


# Source localization in signed networks with effective distance

Zhi-Wei Ma(马志伟), Lei Sun(孙蕾), Zhi-Guo Ding(丁智国), Yi-Zhen Huang(黄宜真), and Zhao-Long Hu(胡兆龙)

**Citation:**Chin. Phys. B, 2024, 33 (**2**): 028902. DOI: 10.1088/1674-1056/ad1482
Journal homepage: http://cpb.iphy.ac.cn; http://iopscience.iop.org/cpb

**What follows is a list of articles you may be interested in**

---

## Identifying multiple influential spreaders in complex networks based on spectral graph theory

Dong-Xu Cui(崔东旭), Jia-Lin He(何嘉林), Zi-Fei Xiao(肖子飞), and Wei-Ping Ren(任卫平)
Chin. Phys. B, 2023, 32 (**9**): 098904. DOI: 10.1088/1674-1056/acac0b

## Important edge identification in complex networks based on local and global features

Jia-Hui Song(宋家辉)
Chin. Phys. B, 2023, 32 (**9**): 098901. DOI: 10.1088/1674-1056/aca6d8

## Identification of key recovering node for spatial networks

Zijian Yan(严子健), Yongxiang Xia(夏永祥), Lijun Guo(郭丽君), Lingzhe Zhu(祝令哲), Yuanyuan Liang(梁圆圆), and Haicheng Tu(涂海程)
Chin. Phys. B, 2023, 32 (**6**): 068901. DOI: 10.1088/1674-1056/acb75f

## AG-GATCN: A novel method for predicting essential proteins

Peishi Yang(杨培实), Pengli Lu(卢鹏丽), and Teng Zhang(张腾)
Chin. Phys. B, 2023, 32 (**5**): 058902. DOI: 10.1088/1674-1056/acb9f9

## Analysis of cut vertex in the control of complex networks

Jie Zhou(周洁), Cheng Yuan(袁诚), Zu-Yu Qian(钱祖燏), Bing-Hong Wang(汪秉宏), and Sen Nie(聂森)
Chin. Phys. B, 2023, 32 (**2**): 028902. DOI: 10.1088/1674-1056/aca208

---

# Source localization in signed networks with effective distance

Zhi-Wei Ma(马志伟)[1], Lei Sun(孙蕾)[2], Zhi-Guo Ding(丁智国)[1], Yi-Zhen Huang(黄宜真)[3], and Zhao-Long Hu(胡兆龙)[1,†]

[1] *Zhejiang Normal University, School of Computer Science and Technology, Jinhua* 321004, *China*
[2] *Shanghai Business School, School of Business and Economics, Shanghai* 200235, *China*
[3] *Jinhua Polytechnic, School of Information Engineering, Jinhua* 321016, *China*

(Received 27 September 2023; revised manuscript received 6 December 2023; accepted manuscript online 12 December 2023)

While progress has been made in information source localization, it has overlooked the prevalent friend and adversarial relationships in social networks. This paper addresses this gap by focusing on source localization in signed network models. Leveraging the topological characteristics of signed networks and transforming the propagation probability into effective distance, we propose an optimization method for observer selection. Additionally, by using the reverse propagation algorithm we present a method for information source localization in signed networks. Extensive experimental results demonstrate that a higher proportion of positive edges within signed networks contributes to more favorable source localization, and the higher the ratio of propagation rates between positive and negative edges, the more accurate the source localization becomes. Interestingly, this aligns with our observation that, in reality, the number of friends tends to be greater than the number of adversaries, and the likelihood of information propagation among friends is often higher than among adversaries. In addition, the source located at the periphery of the network is not easy to identify. Furthermore, our proposed observer selection method based on effective distance achieves higher operational efficiency and exhibits higher accuracy in information source localization, compared with three strategies for observer selection based on the classical full-order neighbor coverage.

**Keywords:** complex networks, signed networks, source localization, effective distance

**PACS:** 89.75.–k, 87.23.Ge, 89.75.Fb **DOI:** 10.1088/1674-1056/ad1482

## 1. Introduction

The uncontrolled dissemination of malicious content through the widespread diffusion and propagation of information in real life can result in significant societal losses. Examples include the rapid spread of false information on social media,[1,2] virus intrusion into computer networks,[3–5] the swift spread of negative or potentially harmful speech affecting national security,[6] as well as the emission of industrial pollutants and hazardous substances.[7,8] While much research has focused on studying information diffusion from the perspective of maximizing influence,[9,10] this paper addresses a reverse problem, namely source localization, which aims to accurately and efficiently identify the initial spreader of information.

Considerable advancements have been achieved in the field of source localization in complex networks. Shah and Zaman were pioneers in systematically investigating the identification of virus sources in networks, utilizing the classical susceptible–infectious–recovered (SIR) model. They introduced the rumor centrality measure in tree-like networks to estimate the source of information propagation.[11] This measure captured the likelihood of a node becoming a source based on its high rumor centrality. The authors extended this measure to general networks by assuming information spreads through breadth-first algorithms. Zhu *et al.* proposed the Jordan centrality algorithm for source localization, which yielded more accurate results for general networks compared to the rumor centrality measure.[12] Brockmann *et al.* introduced effective distance and combined it with maximum likelihood estimation to locate the source, enabling reliable prediction of disease arrival times.[13] Jiang *et al.* further utilized effective distance to transform the original network, facilitating a clear understanding of complex diffusion processes. They employed a $k$-center method to identify multiple sources and the corresponding infection areas in general networks.[14] Other methods have also been proposed for source localization, including maximum likelihood estimation,[11,15–17] statistical physics,[18–21] reverse propagation,[22–34] machine learning,[35,36] etc.[37–39]

However, existing source localization research overlooks the signed nature of node connections, which is frequently encountered in signed networks, such as social networks that incorporate friend and enemy relationships.[40–42] Different from traditional networks, signed networks exhibit distinct properties in many aspects, including the information spread threshold cascades,[43] and the design of immunization strategies.[44] Zhang *et al.* discovered that an increase or clustering of negative interpersonal relationships in signed networks can lead to social unrest, thereby affecting the spread of diseases or information.[45] However, it is not clear whether signed networks show different results from traditional networks in terms of source localization. This paper concentrates on signed networks and proposes corresponding localization

[†]Corresponding author. E-mail: huzhaolong@zjnu.edu.cn

© 2024 Chinese Physical Society and IOP Publishing Ltd 

methods based on their structural characteristics.

Because it is not practical to obtain information from all nodes, quickly and accurately locating the source with a small number of observation nodes (also called as sensors or observers) is the main challenge. In existing research, observers are often chosen randomly or heuristically based on network centrality or community measures. However, these approaches are sensitive to outliers and may lead to false-positive conclusions. Pinto *et al.* were the first to propose a reverse propagation method for source localization.[22] Paluch *et al.* selected observation nodes based on the ordering of node infection times, combined with gradient descent and reverse propagation, which improved localization accuracy and efficiency but required a large number of observers.[29] Hu *et al.* used reverse propagation to infer the initial propagation time matrix and presented a greedy algorithm for optimizing observer selection, significantly reducing the number of observers needed.[46] These methods, however, are limited to single-source localization studies. Fu *et al.* presented the max–min method for addressing multi-source localization based on reverse propagation.[24] Building upon this work, Hu *et al.* observed that the initial propagation time of the true source occurs at its maximum frequency in the referred initial propagation time matrix, leading to the proposal of a maximum–maximum source localization method.[30] Integer programming methods can improve accuracy, but observer selection remains random.[28] Nevertheless, the random observation method is still adopted, requiring a relatively large number of observers. Additionally, researchers have explored the optimal placement problem of observers in complex networks, including betweenness centrality (BC),[47] high coverage,[48] and other approaches.[49] The observation theory can be adopted to achieve the minimum number of observers, but only for linear diffusion systems.[50] Wang *et al.* proposed a full-order neighbor coverage (FONC) strategy to enhance the utilization of observers, significantly improving source localization accuracy.[31] The core idea of the FONC algorithm lies in excluding nodes in the same order as observers from being selected as observers, thereby increasing the utilization of observer nodes. Considering the difference in the propagation ability of positive and negative edges of signed networks, we convert the propagation ability of positive and negative edges into effective distance, and propose a new observer selection method. Meanwhile, for comparison, we propose several observer selection methods based on FONC. Furthermore, we develop a method for source localization in signed networks through reverse propagation. Our contributions are summarized as follows.

(i) We are the first to investigate source localization in signed networks, taking into account the signed nature of information propagation in social networks.

(ii) Based on the structural characteristics of signed networks, we propose a new method to optimize observer selection with effective distance.

(iii) Leveraging the idea of reverse propagation, we provide a source localization method specifically designed for signed networks.

(iv) By using effective distance to select observation nodes, we achieve high accuracy of source localization and high operating efficiency.

(v) The higher the proportion of positive edge, and the higher the ratio of positive edge propagation ability to negative edge propagation ability, the more beneficial to source localization.

The organization of this paper is as follows. In Section 2, we first introduce the signed network model and the propagation model. Then we present the observer selection strategy with effective distance. Finally, combining with the selected observers, we develop the method of source localization in signed networks with reverse propagation. In Section 3, we evaluate the accuracy of the algorithm for source localization. Finally, we summarize and discuss the paper.

## 2. Model and methods

### 2.1. Signed network model

The signed network is denoted as $G = \{V, E\}$, where the node set $V = \{v_1, v_2, \ldots, v_N\}$ and the edge set $(v_i, v_j, w_{ij}) \in E$, $w_{ij} = 1$ indicates the existence of a positive edge between nodes $v_i$ and $v_j$, while $w_{ij} = -1$ indicates the existence of a negative edge. Figure 1(a) is a simple structure of the signed network. In this paper, we assume that the topological structure of network $G$ is known and remains unchanged.

### 2.2. Propagation model

In this paper, the susceptible–infectious (SI) model is adopted. $S$ represents the susceptible state, and $I$ represents the infected state. At time $t_0$, information starts to spread. At a certain time $t$, a node is in either the $S$ or $I$ state. In a signed network, positive edges and negative edges have a different propagation probability. We use $\theta^+$ to represent the propagation probability of positive edges and $\theta^-$ to represent the propagation probability of negative edges. Generally, $\theta^+ > \theta^-$. Here, we assume that the information propagation starts from an unknown node $s^*$, and the initial propagation time $t_0$ is also unknown. The propagation process is as follows.

(i) At the initial time $t_0$, only the source node $s^*$ is in the informed (infected) state.

(ii) The informed nodes pass the information to their neighbors in the next time step based on the corresponding propagation probability of edge.

### 2.3. Observer selection strategies

**Algorithm 1** Define the effective distance matrix $\boldsymbol{S}$, as shown in Fig. 1(b). The elements $S_{ij}$ in matrix $\boldsymbol{S}$ represent the propagation score from node $v_i$ to node $v_j$. The propagation score $S_{ij}$ is obtained by adding up the propagation probability on the shortest path from node $v_i$ to node $v_j$. If there are multiple equal shortest paths, we select the shortest path with the highest number of positive edges. The sum of elements in each row represents the total propagation score of the node. We select the node with the highest total propagation score as the observer and iterate until $\eta$ for all nodes in $V\backslash O$ becomes 0.

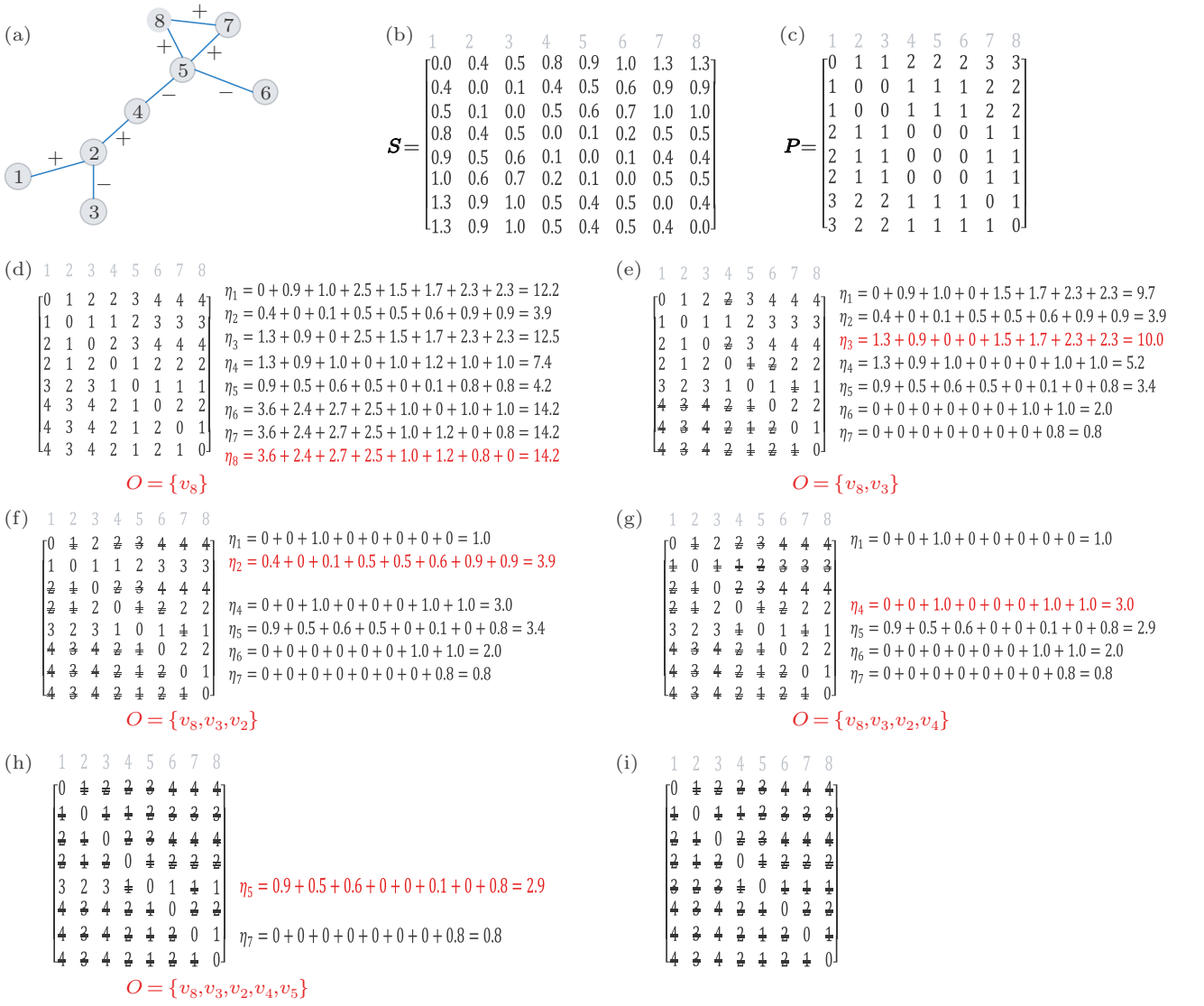



**Fig. 1.** (a) A simple illustration of a signed network. + indicates the positive edge and − indicates the negative edge. (b) Effective distance matrix $\boldsymbol{S}$. (c) Positive edge matrix $\boldsymbol{P}$. (d)–(i) The process of Algorithm 4 for observer selection. Here, we assume that the propagation probability of positive edges $\theta^+$ is 0.4, and the propagation probability of negative edges $\theta^-$ is 0.1.

**Algorithm 2** Full-order neighbor coverage strategy (FONC).[31] We define $d_{v_j,v_k}=m$ as node $v_k$ being the $m$-th-order neighbor of node $v_i$, indicating the shortest distance between $v_j$ and $v_k$. For an observer $o$, we define $V_o^{v_k}$ as the set of nodes that have the same shortest distance to $v_k$ as the observer $o$. In other words, for any $v_j\in V_o^{v_k}$, we have $d_{o,v_k}=d_{v_j,v_k}>0$. The FONC algorithm is an observer selection algorithm based on a greedy strategy, and its main algorithmic idea is as follows.

(i) Let $O$ be the set of observer nodes, initially empty.

(ii) For each node $v_i\in V\backslash O$, calculate $\eta_{v_i}=\sum_{k=1}^{N}|V_{v_i}^{v_k}|$, and add the node with the maximum $\eta$ value to the set of observers $O$. If there are multiple nodes with the same maximum value, the node with the highest degree is selected.

(iii) For all nodes $v_j\in V_o^{v_k}$, set $d_{v_j,v_k}=0$.

(iv) Repeat steps (ii) and (iii) until $\eta$ for all nodes in $V\backslash O$ becomes 0.

Based on the FONC algorithm and considering the characteristics of positive and negative edge propagation capacities in the signed network, we improve the FONC algorithm and

propose three observer selection algorithms as follows.

**Algorithm 3** Define the positive edge matrix $\boldsymbol{P}$, as shown in Fig. 1(c). The elements $P_{ij}$ in matrix $\boldsymbol{P}$ represent the number of positive edges on the shortest path from node $v_i$ to node $v_j$. The sum of elements in each row in the matrix represents the total number of positive edges from the node to other nodes. Based on the FONC strategy, when there are multiple nodes with the same $\eta$ value, we choose the node with more positive edges as the observer. The reason is that positive edges have stronger propagation capabilities, and paths with more positive edges are more likely to be the real propagation paths.

**Algorithm 4** By directly combining the effective distance with the FONC strategy, we no longer calculate the $\eta$ value for each node directly. Instead, we incorporate the propagation probability of edges into the calculation of $\eta$ value and select observers based on the newly calculated $\eta$ value, thereby obtaining the optimal group of observers. To be specific, we simulate this strategy in Figs. 1(d)–1(i). Figure 1(d) is the distance matrix corresponding to the network in Fig. 1(a), and we calculate the $\eta$ values of each node based on the effective distance of the network. If we select $v_1$ as the observer node, then $d_{v_1,v_2}=1$, so none of the first-order neighbors of $v_2$ will be selected as observer nodes. Among them, $v_1$, $v_3$, and $v_4$ are all first-order neighbors of $v_2$. By using the effective distance matrix $\boldsymbol{S}$, we obtain the propagation scores $S_{21}$, $S_{23}$, and $S_{24}$ of all first-order neighbors of $v_2$, which sum to 0.9. The same calculations are performed for the remaining nodes, and we obtain a $\eta$ value of 12.2 for node $v_1$. Nodes $v_6$, $v_7$, and $v_8$ have the highest $\eta$ values, and we randomly select $v_8$ as the observer node for the first step. We add $v_8$ to the set of observers $O$ and delete elements in each column that are equal to the corresponding element in the eighth row of the distance matrix, see Fig. 1(e). Then we calculate the $\eta$ values of the nodes in $V\backslash O$ using the remaining distance matrix and select a new observer node, repeating this process until all elements in the distance matrix are deleted. Figures 1(e)–1(i) show the remaining steps of selecting observer nodes. Finally, we obtain the set of observer nodes $O=\{v_8,v_3,v_2,v_4,v_5\}$.

### 2.4. Source localization method in signed networks based on maximum likelihood estimation

Based on the aforementioned algorithms, we can obtain the observer set $O \triangleq \{o_k\}_{k=1}^{K} \subset V$, representing a set of $K$ observers. Define $t_o$ as the absolute time moment when observer node $o$ receives information. We define $\mathcal{O}$ to represent the observer and the receiving time, i.e., $\mathcal{O} \triangleq \{(o,t_o)\}$, for all $o \in O$.

Accurately locating the source based on the observed node state is equivalent to identifying which node caused the observed data we obtained, making it a maximum likelihood estimation (MLE) problem. Since $s^*$ is uniformly random over the network $G$, the optimal estimator can be expressed as[22]

$$\begin{aligned}\hat{s}(\mathcal{O}) &= \underset{s\in\mathcal{G}}{\operatorname{argmax}}\, \boldsymbol{P}(\mathcal{O} \mid s^*=s) \\ &= \underset{s\in\mathcal{G}}{\operatorname{argmax}} \sum_{\Pi_s} \boldsymbol{P}(\Pi_s \mid s^*=s) \\ &\quad \times \int\cdots\int g(\theta_1,\ldots,\theta_L,\mathcal{O},\Pi_s,s)\,\mathrm{d}\theta_1\cdots\mathrm{d}\theta_L. \end{aligned} \tag{1}$$

Here $\Pi_s$ is the set of all possible paths $\{\mathcal{P}_{s,o_k}\}_{k=1}^{K}$ between sources and observers in graph $G$. The set $\{\theta_l\}_{l=1}^{L}$ represents random propagation delays of all $L$ edges in graph $G$, and $g$ is a deterministic function that depends on the joint distribution of propagation delays in a complex manner. Essentially, the joint probability distribution $g$ in Eq. (1) involves all possible combinations of paths, and computing this combination would be computationally infeasible as it exhibits exponential growth with an increasing number of nodes. In this paper, we incorporate the properties of signed network edges and apply the reverse propagation algorithm to present a method to locate the source in signed networks. The optimal estimator is given by

$$\hat{s} = \underset{s\in G}{\operatorname{argmax}}\, \boldsymbol{\mu}_s^{\mathrm{T}} \boldsymbol{\Lambda}^{-1} \left(\boldsymbol{d} - \frac{1}{2}\boldsymbol{\mu}_s\right), \tag{2}$$

where $\boldsymbol{d}$ is the observed delay, $\boldsymbol{\mu}$ denotes the deterministic delay, and $\boldsymbol{\Lambda}$ represents the delay covariance. Intuitively, when selecting node $s$ as the source, $\boldsymbol{\mu}$ and $\boldsymbol{\Lambda}$ respectively represent the mean and covariance of the observed delay $\boldsymbol{d}$ (a random vector), calculated by the following two formulas:

$$[\boldsymbol{d}]_k = t_{o_{k+1}} - t_{o_k}, \tag{3}$$

$$\begin{aligned}[\boldsymbol{\mu}_s]_k &= \mu_+(|\mathcal{P}_+(s,o_{k+1})| - |\mathcal{P}_+(s,o_1)|) \\ &\quad + \mu_-(|\mathcal{P}_-(s,o_{k+1})| - |\mathcal{P}_-(s,o_1)|), \end{aligned} \tag{4}$$

where $\mu_+$ and $\mu_-$ represent the deterministic delays of positive and negative edges, respectively, and $|\mathcal{P}_+(s,o_1)|$ denotes the number of positive edges on the shortest path $\mathcal{P}(s,o_1)$ between node $s$ and observer $o_1$. Similarly, $|\mathcal{P}_-(s,o_1)|$ represents the number of negative edges on this shortest path

$$[\boldsymbol{\Lambda}]_{k,i} = \sigma^2 \times \begin{cases} |\mathcal{P}(o_1,o_{k+1})|, & k=i, \\ |\mathcal{P}(o_1,o_{k+1}) \cap \mathcal{P}(o_1,o_{i+1})|, & k\neq i, \end{cases} \tag{5}$$

where $k$ and $i$ are positive integers ranging from 1 to $K-1$.

## 3. Results

We verify the proposed method for the source localization with two synthetic networks, namely Erdös–Rényi (ER)[51] and Barabási–Albert (BA)[52] networks. Based on the characteristics of signed networks, we assume that there is a difference in propagation delay between positive and negative edges. We do not consider the balance between positive and negative edges in the signed network.[53,54]

### 3.1. Relationship between observer ratio and network size

First, we investigate the impact of the selected observer ratio $n_O$ on our proposed strategies for different network sizes in both ER and BA networks. The observer ratio $n_O$ is defined as follows:

$$n_O = \frac{N_O}{N}, \tag{6}$$

where $N_O$ represents the number of selected observers, and $N$ is the network size.

According to the analysis presented in Fig. 2, Algorithms 2–4 require similar observer ratio for the same network size in both BA and ER networks. However, Algorithm 1 requires more observers. This discrepancy can be attributed to the fact that Algorithm 1 selects the node with the maximum total propagation score as an observer using the effective distance matrix $\boldsymbol{S}$, and does not take into account the influence of $\eta$, which necessitates a larger number of observers to cover the entire network. As depicted in Fig. 2, as the network scale increases, although the number of observation nodes rises, the proportion of observation nodes decreases. Notably, Algorithm 1 requires more observers in BA networks compared to ER networks. Therefore, the network structure significantly impacts the selection of observers.

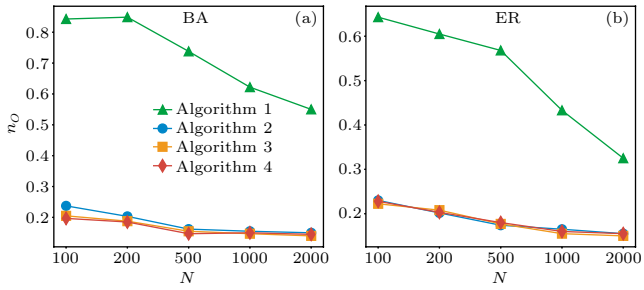



**Fig. 2.** The relationship between the observer ratio $n_O$ under the four algorithms and the network size $N$ in (a) BA networks and (b) ER networks. The network size values are $N = \{100, 200, 500, 1000, 2000\}$, respectively. The average degree $\langle k \rangle = 10$, positive edges account for 80% of the total edges in the network, the propagation probability of negative edges $\theta^- = 0.1$, and the propagation probability of positive edges $\theta^+ = 0.4$. All the results are average of 200 independent simulations.

### 3.2. Relationship between observer proportion and average degree

Here, we further analyze the observer selection ratios for the four algorithms under different average degrees. From the data presented in Fig. 3, it can be observed that Algorithm 1 selects a relatively larger percentage of observers, while Algorithms 2–4 roughly select observers within the range of 20% to 30%. In Fig. 3(a), it can be seen that the observer ratios required by the four algorithms decrease as the network average degree increases. This phenomenon may be attributed to the fact that denser network structures lead to higher average degrees of observation nodes, enabling them to cover a wider range and monitor more information. It is worth noting that for ER networks, as shown in Fig. 3(b), when the average degree $\langle k \rangle$ is less than a certain value, such as 14, a lower number of observers suffices. However, as the average degree continues to increase, more observers are needed. This suggests that for relatively homogeneous networks, as the network becomes denser, it becomes challenging to distinguish $\eta$ values for many nodes, making it difficult to reach $\eta = 0$ and requiring more observer nodes.

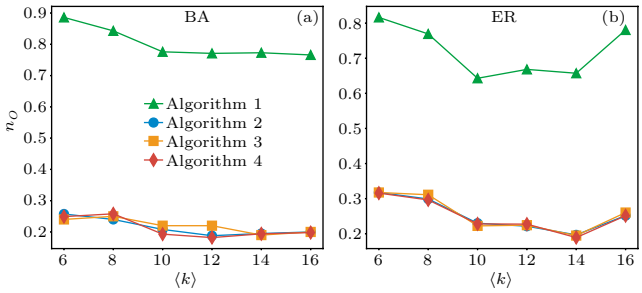



**Fig. 3.** The relationship between the observer ratio $n_O$ under the four algorithms and the network average degree $\langle k \rangle$ in BA networks (a) and ER networks (b). The average degree values are $\langle k \rangle = \{6, 8, 10, 12, 14, 16\}$, respectively. The network size is $N = 100$, the observer ratio is $n_O = 20\%$, positive edges account for 80% of the total edges in the network, the propagation probability of negative edges $\theta^- = 0.1$, and the propagation probability of positive edges $\theta^+ = 0.4$. All the results are average of 200 independent simulations.

### 3.3. Effects of proportion of observation nodes on source localization

To evaluate the performance of our source localization method, we use the area under receiver operating characteristic (AUROC) metric,[55,56] a standard measure from signal processing. The AUROC ranges from 0 to 1, where 1 indicates perfect source localization, i.e., the source can be completely distinguished from other nodes in the network. A random guess corresponds to an AUROC of 0.5.

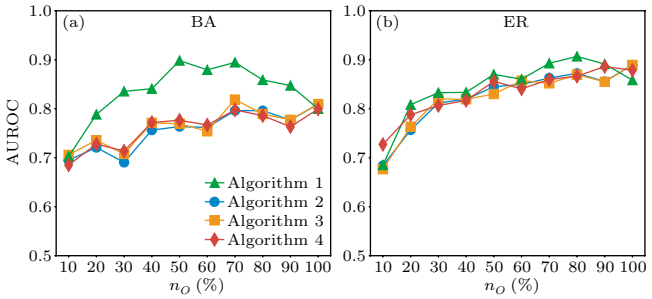



**Fig. 4.** The relationship between the AUROC and different observer ratios $n_O$ in BA networks (a) and ER networks (b). The network size is $N = 100$, and the average degree is $\langle k \rangle = 10$. The positive edges account for 80% of the total edges in the network, the propagation probability of negative edges $\theta^- = 0.1$, and the propagation probability of positive edges $\theta^+ = 0.4$. All the results are average of 200 independent simulations.

First, we study the effect of different observer ratios on the accuracy of source localization in signed networks. Based on the experimental results shown in Fig. 4, it can be observed that the overall accuracy of the four algorithms increases as the observer ratio $n_O$ increases (except for Algorithm 1). In Fig. 4(a), when the observer ratio exceeds 10%,

Algorithm 1 outperforms other algorithms. Similarly, in the results in Fig. 4(b), Algorithm 1 also demonstrates better performance of source localization in ER networks. When the observer ratio $n_O$ ranges from 60% to 80%, the AUROC value of Algorithm 1 is around 0.9, indicating that the observers selected by Algorithm 1 can better locate the source. It should be noted that when $n_O$ exceeds a certain value, such as 80%, the AUROC unexpectedly decreases. This may be due to the increased likelihood of error estimation in the reverse propagation algorithm when there are more observation nodes.

### 3.4. Effects of average degree on source localization

In this section, we investigate the influence of different network average degrees on source localization in signed networks. From the observations depicted in Fig. 5, Algorithm 1 demonstrates a high localization capability in both BA and ER networks under different average degrees. Although there appears to be a range, such as an average degree between 8 and 14, with relatively higher accuracy, it is not significantly outstanding. This suggests that the proposed method exhibits strong robustness to network structure. Furthermore, while the network average degree has a substantial impact on the proportion of observers, as shown in Fig. 3(b), it has a relatively minor impact on source localization.

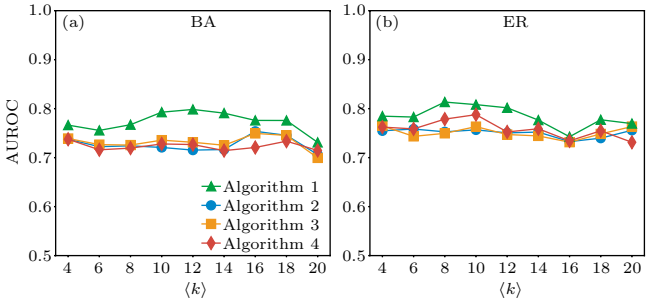



**Fig. 5.** The relationship between the AUROC and different network average degrees $\langle k \rangle$ in BA networks (a) and ER networks (b). The network size is $N = 100$ and the observer ratio is $n_O = 20\%$. The positive edges account for 80% of the total edges in the network, the propagation probability of negative edges $\theta^- = 0.1$, and the propagation probability of positive edges $\theta^+ = 0.4$. All the results are average of 200 independent simulations.

### 3.5. Effects of propagation probability of positive and negative edges on source localization

The propagation probability of positive and negative edges in the signed network influences propagation dynamics, but its impact on source localization remains unclear. Therefore, we further investigate its effects on source localization in signed networks.

From Figs. 6(a) and 6(b), it is evident that as the propagation probability of positive edges increases, the AUROC of all four algorithms exhibits an upward trend. Specifically, Algorithm 1 consistently achieves noticeably higher AUROC value compared to Algorithms 2–4, as observed in Fig. 6(a). Similarly, in Fig. 6(b), Algorithm 1 also demonstrates commendable localization capability in ER networks. This further illustrates the significant influence of the propagation capacities of positive and negative edges on source localization in signed networks. Additionally, we conclude that as the propagation probability of positive edges increases, the performance of source localization also improves. It is worth noting that in ER networks, as shown in Fig. 6(b), when $\theta^+$ is relatively small, the AUROC of Algorithm 1 does not differ significantly from other three algorithms. This may be attributed to the shorter average path lengths in ER networks, resulting in relatively close distances between nodes. As Algorithm 1 selects observers through the effective distance matrix $\boldsymbol{S}$, when $\theta^+$ is small, the shortest path values between nodes in the ER network do not differ significantly, leading to no significant improvement in source localization.

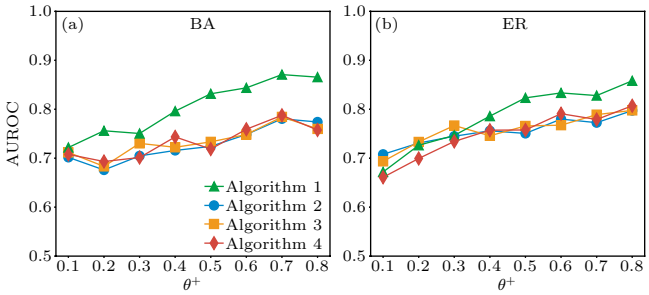



**Fig. 6.** The relationship between the AUROC and the propagation probability of positive and negative edges in BA networks (a) and ER networks (b). The propagation probability of negative edges is $\theta^- = 0.1$ and the propagation probability of positive edges is within the range $\theta^+ = \{0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8\}$. The network size is $N = 100$, the average degree is $\langle k \rangle = 10$, the observer ratio is $n_O = 20\%$, and positive edges account for 80% of the total edges in the network. All the results are average of 200 independent simulations.

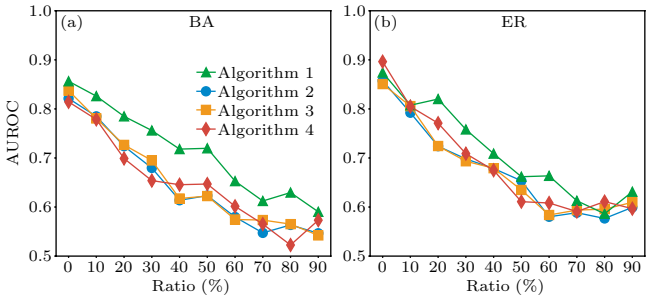



**Fig. 7.** The relationship between the AUROC and the ratio of positive and negative edges in BA networks (a) and ER networks (b). The $x$-axis ratio represents the ratio of negative edges to the total number of edges in the network, ranging from 0 to 0.9. The network size is $N = 100$, the average degree is $\langle k \rangle = 10$, the observer proportion is $n_O = 20\%$, the propagation probability of negative edges is $\theta^- = 0.1$, and the propagation probability of positive edges is $\theta^+ = 0.4$. All the results are average of 200 independent simulations.

### 3.6. Effects of ratio of positive and negative edges on source localization

Furthermore, we investigate the influence of the ratio of positive and negative edges in signed networks on the source localization. As shown in Fig. 7, we can find that as the ra-

tio of negative edges increases, the AUROC of the four algorithms decreases. This indicates that a higher proportion of positive edges enhances the source localization performance of the algorithms. Therefore, we can conclude that these four algorithms are more suitable for signed networks with a higher proportion of positive edges. This finding aligns with real social networks, where individuals typically have more friends than enemies, contributing to better survival in society.

### 3.7. Effects of the location of source on source localization

Considering that the source at different locations may affect the accuracy of the source localization, we select the node with the largest degree, the node with the smallest degree, and the node with the largest betweenness centrality in the network as the propagation source, respectively, to further explore how the source at different locations affect the source localization.

As can be seen from Fig. 8, in BA and ER networks, Algorithm 1 shows good source localization when the source node is the node with the largest degree or the node with the largest betweenness centrality, but when the source node is the node with the smallest degree, the AUROC is lower. It indicates that the source on the periphery of the network is not easy to be located.

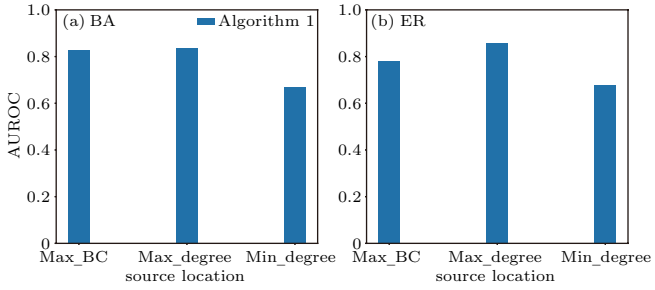



**Fig. 8.** The relationship between the AUROC and the source at different location in BA networks (a) and ER networks (b). The network size is $N = 100$, the observer proportion is $n_O = 20\%$, the average degree is $\langle k \rangle = 10$, the propagation probability of negative edges is $\theta^- = 0.1$, the propagation probability of positive edges is $\theta^+ = 0.4$, and the proportion of positive edges in the network is 80%. All the results are average of 200 independent simulations.

### 3.8. Robustness analysis — propagation probability

The previous results are obtained by using an average delay time of $\mu_+ = \theta^+$ for each positive edge, and an average delay time $\mu_- = \theta^-$ for each negative edge. It actually contains a certain amount of noise, because the propagation delay and the propagation rate of the SI propagation model are different. Here, we assume that the information propagation delay follows a Gaussian distribution $\mathcal{N}(\mu, \sigma^2)$, where the mean $\mu$ incorporates $\mu_+ = \theta^+$ and $\mu_- = \theta^-$. We further increase the noise and study the effect of variance $\sigma^2$ in Eq. (5) on source localization.

From the experimental results in Fig. 9, it can be observed that Algorithm 1 consistently achieves the highest accuracy in source localization, which is consistent with the previous results. Overall, even with an increasing $\sigma^2$, the change in the AUROC is not significant, demonstrating that the proposed source localization method exhibits strong robustness against propagation noise.

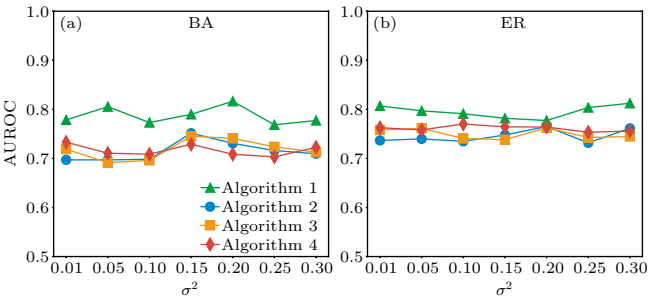



**Fig. 9.** The relationship between the AUROC and the variance $\sigma^2$ in BA networks (a) and ER networks (b). The network size is $N = 100$, the observer proportion is $n_O = 20\%$, the average degree is $\langle k \rangle = 10$, the propagation probability of negative edges is $\theta^- = 0.1$, the propagation probability of positive edges is $\theta^+ = 0.4$, and the proportion of positive edges in the network is 80%. All the results are average of 200 independent simulations.

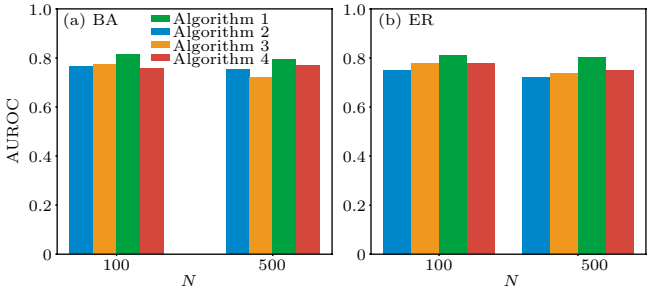



**Fig. 10.** The relationship between the AUROC and network size $N$. The values of $N$ are $\{100, 500\}$, the observer proportion is $n_O = 20\%$, the average degree is $\langle k \rangle = 10$, the propagation probability of negative edges is $\theta^- = 0.1$, the propagation probability of positive edges is $\theta^+ = 0.4$, and the proportion of positive edges in the network is 80%. All the results are average of 200 independent simulations.

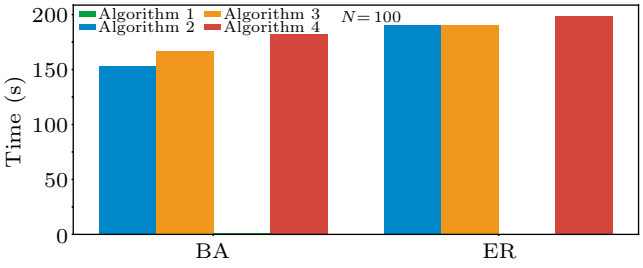



**Fig. 11.** The running time of the four algorithms in BA and ER networks. The network size is $N = 100$, the observer proportion is $n_O = 20\%$, the average degree is $\langle k \rangle = 10$, the propagation probability of negative edges is $\theta^- = 0.1$, the propagation probability of positive edges is $\theta^+ = 0.4$, and the proportion of positive edges in the network is 80%. All the results are average of 200 independent simulations.

### 3.9. Robustness analysis — network size

Additionally, we study the performance of the four algorithms for source localization with larger network sizes. As depicted in Fig. 10, we can find that as the network size increases to 500, the source localization accuracy of the four algorithms remains almost unchanged. This suggests that the

proposed algorithms have strong robustness and can be applied to source localization in large-scale networks. Among them, Algorithm 1 consistently achieves the best source localization results.

Finally, we analyze the operational efficiency of the four algorithms. Figure 11 shows the running time of the four algorithms in BA and ER networks. It is evident that Algorithm 1 requires significantly less time compared to other algorithms. This finding indicates that Algorithm 1 not only has more accurate source localization results but also has higher efficiency.

### 3.10. Error distance

In order to improve the credibility of the experimental results, we introduce error distance to evaluate the effectiveness of source localization. The error distance can be computed by $d_{s^*,\hat{s}}/\langle d\rangle$, where $d_{s^*,\hat{s}}$ is the distance between real source $s^*$ and estimated source $\hat{s}$, $\langle d\rangle$ is the average distance of the network. And further investigate the effects of different ratios of observers and different ratios of positive and negative edges on source localization. The experimental results are presented in Tables 1 and 2.

**Table 1.** The relationship between the error distance and different observer ratios $n_O$ in BA networks.

| $n_O$ (%) | Algorithm 1 | Algorithm 2 | Algorithm 3 | Algorithm 4 |
|---|---|---|---|---|
| 10 | 0.810 | 0.815 | 0.815 | 0.850 |
| 30 | 0.813 | 0.814 | 0.800 | 0.850 |
| 50 | 0.788 | 0.770 | 0.783 | 0.824 |
| 70 | 0.711 | 0.756 | 0.729 | 0.747 |
| 100 | 0.664 | 0.678 | 0.678 | 0.696 |

**Table 2.** The relationship between the error distance and the propagation probability of positive and negative edges in BA networks. The $\theta^+$ represents the propagation probability of positive edges.

| $\theta^+$ (%) | Algorithm 1 | Algorithm 2 | Algorithm 3 | Algorithm 4 |
|---|---|---|---|---|
| 10 | 0.819 | 0.824 | 0.833 | 0.828 |
| 30 | 0.792 | 0.841 | 0.836 | 0.819 |
| 50 | 0.760 | 0.778 | 0.764 | 0.814 |
| 70 | 0.702 | 0.702 | 0.706 | 0.783 |

From Table 1, it can be observed that with different observer ratios, Algorithm 1 has a smaller error distance in most cases compared with other three algorithms. In Table 2, with the increase of positive edge propagation ability in the network, the error distance of Algorithm 1 decreases and outperforms other three algorithms, which further indicates that the stronger the positive edge propagation ability is, the more favorable the source localization is. By introducing the evaluation index of error distance, it further shows that our algorithm can find the source nodes accurately.

## 4. Discussion and conclusion

Accurately and efficiently identifying the source of information propagation is crucial in controlling the spread of harmful content. Many studies have revealed differences in information propagation in signed networks compared to traditional networks. However, the influence of signed network structure, as well as the proportion and propagation probability of positive and negative edges, on source localization remains unclear. In this study, we propose a new optimization strategy with effective distance for observer selection and develop a method for source localization in signed network models using reverse propagation compared with three observer selection methods based on FONC, experimental results demonstrate the overwhelming advantage of observer selection method with effective distance. Interestingly, we find that source localization accuracy increases with an increase in the number of positive edges in the signed network and with an increase in the ratio of positive edge propagation rate to negative edge propagation rate, which seemingly confirms the importance of cooperation. Furthermore, we observe that the average degree of the network has little impact on source localization accuracy. In addition, we find that the source located at the periphery of the network is not favorable for source localization. Moreover, our method exhibits robust source localization accuracy even in the presence of higher levels of noise, indicating its suitability for signed networks. The observer selection method with effective distance not only achieves high accuracy in source localization but also demonstrates superior efficiency.

The observer selection method with effective distance is highly recommended for source localization in signed networks due to its high accuracy, efficiency, and robustness. However, there are several avenues for future research. Firstly, the extension of the single-source problem to a multi-source problem in signed networks requires further investigation. Secondly, considering the impact of changing edge relationships over time, it is important to generalize the signed network source localization problem from static networks to dynamic networks.[57–60] Thirdly, exploring source localization in multi-layer networks by incorporating signed networks opens up new possibilities for research and analysis.[61–63] In this paper, we consider a simple signed network and do not consider the balance of the signed network. However, our hypothesis is feasible to a certain extent, because the enemy of our enemy is not necessarily our enemy, and the friend of our friend is not necessarily our friend. In the future research, we will further study and reveal whether there are new discoveries in the structure of balanced singed networks.

## Acknowledgments

Project supported by the National Natural Science Foundation of China (Grant Nos. 62103375 and 62006106), the Zhejiang Provincial Philosophy and Social Science Planning Project (Grant No. 22NDJC009Z), the Education Min-

istry Humanities and Social Science Foundation of China (Grant Nos. 19YJCZH056 and 21YJC630120), and the Natural Science Foundation of Zhejiang Province of China (Grant Nos. LY23F030003 and LQ21F020005).

## References


[1] Vosoughi S, Roy D and Aral S 2018 *Science* **359** 1146
[2] Ruffo G, Semeraro A, Giachanou A and Rosso P 2023 *Comput. Sci. Rev.* **47** 100531
[3] Lloyd A L and May R M 2001 *Science* **292** 1316
[4] Wang Y Q and Yang X Y 2013 *Chin. Phys. B* **22** 040206
[5] Avcı D and Soytürk F 2023 *J. Comput. Appl. Math.* **419** 114740
[6] Chesney B and Citron D 2019 *Calif. Law Rev.* **107** 1753
[7] Tripathi S, Chandra R, Purchase D, Bilal M, Mythili R and Yadav S 2022 *Environ. Pollut.* **292** 118342
[8] Zhu T, Wang X, Yu Y, Li C, Yao Q and Li Y 2023 *J. Environ. Sci. China* **123** 83
[9] Morone F and Makse H A 2015 *Nature* **524** 65
[10] Yan Z, Xia Y, Guo L, Zhu L, Liang Y and Tu H 2023 *Chin. Phys. B* **32** 068901
[11] Shah D and Zaman T 2010 *ACM Sigmetrics Perform. Eval. Rev.* **38** 203
[12] Zhu K and Ying L 2014 *IEEE ACM Trans. Netw.* **24** 408
[13] Brockmann D and Helbing D 2013 *Science* **342** 1337
[14] Jiang J, Wen S, Yu S, Xiang Y and Zhou W 2015 *IEEE Trans. Inf. Forensics Secur.* **10** 2616
[15] Doerr B, Fouz M and Friedrich T 2012 *Commun. ACM* **55** 70
[16] Zhu K and Ying L 2016 *IEEE ACM Trans. Netw.* **24** 408
[17] Chang B, Chen E, Zhu F, Liu Q, Xu T and Wang Z 2020 *IEEE Trans. Syst. Man Cybern. Syst.* **50** 2242
[18] Lokhov A Y, Mzard M, Ohta H and Zdeborov L 2014 *Phys. Rev. E* **90** 012801
[19] Altarelli F, Braunstein A, Dallsta L, Lage-Castellanos A and Zecchina R 2014 *Phys. Rev. Lett.* **112** 118701
[20] Nino A F, Alen L, Tomislav S, Hrvoje S and Mile S 2015 *Phys. Rev. Lett.* **114** 248701
[21] Liu Y, Li W, Yang C and Wang J 2022 *Sci. Rep.* **12** 5467
[22] Pedro C P, Patrick T and Martin V 2012 *Phys. Rev. Lett.* **109** 068702
[23] Wuqiong L, Wee-Peng T and Mei L 2014 *IEEE J. Sel. Top. Signal Process.* **8** 586
[24] Fu L, Shen Z, Wang W X, Fan Y and Di Z 2016 *Europhys. Lett.* **113** 18006
[25] Shen Z, Cao S, Wang W X, Di Z and Stanley H E 2016 *Phys. Rev. E* **93** 032301
[26] Wang Z, Wang C, Pei J and Ye X 2017 *Proceedings of the AAAI Conference on Artificial Intelligence,* February 4–9, 2017, San Francisco, America, p. 217
[27] Ji F, Tay W P and Varshney L R 2017 *IEEE Trans. Signal Process.* **65** 2517
[28] Hu Z L, Shen Z, Tang C B, Xie B B and Lu J F 2018 *Phys. Lett. A* **382** 931
[29] Robert P, Lu X, Krzysztof S, Boleslaw K S and Janusz A H 2018 *Sci. Rep.* **8** 2508
[30] Hu Z L, Shen Z, Han J, Peng H, Lu J F, Jia R, Zhu X B and Zhao D 2019 *Physica A* **527** 121262
[31] Wang H J and Sun K J 2020 *Europhys. Lett.* **131** 48001
[32] Wang H J, Zhang F F and Sun K J 2021 *Phys. Lett. A* **393** 127184
[33] Zhu P, Cheng L, Gao C, Wang Z and Li X 2022 *IEEE Trans. Netw. Sci. Eng.* **9** 1853
[34] Peng S L, Wang H J, Peng H, Zhu X B, Li X, Han J, Zhao D and Hu Z L 2023 *Chaos* **33** 083125
[35] Wang J, Jiang J and Zhao L 2022 *Proceedings of the ACM Web Conference*, April 25–29, 2022, Lyon, France, p. 1058
[36] Zhao Y, Chen B, Wang X, Zhu Z, Wang Y, Cheng G, Wang R, Wang R, He M and Liu Y 2022 *Inf. Sci.* **588** 67
[37] Wang H J, Hu Z L, Tao L, Shao S and Wang S Z 2023 *Sci. Rep.* **13** 5692
[38] Shao Y, Chen L, Chen Y, Liu W and Dai C 2023 *Inf. Sci.* **635** 375
[39] Wan P, Wang X, Pang G, Wang L and Min G 2023 *Expert Syst. Appl.* **213** 119239
[40] Leskovec J, Huttenlocher D and Kleinberg J 2010 *Proceedings of the SIGCHI conference on human factors in computing systems,* April 10–15, 2010, Atlanta, America, p. 1361
[41] Li L, Fan Y, Zeng A and Di Z 2019 *Physica A* **525** 433
[42] Li A W, Xiao J and Xu X K 2021 *Chin. Phys. B* **30** 038901
[43] Lee K M, Lee S, Min B and Goh K I 2023 *Chaos Solitons Fractals* **168** 113118
[44] Li A W, Xu X K and Fan Y 2022 *Chaos Solitons Fractals* **162** 112489
[45] Zhang P, Zhang X and Xue L 2022 *Int. J. Mod. Phys. C* **33** 2250042
[46] Hu Z L, Wang L and Tang C B 2019 *Chaos* **29** 063117
[47] Brandes U 2001 *J. Math. Sociol.* **25** 163
[48] Li X, Wang X, Zhao C, Zhang X and Yi D 2019 *Appl. Sci.* **9** 3758
[49] Paluch R, Gajewski G, Hoyst J A and Szymanski B K 2020 *Future Gener. Comput. Syst.* **112** 1070
[50] Hu Z L, Han X, Lai Y C and Wang W X 2017 *Royal Soc. Open Sci.* **4** 170091
[51] Erdős P and Rényi A 1960 *Publ. Math. Inst. Hung. Acad. Sci.* **5** 17
[52] Barabási A L and Albert R 1999 *Science* **286** 509
[53] Saeedian M, Azimi-Tafreshi N, Jafari G and Kertesz J 2017 *Phys. Rev. E* **95** 022314
[54] Kirkley A, Cantwell G T and Newman M E 2019 *Phys. Rev. E* **99** 012320
[55] Fawcett T 2006 *Pattern Recogn. Lett.* **27** 861
[56] Hanley J A and McNeil B J 1982 *Radiology* **143** 29
[57] Jiang J, Wen S, Yu S, Xiang Y and Zhou W 2018 *IEEE Trans. Depend. Secur. Comput.* **15** 166
[58] Hu Z L, Shen Z, Cao S, Podobnik B, Yang H, Wang W X and Lai Y C 2018 *Sci. Rep.* **8** 2685
[59] Chai Y, Wang Y and Zhu L 2021 *IEEE Trans. Inf. Forensics Secur.* **16** 2621
[60] Huang Q, Zhao C, Zhang X and Yi D 2017 *Physica A* **468** 434
[61] Hu Z L, Han J, Peng H, Lu J F, Zhu X B, Jia R and Li M 2022 *IEEE Trans. Netw. Sci. Eng.* **9** 3515
[62] Cheng L, Li X, Han Z, Luo T, Ma L and Zhu P 2022 *Chaos Solitons Fractals* **159** 112139
[63] Paluch R, Gajewski Ł G, Suchecki K and Hołyst J A 2021 *Physica A* **582** 126238