PJC: A Multi-source Method for Identifying Information Dissemination in Networks

Yong Ding, Xiaoqing Cui, Huiyong Wang(✉), and Yujue Wang

Guangxi Key Laboratory of Cryptography and Information Security,
Guilin University of Electronic Technology, Guilin 541004,
People's Republic of China
why608@163.com

Abstract. With the development of science and technology, the world has become increasingly closely linked. While enjoying the convenience brought by the Internet, we are also facing the danger of risk dissemination. This problem has become more challenging in real-world networks. In this paper, in view of the outbreak of network threats, such as malware, computer viruses, rumors, etc. It is particularly important to identify the source of network threats. In this paper, we have done the following work. Firstly, we draw on the propagation models from epidemiology and design an algorithm partitioned Jordan Center (PJC) to locate the multiple propagation sources. Then, by establishing an extended model originated from propagation sources, we derive the number of sources of estimation. In order to evaluate the performance of the proposed method, a series of experiments were carried out in real-world network topologies. Experimental results show that the method is more accurate than the existing methods.

Keywords: Information dissemination · PJC · Identification of multi-source

1 Introduction

In today's increasingly interconnected world, while we enjoy the convenience of the world, we are also affected by new diffusion risks. For example, the rapid development of the Internet of Things has made more user contacts more secure. However, we need to guard against loopholes in information transmission technology. "Intelligent" Internet of Things devices may be an entry point for them to attack the network. Similarly, the rapid popularity of social media and mobile Internet devices has enabled people to easily and quickly access news and other information from social networks [1–3]. Rumor spread has entered the new media era, which makes the efficiency and harmfulness of rumor spread reach an unprecedented level. For biological viruses, the capture of highly pathogenic H5N1 influenza virus always threatens people's health. It is essential to identify the location of the source and find the number of the sources.

Assuming that a threatening message may begin to spread over different sources and times, after a certain period of time, we observe that nodes on the network are infected. Due to most real-world networks are complex, most of the previous work on

J. Li et al. (Eds.): SPNCE 2019, LNICST 284, pp. 268–281, 2019.
https://doi.org/10.1007/978-3-030-21373-2_20

source identification was based on single source detection [4, 5] or simple network topology. However, the multi-source heuristic algorithm proposed in recent years can't detect the real source of infection, and the average error is relatively large. For example, [6] proposes a multi-rumor recognition method to identify multiple sources in tree networks, which is difficult to implement in large networks.

In this paper, we propose a novel source identification method to overcome the challenges. First, In the process of information dissemination, we focus on considering that there is a certain gap between the sources when multiple source nodes propagate. When the distances between sources are close, too many overlapping nodes make the same nodes get the same infection, which makes the propagation range smaller. Second, In the real network, the diffusion in the network is complex in time and space. For a clearer understanding, we use effective distance [7]. The concept of effective distance reflects that the small propagation probability between nodes effectively corresponds to the large distance between nodes. The relative arrival time from the source to the node does not depend on any parameters, but is linear to the effective distance between the source and the infected node. In order to determine multiple sources, we firstly need to partition the infection map to minimize the effective distance between the source and the infected node. The node that minimizes the maximum distance to the infected area is considered the source of the propagation.

This paper mainly makes the following contributions: Firstly, we propose a novel method of partitioned Jordan Center to identify multiple sources. We prove that our method is convergent. The experimental results show that this method is superior to other methods. Secondly, by locating the source, we use an effective algorithm to estimate the source diffusion time. Finally, according to the estimated diffusion time, we can accurately estimate the number of diffusion sources.

The rest of this paper is organized as follows. Sect. 2 introduces the preliminary knowledge of the relevant background. Section 3 is the problem formulation of multi-source identification. Section 4 presents a method of Partitioned Jordanian centers for identifying multi-source problems. Section 5 evaluates the proposed method in the actual network topology. Section 6 is related work, and Sect. 7 is a summary of this paper.

2 Preliminary

In this section, we introduce relevant background knowledge, Propagation models and distances. Usually in these Propagation models, we divide the research objects into three categories, each of which has its own state. It mainly includes: the first is the susceptible state (S). Nodes in this state refer to healthy nodes, which are easy to be infected, but not yet infected; the second is Infected state (I). Nodes in this state are infected nodes, which are infectious; the third is Recovered state (R). Nodes in this state are immunized, not infected, or dead. Specific description in the following section. A specific meaning of symbols is given in Table 1 below.

Table 1. Used notations

Notation	Meaning
$P_s(i,t)$	The probability that node i is a susceptible node at time t
$P_I(i,t)$	The probability that node i is the infected node at time t
$v(i,t)$	Probability of node i being infected by neighboring nodes
$\mu_{j,i}$	Probability of node j propagating to node i
$m(i,j)$	Effective distance from node i to node j
$\alpha, \sigma(\alpha)$	Path, The sum of the effective lengths along the edge of the path
$\overrightarrow{S}, S^*$	Estimated source, Propagation source
B_n, D	Infection map, Infected partition

2.1 Propagation Model

Researchers mainly use three propagation models: SIR model, SIS model and SI model. SIR model considers the recovery process. Nodes are initially sensitive. They can be infected with the spread of risk and spread the threat to the next node. But the node can recover and become insensitive. The model deals with infection and curing processes $S \rightarrow I \rightarrow R$. The transmission schematic of SIR is shown in Fig. 1(A). The susceptible person S appears to be in a healthy state. It changes into the infected person I through direct contact with the infected person with a certain probability P. Infected person I regains health status and acquires immunity with the probability of u, thus becoming restorer R.

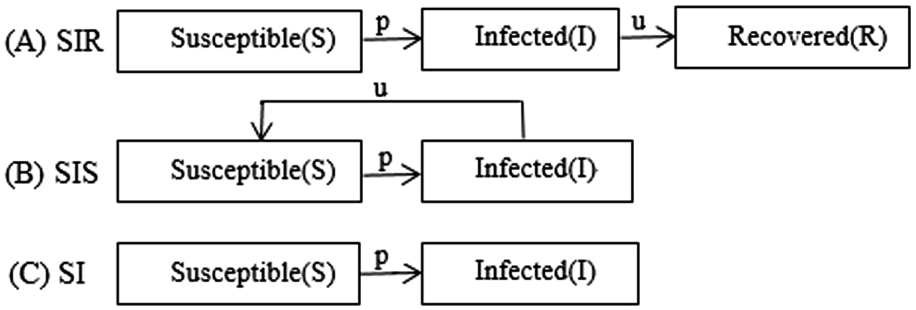

Fig. 1. The propagation diagrams of the three propagation models are (A) SIR (B) SIS (C) SI from top to bottom.

In the SIS model, infected nodes may become susceptible again after being cured. The model represents the process of infection and recover $S \rightarrow I \rightarrow S$. The susceptible node S will be transformed into an infected node with a certain probability p after contact with the infected node I, and some infected nodes will be restored to a susceptible node with a certain probability u. The propagation diagram of SIS is shown in Fig. 1(B).

In this paper, we adopt SI model. SI model is more convenient and applicable. We assume that infection spreads at discrete time steps. Each node is initially sensitive, and it can be infected as the risk spreads. Once a node is infected, it will always be infected. At time t, the probability that a vulnerable node *i* is infected by any infected neighbor is v(i,t). Therefore, we can calculate the probability that node *i* is a susceptible node at time t is

$$P_s(i,t) = [1 - v(i,t)] \cdot P_s(i,t-1) \tag{1}$$

Then, we can get that the probability of node *i* being infected at time t is

$$P_I(i,t) = v(i,t) \cdot P_s(i,t-1) + P_I(i,t-1) \tag{2}$$

We use μ_{ji} to represent the propagation probability of node j to neighbor node *i*, and then we can calculate the probability that node *i* is infected by neighbor node is

$$v(i,t) = 1 - \prod_{j \in N_i} [1 - \mu_{j,i} \cdot P_I(i,t-1)]. \tag{3}$$

Here, N_i represents the set of neighbors of node *i*. This model reflects the probability that any node is in different states at different times. Each time hop can mean one minute, one hour or one day.

2.2 Definition of Distance

Brockmann and Helbing [7] proposed a new concept to solve geometric problems in complex propagation processes by the relationship between propagation probability and effective distance between nodes. The effective distance from a node *i* to the neighbor node j is expressed as:

$$m(i,j) = 1 - \log \mu_{ij} \tag{4}$$

Where μ_{ij} is represented as the propagation probability from node *i* to node j. This formula reflects the small propagation probability from node *i* to node j is equivalent to the large distance between nodes. The length of path $\alpha = \{u_1, \cdots u_l\}$ is defined as the sum of effective length $\sigma(\alpha)$ along the edge of path. The effective distance from any node *i* to node j is defined as the length of the shortest path, which is expressed as:$d(i,j) = \min_\alpha \sigma(\alpha)$. We refer to the effective distance from node *i* to node j as the distance, denoted by $d(i,j)$. Given any set $A \subset V$, the maximum distance between node v and any node j is expressed as:

$$\overline{d}(v,A) = \max_{u \in A} d(v,u) \tag{5}$$

We call $\overline{d}(v,V_e)$ the maximum distance between node v and the infected range of any v. Nodes with the smallest infection range are called Jordan Center [8].

From the above formula, the Jordanian centrality of a node is considered to be the maximum distance from that node to any other infected node [9]. The Jordanian Center represents the node with the smallest Jordanian centrality.

3 Problem Formulation

Suppose that at time T = 0, there are k sources, and the $S^* = \{s_1, s_2, \cdots s_k\}$ diffusion propagation begins at the same time. After a few ticks, n nodes were infected. These nodes form a connected infected graph B_n, and each infected area is $D_i(\subset B_n)$. Let $D = \bigcup_{i=1}^{k} D_i$ be the partition of the infected area, satisfying $D_i \cap D_j = \emptyset$ and $i \neq j$ between partitions. Each partition is a subgraph of the connection of B_n, and the source node s_i can be found in each partition. We try to keep the source node s_i and s_j as far apart as possible. Figure 2 shows a certain distance between source nodes s_1 and s_2, so that many overlapping nodes can be avoided between each region. This can increase the probability of spreading. Assuming the infected node $v_j \in D_i$, the node v_j is infected in a short time. It means that it has a shorter distance to the corresponding source than to other sources. We consider the minimum of the maximum distance from the infected node to any other infected node as the source of propagation.

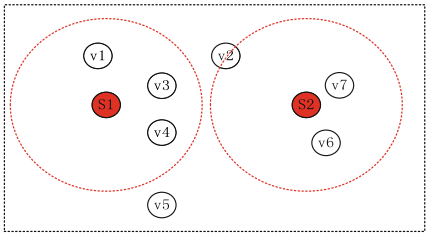

Fig. 2. Topological schematic diagram of separated propagation source nodes.

Our aim is to identify the corresponding partition D of a set of propagation sources S^* and infection graph B_n, so as to minimize the objective function value f.

$$\min f_D = \sum_{i=1}^{k} \max_{u_x \in D_i / s_i} d(u_x, s_i) \tag{6}$$

4 A Partitioned Jordan Center Method for Identifying Multi-source Problems

In this section, we propose a partitioned Jordan Center method (PJC) to identify multiple sources of diffusion and corresponding infected areas. We firstly introduce a method of network partitioning to export the PJC method. Then, it is used to identify multiple sources and estimate the number of sources.

4.1 Multi-source Network Partition

Given the infected graph B_n and a series of source nodes $S^* = \{s_1, s_2, \cdots s_n\}$. The distance between any two source nodes s_i and s_j satisfies $d(s_i, s_j) > \delta$. If $d(s_i, s_j) \leq \delta$, the source node is re-selected until the condition is met. In this section, we firstly divide the infected graph B_n into K infected subgraphs, with the corresponding s_i as the center of the partition D_i. We know that any node $v_j \in B_n$ should be assigned to the partition D_i where the source s_i is located. v_j must satisfy the following conditions:

$$d(v_j, s_i) = \min_{s_t \in S} d(v_j, s_t) \tag{7}$$

For any node $v_j \in B_n$, it needs to correspond v_j to the nearest source s_i. This is similar to the Capacity Constrained Network-Voronoi Diagram (CCNVD) problem [10]. In the K-center [8] method, there is also a similar partitioning method. In future work, a common structure may be used for network partitioning.

Algorithm 1: PJC method to identify multiple sources

Input: An infection graph B_n and the number of sources k.

Initialization: An positive integer T, Randomly select a suitable set $S^{(0)} = \{s_1^{(0)}, \cdots, s_k^{(0)}\} \subseteq B_n$.

For t start from 1 to $T_{\max}$ **do**

$S^{(0)}$ as the center of the partition. Used the partition method to obtain a partition: $D^{(t)} = \bigcup_{i=1}^{k} D_i^{(t)}$.

Find a new partition center in each partition $D_i^{(t)}$ with the following formula:

$s_i^{(t)} = \arg\min_{v_j \in D_i^{(t)}} \max_{v_x \in D_i^{(t)}/v_j} d(v_j, v_x), i = 1, 2 \cdots k$.

if $S^{(t)} = S^{(t-1)}$ **then**

stop.

End

End

Output: A set of estimated sources $S^{(t)} = \{s_1^{(t)}, \cdots, s_k^{(t)}\}$.

4.2 Identifying Multiple Propagation Sources and the Infected Partition

In this section, we propose a partitioned Jordan Center method to identify multiple diffusion sources. We firstly need to find the partition D that changes the infection graph B_n, which can minimize the minimum distance from the infected node to the corresponding partition center. If we randomly select a suitable set of source nodes, Voronoi partition can divide the network into subnets, so that each node is associated with its nearest source node. Therefore, Voronoi partition can find a locally optimal B_n partition. However, to optimize partition D, we need to adjust the center of each partition to minimize the objective function value f in Eq. (6). We adjust the center of each partition by choosing a new partition center to minimize the maximum distance from the partition center to any node in the partition. Detailed partitioning Jordan

central method such as algorithm 1. The following theorem shows the convergence of the proposed method.

Theorem 1: The objective function of Eq. (6) is to decrease iteratively monotonously. Therefore, the partitioned Jordan Center is convergent.

Proof: Suppose that $S^l = \left\{s_1^i, \cdots s_n^l\right\}$ is the sources of estimation in the l times iteration. We use partitioning method to divide the infected graph B_n into $D^l = \bigcup_{i=1}^{k} D_i^l$. In this way, the objective function becomes

$$f^l = \sum_{i=1}^{k} \max_{u_x \in D_i^l / S_i^l} d(S_i^l, u_x) \tag{8}$$

after l iterations.

At the next iteration $l+1$, according to the PJC method, we calculate the center $D^l = \bigcup_{i=1}^{k} D_i^l$ of each partition again and get $S^{l+1} = \left\{s_1^{l+1}, \cdots, s_k^{l+1}\right\}$, which satisfies

$$\max_{u_x \in D_i^l / S_i^{l+1}} d(S_i^{l+1}, u_x) \leq \max_{u_x \in D_i^l / S_i^l} d(S_i^l, u_x) \tag{9}$$

Then the target function becomes

$$\overline{\overline{f}}^l = \sum_{i=1}^{k} \max_{u_x \in D_i^l / S_i^{l+1}} d(S_i^{l+1}, u_x) \tag{10}$$

From Eqs. (8) and (9), we noticed that

$$\overline{\overline{f}}^l \leq f^l \tag{11}$$

We redistribute the infected graph B_n so that the center of each infected partition is $S^{l+1} = \left\{s_1^{l+1}, \cdots s_k^{l+1}\right\}$. Let each infected node $v_j \in B_n$ be associated with the nearest central node s_i^{l+1}, and we get a new partition $D^{l+1} = \bigcup_{i=1}^{k} D_i^{l+1}$ for B_n. Thus, the objective function becomes

$$f^{l+1} = \sum_{i=1}^{k} \max_{u_x \in D_i^{l+1} / S_i^{l+1}} d(S_i^{l+1}, u_x) \tag{12}$$

in the iteration $l+1$ times.

Since each node is associated with its nearest central node s_i^{l+1}, we can know

$$f^{l+1} \leq \overline{\overline{f}}^l \tag{13}$$

From Eqs. (11) and (13), we have

$$f^{l+1} \leq \overline{\overline{f}}^{l} \leq f^{l} \tag{14}$$

Therefore, the objective function of Eq. (6) is consistently reduced, and our proposed partitioned Jordan Center method is convergent.

4.3 Identifying the Number of Propagation Sources

The heuristic algorithm is used to estimate the number of sources. By using the proposed method of source identification, we can obtain the partition D of B_n which is consistent with S^*. If the number of sources is known, the propagation time $T^{(k)}$ can be estimated by Eq. (16). In order to estimate the number of sources, we start from 1 and increase the number of source k in turn, and calculate the corresponding propagation time $T^{(k)}$ until we find $T^{(k)} = T^{(k-1)}$. We choose k (or k–1) as the number of diffusion sources. This is similar to the method [8] in estimating the number of sources.

The propagation time can be determined by the total number of time ticks of diffusion. Given an arbitrary source, the propagation time can be estimated based on the number of time hops between the source and the infected node in each region. In regional D_i, according to any source s_i, any node $v_j \in D_i$ can be found. $g(s_i, v_j)$ represents the sum of the minimum time hops between s_i and v_j, the propagation time of each region can be estimated as

$$t_i = \max\left\{g(s_i, v_j)\middle| v_j \in D_i, i = 1, 2 \cdots k\right\} \tag{15}$$

Then the propagation time of the whole infected area is

$$T = \max\{t_i | i = 1, 2 \cdots k\} \tag{16}$$

This process of propagation is actually simplified. In the real world, the propagation time of different paths with the same hop number is different. Based on SI model, we have solved this time problem in another article [11]. In this field, many current models are based on time hops [12].

5 Experiment Analysis

In this section, we make an experimental analysis of the proposed method of Partitioned Jordan Center. We tested our approach on the following network topologies: Yeast protein-protein interaction network [13], the large-scale web Facebook [14], and the North American Power Grid [15]. The basic attributes of the networks are listed in Table 2. We set the propagation probability μ_{ij} of each edge to follow the uniform distribution on (0,1). Previous work [16, 17] has proved that the distribution of propagation probability will not affect the accuracy of SI model. We randomly select different threat sources that satisfy the conditions to generate the dissemination data as

the basic fact. Under each parameter setting, our method is used to simulate 100 propagation processes and identify the source of each propagation.

Table 2. Basic attributes of real networks

Dataset	Power Grid	Yeast	Facebook
Number of nodes	4,941	2,361	45,813
Number of edges	13,188	13,554	370,532
Average degree	2.67	5.74	8.09
Maximum degree	19	64	223

5.1 Evaluation Source Detection Algorithm

In this section, we test source identification methods. In order to compare the performance with other methods, we match the estimated source $\overrightarrow{S} = \{\overrightarrow{s_1}, \cdots \overrightarrow{s_k}\}$ with the real source $S^* = \{s_1, s_2, \cdots s_k\}$. So that the total error distance between them is the smallest. The average error distance formula is

$$\Delta = \frac{1}{|S^*|} \sum_{i=1}^{|S^*|} g(s_i, \overrightarrow{s_i}) \tag{17}$$

The average error distances of the three real source network topologies are shown in Table 3. Table 3 shows that the average error distance is very small compared with other methods. This shows that our method is superior to other methods. In order to make a clearer comparison, we show a histogram of the average error distance (Δ) as shown in Figs. 3 and 4. Frequency is used to express the percentage of test times when the average error distance is fixed.

We applied the algorithm to the Yeast protein-protein interaction network, the large-scale web Facebook and the North American Power Grid. As shown in Figs. 3 and 4.

We have made histograms for different cases where the true source S is 2 and 3 respectively. When the source is 2, on the Power Grid, the error range is often active in the range of 1 to 2 hops, indicating that our method performs well. And with the Dynamic Age method [4], the average error distance shows 3–5 hops, the maximum number of errors is 3 hops. Multi-rumor-center (MRC) [18] method shows an average error range of 3 to 4 hops. The most common error is 4 hops. The K-center [8] method shows an error range of 1 to 3 hops. On the Yeast network, our method show that the average error is 2 hops. And with the Dynamic Age method, the average error distance is 3–4 hops. The average error distance between the Multi-rumor-center method and the K-Center method is 2–4 hops, and the most experimental results are 3 hops. On the Facebook network, the most performance of our method is 1–3 hops. And with the Dynamic Age method, the average error distance shows 3–5 hops. The average error distance of the Multi-rumor-center method shows 3–4 hops. K-Center method is active around 2–3 hops, but the most performance is 3 hops.

Similarly, when the source is 3, In any network, the average error of our method is more 1–3 hops. Therefore, compared with other methods, our method estimates that the diffusion source is very close to the real source.

Table 3. Accuracy of multi-source identification

Experiment settings		Average error distance			
Network	$\|S^*\|$	PJC	Dynamic age	MRC	K-center
Power grid	2	1.600	3.510	3.135	1.750
	3	2.460	4.626	4.246	2.670
Yeast	2	2.521	3.175	2.710	2.680
	3	2.632	3.146	3.520	2.733
Facebook	2	3.237	3.950	3.433	3.215
	3	3.681	4.763	4.667	4.073

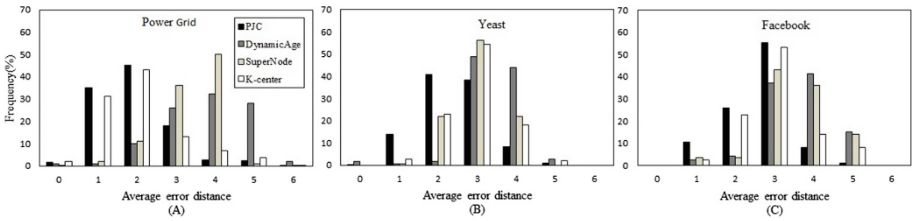

Fig. 3. When the number of sources is 2, the average error distance of different methods on the following three networks. (A) Power Grid; (B) Yeast; (C) Facebook

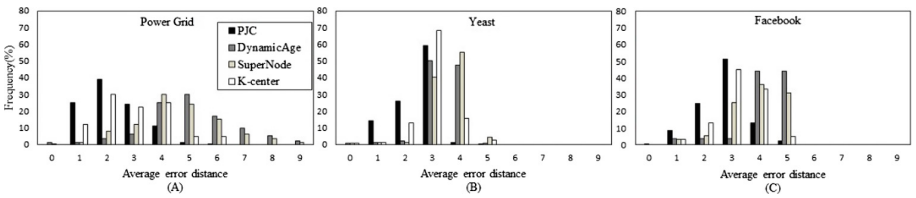

Fig. 4. When the number of sources is 3, the average error distance of different methods on the following three networks. (A) Power Grid; (B) Yeast; (C) Facebook.

5.2 Evaluation of the Number of Sources

In this section, we evaluate the proposed method for estimating the number of sources and predict the diffusion time. When the number of sources is determined, the propagation time of the source can be estimated. Table 4 shows the average and standard variance of estimated time. The results show that the mean of estimated time is close to the true propagation time, and the results of standard deviation are mostly less than 1. This shows that our method can estimate the real diffusion time of the source.

We simulated experiments on North American Power Grid, Facebook and Yeast networks to estimate the number of diffusion sources. As shown in Fig. 5. We let the number of diffusion sources range from 1 to 3. The horizontal axis represents the estimated number of sources and the vertical axis represents the percentage of the estimated number of sources in the experimental operation. On the Power Grid network, when the number of propagation sources is 1, about 78% of the experimental results can accurately estimate the number of sources. When the number of sources is 2, more than 80% of the experimental results can accurately estimate the number of sources. When the number of sources is 3, more than 60% of the experimental results can accurately estimate the number of sources. Similarly, it is obvious that at least half of the experiments can accurately identify the number of sources in the other two networks.

Table 4. Accuracy of spreading time estimation

Experiment settings		Estimated spreading time		
Network	$\|S^*\|$	T = 4	T = 5	T = 6
Power grid	2	4.012 ± 0.710	5.122 ± 1.790	5.987 ± 1.365
	3	4.032 ± 0.580	5.021 ± 0.860	6.025 ± 1.225
Yeast	2	4.521 ± 0.652	5.180 ± 0.420	5.851 ± 0.401
	3	4.340 ± 0.370	5.065 ± 1.210	5.921 ± 1.225
Facebook	2	4.242 ± 0.840	5.273 ± 0.521	5.820 ± 0.725
	3	4.432 ± 0.450	5.120 ± 0.721	5.790 ± 0.414

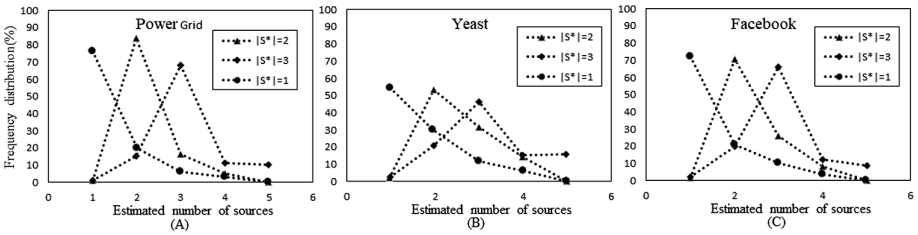

Fig. 5. Estimated number of sources in the following different networks. (A) Power Grid; (B) Yeast; (C) Facebook

5.3 Correlation Between the Real Sources and the Estimated Sources

By the Eq. (6), we detect the correlation between the objective function values on the Power Grid, Year and Facebook networks. As shown in Figs. 6 and 7 below. When the number of sources is 2, the distribution of points on the Power Grid shows an obvious linear relationship. This shows that the objective function values are highly correlated. On Yeast and Facebook, though there are fewer dots scattered, many dots float smaller around a line. This shows that the objective function values are also linearly correlated. Similarly, when the number of sources is 3, the objective function values are linearly correlated regardless of the network. It shows that we can use the proposed source detection method to estimate the location of the real source.

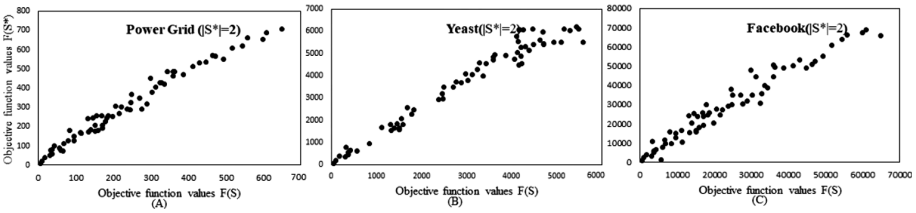

Fig. 6. When the number of sources is 2, the correlation between the objective function values is in the following network. (A) Power Grid; (B) Yeast; (C) Facebook

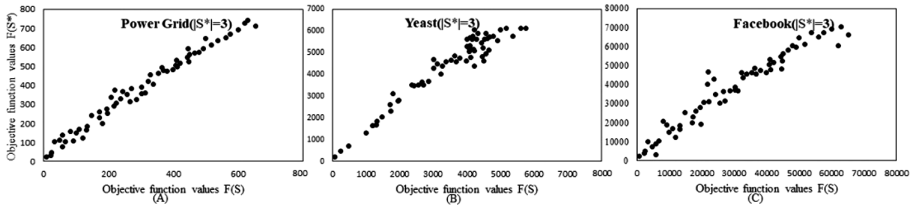

Fig. 7. When the number of sources is 3, the correlation between the objective function values is in the following network. (A) Power Grid; (B) Yeast; (C) Facebook

6 Related Work

In large-scale networks, the problem of outbreak threat propagation has become particularly serious. It becomes very meaningful to identify the source of propagation. However, most of the work focuses on the identification of single source in tree networks. Trees do not contain loops, but only a path between any pair of nodes. This greatly reduces the uncertainty of diffusion and the complexity of propagation, and further reduces the computational complexity. In real networks, threat propagation mostly involves multi-source problems, and the network is more complex. Diffusion processes of different sources are usually interactive and have uncertainties in the propagation process.

We mainly compare with the following multi-source identification methods. Fioriti et al. [4] proposes a dynamic aging method, which takes advantage of the correlation between eigenvalues and the "age" of nodes. The oldest nodes associated with the largest eigenvalues are considered diffusion sources. It essentially calculates the reduction of the maximum eigenvalue of the adjacent matrix after removing nodes. This method is based on the prior knowledge of the number of diffusion sources, and this method is not suitable for large-scale network source identification. Luo et al. [18] identifies multiple sources by expanding a single rumor center. For multiple sources, they propose a two-step method. They divide all infected nodes into different partitions by using Voronoi segmentation algorithm [19]. We need to calculate the number of different propagation paths from the sets. This method is difficult to use in large networks. The K-center [8] method is also a concept of introducing effective distances, using a Voronoi-like partitioning method to partition the network.

7 Conclusion

Most of the current technologies are based on tree networks, and few studies are focused on multi-source problems. In this paper, a novel PJC method is proposed to identify multi-source problems and estimate the number of sources. Considering that there are many overlapping nodes when the source nodes are very close, the same nodes are similarly infected. This makes the dissemination scope is smaller, and it is not suitable to study this propagation mode in large-scale networks. Therefore, this paper considers that there is a certain distance between the source nodes, which can avoid too many overlapping nodes. This can increase the probability of propagation. The experimental results show that our method is very effective.

Acknowledgement. This article is supported in part by the National Natural Science Foundation of China (61772150), the National Cryptography Development Foundation of China (MMJJ20170217) and Guangxi Key Research and Development Program AB17195025, and the open project of Guangxi Key Lab. of crypto, and Info. Security (Grant Nos.GCIS201622), and it is supported by GUET Excellent Graduate Thesis Program(16YJPYSS23).

References

1. Bakshy, E., Hofman, J.M., Mason, W.A., Watts, D.J.: Everyone's an influencer: quantifying influence on Twitter. In: Proceedings of ACM International Conference Web Search Data Mining, pp. 65–74, February 2011
2. Aral, S., Walker, D.: Identifying influential and susceptible members of social networks. Science **337**(6092), 337–341 (2012)
3. Tay, W.P.: The value of feedback in decentralized detection. IEEE Trans. Inf. Theory **58**(12), 7226–7239 (2012)
4. Luo, W., Tay, W.P., Leng, M.: Identifying infection sources and regions in large networks. IEEE Trans. Signal Process. **61**(11), 2850–2865 (2013)
5. Lokhov, A.Y., Mézard, M., Ohta, H., Zdeborová, L.: Inferring the origin of an epidemic with a dynamic message-passing algorithm. Phys. Rev. E **90**(1), 012801 (2014)
6. Luo, W., Tay, W.P.: Finding an infection source under the sis model. In: IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pp. 2930–2934 (2013)
7. Brockmann, D., Helbing, D.: The hidden geometry of complex, network-driven contagion phenomena. Science **342**(6164), 1337–1342 (2013)
8. Jiang, J.J., Wen, S., Yu, S., Xiang, Y., Zhou, W.: K-center: an approach on the multi-source identification of information diffusion. IEEE Trans. Inf. Forensics Secur. **10**(12), 2616–2626 (2015)
9. Dekker, A.H.: Centrality in social networks: theoretical and simulation approaches. In: Proceedings of SimTecT 2008, pp. 12–15 (2008)
10. Yang, K., Shekhar, A.H., Oliver, D., Shekhar, S.: Capacity-constrained network-Voronoi diagram: a summary of results. In: Nascimento, Mario A., et al. (eds.) SSTD 2013. LNCS, vol. 8098, pp. 56–73. Springer, Heidelberg (2013). https://doi.org/10.1007/978-3-642-40235-7_4
11. Wen, S., Zhou, W., Zhang, J., Xiang, Y., Zhou, W., Jia, W.: Modeling propagation dynamics of social network worms. IEEE Trans. Parallel Distrib. Syst. **24**, 1633–1643 (2013)

12. Wang, Y., Wen, S., Xiang, Y., Zhou, W.: Modeling the propagation of worms in networks: a survey. Commun. Surv. Tutorials IEEE **16**(2), 942–960 (2014)
13. Jeong, H., Mason, S.P., Barabási, A.L., Oltvai, Z.N.: Lethality and centrality in protein networks. Nature **411**(6833), 41–42 (2001)
14. Viswanath, B., Mislove, A., Cha, M., Gummadi, K.P.: On the evolution of user interaction in facebook. In: Proceedings of the 2nd ACM Workshop on Online Social Networks, pp. 37–42 (2009)
15. Leskovec, J., Kleinberg, J., Faloutsos, C.: Graph evolution: densification and shrinking diameters. ACM Trans. Knowl. Discovery Data **1**(1) (2007)
16. Wen, S., Zhou, W., Zhang, J., Xiang, Y., Zhou, W., Jia, W.: Modeling propagation dynamics of social network worms. IEEE Trans. Parallel Distrib. Syst. **24**(8), 1633–1643 (2013)
17. Fioriti, V., Chinnici, M., Palomo, J.: Predicting the sources of an outbreak with a spectral technique. Appl. Math. Sci. **8**(135), 6775–6782 (2014)
18. Wolpert, D.H.: The lack of a priori distributions between learning algorithms. Neural Comput. **8**(7), 1341–1390 (1996)
19. Hakimi, S.L., Labbé, M.L., Schmeichel, E.: The Voronoi partition of a network and its implications in location theory. ORSA J. Comput. **4**(4), 412–417 (1992)