

## MODELING INFORMATION SPREADING ON COMPLEX NETWORKS

**Daniel Trpevski, Kire Stamenov, Ljupčo Kocarev**

**A b s t r a c t:** In this article we propose a model for the spread of two types of information in networks. The model is a natural generalization of the epidemic susceptible-infective-susceptible (SIS) model. The two information types have different attractiveness, which affects the nodes' decision on which information type to adopt when both arrive at a node in the same time step. At difference with results from other authors, the model shows simultaneous existence of the two information types in the stable state. We give approximations for the average number of nodes informed with each information type at the end of the spreading process when nodes have high degree.

**Key words:** complex networks; dynamical systems; nonlinear systems; information spreading

### INTRODUCTION

The investigation of social spreading phenomena such as propagation of rumors, the diffusion of fads, the adoption of technological innovations and the success of consumer products mediated by word-of-mouth, has a long tradition in sociology and economics. It has been postulated [1, 2] that the network of contacts between individuals affects the spreading process, and recently, with the development of the theory of complex networks, these effects are gradually unravelled. Indeed, many spreading processes in general have been addressed with the advent of complex network theory, such as virus propagation in social and computer networks [3, 4, 5, 6, 7, 8], the diffusion of innovations [9, 10], the occurrence of information cascades in social and economic sys-

tems [1, 12], disaster spreading in infrastructures [13], or information diffusion in a society through the word-of-mouth mechanism [14].

The most popular model for information or rumor spreading [15] points out the analogy between information spreading and epidemic spreading described by the susceptible-infective-recovered (SIR) model. Agents in the information spreading model are divided in three classes: ignorants, spreaders and stiflers, which correspond to susceptible, infective and recovered individuals, respectively. Epidemiological models have been used to describe information spreading ever since, a prominent example being modeling topic flow through blogspace using the SIR model [16], as well as describing word-of-mouth in product marketing [14] with the SI model. We note that other widely used models for describing collective social behaviour are the threshold models, first proposed by Granovetter [17]. However, we do not treat threshold models in this work. The effects of network topology for threshold models have been analyzed elsewhere [11, 12].

All of the previously mentioned models treat the case where a single agent spreads in a network. In the context of complex network research, the spread of two types of information or epidemic contagions in a network has only recently been considered [18, 19, 20, 21, 22, 23, 24, 25]. The earliest study [18] treats two competing epidemics of the susceptible-infective-susceptible (SIS) type, one of which annihilates the other upon contact with certain probability, i.e. acts as an immunizing agent, [19] treats positive and negative word-of-mouth processes which are in essence two susceptible-infected (SI) processes, with the constraint that the negative word-of-mouth can spread only two hops away from its source. A generalization of the SIS model is proposed in [20] (and extended in [21]) where one disease has priority in the spreading process. B. Karrer and M. E. J. Newman [22], on the other hand, study two competing epidemics of SIR type by mapping the problem to a bond percolation. The model in our work is conceptually the same with that in [23], where two competing epidemics of SIS type are investigated with a continuous-time dynamical model. Also, the probability that a node receives an information type from its neighbours is approximated using the Weierstrass product inequality. In such a setting the authors prove that only one of the epidemics will persist in a network. They later generalize their model to allow coexistence of the two epidemics [24].

The model we use is a natural extension of the susceptible-infective-susceptible (SIS) epidemiological model, where a node in the network can be in one of two states: the *susceptible* ( $S$ ) state, not having contracted the disease, or

in the *infective* ( $I$ ) state, able to spread the disease to each of its neighbours. The infective nodes recover, becoming susceptible to the disease again. In this model each node can be in one of three states: susceptible ( $S$ ) to an information type, and "infective" with two different types of information. Nodes that have adopted information type 1 are said to be in state  $I_1$  and nodes that have adopted information type 2 are said to be in state  $I_2$ . The infective nodes can spread the information they are infected with to their neighbours and can lose interest in their adopted information type, reverting back to the susceptible state. The two information types compete for the nodes, meaning that a node can adopt only one information type at a specific time. The main difference in our model from the one in [23] is that we propose a discrete-time version, and do not use an approximation for the probability of receiving the information types. We obtain that both information types can persist in the network.

The questions of interest for this information spreading model are similar to those in epidemic spreading. In this paper we address two specific questions:

- how many nodes will eventually be reached by each information type, and,
- is there an "epidemic threshold" for the rate of spreading, separating a regime in which the information types remain confined to a small number of nodes from one where they affect a finite fraction of nodes in the network.

A large amount of studies has addressed these same questions for epidemic spreading from a complex network perspective. Using percolation theory ideas and generating function methods, [3] give exact analytical results for the epidemic threshold, outbreak size, and other relevant quantities for the SIR model. The results represent average values over an ensemble of random graphs with an arbitrary degree distribution. In [26, 27], contrary to previous findings [4, 5], the presence of an epidemic threshold has been established for the SIS model and infinitely large networks with a power-law degree distribution. Rather than determining the epidemic threshold for a whole class of networks with a given degree distribution, [7] and [8] propose its calculation for a specific network given with an adjacency matrix, for a SIS and SIR model, respectively. We follow this idea to determine the information spreading threshold in our model.

The paper proceeds as follows. Section 2 gives the model definition and stability analysis by means of which the "epidemic threshold" is deter-

mined. In section 3 we describe the behaviour of the model on regular network topologies. Results on the number of infective nodes from this section are used as approximations for the number of infective nodes in complex network topologies in section 4. The last section concludes the paper and points out potential research directions.

## 2. INFORMATION SPREADING MODEL

### 2.1. Definition of the model

Consider a closed population of  $N$  individuals, whose network of contacts is represented by an undirected unweighted graph  $G = (V, E)$ , with node set  $V$  and link set  $E$ . Let  $A$  denote the adjacency matrix of the graph  $G$ , where  $a_{ij} = 1$  if  $(i, j) \in E$ , i.e. individuals (nodes)  $i$  and  $j$  are in contact with each other, and  $a_{ij} = 0$  otherwise. We propose a discrete stochastic model for information spreading among the nodes in such a network.

At time  $t$  each node  $i$  can be in one of three possible states:  $I_1$ ,  $I_2$  and  $S$ . States  $I_1$  and  $I_2$  signify that the node is a supporter of information type 1 or 2, respectively, and can spread the information to its neighbours, and  $S$  is an undecided or neutral state of the node in relation to the information types circulating in the network. States  $I_1$  and  $I_2$  are analogous to the infective state  $I$  in the SIS model, and state  $S$  is the counterpart of the susceptible state. The state of the node is represented by a state vector containing a single 1 in the component representing the present state and 0 everywhere else,

$$\mathbf{s}_i(t) = [s_i^{I_1}(t) s_i^{I_2}(t) s_i^S(t)]^T,$$

for all  $i \in 1, \dots, N$ . The vector

$$\mathbf{p}_i(t) = [p_i^{I_1}(t) p_i^{I_2}(t) p_i^S(t)]^T$$

is the probability mass function of node  $i$  at time  $t$ , and it states the probability for node  $i$  to be in each of the possible states at time  $t$ .

The model is constructed as follows. First, we delineate all the disjunct events that can occur to node  $i$  in a single time unit when it interacts with its neighbours in Table 1. The events exhaust the space of all possible events of

receiving information from neighbouring nodes, meaning that they have probabilities that sum to 1. In Table 1  $f_i(t)^{I_1}$  and  $f_i(t)^{I_2}$  are given with

$$f_i^{I_1}(t) = 1 - \prod_{j=1}^N [1 - \beta_1 a_{ij} s_j^{I_1}(t)]$$

$$f_i^{I_2}(t) = 1 - \prod_{j=1}^N [1 - \beta_2 a_{ij} s_j^{I_2}(t)]$$

and represent the probability that node  $i$  receives information type 1 or information type 2, respectively, from any combination of its infective neighbours at time unit  $t$ .  $\beta_1, \beta_2 \in [0, 1]$  are parameters which give the probability that an infective node transmits the information upon contact with a neighbour. The assumption that all transmissions of information among nodes are independent is used in the formulation of  $f_i^{I_1}(t)$  and  $f_i^{I_2}(t)$ .

Table 1

*Events that can occur to node  $i$  in a single time unit when interacting with neighbours*

Event	Probability of event
1. Node $i$ does not receive any information type from neighbours.	$(1 - f_i^{I_1}(t))(1 - f_i^{I_2}(t))$
2. Node $i$ receives information type 1 from neighbours, but not information type 2.	$f_i^{I_1}(t)(1 - f_i^{I_2}(t))$
3. Node $i$ receives information type 2 from neighbours, but not information type 1.	$f_i^{I_2}(t)(1 - f_i^{I_1}(t))$
4. Node $i$ receives both information type 1 and type 2 from neighbours.	$f_i^{I_1}(t)f_i^{I_2}(t)$

Next, we assume that receiving information from neighbours is only effective when a node is still in the susceptible state, i.e. it has not adopted any information type. The probability of receiving a single information type is given by events 2 and 3 in Table 1. When a node receives both information

types simultaneously, with probability given by event 4, it will adopt them according to their attractiveness. The attractiveness of the information types is given with the parameters  $a_1, a_2 \in [0,1]$ . If a susceptible node does not receive information from any neighbour, it stays susceptible. This is given with event 1 in Table 1.

Lastly, a node forgets information type 1 or 2 with probability  $\delta_1$  or  $\delta_2$ , respectively. We assume that the parameters are the same for all nodes, a situation which is not true in reality.

Having described that, the equations describing the dynamics of each node, i.e., the evolution of the model, are

$$\begin{aligned} p_i^{I_1}(t+1) &= \left[ f_i^{I_1}(t)(1-f_i^{I_2}(t)) + a_1 f_i^{I_1}(t) f_i^{I_2}(t) \right] s_i^S(t) + (1-\delta_1) s_i^{I_1}(t) \\ p_i^{I_2}(t+1) &= \left[ f_i^{I_2}(t)(1-f_i^{I_1}(t)) + a_2 f_i^{I_1}(t) f_i^{I_2}(t) \right] s_i^S(t) + (1-\delta_2) s_i^{I_2}(t) \\ p_i^S(t+1) &= (1-f_i^{I_1}(t))(1-f_i^{I_2}(t)) s_i^S(t) + \delta_1 s_i^{I_1}(t) + \delta_2 s_i^{I_2}(t) \\ \mathbf{s}_i(t+1) &= \text{Realize}[\mathbf{p}_i(t+1)] \end{aligned} \quad (2)$$

where  $\text{Realize}[\cdot]$  performs a random realization from the probability distribution given with  $\mathbf{p}_i^T(k+1)$ . The parameters  $a_1$  and  $a_2$  which describe the attractiveness of each information type satisfy the constraint

$$a_1 + a_2 = 1,$$

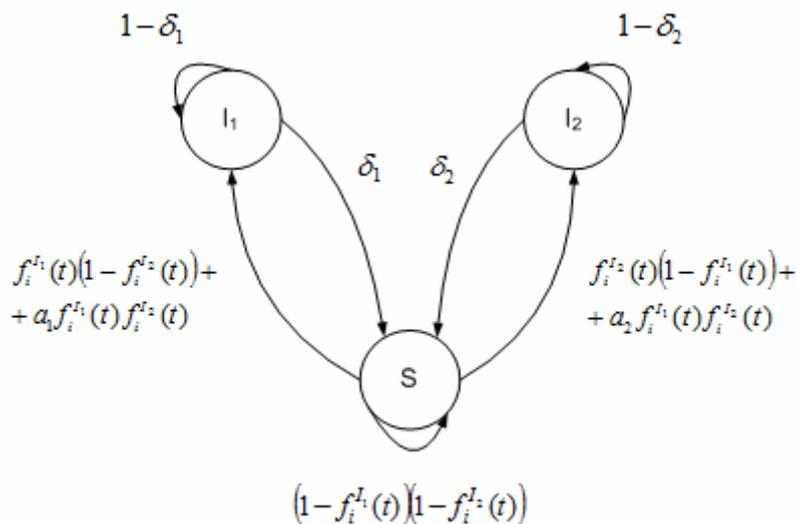
which means that a node divides its opinion among a pool of information types. For a single node, the model equations constitute an inhomogeneous Markov chain, given on Fig. 1.

Let

$$X(t) = \sum_{i=1}^N s_i^{I_1}(t), Y(t) = \sum_{i=1}^N s_i^{I_2}(t), Z(t) = \sum_{i=1}^N s_i^S(t) \quad (3)$$

be the total number of nodes in states  $I_1$ ,  $I_2$  and  $S$  at time  $t$ , respectively. Further, let  $N_{I_1}(t) = \mathbb{E}[X(t)]$ ,  $N_{I_2}(t) = \mathbb{E}[Y(t)]$  and  $N_S(t) = \mathbb{E}[Z(t)]$ . We are interested in the average number of nodes that eventually (when  $t \rightarrow \infty$ ) adopt

information types  $I_1$  and  $I_2$ ,  $N_{I_1} = N_{I_1}(\infty)$  and  $N_{I_2} = N_{I_2}(\infty)$ , compared to the total number of nodes  $N$  in the network.



**Figure 1.** Diagram of the Markov chain which describes the dynamics of a single node  $i$  in the information spreading model

The components of the state vector  $\mathbf{s}_i(t)$  which appear in Eq. (2) are random variables. In order to facilitate the mathematical analysis we shall analyze the evolution of the expected values of these quantities in the remainder of the paper. The random variables  $s_i^{I_1}(t)$ ,  $s_i^{I_2}(t)$  and  $s_i^S(t)$  can be regarded as Bernoulli random variables, and hence their expected values are  $\mathbb{E}[s_i^{I_1}(t)] = p_i^{I_1}(t)$ ,  $\mathbb{E}[s_i^{I_2}(t)] = p_i^{I_2}(t)$  and  $\mathbb{E}[s_i^S(t)] = p_i^S(t)$ . This allows us to rewrite the model given with Eqs. (2) and (1) as

$$\begin{aligned}
 p_i^{I_1}(t+1) &= \left[ f_i^{I_1}(t)(1 - f_i^{I_2}(t)) + a_1 f_i^{I_1}(t) f_i^{I_2}(t) \right] p_i^S(t) + (1 - \delta_1) p_i^{I_1}(t) \\
 p_i^{I_2}(t+1) &= \left[ f_i^{I_2}(t)(1 - f_i^{I_1}(t)) + a_2 f_i^{I_1}(t) f_i^{I_2}(t) \right] p_i^S(t) + (1 - \delta_2) p_i^{I_2}(t) \\
 p_i^S(t+1) &= (1 - f_i^{I_1}(t))(1 - f_i^{I_2}(t)) p_i^S(t) + \delta_1 p_i^{I_1}(t) + \delta_2 p_i^{I_2}(t)
 \end{aligned} \quad (4)$$

and  $f_i^{I_1}$  and  $f_i^{I_2}$  as

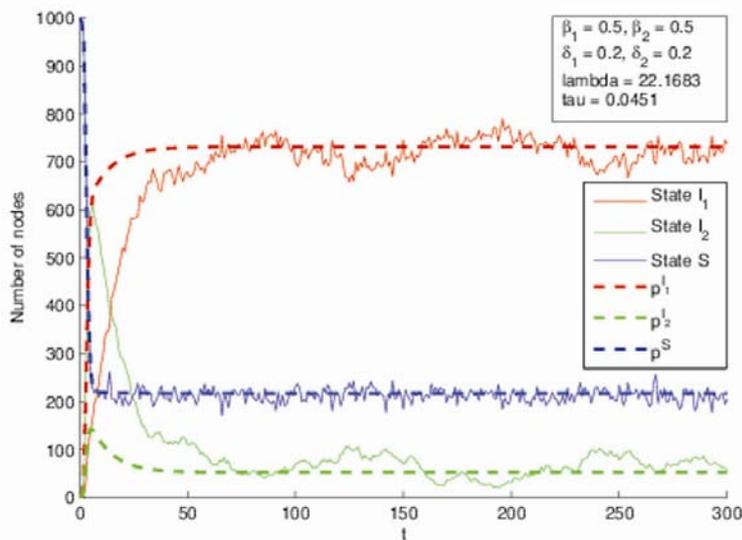
$$f_i^{I_1}(t) = 1 - \prod_{j=1}^N [1 - \beta_1 a_{ij} p_j^{I_1}(t)]$$

$$f_i^{I_2}(t) = 1 - \prod_{j=1}^N [1 - \beta_2 a_{ij} p_j^{I_2}(t)]$$
(5)

Equivalently,  $N_{I_1}(t)$ ,  $N_{I_2}(t)$ ,  $N_S(t)$  can be found using Eq. (4) as

$$N_{I_1}(t) = \sum_{i=1}^N p_i^{I_1}(t), \quad N_{I_2}(t) = \sum_{i=1}^N p_i^{I_2}(t), \quad N_S(t) = \sum_{i=1}^N p_i^S(t)$$
(6)

As a further illustration that the system given with Eqs. (4) and (5) represents the dynamics of the expected-value quantities of the model in Eqs. (2) and (1) consider Fig. 2. As can be seen, the evolution of the summated probability vector according to Eq. (4) corresponds to the evolution of the number of nodes in each state as predicted by Eq. (2). Thus, system (4), at least in the stationary state, describes the probability vectors from which random realizations are made in Eq. (2).



**Figure 2.** Evolution of Eqs. (2) and (4) on a 100-node Barabási-Albert network (see section 4) generated from a fully connected seed with 5 nodes and  $m = 2$ . Initially, node 84 is a supporter of information type 1, and node 415 is a supporter of information type 2. The solid lines show the number of nodes in each state as time progresses, as given by Eq. (3), and the dashed lines show the evolution of the average number of nodes in each state, as given by Eq. (6)

## 2.2. Dynamical systems approach

In this part we apply a dynamical systems approach for the stability analysis of our model. Let us in Eq. (4) replace the probabilities for node  $i$  to be in states  $I_1$  and  $I_2$  with  $x_i(t) = p_i^{I_1}(t)$  and  $y_i(t) = p_i^{I_2}(t)$ , respectively. In these terms, the evolution of the model can be written as

$$\begin{aligned} x_i(t+1) &= [1 - x_i(t) - y_i(t)] \left[ f_i^{I_1}(t)(1 - f_i^{I_2}(t)) + a_1 f_i^{I_1}(t) f_i^{I_2}(t) \right] + (1 - \delta_1) x_i(t) \\ y_i(t+1) &= [1 - x_i(t) - y_i(t)] \left[ f_i^{I_2}(t)(1 - f_i^{I_1}(t)) + a_2 f_i^{I_1}(t) f_i^{I_2}(t) \right] + (1 - \delta_2) y_i(t) \end{aligned} \quad (7)$$

where

$$\begin{aligned} f_i^{I_1}(t) &= 1 - \prod_{j=1}^N [1 - \beta_1 a_{ij} x_j(t)] \\ f_i^{I_2}(t) &= 1 - \prod_{j=1}^N [1 - \beta_2 a_{ij} y_j(t)]. \end{aligned} \quad (8)$$

One needs to rewrite the model in (4) and (5) in such a way since one of the probability equations is dependent on the other two, and would give linearly dependent rows in the Jacobian matrix.

Equation (7) represents a dynamical system  $F : [0,1]^{2N} \rightarrow [0,1]^{2N}$ . To deduce that this dynamical system has only one globally stable fixed point, we use the description of the nodes' dynamics as time-inhomogeneous Markov chains. Indeed, when a node's chain is weakly ergodic, the node will have a stationary distribution of the probabilities  $p_i^{I_1}(t)$ ,  $p_i^{I_2}(t)$  and  $p_i^S(t)$  to a unique globally stable fixed point for each node in the dynamical system (4), or equivalently (7). We use the conditions for weak ergodicity of time-inhomogeneous Markov chains given by Wolfowitz [28], which in this case basically mean that the graph of the Markov chain describing a node's dynamics is ergodic for all time (Fig. 1), and this translates to either

$$\beta_1 \neq 0, \beta_2 \neq 0, \delta_1 \neq 0, \delta_2 \neq 0, \beta_1 \neq 1, \beta_2 \neq 1 \quad (9)$$

or

$$\beta_1 \neq 0, \beta_2 \neq 0, \delta_1 \neq 0, \delta_2 \neq 0, \delta_1 \neq 1, \delta_2 \neq 1. \quad (10)$$

The dynamical system (7) has a fixed point at  $(x_i, y_i) = (0, 0)$  for all  $i$ . The local stability of this fixed point can be analyzed using the Jacobian matrix of the system (7) evaluated at the fixed point

$$DF|_{(0,0)} = \begin{bmatrix} A_{\beta_1} & 0_N \\ 0_N & A_{\beta_2} \end{bmatrix},$$

where  $A_{\beta_1} = \beta_1 A + (1 - \delta_1)I$ ,  $A_{\beta_2} = \beta_2 A + (1 - \delta_2)I$  and  $0_N$  (the matrix of all 0 elements) are  $N \times N$  matrices. Hence the fixed point  $(x_i, y_i) = (0, 0)$  is stable when

$$\max\{\beta_1 \lambda + 1 - \delta_1, \beta_2 \lambda + 1 - \delta_2\} < 1, \quad (11)$$

where  $\lambda$  is the largest eigenvalue of the adjacency matrix. Whenever this condition is fulfilled, no information type will eventually persist in the network, since the system will stabilize to a state where all nodes have probability 1 to be in the susceptible state, and from the aforementioned, this state is the only globally stable state.

Restating condition (11) as

$$\frac{\beta_1}{\delta_1} < \frac{1}{\lambda}, \quad \frac{\beta_2}{\delta_2} < \frac{1}{\lambda}, \quad (12)$$

one can see that the value of

$$\tau = \frac{1}{\lambda}$$

appears as a threshold value for the ratios of information transmission to forgetting  $\beta_1 / \delta_1$  and  $\beta_2 / \delta_2$ . When these ratios are smaller than the network dependent threshold, no information spread will occur in the network. Conversely, when any of them surpasses the threshold, the fixed point  $(x_i, y_i) = (0, 0)$  is unstable, and there will be information spreading in the network. For both of the information types to be able to spread in the network, both transmission to

forgetting ratios need to be larger than the network threshold. Thus, we recover the classical result of the existence of a network specific threshold for the spreading process. It acts as a critical point of the system dynamics, separating the regime where there is no information of any type present in the network from the one where a finite fraction of nodes have adopted an information type.

Lastly, for the cases when (12) is not fulfilled, i.e. when there is information spreading in the network, the average number of nodes reached by each information type (6) can be found by determining the unique globally stable fixed point. This is precisely what we attempt to do in the following sections. Because of the nonlinearities in the dynamical system, we give an approximation to the value of the fixed point when a node's degree is very large. Furthermore, we consider only the case when

$$\frac{\beta_1}{\delta_1} > \frac{1}{\lambda}, \quad \frac{\beta_2}{\delta_2} > \frac{1}{\lambda},$$

since this is the case when both information types are able to spread in the network. The cases when only one information type has transmission to forgetting ratio above the threshold reduce to a SIS epidemic model, and for this model the question of determining bounds on the fixed point has been solved [29].

### 3. BEHAVIOR OF THE MODEL ON REGULAR NETWORK TOPOLOGIES

In this section we present results for the model behaviour on regular networks. The topologies considered are the star and fully connected network. In all numerical simulations there is one node in state  $I_1$  and one node in state  $I_2$  initially. These act as sources for both information types, but can change their state as time progresses. The condition (9) or (10) for the convergence of the model to a unique fixed point will hold for all simulations. In all cases we denote the fixed point with  $x_i$  and  $y_i$  for all  $i = 1, \dots, N$ .

#### 3.1. Star network

In this section results on the star topology are presented. Assume that node 1 is the hub of the star, and nodes  $i = 2, \dots, N$  are the leaves. Eq. (8) for the star becomes

$$f_1^{I_1}(t) = 1 - \prod_{j=2}^N [1 - \beta_1 x_j(t)]$$

$$f_1^{I_2}(t) = 1 - \prod_{j=2}^N [1 - \beta_2 y_j(t)]$$

$$f_i^{I_1}(t) = \beta_1 x_1(t)$$

$$f_i^{I_2}(t) = \beta_2 y_1(t)$$

For  $i = 2, \dots, N$ . In the limiting case when  $N$  is large, i.e. when the hub has a large degree,  $f_1^{I_1}(t)$  and  $f_1^{I_2}(t)$  tend to 1, after some time  $t > t_0$  when the information types will have spread in the network. The fixed point in this case for the hub of the star is

$$x_1 = \frac{\delta_2 a_1}{\delta_2 a_1 + \delta_1 a_2 + \delta_1 \delta_2} \quad (14)$$

$$y_1 = \frac{\delta_1 a_2}{\delta_2 a_1 + \delta_1 a_2 + \delta_1 \delta_2}$$

while for the leaves is

$$x_i = \frac{\delta_2 z_1}{\delta_2 z_1 + \delta_1 z_2 + \delta_1 \delta_2}$$

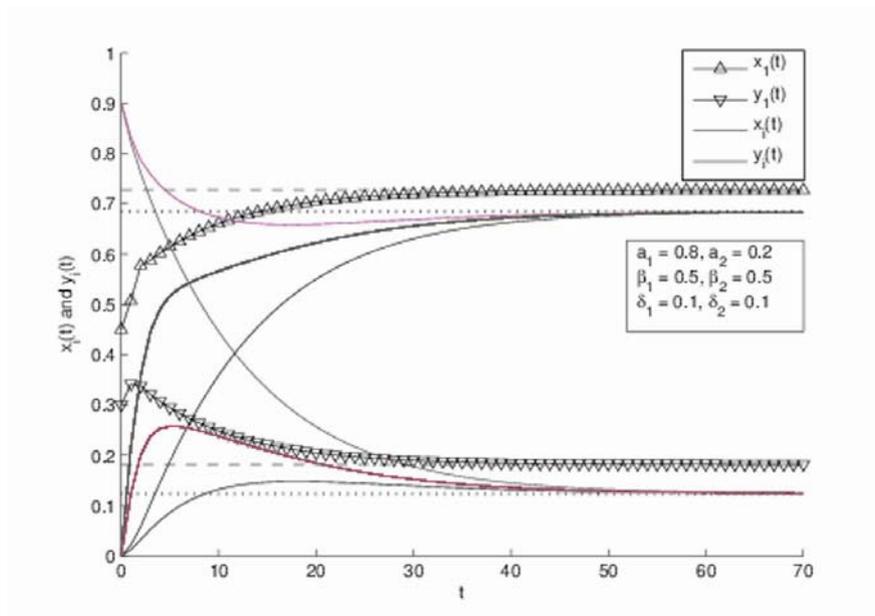
$$y_i = \frac{\delta_1 z_2}{\delta_2 z_1 + \delta_1 z_2 + \delta_1 \delta_2}$$

where

$$z_1 = \beta_1 x_1 (1 - \beta_2 y_1) + a_1 \beta_1 \beta_2 x_1 y_1$$

$$z_2 = \beta_2 y_1 (1 - \beta_1 x_1) + a_2 \beta_1 \beta_2 x_1 y_1$$

Figure 3 shows the results from the model evolution on a star with 1000 nodes. One can see that the approximations for the probabilities of infection with information types 1 and 2 for the star and leaves given with (14) and (15) predict the simulation values well when the hub has a large degree.



**Figure 3.** The evolution of (4) for a 1000-node star. Initially, one leaf is in state  $I_1$  and one leaf is in state  $I_2$ . Lines with triangles show the values for  $x_1(t)$  and  $y_1(t)$ . Dashed lines are the approximations (14) for the hub and dotted lines are the approximations (15) for the leaves

The behaviour of the hub of a star with  $N + 1$  nodes can serve as a simple model for what happens to an arbitrary node with degree  $N$  in an arbitrary network.

### 3.2. Fully connected network

Here we examine the behaviour of the model on fully connected networks. Equation (8) in this case becomes:

$$\begin{aligned} f_i^{I_1}(t) &= 1 - \prod_{j=1, j \neq i}^N [1 - \beta_1 x_j(t)] \\ f_i^{I_2}(t) &= 1 - \prod_{j=1, j \neq i}^N [1 - \beta_2 y_j(t)] \end{aligned} \quad (16)$$

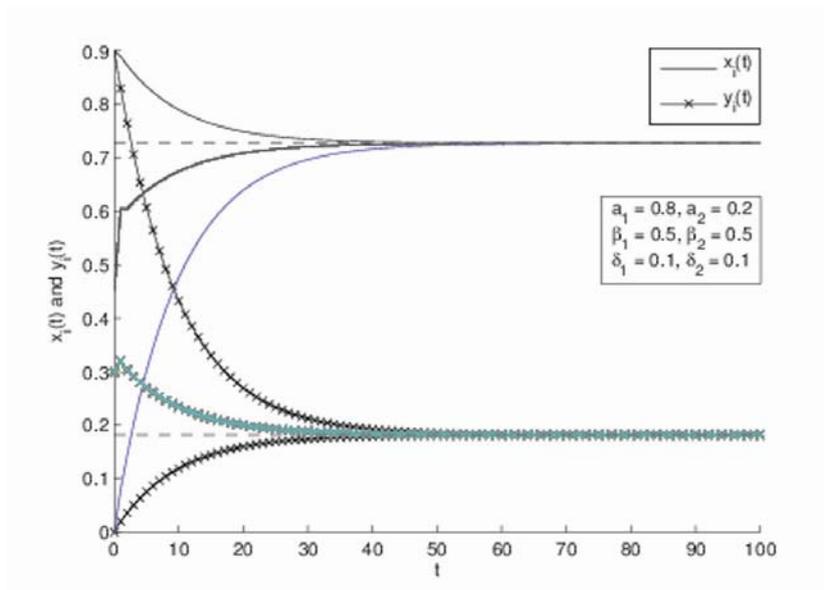
for all  $i$ . When  $N$  is large, each node will have many neighbours, and the product in Eq. (16) will tend to 0, i.e.

$$f_i^{I_1}(t) \rightarrow 1 \quad \text{and} \quad f_i^{I_2}(t) \rightarrow 1$$

after some time  $t > t_0$  when the information types will have spread in the network. This makes the fixed point for the fully connected graph

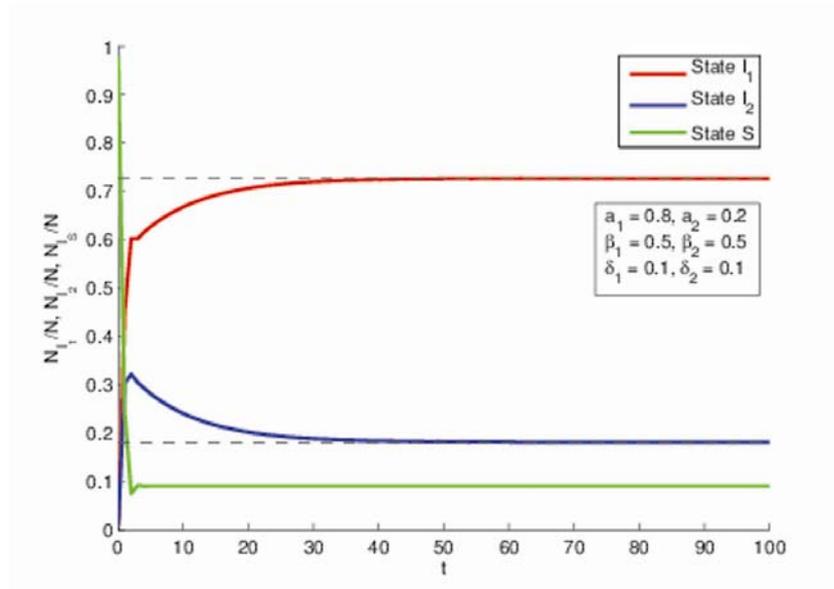
$$\begin{aligned} x_i &= \frac{\delta_2 a_1}{\delta_2 a_1 + \delta_1 a_2 + \delta_1 \delta_2} \\ y_i &= \frac{\delta_1 a_2}{\delta_2 a_1 + \delta_1 a_2 + \delta_1 \delta_2} \end{aligned} \quad (17)$$

for  $i=1, \dots, N$ . Figure 4 shows that (17) approximates the stable values for  $x_i(t)$  and  $y_i(t)$  well when the degree of each node is high enough. Also, note that because of the dense topology, the degree of the nodes does not have to be as high as in the hub of the star for the approximation to hold.



**Figure 4.** The evolution of (7) for a fully connected graph of 100 nodes. Initially, one node is in state  $I_1$  and one node is in state  $I_2$ . Dashed lines represent the approximations  $x_i$  and  $y_i$  given with (17)

Furthermore, the average number of nodes reached by information types  $I_1$  and  $I_2$  in this case after the system stabilizes can be calculated as  $N_{I_1} = Nx_i$  and  $N_{I_2} = Ny_i$ . Naturally, this is also a good approximation to the average number of nodes in the network as calculated by (6) (see Fig. 5).



**Figure 5.** The average number of nodes in each state  $N_{I_1}(t)$ ,  $N_{I_2}(t)$ ,  $N_S(t)$  compared to the total number of nodes in the network  $N$  for the same fully connected graph of 100 nodes as in Fig. 4. Initially, one node is in state  $I_1$  and one node is in state  $I_2$ . Dashed lines are the approximations for the fractions  $N_{I_1}/N$  and  $N_{I_2}/N$  as given by (17)

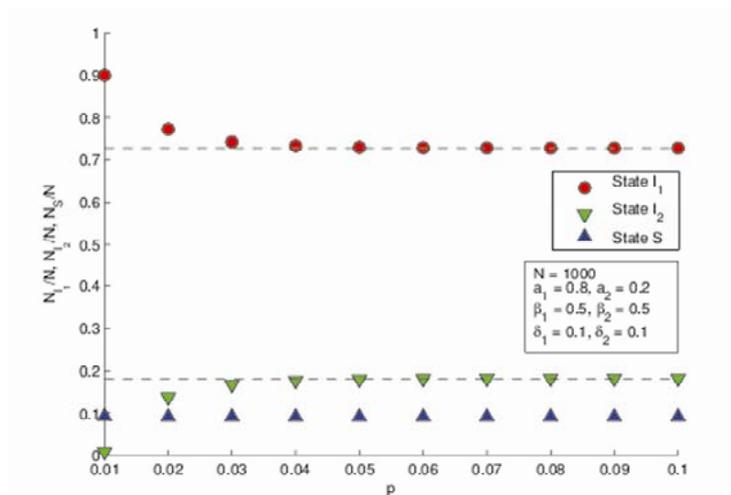
#### 4. BEHAVIOR OF THE MODEL ON COMPLEX NETWORK TOPOLOGIES

In this section we present results for the model behaviour on complex network topologies. All experiments start with one node in state  $I_1$  and one node in state  $I_2$ , which can change their state as time progresses. In particular, we shall see that the results on the regular network topologies can be related to the ones on complex network topologies when the nodes in the networks have high degree.

#### 4.1. Erdős-Rényi random networks

The model proposed by Erdős and Rényi (ER) [30, 31] describes graphs with  $N$  nodes in which every link exists with probability  $p$ . The degree distribution in these networks is Poisson, hence the homogeneous structure in the sense that all nodes have degree close to the average degree  $\langle k \rangle = p(N-1)$ . Also, in this model there is a critical probability value  $p_c = 1/N$  under which the resulting network consists of small disconnected components, and above which there is a giant component in the network containing  $O(N)$  nodes. All the networks are generated with  $p > p_c$  and the sources of information are randomly placed in the giant component. Figure 6 shows the steady-state behaviour of the model for ER networks with 1000 nodes for different values of  $p$ . When increasing  $p$ , the nodes have increasingly higher degree, and the probabilities of being in state  $I_1$  and  $I_2$  are well approximated by Eq. (14). Thus, the fraction of nodes in the network which have adopted each information type is accurately predicted by the approximation

$$\frac{N_{I_1}}{N} = x_i = \frac{\delta_2 a_1}{\delta_2 a_1 + \delta_1 a_2 + \delta_1 \delta_2}, \quad \frac{N_{I_2}}{N} = y_i = \frac{N \delta_1 a_2}{\delta_2 a_1 + \delta_1 a_2 + \delta_1 \delta_2}. \quad (18)$$

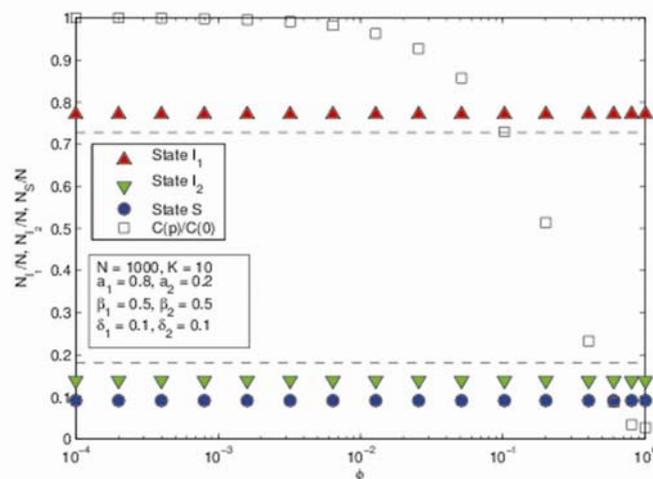


**Figure 6.** The steady state behaviour of the model for ER networks with  $N = 1000$  for different values of  $p$ , depicting the fraction of nodes in each state. For every value of  $p$  the results are averages of 100 network realizations. The model has been run for 250 time units. The dashed lines are the approximations of the fraction of nodes supporting each information type as given with Eq. (18)

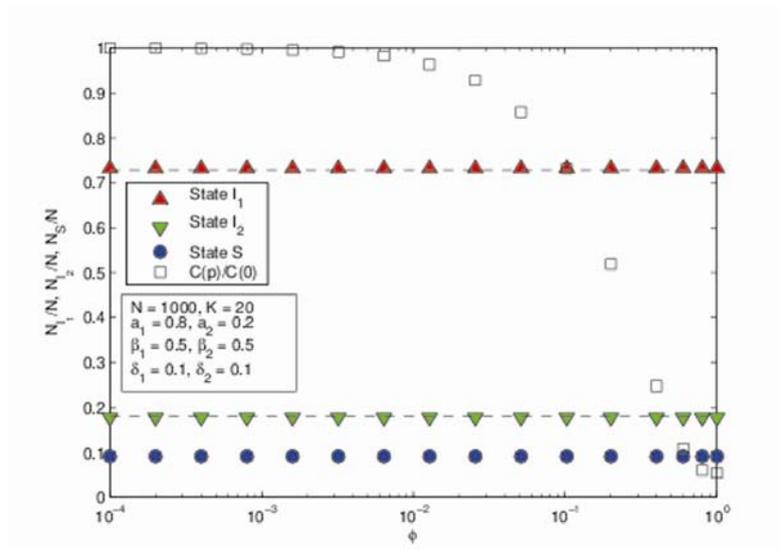
#### 4.2. Watts-Strogatz small-world networks

The Watts-Strogatz (WS) model [32] has been built to reproduce the property that many real-world networks, social networks among them, have very small average distance between nodes, on one hand, and a high clustering coefficient, on the other. The average distance between nodes in these networks is of the order of  $\log N$ , and such networks are said to exhibit the small-world property. However, the small-world characteristic of the WS-model usually implies both the small-world and the high clustering property. The algorithm for constructing a WS small-world network starts from a ring lattice where each node has  $2K$  neighbours,  $K$  in the clockwise and  $K$  in the anticlockwise direction. Each edge is rewired with probability  $\phi$ , not allowing self-loops or multiple edges between nodes. For the values of  $\phi$  that generate a small-world, a network with densely connected neighbourhoods, whose size is regulated by  $K$ , is created, and some of the otherwise distant neighbourhoods are connected by long-range rewired links.

Figures 7 and 8 depict the steady state behaviour of the model when the rewiring parameter  $\phi$  is varied. Results are given for networks generated with two different values of  $K$ .



**Figure 7.** The steady-state behaviour of the model for WS networks for different values of  $\phi$ . The fraction of nodes in each state is depicted, as well as the clustering coefficient for the networks. Results are obtained by averaging over 100 network realizations for each value of  $\phi$ . Networks have  $N = 1000$  nodes and were generated from a starting ring lattice with  $K = 10$ . The model has been run for  $t = 600$  time units. The dashed lines are the approximations of the fraction of nodes supporting each information type as given with Eq. (18)



**Figure 8.** The steady-state behaviour of the model for WS networks for different values of  $\phi$ . The fraction of nodes in each state is depicted, as well as the clustering coefficient for the networks. Results are obtained by averaging over 100 network realizations for each value of  $\phi$ . Networks have  $N = 1000$  nodes and were generated from a starting ring lattice with  $K = 20$ . The model has been run for  $t = 600$  time units. The dashed lines are the approximations of the fraction of nodes supporting each information type as given with Eq. (18)

Figure 8 shows that when nodes in the the starting lattice have degree which is high enough, the product terms in Eq. (8) tend to zero and the average number of nodes in each informed stae is again well approximated by Eq. (18). This does not happen in the case when nodes do not have high enough degree (Fig. 8). Also, the average clustering coefficient of each network, normalized by  $C(0)$  is shown on both figures, to indicate that the amount of clustering does not have an impact of the number of nodes reached by each information type, at difference with what we have observed for an alternate generalization of the SIS model [20].

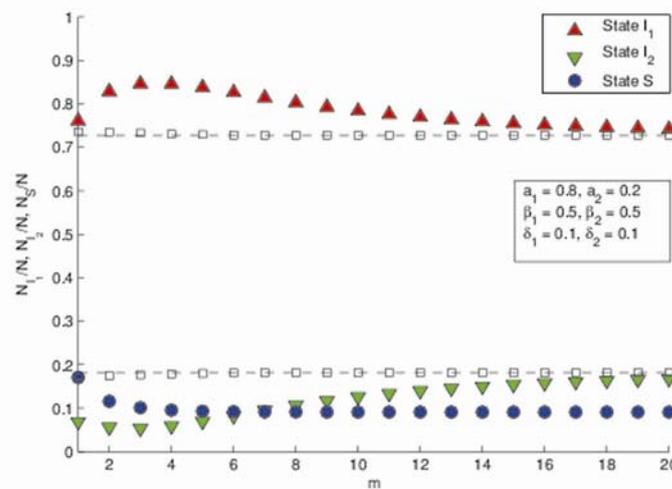
### 4.3. Barabási-Albert power-law networks

The Barabási-Albert (BA) model has been built to mimic another prevalent property of many real-world networks. Namely, it has been observed that real world networks have power-law degree distributions with an exponent

value that has usually lies between 2 and 3. Such degree distributions allow for a statistically significant probability for the existence of hubs, i.e. nodes that have unusually high degree. Networks with power-law degree distributions have been termed power-law networks.

Here we use the original BA algorithm to construct the networks [33]. One starts from a seed of  $m_0$  connected nodes and adds a new node with  $m \leq m_0$  links at each step. The nodes to which a new node is connected are chosen with probability proportional to their degree, and this rule of choosing nodes is known as preferential attachment. This yields a network with average degree  $\langle k \rangle = 2m$ , and degree distribution  $P_k \propto k^{-3}$ .

Figure 9 shows the average number of nodes in each state as a fraction of  $N$ , as the parameter  $m$  and consequently, the average degree  $\langle k \rangle$ , is increased. As  $m$  is increased the networks become denser and the nodes have increasingly higher average degree, so that the approximations (17) for the fraction of nodes reached by each information type hold. An interesting observation is that for low values of  $m$ , the BA topology appears to facilitate the spreading of the more attractive information.



**Figure 9.** The average number of nodes in each state as a fraction of  $N$  for BA networks when  $m$  is varied. For each value of  $m$ , results are averaged over 100 network realizations and the model has been run for  $t = 250$  time units. All networks are generated from a 20-node fully connected seed and have  $N = 1000$ . The dashed lines are the approximations of the fraction of nodes supporting each information type as given with Eq. (18), and squares indicate averages of the stable  $x_i(t)$  and  $y_i(t)$  values for the largest hub in the networks

## 5. CONCLUSIONS

In this paper we make an attempt to study how two different information types propagate and compete in a network. The model presented is a natural generalization of the SIS epidemic model. It describes the interactions among nodes, and general results for the interplay of the two information types on different topologies are given. The key points of this paper are as follows.

- We suggest a discrete stochastic version for the SIS model with two infective states for information spreading. We recover the classical result of an intrinsic network threshold  $1/\lambda$  for the spreading process to occur, where  $\lambda$  is the largest eigenvalue of the network's adjacency matrix. Furthermore, we find that in this model both information types can co-exist in the stable state, as opposed to what is reported in [23] for a continuous-time version of the same model with the Weierstrass product inequality approximating the probability of receiving an information type. The model has a unique stable fixed point, which implies irrelevance of the choice of initial information spreaders.
- We find that when a node  $i$  has high degree, the probability of receiving information of any type from any combination of its infective neighbours tends to 1, and the probabilities to adopt information type 1 or 2 are well approximated by

$$x_i = \frac{\delta_2 a_1}{\delta_2 a_1 + \delta_1 a_2 + \delta_1 \delta_2}$$

and

$$y_i = \frac{\delta_1 a_2}{\delta_2 a_1 + \delta_1 a_2 + \delta_1 \delta_2},$$

respectively.  $\delta_1$  and  $\delta_2$  are the rates of forgetting the information, and  $a_1$  and  $a_2$  denote the attractiveness of each information type. Thus, when the degree of the nodes in an arbitrary network is high enough, the average fraction of nodes adopting each information type is,  $N_{I_1} / N = x_i$  and  $N_{I_2} / N = y_i$ .

Future research directions are numerous. One can apply the methodology of [29] to obtain better bounds on the probabilities of infection for the case when the degree of the nodes is not very high for the aforementioned approximations to hold. Comparing the model predictions to real data is a key question to its usefulness. Also, a generalization for an arbitrary number of information types is in order.

## REFERENCES

- [1] Mark S. Granovetter: "The Strength of Weak Ties", *American Journal of Sociology*, Vol. **78**, No. 6, pp. 1360–1380 (May, 1973).
- [2] Mark. S. Granovetter: "The Strength of Weak Ties: A Network Theory Revisited", *Sociological Theory*, Vol. 1, pp. 201–233 (1983).
- [3] Mark E. J. Newman: "Spread of Epidemic Disease on Networks", *Physical Review E*, Vol. **66**, No. 1, art. 016128 (2002).
- [4] Romualdo Pastor-Satorras, Alessandro Vespignani: "Epidemic Spreading in Scale-Free Networks", *Phys. Rev. Lett.*, Vol. **86**, No 14, pp. 3200–3203 (2001).
- [5] Romualdo Pastor-Satorras, Alessandro Vespignani: "Epidemic Dynamics and Endemic States in Complex Networks", *Phys. Rev. E*, Vol. 63, art. 066117 (2001).
- [6] Victor M. Eguiluz and Konstantin Klemm: "Epidemic Threshold in Structured Scale-Free Networks", *Phys. Rev. Lett.*, Vol. **89**, art. 108701 (2002).
- [7] Deepayan Chakrabarti, Yang Wang, Chenxi Wang, Jurij Leskovec, Christos Faloutsos: "Epidemic Thresholds in Real Networks", *ACM Transactions on Information and System Security*, Vol. **10**, No. 4, art. 1 (Jan. 2008).
- [8] M. Draief, A. Ganesh and L. Massoulié: "Thresholds for Virus Spread on Networks", *Annals of Applied Probability*, Vol. **18**, No 2, pp. 359–378 (2008).
- [9] X. Guardiola, A. Diaz-Guilera, C. J. Pérez, A. Arenas and M. Llas: "Modeling Diffusion of Innovations in a Social Network", *Physical Review E*, Vol. **66**, art. 026121 (2002).
- [10] M. Llas, P. M. Gleiser, J. M. López and A. Diaz-Guilera: "Nonequilibrium Phase Transition in a Model for the Propagation of Innovations among Economic Agents", *Physical Review E*, Vol. **68**, art. 066101 (2003).
- [11] D. J. Watts: "A Simple Model of Global Cascades on Random Networks", *Proceedings of the National Academy of Sciences of the United States of America*, Vol., **99**, No 9, pp. 5766–5771 (2002).
- [12] D. Centola D, V. M. Eguiluz and M. W. Macy: "Cascade Dynamics of Complex Propagation", *Physica A: Statistical Mechanics and its Applications*, Vol. **374**, No 1, pp. 449–456 (2007).
- [13] L. Buzna, K. Peters and D. Helbing: "Modelling the Dynamics of Disaster Spreading in Networks", *Physica A: Statistical Mechanics and its Applications*, Vol. **363**, No 1, pp. 132–140 (2006).
- [14] J. Goldenberg, B. Libai, E. Müller: "Talk of the Network: A Complex Systems Look at the Underlying Process of Word-of-Mouth", *Marketing Letters*, Vol. **12**, No. 3, pp. 211–223 (2001).
- [15] D. J. Daley. D. G. Kendall: "Epidemics and Rumours", *Nature*, Vol. **204**, p. 1118 (1964).
- [16] D. Gruhl, R. Guha, D. Liben-Nowell and A. Tomkins: "Information Diffusion Through Blogspace", In: *Proceedings of the 13th International World Wide Web (ACM)*, New York (2004), pp. 491–501.

- [17] M. S. Granovetter: "Threshold Models of Collective Behavior", *American Journal of Sociology*, Vol. **83**, No 6, pp. 1420–1443 (1978).
- [18] Y-Y. Ahn, N. Masuda, J. D. Noh and H. Jeong: "Epidemic Dynamics of Two Species of Interacting Particles on Scale-Free Networks", *Physical Review E*, Vol. **74**, art. 066113 (2006)
- [19] J. Goldenberg, B. Libai, S. Moldovan and E. Müller: "The NPV of Bad News", *International Journal of Research in Marketing*, Vol. **24**, No 3, pp. 186–200 (2007).
- [20] D. Trpevski, W. K-S. Tang and L. Kocarev: "Model for Rumor Spreading over Networks", *Physical Review E*, Vol. **81**, art. 056102 (2010).
- [21] Z. Zhu, F. Liu and D. Liao: "Rumor Spreading and Degree-Related Preference Mechanism on a Small-World Network", *Sociology Mind*, Vol. **72**, No. 4, pp. 477–481 (2012).
- [22] B. Karrer, M. E. J. Newman: "Competing Epidemics on Complex Networks", *Physical Review E*, Vol. **84**, art. 036106 (2011).
- [23] B. A. Prakash, A. Beutel, R. Rosenfeld and C. Faloutsos: "Winner Takes All: Competing Viruses or Ideas on fair-play Networks", In: *Proceedings of the 21st international conference on World Wide Web*, pp. 1037–1046 (2012).
- [24] A. Beutel, B. A. Prakash, R. Rosenfeld, and C. Faloutsos: "Interacting Viruses in Networks: Can Both Survive?", In: *Proceedings of the 18th ACM SIGKDD international conference on Knowledge discovery and data mining*, pp. 426–434 (2012).
- [25] Y. Wang, G. Xiao and J. Liu: "Dynamics of Competing Ideas in Complex Social Systems", *New Journal of Physics*, Vol. **14**, No 1, 013015 (2012).
- [26] C. Castellano and R. Pastor-Satorras: "Thresholds for Epidemic Spreading in Networks", *Physical Review Letters*, Vol. **105**, art. 218701 (2010).
- [27] R. Parshani, S. Carmi and S. Havlin: "Epidemic Threshold for the Susceptible-Infectious-Susceptible Model on Random Networks", *Physical Review Letters*, Vol. 104, art. 258701 (2010).
- [28] J. Wolfowitz: "Products of Indecomposable, Aperiodic, Stochastic Matrices", *Proceedings of the American Mathematical Society*, Vol. **14**, No. 5, pp. 733–737 (1963).
- [29] D. Smilkov and L. Kocarev: "Influence of the network topology on epidemic spreading", *Physical Review E*, Vol. **85**, art. 016114 (2012).
- [30] P. Erdős and A. Rényi: "On Random Graphs", *Publicationes Mathematicae*, Vol. **6**, 290–297 (1959).
- [31] P. Erdős and A. Rényi: "On the Evolution of Random Graphs", *Publications of the Mathematical Institute of the Hungarian Academy of Sciences*, Vol. **5**, pp. 17–61, 1960.

- [32] D. J. Watts and S. H. Strogatz: "Collective Dynamics of 'Small-World' Networks", *Nature*, Vol. **393**, pp. 440–442, 1998.
- [33] A.-L. Barabási and R. Albert: "Emergence of Scaling in Random Networks", *Science*, Vol. **286**, pp. 509–512, 1999.

Резиме

**МОДЕЛИРАЊЕ НА ШИРЕЊЕТО ИНФОРМАЦИИ  
ВО КОМПЛЕКСНИ МРЕЖИ**

Во оваа статија предложуваме модел за ширење на два типа информации во мрежи. Моделот е природна генерализација на епидемскиот модел susceptible-infective-susceptible (SIS). Двете информации имаат различна атрактивност што, кога истовремено ќе стигнат кај даден јазел, влијае на одлуката на јазелот која информација да ја усвои. Моделот е во дискретно време и, наспроти верзијата на моделот во непрекинато време проучувана од други автори, го покажува истовременото постоење на двете информации во мрежата во стабилна состојба. Во статијата даваме апроксимации за просечниот број на јазли информирани со секоја информација на крајот од ширењето на информациите во случајот кога јазлите имаат голем степен.

**Клучни зборови:** комплексни мрежи; динамички системи; нелинеарни системи; ширење на информации

**Address:**

**Daniel Trpevski**

*Macedonian Academy of Sciences and Arts,  
Boul. Krste Misirkov 2, 1000 Skopje, Macedonia  
dtrpevski@manu.edu.mk*

**Kire Stamenov**

*Macedonian Academy of Sciences and Arts,  
Boul. Krste Misirkov 2, 1000 Skopje, Macedonia  
kirestamenov1989@yahoo.com*

**Ljupčo Kocarev**

*Macedonian Academy of Sciences and Arts,  
Boul. Krste Misirkov 2, 1000 Skopje, Macedonia  
lkocarev@ucsd.edu*