

Experimental Environment to Model, Simulate and Analyze Contagious Diseases as a Diffusion Process in Social Networks

Rafal Kasprzyk (✉ rafal.kasprzyk@wat.edu.pl)

Military University of Technology in Warsaw <https://orcid.org/0000-0002-4938-1874>

Andrzej Najgebauer

Military University of Technology in Warsaw

Research

Keywords: diffusion, complex networks, social networks, contagious diseases, simulation

Posted Date: September 27th, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-923987/v1>

License:  This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Experimental environment to model, simulate and analyze contagious diseases as a diffusion process in social networks

Rafał Kasprzyk^{1(*)}[0000-0002-4938-1874], Andrzej Najgebauer¹ [0000-0002-4268-8697]

¹ Military University of Technology, Faculty of Cybernetics, Gen. S. Kaliskiego Str.2, 00-908 Warsaw, Poland
rafal.kasprzyk@wat.edu.pl, andrzej.najgebauer@wat.edu.pl

Abstract. In this paper the novel model of diffusion on networks and the experimental environment are presented. We consider the utilization of the graph and network theory in the field of modelling and simulating the dynamics of contagious diseases. We describe basic principles and methods and show how we can use them to fight against the spread of this phenomenon. We also present our software solution – CARE (Creative Application to Remedy Epidemics) that can be used to support decision-making activities.

Keywords: diffusion, complex networks, social networks, contagious diseases, simulation.

1 Introduction

Last years have seen a huge interest in network systems. The number of interdisciplinary researches undertaken in this field is affected by the strategic importance of network systems, particularly from the perspective of crisis management. In fact, networks, which are understood as sets of vertices and edges for representing different kinds of relationships between vertices, are ubiquitous. There are many examples of systems modelled with the use of networks, including the Internet, WWW, transport networks, transmission networks, electrical grids, and finally social networks. These systems due to the number of included nodes and edges are called complex networks or large-scale networks (Kasprzyk 2012a).

One of the most important research areas related to complex networks is the issue of diffusion processes on networks. Diffusion is a process by which information, viruses, gossips and any other phenomena spread over networks (Godin 2001; Bartosiak et al. 2013; Kasprzyk 2012a; Kasprzyk 2012b; Lescovec et al. 2007; Lloyd et al. 2001; López-Pintado 2008; Pastor-Satorras and Vespignani 2001) in particular social networks. The standard approach is a simplified assumption that phenomena (information, viruses, gossips) spread in the environment which is modelled using very simple construction of *Regular Graph* like GRID-based graph or similar, very rarely *Random Graphs*.

Standard approaches do not explain the real dynamic of diffusion in real-world networks, in particular:

- Why even slightly virulent phenomena (e.g. contagious diseases) can spread over a network for a long time;
- How to choose nodes to maximize or minimize diffusion range (e.g. how to choose individuals to vaccinate in order to minimize the epidemic's range);
- What is the mechanism of arising secondary phenomena centers (e.g. outbreaks of diseases).

(*) Corresponding author: rafal.kasprzyk@wat.edu.pl

The drawbacks of the standards diffusion models is that they do not take into account underlying real-world networks topology. Who (or what) is connected to whom (what), seems to be a fundamental question. Apparently, networks derived from data on real life cases (most often: networks growing spontaneously), are neither regular graphs nor random ones. As it turned out real networks, which have been intensively studied recently, have some interesting features. These features, whose origins are nowadays discovered, modelled (Barabási and Albert 1999; Barabási and Albert 2000; Erdős and Rényi 1959a; Erdős and Rényi 1959b; Newman 2000; Watts and Strogatz 1998) and examined (Barabási and Albert 2002; Kasprzyk 2012a; Newman 2003; Strogatz 2001; Wang and Chen 2003) significantly affect dynamics of the diffusion processes within real-world networks. Three very interesting models of real-world networks which have been introduced recently e.g. *Random Graph*, *Small World Networks* and *Scale Free Networks*, will be described later in this paper.

We have to also remember that all kinds of phenomena spreading over the networks have their unique properties, and we should be able to model them. The notion of a state machine seems to be useful in this modelling situation. Using probabilistic finite-state machines (Sokolova and de Vink 2004; Vidal et. al. 2005) we can model spreading of vast variety of phenomena. For example, we are able to build models of diseases with any states (e.g. *susceptible*, *infected*, *carrier*, *immunized*, *dead*, etc.), and probabilities of transitions from a state to another resulting from social interactions (contacts). Again, the underlying contacts (social network topology) seems to have a huge impact on the dynamic of diffusion processes, what has been already mentioned.

2 Basic definitions and notations

Let's define network as follows (Kasprzyk 2012a; Korzan 1978):

$$Net(t) = \left\langle G(t) = \langle V(t), E(t) \rangle, \{f_i(v, t)\}_{\substack{i \in \{1, \dots, NF\} \\ v \in V(t)}}, \{h_j(e, t)\}_{\substack{j \in \{1, \dots, NH\} \\ e \in E(t)}} \right\rangle \quad (1)$$

where:

$G(t) = \langle V(t), E(t) \rangle$ - a simple dynamic graph

Comment: The dynamic means that $V(t)$ and $E(t)$ can change over time (Harary and Grupa 1997);

$V(t)$ - sets of graph's vertices;

$E(t)$ - sets of graph's edges, $E(t) \subset \{\{v, v'\} : v, v' \in V(t)\}$;

$f_i : V(t) \rightarrow fVal_i$ - the i -th function describe on the graph's vertices, $i = 1, \dots, NF$, (NF - number of vertex's functions), $fVal_i$ - is a set of f_i values;

$h_j : E(t) \rightarrow hVal_j$ - the j -th function describe on the graph's edges, $j = 1, \dots, NH$, (NH - number of edge's functions), $hVal_j$ - is a set of h_j values.

We assume that values of function's $f_i(\cdot)$ and $h_j(\cdot)$ can also change over time.

In this paper we were particularly interested in relationship between the structure of real-world networks and the dynamic of any phenomena on them. Due to this fact, we focused on the characteristics of the graph $G(t)$, while functions on the graph's vertices (nodes) and edges (links) were omitted.

Simple dynamic graphs are very often represented by a matrix $A(t)$ called adjacency matrix, which is a $V(t) \times V(t)$ symmetric matrix. The element $a_{ij}(t)$ of adjacency matrix equals 1 if there is an edge between vertices v_i and v_j , and 0 otherwise.

The first-neighborhood of a vertex v_i denote as $\Gamma_i^1(t)$ is defined as set of vertices immediately connected with v_i , i.e.:

$$\Gamma_i^1(t) = \{v_j \in V(t) : \{v_i, v_j\} \in E(t)\} \quad (2)$$

The degree $k_i(t)$ of a vertex v_i is the number of vertices in the first-neighbourhood of a vertex v_i , i.e.:

$$k_i(t) = |\Gamma_i^1(t)| \quad (3)$$

The path starting in vertex v_i and ending in vertex v_j is a sequence of $\langle v_0, v_1, \dots, v_{k-1}, v_k \rangle$ where $\{v_{i-1}, v_i\} \in E(t) \quad \forall i = 1, \dots, k$. The length of a path is defined as the number of links in it. The shortest path length starting in vertex v_i and ending in vertex v_j is denoted as $d_{ij}(t)$.

Now we can define diameter D as the longest shortest path i.e.:

$$D(t) = \max_{v_i, v_j \in V(t)} \{d_{ij}(t)\} \quad (4)$$

Let's denote the number of existing edges between the first-neighbourhood of a vertex v_i as $N_i(t)$, i.e.:

$$N_i(t) = |\{v_l, v_k\} : v_l, v_k \in \Gamma_i^1(t) \wedge \{v_l, v_k\} \in E(t)| \quad (5)$$

Now, we can define a very important concept, so called the local clustering coefficient C_i for a vertex v_i , which is then be given by the proportion of $N_i(t)$ divided by the number of edges that could possible exist between first-neighbourhood of a vertex v_i (every neighbour of v_i is connected to every other neighbour of v_i). Formally:

$$C_i(t) = \begin{cases} \frac{2N_i(t)}{k_i(t) \cdot (k_i(t) - 1)}, & |\Gamma_i^1(t)| > 1 \\ 0 & , \quad |\Gamma_i^1(t)| \leq 1 \end{cases} \quad (6)$$

The clustering coefficient C for the whole network is define as the average of C_i overall $v_i \in V$ i.e.:

$$C(t) = \frac{1}{|V(t)|} \sum_{v_i \in V(t)} C_i(t) \quad (7)$$

The degree distribution $P(k, t)$ of a network is defined as the fraction of nodes in the network with degree k .

Formally:

$$P(k, t) = \frac{|V_k(t)|}{|V(t)|} \quad (8)$$

where: $|V_k(t)|$ is the number of nodes with degree k ; $|V(t)|$ is the total number of nodes.

3 Models of real-world networks

From the 50s of the last century, complex networks were modelled using *Random Graphs*, with no organizational principle. Databases introduction together with the need to collect information about genuine network structures revealed in the late twentieth century that these networks have a number of specific features, which have not been known so far. It turned out that although we have been surrounded by exhaustively examined networks, their topology and principles of evolution may still be enigmatic. Analyses carried out on the actual networks proved the existence of their specific characteristics. In particular, these included a relatively small number of edges (sparse graph), a relatively short diameter of the graph (shortest longest path) and a surprisingly short average path length, while high clustering coefficient value is given. Another extremely interesting feature of most genuine networks, is a power law distribution of node's degree, i.e. $P(k, t) \sim k^{-\gamma}$, where γ is a constant typically between 2 and 3.

These features whose origin are nowadays discovered indeed affect the diffusion processes within networks. Understanding the balance of order and chaos in networks is one of the goals of the current research on so called complex networks. Identifying and measuring properties of a real-world networks is a first step towards understanding their topology. The next step is to develop a mathematical model, which typically takes a form of an algorithm for generating networks with the same statistical properties. The above-mentioned features have contributed to creation of a wide range of models which describe genuine networks in a more adequate way than *Random Graphs*. In the 90s, first types of such network models were created i.e. *Small World Networks* and *Scale Free Networks*, and their further development is still underway.

For a long time real networks without visible or known rule of organization were described using Erdős and Rényi model of *Random Graphs* (Erdős and Rényi 1959a; Erdős and Rényi 1959b). Assuming equal probability and independent random connections made between any pair of vertices in initially not connected graph, they proposed a model suffering rather unrealistic topology. Their model has now only a limited usage for modelling real-world network.

Not long ago Watts and Strogatz proposed *Small World* model (Watts and Strogatz 1998) of real-world networks as a result of simple observation that real networks have topology somewhere between regular and random one. They began with regular graph, such as a ring, and then „rewire” some of the edges to introduce randomness. If all edges are rewired a *Random Graph* appears. The idea of this method was depicted in Figure 1.

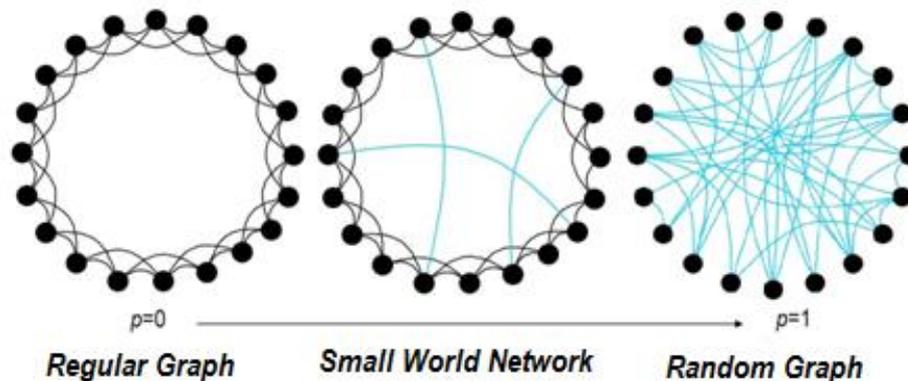


Fig. 1. The idea of *Small World* network model.

The process of rewiring affects not only the average path length but also clustering coefficient. Both of them decrease as probability of rewiring increases. The interesting property of this procedure is that for a wide range of rewiring probabilities the average path length is already low while clustering coefficient remains high. This correlation is typical for real-world networks.

Barabási and Albert introduced yet another model (Barabási and Albert 1999) of real-world networks as a result of two main assumptions: constant growth and preferential attachment. They showed why the distribution of nodes degree is describe by a power law. The process of network generation is quite simple. The network grows gradually, and when a new node is added, it creates links (edges) to the existing nodes with probability proportional to their connectivity. In consequence nodes with very high degree appears (so called *hubs* or *super-spreaders*), which are very important for communication in networks.

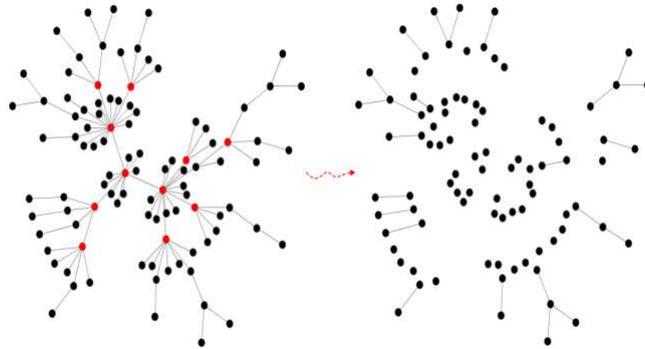


Fig. 2. The role of *hubs* in *Scale Free* network.

There are many modification of this basic procedure for generating networks. Now it is considered that *Scale Free* models of real-world networks are the best ones.

4 Measures of nodes importance

In the Figure 3, there is an example of real social network. Nodes represent individuals and link social interactions.

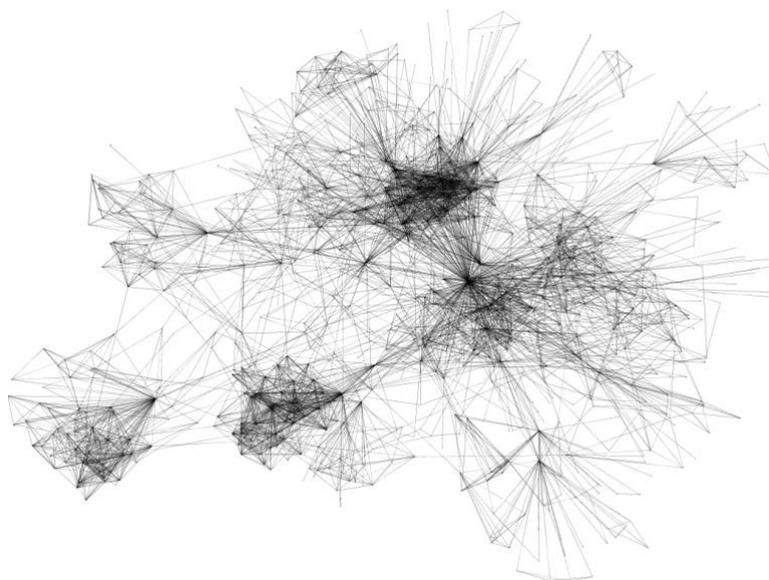


Fig. 3. An example of a real social network.

The most basic and frequently ask question is how to identify the most important nodes. The answer can help maximize or on the other hand minimize diffusion dynamic of any phenomena within networks. We decided to use so called centrality measures to assess nodes importance. No single measure of center is suited for application. Sever noteworthy measures are *degree centrality*, *radius centrality*, *closeness centrality*, *betweenness centrality*, *eigenvector centrality*. Thanks to these measures we can show for example how to disintegrate the network with minimum number of steps and in consequence minimize diffusion area, in particular how to optimize vaccination strategies (Kasprzyk 2009; Kasprzyk 2012a).

- ***degree centrality***

Gives the highest score of influence to the vertex with the largest number of first-neighbours. The degree centrality is traditionally defined analogous to the degree of a vertex, normalized over the maximum number of neighbours this vertex could have:

$$dc_i(t) = \frac{k_i(t)}{|V(t)|-1} \quad (9)$$

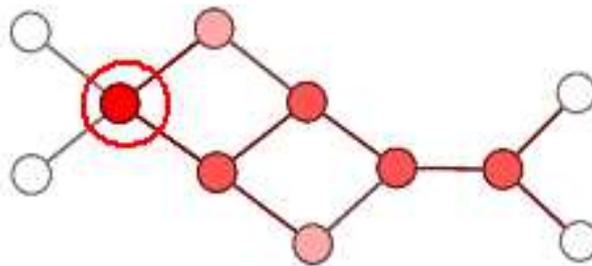


Fig. 4. Importance of nodes according *degree centrality*.

- ***radius centrality***

Chooses the vertex with the smallest value of the longest shortest path starting in each vertex. So if we need to find the most influential node for the most remote nodes it is quite natural and easy to use this measure:

$$rc_i(t) = \frac{1}{\max_{v_j \in V(t)} d_{ij}(t)} \quad (10)$$

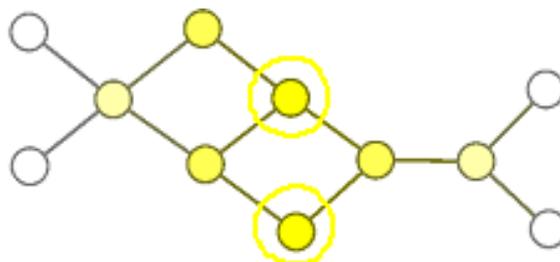


Fig. 5. Importance of nodes according *radius centrality*.

- ***closeness centrality***

Focuses on the idea of communications between different vertices and the vertex, which is "closer" to all vertices gets the highest score:

$$cc_i(t) = \frac{|V(t)| - 1}{\sum_{v_j \in V(t)} d_{ij}(t)} \quad (11)$$

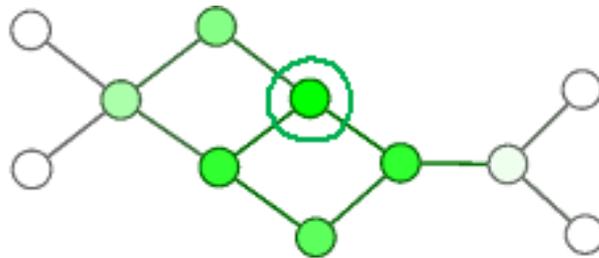


Fig. 6. Importance of nodes according *closeness centrality*.

- ***betweenness centrality***

It can be defined as the percent of shortest paths connecting any two vertices that pass through the considered vertex. If $p_{l,i,k}(t)$ is the set of all shortest paths between vertices v_l and v_k passing through vertex v_i and $p_{l,k}(t)$ is the set of all shortest paths between vertices v_l and v_k then:

$$bc_i(t) = \frac{\sum_{v_l \in V_G(t)} \sum_{v_k \in V_G(t), v_k \neq v_l} \frac{p_{l,i,k}(t)}{p_{l,k}(t)}}{(|V_G(t)| - 2)(|V_G(t)| - 1)} \quad (12)$$

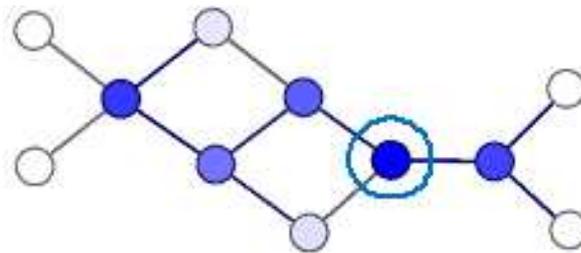


Fig. 7. Importance of nodes according *betweenness centrality*.

- ***eigenvector centrality***

While degree centrality gives a simple count of the number of connection a vertex has, eigenvector centrality acknowledges that not all connections are equal. If we denote the centrality of vertex v_i by $ec_i(t)$ then we can allow for this effect by making $ec_i(t)$ proportional to the centralities of the v_i 's first-neighbours.

$$ec_i(t) = \frac{1}{\lambda} \sum_{j=1}^{|V(t)|} a_{ij}(t) \cdot ec_j(t) \quad (13)$$

Using matrix notation, we have as follows:

$$\vec{ec}(t) = \frac{1}{\lambda} A(t) \cdot \vec{ec}(t) \quad (14)$$

So, we have $A(t) \cdot \vec{ec}(t) - \lambda I \cdot \vec{ec}(t) = 0$ and the λ value we can calculate using $\det(A(t) - \lambda I) = 0$.

Hence, $\vec{ec}(t)$ is an eigenvector of adjacency matrix with the largest value of eigenvalue λ .

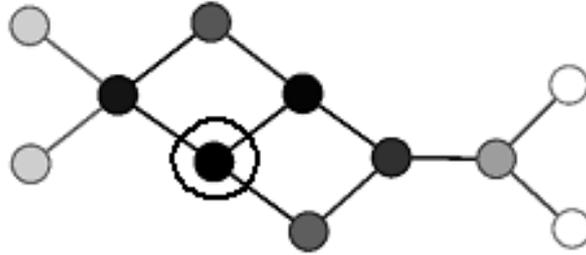


Fig. 8. Importance of nodes according *eigenvector centrality*.

5 Global Connection Efficiency Coefficient

Using centrality measures it is possible to optimize vaccination strategies. At first sight, it seems that the most central nodes should be considered as potential individuals to be vaccinated. However there is an issue connected with the fact that real-life networks' structures are not fully known.

To evaluate how well a G network is connected before and after the removal of a set of nodes we use the *global connection efficiency* (GCE) (Crucitti et. al. 2004). We assume that the *connection efficiency* between vertex v_i and v_j is inversely proportional to the shortest distance:

$$ce_{ij}(t) = \frac{1}{d_{ij}(t)} \quad (15)$$

The global connection efficiency is defined as the average connection efficiency over all pairs of nodes:

$$GCE(t) = \frac{1}{|V_G(t)|(|V_G(t)|-1)} \sum_{v_i \in V_G(t)} \sum_{v_j \in V_G(t), v_j \neq v_i} \frac{1}{d_{ij}(t)} \quad (16)$$

Unlike the average path length, the global connection efficiency is a well-defined quantity as well as in the case of non-connected graphs.

Let $G^-(t, x, rs)$ describe graph G after the removal of $x \in \square$ nodes using $rs \in RS$, where $rs \in RS = \{rn, rrn, dc, rc, cc, bc, ec\}$. Elements of RS describe the removal strategies. The simplest strategy is rn (Random Nodes), which means that random nodes are removed. A simple modification of the random strategy is rrn (Random-Random Nodes), which means that the removal strategy is a two-step one. Firstly, nodes are chosen randomly, and secondly, among all first-neighbours of these nodes random nodes are chosen again. This strategy is often called *Vaccinate Thy Neighbor* (Cohen et al. 2003; Kasprzyk 2009; Madar 2004). All other strategies are based on centrality measures. Using these strategies the nodes with the greatest value of the following measures are removed: dc – *degree centrality*, rc – *radius centrality*, cc – *closeness centrality*, bc – *betweenness centrality*, ec – *eigenvector centrality*.

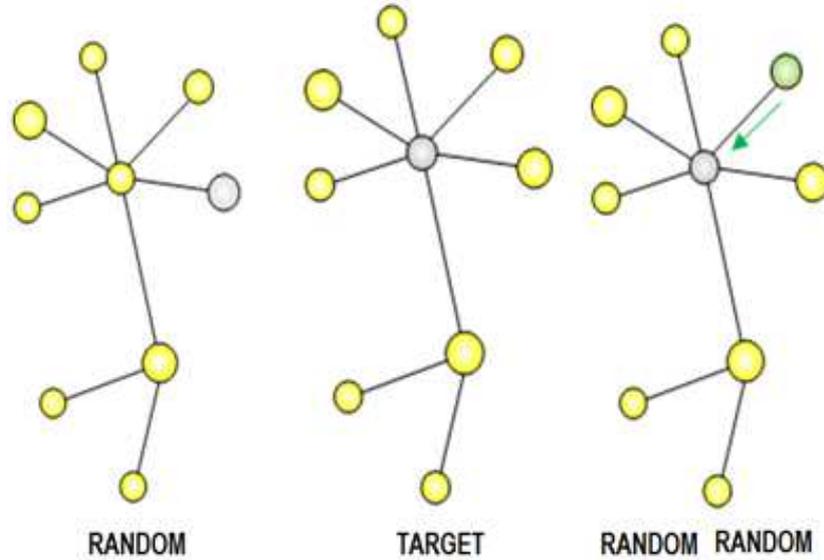


Fig. 9. Projection of removal strategies.

The network durability measure is *global connection efficiency coefficient* defined by the function (Tarapata and Kasprzyk 2010; Kasprzyk 2012a):

$$GCE_{coef}(G(t), x, rs) = \frac{GCE(G^-(t, x, rs))}{GCE(G(t))} \quad (17)$$

The lower the value of the function the higher the effectiveness of the removal/vaccination strategy for a particular graph G .

6 A novel model of diffusion

All in all, who is connected to whom seems to be crucial for diffusion in networks, but all kind of phenomena have their unique properties. In consequence, we defined model of diffusion in network as a vector, with three element (Kasprzyk 2012a; Kasprzyk 2012b):

$$Diff(t) = \langle Net(t), PM_{n=1,2,\dots,N}, Gen(v,t) \rangle \quad (18)$$

where:

$Net(t)$ - network model of system constitutes diffusion environment;

PM_n - probabilistic finite-state machine describing *Phenomenon Model* (information, virus, gossip and so on);

$Gen : V(t) \rightarrow SIG =$ - specific function for simulation needs (*Generator of SIGNALS*), which assigns for each vertex in each simulation step a set of signals as a result of edges status (activated, non-activated) and foremost vertices' first-neighborhood states. These signals are received and processed by PM_n on each vertex.

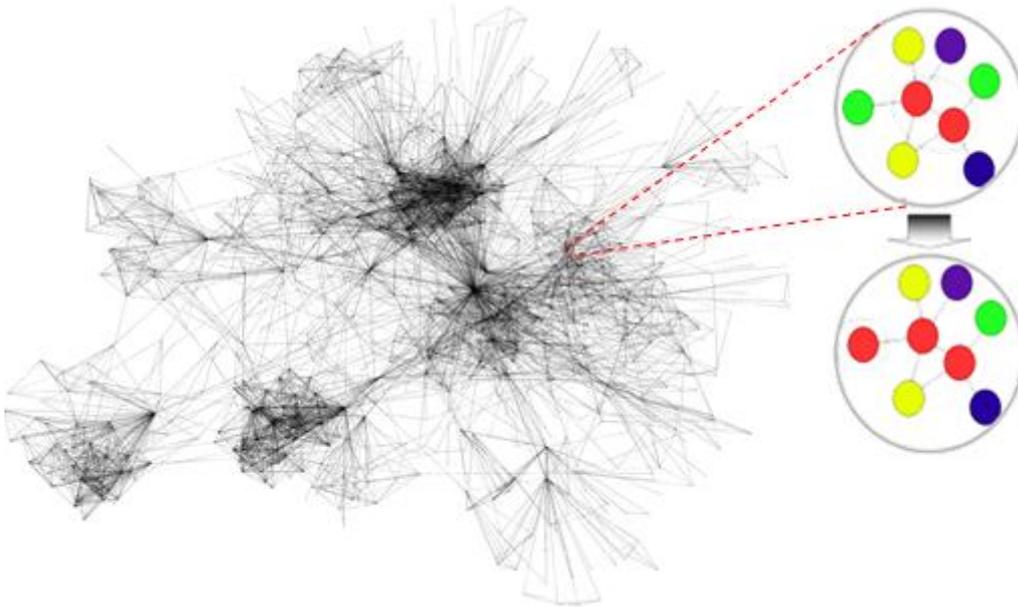


Fig. 10. Projection of the proposed model of diffusion.

Thus both concepts i.e. real-world networks topology and probabilistic state machine models are highly pertaining to the presented idea subject and objectives. The aim is to uncover the diffusion mechanisms hidden in the structure of networks.

7 Experimental environment

The program platform for the development experimental environment is *Framework Gephi* (Bastian 2009, Gephi 2020). *Gephi* is an interactive platform for visualization and exploration of graphs and networks, with modular structure, implemented in the *Model-View-Controller* architecture, by using the *Inversion of Control* pattern. *Gephi* evolves by adding new *plugins* to the existing environment. It is worth noticing that the plugins to the *Gephi* implementation, somehow forced by its designers, are in compliance with the best practices of the object programming, which often comes down to the principle of SOLID (*Single responsibility, Open-closed, Liskov substitution, Interface segregation, Dependency inversion*). Another interesting aspect is also the differentiation between *Application Programming Interface* (API) of the same *Gephi's Framework* and API offered by way of adding of the *Service Provider Interface* (SPI) thereto. *Gephi's API* is created by the platform designers (or under their supervision) and, as a matter of principle, rarely changed. On the other hand, SPI is a set of interfaces or services implemented in the form of special plugins, thus, the designers of the *Gephi's Framework* are not responsible for their proper functioning. Such approach is a tribute to the contemporary needs in terms of the necessity to quickly create software based on the existing components. However, on the other hand, the adopted solution ensures high quality of the software, at the same time guaranteed that its plugins may be used by already numerous *Gephi* users.

The experimental environment was created as a set of original *Gephi* plugins, and its functionality was presented by the *Use Cases*.

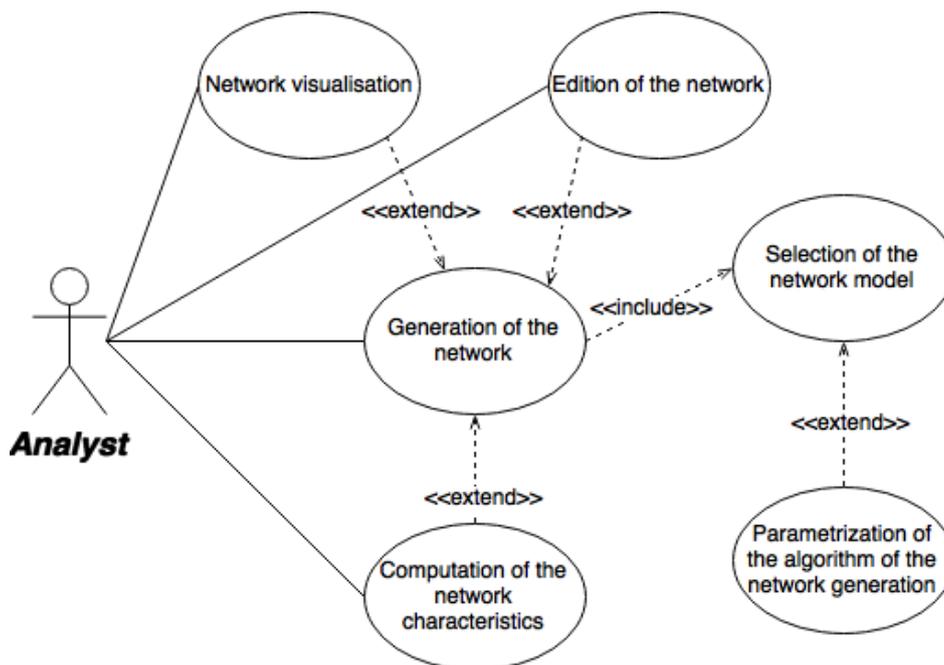


Fig. 11. A use case diagram: “*Generation of synthetic networks (social network models)*”.

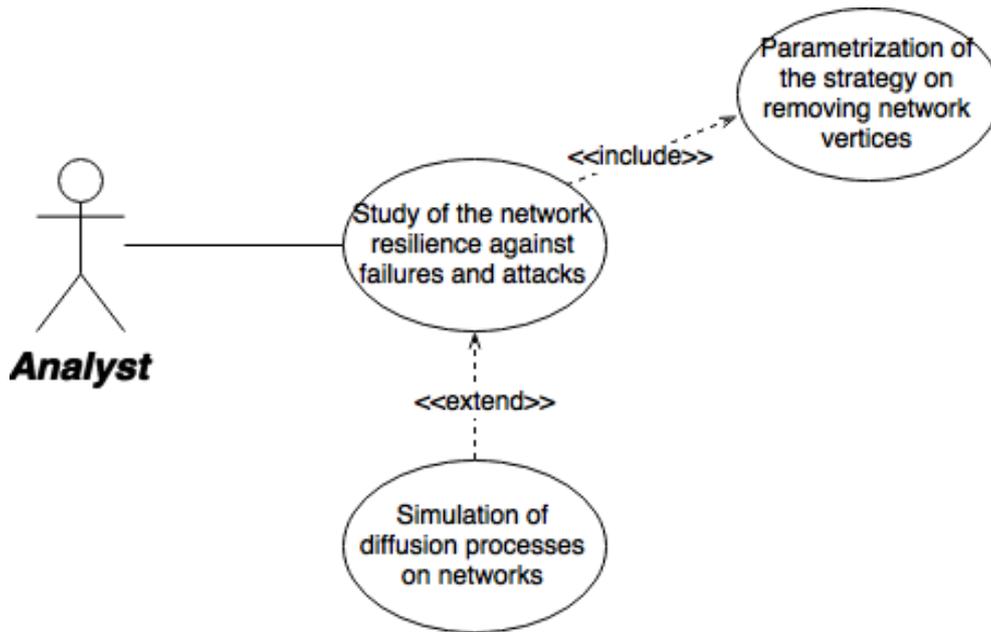


Fig. 12. A use case diagram: “Analyzing effectiveness of removal (vaccination/isolation) strategies”.

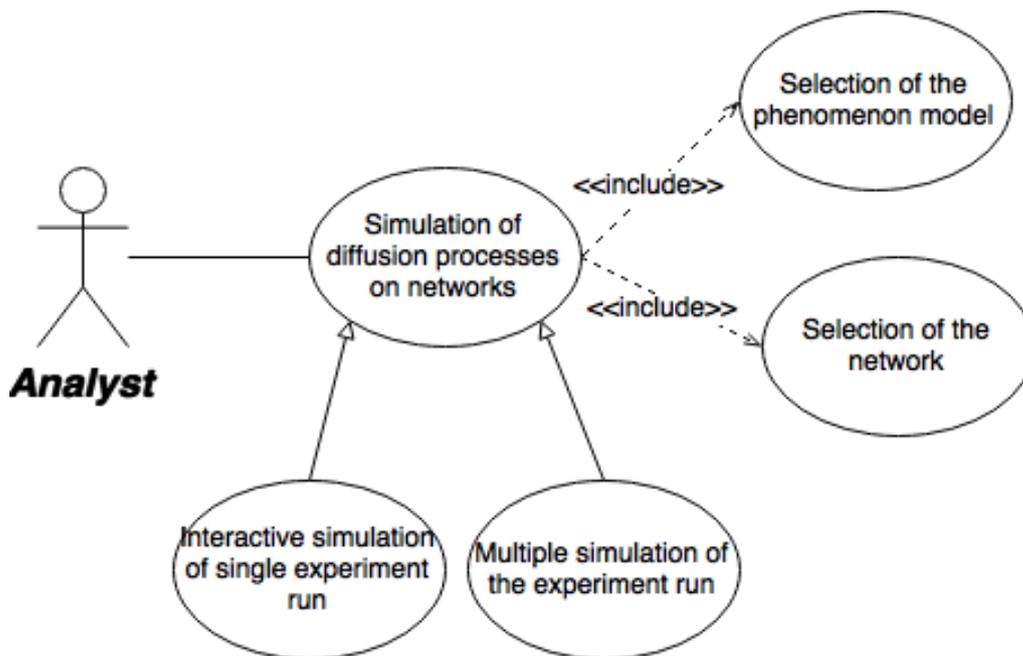


Fig. 13. A use case diagram: “Simulation of contagious disease diffusion on networks”.

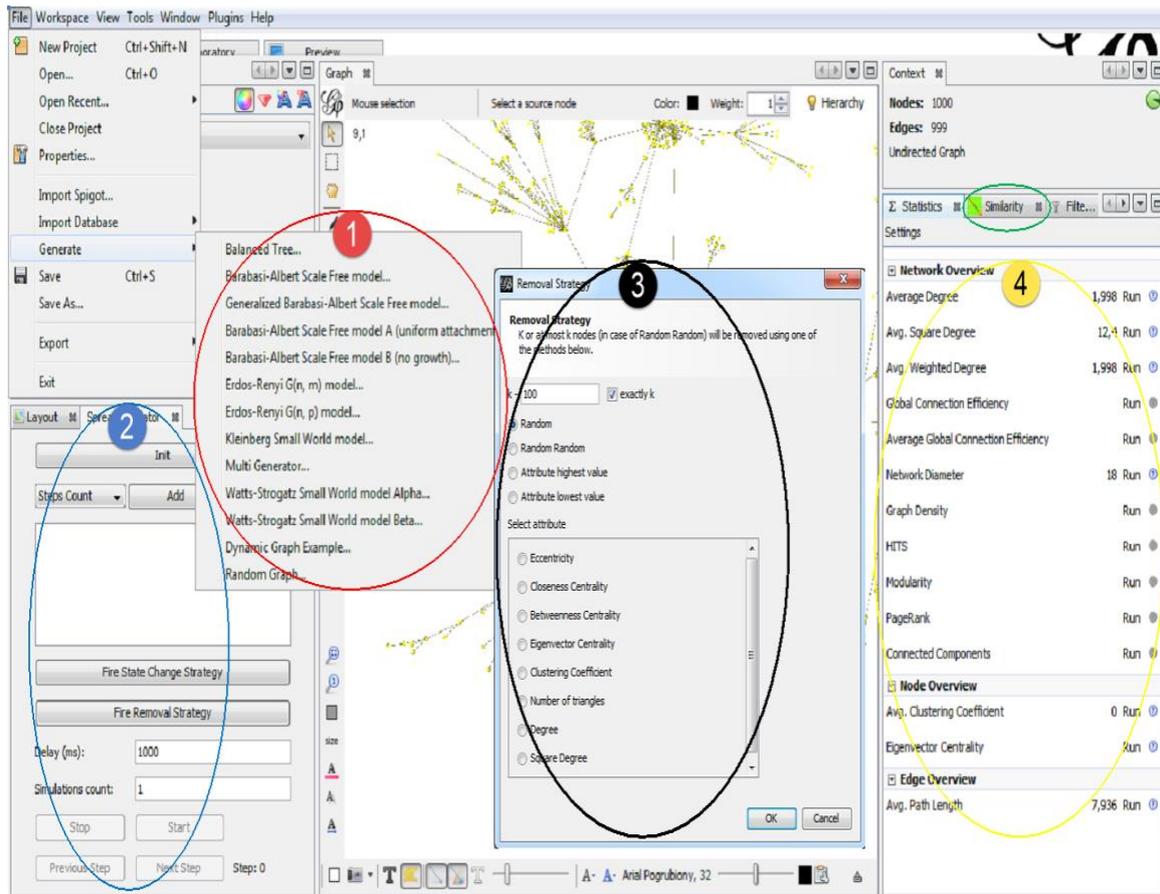


Fig. 14. Graphical User Interface of the experimental environment.

We added to *Gephi* new functionalities like, in particular: complex networks generators, scenarios for centrality measures utilization and finally the ability to simulate diffusion of any phenomena in any networks.

Figure 14 shows the main window of the experimental environment, with the highlighted interface elements, which correspond in terms of their functionalities to the presented use cases. The submenu with a list of implemented synthetic networks generators was marked in red (area no. 1). The tab for parametrization of diffusion processes simulation was marked in blue (area no. 2). Finally, the window allowing to analyze effectiveness of removal (vaccination/isolation) strategies was marked in black (area no. 3). Attention should be also paid to the tab presenting graph statistics (yellow - area no. 4) available on the *Gephi* platform. Such algorithms constitute an integral part of the platform and are successively added and upgraded by the community of *Gephi* programmers, including author of this study.

8 Simple case studies

8.1 Effectiveness of removal strategies in networks

When vaccine supplies for a deadly disease are limited, whom should health workers target? Many researches in *Social Networks* show that human follow a power law degree distributions, which was mentioned earlier. So we decide to analyze the effectiveness of different removal (vaccination/isolation) strategies for different *Scale Free* networks. We take into account three main strategies i.e. random, target and random-random.



Fig. 15. The designation of vaccination strategies.

As we can see on the Figure 16-18, the random strategy is very ineffective. So it is obvious that we should use the target strategy, which is based on centrality measures. In this particular case we take advantage of the betweenness centrality which seems to be most effective. The problem with the target strategy is the fact that it is necessary to know the exact topology of networks. Our knowledge about most real networks is incomplete and uncertain; that's why the target strategy is very often unusable. Then, as experiments prove, the random-random strategy could be utilized, which is much more effective than the random one (Bartosiak et al. 2013; Kasprzyk 2012a).

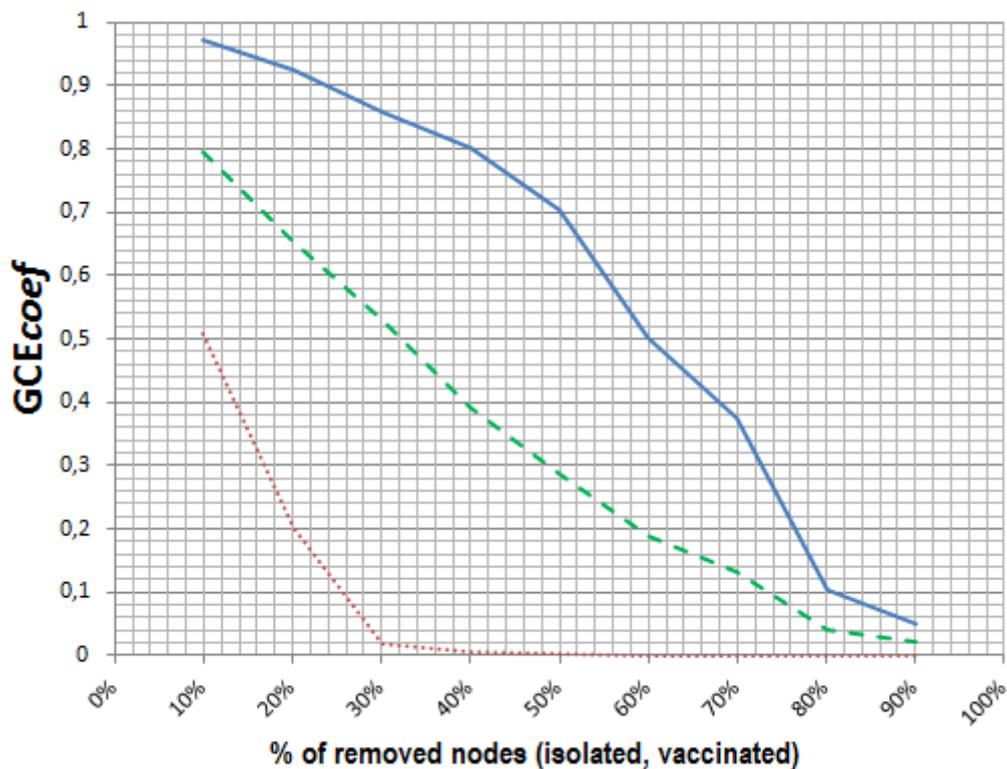


Fig. 16. The effectiveness of removal strategies (*Scale Free* network with 1000 nodes, $\gamma=3$, $\langle k \rangle \approx 6$).

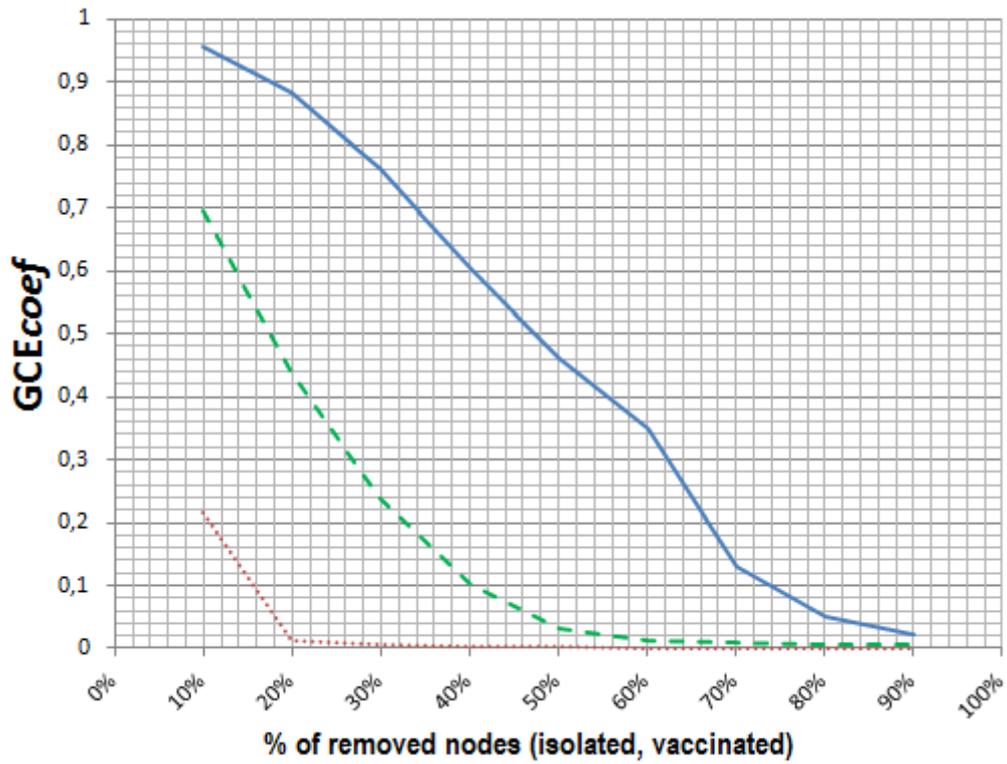


Fig. 17. The effectiveness of removal strategies (*Scale Free* network with 1000 nodes, $\gamma=3$, $\langle k \rangle \approx 4$).

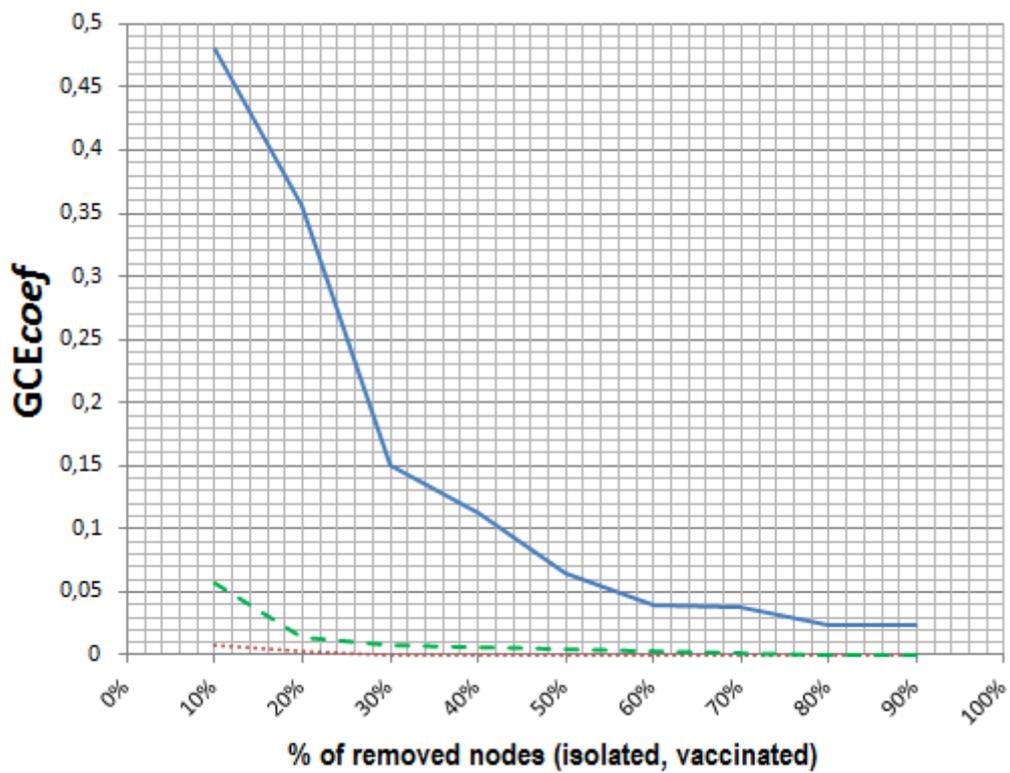


Fig. 18. The effectiveness of removal strategies (*Scale Free* network with 1000 nodes, $\gamma=2$, $\langle k \rangle \approx 2$).

8.2 SIS model of a disease on networks – research question no. 1

Let us now analyse a very simple case study of diffusion process from the field of epidemiology. One of the most extensively studied epidemic model is SIS (*Susceptible-Infected-Susceptible*). In each time step susceptible individuals are infected by each infected neighbours with probability β and the recovering rate of *infected* individuals to *susceptible* ones is α . Parameter λ is known in literature as *speed of spreading* or *virulence* of the disease and is define as:

$$\lambda = \beta / \alpha$$

Figure 19 representing BM_1 diagram of SIS model of a disease prepared in the experimental environment with

$$\lambda = 0.5 / 0.1 = 5.$$

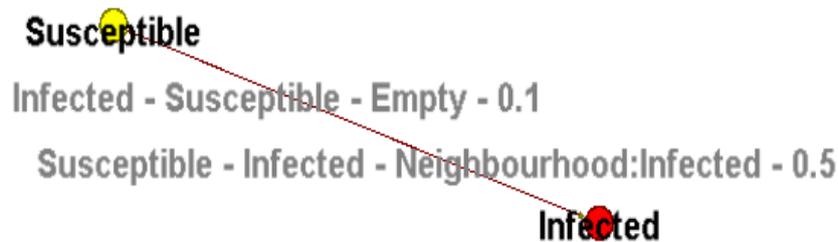


Fig. 19. SIS model of a disease.

We use two networks: *Scale Free* and *Random Graph*. The networks consist of 1 000 nodes and 2 000 edges. The average degrees of the nodes are similar and close to 4. At time 0 a small number of nodes (1%) are infected. The nodes are chosen using various ways:

- the lowest value of the degree centrality;
- the highest value of the degree centrality;
- randomly.

Then simulation of an epidemic is started. This experiment has been repeated 100 times. The dynamics of an infectious disease in different networks is presented in Figure 20 (Bartosiak et al 2013; Kasprzyk 2012a).

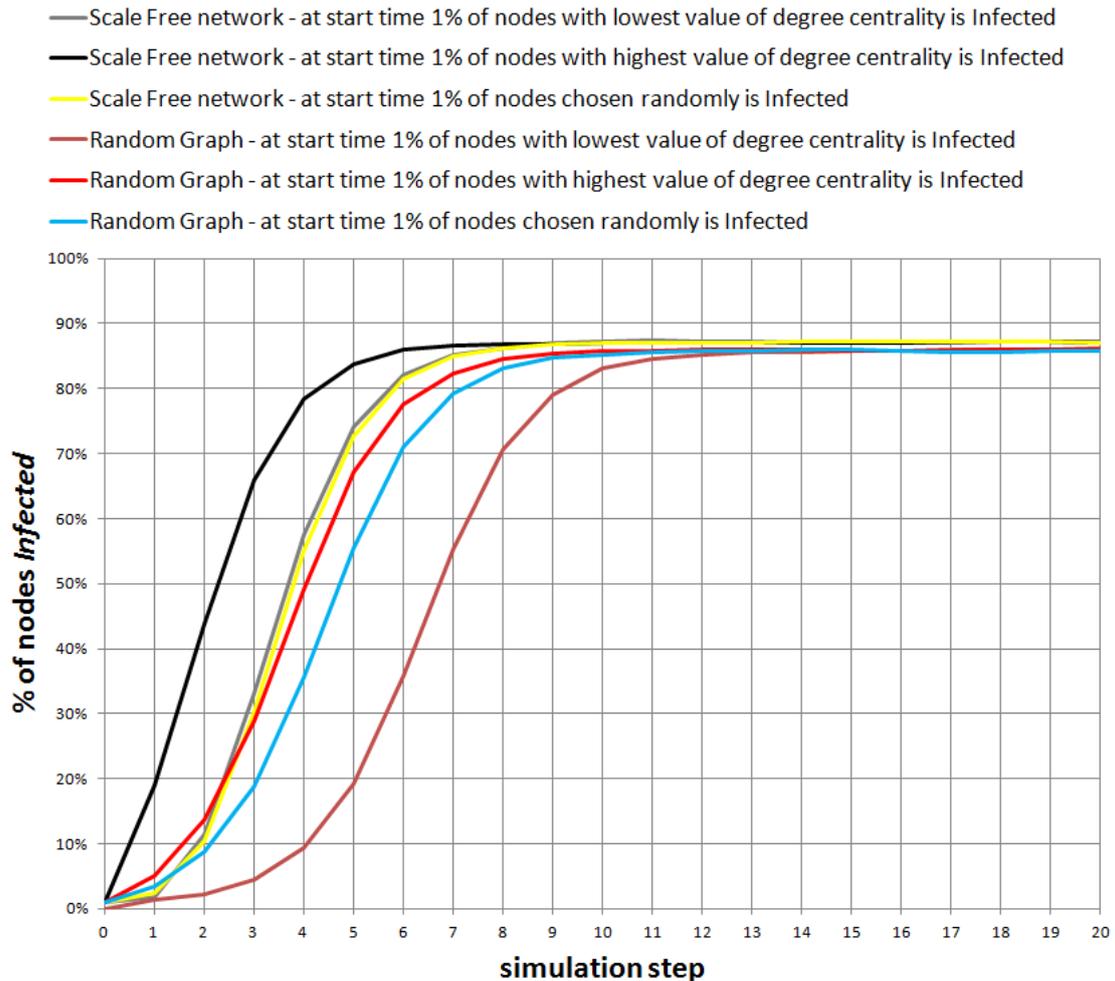


Fig. 20. SIS model of a disease with $\lambda = 5$ in *Scale Free* network and *Random Graph*.

We can see that the topology of the networks as well as the way in which the nodes are chosen to infect at the start time have a great impact on the dynamics of the infectious disease.

8.3 SIS model of a disease on networks – research question no. 2

The central question then becomes: how might network topology affect the disease diffusion process with different *virulence* (λ values). Our focus is still on the SIS model of a disease spreading in networks with different topology. We use three networks: *Scale Free*, *Random Graph* and *Regular Graph* that is exactly GRID-base one (very popular graph used in cellular automates). All networks consist of 10 000 nodes and about 20 000 edges. Average degree of nodes are similar and close to 4. At time 0 small number of nodes (1%) are chosen randomly and infected. Then simulation of the diffusion process is started. Each simulation were repeated 1 000 times. Dynamic of the disease diffusion process in different networks as a function of λ is presented in Figure 21-25 (Kasprzyk 2012a; Kasprzyk 2012b).

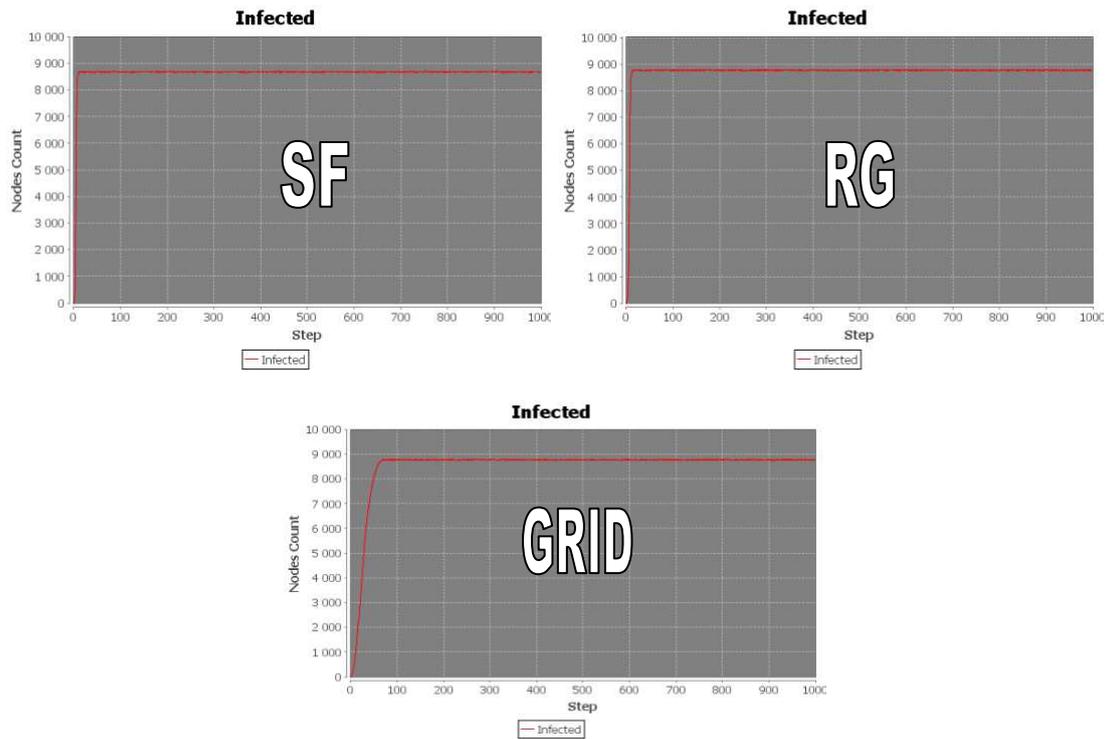
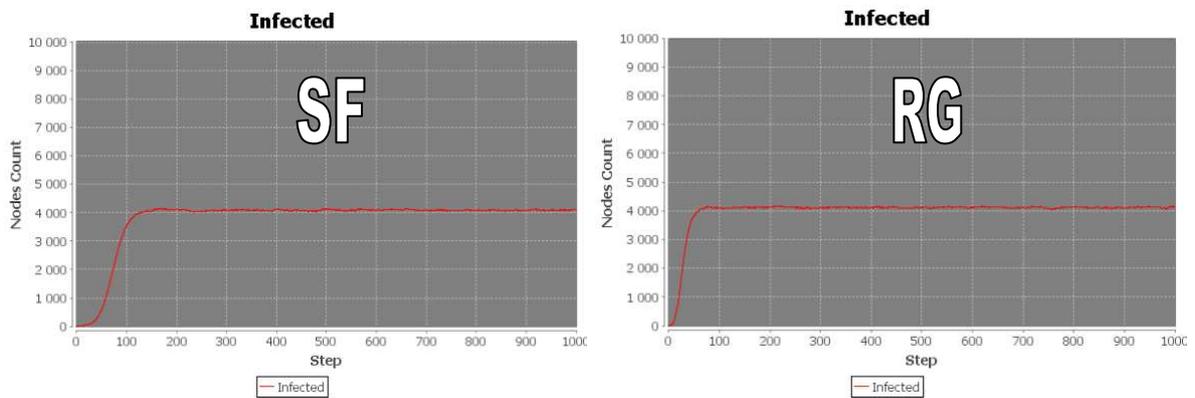


Fig. 21. SIS model of a disease with $\lambda=5$ in networks with different topology.

We can see that if λ is high (e.g. $\lambda=5$), topology of networks have small impact on the diffusion dynamic. According to Figure 20, the number of infected individuals rose sharply and flattened out at a very high level (about 90%).



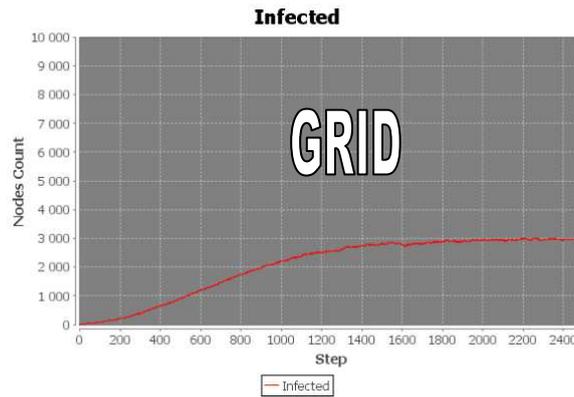


Fig. 22. SIS model of a disease with $\lambda=0.5$ in networks with different topology.

When λ parameter decreases the diffusion dynamic is more and more dependent on network topology. For $\lambda=0.5$ the diffusion dynamic in GRID-based graph is significantly different from the diffusion in *Scale Free* and *Random Graph*. First of all, the number of infected individuals rose slower, secondly flattened out at a lower level (about 30% by contrast with 40% for *Scale Free* and *Random Graph*).

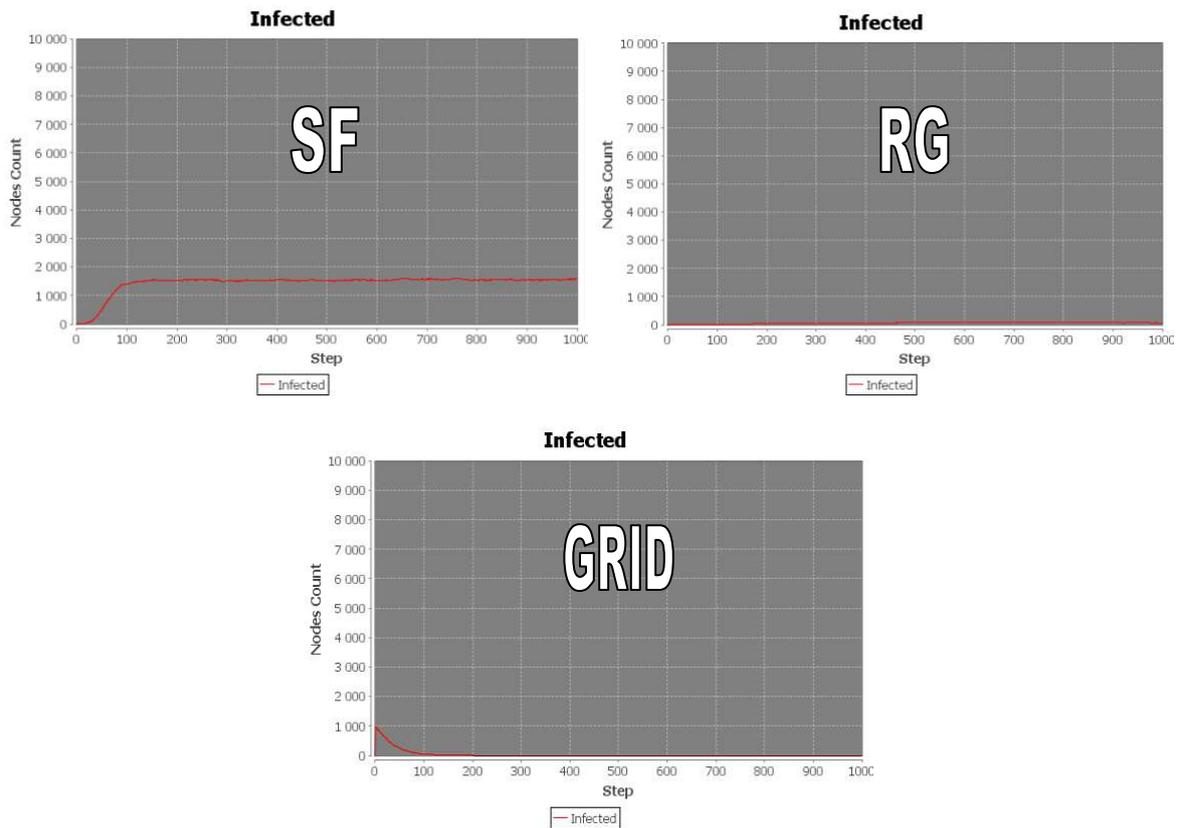


Fig. 23. SIS model of a disease with $\lambda=0.25$ in networks with different topology.

It turns out that for $\lambda=0.25$ the virus of infection diseases disappears from population modelled as GRID-based graph (even though 10% individuals are infected at a start time).

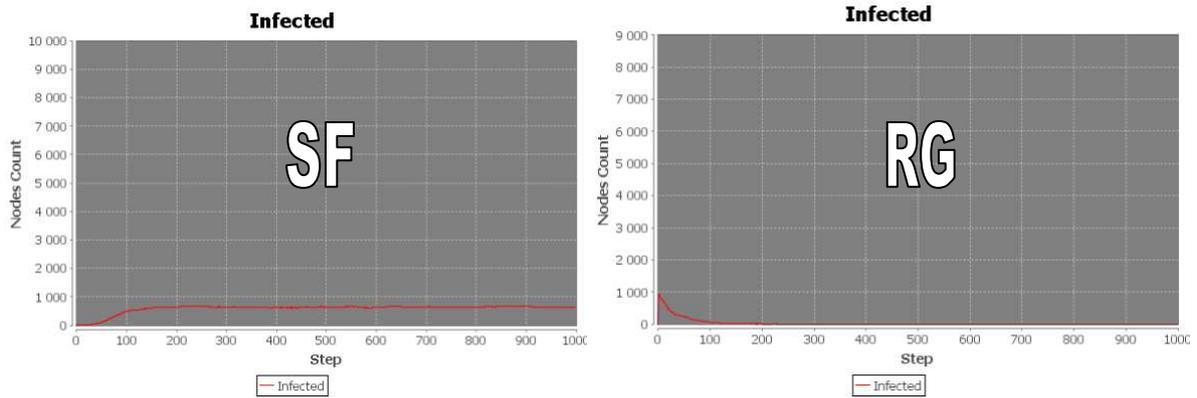


Fig. 24. SIS model of a disease with $\lambda=0.2$ in networks with different topology.

For $\lambda=0.2$ the virus of infection diseases also disappears from population modelled as *Random Graph* (even though 10% individuals are infected at a start time).

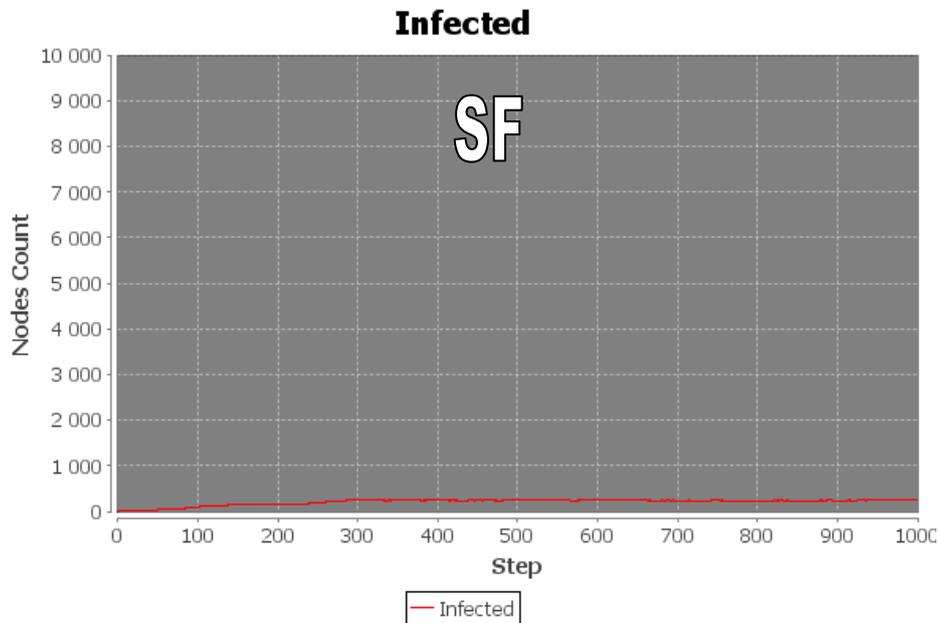


Fig. 25. SIS model of a disease with $\lambda=0.15$ in networks with different topology.

For $\lambda=0.15$ the virus is able to spread only in *Scale Free* network. It is an answer to the question: Why even slightly contagious diseases can plague human population over a long time without being epidemic. It was also analytically proved that in *Scale Free* network there is no epidemic threshold for λ value (Pastor-Satorras and Vespignani 2001).

9 System CARE

As practical utilization of our research system called CARE (*Creative Application to Remedy Epidemics*) was developed (Kasprzyk et al. 2011; Kasprzyk et. al. 2010; Kasprzyk at. al 2009). CARE is *Decision Support System*, which help decision makers to fight with epidemic. CARE contains five modules: *Disease Modelling*, *Social Network Modelling*, *Simulation*, *Vaccination* and *Questionnaires*.



Fig. 26. CARE Graphical User Interface (main screen).

In the *Disease Modelling* module, using probabilistic finite-state machine approach, we can model any kind of disease based on knowledge from the field of epidemiology. We allow to build the models of diseases with any states and transitions in the editor we have proposed. We are able to define some essential parameters like: the transition probability between the states, the minimum/maximum time that an individual spends in each state, the maximum number of neighbours that can be infected by an individual in a given time period and much more.



Fig. 27. Graphical User Interface of the *Diseases Modelling* module.

In *Social Network Modelling* module we can model and generate social networks using complex network theory. Using proposed generators we obtain synthetic networks but with the same statistical properties as real-world social networks. The algorithms generate networks that are *Regular Graphs*, *Random Graphs*, *Small World Networks*, *Scale Free Networks* or modifications thereof.

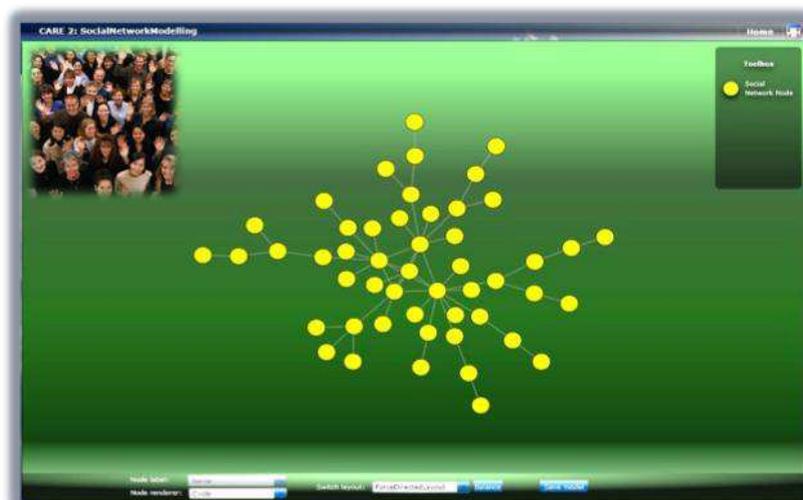


Fig. 28. Graphical User Interface of the *Social Network Modelling* module.

Using *Simulation* module we can visualize and simulate how the epidemic will spread in a given population. The system proposes two ways of information visualization. The first way is called “*Layout*” and helps user to manipulate networks and to set up some parameters of simulation.



Fig. 29. Graphical User Interface of the *Simulation* module with “*Layout*” visualization.

The alternative way is „*Geo-contextual*” one which allows to visualize networks on the world map.

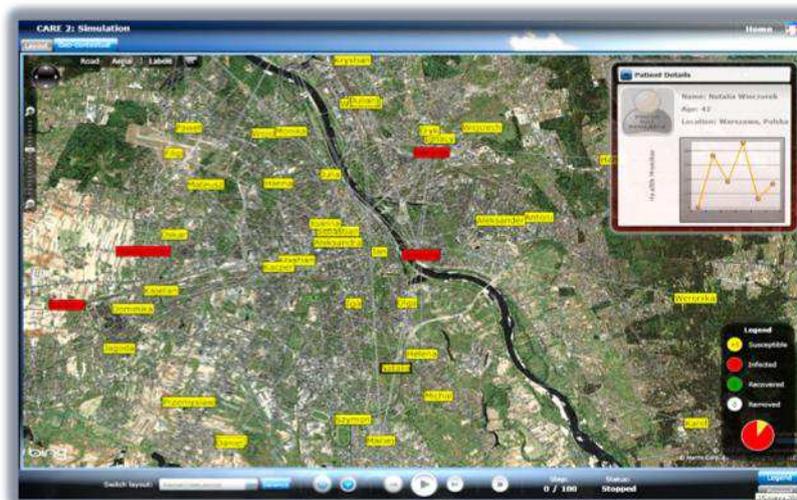


Fig. 30. Graphical User Interface of the *Simulation* module with “*Geo-contextual*” visualization.

The system estimates the expected outcomes of different simulation scenarios and generate detailed reports. The user can assess the results and the effectiveness of the chosen vaccination strategy. A report chart is created on the basis of the simulation. The x-axis represents simulation steps and the y-axis represents a count of individuals in each state in appropriate steps.



Fig. 31. Graphical User Interface of *Reports* generated from the *Simulation* module.

Based on the centrality measures *Vaccination* module helps the user to identify so called “super-spreaders” and to come up with the most efficient vaccination strategy. The identification and then vaccination or isolation of the most important individuals of a given network helps decision makers to reduce the consequence of epidemics or even stop them early in the game. We use a number of centrality measures to address the question “Who is the most important person in a given social network from the epidemic point of view?”. We show how to discover the critical elements of any network, the so-called “super-spreaders” of a disease.

The crucial step in fighting against a disease is to get information about the social network subject to that disease. *Questionnaires* module helps building special polls based on sociological knowledge to help discover network topology. Polls designed in this way are deployed on mobile devices to gather data about social interaction.



Fig. 32. Graphical User Interface of *Questionnaires* module.

10 Conclusion

In this paper we have presented some basic principles and ideas that can provide a deeper understanding of the diffusion processes on networks, in particular the dynamic of contagious diseases. We would like to admit that we are a little bit closer to understand diffusion in networks. As a relevant deliverables the novel model of diffusion on networks and the experimental environment, based on *Gephi* platform, was introduced. These tools can be used in a variety of applications:

- diffusion of any phenomena on any networks;
- identification of nodes in networks that are the most important from a different point of view;
- estimating the effectiveness of removal strategies;
- estimating the amount of resources to stop, slow down or per contra speed up any phenomena on networks.

The solutions presented in the paper have practical implementation as a system to fight with contagious diseases called CARE. It is worth to mentioned that CARE has its counterpart to fight with malwares in the Internet called VIRUS (Kasprzyk 2010).

Declarations

- Ethics approval consent to participate

Not applicable

- Consent for publication

Not applicable

- Availability of data and material

The datasets used and/or analysed during the current study are available from the corresponding author on reasonable request.

- Competing interests

The author(s) declare(s) that they have no competing interests

- Funding

This work was partially supported by Statutory Research Work in the Institute of Computer and Information Systems, Cybernetics Faculty, Military University of Technology in Warsaw, Poland.

- Authors' contributions

Conception and design of the work: Rafał Kasprzyk, Andrzej Najgebauer; Data collection: Rafał Kasprzyk; Data analysis and interpretation: Rafał Kasprzyk; Drafting the article: Rafał Kasprzyk; Critical revision of the article:

Andrzej Najgebauer, Rafał Kasprzyk; Final approval of the version to be published: Andrzej Najgebauer, Rafał Kasprzyk.

- Acknowledgements

Zbigniew Tarapata; Cezary Bartosiak;

References

- Barabási A.L., Albert R. (1999): *Emergence of Scaling in Random Networks*, Science, vol.286, pp.509-512.
- Barabási A.L., Albert R. (2000): *Topology of Evolving Networks: Local Events and Universality*, PRL vol.85(24) pp.5234-5237.
- Barabási A.L., Albert R. (2002): *Statistical mechanics of complex networks*, Review of Modern Physics, vol.74, pp.47-97.
- Bartosiak C., Kasprzyk R., Najgebauer A. (2013): *The graph and network theory as a tool to model and simulate the dynamics of infectious diseases*, BAMS, Vol. 9, Issue 1, 17-28.
- Bastian M., Heymann S., Jacomy M., (2009): *Gephi: an open source software for exploring and manipulating networks*, International AAAI Conference on Weblogs and Social Media.
- Cohen R., Havlin S., Avraham, D. (2003): *Efficient Immunization Strategies for Computer Networks and Population*, PRL, 12 December 2003, Vol. 91, pp. 247901-1.
- Crucitti P., Latora V., Marchiori M., Rapisarda A. (2004): *Error and attack tolerance of complex networks*, Physica A 340 (2004), pp. 388-394.
- Erdős P., Rényi A. (1959a): *On random graphs*, Publicationes Mathematicae 6, pp.290-297.
- Erdős P., Rényi A. (1959b): *On the evolution of random graphs*, Publications of the Mathematical Institute of the Hungarian Academy of Sciences 5, pp.17-61.
- Godin S. (2001): *Unleashing the Ideavirus*, Hyperion, New York.
- Harary F., Gupta G. (1997): *Dynamic Graph Models*, Mathl. Comput. Modelling, vol.25(7), pp.79-87.
- Korzan G. (1978): *Elementy Teorii grafów i sieci – metody i zastosowania*, WNT, Warsaw.
- Kasprzyk R., Najgebauer A., Pierzchała D. (2011). *Modelling and Simulation of an Infection Disease in Social Networks*, 3rd International Conference on Computational Collective Intelligence (ICCCI 2011), *Lecture Notes in Artificial Intelligence* (6922), 388-398.
- Kasprzyk R., Najgebauer A., Pierzchała D. (2010). *Creative Application to Remedy Epidemics*, Proceedings of the 7th International Conference on Computer Simulation in Risk Analysis and Hazard Mitigation, Algarve, Portugal, SEP 13-15, 1545-1561.
- Kasprzyk R. (2010): *Symulator rozprzestrzeniania się złośliwego oprogramowania w sieciach komputerowych*, Symulacja w badaniach i rozwoju, vol.1(2), pp. 139-150.
- Kasprzyk R. (2009): *The vaccination against epidemic spreading in Complex Networks*, Biuletyn ISI, ISSN 1508-4183, No. 3(1/2009), pp.39-43.

- Kasprzyk R., Lipiński B., Wilkos K., Wilkos M., Bartosiak C. (2009): *CARE - Creative Application to Remedy Epidemics*, Biuletyn ISI, ISSN 1508-4183, No. 3(1/2009), pp.45-52.
- Kasprzyk R. (2012a): *Modele ewolucji systemów złożonych i metody badania ich charakterystyk dla potrzeb komputerowej identyfikacji potencjalnych sytuacji kryzysowych*, PhD thesis, supervisor A. Najgebauer, Faculty of Cybernetics, Military University of Technology in Warsaw.
- Kasprzyk R. (2012b): *Diffusion in Networks*, Journal of Telecommunications and Information Technology, 2/2012, 99-106.
- Leskovec J., Adamic L., Huberman B.A. (2007): *The Dynamics of Viral Marketing*, ACM Transactions on the Web, vol. 1(1), Article 5.
- Lloyd A.L., May R.M. (2001): *How Viruses Spread Among Computers and People*, Science, vol.292(5520), pp.1316-1317.
- López-Pintado D. (2008): *Diffusion in complex social networks*, GAMES and Economic Behavior vol.62(2), pp.573-590.
- Madar N., Kalisky T., Cohen R., Ben-Avraham D., Havlin S. (2004): Immunization and epidemic dynamics in complex networks, *Eur. Phys. J. B.*, 2004, Vol. 38, pp. 269-276.
- Newman M.E.J. (2000), *Models of the small world: A review*, J.Stat.Phys., vol.101, pp.819-841.
- Newman M.E.J. (2003), *The structure and function of complex networks*, SIMA Review, vol.45(2), pp.167-256.
- Pastor-Satorras R., Vespignani A. (2001): *Epidemic Spreading in Scale-Free Networks*, PRL vol.86(14), pp.3200-3203.
- Sokolova A., de Vink E. P. (2004): *Probabilistic Automata: System Types, Parallel Composition and Comparison*, LNCS 2925, Springer-Verlag, Heidelberg, pp.1-43.
- Tarapata Z., Kasprzyk R. (2010): *Graph-based optimization method for information diffusion and attack durability in networks*, Lecture Notes in Computer Science, Volume 6086/2010, p. 698-709, Springer.
- Vidal E., Thollard F., de la Higuera C., Casacuberta F., Carrasco R. C. (2005): *Probabilistic Finite-State Machines – Part I*, IEEE Transactions on Pattern Analysis and Machine Intelligence, vol.27(7), pp. 1013-1025.
- Strogatz S.H. (2001): *Exploring complex networks*, Nature, vol.410, pp.268-276.
- Watts D.J., Strogatz S.H. (1998): *Collective dynamics of „small-world” networks*, Nature, vol. 393, pp. 440-442.
- Wang X., Chen G. (2003): *Complex Networks: Small-world, scale-free and beyond*, IEEE Circuits and Systems Magazine, vol. 3(1), pp.6-20.
- Gephi, Project HomePage (2020), [Online] <http://gephi.org/>, last accessed 2020/04/15.