A Bi-objective Source Hiding Method for Network Propagation

Tianyang Gao[1], Danni Qu[1], Liqin Hu[1(✉)], and Zhen Wang[1,2]

[1] School of Cyberspace, Hangzhou Dianzi University, Hangzhou 310018, China
{gaotianyang,qudanni,huliqin,wangzhen}@hdu.edu.cn
[2] Zhuoyue Honors College, Hangzhou Dianzi University, Hangzhou 310038, China

Abstract. In many scenarios, sources want to publish information anonymously and allow it to be distributed on a large scale in a short period of time. To this end, we proposed a bi-objective source hiding method to locate the sources with both hiding and propagation capability in the network. Firstly, three heuristic methods are proposed to evaluate the hiding performance of nodes. Secondly, based on the characteristics of network propagation, breadth first search is applied to measure the propagation capability of nodes. Then, a normalization method is proposed to comprehensively evaluate the hiding ability and propagation ability of nodes, and the node with the strongest comprehensive ability is used as the propagation source. Finally, experiments are simulated based on five different source detection methods on multiple network structures, and the experimental results demonstrate the effectiveness of our method.

Keywords: Complex network · Propagation · Source hiding

1 Introduction

In the face of various network risks, such as public opinion, computer viruses and biological viruses, pinpointing the source of propagation in the network is an effective means to control the risks in a timely manner [1, 2]. Therefore, the identification of propagation sources in complex networks has been a hot topic of interest for researchers. However, there are still risks in locating sources using current source detection methods. They usually target infinite networks and randomly select sources to be simulated in simulation experiments. This makes it possible to successfully evade detection.

Propagation sources have an incentive to evade source detection in multiple scenarios to maintain their privacy. For example, anonymous platforms want to guarantee the anonymity of message senders to enable their freedom of expression [3, 4], and publishers of messages in social networks want to deliver messages on a large scale without revealing the personal information of the publishers [5–7]. Therefore, work on their possible hiding strategies from a source hiding perspective can not only be useful in these scenarios to meet the demand, but also be used to help network administrators analyze the strategies and behavior patterns of evaders to facilitate timely and effective

Y. Chenggang et al. (Eds.): MobiMedia 2022, LNICST 451, pp. 105–125, 2022.
https://doi.org/10.1007/978-3-031-23902-1_9

risk containment. In addition, the ability of nodes to efficiently disseminate information is an important factor in measuring their performance, which illustrates the impact of nodes on the dynamic dissemination process. Revealing this property of nodes will help to better optimize the limited resources for information dissemination [8, 9], which can be used, for example, to describe and predict the dynamic characteristics of financial markets [10] or to effectively market products at low cost [11].

Based on this, we propose a bi-objective source hiding method that satisfies the dual requirements of hiding and propagation capability. The main contributions of this paper are summarized as follows:

- Based on the analysis and replication of existing source detection work, we tested the hiddenness of sources at different locations and found that the closer the location is to the network boundary, the stronger the hiding.
- Heuristic methods are proposed to measure source hiding and propagation capabilities. In scenarios where there is a need for anonymity, locating nodes with high hiding in the network to use them as sources is more operable than methods such as setting propagation protocols [12, 13] and changing the network structure [14].
- A bi-objective strategy is proposed to locate sources in the network with both hiding and propagation, and the node with both high hiding and high propagation capability is selected as the source through a comprehensive evaluation of the nodes.
- Evaluation metrics are proposed to measure the effectiveness of detection from multiple perspectives. The evaluation metrics proposed in this paper are not only generalized to measure the detection effectiveness of arbitrary source detection methods, but also applicable to evaluate the propagation capability and hiding performance of arbitrary nodes.

2 Source Hiding Analysis

In real life, networks are bounded. A paper [17] states that the boundaries of the network can affect the effectiveness of source detection methods. This is indeed the case, a simple example of which is shown in Fig. 1. When node 1 is the source and both nodes 2 and 3 are infected, the infected network is shifted towards the subtree with node 4 as the root. If the tracer uses the central method of rumor [1, 2] to locate the propagation source in the network snapshot shown in the figure, node 4 will be detected as the propagation source, while the real source (node 1) is not detected. It can be concluded that the boundary of the network will have a great impact on the detection of the source.

Further, the location of sources in the finite network will affect the detection effectiveness of the source detection method. As shown in Fig. 2, node A is a source in the network of Fig. 2(A) and node B is a source in the network of Fig. 2(B). Obviously, due to the different positions of nodes A and B in the network, their own hiding is also different. If we use the rumor center method [18, 19] to locate the source in snapshot 3, we can accurately locate the source A in Fig. 2(A), but cannot accurately locate the source B in Fig. 2(B). Therefore, the difficulty of locating the source at different locations based on the same source detection method is different.

To verify the above conclusions, we used several classical source detection methods including the rumor centrality (RC) [2] method, the Jordan centrality (JC) [16] method,

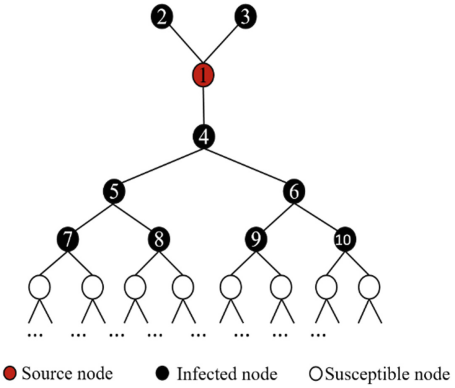

Fig. 1. Example network used to illustrate the impact of network boundaries.

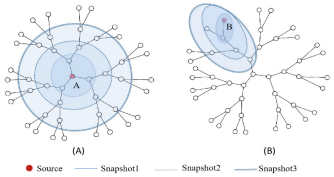

Fig. 2. Example network to illustrate the effect of different location sources on results.

the reverse infection (RI) [20] algorithm, the dynamic age (DA) [21] method and a maximum a posteriori (MAP) estimator [22] for sources at different locations to test the hiding of sources at different locations. First, the underlying network is set as a scale-free network with the number of nodes of 500. Then, the positions of the nodes are described based on the closeness centrality [23] of the nodes. Finally, we use the above method to detect the sources at different locations and calculate the distance between the real source and the detected estimated source, denoted by E. The final test results are shown in Fig. 3.

The horizontal axis of the figure indicates the closeness centrality of the source, and the vertical axis indicates the mean value of the error distance obtained by detecting it M times ($M \geq 500$). The blue scatter points are the actual detection result data points, and the red straight line is the fitted line. It is clear from the figure that the error distance increases with increasing source closeness centrality for any source detection method. This indicates that the effectiveness of the source detection method is directly related to the location of the propagating source, and that sources located at the network boundary have better hiding.

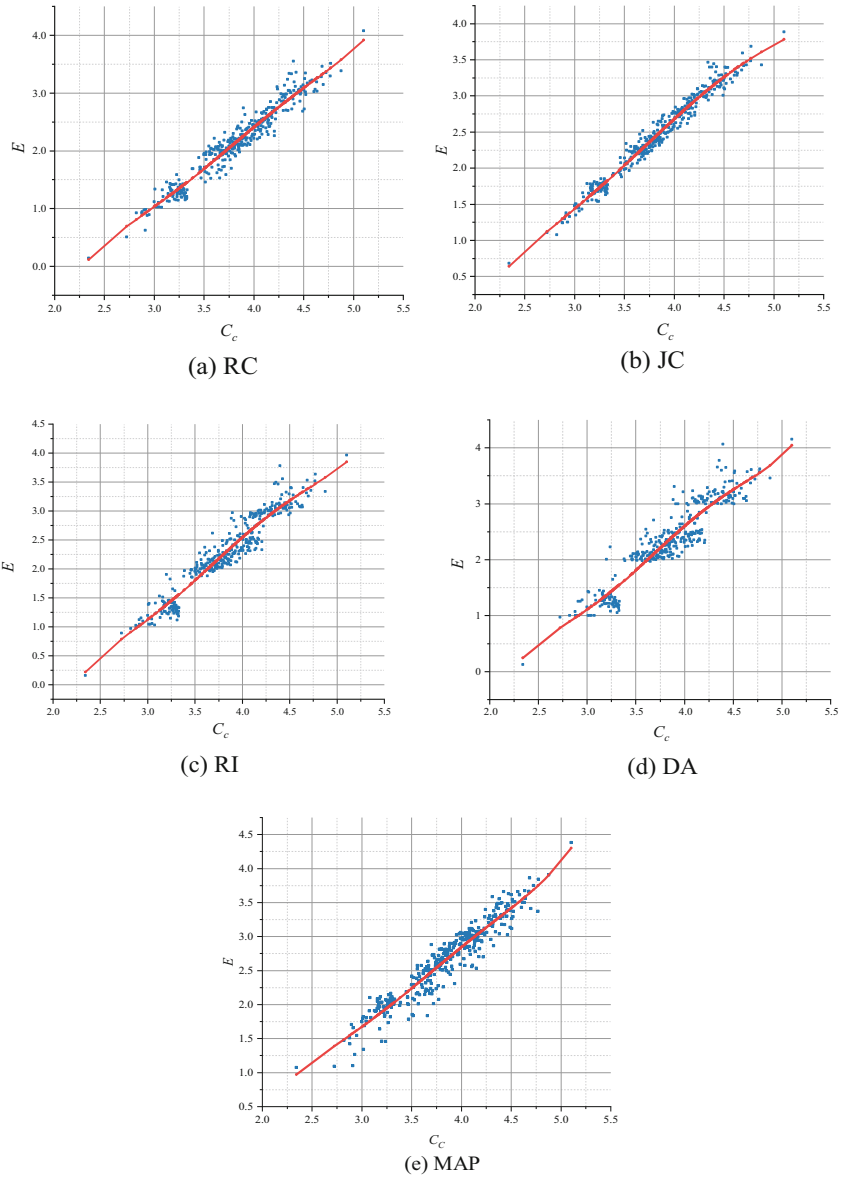

Fig. 3. Detection results for different location sources.

3 Metric and Method

We assume that the propagation process starts at a single source and use an SI model based on the Gillespie algorithm [15] to simulate the propagation process. Table 1 shows some important symbols.

Table 1. Main mathematical symbols and their meanings.

Symbol	Meaning
$G = G(V, E)$	The network structure, V is the set of nodes and E is the set of edges
$G_I = G_I(V_I, E_I)$	An infected subgraph of G
v^*	The real infection source
$\hat{v}$	The detected source
$N(v_i)$	The set of neighbor nodes of node v_i
$N_I(v_i)$	The set of neighbor nodes of node v_i that are in the infected state
$d_G(v_i, v_j)$	The shortest distance between node v_i and v_j
$g_G^{v_i}$	The number of paths through node v_i in graph G

3.1 Metric for Source Hiding

Hiding of Sources Based on Centrality Metric. Through the analysis in Sect. 2, we know that nodes located at the boundary of the network have better hiding. The closeness centrality of a node can indicate the position of the node in the network structure. In a network with N nodes, the closeness centrality of node v_i can be calculated using the following equation.

$$C_C(v_i, G) = \frac{\sum_{v_j \in V} d_G(v_i, v_j)}{N - 1} \tag{1}$$

In addition to this, the betweenness centrality [24] of a node can indicate the intermediary performance of the node in the network. By calculating it can find the endpoint of the shortest path in the network and consider this node more hided compared to other nodes. The formula for the betweenness centrality is as follows.

$$C_B(v_i, G) = \frac{g_G^{v_i}}{\sum_{v_j \in V} g_G^{v_j}} \tag{2}$$

Hiding of Sources Based on Improved K-Shell Algorithm. The K-Shell algorithm is often applied to evaluate the importance of nodes [25]. However, this method is not effective for some networks. As shown in Fig. 4, if the traditional K-Shell algorithm is used to mark the location attributes of nodes in the network, all nodes in Fig. 4(a) will be marked with the same location, and the nodes in Fig. 4(b) are divided into two levels. Obviously, this does not fully confirm the location of the structure where each node is located. In fact, the position of these nodes in the network structure is different. Therefore, we propose an improved K-Shell method for marking the location properties of nodes more efficiently. The specific steps are:

1. Find the nodes in the network with degree 1, mark their position attribute k_{ns} as 1, and then delete these nodes and their connected edges.

2. Step 1 is repeated until the degree of each node in the network is greater than 1. In each round, the k_{ns} values of the marked nodes are added by 1, and then these marked nodes and their connected edges are deleted.
3. Find the node with degree 2 in the network, continue to mark its k_{ns} as the current k_{ns} plus 1, and then delete the marked nodes and their connected edges.
4. Repeat until all nodes in the network are tagged with the corresponding location attributes.

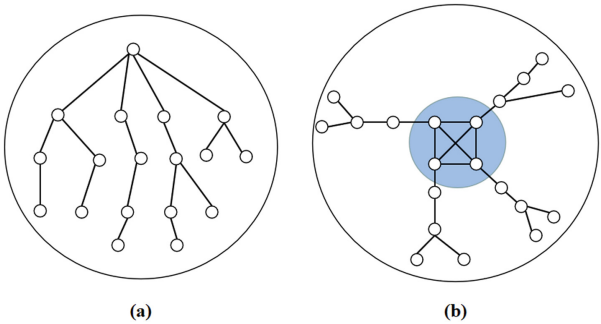

Fig. 4. Example network for which traditional K-Shell methods fail.

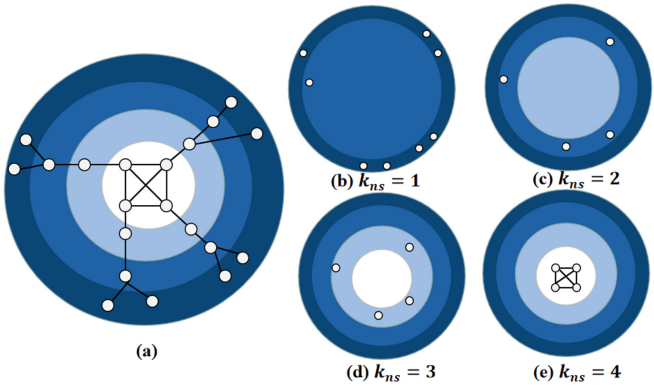

Fig. 5. Results of the improved K-Shell algorithm on the example network.

The node partitioning of our improved K-Shell algorithm for the network shown in Fig. 4(b) is illustrated in Fig. 5. This allows a more efficient description of where the nodes are located in the network structure and thus locates the nodes located at the network boundaries.

Hiding of Sources Based on Network Diameter. The maximum value of the shortest distance between any two nodes in the network is the diameter of the network [26]. As shown in Fig. 6, the shortest path between node 1 and node 7 constitutes diameter 1, and

the shortest path from node 8 to node 13 constitutes diameters 2 and 3. This shortest path spans the entire network, and its length indicates the "depth" of the network. Therefore, the node located at the center of the network diameter must be located at the center of the network structure will be called the central node. Accordingly, the nodes located at the diameter boundary must be located at the boundary of the network. It is worth mentioning that there may be multiple diameters in a network, and each diameter will have corresponding central and boundary nodes. The set of central nodes on all diameters is the set of central nodes of the network.

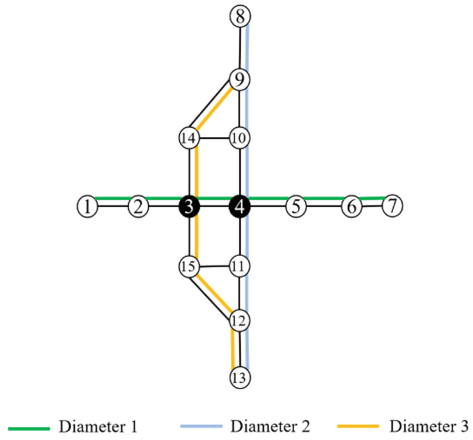

Fig. 6. Example network for illustrating network diameter.

Based on this, we can mark the location attribute l_d of the nodes in the network by the diameter. The larger the value of l_d, the better the node hiding. First, we find all the network diameters and thus the central set of nodes, marking the l_d values of the nodes in the set as 0. Then, the location attributes l_d of other nodes in the network are marked layer by layer according to the central node set. Specifically, the l_d of each non-central node takes the value of its shortest distance from all nodes in the central node set. The l_d values of the nodes of the example network in Fig. 6 are shown in Fig. 7.

3.2 Metric for Source Dissemination Capability

The propagation capability of a source indicates whether it can efficiently disseminate information to other nodes in the network. Therefore, we quantify the propagation capability of a source as the reachable propagation size in a specific time, and apply the breadth-first idea to the metric of node propagation capability. Combining the infection time and infection scale, the propagation capacity of node v_i in the network is denoted as $S(v_i)$, which is defined as follows.

$$S(v_i) = \int_1^{dia} |V_I(v_i, t)| d_t \tag{3}$$

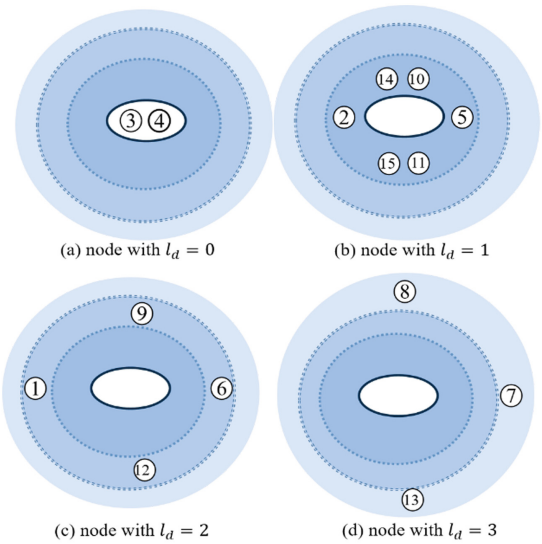

Fig. 7. The l_d values of the nodes in the example network.

where *dia* denotes the diameter of the network and $|V_I(v_i, t)|$ denotes the number of infected nodes at moment t with node v_i as the source and propagation probability of 1. Specially, since the propagation probability is set to 1, the infection process is similar to a breadth-first traversal process with v_i as the root. Specifically, $|V_I(v_i, t)|$ can be expressed as

$$\begin{cases} N_t(v_i) = N(N_{t-1}(v_i)) \\ |V_I(v_i, t)| = 1 + \sum_t |N_t(v_i)| \end{cases} \tag{4}$$

where $N_t(v_i)$ denotes the number of infected nodes added at moment t, i.e., the number of neighboring nodes of the infected nodes added at moment $t-1$, using node v_i as the source.

3.3 Bi-objective Source Hiding Method

We synthesize the above parameters based on the normalization idea to locate the nodes with both hiding and propagation ability. First, the above evaluation metrics are divided into two categories of node hiding evaluation metrics ($C_C(v_i)$, $C_B(v_i)$, $k_{ns}(v_i)$, $l_d(v_i)$) and node propagation ability evaluation metrics ($S(v_i)$), which are normalized separately. The combined calculation of hiding metrics is shown in Eq. (5) and Eq. (6).

In Eq. (5), $C_{C_{min}}$ and $C_{C_{max}}$ denote the minimum and maximum values of closeness centrality in the network, respectively, and $\widetilde{C_C}$ denotes the normalized result of closeness

centrality values. Other hiding metrics have the same meaning as closeness centrality.

$$\begin{cases} \widetilde{C_C(v_i)} = \frac{1}{4} \times \frac{C_C(v_i) - C_{C_{min}}}{C_{C_{max}} - C_{C_{min}}} \\ \widetilde{C_B(v_i)} = \frac{1}{4} \times \frac{C_B(v_i) - C_{B_{min}}}{C_{B_{max}} - C_{B_{min}}} \\ \widetilde{k_{ns}(v_i)} = \frac{1}{4} \times \frac{k_{ns}(v_i) - k_{ns_{min}}}{k_{ns_{max}} - k_{ns_{min}}} \\ \widetilde{l_d(v_i)} = \frac{1}{4} \times \frac{l_d(v_i) - l_{dmin}}{l_{dmax} - l_{dmin}} \end{cases} \tag{5}$$

$$\widetilde{E(v_i)} = \widetilde{C_C(v_i)} + \widetilde{l_d(v_i)} - \widetilde{k_{ns}(v_i)} - \widetilde{C_B(v_i)} \tag{6}$$

In Eq. (6), all the node hiding metrics are combined to obtain $\widetilde{E(v_i)}$. The higher its value is, the better the node is hidden. The normalization of the node dissemination capability metrics is as follows.

$$\widetilde{S(v_i)} = \frac{S(v_i) - S_{min}(v_i)}{S_{max}(v_i) - S_{min}(v_i)} \tag{7}$$

where $S_{min}(v_i)$ and $S_{max}(v_i)$ denote the minimum and maximum quantitative values of node propagation capacity, respectively. $\widetilde{S(v_i)}$ is the normalization of the propagation capacity, and the higher its value the stronger the propagation capacity of the node.

Finally, we combine the parameters $\widetilde{E(v_i)}$ and $\widetilde{S(v_i)}$ to locate the nodes in the network. Since the normalized parameters $\widetilde{E(v_i)}$ and $\widetilde{S(v_i)}$ both take values in the range of 0 to 1, and the larger their values, the stronger the corresponding characteristics. Therefore, we construct a two-dimensional plane based on these two parameters to describe each node in the network. In this plane, the limit optimal value (1, 1) is taken as the center. The five points with the closest distance to the center are the target points with both hiding and propagation ability. Where the calculation of the distance $D(v_i)$ can be expressed as

$$D(v_i) = \sqrt{\left(1 - \widetilde{E(v_i)}\right)^2 + \left(1 - \widetilde{S(v_i)}\right)^2} \tag{8}$$

4 Experiment

4.1 Dataset

To verify the effectiveness of our proposed method, we conducted simulation experiments on two types of generative networks (small-world network and scale-free network) and two types of real networks (E-mail network and Facebook network), respectively.

1. Small-world network: The Watts-Strogatz model was used in the experiments to reconnect each edge on the network with a probability of 0.03, resulting in a small-world phenomenon.
2. Scale-free network: A scale-free network is generated in the experiments based on the Barabasi-Albert model, and its scale-free property is ensured by continuously adding new nodes and preferentially connecting height nodes.

3. E-mail network: This network contains the communication data of about 500,000 E-mails. The addresses of the E-mails are derived as nodes in the network, and an undirected edge will be established between the sender and the receiver of the E-mails [27].
4. Facebook network: contains information about users who registered to the Facebook social platform. The nodes in the network represent users of the Facebook platform, and the edges represent the interactions between users [28].

The specific network parameters used in the experiments are shown in Table 2.

Table 2. Network parameters.

Network	Number of nodes	Number of sides	Average degree	Average clustering coefficient	Network diameter
Small-world	500	5,000	20	0.614	10
Scale-free	500	1,994	7.976	0.246	7
E-mail	1,005	25,571	25.444	0.473	7
Facebook	4,039	88,234	18.816	0.635	7

4.2 Evaluation Metrics

In this paper, the following parameters are applied to evaluate the detection results of source detection methods.

1. $e(v^*, k)$: the distance between the detected true propagation source v^* and the estimated source detected by the source detection method when the infection size is k. It is called the error distance, abbreviated as e. It is usually expressed in terms of hop count, which is obtained by calculating the number of edges passed by the shortest path between two points.
2. $E(v^*, k)$: the mean value of the distance between the real source v^* and the estimated source located by the source detection method, abbreviated as E, when the infection size is k, and the infection and detection process is executed M times with node v^* as the real source. The calculation formula is as follows.

$$E\left(v^*, k\right) = \frac{\sum_{i=1}^{M} e_i(v^*, k)}{M} \tag{9}$$

3. $t(v^*, k)$: the time required to reach the infection size k with node v^* as the source, abbreviated as t.

4. $T(v^*, k)$: the average value of the time required to execute the infection and detection process M times with node v^* as the source, abbreviated as T, and calculated as follows.

$$T\left(v^*, k\right) = \frac{\sum_{i=1}^{M} t_i(v^*, k)}{M} \tag{10}$$

5. $\overline{T}$: the mean value of the time required to traverse each node in the network as a source and reach a fixed infection size k, calculated as follows.

$$\overline{T} = \frac{\sum_{v^* \in V} T(v^*, k)}{|V|} \tag{11}$$

6. $\overline{Error}$: the mean value of the error distance obtained by traversing each node in the network as a source when the infection size k is fixed, which will be abbreviated as $\overline{E}$ in the later section and calculated as follows.

$$\overline{Error} = \frac{\sum_{v^* \in V} E(v^*, k)}{|V|} \tag{12}$$

It is worth mentioning that the detection result obtained when the number of randomizations is large enough will be the same as the result of traversing the source, so we traverse the source to obtain this value in our experiments.

7. $P_e(v^*, k)$: the probability that the error distance in the detection result is e when the node v^* is set as the source and the infection size is k, which will be abbreviated as P_e in the later section and calculated as follows.

$$P_e\left(v^*, k\right) = P\left(e\left(v^*, k\right) = e | v^*\right) \tag{13}$$

4.3 Experimental Results on Scale-Free Network

In the scale-free network, we rate the nodes in the network based on their hiding and propagation capabilities, and the top five nodes are shown in Table 3, and the nodes shown in the table are used as sources in turn for the experiments.

Table 3. Information about nodes in scale-free network.

Source ID	S	k_{ns}	l_d	C_C	C_B	D
255	2,514	4	3	3.980	0.703	0.720
370	2,479	4	3	4.050	0.505	0.724
352	2,462	4	3	4.084	0.549	0.726
473	2,418	4	3	4.172	0.364	0.731
468	2,404	4	3	4.200	0.260	0.733

To test the source hiding, we use the five algorithms described in Sect. 2 to locate the propagation source respectively, and the experimental results are shown in Fig. 8.

The horizontal axis in the figure indicates the infection size, and the vertical axis indicates the error distance between a specific source and a randomly selected source

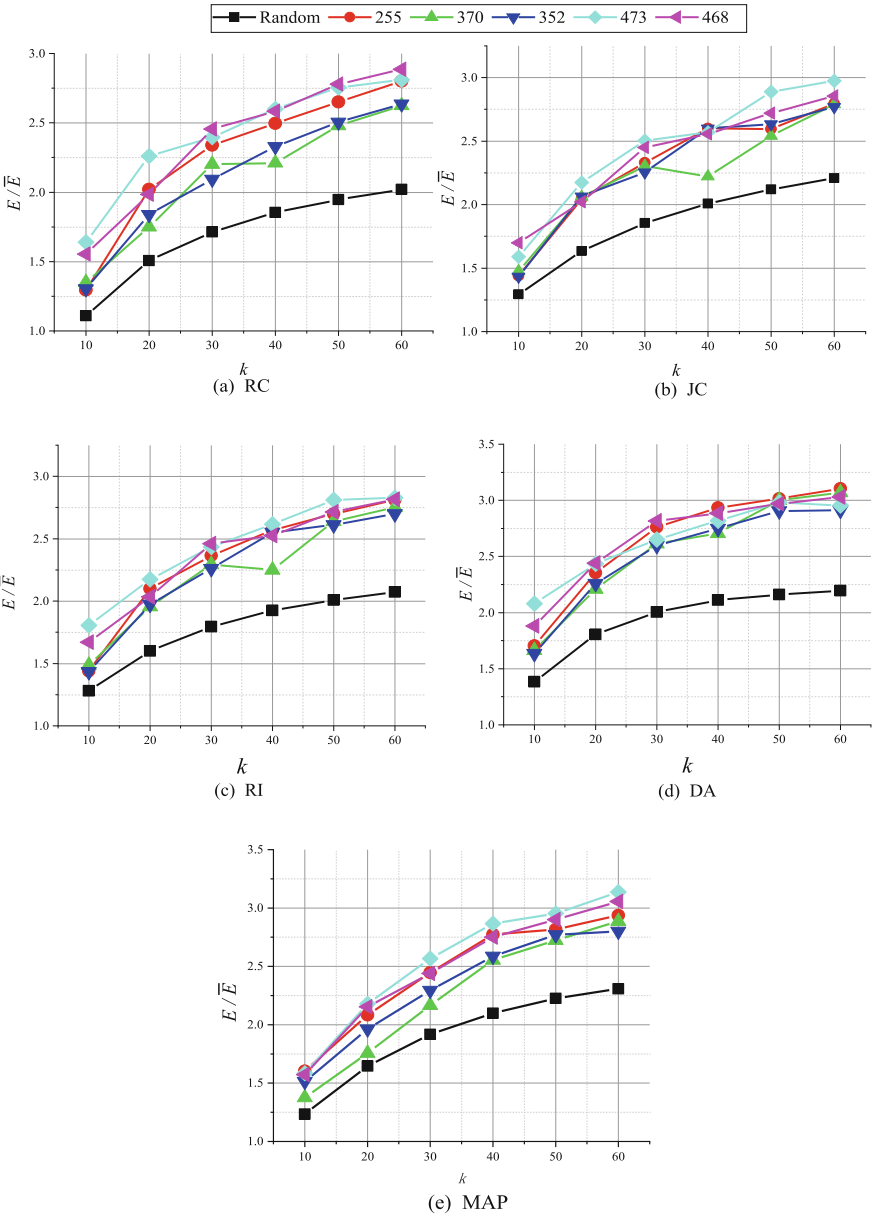

Fig. 8. Experimental results of five source detection methods on scale-free network.

in the detection results. It can be clearly seen that the selection of sources based on the method described in this paper will improve the error distance of the detection results compared to the random selection of sources. This not only shows that sources taken for the experiments have strong hiding, but also demonstrates that our proposed heuristic method can effectively measure the hiding of nodes and thus locate the nodes with strong hiding.

To test the propagation ability of the source, we compare the propagation efficiency of a randomly selected source and a source selected with heuristics at different infection sizes. Specifically, the propagation efficiency is expressed as the time required to propagate to a specified size with the node as the source. A shorter time required indicates a higher propagation efficiency and a higher propagation capability of the source. Similarly, traversing the nodes in the network as sources, the mean value of the infection time is calculated and denoted as $\overline{T}$ and the infection time of other source-specific nodes is denoted as T. The experimental results are shown in Fig. 9. Compared with the randomly selected source, the source selected based on the composite scoring method has a stronger propagation capability, and it can quickly infect the specified number of nodes to reach the expected infection scale.

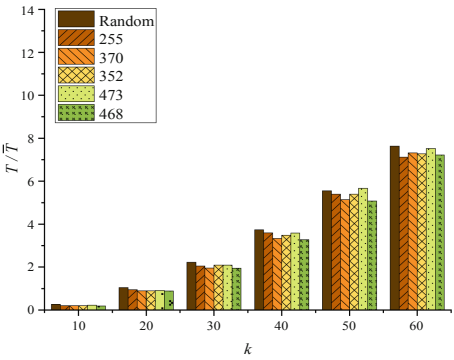

Fig. 9. Source propagation capacity comparison on scale-free network.

Table 4 shows sources selected for this policy for a number of 60 infected nodes. Sources in the table are ranked according to their overall scores.

From the experimental data, it can be seen that the best propagation source for strategy selection is the node with the strongest overall capability. For example, source 255 is not the most well hidden, but its propagation ability is higher than other nodes. Overall, the sources selected by the source hiding strategy have higher hiding and propagation capabilities than the randomly selected sources. In addition, the results in the table show that the detection effect is different when different source detection methods are used with the same node as the source. For the sources selected by the strategy, the JC method is less effective in detecting them compared to other source detection methods. However, the JC method is not the worst performer when looking at the detection results of all nodes in the network. This indicates that the detection results of the JC method are more

Table 4. Comprehensive comparison of experimental results on scale-free networks

Source ID	$T/\overline{T}$	$E/\overline{E}$				
		RC	JC	RI	DA	MAP
255	7.122	2.800	3.105	2.795	2.810	2.937
370	7.320	2.625	3.070	2.790	2.750	2.887
352	7.279	2.635	2.910	2.770	2.700	2.800
473	7.521	2.810	2.955	2.975	2.830	3.137
468	7.220	2.885	3.030	2.855	2.815	3.057
Random	10.878	2.021	2.196	2.209	2.073	2.308

correlated with the source location and its detection is more susceptible to the influence of the source location.

4.4 Experimental Results on Small-World Network

In the small-world network, the integrated capability of each node in the network is measured based on the bi-objective source hiding method, and the node with strong integrated capability is selected as the source. The information of the top five nodes in the comprehensive ability score in the calculation results is shown in Table 5.

The experimental results of testing source hiding using five source detection methods are shown in Fig. 10. The horizontal axis in the figure indicates the infection scale, and the vertical axis indicates the detection results for specific sources and randomly selected sources.

In this case, the larger the value of the error distance, the worse the detection and the better the hiding effect on the source. Obviously, compared to randomly selected sources, the error distance of the detection results of policy-selected sources is greater and the detection is worse. This shows that sources based on the composite scoring method in small-world network can still select sources with high hiding, successfully avoid detection, and ensure the anonymity of the sources.

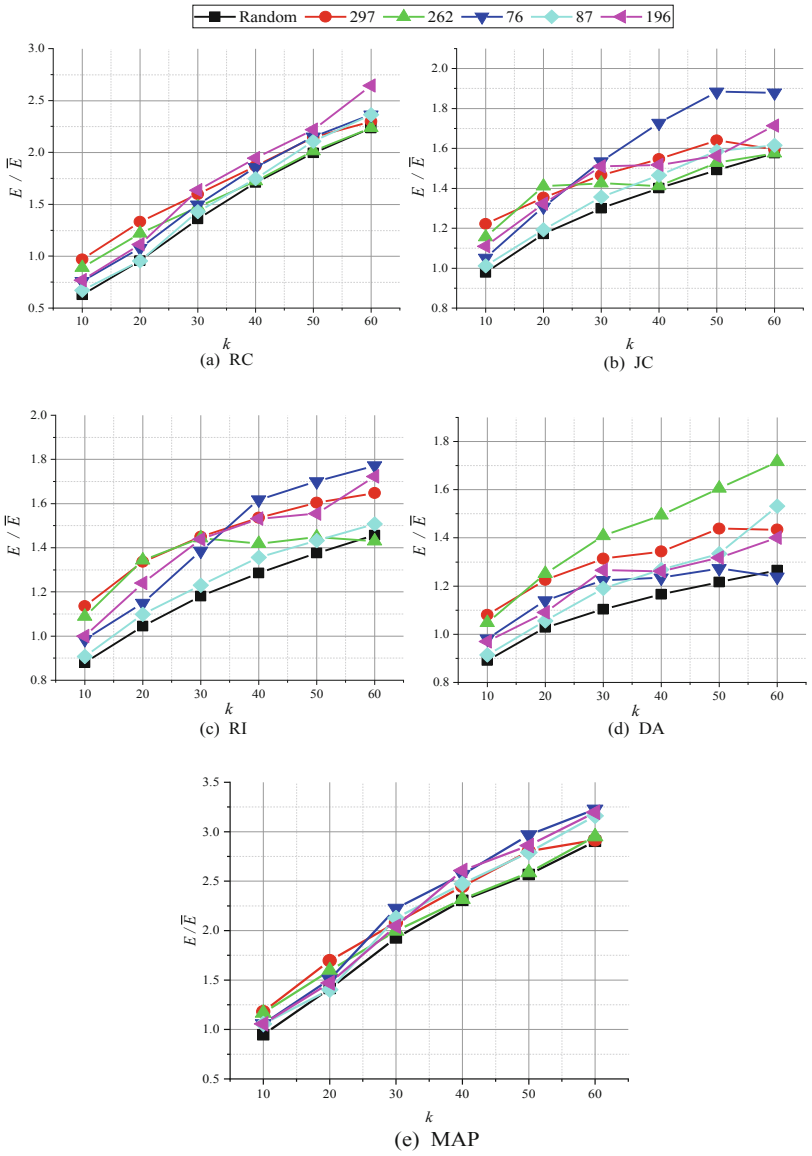

Fig. 10. Experimental results of five source detection methods on small-world network.

The results of testing the propagation capability of the sources selected by the strategy at different infection sizes are shown in Fig. 11. The experimental results vary at different scales, and the detection effect for different sources at the same scale also differs. For example, node 297 and node 262 lags behind other nodes in terms of propagation capability, while nodes 87 and 196 are consistently higher than randomly selected sources.

Table 5. Information about nodes in small-world network.

Source ID	S	k_{ns}	l_d	C_C	C_B	D
297	1,880	16	2	5.251	2.201	0.840
262	1,965	16	1	5.080	19.660	0.914
76	1,775	18	3	5.461	1.707	0.932
87	1,681	18	3	5.649	2.101	0.935
196	1,659	18	3	5.693	0.862	0.939

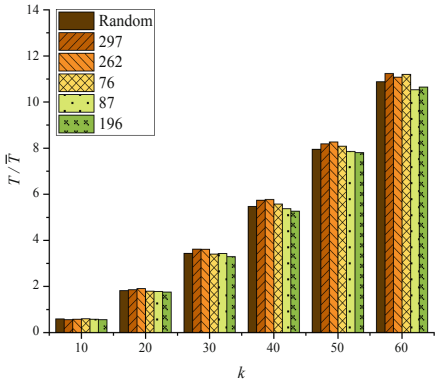

Fig. 11. Source propagation capacity comparison on small-world network.

This is because there are more connected edges between nodes in a small-world, and nodes located at different locations in the network have similar propagation capabilities.

The experimental results of the policy-selected source and the random-selected source in the small-world network when setting the infection size to 60 are shown in Table 6.

Table 6. Comprehensive comparison of experimental results on small-world network.

Source ID	$T/\overline{T}$	$E/\overline{E}$				
		RC	JC	RI	DA	MAP
297	11.233	2.298	1.434	1.596	1.648	2.917
262	11.075	2.235	1.716	1.587	1.466	2.950
76	11.193	2.364	1.238	1.880	1.772	3.227
87	10.530	2.364	1.532	1.616	1.508	3.160
196	10.649	2.644	1.400	1.714	1.722	3.197
Random	10.878	2.234	1.266	1.576	1.457	2.943

The results show that the policy-selected sources have similar propagation ability to the randomly selected sources, but the hiding is higher than that of the randomly selected sources. This is because compared to scale-free network, small-world networks have higher clustering properties and smaller average shortest distance between nodes, making the size of the infected networks generated in the same time similar when different nodes are sources.

4.5 Experimental Results on Real Network

The effectiveness of the source selection method is demonstrated by comparison in the simulations of small-world network and scale-free network. On this basis, we tested the integrated node scoring method in real E-mail networks and Facebook networks. The information of the top five nodes in both networks is shown in Table 7 and Table 8, respectively. The number of nodes in the E-mail network and the Facebook network are 1005 and 4039, the number of edges are 25571 and 88234, respectively.

The number of infected nodes is set to 100 and 300 in the experiments of E-mail network and Facebook network, respectively. The propagation traceability process is then executed 500 times. The results of the different source detection methods are shown in Table 9 and Table 10.

Table 7. Information about nodes in the E-mail network.

Source ID	S	k_{ns}	l_d	C_C	C_B	D
625	6,020	6	2	20,080,761.58	0	0.853
988	5,855	2	2	20,080,761.75	0	0.854
911	5,972	3	2	20,080,761.63	0	0.846
773	5,892	1	2	20,080,761.71	0	0.847
780	5,886	1	2	20,080,761.71	0	0.848

Table 8. Information about nodes in the Facebook network.

Source ID	S	k_{ns}	l_d	C_C	C_B	D
2,814	23,145	2	4	3.270	0	0.744
2,838	23,158	4	4	3.267	0	0.748
2,885	23,158	4	4	3.267	0	0.748
3,003	23,158	4	4	3.267	0	0.748
2,704	22,406	2	4	3.453	0	0.757

The probability of successful detection is 0 for all source detection methods we tested on the E-mail network. There is a high probability that the error distance will

Table 9. Average detection time, average error distance, and the probability (%) of each error distance on the E-mail network.

Source ID	T	Method	E	P_0	P_1	P_2	P_3	P_4
625	0.882	RC	2.028	0.00	1.75	93.75	4.50	0.00
		JC	2.088	0.00	3.00	85.25	11.75	0.00
		DA	1.985	0.00	3.75	94.00	2.25	0.00
		RI	1.993	0.00	2.50	95.75	1.75	0.00
		MAP	2.043	0.00	2.25	91.50	6.00	0.25
988	0.864	RC	2.758	0.00	0.25	24.00	75.50	0.25
		JC	2.425	0.00	0.75	56.00	43.25	0.00
		DA	2.705	0.00	0.00	29.50	70.50	0.00
		RI	2.475	0.00	0.00	52.50	47.50	0.00
		MAP	2.720	0.00	0.25	28.25	70.75	0.75
911	0.888	RC	2.055	0.00	0.75	93.00	6.25	0.00
		JC	2.168	0.00	1.25	80.75	18.00	0.00
		DA	2.013	0.00	0.00	98.75	1.25	0.00
		RI	2.003	0.00	0.25	99.25	0.50	0.00
		MAP	2.105	0.00	0.75	88.00	11.25	0.00
773	0.906	RC	2.15	0.00	0.75	83.50	15.75	0.00
		JC	2.128	0.00	1.75	83.75	14.50	0.00
		DA	2.1	0.00	0.25	89.50	10.25	0.00
		RI	2.013	0.00	1.00	96.75	2.25	0.00
		MAP	2.198	0.00	0.25	80.00	19.50	0.25
780	0.869	RC	2.005	0.00	3.00	93.50	3.50	0.00
		JC	2.148	0.00	0.25	84.75	15.00	0.00
		DA	2.013	0.00	0.00	98.75	1.25	0.00
		RI	2.005	0.00	0.00	99.50	0.50	0.00
		MAP	2.028	0.00	5.75	85.75	8.50	0.00

be 2 or 3. And on the Facebook network, in rare cases the real source can be detected with a very small probability. In the vast majority of cases, the error distance is 1 or 2. The maximum error distance can even reach 5. This shows that the sources selected according to our proposed method are highly hidden and can effectively evade tracking by multiple source detection methods.

Table 10. Average detection time, average error distance, and the probability (%) of each error distance on the Facebook network.

Source ID	T	Method	E	P_0	P_1	P_2	P_3	P_4	P_5
2,814	33.941	RC	1.245	0.00	77.25	21.50	0.75	0.50	0.00
		JC	1.625	0.50	48.50	39.50	11.00	0.50	0.00
		DA	1.333	0.00	69.00	29.75	0.50	0.50	0.25
		RI	1.388	0.00	64.50	32.25	3.25	0.00	0.00
		MAP	1.245	0.00	78.25	20.00	1.00	0.50	0.25
2,838	36.193	RC	1.393	0.00	65.75	31.25	1.00	2.00	0.00
		JC	1.693	0.75	45.75	38.00	14.50	1.00	0.00
		DA	1.535	0.00	53.00	43.00	1.50	2.50	0.00
		RI	1.478	0.00	56.50	39.25	4.25	0.00	0.00
		MAP	1.388	0.00	68.25	27.00	2.50	2.25	0.00
2,885	36.158	RC	1.370	0.00	68.25	27.75	2.75	1.25	0.00
		JC	1.663	0.50	45.00	42.25	12.25	0.00	0.00
		DA	1.510	0.00	55.25	40.25	2.75	1.75	0.00
		RI	1.473	0.00	57.00	38.75	4.25	0.00	0.00
		MAP	1.380	0.00	69.25	25.00	4.25	1.50	0.00
3,003	36.494	RC	1.328	0.00	70.75	27.00	1.00	1.25	0.00
		JC	1.653	1.75	43.75	42.25	12.00	0.25	0.00
		DA	1.488	0.00	56.75	39.75	1.50	2.00	0.00
		RI	1.465	0.00	58.00	38.00	3.50	0.50	0.00
		MAP	1.345	0.00	71.00	24.75	3.00	1.25	0.00
2,704	34.458	RC	1.433	0.00	65.75	28.00	3.50	2.75	0.00
		JC	1.798	1.75	36.00	44.50	16.25	1.50	0.00
		DA	1.570	0.00	58.00	32.50	4.25	5.00	0.25
		RI	1.538	0.00	53.00	40.50	6.25	0.25	0.00
		MAP	1.450	0.00	67.25	24.00	5.25	3.50	0.00

5 Conclusion

In this paper, we propose a bi-objective source hiding method that takes into account both source hiding and propagation capability based on the propagation characteristics of the network. In addition, this paper defines a variety of evaluation metrics for testing the propagation capability and hiding of nodes in different types of network structures. Experiments are conducted in small-world network, scale-free network, E-mail network and Facebook network. The results show that the policy-selected sources have higher propagation capacity compared to randomly selected sources. We also use a variety

of source detection methods to detect the source, and the results show that the source selected by the policy can successfully evade detection.

References

1. Shah, D., Zaman, T.: Detecting sources of computer viruses in networks: theory and experiment. In: Proceedings of the ACM SIGMETRICS International Conference on Measurement and Modeling of Computer Systems, pp. 203–214 (2010)
2. Shah, D., Zaman, T.: Rumors in a network: who's the culprit? IEEE Trans. Inf. Theory **57**(8), 5163–5181 (2011)
3. Wikipedia. Wickr. [EB/OL] (5 February 2022). https://www.wickr.com/
4. Wikipedia. FireChat. [EB/OL] (5 February 2022). http://opengarden.com/firechat/
5. Chen, J., Chen, L., Chen, Y., et al.: GA-based Q-attack on community detection. IEEE Trans. Comput. Soc. Syst. **6**(3), 491–503 (2019)
6. Fionda, V., Pirro, G.: Community deception or: how to stop fearing community detection algorithms. IEEE Trans. Knowl. Data Eng. **30**(4), 660–673 (2017)
7. Waniek, M., Michalak, T.P., Wooldridge, M.J., et al.: Hiding individuals and communities in a social network. Nat. Hum. Behav. **2**(2), 139–147 (2018)
8. Lü, L., Zhou, T., Zhang, Q.M., et al.: The H-index of a network node and its relation to degree and coreness. Nat. Commun. **7**(1), 1–7 (2016)
9. Morone, F., Makse, H.A.: Influence maximization in complex networks through optimal percolation. Nature **524**(7563), 65–68 (2015)
10. Kenett, D.Y., Preis, T., Gur-Gershgoren, G., et al.: Dependency network and node influence: application to the study of financial markets. Int. J. Bifurc. Chaos **22**(07), 1250181 (2012)
11. Conitzer, V., Panigrahi, D., Zhang, H.: Learning opinions in social networks. In: International Conference on Machine Learning, pp. 2122–2132 (2020)
12. Fanti, G., Kairouz, P., Oh, S., et al.: Metadata-conscious anonymous messaging. IEEE Trans. Signal Inf. Process. Over Netw. **2**(4), 582–594 (2016)
13. Fanti, G., Kairouz, P., Oh, S., et al.: Rumor source obfuscation on irregular trees. ACM SIGMETRICS Perform. Eval. Rev. **44**(1), 153–164 (2016)
14. Luo, W., Tay, W.P., Leng, M.: Infection spreading and source identification: a hide and seek game. IEEE Trans. Signal Process. **64**(16), 4228–4243 (2016)
15. Luo, W., Tay, W.P., Leng, M.: On the universality of Jordan centers for estimating infection sources in tree networks. IEEE Trans. Inf. Theory **63**(7), 4634–4657 (2017)
16. Zhou, H., Jagmohan, A., Varshney, LR.: Generalized Jordan center: a source localization heuristic for noisy and incomplete observations. In: 2019 IEEE Data Science Workshop (DSW), pp. 243–247 (2019)
17. Yu, P.D., Tan, C.W., Fu, H.L.: Rumor source detection in finite graphs with boundary effects by message-passing algorithms. In: IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining, pp. 175–192 (2018)
18. Brzozowski, M.J., Adams, P., Chi, E.H.: Google+ communities as plazas and topic boards. In: Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems, pp. 3779–3788 (2015)
19. Anderson, M., Caumont, A.: How social media is reshaping news. Pew Research Center, vol. 24 (2014)
20. Zhu, K., Ying, L.: Information source detection in the SIR model: a sample-path-based approach. IEEE/ACM Trans. Netw. **24**(1), 408–421 (2014)
21. Fioriti, V., Chinnici, M., Palomo, J.: Predicting the sources of an outbreak with a spectral technique. Appl. Math. Sci. **8**(135), 6775–6782 (2014)

22. Chang, B., Chen, E., Zhu, F., et al.: Maximum a posteriori estimation for information source detection. IEEE Trans. Syst. Man Cybern.: Syst. **50**(6), 2242–2256 (2018)
23. Albert, R.: Scale-free networks in cell biology. J. Cell Sci. **118**(21), 4947–4957 (2005)
24. Luo, W., Tay, W.P., Leng, M.: Identifying infection sources and regions in large networks. IEEE Trans. Signal Process. **61**(11), 2850–2865 (2013)
25. Namtirtha, A., Dutta, A., Dutta, B.: Weighted kshell degree neighborhood: a new method for identifying the influential spreaders from a variety of complex network connectivity structures. Expert Syst. Appl. **139**, 112859 (2020)
26. Chen, B., Zhu, W.X., Liu, Y.: Algorithm for complex network diameter based on distance matrix. J. Syst. Eng. Electron. **29**(2), 336–342 (2018)
27. Jure Leskovec. http://snap.stanford.edu/data/E-mail-Enron.html. Accessed 5 Jan 2021
28. Jure Leskovec. https://snap.stanford.edu/data/egonets-Facebook.html. Accessed 5 Jan 2021