ISSN 2073-2244
www.alb-shkenca.org

AKTET
Journal of Institute Alb-Shkenca
Revistë Shkencore e Institutit Alb-Shkenca

CAN WE HAVE PANDEMICS IN THE INTERNET?
A MUND TË KEMI PANDEMI NË INTERNET?

MARGARITA IFTI[a*] and PAUL NEUMANN[b]
[a]Department of Physics, Faculty of Natural Sciences, University of Tirana, Bul. Zog I, Tirana, Albania
[b]Department of Computer Sciences, University"Kristal", Highway Tirana-Durres, km.5, Tirana, Albania
Email: margarita.ifti@unitir.edu.al

ABSTRACT

We use the three-species cyclic competition model (Rock-Paper-Scissors), described by reactions A + B → 2B, B + C → 2C, C + A → 2A, for emulating a computer network with e-mail viruses. Different topologies of the network bring about different dynamics of the epidemics. When the parameters of the network are varied, it is observed that very high clustering coefficients are necessary for a pandemics to happen. The differences between the networks of computer users, e-mail networks, and social networks, as well as their role in determining the nature of epidemics are also discussed.

PERMBLEDHJE

Ne përdorim modelin e garës ciklike mes tri specieve (Guri-Letra-Gërshërët), i cili përshkruhet nga reaksionet A + B → 2B, B + C → 2C, C + A → 2A, për të emuluar një rrjet kompjuterik me viruse që përhapen nëpërmjet postës elektronike. Topologji të ndryshme të rrjetit shkaktojnë dinamikë të ndryshme të epidemisë. Kur variojmë parametrat e rrjetit, vërejmë se, që të kemi pandemi, nevojiten koeficientë tepër të lartë agregimi. Ne diskutojmë ndryshimet ndërmjet rrjetave të përdoruesve të kompjuterave, të postës elektronike e rrjetave shoqërore, si edhe rolin që luajnë ato në përcaktimin e natyrës së epidemisë.

Key words: computer networks, computer viruses, topology, clustering.

INTRODUCTION

The Internet is a complex network of computers and routers, connected by wires or wireless links, the World Wide Web is a network of web pages connected by hyperlinks, the social networks are composed of human beings and social connections. All these are examples of complex networks, whose structure has interested scientists, particularly physicists, in recent years. Three concepts are central to the modern understanding of complex networks:

Small worlds: Despite the large size of the network, there is always a relatively short path between any two nodes. The best known example is the "six degrees of separation" idea, discovered by Stanley Milgram [15] who stated that there is a path of acquaintances with typical length of six between any pair of people in the United States.

Clustering: In social networks, particularly, we observe the formation of "cliques", representing circles of friends or acquaintances in which everybody knows everybody. This attribute is characterized by the clustering coefficient, defined for a node i with $k_i>1$ edges as follows: $C_i = 2E_i / [k_i (k_i-1)]$, where E_i is the number of links between the k_i nodes connected to the *i*-th node [20]. Their average over the network gives its clustering coefficient.

Degree distribution: The number of edges (links) for a particular node is known as the degree of that node. The degree distribution $P(k)$ for the real networks (World Wide Web [3], Internet [10], and metabolic networks [14]) has a power-law tail: $P(k) \approx k^{-\gamma}$. Networks with the above degree distribution (independent of N) are called scale-free [1].

The topology of the Internet is studied at the router level, and at the interdomain level. Faloutsos et al. [10] report power-law distributions with exponents around 2.2 at the interdomain level, and 2.48 at the router level. The clustering coefficient is 0.18 and 0.3 [17]. Similar topology has been observed in many social networks.

In previous works we have considered a non-spatial version of three-species competition models, and studied their long-term behaviour [12, 13]. They exhibit the extinction of one of the species, when there are no migrations, or survival of all three, when there are enough migrations into and out of the system.

However, these models do not correctly represent the real spatial structure of the interaction and migration networks [2]. In this paper we will present a model that relates to the spread and survival of e-mail viruses in the Internet.

THE *ABC* CYCLIC MODEL ON A SCALE-FREE NETWORK

The *ABC* model we have previously studied in the non-spatial version has relevance to epidemiological studies, since it resembles the case when the immunity to a particular infectious disease is only temporary, otherwise known in epidemiology as the *SIRS* (Susceptible-Infected-Recovered-Susceptible) model [11]. (The similarity is not present at all the steps within the *SIRS* model.) This is particularly relevant when one is interested in the spread of computer viruses in the Internet. We would like to be able to predict, just as for ordinary infectious diseases, whether they will flare up and die out, or remain endemic. Several questions arise, the answer to which will be determined by the topology of the system. First, let us discuss the relevance of the *SIRS* model to the computer viruses' spread.

COMPUTER AND E-MAIL VIRUS FEATURES

The very name of computer viruses comes from the obvious resemblance to natural viruses: their ability to self-replicate, high speed of spreading, the ability to select the target system (i.e. each virus only targets certain systems, or groups of systems), the ability to infect non-infected systems, our difficulties in fighting them, etc. Recently one can also add an increasing rate of mutation and the appearance of new generations of viruses. The global distribution of the viruses in a large degree is determined by the mass-production and popularity of personal computers, which are particularly vulnerable to infection, due to their standardised architecture and software.

Many viruses are able to mutate by modifying their own code randomly, so that it is difficult to establish certain fingerprinting characteristics of a given virus. It makes a reappearing virus unrecognisable to the antivirus software, causing reinfection. This circumstance justifies the use of a *SIRS* model for this class of viruses. The spread of computing networks created an environment on which the viruses would thrive. But the contrary is also true: computing networks create the environment in which the antiviruses can spread very quickly. Hence the spread of some sorts of computer infections (but not all of them) can be epidemiologically modelled by a *SIRS* model. Our *ABC* cyclic model, described by reactions A + B → 2B, B + C → 2C, C + A → 2A (all three at equal rates), is a good first realisation of the *SIRS* model.

MODELS OF NETWORKS WITH DIFFERENT TOPOLOGIES

Very good reviews in the field have been written by R. Albert and A.-L. Barabási [2], S.N. Dorogovtsev and J.F.F. Mendes [7], M.E.J. Newman [16], etc. We will give very short descriptions of relevant models.

A network is represented by a graph, i.e. a set of nodes and edges (links) between them. The theory of random graphs was founded by Erdös and Rényi [8, 9]. A random graph is defined as N nodes connected by n edges, randomly chosen among all possible edges [8]. Erdös and Rényi [9] showed that there is a sudden change in the cluster structure as the average degree per node $\langle k \rangle$ goes through 1: from a loose collection of small clusters, a large cluster is formed, which for $\langle k \rangle_c = 1$ has approximately $N^{2/3}$ of the nodes, and for larger $\langle k \rangle$ has a number of nodes proportional to network size N. For large N the degree distribution of a random network is Poisson with mean $\langle k \rangle$. Random networks have clustering coefficients that are much smaller than those observed in real-world networks.

The origin of the scale-free behaviour in networks was first addressed by Barabási and Albert [3]. They attributed the scale-free nature of the networks to two mechanisms present in many networks. The Barabási-Albert (BA) algorithm has two important ingredients:

(i) Growth: Start with a small (m_0) number of nodes; every step add a new node with $m(\leq m_0)$ edges which link the newly added node to m nodes already in the network.

(ii) Preferential attachment: The probability that the new node will be connected to node i depends on the degree k_i of node i: $P(k_i) = k_i / \Sigma k_j$.

This algorithm produces a scale-free network with an exponent equal to 3 for the power-law degree distribution. This model is a minimal one that captures the mechanisms responsible for the power-law degree distribution.

The CM is already the standard for simulating networks of a given γ. For each node, its degree k is drawn from the power-law probability distribution function. Then pairs of links are randomly placed between the nodes, until they get "saturated" [5, 6]. This model creates correlations in the connectivity distribution between the system nodes for γ<3. UCM uses the same algorithm for the construction of the network, except that the upper cutoff is fixed beforehand at $k_{max}=N^{1/2}$. Then the connectivity correlations disappear [4].

The spread of ideas, innovations, and computer viruses is influenced immensely by the network structure. Many studies of dynamics on networks have been and are being done [18, 19]. This area of study is a rapidly developing one.

SIMULATIONS OF *ABC* MODEL IN NETWORKS WITH DIFFERENT TOPOLOGIES

We tried to simulate the behaviour of the *ABC* model in a network, generated according to the rules described above. Of course, the above model does not completely describe real life networks, but it is a good toy model. In each case, the random networks, as well as scale-free networks (generated with BA, CM and UCM algorithm), were used for comparison.

In our approach to the simulation of our systems, we have to simulate continuous time Markov processes. In these processes the event times are exponentially distributed, and the characteristic time depends on the current configuration of the system. First, we initialize the simulation: the nodes are assigned an initial state at random, i.e. A, B, or C with probability 1/3, and then they are assigned a "wake up" time with exponential distribution. The node with the smallest time will be activated: if its state is, say, B, it will convert the nodes at state A among the ones it is connected to (emulating infection at contact), and analogously for other states. The process is then iterated.

NO PANDEMICS IN THE INTERNET!

We started our simulations with the scale-free (BA, CM, UCM) models. In all cases the number of A, B, C individuals drifts with time, but none of them becomes zero. This situation is observed in network sizes up to 50 000 and $\langle k \rangle$ up to 200. A typical time series of the simulations of the *ABC* cyclic system in a scale-free network is shown in Fig. 1.

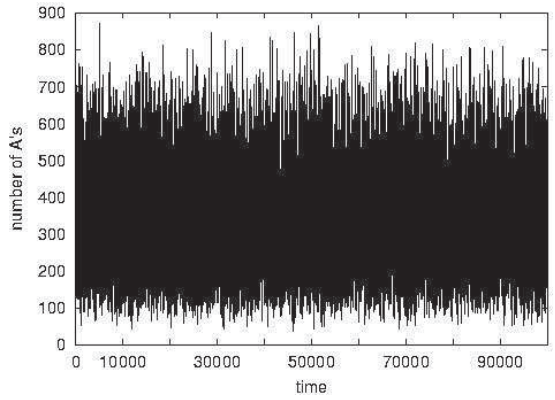

Figure 1. A typical time series for the *ABC* model in a scale-free network of size *N=900.*

This situation is very applicable to network viruses. Indeed we have seen viruses reappear several times in the real life. But does this mean that we cannot have pandemics in small-world and scale-free networks?

The computer users and e-mail networks appear as "diluted" versions of the social networks, so their clustering coefficient is lower, close to those obtained with the Barabási-Albert model, for which the *ABC* model does not reach fixation. As far as clustering goes, in one end of the spectrum are the social networks, which are highly clustered, with clustering coefficients around 0.2, while in the other end are the random networks, which are "maximally distributed". Runs performed on random networks were quite surprising: one would require enormous connectivities *(p)* to reach fixation on them! A plot of critical connectivity vs. network size for the random networks is shown in Fig. 2. The critical connectivity reaches values close to 0.36, which is enormous.

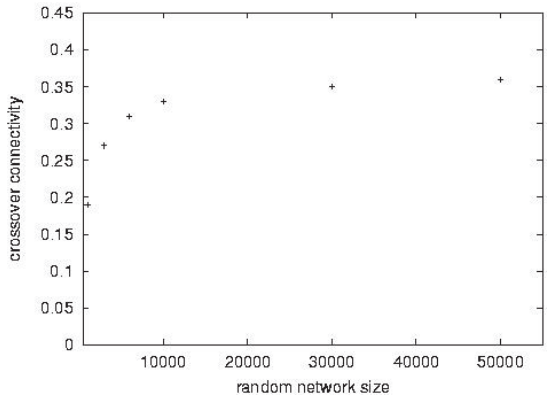

Figure 2. Critical connectivity vs. network size for random networks.

CONCLUSIONS

The modern computer and e-mail viruses, with their ability to mutate, or the absence of a reliable antivirus, call for *SIRS* epidemiologic models, rather than simply *SIR*. Simulations of the spread and evolution of such systems in different scale-free topologies show that all three types, (A, B, or C) are present in the network at all times. This result would relate to the observed peaks and valleys of epidemics. Runs performed on random networks show that the critical connectivities approach 0.36, which is quite unreachable in the real life networks. Apparently, the computer viral epidemics do not exhibit the same characteristics as human epidemics, because the clustering coefficients for the computer users/e-mail networks cannot achieve the critical values. It seems that the pandemics of a virus in the Internet is an unlikely event, unlike pandemics in human social networks. The Internet is threatened more by the side-effect of extreme congestion from the uncontrolled multiplication of viruses.

REFERENCES

1. R. Albert and A.-L. Barabási, Phys. Rev. Lett. **85**, 5234-37 (2000).
2. R. Albert and A.-L. Barabási, Rev. Mod. Phys. **74**, 47-97 (2002).
3. R. Albert, H. Jeong and A.-L. Barabási, Nature **401**, 130-131 (1999).
4. M. Catanzaro, M. Boguñá, and R. Pastor-Satorras, Phys. Rev E **71**, 027103 1-4 (2005).
5. R. Cohen, K. Erez, D. ben Avraham, and S. Havlin, Phys. Rev Lett. **85**, 4626-4629 (2000).
6. S.N. Dorogovtsev, J.F.F. Mendes, and A.N. Samukhin, Phys. Rev E **63**, 062101 1-4 (2001).
7. S.N. Dorogovtsev and J.F.F. Mendes, Adv. Phys. **51(4)**, 1079-1187 (2002).
8. P. Erdös and A. Rényi, Publ. Math. Debrecen **6**, 290-297 (1959).
9. P. Erdös and A. Rényi, Publ. Math. Inst. Hung. Acad. Sci. **5**, 17-61 (1960).
10. M. Faloutsos, P. Faloutsos and C. Faloutsos, Proc. ACM SIGCOMM, Comput. Commun. Rev. **29**, 251-263 (1999).
11. H.W. Hethcote, SIAM Review **42**, 599-653 (2000).
12. M. Ifti and B. Bergersen, Eur. Phys. J. E **10(3)**, 241-248 (2003).
13. M. Ifti and B. Bergersen, Eur. Phys. J. B **37**, 101-107 (2004).
14. H. Jeong, B. Tombor, R. Albert, Z.N. Oltvai and A.-L. Barabási, Nature **407**, 651-654 (2000).
15. S. Milgram, Psych. Today **2**, 60-67 (1967).
16. M.E.J. Newman, SIAM Review **45(2)**, 167-256 (2003).
17. R. Pastor-Satorras, A. Vázquez and A. Vespignani, Phys. Rev. Let. **87**, 258701 1-5 (2001).
18. R. Pastor-Satorras and A. Vespignani, Phys. Rev. Lett. **86**, 3200-3203 (2001).
19. R. Pastor-Satorras and A. Vespignani, Phys. Rev E **65**, 036104 1-9 (2002).
20. D.J. Watts and S.H. Strogatz, Nature **393**, 409-410 (1998).