

## SOCIAL NETWORKS FOR EPIDEMIC SPREADING: A CASE STUDY

SILVIA RAUSANU AND CRINA GROSAN

ABSTRACT. The study of social networks and the spread of disease on networks has recently attracted considerable attention and there is a huge amount of work dealing with this matter. The current work presents and develops a social network model based on the multigraph idea in order to simulate the spread of an epidemic. The paper deals with a real example which refers to the spread and evolution of A/H1N1 virus in Romania. The simulations obtained by applying the proposed model approximate very well the real evolution of the virus studied over a period of 10 months.

### 1. INTRODUCTION

The epidemiologists have developed mathematical models to estimate the size and other key features of an epidemic and social networks seem to be a very good candidate in simulating the spread of such epidemics [4, 9]. There are several approaches which simulate how a disease is transmitted among the nodes of a network which has the proprieties of a social network [3, 11, 10].

In our paper we develop a very general model which uses more than a simple social network but a set of connected social networks in order to simulate the spread of a well know epidemic of the past year and the begging of the current one: the A/H1N1 virus.

The appearance of the virus A/ H1N1 was dated at the end of April 2009 in Mexico, ever since this virus has spread around the globe with an amazing speed. Romania was reached by the virus only on the 23<sup>rd</sup> of May 2009. As this virus can be contacted by air, the spread among the countries was facilitated by the multiple means of transportation, but not only among the countries, but also inside them.

The case of Romania did not make any exception: the virus was contacted from abroad, firstly in the main cities of the country, then, following the same

---

Received by the editors: August 13, 2010.

2010 *Mathematics Subject Classification.* 91D30.

1998 *CR Categories and Descriptors.* code [I6.3]: Computing Methodologies – *Simulation and Modeling*; code [G2.2]: Mathematics of Computing – *Graph Theory*.

*Key words and phrases.* social networks, epidemics spread modeling.

rule of moving masses of population, in almost each county. In addition to these factors, the time in which the spread got a higher speed was in the months when the likelihood of flu infection was greater and the crowding in different areas was much more common.

These factors can form a pattern in the spread of the virus, at any level of community: inside a county or inside a country. However, in Romania the spread did not respect entirely the patterns, nor the expectations of the Health Minister of Romania, as some counties, during the entire epidemic development, did not even had a case of infection with the virus A/H1N1.

The paper is organized as follows: Section 2 presents the basic social networks notions used in our model, in Section 3 some of the existing work in epidemics spreading is analyzed, in Section 4 the proposed model is presented in detail. Section 5 presents details about the simulations followed by experiments. Section 6 contains conclusions and future work ideas.

## 2. SOCIAL NETWORKS

The notion of social network and the methods of social network analysis have attracted considerable interest and curiosity both from the social-behavioral community and from the computer science and mathematical community in the recent decades. Research in the social studies is mostly interested in the reason behind the network connection rather than the properties of the network structure itself. On the other side, the research in graph-theory has provided a wealth of quantitative tools and mechanism for describing networks mostly by analyzing the type of relations and the general structure.

Along with the growing interest and amount of research – in distinct domains – over the use and properties of social networks, it has been established a general set of principles for the social network analysis suitable for any domain [9].

During the studies of real-world social networks, some research was oriented towards the statistical properties of these networks. One important and fundamental result that has emerged from these studies concerns the numbers of ties that actors have to other actors, their so-called “degrees.” It has been found that in many networks, the distribution of actors’ degrees is highly skewed, with a small number of actors having an unusually large number of ties. Simulations and analytic work have suggested that this skewness could have an impact on the way in which communities operate, including the way information travels through the network and the robustness of networks to removal of actors. Among the developed models for networks, three of them will be recalled in what follows.

We first give some basic definitions which will be further used.

Consider a *graph*  $G(V, E)$  is a plotting structure for the set of vertices  $V$  (nodes) and the set of edges  $E$  for the connection of pairs of vertices. The cardinality of the set  $V$  is denoted by  $n$  and is known to be the *order* of the graph, while the cardinality of  $E$ ,  $m$ , is called the *size* of the graph. Two joined vertices by an edge are called *end-vertices*. Any two vertices connected by an edge are *adjacent* or *neighbors*.

*Definition 1 (multigraph)*

A *Multigraph* is a directed or undirected graph in which is allowed to have in the set of edges some edges contained more than once. If the edge  $e$  occurs several times in  $E$ , then its copies are called *parallel edges*.

*Definition 2 (degree of a vertex)*

The number of edges having as end-vertex the node  $v_1$  in the graph is called the *degree of vertex  $v_1$* .

*Remark 1:*

If the graph is a multigraph, then the parallel edges are counted according to their multiplicity.

*Centrality indices* exist for the quantification of an intuitive feeling that in most networks some vertices or edges are more central, important, than others. The computations of such indices provide new methods for the analysis of the social network.

*Definition 3 (degree centrality)*

*Degree centrality* is a characteristic of the node with the highest degree – the connector/hub of the network.

*Definition 4 (Betweenness centrality)*

*Betweenness centrality* characterizes a node that influences the flow of the network but not the network itself, in graph theory being a critical point for connectivity of the graph.

*Definition 5 (closeness centrality)*

*Closeness centrality* is the quality of a node that has the shortest path to every other node.

*Network centralization* is the property of keeping the connectivity of the graph around one or few vertices. High centralization on a node may lead to fragmentation of the network, as there is one critical node, while low centralization guarantees smaller changes of such thing to happen.

*Definition 6 (clustering coefficient)*

The *clustering coefficient* [3] refers to the tendency (observed in many natural networks) of forming *cliques* in the neighborhood of one given vertex. It measures the average probability that two neighbors of vertex  $i$  are also connected.

The clustering coefficient is computed as follows: let  $i$  be a vertex, with degree  $k$  and  $e$  the number of edges formed among its neighbors, then the

clustering coefficient is the ratio between the actual number of edges among neighbors and the maximum number of possible edges between neighbors:

$$c = \frac{e}{\frac{k \cdot (k+1)}{2}}$$

*Definition 7 (degree distribution)*

*Degree distribution* is a statistical characterization of the graph given by the sequence of degrees or by the relative probability distribution of degrees.

**2.1. Random networks – Erdos-Renyi model.** The random network model of *Erdos* and *Renyi* (ER) is considered as the first model able to coherently describe networks of arbitrary size [3]. The rules to create a random graph according to *Erdos* and *Renyi* prescriptions are simple. The network is a set  $N$  of  $n$  different vertices, and  $m$  edges ( $E$ ) joining the nodes, such that between any two vertices  $i, j$  exists a connecting edge with the independent probability  $p$ . In these conditions, the maximum number of edges in the network is:

$$\max(m) = \frac{n \cdot (n - 1)}{2} \cdot p$$

and the maximum vertex degree is:

$$\max(d(i)) = n-1, \forall i \in N.$$

The average degree is easily computed:

$$k_{average} = (n-1) \cdot p$$

The clustering coefficient is  $c = p$ , since it is simply the probability that two nearest neighbors of a vertex of degree  $k$  have an edge between them. The probability that a randomly chosen node has exactly the degree  $k$  (that is that only  $k$  of its possible edges are present and  $n - 1 - k$  are absent) is given by the *binomial distribution*:

$$P(k) = \binom{n-1}{k} \cdot p^k (1-p)^{n-1-k}$$

However, as a model of real-world social network, it has some serious shortcomings. Perhaps the most serious is its degree distribution, which is quite unlike for those seen in most real-world networks. The approximation to the real-world model is poor as the degree distribution is highly skewed. On the other hand, the random graph has desirable properties, particularly the fact that many features of its behavior can be calculated exactly.

Some algorithms might work in creating a more appropriate approximation to real-world networks. Starting from knowing only the degree distribution of the desired network, the probabilities for a random vector to have degree  $k$ , must be normalized. Take a number of  $N$  vertices and assign to each a number of  $k$  ends of edges, where  $k$  is a random number drawn independently from

the distribution  $p_k$  for each vertex. Then we take the ends of edges randomly in pairs and join them up to form edges between vertices. This procedure will produce a graph with exactly the desired degree distribution, but which is in all other respects random. To put it another way, we have generated a graph that is drawn uniformly at random from the set of graphs with the given degree distribution.

**2.2. Scale-free model.** Scale-free networks are characterized by the power law distribution in degree distribution, which can be expressed mathematically as  $P(k) \approx k^{-\gamma}$ . From the form of the distribution it is clearly that when  $\gamma < 2$ , the average degree diverges and when  $\gamma < 3$ , the standard deviation of the degree diverges [8]. It has been found that most scale-free networks have exponents between 2 and 3, thus, they lack a characteristic degree or scale, and therefore their name.

The clustering coefficient distribution is in a strong reverse connection with the degree distribution, as it decreases with the increase of the degree of the node.

The *Erdos-Renyi* model is less likely to produce a scale-free network as their properties are not consistent with those of the ER model; consequently a model for growth process is needed. The most widely generative model is the one of *Barabási* and *Albert's* (*'rich and get richer'*) in which each new vertex connects to another one with a probability that is not uniform, but proportional with the in-degree of the old-vertex. Later development led to respecting the power-law distribution by making a supposition according to which a vertex with high in-degree will tend to attract more vertices than the other ones.

**2.3. Small world model.** The small-world property refers to the fact that in many large scale networks the average distance between vertices is very small compared to the size of the graph. The distance between two vertices in a graph is measured as the shortest path length  $l$  among them. The small-world property is present when  $l$  scales logarithmically (or even slower) with the number of vertices.

Inheriting from sociology, this model is known as well for "*six degree separation*", stating that a short number of acquaintances (on average six) is enough to create a connection between any two 'actors' chosen randomly.

From the mathematical point of view, the small-world effect describes those graphs whose diameter and average path-length grow much more slowly than the number of nodes  $n$ , typically,  $\ln n$ , just as in random ER model. Yet, a random graph has a very small local interconnectedness, captured by the clustering index.

Many scale-free networks are embracing the small-world model [3]: the average path length and the clustering coefficient of the network are compared with the same quantities calculated over random graphs with the same number of nodes and vertices. However, the reverse is not true: small-world networks are not scale-free networks.

### 3. RELATED WORK ON EPIDEMIC SPREADING

In [1], the impact of the clustering coefficient and of the assortativity coefficient in a randomly generated network over epidemic behavior is analyzed. The authors start with a generation algorithm for the social network, emphasizing the positions of the already infected nodes, by placing them on an imaginary ring. For the simulation, it is used the SIR epidemic model which is run for distinct configuration of the network: different indices or different infected population size. The results of the experiment prove that the infection size, after a number of simulations, is negatively proportional with the values of the clustering and assortativity indices. The greater is the clustering index of the network, the smaller is the size of the newly infected population, this statement being true for the assortativity index, as well.

A project closer to the theme discussed in this paper is [2] which propose a model for the progression of pandemic influenza A (H1N1) in Vietnam. The main difference which arises from this paper is the used structure, distinct from social networks. It was developed an age- and spatially-structured mathematical model in order to estimate the potential impact of pandemic H1N1 in Vietnam and the opportunities for reassortment with animal influenza viruses. The model tracks human infection among domestic animal owners and non-owners and also estimates the numbers of animals may be exposed to infected humans.

In [6], the simulation of the pandemic influenza is performed using graphical representation on the maps. Maps are built by stitching the counties that contain cities and localities, the encoding of these elements using different colors on the map and the generation of the neighbor relationship.

### 4. MODEL DESCRIPTION

The case of spreading the virus A/H1N1 in Romania might be modeled at the level of counties taking into account their connections with the other counties in Romania, but also the connections with other countries affected by the this epidemic. However, this approach would not have been enough as in the high peek of spreading, most of the infection were occurring only inside a county due to some extra crowding coefficients.

This is why the epidemics must be modeled as well at the level of a county by tracking (or assuming) the connections between the locators of the county.

The system will be represented as a network of networks.

The main network has as nodes the counties of Romania and some other important countries of the world with which Romania has a stronger connection. The edges are the connections between the counties created on the basis of neighborhoodness, collegial nodes, railroad nodes, important airports, tourism attractions. The influence of these connections is underlined by some probability of infection in one current node with the help of the number of infection cases in the counties at the other end of the edge. The networks inside the counties are advisable in simulating the epidemic because are nearer to the real situation and, furthermore, are a must in obtaining some results closer to reality. These inner-networks are impossible to be created based to reality, as tracking millions of connections is way out of any league, so they will be randomly generated based on a known model of networks: *Erdos-Renyi*. According to this model, in a network of  $n$  nodes, any two vertices are connected with an independent probability  $p$  which means that the network has as clustering index that initial probability  $p$  since it is simply the probability that two nearest neighbors of a vertex of degree  $k$  have an edge between them.

In the real situation, a county of  $m$  individuals with a population density  $d$ , is manipulated according to the *Erdos-Renyi* model [5] by generating a random network with  $m'$  nodes and clustering index  $d'$ .  $m'$  is a scale of the size of population  $m$  according to the maximum size of the social network permitted by the implementation.  $d'$  is also a scale of the density but according to the domain values of the clustering index :  $[0, 1]$ , which means that the county with the highest population density value will have the highest clustering index; although this would imply that the maximum clustering index between all the counties is 1, the maximum clustering index will actually be around 0.6 as any higher value will transform the network into an almost complete graph, situation which is mostly uncommon for large real-world social network (like a county, in this situation). In this way some connection between individuals will be assumed in the scope of gaining the clustering index. No further properties are attached to the network during its construction as the purpose of the network is oriented towards the direct connections between nodes (not by paths, centrality properties, etc).

When a county has some cases of infection, it appears the issues of positioning these infected nodes in the network. In order to gain some local results, the infected node must have an important position in the network, one idea could be selecting the first  $i$  nodes with the highest degree. This decision is taken on the basis that in reality the spread of the virus is done by air, usually in crowded places, which in network-terms is translated as clusters.

Usually in clusters there are some nodes with high degree which are important to the network, consequently, for the current case, there should be placed the already infected individuals. After setting the infected nodes, their position is no longer of interest, but only their ‘identifier’ in the entire network.

The spreading simulation is performed at two levels, one at a level of counties and one at the level of individuals. The individuals are tightly connected to the county they belong to, their individual evolution being influenced by the same characteristics the county is, but in different proportions. One individual can be connected to one county only, in this way a much complex network is created, although the homogeneity of the nodes’ types and relevance lacks completely.

The factors of influence in the spreading are transmitted from the county in general, to each individual located in it, in particular. This transmission order classifies the network in two hierarchies, corresponding to the levels of spreading simulations, in this way the network for spreading fits the model of the hierarchical social networks. However, the flow of transmission is bidirectional, not only a county distributes its characteristics, but also the mass of individuals contribute to the final statistics computing for a county. Having the interdependence described above, the two hierarchies can be isolated and separated physically, but sharing the context of spreading, keeping the exchange of information during the simulation.

The spreading simulation is performed at two levels, one at a level of counties - the outer network - and one at the level of individuals - the inner network.

**4.1. The outer network.** The *outer network* is the network composed of the counties of Romania, some type of map of the country. Although a social network has usually as nodes sole individuals, a group or an organized formation of individuals can be at the basis of a network, in the current case, the mass of inhabitants of a certain country. The connections between these nodes are created on many criteria fact that change the structure of the network from a simple graph - directed or undirected - to a directed multigraph as between two nodes exists more than one type of links.

The outer network suits the scale-free model as it gathers its most important features. Firstly, the existence of hubs is underlined by some counties that tend to have lots of connections with the other nodes; these nodes have actually a high importance not only in the network but also in the country (collegial centers, main city, tourism nodes, etc). Secondly, the connectivity of the network is easily assured only by one type of connection: the neighborhood between counties, which makes the multigraph underlying it strongly



connected. The multigraph structure is defending the network from fragmentation as the important counties do not keep in their links towards others the key of connectivity.

The size of of the outer network is 41 (the number of counties in Romania).

The links between the vertices are formed on some pre-established conditions which imply only characteristics taken from reality. The most important reasons for putting an edge between two nodes are: geographical closeness; collegial surroundings; nodes of means of transportation - railway, airport; tourism attractions; poverty level.

**4.2. The inner network.** The "inner network" is developed from one internal node of the outer network, inheriting some computed or native attributes from it.

The nodes of the inner network (corresponding to each county) will represent the individuals of a county without containing any extra information. The edges will be simple connections between individuals, generated according to the chosen model.

The two parameters required for the construction of the network are received from the corresponding node in the outer network and they are:

- the number of nodes in the network and
- the probability that any two nodes are connected.

Seeing clusters as crowds of individuals makes a logical connection with the population density which is known from the very beginning.

the number of nodes in the network, in an ideal programming environment, could have been taken raw as the size of the population in the current county. Unfortunately, the designed implementation does not permit such a vast memory usage; thus this parameter should be scaled as well according to the maximum size allowed and the maximum size of population among all counties. Applying this theory, an individual will actually represent  $n$  individuals which could cause problems during simulations when scaling the number of already infected individuals, for example: there is one infected individual, but following the scaling rule, one real individual is 0.2 of one individual in the used network, therefore there will be 0.2 individuals infected.

Due to the memory limitation, each inner network will be split in  $m$  smaller inner network which can be considered to be independent communities or clusters of individuals inside a county. The number of communities in a county will be equivalent with the population density in that county.

## 5. SIMULATIONS

An epidemic is characterized firstly by the way the virus can be contacted; the easier the virus is contacted, the more factors encourage the epidemics.

In the case of the virus A/ H1N1, the spreading is done by air, a common and successful medium for an epidemic to pass to a pandemic spreading. The influence of the continuous movement of masses of population is sustained by the multiple means of transportation between communities of individuals and, moreover, by the increase of the crowding coefficient that traveling with most of those means presumes. The considered situations were explained in the previous chapter by describing the way they were modeled and selected from the multitude existing.

The types of edges are the form chosen for codifying the way individuals change their current node location to another, carrying along the virus from one infected community to another one still healthy. Among the percentage of individuals that move from one node to another there exists the possibility of existing individuals that are infected, and as the virus is transmitted by air, any short or long contact of that individual can add a new victim to the general statistics. However, this theory is not totally real, depending actually on the particular characteristics of the individual, e.g. the power of its immunity system.

The state of the weather is another positive factor of influence for the spreading. In the cold months when it is often raining, snowing or wind blowing, the human immunity system fails to keep the same properties as it used to in the warmer months, so the probability of viral infection is increased for the majority of the population.

One factor that independently rises from the context is the apparition of a vaccine against this virus. The factors of influence remain valid for all the cases, but the number of susceptible individuals to the disease decreases drastically as an enormous part of the population of Romania has taken this vaccine. Consequently, in the final months of the epidemic, the spreading has reached the lowers level of activity and finally became inactive.

**5.1. Numerical experiments.** Numerical experiments are performed in Romania, over a hierarchical network with two layers corresponding to the country level (this network has 41 nodes corresponding to the 41 counties) and the county level (for each of the 41 counties, the network has a variable number of nodes according to the population size of each of them).

The results are simulated over 9 months, between May 2009 (the starting months which is not taken into consideration for simulations) and February 2010.

From the input set it is generated, for each test of the application (each input month), another set of data, used in the following computations. Besides the data gathered from different sources, there are other parameters which

influence the spreading simulations, parameters which are established from the very start as constants for the entire application.

**5.2. Data.** The collected data includes the monthly situation of newly infected population size of each county from Romania and of the countries included in the network with which our country has stronger connections.

Tables 1 and 2 present the data concerning each node of the outer network during the studied period of time: May 2009 – February 2010.

The data organized in these tables was connected from the Romanian Health Minister official site, the section of press communicates [7].

Another set of data which remains constant during all the executions are the characteristics of an internal node in the outer network. These data consist of the population size and population density of each county.

Table 2 does not contain population information about the countries with which Romania has connections as the size and density of population corresponding to their country does not arise any interest as the simulation is not performed on those nodes, it are only used for helping the simulation inside Romania.

**5.3. Results of the simulations.** We present the results of the simulations for two of the nine months: August and October 2009. Simulations are performed in an identical manner for the other remaining 7 months.

*Experiment 1 – Month August*

The first experiment has as purpose to simulate the evolution of the epidemic over the counties of Romania during the month August 2009. The month which will be given as input will be *July*. The data for input, besides the one given by the user, is read from the local database and will be used for the construction of the network.

The construction of the network will proceed as follows: the nodes – counties of Romania and some countries – will be taken with the attached information: the name, population size and density (the data from the Table 2 will be loaded). To the loaded nodes there will be added some extra information: the situation for the month given as input: the infected population size of each node, the corresponding column for the month July 2009 from the Tables 1 and 2; after the nodes are set in the network, the connections are loaded.

At this moment, the network is loaded into memory and it is ready for starting the simulations over it. However, the simulation influence factors must be taken into account: the “temporal constants”. The month August is known to be in Romania as the month when most of the people go in a holiday, consequently there will be slight modifications on the computations.

TABLE 1. Infection size on nodes - counties

County	Infection population size									
	May 2009	June 2009	July 2009	Aug. 2009	Sept. 2009	Oct. 2009	Nov. 2009	Dec. 2009	Jan. 2010	Feb. 2010
Alba	0	0	0	0	0	0	22	44	13	0
Arad	0	0	0	0	0	0	80	50	10	0
Arges	0	0	0	1	1	0	29	54	23	1
Bacau	0	0	0	2	0	0	130	111	29	1
Bihor	0	0	0	0	0	1	10	4	2	0
Bistrita- Nasaud	0	0	0	0	0	0	0	1	3	3
Botosani	0	0	0	0	0	0	269	133	111	0
Brasov	0	0	27	9	3	1	29	43	14	4
Braila	0	0	0	6	0	0	2	11	14	1
Buzau	0	0	0	0	0	0	25	121	5	0
Caras- Severin	0	0	0	0	0	0	19	63	14	1
Calarasi	0	0	1	2	0	0	3	0	5	0
Cluj	0	0	0	6	1	3	58	84	45	1
Constanta	0	0	5	2	1	0	21	55	48	4
Covasna	0	0	0	0	0	0	5	22	11	4
Dambovita	0	0	0	2	0	0	94	107	52	1
Dolj	0	0	8	0	0	1	148	98	8	0
Galati	0	0	1	4	0	0	85	33	22	1
Giurgiu	0	0	0	0	0	0	20	16	8	0
Gorj	0	0	0	0	0	0	0	0	0	0
Harghita	0	0	0	0	0	0	52	39	23	0
Hunedoara	0	0	0	8	0	0	94	40	27	0
Ialomita	0	0	0	0	4	0	16	22	5	0
Iasi	0	7	8	6	8	40	171	58	39	3
Ilfov	5	15	62	63	9	12	725	612	260	15
Maramures	0	0	0	0	0	0	18	19	3	0
Mehedinti	0	0	2	1	1	0	11	12	1	0
Mures	0	0	7	7	0	1	27	66	56	3
Neamt	0	0	0	0	0	0	56	33	24	1
Olt	0	0	0	0	0	0	10	24	40	0
Prahova	0	0	1	4	0	41	51	83	72	0
Satu Mare	0	0	0	0	0	0	0	0	3	1
Salaj	0	0	0	0	0	0	3	3	1	0
Sibiu	0	0	1	3	0	1	24	79	16	0
Suceava	0	0	0	2	0	0	10	70	30	1
Teleorman	0	0	0	3	1	0	10	10	3	0
Timis	0	5	2	8	0	0	23	62	38	3
Tulcea	0	0	0	0	0	0	6	23	5	1
Vaslui	0	0	0	0	0	0	45	27	4	0
Vâlcea	0	0	0	3	3	0	22	11	6	0
Vrancea	0	0	0	0	0	0	62	43	4	0

TABLE 2. Infection size on nodes - connected countries

Country	Infection population size									
	May 2009	June 2009	July 2009	Aug. 2009	Sept. 2009	Oct. 2009	Nov. 2009	Dec. 2009	Jan. 2010	Feb. 2010
Ukraine	0	1	1	0	6250	850000	11005	1230	301	22
Hungary	0	7	11	138	1250	1877	1107	203	70	3
Bulgaria	0	5	10	47	470	100000	2307	967	111	25
SUA	2254	20000	33902	6700	3200	1050	320	115	67	5
Canada	280	5438	7983	2060	986	320	98	20	0	0
UK	40	1540	7447	5957	8960	17325	6015	2000	200	0
Spain	93	430	760	838	2600	17303	1230	700	183	7
Mexico	1626	4957	10262	2350	1739	600	121	67	21	0
France	12	171	300	825	3024	7017	659	226	105	0
Turkey	0	27	40	50	180	625	303	29	15	0
Greece	0	58	109	1340	2506	2030	270	37	25	0
Germany	11	291	470	12320	1445	4445	750	217	32	12
Italy	9	86	130	1138	7213	21207	3070	375	93	31
Portugal	1	6	27	1960	1530	1248	625	123	75	1
Netherlands	3	100	134	1368	1020	2364	950	99	65	5

After the application has been executed, the values of the simulation will be as displayed in Figure 1 (there are displayed also the real results for a more obvious comparison).

The error expresses the number of extra or missing cases from the simulated results, compared with the real results. The value of 2.07 is an acceptable one as for the entire country there have been a mistake of this size which reported to the population of the country is meaningless.

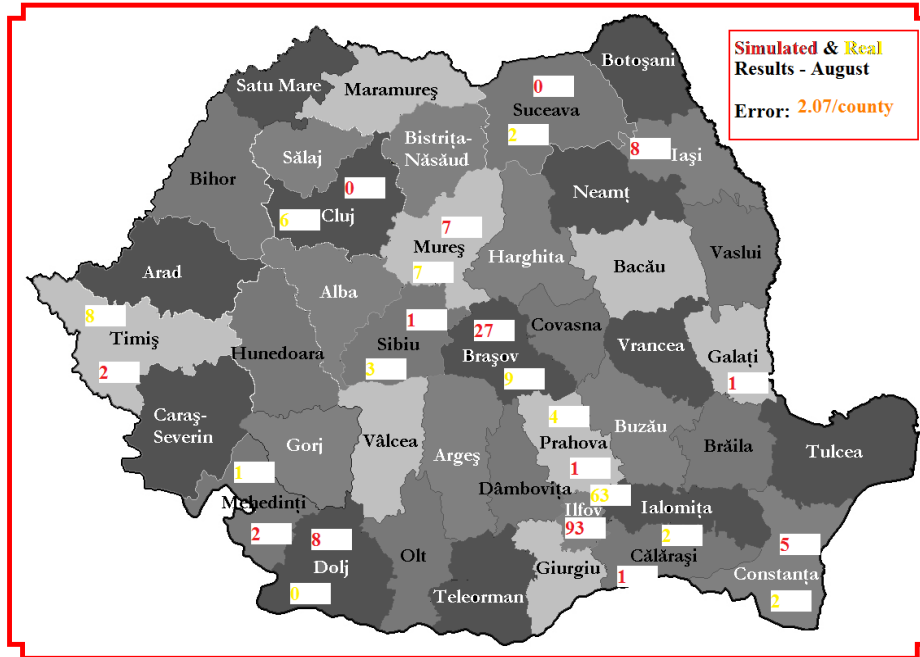
The arguments for the appearance of this error have a vast range of foundering reasons; however the simplest one is that the evolution of a virus on a real network is rather unpredictable. Still, one of the explanations is that during the month of August it is possible that the theory of moving is not sustained as people can make their holidays abroad, not only in the country.

#### *Experiment 2*

The second experiment has as purpose to simulate the evolution of the epidemic over the counties of Romania during the month October 2009. The month which will be given as input will be *September*. The data for input, besides the one given by the user, is read from the local database and will be used for the construction of the network.

The network is constructed in a similar way to the one explained for the previous experiment, except for the fact that the size of the infected population

FIGURE 1. Results for the month August.



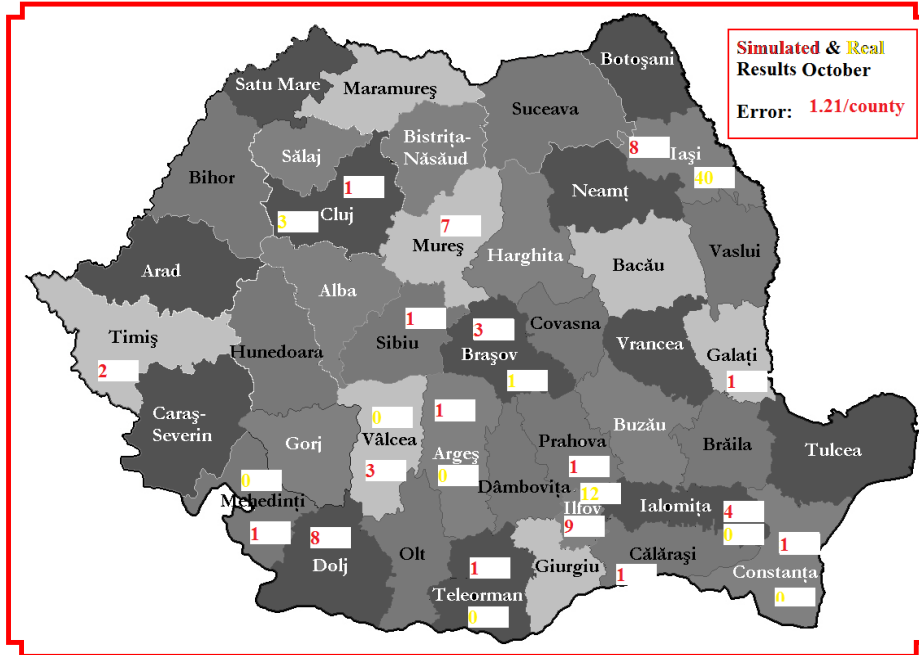
is taken from the column corresponding to the month September 2009 of the Tables 1 and 2.

The month October imposes some other limitations over the computations for the epidemic simulation. This month is known to be the month when the collegial year begins. The map for the month of October 2009 is depicted in Figure 2.

For this experiment there has been obtained a better results in means of distributing the bad-placed cases: 1.21 cases/county. In comparison with the country population, the difference is hardly noticeable.

One of the most relevant reasons for having these errors is the fact that October is an autumn month when the temperatures decrease and the infection likelihood increases, but slowly than in the succeeding months. In this way one of the draw-backs of the algorithm is that the “temporal constants” range are far too restrictive. At the moment a month has or not a property, it can not have it in a certain proportion, as it might be case for the month October.

FIGURE 2. Results for the month October.



## 6. CONCLUSIONS

The paper proposes a new approach for analyzing epidemic spreading over social networks by introducing a new model tested against real-world results. The model is based on intensive research in social networks and epidemic spreading, viewed from different aspects: the mathematical way and the sociologic-statistical way. The selected data for taking part in the model has been restricted to a number of characteristics and support further extension.

The developed model is general and can be applied to any hierarchical structure similar to the one of Romania and it is valid for the simulation of the spreading of any other virus. The simulation algorithm can support modifications to fit any other epidemiological model. The application, although it is presented as a case of study, can be modified to have a general character: it can suit any country of the world, only with the change of data from the database and of the characteristics deduced from the time of year, which rather seem to be specific to Romania.

**Acknowledgement.** The second author acknowledges the support from the CNCSIS grand IDEI 2412/2009.

## REFERENCES

- [1] J. Badham, R. Stocker, *The impact of network clustering and assortativity on epidemic behaviour*, Theoretical Population Biology, 77 (2010), pp. 71-75.
- [2] M.F. Boni, B.U. Manh, P.Q. Thai, J. Farrar, T.T Hien, N.T. Hien N. Van Kinh, P. Horby, *Modelling the progression of pandemic influenza A (H1N1) in Vietnam and the opportunities for reassortment with other influenza viruses*, BMC Medicine, 7 (2009), pp. 43-47.
- [3] G. Caldarelli, A. Vespignani, *Large Scale Structure and Dynamics of Complex Networks from Information Technology and Natural Science*, World Scientific, London, 2007.
- [4] A. Degenne, M. Forsé, *Introducing social networks*, Sage, London, 1999.
- [5] M.J. Keeling, K.T.D. Eames, *Networks and epidemic models*, J. R. Soc Interface, 2, (2005), pp. 295-307.
- [6] V. Prejmerean, M. Frentiu, V. Cioban, and O. Ghiran, *Graphical representation of the pandemic spreading*, First International Conference on Complexity and Intelligence of the Artificial and Natural Complex Systems, Medical Applications of the Complex Systems, Biomedical Computing, Targu Mures, Romania, 2008, pp. 197-202.
- [7] Romanian Health Ministry, Press release, <http://www.ms.ro/?pag=62>
- [8] *Scale-free Networks*, [http://www.scholarpedia.org/article/Scale-free\\_networks](http://www.scholarpedia.org/article/Scale-free_networks)
- [9] B. Wellman, *Structural Analysis: From Method and Metaphor to Theory and Substance* Cambridge: Cambridge University Press, London, 1988.
- [10] S. Wasserman, K. Faust, *Social network analysis: methods and applications*, Cambridge University Press, 1994.
- [11] D.J.Watts, S.H. Strogatz, *Collective dynamics of 'small-world' networks*, Nature, 393 (1998), pp. 440-442

DEPARTMENT OF COMPUTER SCIENCE, BABES-BOLYAI UNIVERSITY, KOGALNICEANU  
1, 400084, CLUJ-NAPOCA, ROMANIA

*E-mail address:* [silvia.rausanu@gmail.com](mailto:silvia.rausanu@gmail.com), [cgrosan@cs.ubbcluj.ro](mailto:cgrosan@cs.ubbcluj.ro)