

A Graph Cellular Automata Model to Study the Spreading of an Infectious Disease

Maria Jose Fresnadillo Martínez¹, Enrique García Merino²,
Enrique García Sánchez¹, Jose Elias García Sánchez¹,
Angel Martín del Rey³, and Gerardo Rodríguez Sánchez⁴

¹ Department of Preventive Medicine, Public Health and Clinic Microbiology
Faculty of Medicine, University of Salamanca
C/Alfonso X "El Sabio" sn, 37007-Salamanca, Spain

{jofrema,engarsan,joegas}@usal.es
² I.E.S. Politécnico de Soria, Soria, Spain
engarme@gmail.com

³ Department of Applied Mathematics
E.P.S. de Ávila, Universidad de Salamanca
C/Hornos Caleros 50, 05003-Ávila, Spain
delrey@usal.es

⁴ Department of Applied Mathematics
E.P.S. de Zamora, Universidad de Salamanca
Avda. Cardenal Cisneros 34, 49022-Zamora, Spain
gerardo@usal.es

Abstract. A mathematical model based on cellular automata on graphs to simulate a general epidemic spreading is presented in this paper. Specifically, it is a SIR-type model where the population is divided into susceptible, infected and recovered individuals.

1 Introduction

As is well known infectious diseases are those caused by pathogens (virus, bacteria, epiphytes) or parasites (protozoans, worms) and which can spread in the population. They have been an human enemy from time immemorial. Epidemics and pandemics can place sudden and intense demands on health systems: communicable diseases such as measles, influenza, tuberculosis, etc. are a common fact of modern life. Currently, they are events of concern and interest to many people worldwide: Remember epidemics such as Lyme diseases, toxic-shock syndrome, hepatitis C and E, AIDS, SARS, the Ebola virus, Avian Flu, and more recently the outbreak due to N1H1 virus. They can disrupt economic activity and development. The effects of high disease mortality on mean life span and of disease debilitation and mortality on the economy in afflicted countries are considerable.

As a consequence, the importance of understanding the dynamics and evolution of infectious diseases is steadily increasing in the contemporary world. The study, design and analysis of mathematical models to simulate epidemic spreading has a long history (see [9,19] although the crucial moment in the mathematical epidemiology was reached in 1927 when Kermack and McKendrick (see [13])

introduced its famous model based on a system of ordinary differential equations to study the transmission of the Great Plague occurred in London from 1665 to 1666. It is the first compartmental model; the population is divided into different compartments or classes: susceptibles (individuals which are susceptible to the disease), infected (individuals which have been infected by the disease and are infectious) and recovered (individuals which are removed from infected compartment). Consequently, it is a SIR model where susceptible individuals are infected and the individuals leaving the infective compartment become immune, dead or removed by an isolation policy.

Since then, several mathematical models have been appeared in the literature (see [12,23] and references therein). The majority of mathematical models to simulate epidemic spreading are based on the use of differential equations (see, for example, [6] and references therein). Unfortunately, these models exhibit some important drawbacks since they do not take into account spatial factors such as population density, they neglect the local character of the spreading process, they do not include variable susceptibility of individuals, they cannot comprehensively depict complex contagion patterns (which are mostly caused by the human interaction induced by modern transportation), etc. As a consequence, this can lead to unrealistic results, such as, for example, endemic patterns re-laying on very small densities of individuals, which are called “atto-foxes” or “nano-hawks” (see [16]).

Other mathematical models are based on a particular type of finite state machines called cellular automata. Cellular automata (CA for short) are simple models of computation capable to simulate physical, biological or environmental complex phenomena (see, for example, [22,25]).

CA were introduced by J. von Neumann and S. Ulam in the 50's and their motivation was to obtain a better formal understanding of biological systems that are composed of many identical objects that are relatively simple. The pattern evolution of a cellular automata is the result of the interactions of its objects. Cellular automata have been studied from a dynamical system perspective, from a logic, automata and language theoretic perspective and through ergodic theory.

Roughly speaking, a cellular automaton consists of a discrete spatial lattice of sites called cells, each one endowed at each time t with a state from a finite state set. The state of each cell is updated in discrete time steps according to a local transition function which depends on the states of the cells in some neighborhood around it. As the lattice is finite, some type of boundary conditions must be imposed. As is mentioned above, the usual topologies of CAs are chains and regular lattices; nevertheless, particular properties of two-dimensional lattice space exhibit some drawbacks: connection topology among the cells is restricted to predetermined homogeneous lattice, etc. As a consequence, although that is the standard paradigm for cellular automata, other topologies must be considered when the phenomenon to simulate therefore requires it. In this sense the topologies based on graphs are very important and useful: the nodes of the graph stand for the cells of the CA, and the neighborhood of a particular cell/node is constituted by the nodes adjacent to that one.

The CA-based models for epidemiological spreading eliminate the last mentioned shortcomings exhibited by the models based on ODEs, and are specially suitable for computer simulations. They have been used by several researches as an efficient alternative method to simulate epidemic spreading (see, for example, [1,3,5,7,8,17,18,20,24]), apart from another works appeared in the life sciences and computing literature). Of special interest are the CA-epidemic proposals modeling the motion of individuals (see, for example [2,4,14]). In the majority of these CA models the individuals are assumed to be distributed in the cellular space (defined as an homogeneous lattice) such that each cell stands for an individual of the population.

Here, we introduce a new mathematical model to simulate epidemic spreading. It is a SIR model and is based on cellular automata on graphs. In each cell several individuals are considered instead of only one individual, as is stated in the majority of proposals appeared in the literature. Consequently, each cell stands for a town or a city and its state is obtained from the fraction of the number of individuals which are susceptible, infected, or recovered from the disease.

The model introduced in this work can be considered as the continuation and improvement of the models shown in previous works of the authors. Specifically, in [10,11] a SIS and SIR models based on cellular automata endowed with the traditional topology (Von Neumann and Moore neighborhoods) was presented. Also, an improved SIS model was published in [15] considering topologies based on graphs. There are few works dealing with the use of cellular automata to simulate epidemic spreading considering each cell as an urban centre or a portion of land. Maybe the first paper was due to Sirakoulis, Karafyllidis and Thanailakis (see [21]) and in this work the basis concepts was stated although some drawbacks (related to the motion of individuals) was also presented.

The rest of the paper is organized as follows: In section 2 the basic theory about cellular automata on graphs is stated; The mathematical model to simulate the epidemic spreading is introduced in section 3. An illustrative simulation is presented in section 4, and finally, the conclusions and further work is presented in section 5.

2 Cellular Automata on Graphs

A graph G is a pair (V, E) where $V = \{v_1, v_2, \dots, v_n\}$ is an ordered non-empty finite set of elements called nodes (or vertices), and E is a finite family of pairs of elements of V called edges. Two nodes of the graph, $v_i, v_j \in V$, are said to be adjacent (or neighbors) if there exists an edge in E of the form (v_i, v_j) . We consider undirected graphs, that is, $(v_i, v_j) = (v_j, v_i) \in E$. A graph G is called simple if there is not two edges of G with the same ends and no loops exist, *i.e.* edges whose start and end is located at the same node.

If $V = \{v_1, \dots, v_n\}$, the adjacency matrix of G is the $n \times n$ matrix, $A = (a_{ij})$, where

$$a_{ij} = \begin{cases} 1, & \text{if } (v_i, v_j) \in E \\ 0, & \text{if } (v_i, v_j) \notin E \end{cases}$$

As this work deals with undirected graphs, the adjacency matrix is symmetric.

The neighborhood of a node $v \in V$, N_v , is the set of all nodes of G which are adjacent to v , that is, $N_v = \{u \in V \text{ such that } (v, u) \in E\}$. The degree of a node v , d_v , is the number of its neighbors.

A cellular automaton on an undirected graph $G = (V, E)$ is a 4-tuple $\mathcal{A} = (V, S, N, f)$. The set V defines the cellular space of the CA such that each node stands for a cell the cellular automaton. S is the finite set of states that can be assumed by the nodes at each step of time. The state of the node v at time step t is denoted by $s_v^t \in S$, and it changes accordingly to the local transition function f . N is the neighborhood function which assigns to each node its neighborhood, that is:

$$N: V \rightarrow 2^V$$

$$v_i \mapsto N(v_i) = N_{v_i} = \{v_{i_1}, v_{i_2}, \dots, v_{i_{d_v}}\}$$

Finally, the local transition function f calculates the state of every node at a particular time step $t + 1$ from the states of the its neighbors at the previous time step t , that is:

$$s_v^{t+1} = f(s_{v_{i_1}}^t, s_{v_{i_2}}^t, \dots, s_{v_{i_{d_v}}}^t) \in S,$$

where $N_v = \{v_{i_1}, v_{i_2}, \dots, v_{i_{d_v}}\}$.

3 The SIR Mathematical Model

In the mathematical epidemiological model introduced in this work the population is divided into three classes: those who are susceptible to the disease, those who are infected and those who have recovered and are immune to the disease. Moreover, the population is located at city centres which stand for the nodes of a graph G . If there is some type of transport connection between two of these cities, the associated nodes are connected by an edge. The following assumptions are also made:

1. The population of each node remains constant over time, that is, no births or deaths are taking into account. Moreover, the population distribution is inhomogeneous where P_u is the number of individuals of the node $u \in V$, and $P = \max \{P_u, u \in V\}$.
2. The transmission of the disease is through direct physical contact: touching an infected person, including sexual contact.
3. The population are able to move from its node to another one and return to the origin node at every step of time.

As the model introduced in this work is a SIR model, then the state of the node $u \in V$ at time step t is the triple $s_u^t = (S_u^t, I_u^t, R_u^t) \in Q \times Q \times Q = S$, where $S_u^t \in [0, 1]$ stands for the fraction of susceptible individuals of the node u at time t , $I_u^t \in [0, 1]$ stands for the fraction of infected individuals of the node u at time

t , and $R_u^t \in [0, 1]$ stands for the fraction of recovered individuals of the node u at time step t . Consequently, the transition function of the CA is as follows:

$$s_u^t = f(s_{v_1}^{t-1}, \dots, s_{v_{d_u}}^{t-1}) = (S_u^t, I_u^t, R_u^t) \\ = ((d \circ f_S)(s_{v_1}^{t-1}, \dots, s_{v_{\gamma_u}}^{t-1}), (d \circ f_I)(s_{v_1}^{t-1}, \dots, s_{v_{\gamma_u}}^{t-1}), (d \circ f_R)(s_{v_1}^{t-1}, \dots, s_{v_{\gamma_u}}^{t-1}))$$

The ground where the epidemic is spreading is modeled as a weighted graph where each node stands for a city or a town, and the arc between two nodes represents the connection between the corresponding cities. In this sense, the connection factor between the nodes u and v is the weight associated to the edge $(u, v) \in E$ and it is denoted by w_{uv} . It depends on the transportation capacity of the public and non-public transport; Consequently:

$$w_{uv} = \frac{h_{uv}}{\max\{h_{xy}, \forall x, y \in V\}} \in [0, 1],$$

where h_{uv} is the total amount of population which move from u to v during a time step.

The infected individuals of u at time step t is given by the sum of the following terms:

- The infected individuals at the previous time step which have not been recovered.
- The susceptible individuals which have been infected during the time step. In this case we have to take into account the recovery rate $r \in [0, 1]$. These new sick individuals of u can be infected both by the infected individuals of u or by the infected individuals of the neighbor nodes of u which have moved to u during the time step. In the first case, only the rate of transmission, $p \in [0, 1]$, is involved, whereas in the second case we have to consider the connection factors between the nodes, and the population and movement factor of each node. Moreover we also consider the susceptible individuals of u moved to a neighbor node during the step of time and infected in this neighbor node by its corresponding infected individuals; in this case $\eta_u \in [0, 1]$ yields the portion of moved susceptible individuals from u to its neighbor nodes. Note that $\sum_{v \in V_u} \eta_{uv} = \eta_u$.

As a consequence the mean-field equation for infected individuals is the following:

$$f_I(s_{v_1}^{t-1}, \dots, s_{v_{\gamma_u}}^{t-1}) = (1 - r) I_u^{t-1} + p(1 - \eta_u) S_u^{t-1} I_u^{t-1} \\ + p(1 - \eta_u) S_u^{t-1} \sum_{v \in V_u} \frac{P_v}{P} w_{vu} I_v^{t-1} \\ + p S_u^{t-1} \sum_{v \in V_u} (1 - w_{vu}) \eta_{uv} I_v^{t-1}. \tag{1}$$

On the other hand, the susceptible individuals of each node is given by the difference of the susceptible individuals of the node at the previous time step

and the susceptible individuals which have been infected as is mentioned above. As a consequence, the following equation holds:

$$\begin{aligned}
 f_S \left(s_{v_1}^{t-1}, \dots, s_{v_{\gamma_u}}^{t-1} \right) &= S_u^{t-1} - p(1 - \eta_u) S_u^{t-1} I_u^{t-1} \\
 &\quad - p(1 - \eta_u) S_u^{t-1} \sum_{v \in V_u} \frac{P_v}{P} w_{vu} I_v^{t-1} \\
 &\quad - p S_u^{t-1} \sum_{v \in V_u} (1 - w_{vu}) \eta_{uv} I_v^{t-1}.
 \end{aligned} \tag{2}$$

Finally, the recovered individuals of a node at a particular time step is given by the recovered individuals at the previous time step plus the infected individuals which have been recovered during the time step, that is:

$$f_R \left(s_{v_1}^{t-1}, \dots, s_{v_{\gamma_u}}^{t-1} \right) = R_u^{t-1} + r I_u^{t-1}. \tag{3}$$

Note that, as a simple calculus shows:

$$I_u^t + S_u^t + R_u^t = I_u^{t-1} + S_u^{t-1} + R_u^{t-1} = P_u, \tag{4}$$

and consequently equation (3) can be substitute for the following equation:

$$f_R \left(s_{v_1}^{t-1}, \dots, s_{v_{\gamma_u}}^{t-1} \right) = 1 - I_u^t - S_u^t. \tag{5}$$

Moreover, as

$$\begin{aligned}
 f_S \left(s_{v_1}^{t-1}, \dots, s_{v_{\gamma_u}}^{t-1} \right) &\in [0, 1], \\
 f_I \left(s_{v_1}^{t-1}, \dots, s_{v_{\gamma_u}}^{t-1} \right) &\in [0, 1]
 \end{aligned} \tag{6}$$

and $f_R \left(s_{v_1}^{t-1}, \dots, s_{v_{\gamma_u}}^{t-1} \right) \in [0, 1]$, then a discretization function $d: [0, 1] \rightarrow Q$ must be used in order to get a finite state set. In our case, the discretization function used is the following:

$$\begin{aligned}
 d: [0, 1] &\longrightarrow Q \\
 x &\longmapsto d(x) = \frac{[100 \cdot x]}{100}
 \end{aligned} \tag{7}$$

where $[m]$ stands for the nearest integer to m . As a consequence,

$$Q = \{0, 0.01, 0.02, \dots, 0.99, 1\}. \tag{8}$$

4 An Illustrative Simulation

In this simulation we will suppose that the epidemic is spreading over n cities forming a complete graph K_n . A complete graph is a graph in which each pair

of graph nodes is connected by an edge (that is, each city is connected with each others). The complete graph with n nodes is denoted by K_n and has $\frac{n(n-1)}{2}$ edges. The adjacency matrix A of the complete graph K_n takes the particularly simple form of all 1s with 0s on the diagonal.

For the sake of simplicity this example deals with the complete graph K_6 , that is, only $n = 6$ cities are involved in the spreading of the epidemic: u_1, \dots, u_6 . In Figure 1 the graph topology of this example is shown.

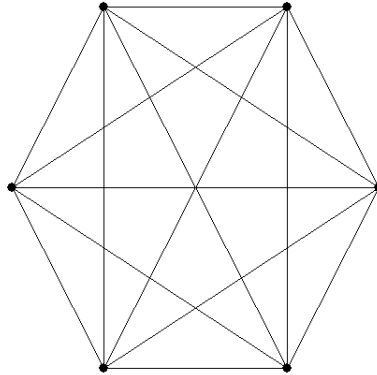


Fig. 1. Complete graph K_6

Moreover, the parameters used in this example are merely illustrative and they do not correspond to a particular infectious disease. We will consider the following initial configuration:

$$S_{u_1}^0 = 0.8, I_{u_1}^0 = 0.2, R_{u_1}^0 = 0,$$

$$S_{u_i}^0 = 1, I_{u_i}^0 = R_{u_i}^0 = 0, \quad 2 \leq i \leq 6.$$

That is, there is only one node at time $t = 0$ with infected population. Moreover, the parameters used are:

$$p = 0.5, r = 0.6,$$

$$\eta_{u_i} = 0.25, \quad 1 \leq i \leq 6 \text{ with } \eta_{u_i v} = 0.05 \quad \forall v \in N_{u_i}.$$

Note that it is assume that $\eta_{u_i v} = \eta_{u_i} / d_{u_i}$ for each i . Moreover, let us suppose that the population of each node is the same: $P_{u_i} = 100$ with $1 \leq i \leq 6$, and also the transport capacity between two nodes is the same: $w_{u_i u_j} = 1$ for $1 \leq i, j \leq 6$. Note that this example deals with an homogeneous-symmetric case.

In Figure 2, the evolution of the number of susceptible, infected and recovered individuals is shown.

In Table 1, the necessary conditions for epidemic spreading from a single node u to a neighbor node v are shown. In the first, second and third column some different and arbitrary values of the parameters p , η_u and w_u are taken. In the

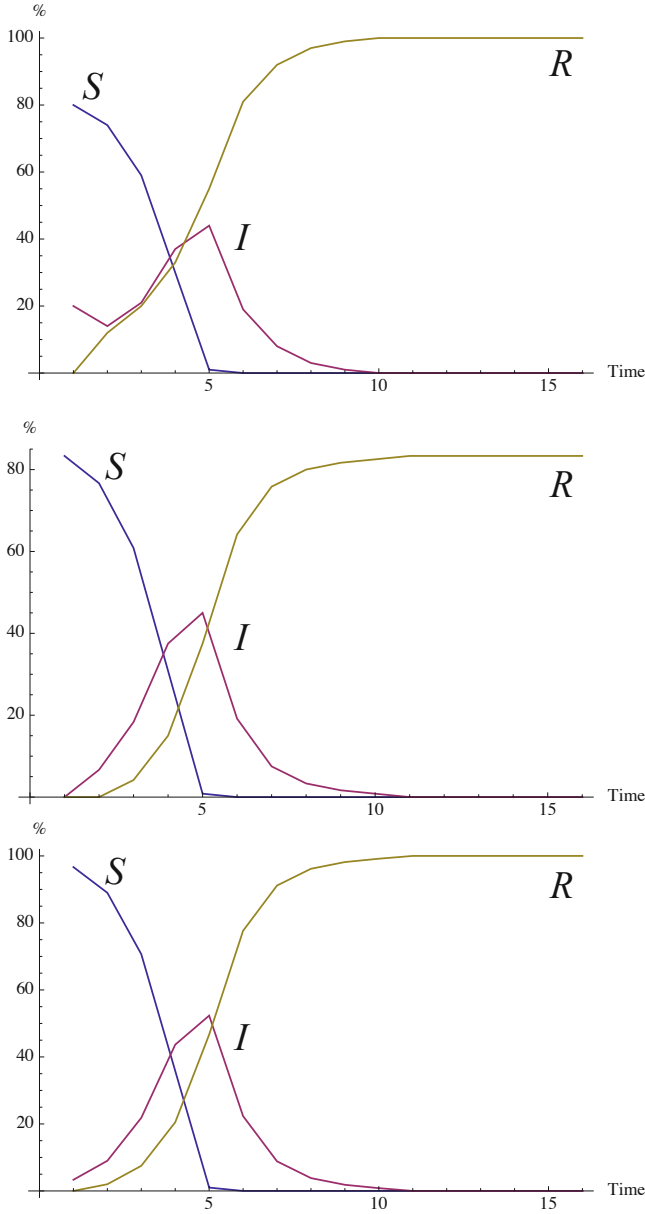


Fig. 2. Top: Evolution of the proportion of the susceptible, infected and recovered population in the node u_1 . Middle: Evolution of the susceptible, infected and recovered population in the nodes u_2, \dots, u_6 . Bottom: Evolution of total number of susceptible, infected and recovered individuals.

fourth column, the minimum state of the node u to produce epidemic spreading is shown, and finally in the fifth column the state of the neighbor node (when the spreading occurs) at the following state of time is given.

Table 1. Necessary conditions for epidemic spreading in the case of K_6

p	η_u	w_u	I_u^0	I_v^1
0.25	0.25	0.5	0.84	0.08
0.25	0.25	1	0.53	0.1
0.25	0.5	1	0.8	0.1
0.5	0.25	0.5	0.47	0.09
0.5	0.25	1	0.27	1
0.5	0.5	0.5	0.67	0.09
0.5	0.5	1	0.4	0.1
0.5	0.75	1	0.8	0.1
0.75	0.25	0.5	0.33	1
0.75	0.25	1	0.18	0.1
0.75	0.5	0.5	0.47	1
0.75	0.5	1	0.27	0.1
0.75	0.75	0.5	0.84	0.09
0.75	0.75	1	0.53	0.1
1	0.25	0.5	0.25	0.1
1	0.25	1	0.13	0.1
1	0.5	0.5	0.36	0.1
1	0.5	1	0.2	0.1
1	0.75	0.5	0.67	0.1
1	0.75	1	0.4	0.1

5 Conclusions

In this work a new SIR-epidemiological model based on cellular automata on graphs has been proposed. The main characteristics of this model are the following:

- Each node of the graph stand for a group of individuals placed on a city or town.
- These individuals are classified into three compartments: susceptible, infected and recovered. As a consequence, the state of the each node at a particular time step is the 3-tupla formed by the portion of susceptible, infected and recovered individuals at this time.
- It is suppose that the transmission of the disease is through direct physical contact between an infected and a susceptible individual.
- The population is able to move from a node to another one.
- The local transition function of the cellular automata is non-linear and it involves the following parameters: the recovery rate, the rate of transmission, the movement factor for susceptible individuals, the connection factor between the nodes and the population of each node.

The laboratory simulations obtained seem to be in agreement with the expected behavior of a real epidemic.

Future work will aim to extend the paradigm presented in this work to other compartmental models as SIRS, SEIR, etc. Moreover, the study of the introduction in the model of new parameters and vaccination effect must be taken into account.

Acknowledgements. This work has been supported by Fundación “Memoria D. Samuel Solórzano Barruso” (University of Salamanca, Spain) under grant YA7Y / 463AC06.

References

1. Ahmed, E., Agiza, H.N.: On modelling epidemics including latency, incubation and variable susceptibility. *Physica A* 253, 247–352 (1998)
2. Ahmed, E., Elgazzar, A.S.: On some applications of cellular automata. *Physica A* 296, 529–538 (2001)
3. Beauchemin, C., Samuel, J., Tuszynski, J.: A simple cellular automaton model for influenza A viral infections. *J. Theor. Biol.* 232, 223–234 (2005)
4. 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, 3707–3717 (1993)
5. Boccara, N., Cheong, K., Oram, M.: A probabilistic automata network epidemic model with birds and deaths exhibiting cyclic behaviour. *J. Phys A-Math. Gen.* 27, 1585–1597 (1994)
6. Diekmann, O., Heesterbeek, J.O.P.: *Mathematical Epidemiology of Infectious Diseases: Model Building, Analysis and Interpretation.* Wiley (2000)
7. Fuentes, M.A., Kuperman, M.N.: Cellular automata and epidemiological models with spatial dependence. *Physica A* 267, 471–486 (1999)
8. Fuks, H., Lawniczak, A.T.: Individual-based lattice model for spatial spread of epidemics. *Discrete Dyn. Nat. Soc.* 6, 191–200 (2001)
9. Hamer, W.H.: Epidemic disease in England. *Lancet* 1, 733–739 (1906)
10. Martín del Rey, A., Hoya White, S., Rodríguez Sánchez, G.: A Model Based on Cellular Automata to Simulate Epidemic Diseases. In: El Yacoubi, S., Chopard, B., Bandini, S. (eds.) *ACRI 2006. LNCS*, vol. 4173, pp. 304–310. Springer, Heidelberg (2006)
11. Hoya White, S., Martín del Rey, A., Rodríguez Sánchez, G.: Modeling epidemics using cellular automata. *Appl. Math. Comput.* 186, 193–202 (2007)
12. Keeling, M.J., Rohani, P.: *Modeling Infectious Diseases in Humans and Animals.* Princeton University Press (2008)
13. Kermack, W.O., McKendrick, A.G.: Contributions to the mathematical theory of epidemics, part I. *Proc. Roy. Soc. Edin. A* 115, 700–721 (1927)
14. Mansilla, R., Gutierrez, J.L.: Deterministic site exchange cellular automata model for the spread of diseases in human settlements. *Complex Systems* 13, 2 (2001)
15. Fresnadillo, M.J., García, E., García, J.E., Martín, Á., Rodríguez, G.: A SIS Epidemiological Model Based on Cellular Automata on Graphs. In: Omatu, S., Rocha, M.P., Bravo, J., Fernández, F., Corchado, E., Bustillo, A., Corchado, J.M. (eds.) *IWANN 2009, Part I. LNCS*, vol. 5518, pp. 1055–1062. Springer, Heidelberg (2009)

16. Molisson, D.: The dependence of epidemic and population velocities on basic parameters. *Math. Biosci.* 107, 255–287 (1991)
17. Quan-Xing, L., Zhen, J.: Cellular automata modeling of SEIRS. *Chinese Phys.* 14, 1370–1377 (2005)
18. Ramani, A., Carstea, A.S., Willox, R., Grammaticos, G.: Oscillating epidemics: a discrete-time model. *Physica A* 333, 278–292 (2004)
19. Ross, R.: *The prevention of malaria*, 2nd edn. Murray (1911)
20. Rousseau, G., Giorgini, R., Livi, H., Chaté, H.: Dynamical phases in a cellular automaton model for epidemic propagation. *Physica D* 103, 554–563 (1997)
21. Sirakoulis, G.C., Karafyllidis, I., Thanailakis, A.: A cellular automaton model for the effects of population movement and vaccination on epidemic propagation. *Ecol. Model.* 133, 209–223 (2000)
22. Toffoli, T., Margolus, N.: *Cellular Automata Machines: A New Environment for Modeling*. The MIT Press (1987)
23. Vynnycky, E., White, R.G.: *An Introduction to Infectious Disease Modelling*. Oxford University Press (2010)
24. Willox, R., Grammaticos, B., Carstea, A.S., Ramani, A.: Epidemic dynamics: discrete-time and cellular automaton models. *Physica A* 328, 13–22 (2003)
25. Wolfram, S.: *A New Kind of Science*. Wolfram Media Inc. (2002)