Applied Mathematics and Computation 254 (2015) 229–239

Contents lists available at ScienceDirect

Applied Mathematics and Computation

journal homepage: www.elsevier.com/locate/amc

Influence maximization of informed agents in social networks

Omid AskariSichani, Mahdi Jalili *

Department of Computer Engineering, Sharif University of Technology, Tehran, Iran

ARTICLE INFO

Keywords:
Complex networks
Opinion formation
Influence maximization
Influential identification
Informed agents
Multi-agent social simulation

ABSTRACT

Control of collective behavior is one of the most desirable goals in many applications related to social networks analysis and mining. In this work we propose a simple yet effective algorithm to control opinion formation in complex networks. We aim at finding the best spreaders whose connection to a reasonable number of informed agents results in the best performance. We consider an extended version of the bounded confidence model in which the uncertainty of each agent is adaptively controlled by the network. A number of informed agents with the desired opinion value is added to the network and create links with other agents such that large portion of the network follows their opinions. We propose to connect the informed agents to nodes with small in-degrees and high out-degree that are connected to high in-degree nodes. Our experimental results on both model and real social networks show superior performance of the proposed method over the state-of-the-art heuristic methods in the facet of opinion formation models.

1. Introduction

Network science has attracted much attention in recent years, which is mainly due to the recent advances in graph theory and availability of huge datasets [1]. Sociology has benefited from these advances in network science and many scholarly works have been carried out on social networks analysis and mining. Suppose that we are about to market a new product, to promote an innovation or to debate the candidacy of a novice in the parliament. Or even in other applications, such as controlling a set of dynamical systems, in which, similarly each person is considered as a single dynamical system and the whole environment is a complex network, different systems start to affect each other [2]. All of which, explicitly or implicitly, contain an opinion formation process within the interference in a social network of people in which some kind of consensus process happens between the agents. In this literature, researchers have designed a suitable and accurate system that in which feedback loops, instabilities and cascade effects are considered in order to simulate a natural system. In such a paradigm, called complexity science, models and eventually results would be more realistic. [3]. The ultimate goal is to efficiently model this phenomenon and find the best strategy to propagate a desired opinion between the individuals with the lowest possible cost. Such a goal can be achieved by employing informed agents in the society [4]. The society is considered as a graph in which nodes represent individuals (or agents) and edges represent the interactions between them. Every agent has a value indicating his/her interest, taste or tendency toward an objective. This value is referred to as the *opinion* of the agent.

As two persons meet and interact, they make their opinions closer by some kind of update process. Opinion formation is a generic form of information diffusion, information cascade or innovation propagation which is proposed to mathematically

* Corresponding author at: Department of Computer Engineering, Sharif University of Technology, Azadi Avenue, Tehran, Iran.
E-mail address: mjalili@sharif.edu (M. Jalili).

http://dx.doi.org/10.1016/j.amc.2014.12.139
0096-3003/

describe the concepts observed in real-world networks. There are several applications regarding such models such as community detection [5], large-scale non-convex optimization [6], analysis of opinion formation by informed agents [4], and forecasting final opinions in a social network [7]. In this work, we consider each person as a particle and propose a measure for selecting the individuals maximizing the desired diffusion. The result of this study could be useful in influence maximization in terms of economics, targeted virus immunization and finding the optimal strategy in social advertisements. A number of methods have been proposed to model this opinion formation process [8]. For instance, Voter model represents selection between two distinct opinions; a political electoral system among two parties is of this kind [9,10]. Likewise, there are other models for opinion dynamics in the literature that are based on binary opinions [11–16]. In more recent models, opinions are represented as the outcome of social or cultural traits of the people who live together as discrete [17] or continuous values [18,19]. Related to this, one can also mention co-evolutionary games that integrate cooperation between competitive agents and concepts of network consensus [20,21]. These methods provide a theoretical framework to cover the phenomena, such as herding effects, social shifts and interactions among living organisms [22]. The viewpoint of these game theoretical methods sheds more light on several cooperation problems. For instance, in the Prisoner's Dilemma game, this knowledge could be used to find the optimal maximal degree warranting the best information exchange among influential players [23].

In this manuscript, we aim at finding the best communication strategy in order to maximize the influence of informed agents on the opinions of other agents. Given the structure of a social network, we would like to determine a subset of agents whose connecting to the informed agents will cause the best consensus properties (e.g., largest portion of the agents getting into a consensus towards the desired opinion) in the network. In this regard, there are some heuristic methods such as high-degree, high-betweenness and high-closeness nodes [24,25], and even more sophisticated metrics that we take into account [26]. In this manuscript, we introduce a heuristic algorithm and show that it outperforms a number of other algorithms. This model involves agents with continuous opinions in a range between two boundaries and a synchronous discrete time update scheme based on bounded confidence model. The model is numerically simulated on both artificially generated and real social networks and the results reveal its effectiveness.

2. Background

2.1. Opinion formation with social power

A class of opinion formation models is defined based on continuous opinion values and a special opinion update rule [27–34]. In these models, each individual takes two deterministic continuous values: the opinion and the uncertainty. Each agent updates his/her opinion through communications with his/her neighbors. The opinion of agent i can change over a range as [28]

$$[x_i - u_i, x_i + u_i], \tag{1}$$

where x_i is the opinion of node i and u_i is its uncertainty. The uncertainty can model personal characteristics of the agents. For instance, sociable agents have higher values of uncertainty than unsociable ones.

Let us consider a directed network on which the individuals interact with each other. These individuals discuss their opinions, and if they meet some criteria, their opinion values will be updated [30,35]. Deffuant et al. introduced an opinion formation model, denoted as bounded confidence model, that has an iterational process [28]. In each iteration, a node and one of its neighbors are selected randomly. If they have close enough opinions, they will make their opinions closer to each other. In general, two sets of information are important in updating the opinion values: neighborhood between the individuals and the difference between their opinion values. Individuals are affected only by those with a direct link on the network, i.e., their neighboring nodes. Furthermore, an agent influences the opinion of its neighbors if the distance between their opinions is less than a certain threshold. Let us consider a network of size N. At each step, one of the existing links between nodes i and j having the opinions x_i and x_j (that each can take a value in the range $[-1, +1]$) is randomly chosen. If the difference between the opinions of the two neighboring agents is less than a specified threshold, commonly called uncertainty of the one who is influenced, i.e., $|\mathrm{x_i} - \mathrm{x_j}| < u_i$, they will influence each others' opinions through the following opinion update equations

$$\begin{cases} x_i(t+1) = x_i(t) + \mu\left[x_j(t) - x_i(t)\right] \\ x_j(t+1) = x_j(t) + \mu\left[x_i(t) - x_j(t)\right] \end{cases}, \tag{2}$$

where t indicates time-step of the simulation process. If the bounded confidence condition does not hold, i.e., the difference of the opinions is larger than u_i indicating the uncertainty of agent i, no update will be made in the opinion values. μ is the convergence rate (opinion change rate) parameter, which often takes a value between 0 and 1. We implement a synchronous update scheme for the agents and perform the numerical simulations with large enough iterations to ensure that the maximum possible consensus has happened. We then monitor the final opinion values. In this way, we make sure the method is converged as the opinion trend do not change dramatically afterwards.

In the original bounded confidence model, all of the agents have the same influence on their neighbors. However, in real networks, people have different social capabilities, political situations, incomes, or physical features, and thus different *social* *power* on their neighbors. Social power is also known as social influence, social pressure or social impact and has been used in the literature [17,36,37]. In order to take into account social diversity, the following term has been proposed as the social power of the agents [38], as follows

$$S_i = (k_i^{in})^{\alpha}, \tag{3}$$

where S_i is the social power of agent i, k_i^{in} is its in-degree (or any other centrality measure) and α is a control parameters. We set $\alpha = 0$ and $\alpha = 1$ in this work. Comparing the results of this configuration is useful to understand the importance of social power in diffusion. The value $\alpha = 1$ is suggested in [38], which means that the social power of an agent is equal to its in-degree. Taking into account the social power, the update equation is rewritten as follows

$$\begin{aligned} x_i(t+1) &= \begin{cases} x_i(t) + \mu \frac{S_j}{S_i+S_j}\left(x_j(t) - x_i(t)\right) & \textit{if } \left|x_i(t) - x_j(t)\right| < u_i \\ x_i(t) & \textit{else} \end{cases} \\ x_j(t+1) &= \begin{cases} x_j(t) + \mu \frac{S_i}{S_i+S_j}\left(x_i(t) - x_j(t)\right) & \textit{if } \left|x_i(t) - x_j(t)\right| < u_j \\ x_j(t) & \textit{else} \end{cases}, \end{aligned} \tag{4}$$

where $x_i(t)$ is the opinion of agent i in time-step t, μ is the convergence rate, S_i is the social power of agent i and u_i is its uncertainty.

2.2. Existing metrics

Finding influential individuals in a network could conclude to a wider and faster diffusions through complex networks; as a consequence, this problem is considered in this work in opinion formation circumstances. We need to identify the level of persuasiveness for each node through a cost-effective method. There are many heuristics that uncover the influential nodes in a complex networks. We used degree, betweenness and closeness centrality measures as the heuristic methods in order to compare their performance with the proposed heuristic in the following. Degree is the simplest measure that is defined as the number of connected nodes. Betweenness centrality for each node i is defined as the number of shortest paths connecting each pair of nodes passing through node i, divided by the number of all shortest paths; it is obtained as

$$B_i = \sum_{j<k} \frac{g_{jk}(i)}{g_{jk}}, \tag{5}$$

where g_{jk} is the number of shortest paths connecting the nodes j and k, and $g_{jk}(i)$ is the number of these path passing through i.

$$C_i = \frac{N-1}{\sum_{j=1}^{N} d(i,j)}, \tag{6}$$

Closeness is one of the centrality metrics which is based on the inverse of average shortest path length. The concept is inspired from the question it asks what if it is not important to have many direct friends (high degree) or to be in a vital path connecting people together (high betweenness) and just it is prominent to be in the middle of the network (high closeness). Indeed, closeness describes the distance from the center of the network. Eq. (6) is a normalized from of closeness where $d(i,j)$ is the average path length between nodes i and j.

Apart from the well-known centrality measures, in order to identify influential nodes in a social network, Chen et al. in 2012 introduced a measure, named local centrality, as a trade-off between degree and the time-consuming centrality measures. This metric considered both the nearest and the next nearest neighbors and produced a semi-local measurement which is effective in comparison with the well-studied heuristics [26]. We call this metric *Chen* in this study owing to the first author's name. Chen's measure, or technically speaking, $C_L(v)$ of node v is defined as follows

$$C_L(v) = \sum_{\forall u \in N_v} \sum_{\forall w \in N_u} \Gamma(w), \tag{7}$$

where N_v is the neighbors list of node v and $\Gamma(w)$ is defined as

$$\Gamma(w) = N_w \bigcup \left(\bigcup_{\forall p \in N_w} N_p \right) - \{w\}, \tag{8}$$

where $\Gamma(w)$ is a set of nodes which are neighbors or neighbors of neighbors of node v. Chen et al. used Susceptible-Infected-Recovered (SIR) model to evaluate the mentioned measure and showed it could outperform to some extent other existing heuristics to spread the desired idea, especially in heterogeneous networks [26]. In this study, we implement local centrality metric and compare it with the proposed method in an opinion formation manner to find out which method could help the desired opinion to be propagated the best in various network structures. It is worth mentioning that in another experiment which its results are not discussed here, we used in-degree instead of degree measure in Eq. (8) in order to take into account of directed networks; nonetheless, the results were quite similar.

3. The proposed method for influence maximization

In this section, we describe an effective algorithm to control the opinion formation process in complex networks. Accordingly, a number of informed agents whose opinion values are fixed at the desired value are added to the initial population. These agents try to propagate the desired opinion through the network. Since creating new links requires cost, the informed agents create minimal number of connections (the most rational and possible option, i.e., one connection per agent in this work) such that maximum performance is obtained. It is conspicuous that more connections each agent make, the easier the solution is obtained. In this study, We consider a network with N nodes which each individual has an opinion falls in $[-1, +1]$ and uncertainty is defined constant as 0.5. Our aim is to find those agents that the informed agents should create connections to, such that the maximum performance (i.e., the maximum number of the agents who have reached to a consensus at +1) eventually will be obtained. Node degree is the simplest quantity that can be obtained in social networks. Our proposed method is based on knowledge of the degree of every node and that of the neighbors.

In this literature, an extreme agent is defined as an agent with an opinion close to the extreme values (boundaries) whose uncertainty is quite small (i.e., difficult to change the opinion) [35]. Often, there are a few extremists in social networks, they are not flexible and hardly change their opinions. We consider such extreme agents to take an opinion value close to -1 or $+1$ (which are the considered boundaries).

The substantial point is that high-degree nodes are more difficult to be influenced than those with low-degrees. Therefore, if the informed agents are directly connected to hub nodes, they are less likely to be able to change these nodes' opinion. Furthermore, in the opinion formation process, the selection probability of informed agents, when they are attached to high-degree nodes, is very low; and this may decrease their efficiency. On the other hand, high-degree nodes have greater influence on the society than low-degree nodes and could propagate the desired opinions faster. In this work, we propose a strategy to choose the target set in order to create direct links with informed agents. This strategy is just based on the knowledge of node degrees and degree of their first neighbors. Since low in-degree nodes are highly negotiable, the target set is chosen among nodes with small in-degrees, i.e., these nodes will be considered as the interface of informed agents with the society. On the other hand, these nodes should be able to efficiently spread the opinion of informed agents to the whole society. Therefore, they should have direct links to nodes with many neighbors (i.e., high in-degree agents or nodes with high prestige). We propose the following metric called EPN (Effective Potential Nodes) measure for the nodes in order to choose them in the target set to make a direct link with informed agents. It is called EPN1 since every node just take into account of his first neighbors.

$$\text{EPN1}: \qquad f_i = \frac{k_i^{out} + \varepsilon}{k_i^{in} + \varepsilon} \times \frac{\frac{1}{|N_i^{out}|} \sum_{j \in N_i^{out}} k_j^{in}}{\arg\max\limits_{l} \left(k_l^{in}\right)}, \tag{9}$$

where k_i^{out} is out-degree of node i and N_i^{out} is a list of nodes to whom i has outgoing links. The probability of selecting node i in the target set would be f_i, if we normalize Eq. (9) with sum over all of its values. Nodes with small in-degree (easy to be influenced) and large out-degree (appropriate for propagation) that are connected to nodes with large in-degree will have high value for f, and thus, are likely to be chosen in the target set. In the experiments, first, the above metric is calculated for all the nodes to obtain the f-values. Then, for m informed agents added to the network, m nodes with the highest f-value are selected and each of the informed agents (that all have opinion value of +1) is connected to these nodes, one by one. The proposed measure has lower computation than Chen's but because of the idea behind it, is quite more effective as one could see in the section (4), i.e., Results section.

EPN measure has the ability of being applied with multiple levels (in this study, we consider only 1- and 2-level measure). In order to consider 2 levels for EPN, instead of Eq. (9), one should perform the computations as

$$\text{EPN2}: \qquad f_i = \frac{k_i^{out} + \varepsilon}{k_i^{in} + \varepsilon} \times \frac{\frac{1}{|N_i^{out}|} \sum_{j \in N_i^{out}} \left(\beta k_j^{in} + (1-\beta)\frac{1}{|N_j^{out}|} \sum_{p \in N_j^{out}} k_p^{in}\right)}{\arg\max\limits_{l} \left(k_l^{in}\right)}, \tag{10}$$

where β is a scale parameter which is chosen from $[0, +1]$. The only difference between Eq. (9) and Eq. (10) is that in order to obtain the values through Eq. (10), one should go one level deeper by considering the mean degree of neighbors of the neighbors.

4. Simulation results

In this work, we use both artificially-generated networks and real-world social networks as well. As in model networks, we consider scale-free, small-world and random networks. As a benchmark for scale-free networks, we take into account Barabasi–Albert (BA) network that is based on preferential attachment algorithm [39]. In this growing model, an initial network grows and at each step a node with m links is added to the network. These new nodes are linked to old nodes with a probability that is proportional to their degree. This process is repeated until the size of the network reaches to N. The model results in a network with power-law degree distribution that is common in many real systems. Real networks have been shown to have not only scale-free property, but also concretion and shrinking diameter [40]. We use the model proposed in [40], called Forest Fire (FF) model, to produce such networks.

Another well-known structure is the one introduced by Watts and Strogatz (WS) resulting in networks with both small-world property and high transitivity [41]. WS networks are constructed as follows. First a regular ring graph with N nodes each connected to its m-nearest neighbors is constructed. Then, the connections are rewired with a probability P provided that self-loops and multiple connections are prohibited. We set $N = 2000, P = 0.1$ and $m = 6$ in this work. We also consider pure random networks produced by the model proposed by Erdős and Rényi (ER) [42]. In ER networks, every pair of nodes are independently set to share a link with probability P. Here we set $N = 2000$ and $P = 0.012$ resulting in networks with expected average degree of 12, similar to other networks. Many social networks show community structure, where the intra-community links are dense while there are sparse inter-community connections. We use the model proposed in [43] in order to construct such networks. In this model, first a number of similar isolated networks (BA or WS networks in this work) are considered. Then, each intra-community link is rewired to an inter-community link with probability P_{inter}. In this way, the average degree of the final network will be the same as the isolated networks. We set $P_{inter} = 0.1$ in this work.

Real networks might have properties that cannot be captured by existing models. Therefore, we also consider a number of freely available social networks data. The adjacency matrix of these networks can be downloaded at (http://konect.uni-koblenz.de/networks/). (Table 1 summarizes the demographic information of these networks. Advogato is an online social network of free software developers that describes itself as the advocate for open source programmers' community. It is free and online since 1999. Google Plus network is a directed network containing Google+ user-user links. Each node represents an user, and each directed edge denotes that one user has the other user in his circles. It is one of the original datasets from SNAP project [44]. Filckr is a photo sharing website owned by Yahoo Corporation and we used the social network between its users and their friends. Facebook is one of most famous social networks in the world. We used a simple and small subset of which that is a directed network containing friendship data of more than 63000 users.

In order to perform the experiments, first, the agents are given an initial opinion values at random. Then, a number (10% of the network size N) of informed agents with desired opinion value (+1 here) is added to the network. Each informed agent creates one link to the network and our aim is to find to whom they should make a connection. The connections should be made in a way such that the population opinion could be pinned in a short time and the targeted opinion goes viral very fast.

Parameter μ of the opinion update model (in equations (2) and (4)) plays a pivotal role in the convergence process; accordingly, the first step is to fix it to an optimal value. (Fig. 2) shows the mean opinions in BA networks as a function of μ with the maximum simulation steps set as 105. The results show averages over 100 realizations. As expected, by increasing μ the performance improves, i.e., larger number of agents gets into synchrony. However, no significant improvement is obtained for valued $\mu > 0.8$; thus, we fixed $\mu = 0.8$ for the rest of the experiments. The behavior of μ in other networks is similar to BA ones (data is not shown here).

In order to study the effect of parameter β in EPN convergence with two levels, we employed different values for this parameter from the range $[0, +1]$ in a similar network configuration. (Fig. 1) shows the average of opinions at the end of the opinion formation process in which the goal is to reach the desired opinion value +1.

Increasing the value of parameter β makes EPN2 more local. If we set it to 1, EPN2 will equal EPN1. This means that each agent considers only his/her direct friends. If we set it to 0, then each agent ignores the direct friends and takes into account only the friends of his/her friends. Therefore, a value in between is the best choice. It is worth mentioning that if higher levels of EPN are used, this control parameter plays more important role; as this level parameter increases, the number of friends of friends to be considered is more, and consequently, we need more computational power.

Given (Fig. 1) for tuning parameter β, we fix the best obtained value for it which is 0.1 after this phase. These results show that the second term of Eq. (10) is more significant for diffusion, i.e., as EPN goes more global, it performs much better.

Informed agents have the role to intelligently convince the society to their opinions. In this work, we compare the performance of the proposed method (i.e., choosing the nodes to which the informed agents should create a connection) with three heuristic methods: choosing these target nodes based on their betweenness, their closeness and their degree, i.e., the larger is the degree of a node, the higher the chance to be linked to an informed agent. We also compare diffusion trends with Chen's algorithm [26]. Note that in our strategy, we connect the informed agents to low in-degree and high out-degree agents that have high in-degree agents in their neighborhood. (Fig. 2) shows the average opinions as a function of the ratio

Table 1
Structural information about the real-world networks which has been used in this work. N is the network size, $<k>$ average degree, C clustering coefficient, L average path length in the largest connected component and *LCCS* is the largest connected component size. All of which are real social networks with different sizes.

Network name	N	$<k>$	C	L	*LCCS*
Advogato	6,551	15.67	0.15	3.27	5,054
Google Plus	23,628	3.3217	0.1741	4.71	23,613
subset of Facebook friendships	63,731	25.642	0.1250	5.14	63,392
Flickr	2,302,925	28.781	0.2457	6.6790	2,173,370

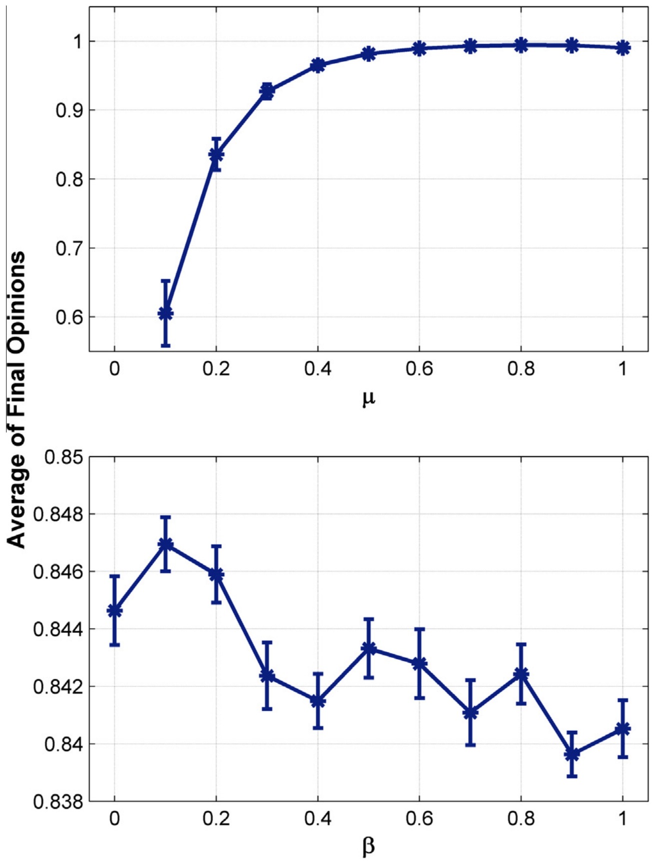

Fig. 1. Average of final opinion values and standard error as a function of parameters μ and β. 200 informed agents are added to a Barabasi–Albert network with $N = 2000$ nodes and $m = 6$ based on proposed heuristics expressed in Eq. (9) and Eq. (10). Data show averages over 100 simulations. μ is the convergence rate and the figure shows that for $\mu = 0.8$, the average opinion with lower uncertainty almost reaches to the highest plausible portion. β on the grounds of Eq. (10), depicts the influence of first-layer (direct) neighbors in comparison with the second ones.

of informed agents. It is expected that as more informed agents with desired opinion value are added to the network, the performance improves; however, there is no mentionable differences between EPN1 and EPN2.

(Fig. 3) shows the average of opinions in the last iteration of simulation as a function of social power parameter α that varies in the range $[0, +2]$. The informed agents are connected to target nodes that have been chosen based on different strategies including the ones proposed in this work. It is seen that as the exponent increases, the performance of informed agents reduces, which can be explained as follows. Each informed agent is just connected to one node in the network, and thus, owing to this fact that the degree keeps constant equal to 1, its influence will not change by increasing the social power exponent. However, by increasing the exponent, the internal nodes of the network have more power to influence others, and of course, need more power to be influenced. The results show that the proposed EPN strategy outperforms other heuristics methods.

Assortativity of a graph is defined as

$$r = \frac{\frac{1}{E}\sum_{j>i} k_i k_j a_{ij} - \left[\frac{1}{E}\sum_{j>i}\frac{1}{2}(k_i + k_j)a_{ij}\right]^2}{\frac{1}{E}\sum_{j>i}\left(k_i^2 + k_j^2\right)a_{ij} - \left[\frac{1}{E}\sum_{j>i}\frac{1}{2}(k_i + k_j)a_{ij}\right]^2}, \tag{11}$$

where r is assortativity measure, E is the number of edges, a_{ij} is the (i, j) entry of adjacency matrix. if $r > 0$ we call it assortative network and if $r < 0$ we call it disassortative one, else there is no specific intention in the connection between the nodes in the sense of their degrees.

We used the standard edge-swapping method to change the assortativity of the networks [45,46]. This results in networks with a desired assortativity level in the range $[-1, +1]$. In this method, in each iteration, two edges are selected at random. Then, these edges are swapped, resulting in two new edges. These new edges are replaced with previous ones if they change assortativity in the desired direction. Otherwise, the new edges are discarded and the old edges are kept. Such

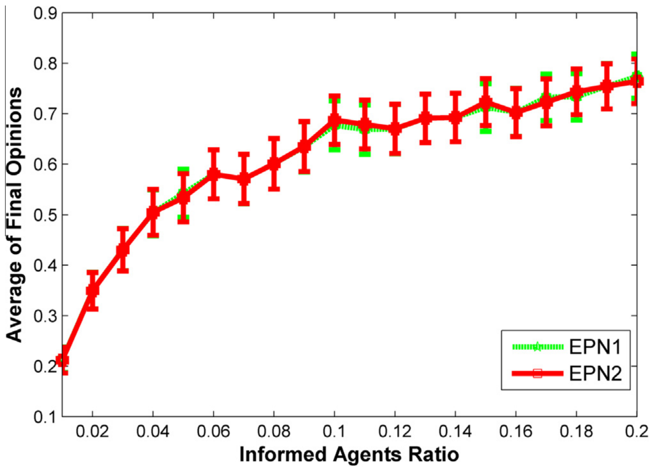

Fig. 2. Average and standard error of final opinions as a function of ratio of informed agents varies between 0.01 and 0.2 with 0.01 steps. Parameters $\mu = 0.8$ and $\beta = 0.1$. Other designations are as (Fig. 1). Increasing the ratio of informed agents obviously results in more unanimous society members, and furthermore, it eventuates in a higher average final opinions, when our desired opinion is considered as +1.

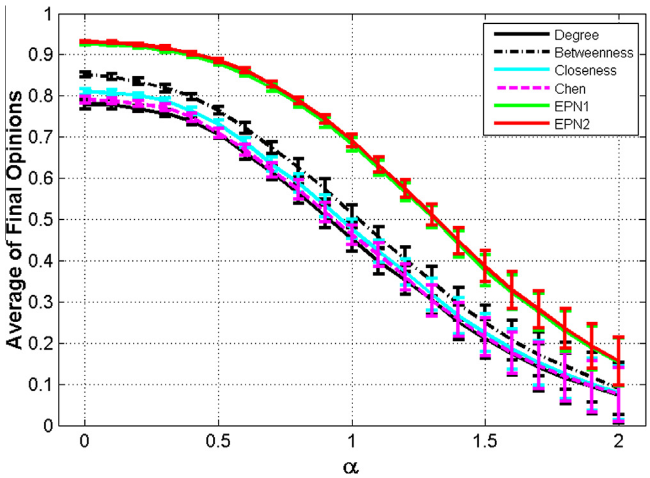

Fig. 3. Average and standard error of final opinions as a function of social power exponent (α). 200 informed agents with the desired opinion are connected to target nodes in Barabasi–Albert network with 2000 nodes (means they will have 2200 nodes). When an informed agent is added to the network, it is connected to the nodes in the original network with a probability that is proportional to the degree (Degree), to the betweenness centrality (Betweenness), to the closeness centrality (Closeness), to Chen's centrality measure (Chen), based on the EPN1 (1 level – as expressed in Eq. (9)) or EPN2 (2 level – as expressed in Eq. (10). Average degree is 12 and the desired opinion value equals to +1. Parameters $\mu = 0.8$ and $\beta = 0.1$ were chosen for this experiment due to the results of parameter tuning phase depicted in (Fig. 1). Data show averages over 50 runs. It stands to the fact that when members' social power increases, it becomes harder to convince them to change their opinion; thus, the trend is decreasing.

edge swaps preserve the in- and out-degrees of all nodes involved, thereby keeping the degree distribution intact [47]. The result is shown in (Fig. 4), where, expecting, EPN strategy works better in disassortative networks. In disassortative networks, existence of the potential nodes with low degree and neighbors with high degree is more probable. Therefore, we expect that EPN measure to work better in such network structures, and indeed, this result is an endorsement to the idea behind of proposed heuristic.

We applied the algorithms on various model networks and the results are shown in (Fig. 5: A). It shows standard error and average of final opinion values as a function of the simulation steps. The desired opinion value is +1, and the initial opinion values are uniformly chosen in the range $[-1, +1]$. The results show that EPN strategy (with 1 or 2 levels) work better than others in all network structures. Moreover, we found that without using any social power (i.e., $\alpha = 0$), the propagation experiment converges in time-steps about 5 to 10 times of network size. However, when the nodes have social power (i.e., $\alpha = 1$), the convergence time increases. Indeed, when the nodes have social power, they resist against changing their opinions, and consequently the time needed for convergence to the desired opinion increases. Additionally, in networks with community

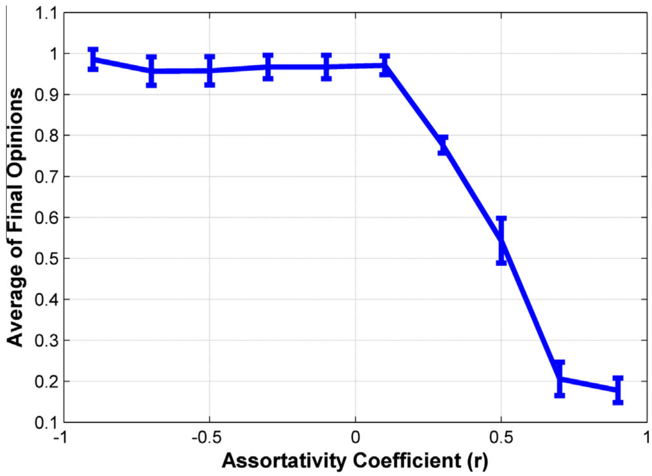

Fig. 4. Average and standard error of final opinions as a function of assortativity coefficient (r) in Barabasi–Albert scale-free networks with $N = 2000$ and $m = 6$. Data show mean value with bars corresponding to standard deviation over 100 realizations. Thanks to the idea behind EPN heuristics (depicted in Eqs. (9)), those networks enjoying a substantial population of low-degree nodes connected to high-degree ones (disassortative), are more convenient targets for the proposed method. This figure shows that networks with assortativity coefficient lower than 0.3 are appropriate in order to propagate desired opinions to more than 80% of the society members.

structure, when the informed agents are connected based on betweenness measure, the convergence performance is better than the case when the connections are made based on degree and closeness measures for instance in (Fig. 5, a6 and b6). (Fig. 5: B) shows the results when the nodes have a social power with exponent $\alpha = 1$. While in some cases, the degree-, betweenness-, closeness-based and Chen strategies could not converge to $+1$ (the desired opinions), the EPN-based strategies could and in most cases, it has higher speed than other methods. Our results indicate that degree distribution and community structure of the networks are important factors in determining their propagation properties.

(Fig. 5) demonstrates that when higher social power is taken into account, the standard error increases and wide-spread cascade among the agents is more difficult to be attained. However, EPN and especially those with higher levels are better than other methods. In the case when the network has strong community structure, the metric based on betweenness could provide better results that the proposed method. However, betweenness is a metric that need global information on the network, and its computation is much more expensive that the proposed metrics.

In addition, as seen in (Fig. 5), the metrics based on degree and Chen measure have a higher standard error than others. Selecting a high-degree node to lobby can be inefficient, since such an agent is often have high social power. As a result, it can frequently happen that a high-degree agent do not pay attention to the lobbyist's opinion resulting in a poor opinion propagation. As shown in Fig. 5 and Fig. 6, there is not much a difference between the metrics based on EPN1 and EPN2 since their concept is the same. However, when the networks have strong community stronger, EPN2 outperforms EPN1.

As it has been frequently shown, artificially-generated networks through models cannot capture all properties of real-world networks. Therefore, we implement the algorithms on a number of real-world social networks for which the information is given in Table (1). (Fig. 6) shows average opinions as a function of simulation time-steps in four real networks. Comparing (Fig. 5: A) and (Fig. 6), it is seen that the gap between the proposed method and the other three methods is more pronounced in real networks than those obtained through models. It is seen that the final opinions do not converge to the desired opinions (+1) in Advogato and Flickr networks, which is due to the fact that these networks are disconnected and the desired opinion cannot be well-propagated through the network. The proposed EPN strategies outperformed the other three algorithms in Advogato, Facebook and Flickr networks; however, the betweenness-based strategy was the best-performed in Google Plus network. Average path length plays an important role in diffusion properties of networks. In the real networks studies in this work, Advogato has the smallest average path length, while Flicker and Facebook have the largest. Advogato has faster convergence than Flicker and Facebook networks. However, while Flicker has larger average path length than Facebook network, its convergence is faster, which is due it its higher clustering coefficient than Facebook. High clustering coefficient makes the local connections stronger, and thus, the propagation properties of networks.

As discussed, when the informed agents are connected to high-degree nodes, they try to influence the society through these nodes. To this end, first, these high-degree nodes should be influenced by informed agents, and then, their opinion spread to the whole society. However, high-degree nodes are not flexible and hardly change their opinions. Therefore, the opinion spread faces a difficulty at the very first step, if such a strategy is used.

Our proposed method tries to target bridge nodes which are connected to hub nodes. Since hub nodes have large degree, and thus probably small uncertainty and small probability for the informed agent to be selected. In contrast the nodes which are connected to hub nodes and have small degree are not difficult to be convinced to change their opinions. On the other hands, they connect to high-degree nodes, and thus they can communicate the desired opinion value to many other nodes

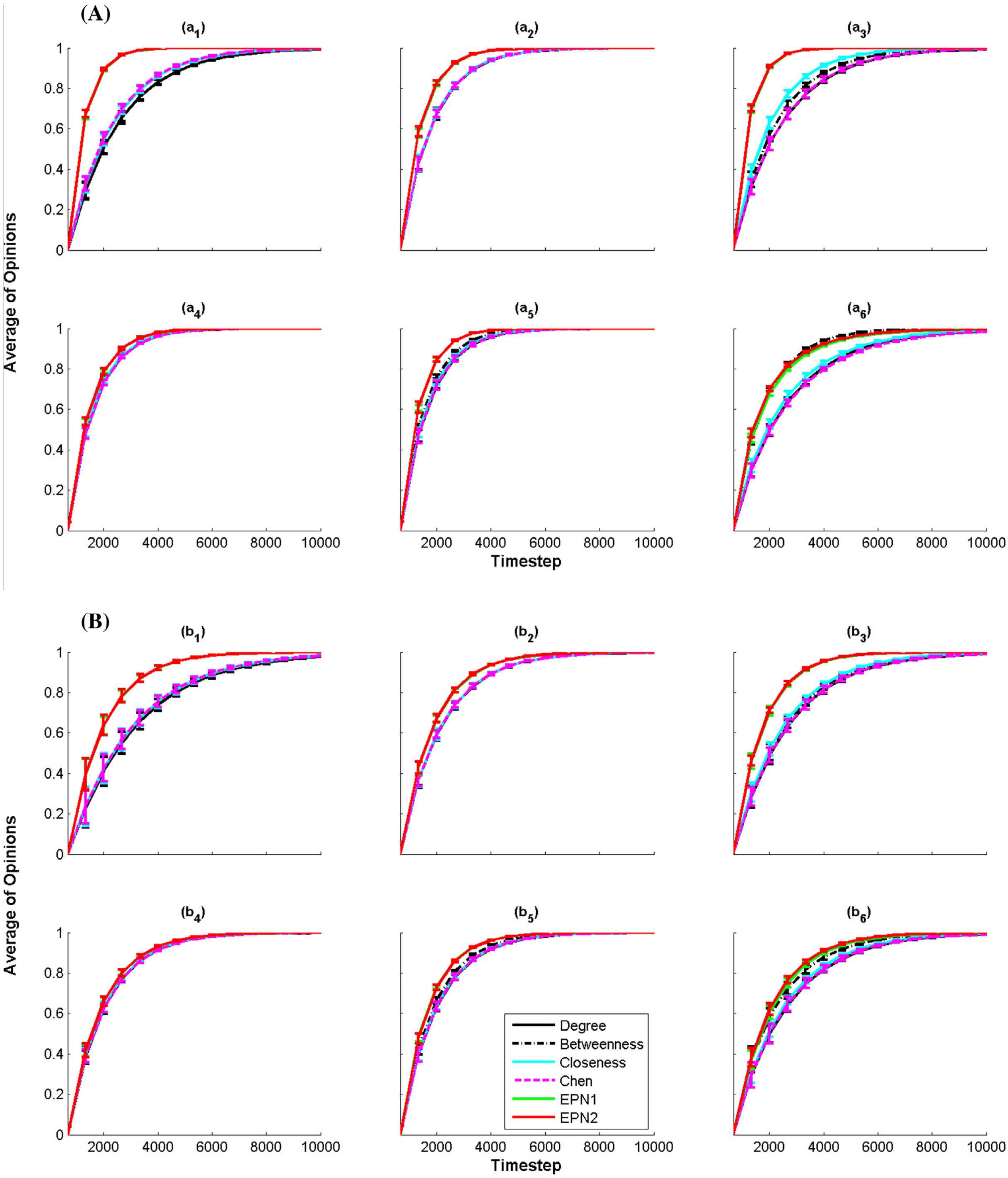

Fig. 5. Average opinion and standard error as a function of time-step in model networks (A) without ($\alpha = 0$), and (B) with ($\alpha = 1$) social power. The network types are (a1,b1) Barabasi–Albert, (a2,b2) Erdős–Rényi, (a3,b3) Forest-Fire, (a4,b4) Watts–Strogatz, (a5,b5) Barabasi–Albert with community structure and (a6,b6) Watts–Strogatz with community structure. Other designations are as (Fig. 3). The main message of this figure is that when the social influence is considered in a social network, the propagation of the desired opinion (of the informed agents, which often do not have high social power) is harder to reach. However, EPN-based methods illustratively surpass other existing methods and base-line heuristics in reaching the desired opinion (+1).

through these hub nodes. The proposed metric somehow acts as betweenness centrality; however, it only needs local information (i.e., degree), while computation of betweenness centrality needs global information from the network. As the results show, the consensus time is always faster using EPN measure with one or two levels as compared to the cases when the

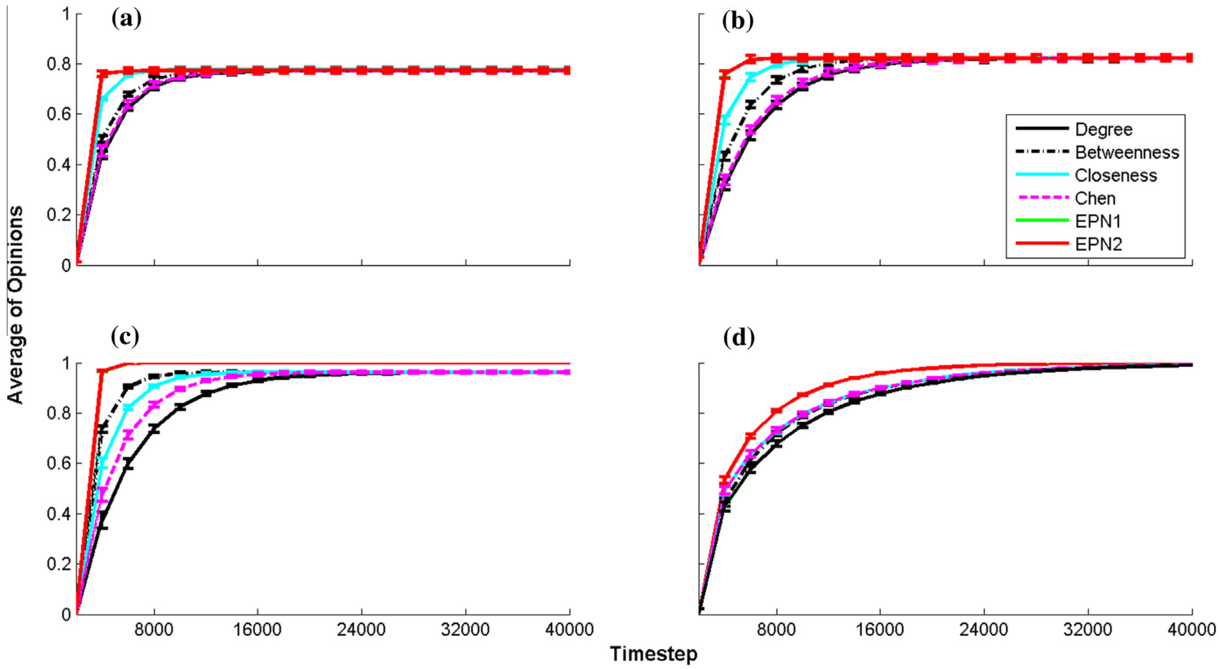

Fig. 6. Average opinion and standard error as a function of time-step in real networks including (a) Advogato, (b) Google Plus, (c) Flickr and (d) a subset of Facebook friendships networks. Data show averages over 20 runs and with $\alpha = 0$. Other designations are as (Fig. 3). The information about these real networks can be found in Table (1).

informed agents are connected to high degree, betweenness, closeness nodes or even Chen's method. The strategy based on betweenness centrality, as a global centrality measures, performs better than degree-based strategy in almost all and better than closeness in most of the time. Chen's centrality measure is a middle metric which is cost-effective and appropriate in various networks; however, the proposed method outperforms all. Networks with community structure are better spreaders as compared to those without such a structure.

As the compendium of this study, we need to highlight that the proposed heuristic method (EPN), has its roots in a set of ideas. First, to influence the society, we target the nodes that have low prestige (in-degree), and thus can easily be influenced. In order to have maximal influence on the society, these low in-degree nodes should be connected to the nodes that have high in-degree. Additionally, these nodes have better efficiency when they enjoy high out-degree values. If we consider all of these ideas together and properly perform a normalization, the final measure would be EPN1. It should be noted that affecting high-prestige nodes with a third-party is a wiser choice than directly connecting them. This is due to the fact that these third party nodes have already made a trust relations with others and have usually higher degree than the informed agents.

Based on the results in Section (4), after tuning the algorithm with the best possible parameters (Fig. 1), we can verify the implementation in order to analyze the behavior of the model as the number of informed agents increases. It is almost monotonically increasing which perfectly aligns with our expectation (Fig. 2). Moreover, we can verify our method with the assortativity measure and depict its strength circumstances (Fig. 4). In addition, we learn that although considering social power for the agents makes the problem harder, it can still be easily solved by EPN-based method, emphasizing superior performance of the proposed method over state of the art algorithms. (Fig. 3). As a consequence, after applying the base-line methods and EPN with different levels over manifold real (Fig. 6 and a variety of artificial networks (Fig. 5), we can potentially address this method as a superior solution and a cost-effective heuristic in the opinion formation literature.

5. Conclusion

In this manuscript, we studied the problem of influence maximization in social networks with informed agents. We used bounded confidence model in order to update the opinion of interacting agents; if two agents are chosen and their opinion values are close enough, they update their opinions. We also assumed a hypothesis; the higher the in-degree of an agent, the higher social power and as a result the lower uncertainty will be. Furthermore, a number of informed agents with known opinion value were added to the network and connected to a set of nodes. These informed agents try to shift the opinion of other agents to their opinion values. We proposed an algorithm to choose the target nodes in order to connect to informed agents. These nodes have small in-degree and large out-degree values which have high in-degree nodes in their neighborhood. Our experiments on a number of model and real social networks showed that the proposed method outperformed heuristic methods including connecting the informed agents to high-degree nodes or those with high betweenness or closeness centrality. We applied a recent measurement proposed by Chen et al. (2012) [26]; nonetheless, the proposed method outperforms that, as well.

As future works, we will consider unsymmetrical (biased) initial distribution of the opinions for the agents. When the agents' opinion values are initialized in a certain way, the way the informed agents should lobby would certainly be different (which will depend on the initial opinion values).

References

[1] S.V. Boccaletti, Y. Latora, M. Moreno, D.H. Chavez, Complex networks: structure and dynamics, Phys. Rep. 424 (4) (2006) 175–308.
[2] M. Jalili, O. AskariSichani, X. Yu, Optimal pinning controllability of complex networks: dependence on network structure, to appear in J. Phys. Rev. E.
[3] D. Helbing, D. Brockmann, T. Chadefaux, K. Donnay, U. Blanke, O. Woolley-Meza, M. Moussaid, A. Johansson, J. Krause, S. Schutte, M. Perc, Saving human lives: what complexity science and information systems can contribute, J. Stat. Phys. (2014) 1–47, http://dxdoi.org/10.1007/s10955-014-1024-9. <http://dx.doi.org/10.1007/s10955-014-1024-9>.
[4] M. Afshar, M. Asadpour, Opinion formation by informed agents, J. Artif. Soc. Soc. Simul. 13 (4) (2010) 5.
[5] I.-C. Morarescu, A. Girard, A model of opinion dynamics for community detection in graphs, in: Second IFAC Worksh. on Dist. Est. and Cont. in Networked Syst., 2010, pp. 251–256.
[6] O. AskariSichani, M. Jalili, Large-scale global optimization through consensus of opinions over complex networks, Compl. Adap. Syst. Model. 1 (1) (2013) 11.
[7] M. Kimura, K. Saito, K. Ohara, H. Motoda, Learning to predict opinion share in social networks, in: AAAI, 2010.
[8] C. Castellano, S. Fortunato, V. Loreto, Statistical physics of social dynamics, Rev. Mod. Phys. 81 (2) (2009) 591.
[9] R.A. Holley, T.M. Liggett, Ergodic theorems for weakly interacting infinite systems and the voter model, Ann. Prob. (1975) 643–663
[10] P. Clifford, A. Sudbury, A model for spatial conflict, Biomet 60 (3) (1973) 581–588.
[11] H. Follmer, Random economies with many interacting agents, J. Math. Econ. 1 (1) (1974) 51–62.
[12] W.B. Arthur, Increasing Returns and Path Dependence in the Economy, University of Michigan Press, 1994.
[13] A. Orlan, Bayesian interactions and collective dynamics of opinion: Herd behavior and mimetic contagion, J. Econ. Behav. Org. 28 (2) (1995) 257–274.
[14] B. Latané, A. Nowak, Self-organizing social systems: Necessary and sufficient conditions for the emergence of clustering, consolidation, and continuing diversity, Progr. Commun. Sci. (1997) 43–74
[15] S. Galam, Rational group decision making: a random field ising model at t= 0, Phys. A Stat. Mech. Appl. 238 (1) (1997) 66–80.
[16] G. Weisbuch, G. Boudjema, Dynamical aspects in the adoption of agri-environmental measures, Adv. Compl. Syst. 2 (01) (1999) 11–36.
[17] R. Axelrod, The dissemination of culture a model with local convergence and global polarization, J. Conflict Resolut. 41 (2) (1997) 203–226.
[18] A. Alaali, M.A. Purvis, B.T.R. Savarimuthu, Vector opinion dynamics: an extended model for consensus in social networks, in: IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology, 2008. WI-IAT'08., 3, IEEE, 2008, pp. 394–397.
[19] G. Huang, J. Cao, G. Wang, Y. Qu, The strength of the minority, Phys. A Stat. Mech. Appl. 387 (18) (2008) 4665–4672.
[20] M. Perc, Evolution of cooperation on scale-free networks subject to error and attack, New J. Phys. 11 (2009) 033027.
[21] M. Perc, P. Grigolini, Collective behavior and evolutionary games: an introduction, Chaos, Solitons Fractals 56 (2013) 1–5.
[22] M. Perc, J. Gómez-Gardeñes, A. Szolnoki, L.M. Floría, Y. Moreno, Evolutionary dynamics of group interactions on structured populations: a review, J. R. Soc. Interface 10 (80) (2013) 20120997.
[23] A. Szolnoki, M. Perc, Z. Danku, Making new connections towards cooperation in the prisoner's dilemma game, EPL 84 (5) (2008) 50007. http://stacks.iop.org/0295-5075/84/i=5/a=50007.
[24] S. Wasserman, K. Faust, Social Network Analysis: Methods and Applications, vol. 8, Cambridge University Press, 1994.
[25] D. Kempe, J. Kleinberg, E. Tardos, Maximizing the spread of influence through a social network, in: Proceedings of the Ninth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, ACM, 2003, pp. 137–146.
[26] D. Chen, L. Lü, M.-S. Shang, Y.-C. Zhang, T. Zhou, Identifying influential nodes in complex networks, Phys. A Stat. Mech. Appl. 391 (4) (2012) 1777–1787.
[27] K. Sznajd-Weron, J. Sznajd, Opinion evolution in closed community, Int. J. Mod. Phys. C 11 (06) (2000) 1157–1165.
[28] G. Deffuant, D. Neau, F. Amblard, G. Weisbuch, Mixing beliefs among interacting agents, Adv. Comp. Syst. 3 (1–4) (2001) 87–98.
[29] G. Deffuant, F. Amblard, G. Weisbuch, T. Faure, How can extremism prevail? a study based on the relative agreement interaction model, J. Artif. Societ. Soc. Simul. 5 (4) (2002).
[30] R. Hegselmann, U. Krause, Opinion dynamics and bounded confidence models, analysis, and simulation, J. Artif. Soc. Soc. Simul. 5 (3) (2002) 2.
[31] D. Stauffer, The sznajd model of consensus building with limited persuasion, Int. J. Mod. Phys. C 13 (03) (2002) 315–317.
[32] F. Caruso, P. Castorina, Opinion dynamics and decision of vote in bipolar political systems, Int. J. Mod. Phys. C 16 (09) (2005) 1473–1487.
[33] G. Weisbuch, G. Deffuant, F. Amblard, J.P. Nadal, Interacting Agents and Continuous Opinions Dynamics, Springer, 2003. pp. 225–242.
[34] S. Galam, Sociophysics: a review of galam models, Int. J. Mod. Phys. C 19 (03) (2008) 409–440.
[35] G. Deffuant, Comparing extremism propagation patterns in continuous opinion models, J. Artif. Societ. and Soc. Simul. 9 (3) (2006).
[36] A.H. Rodrguez, Y. Moreno, Effects of mass media action on the axelrod model with social influence, Phys. Rev. E 82 (1) (2010) 016111.
[37] A. Flache, M.W. Macy, Local convergence and global diversity from interpersonal to social influence, J. Conflict Resolut. 55 (6) (2011) 970–995.
[38] M. Jalili, Effects of leaders and social power on opinion formation in complex networks, Simul.
[39] A.-L. Barabasi, R. Albert, Emergence of scaling in random networks, Science 286 (1999) 5009–5012.
[40] J. Leskovec, J. Kleinberg, C. Faloutsos, Graphs over time: densification laws, shrinking diameters and possible explanations, in: Proceedings 11th ACM SIGKDD International Conference on Knowledge Discovery in Data Mining, ACM, 2005, pp. 177–187.
[41] D.J. Watts, S.H. Strogatz, Collective dynamics of 'small-world' networks, Nature 393 (1998) 440–442.
[42] P. Erdős, A. Rényi, On the evolution of random graphs, Magy. Tud. Akad. Mat. Kutató Int. Közl 5 (1960) 17–61.
[43] M. Babaei, H. Ghassemieh, M. Jalili, Cascading failure tolerance of modular small-world networks, IEEE Trans. Circuits Syst. II Express Briefs 58 (8) (2011) 527–531.
[44] J. Leskovec, Stanford network analysis package (snap), URL <http://snap.stanford.edu>.
[45] R. Milo, N. Kashtan, S. Itzkovitz, M. Newman, U. Alon, On the uniform generation of random graphs with prescribed degree sequences, ArXiv preprint cond-mat/0312028.
[46] J.L. Payne, M.J. Eppstein, Evolutionary dynamics on scale-free interaction networks, Evol. Comput. IEEE Trans. 13 (4) (2009) 895–912.
[47] D.A. Pechenick, J.L. Payne, J.H. Moore, The influence of assortativity on the robustness of signal-integration logic in gene regulatory networks, J. Theor. Biol. 296 (2012) 21–32.