahault A, Evidence for global mixing in real influenza epidemics, *J Phys A Math Gen* **31**:L361 – L365, 1998.
19. Bagni R, Berchi R, Cariello P, A comparison of simulation models applied to epidemics, *J Artif Soc Soc Simul* **5**(3): 2002.
20. Boccara N, Cheong K, Critical behavior of a probabilistic automata network SIS model for the spread of an infectious disease in a population of moving individuals, *J Phys A Math Gen* **26**(5):3707 – 3717, 1993.
21. Boccara N, Cheong K, Oram M, A probabilistic automata network epidemic model with births and deaths exhibiting cyclic behavior, *J Phys A Math Gen* **27**:1585 – 1597, 1994.
22. Ahmed E, Elgazzar A, On some applications of cellular automata, *Physica A* **296**: 529 – 538, 2002.
23. Bailey N, The simulation of stochastic epidemics in two dimensions, in *Proceedings of the 5th Berkeley Symposium on Mathematics and Statistics*, Vol. 4, University of California, Berkeley and Los Angeles, CA, 1967.
24. Schönfisch B, *Zelluäre Automaten und Modelle für Epidemien*, PhD dissertation, University of Tübingen, 1993.
25. Fu S, *Modelling Epidemic Spread through Cellular Automata*, Master's thesis, The University of Western Australia, 2002.
26. Situngkir H, *Epidemiology through Cellular Automata*, Bandung Fe Institute, Tech. Rep., March 2004.
27. Duryea M, Caraco T, Gardner G, Maniatty W, Szymanski B, Population dispersion and equilibrium infection frequency in a spatial epidemic, *Physica D* **132**:511 – 519, 1999.
28. Benyoussef A, Boccara N, Chakib H, Ez-Zahraouy H, Lattice three-species models of the spatial spread of rabies among foxes, *Int J Mod Phys C* **10**:1025 – 1038, 1999.
29. Fuks H, Lawniczak A, Individual-based lattice model for spatial spread of epidemics, *Discr Dyn Nat Soc* **6**:191 – 200, 2001.
30. Kleczkowski A, Grenfell B, Mean-field-type equations for spread of epidemics: the 'small world' model, *Physica A* **274**:355 – 360, 1999.
31. Abbas K, Mikler A, Ramezani A, Menezes S, Computational epidemiology: Bayesian disease surveillance, in *Proceedings of the International Conference on Bioinformatics and Its Applications*, FL, USA, December 2004.
32. O'Leary D, Models of infection: person to person, *Comput Sci Eng* **6**(1):68 – 70, Jan-Feb 2004.