HETEROGENEOUS EPIDEMIOLOGIC INFLUENCE SPREADING MODELS IN COMPLEX NETWORKS

Conference Paper · January 2009

CITATIONS
0

READS
28

3 authors:

Saso Gramatikov
Ss. Cyril and Methodius University in Skopje
33 PUBLICATIONS **77** CITATIONS
SEE PROFILE

Dimitar Trajanov
Ss. Cyril and Methodius University in Skopje
158 PUBLICATIONS **624** CITATIONS
SEE PROFILE

Ljupco Kocarev
Macedonian Academy of Sciences and Arts
416 PUBLICATIONS **15,434** CITATIONS
SEE PROFILE

HETEROGENEOUS EPIDEMIOLOGIC INFLUENCE SPREADING MODELS IN COMPLEX NETWORKS

Sasho Gramatikov[1], Dimitar Trajanov[1], Ljupco Kocarev[1]

[1] Ss. Cyril and Methodius University – Skopje, Macedonia
sasho@feit.ukim.edu.mk, mite@feit.ukim.edu.mk, lkocarev@feit.ukim.edu.mk

***Abstract* – Many of the systems around us are connected in networks with complex patterns forming a complex networks. The individuals of these networks interact among each other tending to impose theirs own state to the surrounding individuals. Such tendencies cause dynamic processes in complex networks defined as influence spreading. There is a big diversity of such processes, and their research gives important results for predicting the speed and rate of spreading of many processes like natural viruses, computer viruses, social processes etc.**

***Keywords* – complex networks, influence**

1. INTRODUCTION

The necessity of modeling natural deceases spreading gives an outbreak of many mathematical models [1][2][3][4] which analytically predict the behavior of systems where a population is initially infected. The main purpose of these models is predicting the scale of spreading in certain moment of time, given the initial state and the infection and healing rates. Although these models are initially dedicated to disease spreading, they quickly become popular for describing computer virus spreading, viral marketing, gossip spreading and other social processes.

The main issue of the analytical models of disease spreading is the fact that they do not put an accent on the connectivity of the population. Therefore we propose a nondeterministic model where we observe the population as a complex network of nodes connected to each other. The infection and recovery rate are defined through probabilities of changing state of a single node, giving the opportunity to define more complex behavior of the model. Depending on the states that one node can have, several epidemiologic models are defined: SIR (Susceptible - Infected - Recovered), SIS (Susceptible - Infected - Susceptible) and SI (Susceptible -Infected).

2. SIR

Let G(*D*) is symmetric directed graph defined with the adjacency matrix *D* with dimensions *n*x*n* where *n* is the size of the network. *D* is such a matrix that contains value 1 on position (*i*,*j*) only if a link from *i* to *j* exists. At any given time, each node can be in one of the three possible states: S – susceptible to infection, I – infected and R – recovered or removed. Initially nodes are susceptible to infection, and due to the influence from the nodes they are connected to, there is a positive probability that each node gets infected. Once a node is infected it can only stay in the same state or become recovered with certain probability. The process is irreversible, meaning that infected node can not become susceptible, nor can recovered node become infected.

Let $S_i(k)=[s_i^S(k)\ s_i^I(k)\ s_i^R(k)]$ is state vector which represents the state of node *i* in time *k*. This vector can contain only one entry with value 1. The other values equal 0. If node *i* is in state S at time *k*, then $s_i^S(k)=1$, if *i* is in state I than $s_i^I(k)=1$ and if it is in stat R than $s_i^R(k)=1$. Let $P_i(k)=[p_i^S(k)\ p_i^I(k)\ p_i^R(k)]$ is probability vector of node *i* at time *k*. Each value of this vector represents a probability that node *i* will be in one of the three possible states. Thus $p_i^S(k)$ represents the probability that node *i* will be in state S at time *k*. Seemingly, $p_i^I(k)$ and $p_i^R(k)$ represent the probabilities of being in state I and R, respectably. The evolution of the SIR model in time is defined with the following equations:

$$p_i^S(k+1)=1-p_i^I(k+1)-p_i^R(k+1) \tag{1}$$

$$p_i^I(k+1)=s_i^S(k)\left[1-\prod_{j=1}^{n}(1-\beta d_{ij}s_i^I(k))\right]+(1-\alpha)s_i^I(k) \tag{2}$$

$$p_i^R(k+1)=\alpha s_i^I(k)+s_i^R(k) \tag{3}$$

In these equations β is the probability that infected node will infect the neighboring susceptible nodes and α is the probability that infected node will become recovered. The probability that node *i* will be in state I at time *k*+1 equals the probability that the node will be infected if it was previously susceptible

or the probability that the node will remain in state I if it was previously infected. The probability that a susceptible node will be infected equals 1 minus that probability that it will not be infected. The last probability is product of the probabilities that none of the neighbor nodes will infect the susceptible node. The probability that a node will be in state R at time k+1 equals the probability of recovery if the node was previously in state I or 1 if it was already in state R. d_{ij} in (2) is an entry of the adjacency matrix D and has value 1 only if a link between the nodes of interest exist. Equation (3) shows that once in a state R, the node will remain in that state till the end of the process, and that does not depend on the network structure. Once the probabilities are determined, we can easily calculate the probability of state S, since each node must be in any of the three states, and therefore the sum of all probabilities equals 1.

The definition of the model is not complete without laying out its initial state. Let all the nodes are healthy and susceptible to infection. In such case, according to the definition of state vector, the initial state of each node is $S_i(0) = [1\ 0\ 0]$. In order to initiate dynamic process in the network we infect one node which results with a state described with the state vector $S_i(0) = [0\ 1\ 0]$. After applying (1)-(3) the probabilities of the states are calculated, and the next state of the network is determined with a process similar to throwing a dice with three sides i.e. for each node a random number in the range [0,1] is generated, and compared to the state probabilities of that node. If the number is in the range [0, $p_i^S(k+1)$] the final state well be S, if the number is in the range [$p_i^S(k+1)$, $p_i^S(k+1) + p_i^I(k+1)$], the node will be in state I. Otherwise it will be in state R.

Equations (1)-(3) define a homogeneous model where the probability β that a node will infect/influence other nodes is the same for all nodes in the network. However, in reality not all the nodes have the same importance and capability of influencing other nodes. Therefore, we define a heterogeneous model where the probability β will be function of the node importance. The greater the importance of a node, the greater the probability that it will influence the neighboring nodes.

The main question that arises is how to determine the importance of a node. What we propose in this paper, is determining the importance of a node according to some of the properties of networks defined in graph theory [5]. The simplest form of defining the importance of nodes is considering their out degree i.e. the number of outgoing links. Nodes with large out degree are influent because they are connected to a large number of nodes in the network. However, this might not always be true because the degree of a node is a property with a local character. It does not give any information about the position of the node in the overall network. A property that has a global character is the node betweenness centrality [6]. This property is a measure of the importance of a node in a network, and is calculated as the fraction of shortest paths between node pairs that pass through the node. Let G is a graph given with set of nodes V and set of edges E. Let s and t are two nodes of the graph. σ_{st} is the number of paths that pass from s to t. Let $\sigma_{st}(v)$ is the number of shortest paths that pass through the node v. The central betweenness of node v is:

$$C(v) = \sum_{s \neq v \neq t \in V} \frac{\sigma_{st}(v)}{\sigma_{st}} \qquad (4)$$

Once we have the importance of the nodes, the next step would be defining the probability of infecting the neighboring nodes β_i. We assign a value of β_i which will be proportional to the node betweenness centrality. The values of β_i are within the range [$\beta_{\min}$, $\beta_{\max}$] where $\beta_{\min}$ and $\beta_{\max}$ are the minimum and maximum values that can be assigned to any node in the network . Let $C(i)$ is the betweenness centrality of node i defined in (4) and C_{min} and C_{max} are the values of betweenness centrality of the least and most important node in the network. The infection probability β_i is defined as:

$$\beta_i = \beta_{\min} + \frac{\beta_{\max} - \beta_{\min}}{C_{\max} - C_{\min}} (C(i) - C_{\min}) \qquad (5)$$

According to (5) each node gets portion of infection probability proportional to the node betweenness.

3. COMPLEX NETWORKS

Since one of our main goals is to analyze the behavior of different real network topologies, we use different types of complex networks. Complex networks have certain properties that make them different from aspect of topology. The difference comes from the way nodes are connected among each other. According to the inter-link dependences several types of network topologies are defined.

3.1. Random Networks

The simplest and most straightforward realizations of complex networks are random networks. These networks are characterized by nodes that are randomly connected to each other, with certain probability p *[9]*. For networks with large number of links, the average number of links per node is the same and the degree distributed follows the Poisson distribution. This fact shows that the probability that a node will have large deviation from the average value is exponentially small.

3.2. Geographic Random Networks

A special case of random networks are geographically random networks. These networks are characterized by nodes that are randomly distributed in the space, and are connected only to the nodes in

their proximity. A typical example of random geographic network is wireless ad hoc network where each wireless station is connected to the stations that are within its range of coverage.

3.3. Small-world Networks

According to the link structure, small world networks stand between random and lattice connected network. They are generated by randomly replacing fraction of links from d-dimensional lattice structure [7]. If the fraction equals zero, than the network is lattice, and if the fraction is one, than the network is random network. For fraction between the extreme values, we get a small-world network. The name of these networks comes from the property that the average shortest distance between two nodes increases logarithmically with the number of nodes. Therefore the wider the network, it is easier to connect two distant nodes with just a few links. Thus, although the network is large, at the same time it is small because any node is reachable in average a few steps. Small world networks are composed of highly connected clusters, in which very few nodes provide connectivity to the rest of the world by setting links with other clusters

3.4. Scale-free Networks

Scale-free networks have distribution of connectivity that decays with power low. The number of nodes with exactly k links follows a power law, each with a unique degree exponent. These networks are characterized by presence of nodes called hubs, with large number of links. These nodes are dominant in the structure of all scale-free networks, making each node from the network easily reachable from any point [7][8].

4. RESULTS

In the following section we present the results obtained by simulating the SIR model. In the simulations we observe four different network topologies: scale-free, small world, random and geographic random networks. Each of the networks is of size 500. Each simulation is executed in 10 independent iterations, and the mean value is presented as a final result. One of the key elements that has to be defined in order to start the dynamic process is the initial state. In our simulation we infect a node with average importance. After calculating the importance of the nodes according to (4) they are sorted and the node with importance closest to the mean importance is selected as an initiator of the SIR process.

Fig 1 shows the number of susceptible, infected and recovered nodes during time for a homogeneous SIR model for scale-free network where the probability of infection is $\beta = 0.1$ and the recovery probability $\alpha = 0.2$. At the beginning all the nodes are in susceptible state and during time, due to the infection spread by the initially infected node, the number of infected nodes increases. As this number increases, so does the number of recovered nodes. After reaching a peak value, the number of infected nodes falls to 0 value, and the number of recovered nodes reaches its stationary state. From the figure it is obvious that the stationary state of the homogeneous SIR model does not result in complete spreading of the disease. It reaches only portion of the nodes because the rate of recovery is larger than the rate of infection and after a certain period there are not infected nodes that can cause any further spread of the disease.

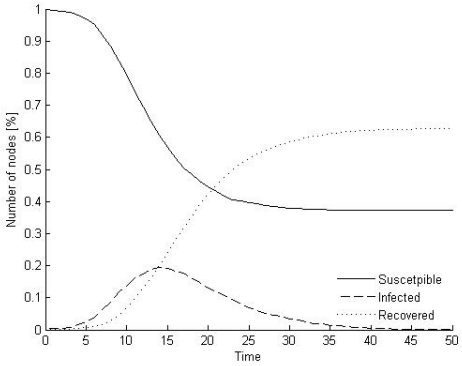

Fig. 1 - Number of population in state S, I and R for a homogeneous SIR model for $\beta = 0.1$ and $\alpha = 0.2$

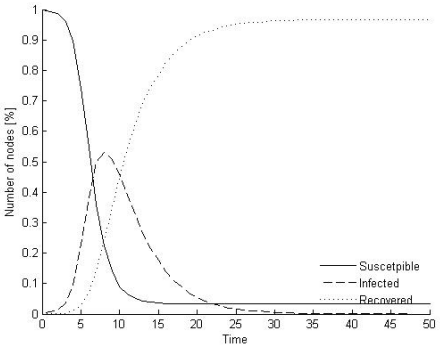

Fig. 2 - Number of population in state S, I and R for a homogeneous SIR model for $\beta = 0.25$ and $\alpha = 0.2$

Fig 2 shows the number of nodes in certain states for a homogeneous SIR model for the same network with probability of infection $\beta = 0.25$ and recovery probability $\alpha = 0.2$. In this case the infection reaches almost every node resulting with network where most of the odes are recovered. It can be also seen that the stationary state is reached much faster than the previous case.

The number of nodes in certain states for a heterogeneous SIR model for a scale-free network is shown on fig. 3. The heterogeneity is determined according to (2) and the range of values for infection probability is [0.05, 0.45]. The recovery probability for the model is $\alpha = 0.25$. From the figure we can see that the peak of the number of infected nodes reaches higher values for a shorter period of time. This number is efficient for reaching stationary state where all the nodes are in the recovered state.

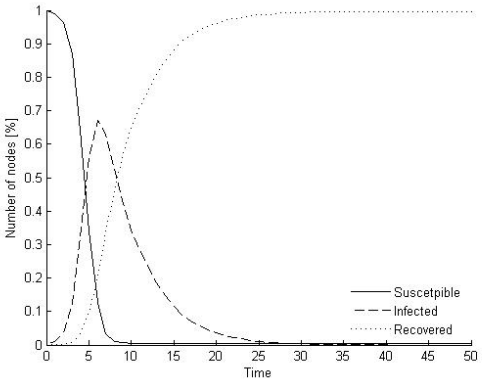

Fig.3 - Number of population in state S, I and R for a heterogeneous SIR model for $\beta = [0.05, 0.45]$ and $\alpha = 0.2$

The number of recovered nodes for different network topologies for homogeneous SIR model is shown on figure 4. The SIR model is defined with probability of infection $\beta = 0.25$ and recovery probability $\alpha = 0.25$. From the figure we can see that random networks are most prone to spreading influence because they reach highest percentage of recovered nodes. Scale-free and small world networks reach almost the same values of recovered nodes, however small world networks reach the stationary state faster than any topology. Geographic random networks are the most inert to spreading the influence both from aspect of coverage and speed of convergence.

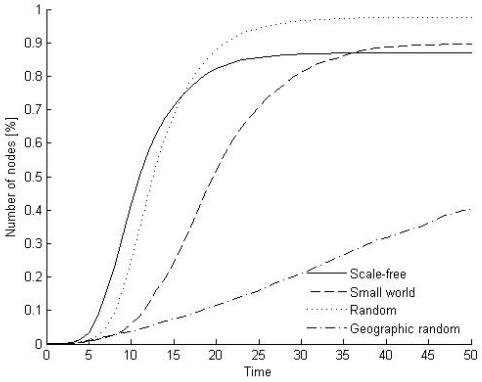

Fig. 4 - Number of recovered nodes for a homogeneous SIR model for $\beta = 0.25$ and $\alpha = 0.2$

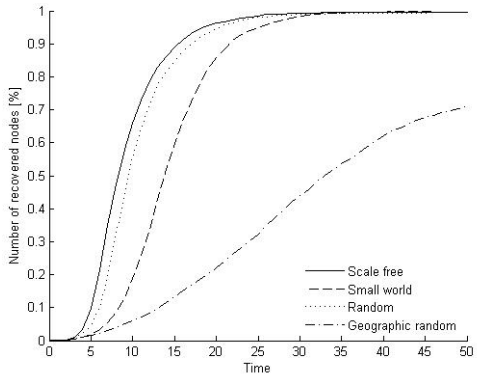

Fig. 5 - Number of recovered nodes for a heterogeneous SIR model for $\beta = [0.05, 0.45]$ and $\alpha = 0.2$

Fig 5 presents the number of recovered nodes for different network topologies for a heterogeneous SIR model. In this case all the topologies except the geographic random reach a complete spreading of influence. The speed of spreading is fastest for the scale-free networks. Slightly slower is the speed of random networks, and than follow small world networks. Although geographic random networks in the heterogeneous model spread the influence with greater intensity and speed than the homogeneous model, the influence is not spread in the entire network.

5. CONCLUSION

Results of the simulations show that the spreading of the SIR model depends mainly on the infection and recovery probabilities. When the recovery probability has higher value than the infection probability, the process is spread almost to a completion. The topology has also major impact on the speed and scale of spreading the process. Scale-free networks are the fastest spreading topologies whilst random networks reach the highest values of recovered nodes. In any of the observed models, the geographic random networks spread the SIR model with smallest rate and scale. From the comparison of the results of homogeneous and heterogeneous SIR model we can conclude that the first model converges to stable state with lower values of recovered nodes and with lower speed, although the results do not differ significantly.

6. REFERENCES

[1] R. Pastor-Satorras, A. Vespignani: Epidemic Spreading in Scale-Free Networks, Physical review letters, vol. 86, no14 , 2001

[2] T. Tassier: SIR Model of Epidemics, Journal of Mathematical Biology, 39 (6), 1999

[3] T.Tassier: Introduction to The Modelling of Epidemics – SIS Models, Duke University, 2005

[4] N.T.J. Bailey, The Mathematical Theory of Infectious Diseases and its Applications. New York: Hafner Press, 1975.

[5] D. B. West, Introduction to Graph Theory, *Prentice Hall*, 2001

[6] Freeman, L. C. Centrality in social networks: Conceptual clarification. Social Networks 1, 215–239, 1979

[7] L. A. N. Amaral, A. Scala, M. Barthe lemy, H. E. Stanley: Classes of small-world networks, *PNAS 97:11149-11152*

[8] A. L. Barabasi: Linked, Penguin Group, London, May, 2003

[9] Michael Karonski and Adrzej Rucinski: The Origins of the Theory of Random Graphs, *The Mathematics of Paul Erdos, Berlin, Springer, 1997*