

The vaccination against epidemic spreading in Complex Networks

Rafał Kasprzyk
e-mail: rkasprzyk@wat.edu.pl

Institute of Information Systems
Cybernetics Faculty, Military University of Technology
Kaliskiego Str. 2, 00-908 Warsaw

The paper focuses special attention on research of Complex Networks. Complex Networks have *Scale Free* and *Small World* features, what make them accurate model of many networks such as social networks. These features, which appear to be very efficient for communication networks, favor at the same time the spreading of many diseases.

Based on defined centrality measures, we show how to discover the critical elements of any network. The identification and then vaccination of the critical elements of a given network should be the first concern in order to reduce the consequence of epidemics.

Keywords: complex networks, small world, scale free, vaccination strategy

1. Definition and notation

Complex Networks are commonly modeled by means of simply or directed graphs that consist of sets of nodes representing the objects under investigation, joined together in pairs by links if the corresponding nodes are related by some kind of relationship.

Formally graph is a vector $G = \langle V, E, P \rangle$ where: V is a set of vertices, E is a set of edges, and P is an incident relation e.g. $P \subset V \times E \times V$.

The degree k_i of a vertex v_i is the number of edges originating from or ending in vertex v_i . The shortest path length d_{ij} is an alternating sequence of nodes and links, starting and ending with a node. The length of a path is defined as the number of links in it. Now we can define diameter D as the shortest longest path e.g. $\max \{ d_{ij} \}$. Networks very often are represented by a matrix called the adjacency matrix \mathbf{A} , which in the simplest case is an $n \times n$ symmetric matrix, where n is the number of vertices in the network. Element of adjacency matrix $A_{ij}=1$, if there is an edge between vertices i and j , and 0 otherwise.

In some cases the use of graph to represent Complex Networks does not provide a complete description of the real-world systems under investigation. For instance, if contacts in social networks are represented as a simple graph, we only know whether individuals are connected, but we cannot model strength of these connection. However, for further consideration, we temporally use only formal graph definition.

2. Networks generators

In 1960 Erdős and Rényi [1] described their investigations of random graphs. Assuming equally probable and independent random connections made between any pair of vertices in initially not connected graph, they derived a model suffering unrealistically topology. Because of that their model had limited usage for modeling real life social network, nevertheless they proved a number of interesting result about random graphs. There are a few models of random graphs, some of them we have implemented.

Identifying and measuring properties of a Social Networks is a first step towards understand their topology, structure and dynamics. 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. Apparently, networks derived from real data (most often spontaneously growing) have “*six degree of separation*”, power low degree distributions, hubs occurring, tendency to form clusters and many other interesting features. Two very interesting models capture these feature, have been introduced recently.

First, Watts and Strogatz in 1998 [2] deal with mentioned features by a strategy that seems perfectly obvious once someone else has thought of it. They interpolate between two known models. They begin with regular lattice, such as a ring, and then ‘rewire’ some of the edges to

introduce randomness. If all edges are rewired a random graph appears. The process of rewiring effects not only the average path length but also clustering coefficient. Both of them decrease as probability of rewiring increases. The striking features of this procedure is that for relatively wide range of rewiring probability the average path length is already low while clustering coefficient remains high. It is called *Small World* model, more precisely Beta-model of *Small World* network. Because Beta-model exists so there is also Alfa-model of *Small World* network. This model try to capture, of course in some abstract sense, the authentic/real way of social connections formation. What is surprising is not that real Social Networks are *Small World* but that people are able to find the shortest path between each other so easily. Kleinberg [3] explained it using his own model of *Small World* networks.

Second, Barabasi and Albert in 1999 [4] introduced their model of networks as a result of two main assumption: constant growth and preferential attachment. They expressed the degree sequence – the number of vertices with each possible number of edges. They show why the distribution of degrees 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 to the existing nodes with probability proportional to their connectivity. In this way, high connected individuals receive more new links than low connected ones, and hence, ‘old’ nodes are more connected than ‘young’ ones. It is called *Scale Free* model. The process of *Scale Free* networks generation has many modifications [9].

3. Centrality measures

We start the analysis of epidemic spreading by introducing centrality measures [8], which are the most fundament and frequency used measures of network structure. The central vertices in Complex Networks are of particular interest because they might play the role of organization hubs. Centrality measures address the question “Who (what) is the most important or central person (node) in given social network?”. No single measure of center is suited for all application. Based on defined centrality measures, we show how to discover the critical elements of any network so-called “*super-spreaders*” of a disease.

When a vaccination for a disease exist, immunizing certain individuals against being infected by a disease may be the most efficient

way to prevent loss of time and funds due to the disease. Obviously, immunization of the entire population will eradicate the disease entirely, but this is not always possible, or may involve high cost and effort. Therefore, the choice of which individuals to immune is an important step in the immunization process, and may increase the efficiency of the immunization strategy. So the identification and then vaccination of the critical elements of a given network should be the first concern in order to reduce the consequence of epidemics.

We considered five most important centrality measures e.g. degree (gives the highest score of influence to the vertex with the largest number of first neighbors), radius (chooses the vertex with the smallest value of shortest longest 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), closeness (focuses on the idea of communications between different vertices and the vertex which is ‘closer’ to all vertices gets the highest score), betweenness (it can be defined as the percent of shortest paths connecting any two vertices that pass through consider vertex) and eigenvector (acknowledges that not all connections are equal so connections to people who are themselves influential will lend a person more influence than connections to less influence people). Two of these measure will be presented in details.

4. Degree centrality

The simplest of centrality measures is degree centrality, also called simply degree. The degree centrality is traditionally defined analogous to the degree of a vertex, normalized over the maximum number of neighbors this vertex could have. This agrees with the intuitive way to estimate someone’s influence from the size of his immediate environment. Thus, in a network of n vertices, the degree centrality of vertex v_i , is defined as:

$$\text{center}_{[0,1]}^{\text{degree}} = \frac{k_i}{n-1}$$

The normalization in the region $[0, 1]$ is used here to make the centrality of different vertices comparable, and also independent of the size of the network.

5. Betweenness centrality

This measure assumes that the greater number of paths in which a vertex participates, the higher the importance of this vertex for the network is.

Betweenness centrality refines the concept of communications, introduced in closeness centrality.

Informally, betweenness centrality of a vertex can be defined as the percent of shortest paths connecting any two vertices that pass through that vertex. If $P_{ik}(i)$ is the set of all shortest paths between vertices v_i and v_k passing through vertex v_i and $P_{ik}(\text{absence } i)$ is the set of all shortest paths between vertices v_i and v_k not including v_i , then:

$$\text{center}_i^{\text{betweenness}} = \sum_{i \neq k} \frac{P_{ik}(i)}{P_{ik}(\text{absence } i)}$$

The normalized version divides this value with the maximum possible betweenness centrality, that is all possible shortest paths in a completely connected graph.

$$\text{center}_{[0,1]}^{\text{betweenness}} = \frac{2 \sum_{i \neq k} P_{ik}(i)}{(n-2)(n-1)}$$

This definition of centrality explores the ability of a vertex to be ‘irreplaceable’ in the communication of two random vertices. It is of particular interest in the study of network immunization, because at any given time the removal of the maximum betweenness vertex seems to cause maximum damage in terms of connectivity and mean distance in the network. Its main disadvantages is that the summation operator practically means that it needs global information about the network, in order to compute the betweenness of a single vertex, and that is simply not possible in many contexts. For the same reason it is expensive in computing time to compute the score of a vertex, although this disadvantage was significantly improved recently.

6. Connection efficiency

To evaluate how well a network is connected before and after the removal of a set of nodes we use the global connection efficiency (GCE), a measure introduced by Latora and Marchiori [7]. We assume that the connection efficiency between vertex v_i and v_j is inversely proportional to the shortest distance:

$$\text{connection}_{ij}^{\text{efficiency}} = \frac{1}{d_{ij}}$$

When there is no path in the graph between vertex v_i and v_j we have $d_{ij} = \text{infinity}$ and consequently connection efficiency is equal zero. The global connection efficiency is defined

as the average of the connection efficiency over all couples of nodes.

$$\text{global_connection}^{\text{efficiency}} = \frac{2 \sum_{i \neq j} \text{connection}_{ij}^{\text{efficiency}}}{n(n-1)} = \frac{2}{n(n-1)} \sum_{i \neq j} \frac{1}{d_{ij}}$$

Unlike the average path length, the global connection efficiency is well-defined quantity also for non-connected graph.

7. How to stop disease without full knowledge about network

When vaccine supplies for a deadly (serious, common) disease are limited, whom should health workers target? Our strategies for vaccination with limited resources suggests that the “super-spreaders” should be immunized. But how these individuals can be reached without knowing who they are in advanced? We know how and we believe we can do it!

Many researches in Social Networks show that human contacts have *Scale Free* feature, what was mentioned earlier. In consequence most of the individuals have only a few links to others, while a few individuals have a very large number of contacts. These features, which appear to be very efficient for communication networks, favor at the same time the spreading of many diseases [6]. What now... if one of these highly connected individuals in a human population become infected, he becomes a “super-spreader” infecting all of his numerous neighbors and distributing the disease very fast. This suggests a deceptively simple solution like immunizing all the “super-spreaders” in a network of human contacts slows or even stops the spread. Unfortunately, it is rarely obvious which individuals have many contacts. People infrequently know exactly how many contacts they have and sometimes they lie, particularly about sensitive subject like number of sexual partners.

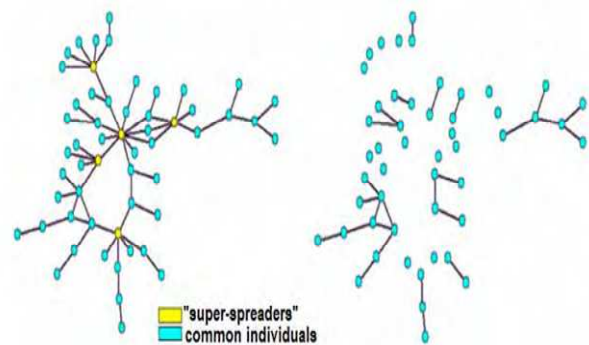


Fig. 1. Importance of “super-spreaders” identification. If we vaccinate the “super-spreaders” e.g. with highest degree, the network rapidly disintegrates and we stop disease spreading

Random immunization is almost useless because *Scale Free* networks remain connected even after up to 80% of their all nodes are removed (immunized or isolated). So almost whole population must be vaccinated to prevent the disease's spread. The word "randomly" is the keyword here. Any smart attacks i.e aimed at "super-spreader" disintegrate the network rapidly.

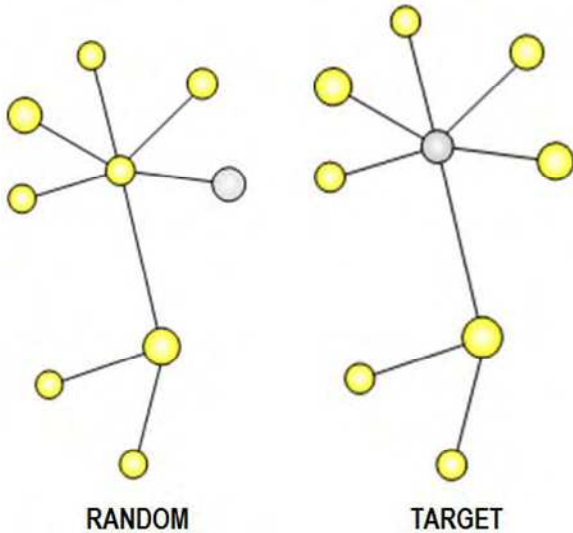


Fig. 2. Two most popular vaccination strategy

What does it mean smart if we do not know who the "super-spreaders" are and/or where can we find them? We can do it just put in a simple modification of random vaccination that is much more effective, according to our computer simulation and based on new concept introduced in [10, 11] with few modification.

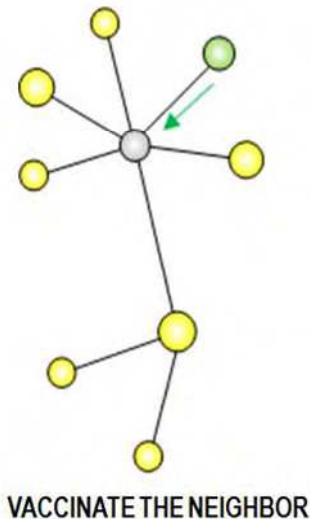


Fig. 3. New vaccination strategy very effective even the knowledge about network topology is uncertain and incomplete

The idea is to randomly choose, say, 20% of the individuals and ask them to fill out our special

questionnaires. One of the most important question in all forms for any disease is to name at least one acquaintance/friend/boy/colleague etc., then vaccinate those identified individuals. Potential "super-spreaders" have such a large number of contacts that they are very likely to be named at least once. On the other hand, the "super-spreaders" are so few in number that the random sample of individuals is unlikely to include many of them. Using this vaccination strategy, a disease can be stopped by vaccinating less than 20% of the individuals. If a larger sample is asked for names, or those named twice are vaccinated the total number of vaccinations required can be even lower. We could modify this basic method in many ways and adopt it to specific virus.

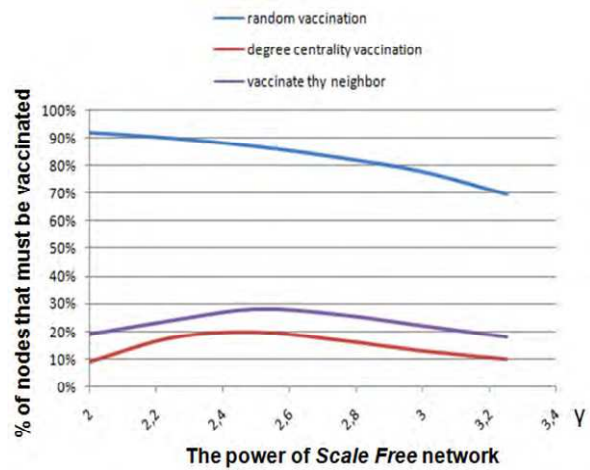


Fig. 4. The effect of three vaccination strategy in case *Scale Free* networks with different power

8. Summary

Presenting idea is a new attempt at integrating theories and practices from many area, in particular: social networks, graph and network theory, decision theory, data mining and security. It utilizes that theoretical basis for very practical purpose of growing importance and demand: widely understood countering high contagious diseases like HIV/AIDS, SARS and others. Our work have enormous practical potential in region such as Africa, where there are not enough medicine to treat all who are at risk

9. References

- [1] P. Erdős, A. Rényi: On random graphs, Publ. Math Debrecen 6 (1959), 290-297.
- [2] Watts Duncan J., Strogatz Steven H.: *Collective dynamics of „small-world” networks*, Nature, 393:440-442, 1998.
- [3] Jon M. Kleinberg: *Navigation in small world*, Nature 406, 845 (24 August 2000)
- [4] Barabási Albert László, Albert Réka: *Emergency of Scaling in Random Networks*, Science, 286:509-512, 1999.
- [5] Brandes U., Kenis P., and Raab J.: *Explanation Through Network Visualization*, Methodology 2006, Vol. 2(1): 16-23.
- [6] Romualdo Pastor-Satorras & Alessandro Vespignani: *Epidemic Spreading in Scale-Free Networks*, PRL Volume 86, Number 14 p. 3200 (2 April 2001).
- [7] Paolo Crucitti, Vito Latora, Massimo Marchiori, Andrea Rapisarda: *Error and attack tolerance of complex networks*, Physica A 340 (2004), 388-394.
- [8] Stefan Wuchty, Peter F. Stadler: *Centers of complex networks*, Journal of Theoretical Biology 222 (2003), 45-53.
- [9] Barabási Albert László, Albert Réka, *Topology of Evolving Networks: Local Events and Universality*, PRL Volume 85, Number 24 p.5234 (11 December 2000).
- [10] Reuven Cohen, Shlomo Havlin, Daniel ben-Avraham: *Efficient Immunization Strategies for Computer Networks and Population*, PRL Volume 91 Number 24 p.247901-1 (12 December 2003).
- [11] N. Madar, T. Kalisky, R. Cohen, D. ben-Avraham, S. Havlin: *Immunization and epidemic dynamics in complex networks*, Eur. Phys. J. B 38, 269-276 (2004).d