Evolutionary Algorithm for Optimal Vaccination Scheme

To cite this article: K J Parousis-Orthodoxou and D S Vlachos 2014 *J. Phys.: Conf. Ser.* **490** 012027

View the article online for updates and enhancements.

Evolutionary Algorithm for Optimal Vaccination Scheme

K.J. Parousis-Orthodoxou and D.S. Vlachos

Department of Computer Science and Technology,
Faculty of Sciences and Technology,
University of Peloponnese,
GR-221 00 Tripolis, Greece

E-mail: `kp0100@uop.gr`

Abstract.   The following work uses the dynamic capabilities of an evolutionary algorithm in order to obtain an optimal immunization strategy in a user specified network. The produced algorithm uses a basic genetic algorithm with crossover and mutation techniques, in order to locate certain nodes in the inputted network. These nodes will be immunized in an SIR epidemic spreading process, and the performance of each immunization scheme, will be evaluated by the level of containment that provides for the spreading of the disease.

1. Introduction

The field of complex systems, has taken a lot of attention the recent years ([1], [2] and [3]), because of the number of applications that can be found in everyday life. Some examples of complex systems that are have been analyzed over the literature, are the Internet (in router level), the World Wide Web [4], the protein interaction network [5], the power grid of a country, the colonization of ants and several other examples. Many of these networks have shown very interesting properties, with the most important being the small-world effect [6] and the scale-free property [7].

The dynamical operations that can be applied in various complex systems, are the cornerstone of the research in the field. Many studies have focused on the approaches of the epidemic spreading ([8], [9] and [10]) of various types (SIS, SIR etc) and in different types of networks (scale-free, random etc). Also, many of these studies have been concentrated in locating optimal vaccination schemes to control the spreading of a disease or information in distinct network topologies ([11], [12], [13] and [14]). These two different cases might be focused on maximizing the diffusion (information) or minimizing it (disease) within the network.

Another dynamical operation, which has been proven to be an extremely useful tool in obtaining optimal solutions to various problems, are the evolutionary algorithms. These type of process, mimics to a certain point, the process of evolution that has been met in most ecosystems. Using certain concepts such as mutation, chromosome or crossover, these algorithms can be modified in such a way as to locate global extrema and minimize the time of the operation ([15] and [16]).

There have been some attempts to merge the two fields, either for optimizing the effect of the application of a genetic algorithm in a problem, or by using a genetic algorithm in order

to explore a certain property of a network. An example of the first type is the attempt to distribute the population of a genetic algorithm onto the nodes of a complex network, and allow the crossover and mutation processes only between the population members that reside on these nodes [17]. The second type of merging of the two pioneering topics involves using a certain type of genetic algorithm as a tool that balances the exploration-exploitation trade off, in order to explore a specific property of a network [18], and the study that is presented here falls into that category.

2. Algorithm Description

In this section, we will analyze the parts of the algorithm that operate simultaneously. Our algorithm is divided in two categories, the genetic algorithm and the epidemic spreading process. Aiming for the optimal vaccination scheme, the genetic algorithm creates random vaccination strategies in a binary form. The fitness criteria of the genetic algorithm are given by runs of SIR epidemic spreading model, which is described below.

2.1. The SIR Model

The specific model of epidemic spreading operates with basic principle that there are three pools which contain all the nodes of the network. At the beginning of the spreading all the nodes (except from the vaccinated ones) are in the Susceptible pool. As the spreading progresses, the nodes are transferred with probability $p_i = 0.76$ to the Infected pool. Then with probability $p_c = 0.002$ are transferred to the Recovered pool. The key component of the SIR model is the fact that, as soon as a node has recovered from the infection, it can no longer carry it again. This is why the three states of the model are Susceptible Infected Recovered. The vaccinated nodes are considered to be in the Recovered state as soon as the infection starts to spread. For the purposes of fairness to the vaccination schemes that are produced, the patient zero of the spreading, is chosen in the beginning of the process and is the same for all fitness evaluations that are performed during the run of the genetic algorithm.

2.2. The Fitness Function

The fitness function produces outcomes that are based on the cost of each vaccination scheme, which factors the number of vaccinated nodes of the scheme and the number of infected nodes during the spreading process. The cost of each vaccine is 5 whereas the cost of treating each infected node is 1. Therefore, the dominating scheme in each generation should use an optimal number of vaccines, that would allow for a minimal percentage of infected population. Note that since the cure/vaccine ratio is 1/5, it is never profitable to vaccinate more than 1/5 of the population. Therefore, it is in each case, the maximum allowed vaccination size.

2.3. The Genetic Algorithm

The process of the genetic algorithm initializes with locating certain basic attributes of the network that is inputted (number of nodes, degrees etc). These values will define some of the parameters of the process, such as the population (number of nodes). Then, the algorithm fills the population with vaccination schemes in binary form (0 for the non-vaccinated nodes and 1 for the vaccinated ones) and evaluates their fitness with the use of the function described above. The fitness is then stored uniformly in a list that will be used as reference for obtaining the fittest. Note that one of the random parameters for the vaccination scheme is the number of nodes that will be vaccinated.

When the initial population is set, the algorithm begins the evolutionary process. We choose two nodes with the use of the *roulette function*. That is, by choosing randomly, but with greater possibilities for the fittest of the population. This ensures that the choice will be fair, since

the fittest have an advantage, but also not excluding the share of the population that did not perform so well and might yield more promising results if coupled with someone else.

After the two 'parents' are chosen, they crossover with probability $p_{cross} = 0.7$ and produce two offsprings with half their chromosomes (single point crossover in the middle of the chromosome). If crossover is not performed, the mutation process will be explored with probability $p_{mut} = 0.002$ for each of the initially chosen members of the population. If the mutation is a go, one of the bits of the chromosome are reversed and the result is forwarded to the next generation. In each case, if one of the operations is performed (crossover,mutation) the resulting chromosomes take the place of the initially chosen, whereas, if no operation is performed, the original chromosomes are transfered into the population intact. This function continues until the entire population is rebuilt. The number of times that the function is executed is half the population size (since it produces two members at each time step).

When the population for each time step is completed, the algorithm creates the uniform fitness for it, and restarts the process for the next time step. This carries on for user defined generations. After the final generation, the algorithm locates and returns the optimal generation scheme for the specific input parameters and its cost.

3. Experimental Procedure and Results

In this section, we will present the networks that were used for the experimental procedure of the algorithm, the results they yielded as well as an analysis of the vaccination schemes that were ultimately chosen. Each network has been tested 50 times and the results allow us to locate the best fitted schemes.

The experiments that are carried out, are performed in three distinct types of networks, which mean to cover a broad spectrum of the field. The first network that is presented, is produced by the preferential attachment model of Barabasi and Albert [7]. This infamous model in complex networks, produces undirected networks with the PA rule, which gives scale-free networks. The number of nodes for this graph, as well as the rest, is 500, whereas the number of edges differ from graph to graph and in this case they are 1491. The mean number of edges is 3. The BA graph has given some interested results regarding the optimal vaccination schemes in each of the 50 runs. Figure 1 shows the number of vaccinated nodes in each of the 50 runs that were performed in the BA network. The largest number of vaccinated nodes is 89, which means that only for vaccination, the cost of this scheme is 445. It is assumed that in that specific scheme, the epidemic was contained. In 14 of the cases, the no of vaccinated nodes was between 42 and 89. In all the rest (36), it was below 5.

The second network is based in the creation of triangles [19]. The newly introduced node, chooses at random among the preexisting edges of the network and it then connects to both ends of this edge. The control parameter q decides if the old edge between the two preexisting end-nodes, stays or not. This model operates in such a way as to create network with embedded small-world property. In the network examined in this paper, the control parameter q has the value of 0.5, which means that approximately half of the edges in question, are erased. This network is also undirected and the number of edges are 735.

For this network, the results are more condensed.As it shown in Figure 2, the no of vaccinated nodes is wrapped around 30 (varies between 7 and 51). This means that the algorithm manages to locate a fairly good vaccination scheme almost every time.

The third and final network that is used in this work is an Erdos-Renyi random network [20]. For the network at hand, the probabillity to establish a link is $p_L = 0.02$ which returns a network of 2530 edges.

For the ER network, it is clear in Figure 3 that the vaccination schemes keep the number of vaccinated nodes low. Half of these schemes returned as best option, 0 vaccinated nodes. Also, the maximum value of vaccinated nodes is 3, which is found only in four occasions. Because
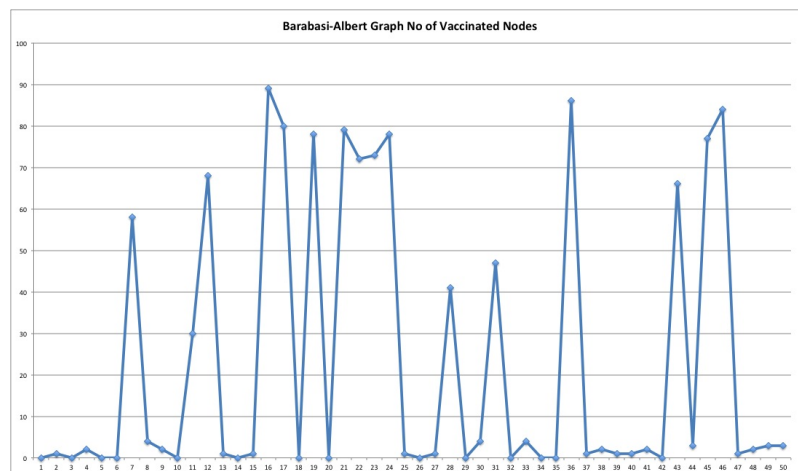
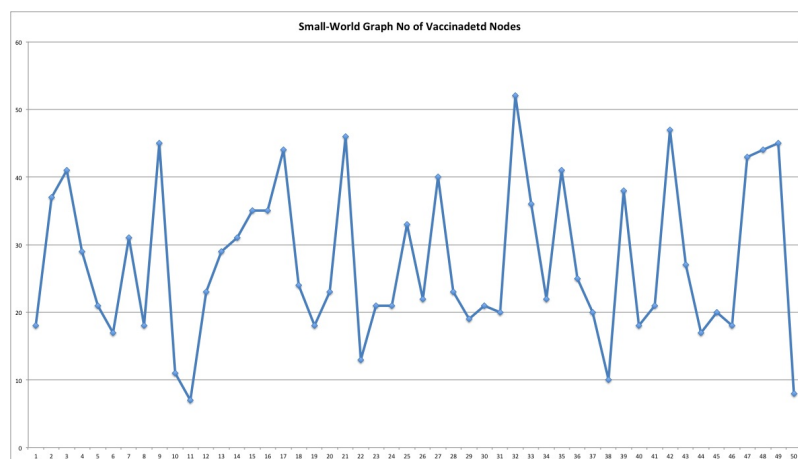**Figure 1.** The number of vaccinated nodes for the 50 runs of the algorithm, in the BA network.

Figure 2. The number of vaccinated nodes for the 50 runs of the algorithm, in the small-world network.

of the significantly greater number of edges than the other networks, the epidemic spreading manages to find a way through the network and this just increases the cost for any vaccination scheme.

4. Conclusions and Future Work

The work that has been analyzed in this paper, is an effort to combine an evolutionary process with a dynamical process in complex networks, in order to obtain an optimal scheme for various types of networks. The results where promising for the structured networks, whereas the behavior of the algorithm in the different types of networks shows that the process converges to desirable values. For the future, the algorithm must be tested more extensively and the nodes that were chosen in each network, must be analyzed further with respect to their role in it.
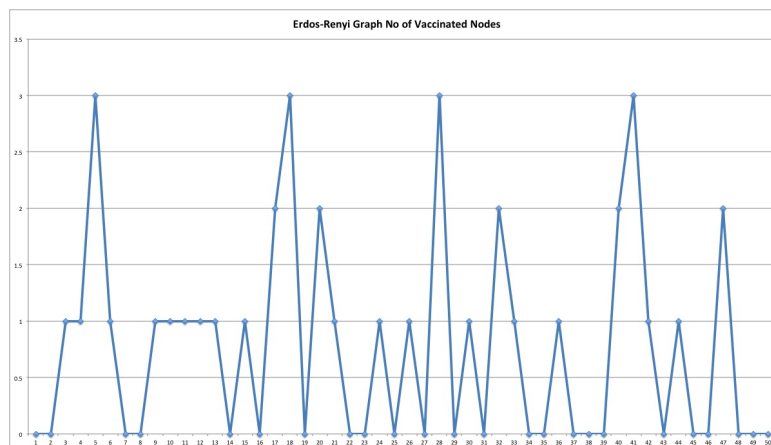
Figure 3. The number of vaccinated nodes for the 50 runs of the algorithm, in the ER random network.

[1] S. H. Strogatz, "Exploring complex networks", Nature, vol. 410, pp. 268-276, March 2001.

[2] R. Albert and A. - L. Barabasi, "Statistical Mechanics of Complex Networks", Review of Modern Physics, vol. 74 pp. 47-91, January 2002.

[3] X. F. Wang, "Complex networks: topology, dynamics and synchronization", Int. J. Bifurcation & Chaos, vol.12 no. 5 pp. 885-916, May 2002.

[4] R. Albert H. Jeong, and A.-L. Barabasi, Nature 401, 130 (1999).

[5] Jing-Dong J. Han, Nicolas Bertin, Tong Hao, et al, "Evidence for dynamically organized modularity in the yeast proteinprotein interaction network" Nature 430, 88-93 (1 July 2004).

[6] L. A. N. Amaral, A. Scala, M. Barthlmy, and H. E. Stanley, "Classes of small-world networks", PNAS vol. 97 no. 21, 1114911152.

[7] Albert-Laszlo Barabasi, Reka Albert, "Emergence of Scaling in Random Networks", Science 15 October 1999, Vol. 286 no. 5439 pp. 509-512 .

[8] Romualdo Pastor-Satorras and Alessandro Vespignani, "Epidemic Spreading in Scale-Free Networks", Phys. Rev. Lett. 86, 32003203 (2001).

[9] Marian Boguna and Romualdo Pastor-Satorras, "Epidemic spreading in correlated complex networks", Phys. Rev. E 66, 047104 (2002).

[10] Y. Moreno, R. Pastor-Satorras, A. Vespignani, "Epidemic outbreaks in complex heterogeneous networks", The European Physical Journal B - Condensed Matter and Complex Systems, April 2002, Volume 26, Issue 4, pp 521-529

[11] Romualdo Pastor-Satorras and Alessandro Vespignani, "Immunization of complex networks", Phys. Rev. E 65, 036104 (2002).

[12] N. Madar, T. Kalisky, R. Cohen, D. ben-Avraham, S. Havlin, "Immunization and epidemic dynamics in complex networks", The European Physical Journal B - Condensed Matter and Complex Systems, March 2004, Volume 38, Issue 2, pp 269-276.

[13] P. Holme, "Efficient local strategies for vaccination and network attack", EPL (Europhysics Letters) Volume 68 Number 6 P. Holme 2004 Europhys. Lett. 68 908.

[14] J. Gomez-Gardenes, P. Echenique, Y. Moreno, "Immunization of real complex communication networks", The European Physical Journal B - Condensed Matter and Complex Systems, January 2006, Volume 49, Issue 2, pp 259-264.

[15] Melanie Mitchell, "An Introduction to Genetic Algorithms", 06 February 1998.

[16] Davis, L., "Genetic algorithms and simulated annealing", Research notes in artificial intelligence (ISSN 0268-7526), 1987.

[17] D S Vlachos and K J Parousis-Orthodoxou, "Evolutionary Processes in Complex Networks and Small Worlds", Journal of Physics: Conference Series Volume 410 conference 1

[18] Anca Gog, D. Dumitrescu, Bat Hirsbrunner, "Community Detection in Complex Networks Using Collaborative Evolutionary Algorithms", Advances in Artificial Life, 2007

[19] Zhongzhi Zhang, Shuigeng Zhou, Zhen Shen, From regular to growing small-world networks, Physica A 385 (2007) 765-772.

[20] Erdos, P.; Renyi, A. (1959). "On Random Graphs. I". Publicationes Mathematicae 6: 290297.