Interest aware influential information disseminators in social networks

Santhoshkumar Srinivasan[1] · L. D. Dhinesh Babu[1]

Abstract

In recent days, finding influential disseminators in social networks has become a crucial issue due to its importance in the spread control of information, product advertisement, and rumor control. Most current researches on influencer identification are focused on topological factors such as coreness, centrality, and degree distribution. But these methods do not consider the interest of receiver though it plays a vital role in carrying information forward. To consider the receiver's interest in finding influential spreaders, this paper proposes a robust and reliable two-step influencer finder model which considers the individual's interest on spreader as well as the spreading information. This approach combines the individual's location and the interest on the neighbor/topic. In step 1, a novel method to find a trust vertex of spreaders is proposed. In step 2, a weighted neighborhood centrality method is proposed to identify one or more influential spreaders using the trust vertex. The experiments conducted on six different datasets to prove the effectiveness of the proposed approach. The results show that the proposed approach is better than other recent and well-known state-of-the-art algorithms.

1 Introduction

The proliferation of social media applications and ample innovation on new age devices such as IoT devices and smartphones has led to the increased research interest on real-time analysis of information diffusion in social networks. Also, information spread in social networks is faster due to the flow of novel information from various sources with diverse viewpoints [1]. In this type of networks, the influence of some spreaders is often hinder or rapidly increase the spreading of Information on an unprecedented scale. So, understanding spreader's influence on receivers has become crucial to study or control the information diffusion.

Influential spreaders are participants in social networks who can spread information/message rapidly among other participants. Identifying such most significant spreader in social media networks has been a critical factor in: (i) fast spreading the vital information and/or (ii) drastically reduce the rumor spreading.

At any point in time, a spreading process will have one or more influential spreaders across the communities or sub-networks. Finding u number of the influential spreader in a network can be defined as [2]:

$$f : S \rightarrow \text{Influenced Recivers} \forall SP$$
$$IN = \{u | \max(f(u)) \forall SP\} \tag{1}$$

In Eq. (1), SP is a spreading process, u is a set of influential originating nodes which can maximize the spreading process.

From Eq. (1), an influential spreader is having a higher probability of information acceptance by the neighbors. Hence, Information spreads faster through influential originators compared to the normal originators. The issue

✉ Santhoshkumar Srinivasan, santhoshramuk@gmail.com; L. D. Dhinesh Babu, lddhineshbabu@gmail.com | [1]School of Information Technology and Engineering, Vellore Institute of Technology, Vellore, India.

of finding a *u*-set of participants who can lead to maximal spread is an NP-hard optimization problem [3].

Most of the influencer identifiers, discussed in Sect. 2, focus on topological factors such as centrality, coreness, and h-index. Topological factors clarify the information flow pattern within the network. But these are not enough since such factors do not consider the spreader or receiver's temporal behavior in the spreading process. In social networks, temporal behaviors like opinions of individual on neighbors affect the individual's participation in information diffusion. To the best of our knowledge, no model has been proposed to identify the influential spreaders by considering the temporal aspects. This paper focuses on evaluating the importance of a person with his/her neighbors by considering both topological and temporal aspects of the network.

In a typical social network, once the receiver receives the message, he/she decides to spread or ignore the message. If the person decides to spread, the receiver becomes the spreader; otherwise, the message will be ignored. The same approach follows henceforth. This is generally called as spreading process. The interest of the receiver decides whether the person receives or ignores the message from the spreader. Therefore, the interest of the receiver on the spreading topic or spreader is the most important. Finding the impact of receiver's interest in the spreading process is one of the important challenges in a dynamic social networking system. Receiver's interest on topic or spreader changes over time in any dynamic social network. Hence, an Individual's interest should be updated in a dynamic fashion. The proposed approach considers an individual's interest on the topic as well as the person while finding influential spreaders.

We study the interest of interacting participants using opinion dynamics. Opinion dynamics is one of the widely studied temporal social behavior that impacts the individual persons in any social network. The opinion of a person in a network can be affected by their neighbors' opinion. So, the person updates the opinion after each interaction with neighbors. Complex networks like Social networks can be attributed with continuous opinion update model [4]. A continuous update model can be categorized as linear and non-linear models [5]. In non-linear update model, first a confidence level value is set, and an individual will perform the synchronous update only when an average of all opinions is lower than the confidence level. In linear update model, the average of all opinions is considered as the weight of the individual.

In this paper, we propose a novel influential nodes finder measure, Interest Aware Neighborhood Centrality (IANC), to quantify and rank the spreading ability of one or more nodes using the trust values of neighbors. This is a two-step framework to rank influential nodes by

acquiring topological and temporal aspects of social networks. The proposed approach first calculates the receiver interest after each interaction and consider the calculated value as the weight of an individual. We use a continuous, non-linear opinion update model inspired by the *Hegselmann-Krause* model [4, 6, 7] for this calculation. This is a completely localized approach which updates a pair of participants who involves in interaction at any given time. Then we rank influential spreaders based on a novel topological measure called Neighborhood Opinion Centrality (NPC) model. The global algorithm, NPC, is applied at the network level to rank the influential spreaders. The contribution of this paper is three-fold: (i) We provide a domain-based opinion update model that characterize the communication of the network (ii) We formulate a matrix of opinions for each node in the network that acts as weight of the node (iii) Then we use this weight to identify the *u*-influential spreaders who can disseminate information in quick fashion.

The paper starts by discussing the various models proposed so far on identifying the influential spreaders. Next, in Sect. 3, we define the nature of the proposed model and assumptions. Section 4, elaborates the proposed approach and explain how it identifies the influential spreaders. In Sect. 5, we will discuss the experimental approach and results and compare with existing models. This paper ends with conclusions and future works, in Sect. 6.

2 Related works

The spreading maximization process can be characterized by the topological structure of the network and interaction dynamics among participants in the network [8, 9]. Most of the models which predict the number of influencers is based on the topological structure of the network. Various classical centrality measures have been proposed to identify the influential spreaders, such as degree centrality, betweenness centrality [10], closeness centrality [11], k-shell coreness [12]. Among these, k-shell centrality [12] has been widely accepted and first majorly applied method for finding the influential spreader. But this method does not suit the problem for various reasons: 1. It tends to find core k set of influencing nodes with spreading capacity overlaps each other, 2. It does not consider the impact of neighbors of influential spreaders. Naturally, neighbors of influential spreaders are also having a significant impact on the spreading process.

Many different types of models have been proposed on the shortcomings of k-core decomposition such as k-hop [13], k-truss [14], Distributed k-core decomposition [15], IC and k-shell method for weighted networks [16]. Liu et al. [17] have shown the accuracy of k-shell decomposition

can be improved by removing the redundant edges from the network. PageRank [18], LeaderRank [19] and HITS [20] are proposed to rank the nodes in the directed network. But, coreness of a network does not accurately define the higher influential nodes as the low-core nodes act as bridge elements that control the spreading process [21].

Kemp et al. [3] proved that the problem of identifying the optimal influencers is NP-hard problem and proposed a greedy algorithm. But this algorithm is not suitable for complex and high-density networks as the computation cost is high. Succeeding to that effort, there have been various greedy/heuristic based algorithms proposed such as CELF [22], DDH [23], DiDH [23] and IC [24]. CELF is a greedy algorithm, proposed by Dk et al., selects the seed nodes to reduce the running complexity. But, it does not improve much in terms of computational overhead. Degree discount heuristic (DDH) is a heuristic based approach where selected seed node's degree should be discounted from neighbors' degree while calculating the influence maximization. Sankar et al. used Diffusion degree heuristic (DiDH) for influence maximization in large-scale complex social networks. Independent cascade (IC) model is a k-medoid clustering heuristic-based model that acts on community structured networks.

Neighborhood coreness centrality introduced by Bae et al. [25], considers the k-shell coreness index of neighbors to identify the important nodes. Ma et al. [26] proposed a neighborhood centrality-based model by the inspiration of gravity formula. Neighborhood centrality outperforms the other centrality methods in most of the cases [27].

All above researches are considering the static properties of the network in identifying the maximum spreader. But information spread in any complex network is also based on the interest of receivers. The opinion of the receiver on spreader and information plays a vital role in the spreading process. We are addressing the problem of influential spreaders by considering the opinion of participants in the network. The opinion of a participant keeps changing at every contact with the neighbor as the opinion is time bounded based on communication between participants. The opinion of an individual is a continuous value based on various interactions with neighbors in the network. So, updating the opinion of each person is a local heuristic.

Basically, opinion models are used to study the convergence or divergence of social behavior in social networks. Many different models have been emerged to study the phenomenon. Complex networks use local belief update models to characterize the interaction pattern in the network. Local belief update is broadly classified into a. linear update and b. non-linear update models. In linear update model [28, 29] the weight for individuals is calculated based on their relative importance in the network. In non-linear update models [30], a confidence level is set, and the synchronous interactions update the belief only when the average opinion value is above the confidence level. These models are widely called the Hegselmann-Krause model.

Non-linear update models are well suited for social networks, due to the complex nature of these networks. The first non-linear model was introduced and analyzed by Krause et al. [6]. Deffuent et al. [31]. has extended the model stating the agents can communicate with their immediate four neighbors. Weisbuch [32] has analyzed the convergence property on scale-free networks. The opinion dynamics discussed so far considers the opinion of an individual on a neighbor is the same for all the topics. But a person can have a different kind of opinions with the same neighborhood on different topics. So, the proposed approach considers the domain-specific opinion of an individual that is being updated after each interaction with other participants in the social network.

3 Model environment

We focus on unweighted and undirected networks to examine the proposed approach, IANC. For opinion modeling in step 1, we consider the network as a social interaction network. For weighted neighborhood modeling, we consider the network as SIR model. Let $G_c = \{\gamma, \delta_c\}$ be the connected graph network, where $\gamma = \{1, 2, 3, \dots n\}$ be the set of nodes in the networks and δ_c be the set of edges of graph G_c. We denote $n = |\gamma|$ and $m = |\delta_c|$ as the number of nodes and edges respectively. N_j is the set of neighbors of the node j. i.e., $k_j = |N_j|$ is the degree of node j. Interaction takes place in a random fashion in this connected graph.

3.1 Hegselmann–Krause (HK) model

In this section, the social interaction network for step 1 is explained. To update the opinions of individuals after each interaction, we follow non-linear opinion update model called *Hegselmann-Krause* (HK) model: A finite number of individuals synchronously update their opinion based on the possible interactions among them. The opinion can be updated only if that differs by less than a confidence level €. In this section, we describe the discrete synchronous opinion update model. Let us assume that we model the interactions of nodes of G_c that updates their opinion synchronously. It is assumed that each time step t = 0,1,2,.., the opinion of node $i \in [\gamma]$ can be represented by a vector $O_i(t) = [O_1, O_2, \dots, O_q]$. Here q is the possible number of outcomes of any interaction in the network. Neighborhood function with opinion vectors can be defined as,

$$N_i(t) = \{j \in [\gamma] \mid \left| O_i(t) - O_j(t) \right| > \epsilon\}$$

With this, the opinion update model can be defined as,

$$O_i(t+1) = \frac{1}{|N_i(t)|} \sum_{j \in N_i(t)} O_j(t) \qquad (2)$$

This model is based on the idea that the topological connection has a lesser influence on the information dissemination. The interaction between two persons is based on the common interest they possess. Though a person closely connects with the neighbor, if the interest differs by more than a threshold, communication between the pair does not occur or the possibility of interaction is negligible [5].

3.2 SIR epidemic model

We follow Susceptible–Infected–Recovered (SIR) model to study the spreading process for calculating NPC in step 2. Recently, the SIR model is widely used in network related researches on studying information and rumor spreading in social interaction networks compared to other models such as SIS, Independent Cascade [16, 33]. Nodes in SIR model belongs to any of three states: susceptible (S), infected (I) and recovered (R). S set of nodes are susceptible to information, I set of nodes are spreading information to susceptible nodes and R set of nodes are recovered from the information spreading process and not affected by the information anymore. At any given time t, $S(t) + I(t) + R(t) = \gamma$. Figure 1 represents the model. At time 0, all the nodes in the network are in susceptible set S except only one node act as Spreader I. As time increases, the I nodes infect their neighbors at rate β. The infected nodes become recovered at a recovery rate ¥. At any given time, t, S(t) denotes the number of susceptible nodes, I(t) denotes the number of Infected nodes and R(t) denotes the number of recovered nodes and the total number of nodes in this model is $S(t) + I(t) + R(t) = \gamma$.
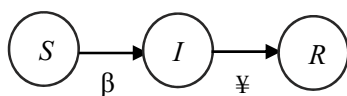
4 Interest aware neighborhood centrality

In this paper, we propose a novel influential nodes finder measure called Interest aware neighborhood centrality (IANC) as a two-step approach. The two-step measure shown in Fig. 2 is proposed as an amalgam of network topology and temporal aspects of the network. In step 1, we propose a weight update model based on the interaction between any pair of nodes in the network. This process of updating the opinion weight of nodes occurs at the local level during each and every communication.

In step 2, the proposed approach uses this node weight in a novel node weighted neighborhood centrality measure called Neighborhood Opinion Centrality (NPC) to rank the influential spreaders in the network.

NPC is a global centrality measure combines weights of each node and degree centrality.

4.1 Neighborhood opinion distance calculator step

In social networks, a participant of information diffusion process might be interested in neighbors and/or the topic the neighbor shares. This interest plays a major role in the spreading of the information to neighbors in the network. As already discussed, a person decides to spread the information based on the interest on two major criteria: (i) The neighbor who passed the information to him/her, and (ii) the topic on which the information is being shared. Opinion dynamics are used to measure the interest factor of an individual in interaction networks. In general, a person neither completely agree nor disregard the opinion of the neighbor [6]. In the proposed model, we first calculate the opinion matrix of individuals i.e., nodes in the network and consider this
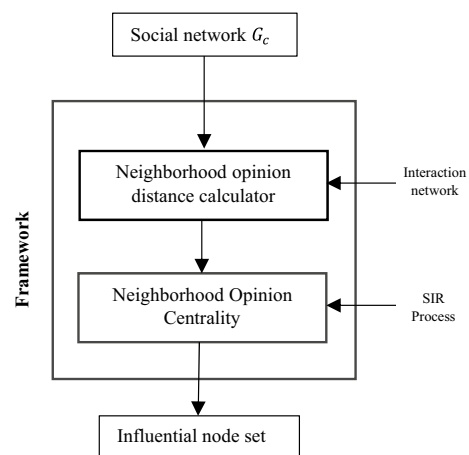

Fig. 1 SIR epidemic model

Fig. 2 Interest aware neighborhood centrality framework

opinion as the weight of the node and rank the influential person by using weighted neighborhood ranker method.

The opinion matrix is based on the finite set of common topic domains every individual possesses. The proposed opinion update model is inspired by the famous non-linear continuous opinion update model called the *Hegselmann–Krause* (HK) model. This opinion update model acts during every interaction between any pair of nodes in the network. After each interaction, the opinion matrix of both the individuals is updated.

An individual in the network might be more interested in one or more topic domains. The topic domain of an individual j in the network is a vector of p dimension $d^j = [d_1, d_2, \ldots d_p]$. Here p is a finite number of topic domains we consider in this model. Opinion of an individual j on any topic is a q dimension vector $O^j = [O_1, O_2, \ldots, O_q]$. Here q is a uniform distribution as the possible outcomes of an interaction between any pair of interactions for every topic domain.

$$O_j = \left\{ O_1, O_2, O_3, \ldots, O_q \,\Big|\, \sum_{i=1}^{q} O_i = 1 \,\&\, O_i \int [0,1] \right\} \qquad (3)$$

Here $O_j = 0$ indicates the complete disagreement with other party on the given topic d_i and $O_j = 1$ indicates complete agreement with other party on the given topic d_i. Hence, the opinion of an individual j is denoted as

$$DO^j = d^j \times O^j \qquad (4)$$

It is a $\chi : p \times q$ matrix. This matrix is an adjacent matrix with values representing the outcome of every interaction of an individual with neighbors in topics. DO_d^j is opinion vector of topic d of node j.

The opinion of each node is updated on every communication on the respective topic domain. We measure this opinion factor as neighborhood opinion distance (NPD). The NPD between any pair of nodes for topic d is defined as,

$$NPD_d(i,j) : DO_d^j.DO_d^j \rightarrow R^+ \qquad (5)$$

NPD updates if threshold τ is larger than the distance between the pairs. We assume threshold τ is a distance threshold value for every topic domain beyond which the communication does not take place.

From NPD value in Eq. (5), we derive the trust value of a node j as ρ. This is a non-increasing function of NPD.

$$P_d^j = 1 - \frac{\sum_{i \in Neighbours}^{N} NPD_d(i,j)}{N} \qquad (6)$$

At any given time, the weight of a node j in the network is measured with the trust value vector

$\rho^j = \left\{ \rho_1^j, \rho_2^j, \rho_3^j, \ldots, \rho_p^j \right\}$ respective to the topic domain vector $d^j = [d_1, d_2, \ldots d_p]$. i.e.,

$$\rho^j = \{ \rho_1^j, \rho_2^j, \rho_3^j, \ldots, \rho_p^j \,|\, \rho_i^j \le 1 \} \qquad (7)$$

where ρ_d^j is trust for $d_i^j \rho^j$ in Eq. (7) is a p dimensional vector representing the weight of node j. Finally, we get the Weight of a node is ρ_d^j is a continuous value and keep updating with respect to time.

4.2 Neighborhood opinion centrality model

Neighborhood Opinion Centrality model considers the opinion on different topics and degree of neighbors to quantify the spreading ability of nodes. This model is derived based on the assumption that contribution of neighbors in information spread is vital to increase the flow. The basic idea is that spreader with more trust value neighbors is having the ability to spread faster than other spreaders. Time-bounded trust value is considered as the weight of each node. The proposed model also focuses on topological factors such as degree distribution and centrality of a node. In contrary to other influential spreader-finding methods, this method considers topic domain while finding trust value of a node. So, the proposed method identifies different sets of influential spreaders specific to the topic domain. This algorithm applies at the global level.

In step two, trust vector ρ_j and degree centrality is considered to analyze the global features of the network. A high influential node indicates that it has high trust value among its neighbors and it is surrounded by the nodes which can spread the message faster. This model is described in below algorithm.

Algorithm − **Neighborhood Opinion Centrality model:**

Input: $G_c = \{\gamma, \delta_c\} \,|\, \gamma = \{1,2,3,\ldots n\}$ *nodes* $\&$ δ_c *set of edges*
Every node j with trust value vector $\rho^j = \{\rho_1^j, \rho_2^j, \rho_3^j, \ldots, \rho_p^j\}$ *for the topic domains*
Output: $I = \{I^1, I^2, I^3, \ldots, I^u \,|\, I \in \gamma \,\&\, |I| \le |\gamma|\}$
For every node j in G_c:
    For every topic domain d:
        For all neighbors i of node j:
            $di_{ij} = k_i.k_j$
            $di_d^j = (di_{ij}/\langle di \rangle)^\alpha$
            $IF_d^j = IF_d^j + di_d^j$
        END
        Influence factor $G(j,d) = \vartheta.IF_d^j + \rho_d^j$
    END
END
Sort nodes by $G(j,d)$ in descending order
Return first u nodes as influential nodes $I_d = \{I^1, I^2, I^3, \ldots, I^u\}$

Here α is equilibrium factor, in this case, we set this value as 1 and di is average degree coefficient. It is defined as,

$$\langle di \rangle = \frac{\sum_{i \ f \ Neighbors}^{N} k_i . k_j}{N} \tag{8}$$

To provide equal weight to degree centrality and trust value, we introduce a normalization factor ϑ. ϑ normalizes the degree influence value IF_d^j to be a fractional value, i.e.,

$$\vartheta.IF_d^i \leq 1 \tag{9}$$

Finally, I_d nodes in Eq. (10) represent the set of influential nodes by the rank for topic d.

$$I_d = \left\{ l^1, l^2, l^3, \dots, l^u \right\} \tag{10}$$

5 Experimental results and discussions

We considered six social networking datasets from public large dataset library SNAP [34] for the experiments to prove the effectiveness of the proposed model. The proposed model is compared with state-of-the-art and recent influential finder approaches and models. The experiments are performed on a server with 16 GB ram and a 4.0 GHz octa-core processor running 64-bit Python 3.6. We have used NetworkX [35] to generate and load the datasets as the large graph. For the experiments in step 1 and step 2, we have simulated the SIR epidemic model. The opinion of receivers on spreaders are randomly chosen with values ranging from 0 to 1 during the interaction between them. Throughout the evaluation, we assume the neighbors interact on only one topic for every dataset and the spreading rate of infected is set to be 1. i.e., the infected node can infect maximum of 1 person and goes to the recovered state. The evaluation results presented at Sect. 5.2 are averaged at least for fifty runs. A single run ends only after a minimum 80% of the population goes into the recovered state.

5.1 Datasets

To evaluate the effectiveness and efficiency of the proposed method, we apply it to six standard real-world networks. These real networks are all undirected and unweighted, investigated on SIR spreading process. Topological features of these networks are summarized in Table 1.

5.2 Results

We have compared recent as well as well-known influential finders such as degree centrality (k), k-shell decomposition (ks), and Weighted neighborhood centrality (Cnc+) with the proposed influential finder IANC. In every comparison, we have used box-plot to quantify and visualize the result distribution in all the algorithm. Box-plot helps to visualize the distribution of a set of values.

To measure the effectiveness of spreading capability of the proposed two-step influential finder, we use the imprecision function [38, 39]. The imprecision function is used to quantify the importance of influence finder measure. Imprecision function estimates the spreading efficiency of a node chosen as the originator. We can identify the importance of the proposed measure by estimating chosen nodes spreading ability. Imprecision function is defined as,

$$\varepsilon(u) = 1 - \frac{S(u)}{S_{\max}(u)} \ \dots \tag{11}$$

Here, u is the set of nodes identified as influential spreaders, $S(u)$ is the average spreading efficiency of u nodes and $S_{\max}(u)$ is the highest spreading efficiency in u nodes. A lower value indicates the efficiency of the method.

(a) Dolphin
(b) Karate Club
(c) PowerGrid
(d) ca-HepPh
(e) ca-Condmat
(f) ego-Twitter

Table 1 Dataset for experiments

	n	E	<k>	H	β_{th}	β
Karate Club	34	78	4.588	1.6895	0.129	0.242
Dolphins	62	159	5.129	1.3268	0.1723	0.2351
PowerGrid	4,941	6,594	2.669	1.4522	0.258	0.491
ca-HepPh	12,008	118,521	19.74	5.9823	0.01	0.02
ca-Condmat	21,363	196,972	22	2.99	0.02	0.035
ego-Twitter	81,306	1,768,149	42.02	3.45	0.134	0.15

Here <k> is average degree, H = <k^2>/<k>^{2}—Degree Heterogenicity index [36], β_{th} = <k>/<k^2>—epidemic threshold [37]

Figures 3, 4, 5, 6, 7, and 8 are the comparison of imprecision with various algorithms. We have analyzed the algorithms with influence node proportion between 0.01 and 0.2. The proposed algorithm outperforms other algorithms in all proportional values for the bigger network, ca-Condmat, and ego-Twitter. This algorithm gives better results when the number of people in the network is huge comparing to smaller networks. i.e., This implies an individual's opinion gives the bigger difference in finding influential nodes for such networks. Overall, in all network studies, imprecision of the proposed method is at a lower level. This proves that IANC outperforms the other algorithms studied.

Next, to quantify the correctness of the proposed method, we use Kendell's tau [40] as a rank correlation coefficient method. Kendall's tau is a correlation metric that identifies the relationship between two set of values. This method quantifies the correlation between two different sets in spreading ability. A high tau value indicates the performance of the algorithm is better.

(a) Dolphin
(b) Karate Club
(c) PowerGrid
(d) ca-HepPh
(e) ca-Condmat
(f) ego-Twitter

Figures 9, 10, 11, 12, 13, and 14 show the Correlation between SIR Epidemic model and various algorithms. Infection rates 0.01–0.15 are considered in this study. We set recovery rate ¥ as 1 to let an infected node to infect
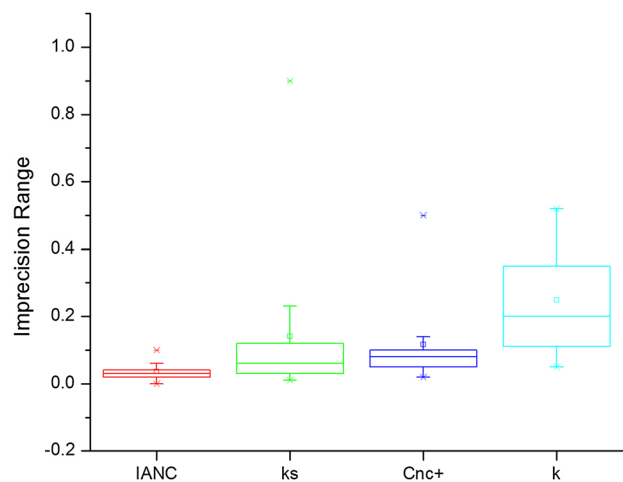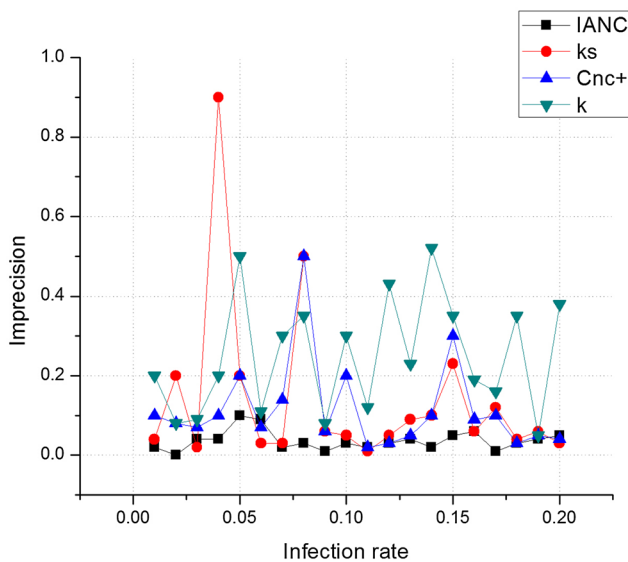
Fig. 3 Dolphin imprecision
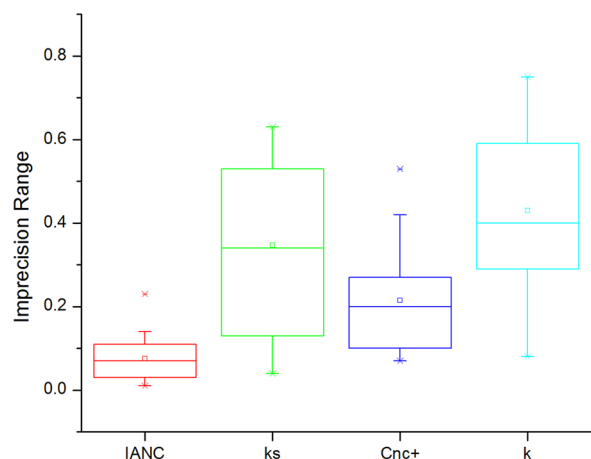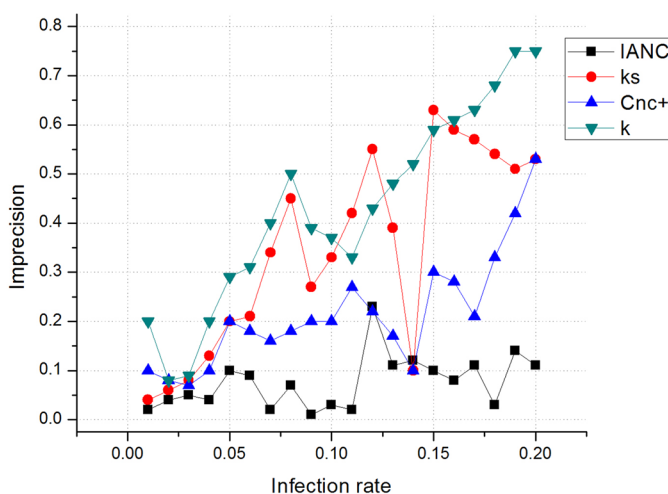
Fig. 4 Karate Club imprecision
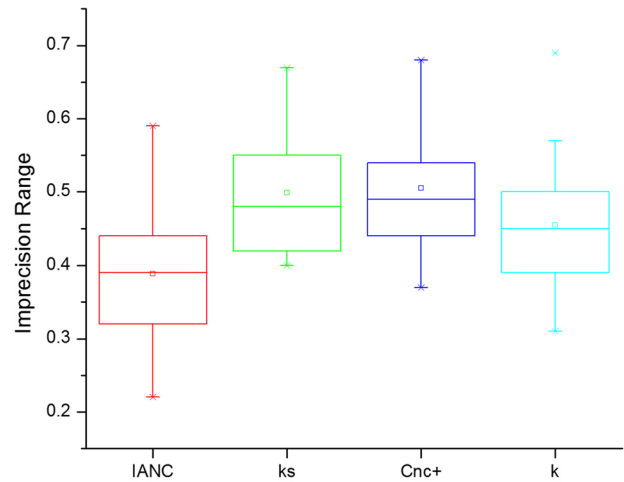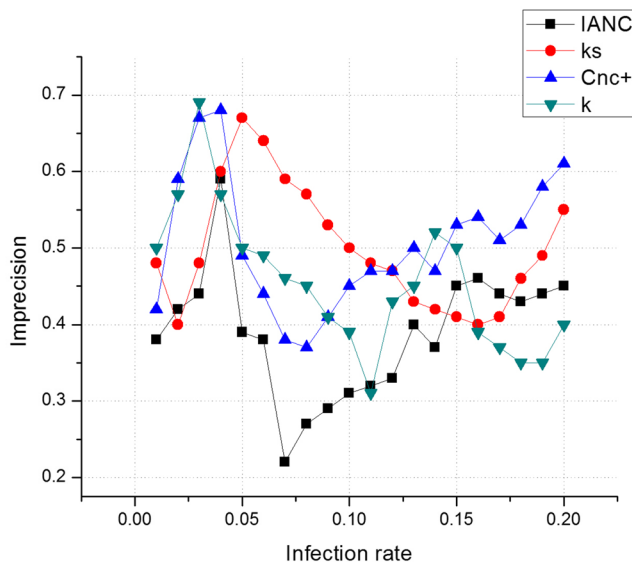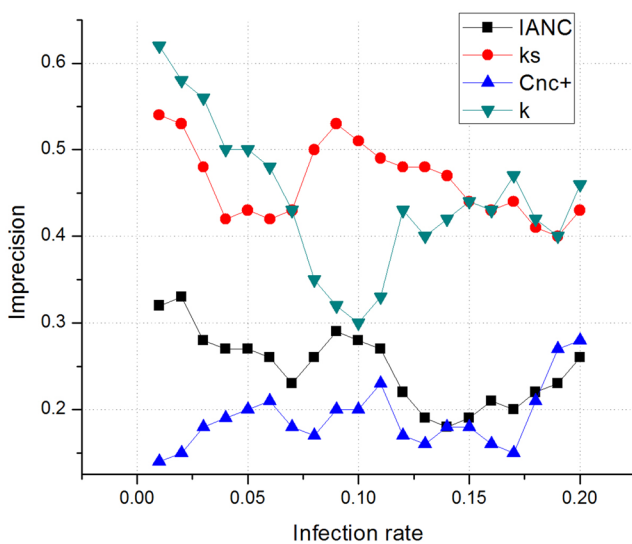
Fig. 5 PowerGrid imprecision
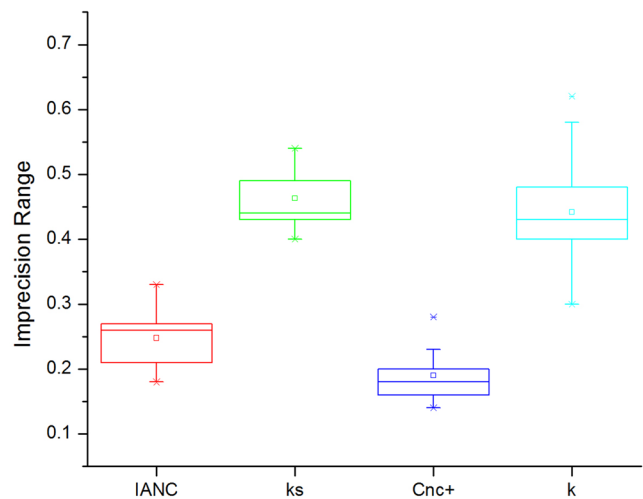
Fig. 6 ca-HepPh imprecision

only immediate neighbors and move to the recovered state. In all studies, IANC outperforms other algorithms. The Positive correlation between SIR and IANC is always high. This proves IANC algorithm gives better results compared to other algorithms studied.

Next, the monotonicity of IANC ranking method is investigated. We rank the influential nodes by their influential factor which is directly proportional to neighbor opinion distance. Nodes with same influential factor are considered to have the same rank. With monotonicity M, we are quantifying the fraction of a tie in the ranking list. Higher monotonicity implies the number of tied ranking

nodes is lesser in the ranking list. The method with higher monotonicity is more robust than other ranking methods. Figure 15 explains the monotonicity of different algorithms. Monotonicity of IANC is higher as the tied influenced nodes are very less compared to other algorithms. The influence originators set identified by IANC is capable of affecting a large number of nodes as the influencers are diverse across different layers/communities of the network.

Imprecision rate, Kendell's tau correlation and monotonicity presented above show the effectiveness of our proposed approach in all six social networks. From the results,
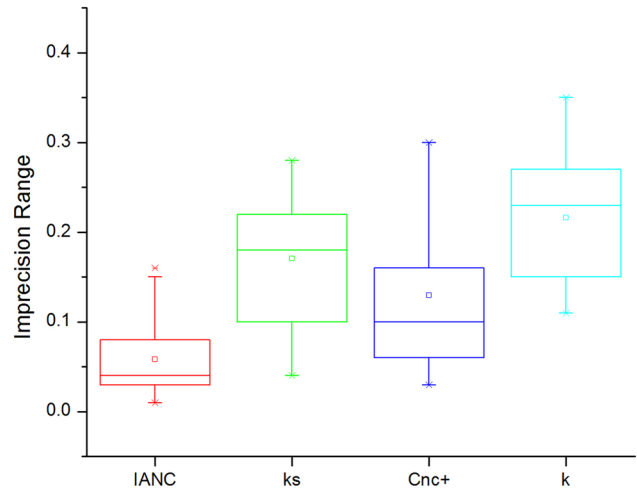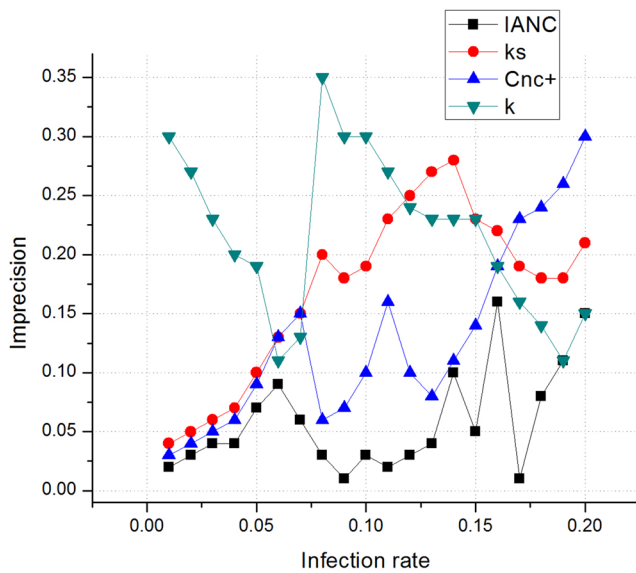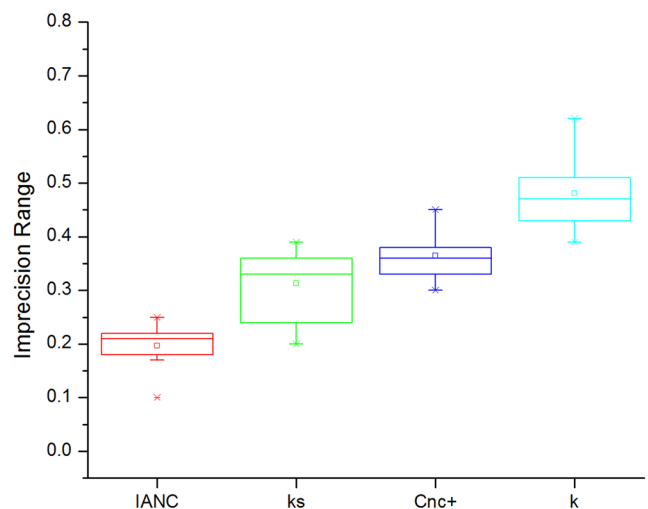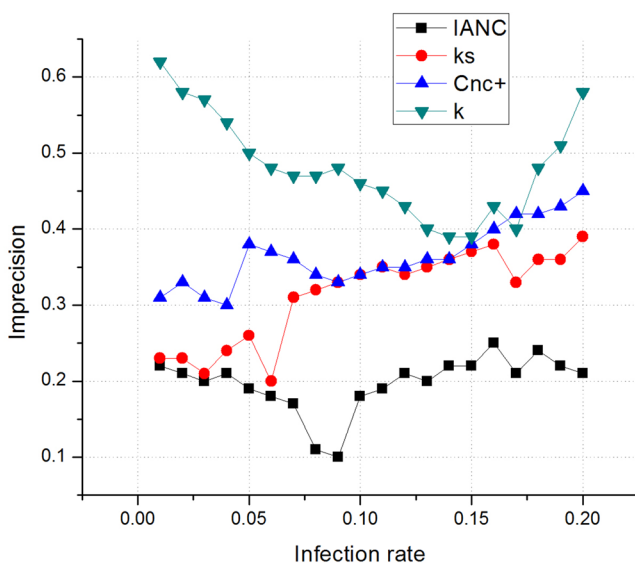
Fig. 7 ca-Condmat imprecision

Fig. 8 ego-Twitter imprecision

IANC provides the most influential spreaders for larger datasets. Such influential spreaders diffuse the information to a maximum number of people at the earliest. IANC perform moderately for smaller datasets when influencing circle of influencers overlaps more. But IANC outperforms state-of-the-art influence finder approaches in such cases also. Considering temporal features along with topological features have shown a greater result in finding influential spreaders.

6 Conclusions

Spread maximization/Control of spreading is an interesting research topic in recent ever increasing social networking environments. Among all spread maximization researches, finding the influence spreaders in a complex network is a challenging research topic. In this paper, a two-step framework to find a set of influential originators in social networks is proposed. The proposed
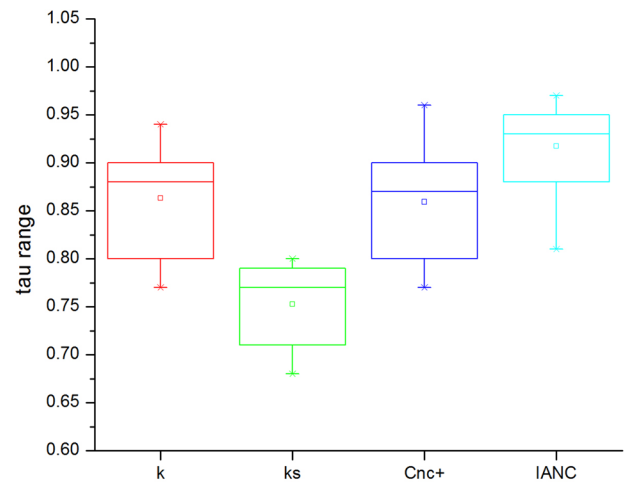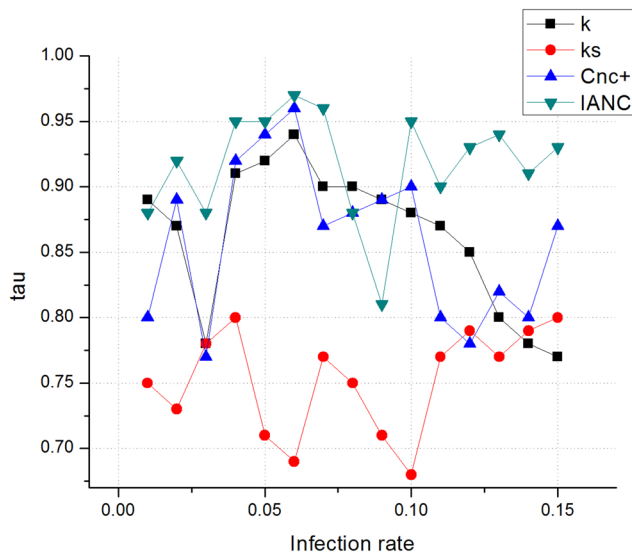
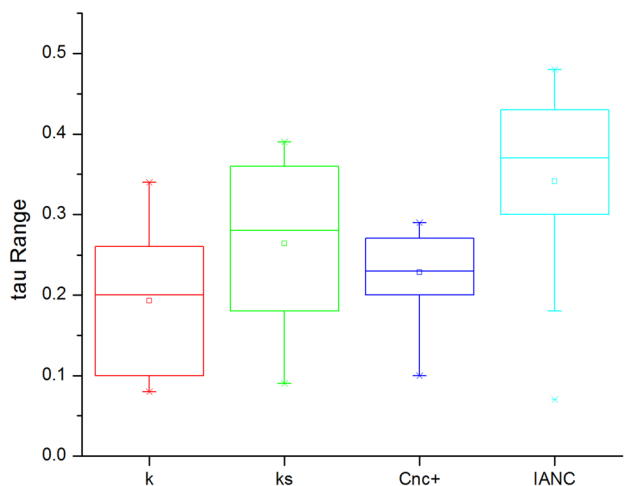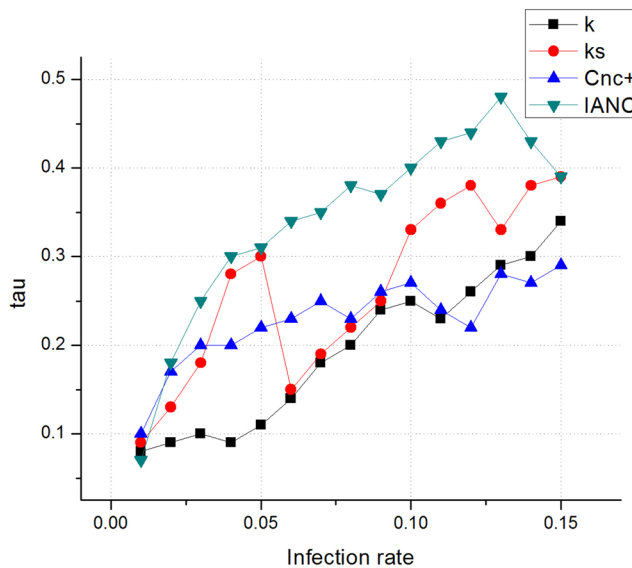**Fig. 9** Dolphin correlation

Fig. 10 Karate Club correlation

approach considered the interest of receivers and their location with neighbors in identifying the influential spreaders. Unlike other methods, the proposed method focuses on temporal aspects of a network along with topological attributes. i.e., This method is an amalgam of opinion update model and neighborhood centrality model. To evaluate the performance of the proposed framework, we apply it on six real-world networks comparing with recent and well-known influential finders. We evaluated imprecision function to quantify the spreading efficiency, Kendall's tau correlation with SIR model to identify the correctness and monotonicity index to measure the ranking resolution. The experimental evaluation shows the proposed model gives a more accurate set of influential spreaders compared to other recent and well-known influential finders.
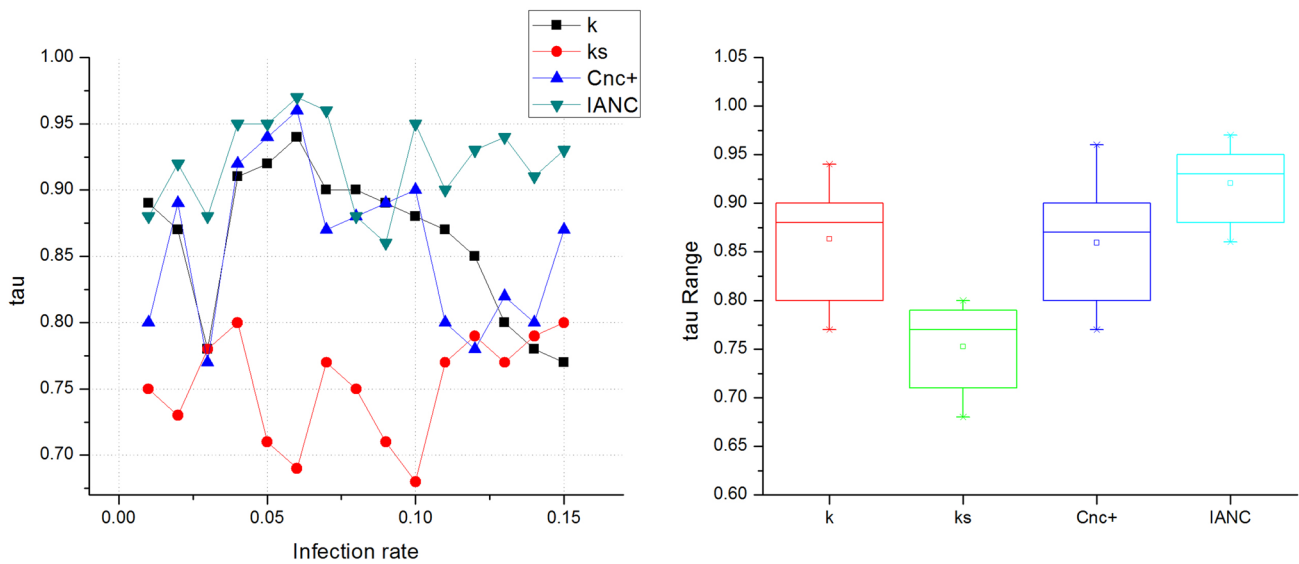
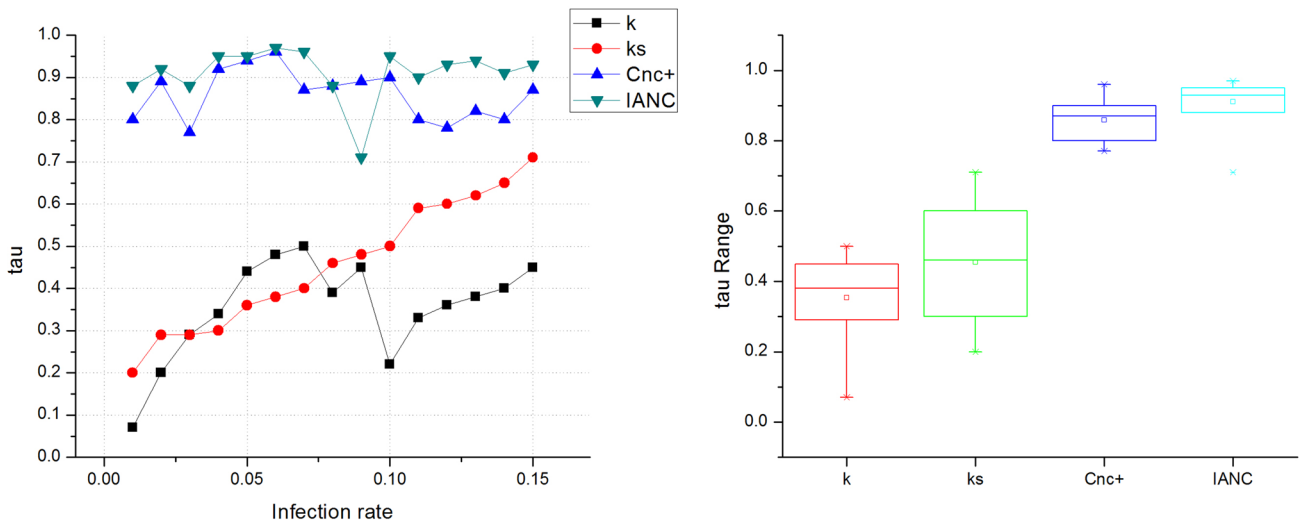**Fig. 11** PowerGrid correlation

Fig. 12 ca-HepPh correlation

This work shed some light on considering temporal aspects of diffusion will provide reliable results in devising effective strategies to identify influential spreaders in social networks. This opens various future research possibilities on considering the temporal and topological attributes of the social network in this area. Other possible future work is to consider homogenous and heterogeneous natures of the network in parallel to identify the important spreader as the experiments in this paper considered heterogeneous nature of networks in this paper.
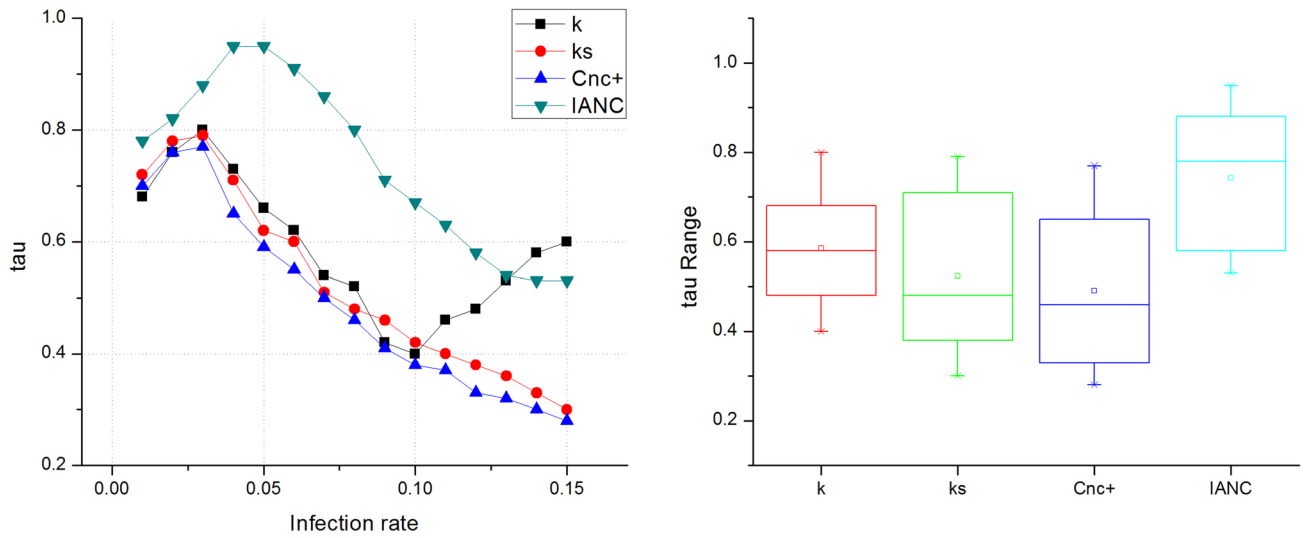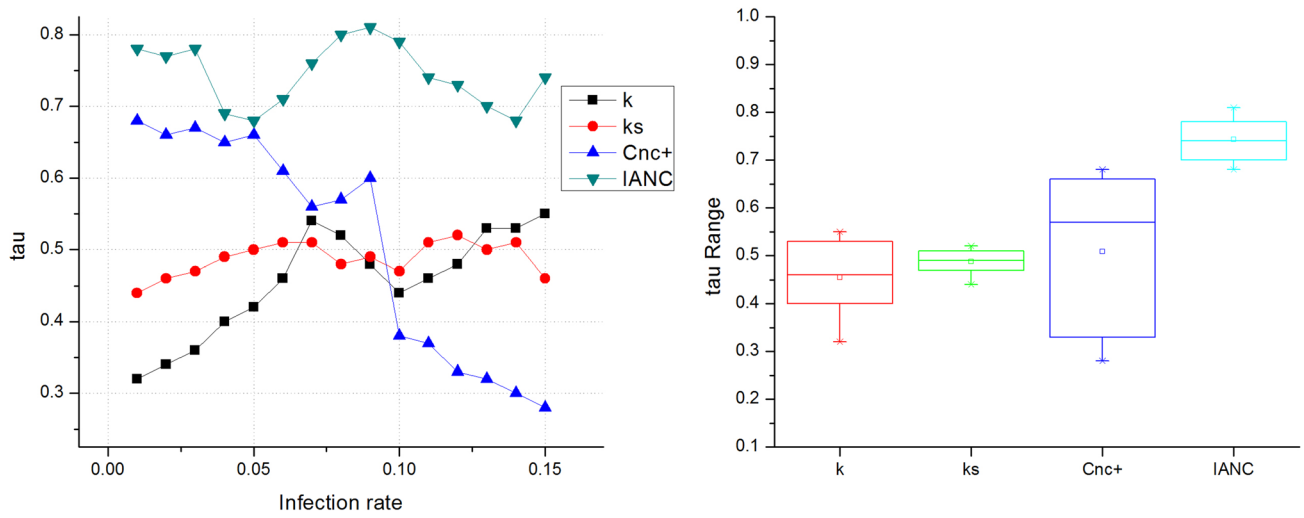
Fig. 13 ca-Condmat correlation

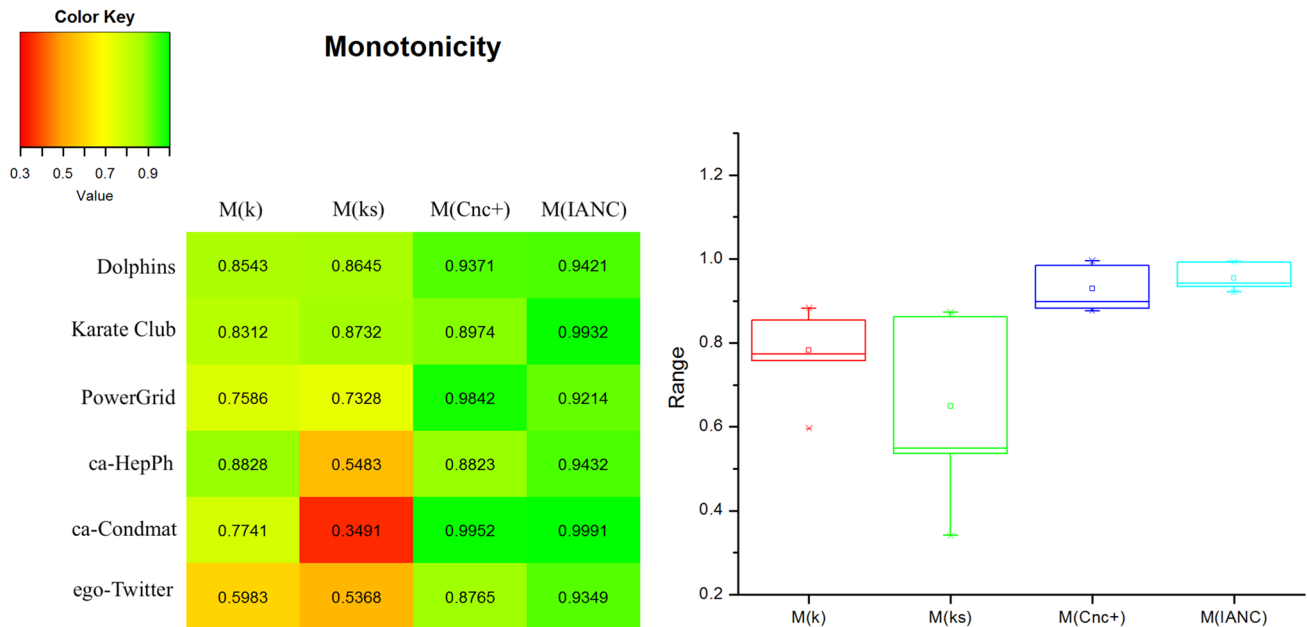

Fig. 14 Ego-Twitter correlation

Fig. 15 Comparison of monotonicity

Compliance with ethical standards

Conflict of interest On behalf of all authors, the corresponding author confirms that there is no conflict of interest.

References

1. Bakshy E, Rosenn I, Marlow C, Adamic L (2012) The role of social networks in information diffusion. In: Proceedings of the 21st international conference on World Wide Web, pp 519–528
2. Pei S, Makse HA (2013) Spreading dynamics in complex networks. J Stat Mech: Theory Exp 2013(12):P12002
3. Kempe D, Kleinberg J, Tardos E (2003) Maximizing the spread of influence through a social network. In: Proceedings of the ninth ACM SIGKDD international conference on Knowledge discovery and data mining, pp 137–146
4. Weisbuch G, Deffuant G, Amblard F, Nadal J-P (2003) Interacting agents and continuous opinions dynamics, Heterogenous agents, interactions and economic performance, pp 225–242
5. Li L, Scaglione A, Swami A, Zhao Q (2012) Phase transition in opinion diffusion in social networks. In: 2012 IEEE international conference on Acoustics, speech and signal processing (ICASSP), pp 3073–3076
6. Hegselmann R, Krause U et al (2002) Opinion dynamics and bounded confidence models, analysis, and simulation. J Artif Soc Soc Simul 5(3):1–33
7. Blondel VD, Hendrickx JM, Tsitsiklis JN (2009) On Krause's multi-agent consensus model with state-dependent connectivity. IEEE Trans Autom Control 54(11):2586–2597
8. Guille A, Hacid H, Favre C (2013) Predicting the temporal dynamics of information diffusion in social networks, arXiv preprint arXiv:1302.5235
9. Guille A, Hacid H, Favre C, Zighed DA (2013) Information diffusion in online social networks: a survey. ACM Sigmod Record 42(2):17–28
10. Freeman LC (1977) A set of measures of centrality based on betweenness. Sociometry 35–41
11. Newman ME (2005) A measure of betweenness centrality based on random walks. Soc Netw 27(1):39–54
12. Batagelj V, Zaversnik M (2011) Fast algorithms for determining (generalized) core groups in social networks. Adv Data Anal Classif 5(2):129–145
13. Cai L, Meng T, He T, Chen L, Deng Z (2017) K-Hop community search based on local distance dynamics. In: International conference on neural information processing, Springer, pp 24–34
14. Cohen J (2008) Trusses: cohesive subgraphs for social network analysis. National Security Agency Technical Report, vol 16
15. Montresor A, De Pellegrini F, Miorandi D (2013) Distributed k-core decomposition. IEEE Trans Parallel Distrib Syst 24(2):288–300
16. Garas A, Schweitzer F, Havlin S (2012) A k-shell decomposition method for weighted networks. New J Phys 14(8):083030
17. Liu J-G, Ren Z-M, Guo Q (2013) Ranking the spreading influence in complex networks. Phys A Stat Mech Appl 392(18):4154–4159
18. Bryan K, Leise T (2006) The $25,000,000,000 eigenvector: the linear algebra behind Google. SIAM Rev 48(3):569–581
19. Lu L, Zhang Y-C, Yeung CH, Zhou T (2011) Leaders in social networks, the delicious case. PLoS ONE 6(6):e21202
20. Kleinberg JM (1999) Authoritative sources in a hyperlinked environment. J ACM (JACM) 46(5):604–632
21. Reppas A, Lawyer G (2012) Low k-shells identify bridge elements critical to disease flow in small-world networks. AIP Conf Proc 1479(1):1426–1429
22. Leskovec J, Krause A, Guestrin C, Faloutsos C, VanBriesen J, Glance N (2007) Cost-effective outbreak detection in networks. In: Proceedings of the 13th ACM SIGKDD international conference on knowledge discovery and data mining, pp 420–429
23. Pal SK, Kundu S, Murthy C (2014) Centrality measures, upper bound, and influence maximization in large scale directed social networks. Fund Inf 130(3):317–342
24. Saito K, Nakano R, Kimura M (2008) Prediction of information diffusion probabilities for independent cascade model. In:

International conference on knowledge-based and intelligent information and engineering systems, pp 67–75

25. Bae J, Kim S (2014) Identifying and ranking influential spreaders in complex networks by neighborhood coreness. Phys A Stat Mech Appl 395:549–559

26. Ma L-L, Ma C, Zhang H-F, Wang B-H (2016) Identifying influential spreaders in complex networks based on gravity formula. Phys A Stat Mech Appl 451:205–212

27. Liu Y, Tang M, Zhou T, Do Y (2016) Identify influential spreaders in complex networks, the role of neighborhood. Phys A Stat Mech Appl 452:289–298

28. French JR Jr (1956) A formal theory of social power. Psychol Rev 63(3):181

29. DeGroot MH (1974) Reaching a consensus. J Am Stat Assoc 69(345):118–121

30. Tsitsiklis JN (1984) Problems in decentralized decision making and computation. Massachusetts Inst of Tech Cambridge Lab for Information and Decision Systems

31. Deffuant G, Amblard F, Weisbuch G, Faure T (2002) How can extremism prevail? A study based on the relative agreement interaction model. J Artif Soc Soc Simul 5(4):1

32. Weisbuch G (2004) Bounded confidence and social networks. Eur Phys J B 38(2):339–343

33. Zhao L, Cui H, Qiu X, Wang X, Wang J (2013) SIR rumor spreading model in the new media age. Phys A Stat Mech Appl 392(4):995–1003

34. "SNAP". http://snap.stanford.edu/

35. N. Developers, "NetworkX," networkx. lanl. gov (2010)

36. Hu H-B, Wang X-F (2008) Unified index to quantifying heterogeneity of complex networks. Phys A Stat Mech Appl 387(14):3769–3780

37. Castellano C, Pastor-Satorras R (2010) Thresholds for epidemic spreading in networks. Phys Rev Lett 105(21):218701

38. Kitsak M, Gallos LK, Havlin S, Liljeros F, Muchnik L, Stanley HE, Makse HA (2010) Identification of influential spreaders in complex networks. Nat Phys 6(11):888

39. Liu Y, Tang M, Zhou T, Do Y (2015) Core-like groups result in invalidation of identifying super-spreader by k-shell decomposition. Sci Rep 5:9602

40. Knight WR (1966) A computer method for calculating Kendall's tau with ungrouped data. Journal of the American Statistical Association 61(314):436–439