INTELLIGENT MODEL FOR VIRUS SPREADING

Honoriu Vălean[1], Adina Pop[2], Camelia Avram[3]

[1, 2, 3] Technical University of Cluj – Napoca, Automation Department, Ghe. Baritiu st, nr. 26, Cluj – Napoca, Romania, {honoriu.valean,adina.pop,camelia.avram}@aut.utcluj.ro

Abstract: Epidemic propagation models have been applied on modeling the propagation of viruses. Some mathematical models and computer simulations deal with the spatial distribution of susceptible along a line, across a lattice or over a network to overcome the inaccuracies due to the assumption of random mixing of the population. The viral propagation is determined by intrinsic characteristics of the network. This paper presents a model of the co-evolution of transmissible disease and a population of non-randomly mixed susceptible agents. The simulation applies a modified mathematical SIR epidemiological model of disease transmission. In this paper the authors propose an intelligent model for avian influeza spreading.

Keywords: intelligent models, virus spreading, simulation, artificial intelligence, monitoring

1. INTRODUCTION

Since the 1920's, stochastic models of epidemics have been used for viruses spreading modeling and simulation. Epidemic propagation models (Bailey, 1975) have been applied on modeling the propagation of viruses (Kephart and White, 1993a). Simulation models have been used to discuss the influence of the network topology (Kephart, 1994)(Wang, 2000)(Pastor-Satorras, 2001a).
Virtual experiments are conducted by varying the type of topology, the number of nodes, density and isolation. Experiment results show that random graph topology generated by the same density and isolation as real world data set could be used on modeling the viruses' propagation. Kephart and White (1991, 1993b) are among the first who propose epidemiology-based analytic models. Their studies, however, are based on topologies that do not represent modern networks. Staniford et al. (2002) reported a study of the Code Red worm propagation, but did not attempt to create an analytic model. The more recent studies by Pastor-Satorras et al. (2001a, 2001b, 2002a, 2002b, 2002c) and Barabási et al. (1999, 2002) focused on epidemic models for power-law networks. The viral propagation is largely determined by intrinsic characteristics of the network.
This work aims to develop a general analytic model of avian influenza virus propagation and based on this model a system that will help to prevent the spreading in the populated area and to warn the population. This paper is proposing the use of existing models to determine the avian influenza spreading for Romania. A similar system exists in SUA, HPAI (Highly Pathogenic Avian Influenza), which with the program for pandemic response are part of the National Strategy for pandemic influential.

2. THE SIMULATION MODEL

The standard model is based on a population of individuals who are either susceptible to a specific disease (*susceptible* denoted **S**) or infected with the disease and capable of transmitting it to others (*infected* denoted **I**) (*** and Tassier, 2005). An SIR model is an epidemiological model that computes the theoretical number of people infected with a contagious illness in a closed population over time. The name of this class of models derives

from the fact that they involve coupled equations relating the number of susceptible people S(t), number of people infected I(t), and number of people who have recovered R(t). One of the simplest SIR models is the Kermack-McKendrick model. Population members who overcome a disease may become immune to further infection or may become susceptible once again depending on the particular disease. Population members who are immune to a disease or remain infected but through isolation cannot transmit it, are considered *removed* (denoted **R**). The model described above is known as an *SIR* model. It may be modified slightly to provide fresh susceptible through birth or immigration. Some pertinent parameters of epidemic models are as follows. The period of time during which a disease exists entirely within an organism is known as the disease's *latent period*. The organism is not infected during this period. An *incubation period* often follows latency. During incubation the organism may not show outward sign of infection but is nevertheless infected. Usually once the incubation period is over, the victim of the disease is clearly marked by symptoms and can therefore be avoided by susceptible. Probabilistic epidemiological models that operate in discrete time steps are particularly suited for a software implementation. At any time step, the probability of a new case of the disease appearing is proportional to the number of susceptible multiplied by the number of infected. This basic model assumes random mixing of individuals in the population and does not allow for the complex interactions between physically separated sub-populations, nor for variable incubation or latent periods of a disease. The problems inherent in models that make simplifying assumptions concerning the nature of spatial distributions are discussed in Durrett et al (1994). Various extensions to the SIR model to allow for these phenomena have been added over the last fifty years. Some mathematical models and computer simulations deal with the spatial distribution of susceptible along a line, across a lattice or over a network to overcome the inaccuracies due to the assumption of random mixing of the population. Cellular-automata and other discrete versions of the SIR method have been utilized also, Willox et al (2003) and Martins et al (2003). Some of these models have also incorporated disease *carriers*, and non-homogeneous populations.

The model proposed in the current paper is a modified SIR model for epidemics and allows all of these phenomena to emerge from the simulation without hard-coding their behavior.

The virus propagation mechanisms are categorized as one-to-one, one-to-many, many-to-one, and many-to-many. The category refers to the number of infection source and the number of infection targets at once. A new measure, isolation, is needed to describe the virus propagation topology since the isolation nodes are critical to in the propagation process. The change of states is determined by the propagation of viruses through the network and the propagation of warning messages through the local community. We assume that the population will receive an automatic warning message if their local authorities receive a warning about the virus propagation.

2.1 Model for virus spreading into a populated area

S_t - number of persons that can be infected
I_t - number of persons that are infected
R_t - number of persons that are cured
N – Total population
$s_t = S_t / N$
$i_t = I_t / N$
$r_t = R_t / N$
$s_t + i_t + r_t = 1$
$\beta = \gamma \bullet \alpha$

β – Persons infected by one infected person during a time interval
γ – Persons that are taking contacts with a infected person
α – infected persons from γ, in %
s_t, i_t, r_t – population at the moment t in time
k – Recovering coefficient

$$t+1: \begin{cases} S_{t+1} = S_t - \beta S_t I_t \\ R_{t+1} = R_t + kI_t \\ I_{t+1} = I_t + \beta S_t I_t - kI_t = I_t(1 + \beta S_t - k) \end{cases}$$

$$t+1: \begin{cases} s_{t+1} = s_t - \beta s_t i_t \\ r_{t+1} = r_t + ki_t \\ i_{t+1} = i_t + \beta s_t i_t - ki_t = i_t(1 + \beta s_t - k) \end{cases}$$

$$s_{t+1} + i_{t+1} + r_{t+1} = s_t + i_t + r_t = 1$$

2.2 Model virus spreading between two populated areas

A Bayesian network is a probabilistic graphical model that represents a set of variables and their probabilistic independencies, and are used in the domain of artificial intelligence and statistics. Formally, Bayesian networks are directed acyclic graphs whose nodes represent variables and whose arcs encode conditional independencies between the variables.

If there is an arc from node *A* to another node *B*, *A* is called a *parent* of *B*, and *B* is a *child* of *A*. The set of parent nodes of a node X_i is denoted by parents(X_i). A directed acyclic graph is a Bayesian

Network relative to a set of variables if the joint distribution of the node values can be written as the product of the local distributions of each node and its parents (***):

$$P(X_1,\ldots,X_n)=\prod_{i=1}^{n}P(X_i \mid parents(X_i))$$

f node X_i has no parents, its local probability distribution is said to be *unconditional*, otherwise it is *conditional*. If the value of a node is *observed*,

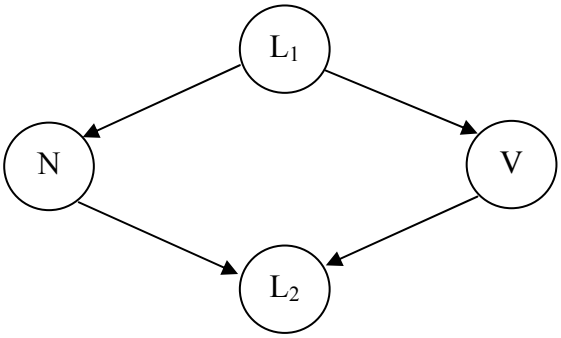

Figure 1. The Bayes network for virus spreading.

then the node is said to be an *evidence* node.
In figure 1 the Bayes network for virus spreading is presented.

The spread model equations for avian influenza are:

$$P(N\Lambda V\Lambda L_2)=P(L_2 \mid N\Lambda V)P(N)P(V)$$
$$P(\overline{N}\Lambda V\Lambda L_2)=P(L_2 \mid \overline{N}\Lambda V)P(\overline{N})P(V)$$
$$P(N\Lambda \overline{V}\Lambda L_2)=P(L_2 \mid N\Lambda \overline{V})P(N)P(\overline{V})$$
$$P(\overline{N}\Lambda \overline{V}\Lambda L_2)=P(L_2 \mid \overline{N}\Lambda \overline{V})P(\overline{N})P(\overline{V})$$
$$P(L_2)=P(N\Lambda V\Lambda L_2)+P(\overline{N}\Lambda V\Lambda L_2)+$$
$$+P(N\Lambda \overline{V}\Lambda L_2)+P(\overline{N}\Lambda \overline{V}\Lambda L_2)$$

$$P(L_2 \mid N\Lambda V)=p_1$$
$$P(L_2 \mid \overline{N}\Lambda V)=p_2$$
$$P(L_2 \mid N\Lambda \overline{V})=p_3$$
$$P(L_2 \mid \overline{N}\Lambda \overline{V})=p_4$$
$$P(N)=p_N$$
$$P(\overline{N})=1-p_N$$
$$P(V)=p_V$$
$$P(\overline{V})=1-p_V$$

L_1 – populated area from where the virus is started
L_2 – populated area where we which the virus can appear
N – Commuters from L_1
V – Potential tourists from L_1

P(N) – probability that in L1 exist commuters
P(V) – probability that in L1 exist tourists

$P(L_2 \mid N\Lambda V)$ – probability of arriving in L_2 commuters and tourists
$P(L_2 \mid \overline{N}\Lambda V)$ – probability of arriving in L_2 only tourists
$P(L_2 \mid N\Lambda \overline{V})$ – probability of arriving in L_2 only commuters
$P(L_2 \mid \overline{N}\Lambda \overline{V})$ – probability of not arriving in L_2 tourists and commuters
p_1, p_2, p_3 and p_4 – statistical data

P(N), P(V), p_1, p_2, p_3 and p_4 are chose arbitrary.

3. SIMULATION AND TESTING

A simulation based on agent technology is proposed. The models presented above are used to determine the evolution in time of the infected individuals in a populated area and between them.
Agents may wander randomly over the space and have a position and velocity. During each time step of the simulation, agents expend an amount of energy proportional to their volume to move and metabolise. At each simulation time step, energy is gained by an agent from the environment. Agents exhausting their energy supply "die" and are removed from the simulation. Agents also age throughout a simulation and are removed if they reach the end of their lifespan.
The agents in the model may carry virtual diseases, transmit them to other agents and succumb to infection themselves. The diseases in the simulation co-evolve alongside the agent population but may only exist *within* a host agent i.e. disease does not persist in the environment. A susceptible agent is exposed to a disease when it intersects with an infective agent. An agent that is carrying a disease cannot be infected by a second disease (i.e. an active disease blocks secondary infection). If an agent is not carrying a disease its susceptibility is determined. Simulation diseases also possess a *devastation* value that measures the virulence of a disease. This parameter is used to scale the probability of infection and the amount of energy required of a host to survive a time step of infection.
A parameter determines the lifespan of a simulation disease in each host. Long-lived diseases require a host to invest substantial amounts of energy to overcome infection. If a disease is overcome without the death of the host, the agent acquires immunity to the strain of the disease by adding it to an *immunity list*. Any further contact with this disease will result in an *immune response* that prevents the disease from infecting the agent a

second time. If a disease kills its host, or the host dies for any reason, the disease it carries dies also, irrespective of its lifespan.
The parameters for the disease and agents outlined fully specify the features of a epidemic models discussed above. A complex and flexible simulation has been devised that allows for studies of epidemics in non-homogeneous populations with nonrandom mixing. This model eliminates many of the problems in earlier epidemiological models.
Depending on the parameters of the disease, the traits of the infected agent and the population as a whole, the disease may or may not cause an epidemic.

Some observed outcomes are described below along with the conditions giving rise to them in the present simulation environment.

Disease elimination (immediate). If the disease is insufficiently long-lived, or the population is insufficiently dense, or the host does not co-habit with others, then the disease may fail to contact any susceptibles before it dies within the host. The disease will be eliminated from the population immediately.

Disease spread (immediate). A disease may mutate sufficiently within a host to infect susceptibles neighbors significantly different to the original host. If the host mixes amongst others of its kind they may become infected with the disease also.

Disease elimination (eventual). If the disease manages to take a hold in the population it may nevertheless die out eventually if the number of susceptibles is reduced. This may happen when a sizeable proportion of the agents encountered by infectives is immune to the disease (even though the population as a whole may not have a significant number of immune members). Circumstances like this arise when agents overcome the disease and acquire immunity, or when the disease is so devastating that it rapidly wipes out the supply of susceptibles before the agents are able to produce many offspring.

Disease spread (continual). A disease well-suited to its environment has sufficient lifespan to ensure it is passed from one susceptible agent to another. Such a disease also needs to be sufficiently devastating that it can be transferred successfully, but not so devastating that it kills off its supply of susceptibles. Diseases that fit these criteria also have to be sufficiently stable to avoid unwanted mutations that would render them ineffective, but sufficiently mutable so that they can keep infecting an evolving population of hosts. The simulation has given rise to diseases that meet all of these criteria and persist in the population for long periods of time.

3.1 Test case

A. Spreading inside of a populated area

Iteration step	Susceptible	Infected	Cured
1	0.9998	0.0002	0.0
2	0.9997	0.0003	0.0
3	0.9996	0.0004	0.0
4	0.9995	0.0005	0.0
5	0.9993	0.0007	0.0
…	…	…	…
20	0.9732	0.0118	0.0148
21	0.9662	0.0149	0.0187
22	0.9575	0.0187	0.0237
23	0.9468	0.0232	0.0298
24	0.9335	0.0288	0.0375
…			
40	0.3801	0.1037	0.5161
41	0.3564	0.0931	0.5503
42	0.3365	0.0823	0.5811
43	0.3198	0.0718	0.6082
44	0.3060	0.0619	0.6320
45	0.2947	0.0528	0.6524
46	0.2853	0.0447	0.6698
47	0.2777	0.0376	0.6846
48	0.2714	0.0314	0.6970
49	0.2663	0.0262	0.7074
50	0.2621	0.0217	0.7161

In figure 2 the evolution of infected population is presented.

B. Spreading between two populated area

Suppose that avian influenza appeared in Floresti, Cluj County and the simulation model was applied to test the apparition of virus in Cluj-Napoca, near Floresti. Figure 3 represents the Bayes network for virus spreading.

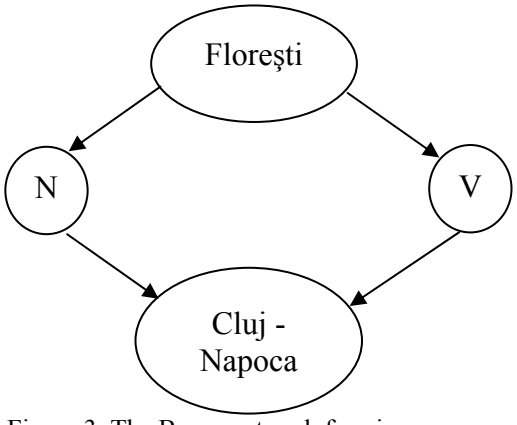

Figure 3. The Bayes network for virus spreading. Test Case.

Input data:

$$P(N) = 0.8$$

$$P(V) = 0.5$$
$$P(\overline{N}) = 1 - 0.8 = 0.2$$
$$P(\overline{V}) = 1 - 0.5 = 0.5$$

$$P(Cluj - Napoca \mid N \Lambda V) = 0.8$$
$$P(Cluj - Napoca \mid \overline{N} \Lambda V) = 0.2$$
$$P(Cluj - Napoca \mid N \Lambda \overline{V}) = 0.7$$
$$P(Cluj - Napoca \mid \overline{N} \Lambda \overline{V}) = 0.1$$

$$P(N \Lambda V \Lambda Cluj - Napoca) =$$
$$= 0.8 \bullet 0.5 \bullet 0.8 = 0.32$$

$$P(\overline{N} \Lambda V \Lambda Cluj - Napoca) =$$
$$= 0.2 \bullet 0.5 \bullet 0.2 = 0.02$$

$$P(N \Lambda \overline{V} \Lambda Cluj - Napoca) =$$
$$= 0.8 \bullet 0.5 \bullet 0.7 = 0.28$$

$$P(\overline{N} \Lambda \overline{V} \Lambda Cluj - Napoca) =$$
$$= 0.2 \bullet 0.5 \bullet 0.1 = 0.01$$

$$P(Cluj - Napoca) =$$
$$= 0.32 + 0.02 + 0.28 + 0.01 = 0.63$$

CONCLUSIONS AND FUTURES WORKS

In this paper an intelligent model is proposed. Several applications of the intelligent models were used to simulate different types of systems but not for avian influenza. Previous works in this filed are open for public and makes the documentations difficult.

The models for virus spreading into a populated area and between two populated are presented. The proposed models were tested and the results are presented above.

REFERENCES

Bailey, N.J.T. (1975). *The Mathematical Theory of Infectious Diseases and Its Applications.* New York: Oxford University Press.

Barabási A.-L. and Albert R. 1999. *Emergence of scaling in random networks*. Science, 286:509, 512.

Barabási. A.-L. and Dezsö Z. (2002). *Halting viruses in scale-free networks*. Physical Review E.

Durrett, R., Levin, S. (1994). *The Importance of Being Discrete (and Spatial).* Theoretical Population Biology 46:3 (1994) 363-394

Jones, D. S. and Sleeman. (1983). *B. D. Ch. 14 in Differential Equations and Mathematical Biology.* London: Allen & Unwin.

Kephart J. O. and White. S. R. (1991). *Directed-graph epidemiological models of computer viruses.* In Proceedings of the 1991 IEEE Computer Society Symposium on Research in Security and Privacy.

Kephart, J.O. and White, S.R. (1993a). *Measuring and Modeling Computer Virus Prevalence*. IEEE Computer Security Symposium on research in Security and Privacy. Oakland, California.

Kephart J. O. and White. S. R. (1993b). *Measuring and modeling computer virus prevalence*, Proceedings of the 1993 IEEE Computer Society Symposium on Research in Security and Privacy, May 1993.

Kephart, J.O., (1994). *How Topology Affects Population Dynamics*. Langton, C.G. (ed.) Artificial Life III. Reading, MA: Addison-Wesley, 1994.

Martins, M.L., Ceotto, G., Alves, S.G., Bufon, C.C.B., Silva, J.M., Laranjeira, F.F. (2000). *A Cellular Automata Model for Citrus Variagated Chlorosis.* eprint arXiv : cond-mat/0008203

Pastor-Satorras, R. and Vespignani, A. (2001a). *Epidemic Dynamics and Endemic States in Complex Networks*. Computer Viruses as Artificial Life. Journal of Artificial Life.

Pastor-Satorras R. and Vespignani A. (2001b). *Epidemic spreading in scale-free networks*. Physical Review Letters.

Pastor-Satorras R. and Vespignani A. (2002a). *Epidemic dynamics in finite size scale-free networks.* Physical Review E, 65:035108.

Pastor-Satorras R. and Vespignani. A. (2002b). *Epidemics and immunization in scale-free networks*. In S. Bornholdt and H. G. Schuster, editors, Handbook of Graphs and Networks: From the Genome to the Internet. Wiley-VCH, Berlin.

Pastor-Satorras R. and Vespignani A. (2002c). *Immunization of complex networks*. Physical Review E, 65:036104.

Staniford S., Paxson V. and Weaver N. (2002). *How to own the internet in your spare time.* In Proceedings of the 11th USENIX Security Symposium.

Tassier T. (2005). *SIR Model of Epidemics*, Anual report.

Wang, C., Knight, J.C. and Elder, M.C. (2000). *On Computer Viral Infection and the Effect of Immunization*. IEEE 16th Annual Computer Security Applications Conference.

Willox, R., Grammaticos, B., Carstea, A.S., Ramani, A. (2003). *Epidemic Dynamics:*

Discrete-Time and Cellular Automaton Models. Physica A 328.

***, *A Brief Introduction to Graphical Models and Bayesian Networks* - http://www.cs.berkeley.edu/~murphyk/ Bayes/bayes.html

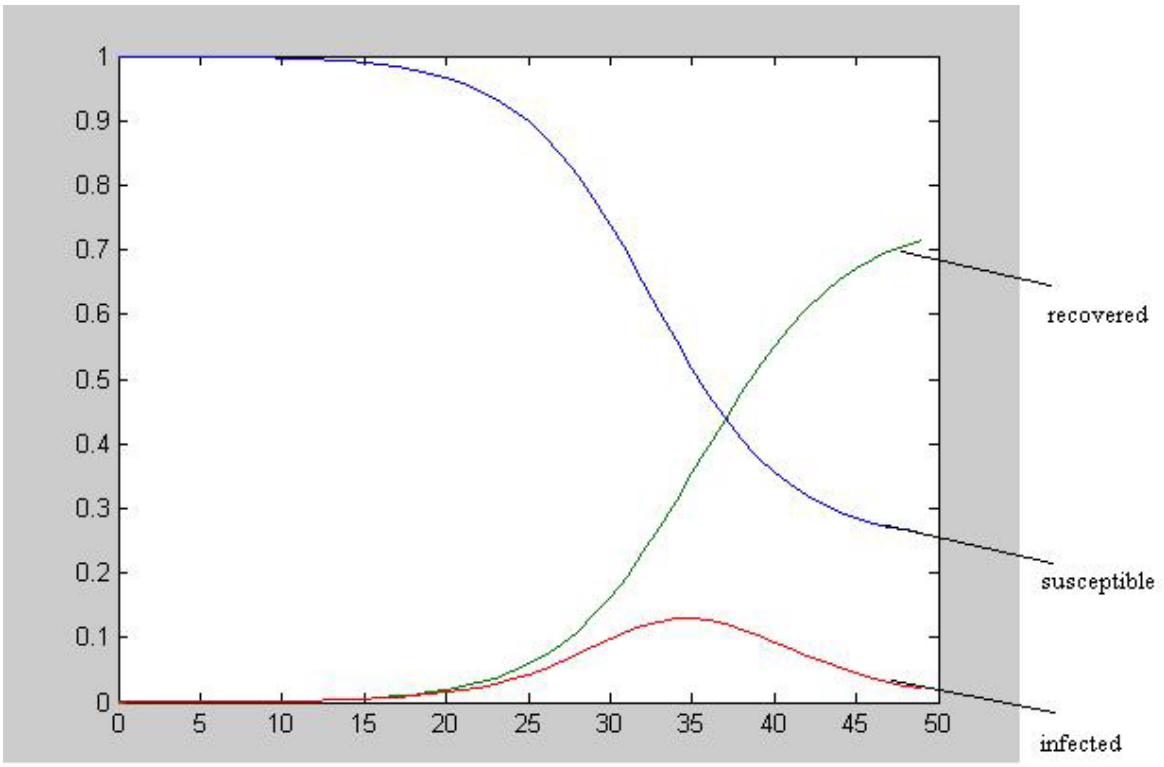

Figure 2. The evolution of infected population