IEEE Access
Multidisciplinary : Rapid Review : Open Access Journal

Received March 2, 2018, accepted April 1, 2018, date of publication April 4, 2018, date of current version April 25, 2018.

Digital Object Identifier 10.1109/ACCESS.2018.2822844

CDA: A Clustering Degree Based Influential Spreader Identification Algorithm in Weighted Complex Network

QIAN WANG[1,2], JIADONG REN[1,2], YU WANG[1,2], BING ZHANG[1,2], YONGQIANG CHENG[3], AND XIAOLIN ZHAO[4]

[1]College of Information Science and Engineering, Yanshan University, Qinhuangdao 066000, China
[2]Computer Virtual Technology and System Integration Laboratory of Hebei Province, Qinhuangdao 066000, China
[3]Computer Science, University of Hull, Hull HU6 7RX, U.K.
[4]Beijing Key Laboratory of Software Security Engineering Technique, Beijing Institute of Technology, Beijing 100081, China

Corresponding author: Qian Wang (wangqianysu@163.com)

This work was supported in part by the National Key R&D Program of China under Grant 2016YFB0800700, in part by the National Natural Science Foundation of China under Grant 61472341, Grant 61772449, and Grant 61572420, in part by the Natural Science Foundation of Hebei Province China under Grant F2016203330, in part by the Advanced Program of Postdoctoral Scientific Research under Grant B2017003005, and in part by the Doctoral Foundation of Yanshan University under Grant B1036.

ABSTRACT Identifying the most influential spreaders in a weighted complex network is vital for optimizing utilization of the network structure and promoting the information propagation. Most existing algorithms focus on node centrality, which consider more connectivity than clustering. In this paper, a novel algorithm based on clustering degree algorithm (CDA) is proposed to identify the most influential spreaders in a weighted network. First, the weighted degree of a node is defined according to the node degree and strength. Then, based on the node weighted degree, the clustering degree of a node is calculated in respect to the network topological structure. Finally, the propagation capability of a node is achieved by accounting the clustering degree of the node and the contribution from its neighbors. In order to evaluate the performance of the proposed CDA algorithm, the susceptible-infected-recovered model is adopted to simulate the propagation process in real-world networks. The experiment results have showed that CDA is the most effective algorithm in terms of Kendall's tau coefficient and with the highest accuracy in influential spreader identification compared with other algorithms such as weighted degree centrality, weighted closeness centrality, evidential centrality, and evidential semilocal centrality.

INDEX TERMS Clustering degree, influential spreaders, weighted complex network.

I. INTRODUCTION

Complex networks are pervasive in our daily life [1], such as social network, traffic network, power grid, and software network etc.. While these networks provide convenience for our lives, certain risk still exists. If a failure occurs in the important nodes, it can cause catastrophic failure of the entire network. For example, in early 2008, there was a massive snowfall lasted for many days in southern China and it caused large-scale blackout because the critical towers and major transmission lines were severely damaged [2]. Therefore, it is of theoretical and practical significance for identifying the important nodes in complex networks. In the customer network, the most influential customers can promote new products to achieve the maximum benefits [3]. In the microblog network, the most influential microblog leaders can attract lots of attention and spread information in a short period of time to a greater range [4]. In the software network, the key functions need to be protected, and this can reduce the overhead of software fault location and daily maintenance [5]. Therefore, how to accurately and effectively identify the influential spreaders in complex network becomes a hot topic in many fields.

In the real-world networks, the objects can be represented as nodes and the connections between the objects are edges. Different real-world networks have different importance of nodes and edges. Therefore, the research of weighted complex networks should be paid more attention. In the current literature, researchers have proposed many algorithms for identifying key nodes. In [6], classical centrality algorithms including degree centrality (DC) [7], betweenness centrality (BC) [8] and closeness centrality (CC) [9] are extended to be applied in weighted complex networks, but they are circumscribed in weighted networks. A variety of algorithms are proposed to identify influential spreaders in weighted networks. Wei *et al.* [10] proposes the evidential centrality algorithm (EVC) based on the Dempster–Shafer evidence theory [11], [12]. The propagation capability of a node is obtained by the node degree and strength in a weighted network. This algorithm is simple and similar to DC, i.e. the global structure of the network is ignored. Gao *et al.* [13] propose the evidential semi-local centrality algorithm (ESC) to consider not only the degree and strength of every node, but also the EVC value of its neighbors within 2 steps range. However, the topological connection of the node neighbors is ignored. If a node is closely connected with its neighbors, there must be more affection between each other hence the more influential the node is. Ren *et al.* [14] propose the evidential local structure centrality algorithm (ELSC) that considers the EVC value of the neighbors within 2 steps range as well as the topological connections between neighbors to identify influential spreaders based on the Dempster–Shafer evidence theory. There are also some other node rank algorithms that perform well in unweighted network and further applied to the weighted network with promising performance. Such as eigenvector centrality [15], k-shell decomposition [16], PageRank [17] and LeaderRank [18], and so on. However, all above algorithms only consider the sum of the edge weights or ignore the difference of neighbor contribution.

The number and the weight of the edge should be considered when evaluating the propagation capability of a node in weighted complex network, since they both play important roles in different aspects [6]. For example, a salesman is expected to have a large number of the connected edges in the social network in order to achieve maximize product impact, a high possibility is that a high edge weight may mean a potential important customer. So, it is important to consider the contribution degree of the node's neighbors for identifying influential nodes in complex networks.

In this paper, a novel algorithm called clustering degree algorithm (CDA) is proposed to identify the influential spreaders by considering comprehensively the degree and strength of node as well as the network topology and the differentiated contribution degree of its neighbors. In order to compare the performance between CDA and other algorithms, the Kendall's tau coefficient [27] is employed for the node rank correlation; and the SIR model [19] is applied for the propagation ability of the top-k nodes in each algorithms.

The rest of this paper is organized as follows. Section 2 presents the relevant definitions of the algorithm. A clustering degree algorithm is proposed in Section 3 and a case study is presented in this section as well. Thereafter, in Section 4, the performance of CDA is verified by experiments with real data sets. Finally, the conclusions are summarized in Section 5.

II. DEFINITIONS

A weighted and undirected network $G = (V, E, W)$ is considered, where V represents the node set, E represents the edge set of the network and W represents the weight set of edge. |V| indicates the number of node, and |E| indicates the number of edge. $E(i, j)$ represents the connected edge between node i and j, and $W(i, j)$ represents the weight of edge $E(i, j)$.

Definition 1: Node Strength. The strength of node i is the sum of weights of all edges connected to node$i$.

Node Strength (*NS*) is defined as follows.

$$NS(i) = \sum_{j\in\Gamma_i} W(i, j) \tag{1}$$

where Γ_i represents the set of direct neighbors of node i.

Definition 2: The Weighted Degree of node. The weighted degree of node i is the combination of the degree and the strength of node i.

The Weighted Degree (*WD*) of a node is defined as follows.

$$WD(i) = \alpha K(i) + (1 - \alpha)NS(i) \tag{2}$$

where $K(i)$ is the degree of node i and α is a tuning parameter. Without losing generality, the tuning parameter α is set to 0.5 [20].

Definition 3: The Clustering Degree of node. The clustering degree of node i is determined by the weighted clustering coefficient and the weighted degree of node i.

The Weighted Clustering coefficient (*CW*) of node is defined as follows [21].

$$CW(i) = \frac{1}{NS(i)[K(i) - 1]} \sum_{j,k} \frac{W(i, j) + W(i, k)}{2} a_{ij}a_{jk}a_{ki} \tag{3}$$

$a_{ij} = 1$, if node i is connected with node j. Otherwise, $a_{ij} = 0$. Node j and node k are arbitrary direct neighbors of node i. The number of closed triplets in the neighborhood of a node and the total relative weight with respect to the strength of the node are considered. The normalization factor $NS(i)[K(i) - 1]$ accounts for the maximum possible number of triplets the edge weight may participate, it ensures that $0 \leq CW(i) \leq 1$.

In a complex network, node propagation capability is attributed directly to node position. In general, the node that locates closer to the core of the network will have larger weighted clustering coefficient and greater propagation capability.

The Clustering Degree (*CD*) of a node is defined as follows.

$$CD(i) = WD(i) \times f(CW(i)) \tag{4}$$

The function $f(.)$ is a Sigmod function $f(x) = 1/(1 + e^{-x})$. It ensures a monotonic increasing between 0 and 1. Because the minimum value of the *CW* of the node is 0 and its maximum value is 1. Thus the value of $f(x)$ function ranges between 0.5 and 0.7311. Then, CW can be well used to prove the topology affection for the node propagation.

Therefore, the Clustering Degree (CD) of a node can be rewritten as follows.

$$CD(i) = WD(i) \times \frac{1}{1 + e^{-CW(i)}} \tag{5}$$

Definition 4: The Propagation Capability of node. The propagation capability of node i is determined by the clustering degree of node i and the neighbors of node i.

The Propagation Capability (PC) of a node is defined as follows.

$$PC(i) = CD(i) + \sum_{j \in \Gamma_i} \frac{W(i,j)}{W_{\max}} CD(j) \tag{6}$$

where W_{max} is the maximum edge weight, $W(i,j)/W_{max}$ is a standardized factor that ranges from 0 to 1.

III. CDA ALGORITHM

A. THE CLUSTERING DEGREE ALGORITHM

When the propagation capability of a spreader is measured in weighted network, the weighted degree of node is defined according to the node degree and the strength. It is a simple local index and similar to degree centrality algorithm. This paper combines the information of the network topology and the weighted clustering coefficient that used to measure the core and the peripheral location of a node in the network. Further consideration is given to the propagation capability of the spreader greatly influenced by its direct neighbors where each neighbor's contribution to the corresponding spreader is measured by the edge weight.

In CDA algorithm, firstly, the weighted degree (WD) of a node is calculated according to the degree and the strength of the node. Then, based on the weighted degree, the location of the node is measured by the weighted clustering coefficient. A larger clustering coefficient means that the node has a closer distance to the core of the network and a stronger propagation capability. The clustering degree (CD) of a node is obtained based on the weighted clustering coefficient and the weighted degree. Finally, the propagation capability (PC) of a node is achieved by the clustering degree and the different neighbor contributions (edge weight). The pseudo-code of the algorithm is shown as Algorithm 1.

In line 1, α is initialized. This is the tuning parameter of the node degree and strength. Lines 2-4 are the process to compute the weighted degree (WD) of a node according to the node degree and strength. Lines 5-7 are the process to calculate the clustering degree (CD) of a node according to the weighted clustering coefficient and the weighted degree of a node. The propagation capability (PC) of a node is calculated in lines 8-12. The node is sorted in line 13, the higher the value of PC is, the more influential the node is.

B. AN SAMPLE CASE STUDY

To better explain the algorithm CDA, a simple network with 10 nodes and 13 edges is constructed to demonstrate how

Algorithm 1 CDA

Input: adjacent matrix $graph = (a_{ij})_{N \times N}$ and adjacent table adjList corresponding to weighted complex network
Output: The ranked list and the PC of every node
Process:
01) **Initialize** α
02) node strength $NS(i) \leftarrow$ the weights of all the connected edges of node i.
03) $K(i) \leftarrow$ the degree of node i
04) weight degree $WD(i) \leftarrow \alpha \times K(i) + (1 - \alpha) \times NS(i)$
05) $CW(i) \leftarrow$ the local clustering coefficient of node i
06) $f(i) \leftarrow$ the CW(i) is mapped to the sigmod function
07) clustering degree $CD(i) \leftarrow WD(i) \times f(CW(i))$
08) $W_{max} \leftarrow$ maximum all edge weight W
09) **for**each direct neighbor node j of i
10) contribution of j to $iI_{j \to i} += \frac{W(i,j)}{W_{\max}} CD(j)$
11) **end for**
12) $PC(i) \leftarrow CD(i) + I_{j \to i}$
13) sort(PC,'descend')

CDA works. The network is undirected and weighted as shown in Figure 1. The calculation process of CDA is shown in Figure 2. Take node 5 as an example to illustrate the implementation process of CDA.

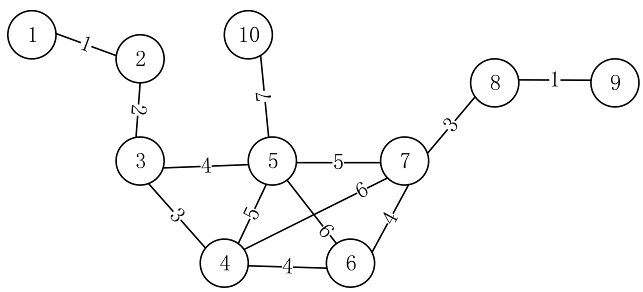


FIGURE 1. An example network schematic diagram.

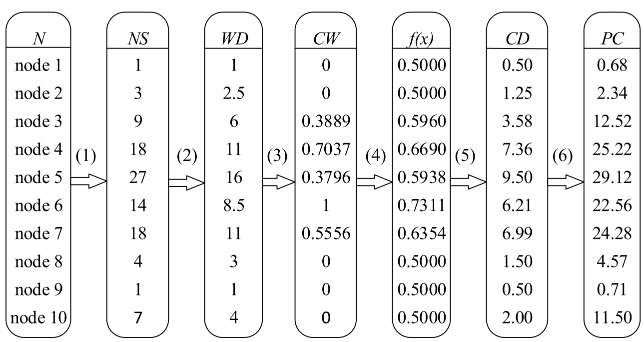


FIGURE 2. The calculation process of CDA in example network.

Step 1: The strength of node 5 is calculated by (1).

$$NS(5) = \sum_{j \in \Gamma_i} W(i,j) = 27$$

Step 2: The weighted degree of node 5 is calculated by (2).

$$WD(5) = \alpha \times K(i) + (1 - \alpha) \times NS(5) = 16$$

Step 3: The weighted clustering coefficient of node 5 is calculated by (3).

$$CW(5) = \frac{1}{NS(5)[K(5)-1]} \times \sum_{j,k} \frac{W(i,j)+W(i,k)}{2} a_{ij}a_{jk}a_{ki} = 0.3796$$

Step 4: The clustering degree of node 5 is calculated by (5).

$$CD(5) = WD(5) \times f(5) = 9.50$$

Step 5: The propagation capability of node 5 is calculated by (6).

$$PC(5) = CD(i) + \sum_{j \in \Gamma_i} \frac{W(i,j)}{W_{\max}} CD(j) = 29.12$$

The propagation capability of all nodes in the example is shown in Figure 2. If only the node degree and strength are considered, the propagation capability of node 4 and node 7 can not be distinguished as $WD(4) = WD(7) = 11$. Moreover, as seen in Figure 1, node 4 is closer to the core location of the network than node 7, so node 4 should have stronger propagation capability than node 7. By further considering the topology of the node in the network node 4 and node 7 can be distinguished successfully as the clustering degree of node 4 is greater than the clustering degree of node 7, which is consistent with the actual situation. However, if only the clustering degree of node is considered, the propagation capability of node 1 and node 9 can not be distinguished as $CD(1) = CD(9) = 0.5$. By fully incorporating contribution of neighbors to a spreader, the proposed CDA algorithm again successfully distinguished node 1 and 9 by different propagation capability.

TABLE 1. Rank of node propagation capability with different algorithms.

Rank	WDC	WCC	EVC	ESC	CDA	SIR
Rank1	5	5	5	5	5	5
Rank2	4,7	4	4,7	4	4	4
Rank3	6	3	6	7	7	7
Rank4	3	7	3	6	6	6
Rank5	10	6	10	3	3	10
Rank6	8	2	8	10	10	3
Rank7	2	1,8	2	8	8	8
Rank8	1,9	9	1,9	2	2	2
Rank9		10		9	9	9
Rank10				1	1	1

For the above example, Table 1 shows the sorted node propagation capability calculated by different algorithms, including WDC, WCC, EVC, ESC, CDA and the SIR model. The rank of WDC and EVC is consistent and they can not distinguish the propagation capability of node 4 and 7 as well as the propagation capability of node 1 and 9. WCC can not distinguish the propagation capability of node 1 and 8. The propagation capability of each node can be distinguished by ESC and CDA. Among them, node propagation capability rank is consistent based on ESC and CDA. In addition, comparing with the real propagation capability of node based on SIR model, only ESC and CDA can accurately identify the node rank in the network (except node 3 and 10). The superiority of CDA over ESC will be further demonstrated in section 4.4. So this example shows that CDA is feasible and objective for identifying the influential nodes in weighted networks.

IV. EXPERIMENTAL ANALYSIS

A. EXPERIMENTAL DATA

The real datasets used to evaluate our proposed algorithm are taken from http://www-personal.umich.edu/~mejn/netdata/, including Zachary, Netscience and Hep. The Zachary [22] network is a weighted network consisting of 34 nodes, each node represents a member in the club, and each edge indicates a friend relationship outside the clue between that member and other members. The edge weight indicates the degree of closeness between members. The Netscience [23] network is a weighted network consisting of 1589 nodes, each node represents a scientist, the edges represent collaboration on papers between scientists, and the edge weight represents their collaboration times. Here, the largest component with 379 scientists is considered. The Hep [24] network is a weighted network containing 8361 nodes, each node represents a scientist, the edges represent collaboration posting preprints between scientists on the High-Energy Theory E-Print Archive between Jan 1, 1995 and December 31, 1999, and the edge weight represents their collaboration times.

The basic topological properties are shown in Table 2.

TABLE 2. The basic topological properties of the three real networks.

Network	|V|	|E|	Kavg	K_{max}	W	W_{max}	C
Zachary	34	78	4.58	17	2.9615	7	0.5817
Netscience	379	914	4.82	34	0.5356	4.75	0.7610
Hep	8361	15751	3.88	50	0.9731	34.01	0.605

In Table 2, the |V| and |E| represent the total number of nodes and edges; K_{avg} and K_{max} represent the average and the maximum node degree; W and W_{max} represent the average and the maximum edge weight; and C indicates the clustering coefficient of the network.

B. THE SIR MODEL

The SIR model is widely used for examining the propagation ability of nodes [25]. In this model, each node has three possible states: S (Susceptible), I (Infected) and R (Recovered). At the initial stage, only one node is infected and the other nodes are susceptible. During the propagation process of each step, each infected node randomly selects its direct susceptible neighbors with probability P, and then enters the recovered state with probability equal to 1. The propagation process will stop when there isn't any new node to be infected. In weighted networks [26], the susceptible neighbor node j is

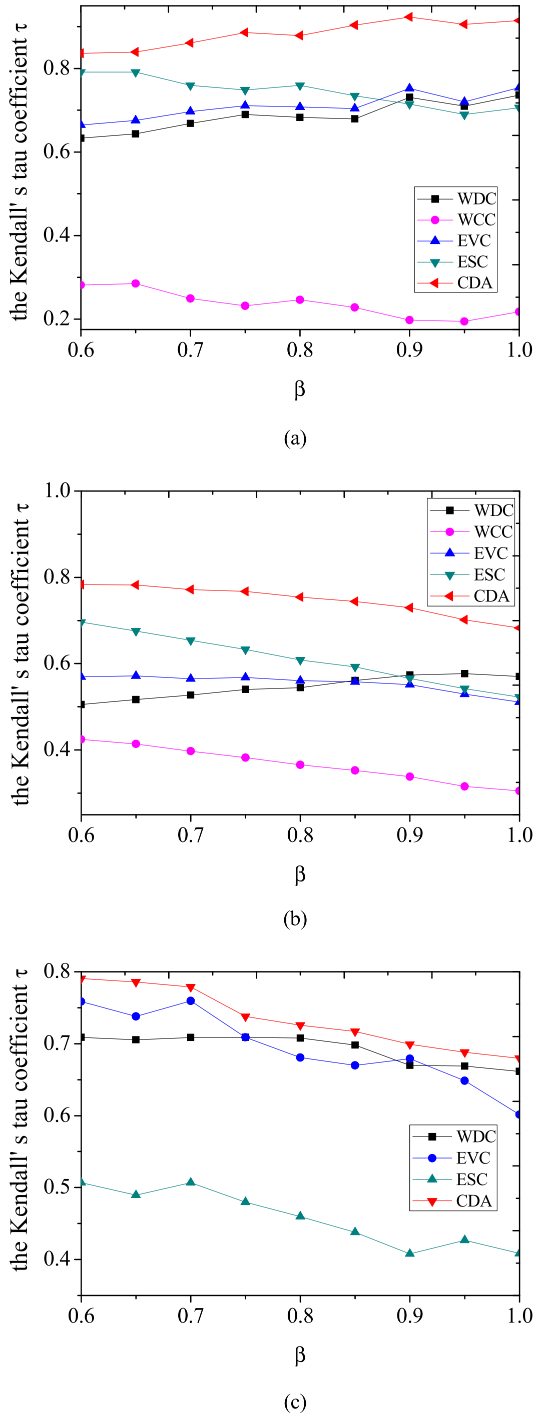


FIGURE 3. The Kendall's tau τ values corresponding to different algorithms in different networks. (a) In the Zachary network. (b) In the Netscience network. (c) In the Hep network.

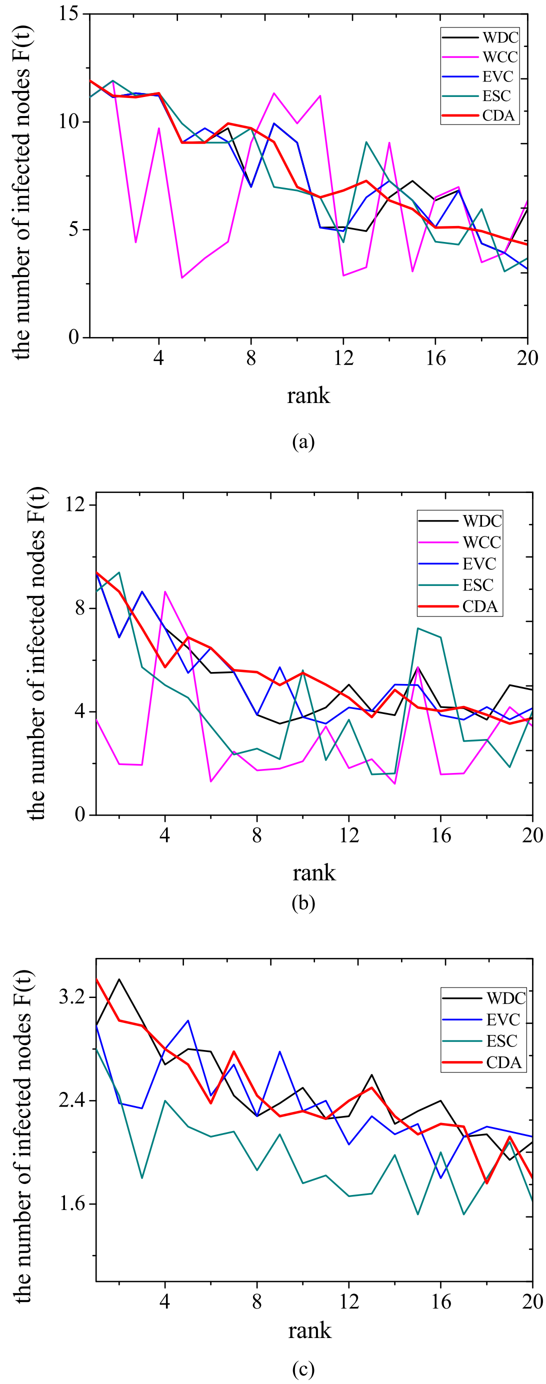


FIGURE 4. The average number of infected nodes by top-k nodes of different algorithms. (a) In the Zachary network. (b) In the Netscience network. (c) In the Hep network.

infected by the infected node i with probability P.

$$P = \left(\frac{W(i,j)}{W_{\max} + 1} \right)^{\beta}, \quad \beta > 0 \tag{7}$$

where $W(i, j)$ represents the weight of edge $E(i, j)$; W_{max} indicates maximum edge weight and β denotes the regulatory factor of propagation speed. As $W(i,j)/(W_{\max}+1) < 1$, the propagation speed and value of β are inversely proportional. Finally, the actual propagation capacity of the initial infected nodes is quantified according to the number of infected nodes. At each step t of the propagation process, the number of infected nodes is represented by F(t). The reliability of the results is ensured by averaging over 500 independent experiments.

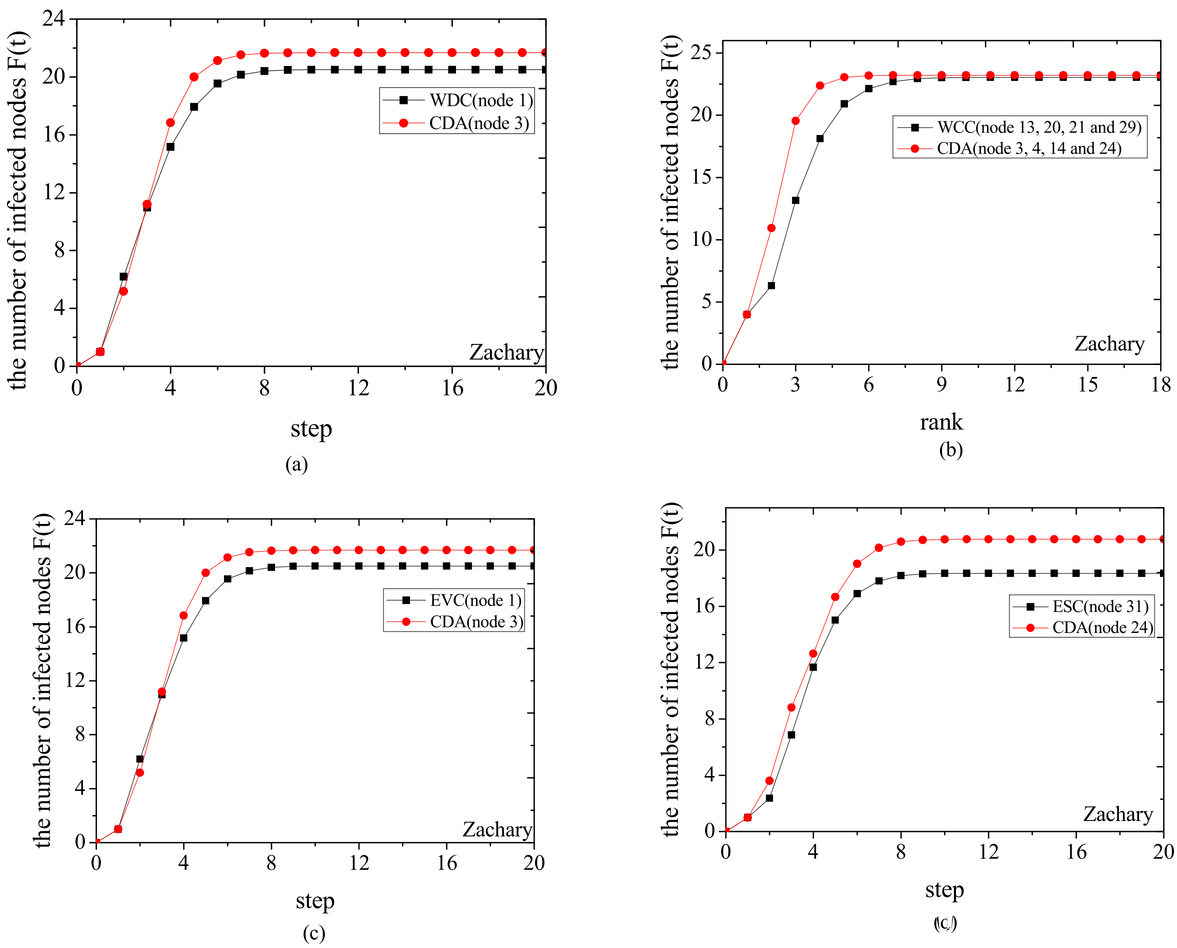


FIGURE 5. The number of infected nodes by initially top-10 nodes in Zachary network. (a) Comparison between WDC and CDA. (b) Comparison between WCC and CDA. (c) Comparison between EVC and CDA. (d) Comparison between ESC and CDA.

C. THE KENDALL'S TAU COEFFICIENT

The Kendall's tau coefficient τ is widely used in the correlation analysis. The bigger the Kendall's tau coefficient, the better the correlation between the observations. In this paper, it is used to analyze the rank correlation of the nodes achieved by a certain algorithm and that of the SIR model. It considers a set of joint observations consisted of random variables X ={x1, x2, ..., xn} and Y = {y1,y2, ..., ym}, n,m $\in$R. Here, X represents the node rank based on a certain algorithm and Y represents the node rank based on the SIR model. If $(x_i - x_j)(y_i - y_j) > 0$, the observations have concordant rank in X and Y. If $(x_i - x_j)(y_i - y_j) < 0$, they are considered to be inconsistent rank in X and Y. If $(x_i - x_j)(y_i - y_j) = 0$, they have the same rank in X and Y. Then the Kendall's tau coefficient τ is defined as follows [27].

$$\tau = \frac{n_c - n_d}{0.5n(n-1)} \tag{8}$$

where nc and nd indicate the number of concordant pairs and inconsistent pairs, respectively.

D. EFFECTIVE VERIFICATION AND COMPARISON BASED ON THE SIR MODEL

In this section, different real networks are applied for the effectiveness comparison between CDA and other algorithms, i.e. WDC, WCC, EVC and ESC, in terms of Kendall's tau coefficient, rank of top-k nodes and top-10 node propagation score. However, WCC needs to calculate the distance between any two nodes in the network, so it is not suitable for applications in disconnected networks. Disconnected network means a network with isolated nodes. The Hep network is a disconnected network, therefore, WCC is incapable in the Hep network, but it can be used in other networks.

1) KENDALL'S TAU COEFFICIENT

The Kendall's tau coefficient τ is used to analyze the rank correlation of the nodes between a certain algorithm and the SIR model in three real networks. The corresponding values of τ for different β in different networks are shown in figure 3, β ranges from 0.6 to 1.

In Zachary network, the curve of CDA is above the curves of other four algorithms, it indicates that CDA has the highest

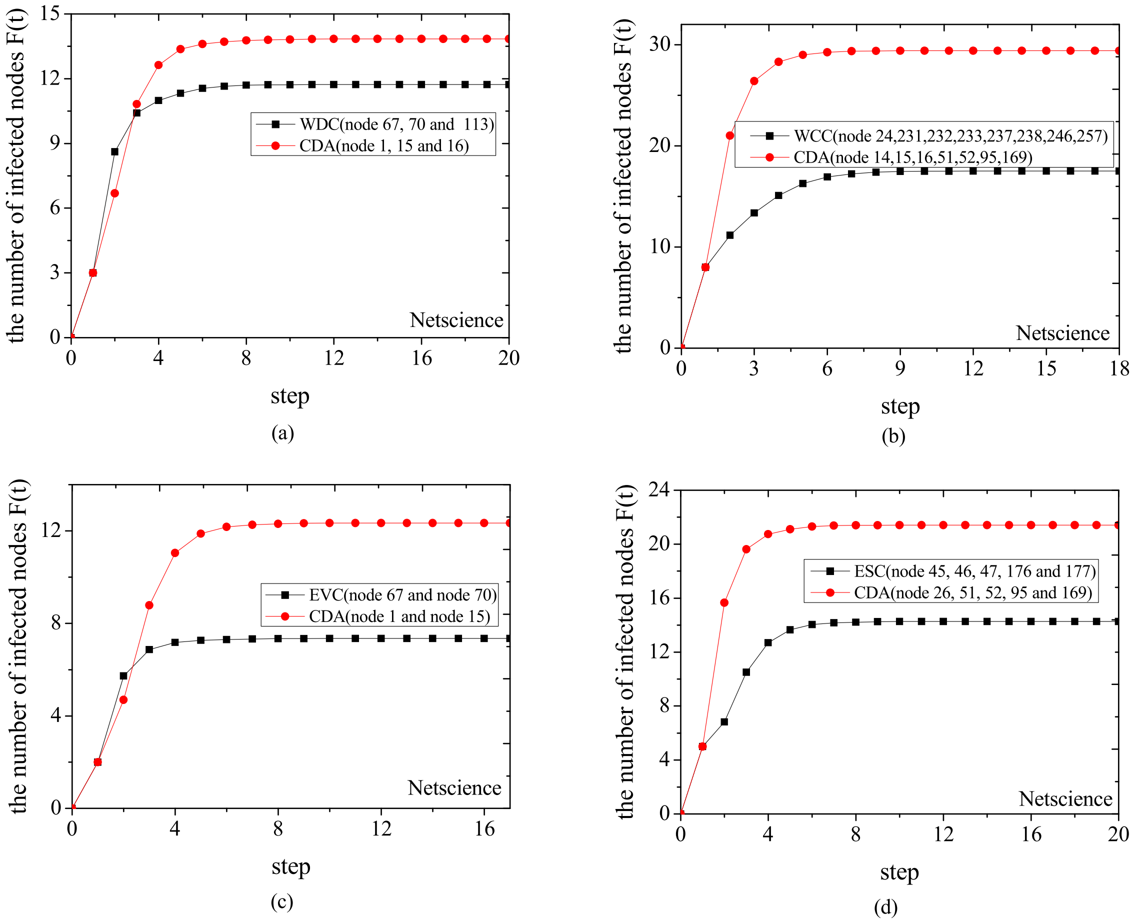


FIGURE 6. **The number of infected nodes by initially top-10 nodes in Netscience network. (a) Comparison between WDC and CDA. (b) Comparison between WCC and CDA. (c) Comparison between EVC and CDA. (d) Comparison between ELSC and CDA.**

accuracy of node rank comparing to the other four algorithms. In Netscience network, CDA performs the best. In Hep network, CDA performs the best; ESC is relatively poor performed. The experiment shows that the rank correlation of nodes between CDA algorithm and the SIR model of different β is the best, which is the node rank of CDA is the most accurate among the five algorithms. However, the Kendall's tau coefficient τ can only evaluate the rank correlation of nodes between an algorithm and the SIR model, and cannot evaluate the propagation ability of nodes.

2) RANK OF TOP-K NODES

In order to further verify the effectiveness of CDA and other four algorithms, the rank of top-k nodes is obtained by WDC, WCC, EVC, ESC and CDA. Then the SIR model is applied to evaluate the number of infected nodes by the top-k nodes, F(t) represents the number of infected nodes by SIR model as mentioned above and t is the step value which is set to 3. Here, $\beta = 1$. The average number of infected nodes by each top-k node of different algorithms are shown in Figure 4. The curve of CDA is downward sloping the most gently among the five algorithms. Namely, the propagation ability of the top-k nodes based on CDA decreases most steadily with the increasing of k. It also illustrates that the top-k node rank of CDA is consistent with the ability of node infection. So the top-k nodes of CDA is the most accurate among the five algorithms.

3) TOP-10 NODE PROPAGATION SCORE

The propagation ability of the different top-10 nodes between CDA and other algorithms is compared by the SIR model. The rank of top-10 nodes by WDC, WCC, EVC, ESC and CDA in different networks are listed in Tables 3, Table 4 and Table 5. WCC is incapable in the HEP network, so WCC does not appear in Table 5.

The SIR model is applied to evaluate node propagation capability, the top-10 nodes are used as initial nodes to infect other nodes in the network, the infection situation of each step in the propagation process is analyzed, and the number of infected nodes in the network are compared when the propagation stops. The simulation results are shown in Figures 5-7.

The top-10 nodes selection by different algorithms may be the same or different, and the propagation ability of the same nodes does not need to be compared. For example, in Zachary

TABLE 3. The top-10 nodes ranked by five algorithms in Zachary network.

Rank	WDC	WCC	EVC	ESC	CDA
Rank 1	34	1	34	1	34
Rank 2	1	34	1	34	3
Rank 3	33	20	33	3	1
Rank 4	3	32	3	33	33
Rank 5	2	13	2	9	2
Rank 6	24	21	32	14	14
Rank 7	32	29	24	2	9
Rank 8	4	2	4	32	32
Rank 9	9	33	9	4	24
Rank10	14	9	14	31	4

TABLE 4. The top-10 nodes ranked by five algorithms in Netscience network.

Rank	WDC	WCC	EVC	ESC	CDA
Rank 1	4	231	4	5	4
Rank 2	26	232	26	4	5
Rank 3	5	233	5	16	51
Rank 4	51	5	51	15	16
Rank 5	52	26	95	45	26
Rank 6	95	257	52	46	52
Rank 7	169	24	169	47	1
Rank 8	67	237	67	176	169
Rank 9	113	238	16	177	15
Rank10	70	246	70	1	95

TABLE 5. The top-10 nodes ranked by five algorithms in Hep network.

Rank	WDC	EVC	ESC	CDA
Rank 1	24	24	480	546
Rank 2	546	87	168	547
Rank 3	547	546	481	24
Rank 4	530	480	473	480
Rank 5	480	547	956	530
Rank 6	763	168	884	87
Rank 7	168	530	39	763
Rank 8	997	997	656	168
Rank 9	87	763	123	997
Rank10	1754	926	1571	926

network, the top-10 different nodes based on ESC and CDA are node 24 and node 31, respectively. Therefore, only the propagation capability of node 24 and node 31 need to be compared (As shown in Figure 5 (c)). If the top-10 nodes of the two algorithms are identical, and the total propagation ability of the top-10 nodes is the same. At this point, a better algorithm is determined by verifying the consistency of node rank and propagation ability. For example, in Zachary network, the top-10 node of WDC is exactly the same as the top-10 node of CDA. The top-1 nodes of the two algorithms are node 34, while the top-2 nodes are node 1 and node 3, respectively. Therefore, the propagation capability of node 1 and node 3 need to be compared. The experimental results show that the propagation capacity of node 3 is stronger than that of node 1 (As shown in Figure 5 (a)). Therefore, the top-10 node rank of CDA is more accurate than WDC.

From Figures 5-7, it is clear that the curve of CDA is above the curves of other algorithms at the node propagation process in different networks. In other words, CDA is the best among the five algorithms for the number of infected nodes at each step during the propagation process. What's more, when the propagation of the infected nodes stops, the final number of infected nodes based on CDA is the largest among the five algorithms. It further illustrates that CDA can identify the propagation ability of top-10 nodes better than WDC,

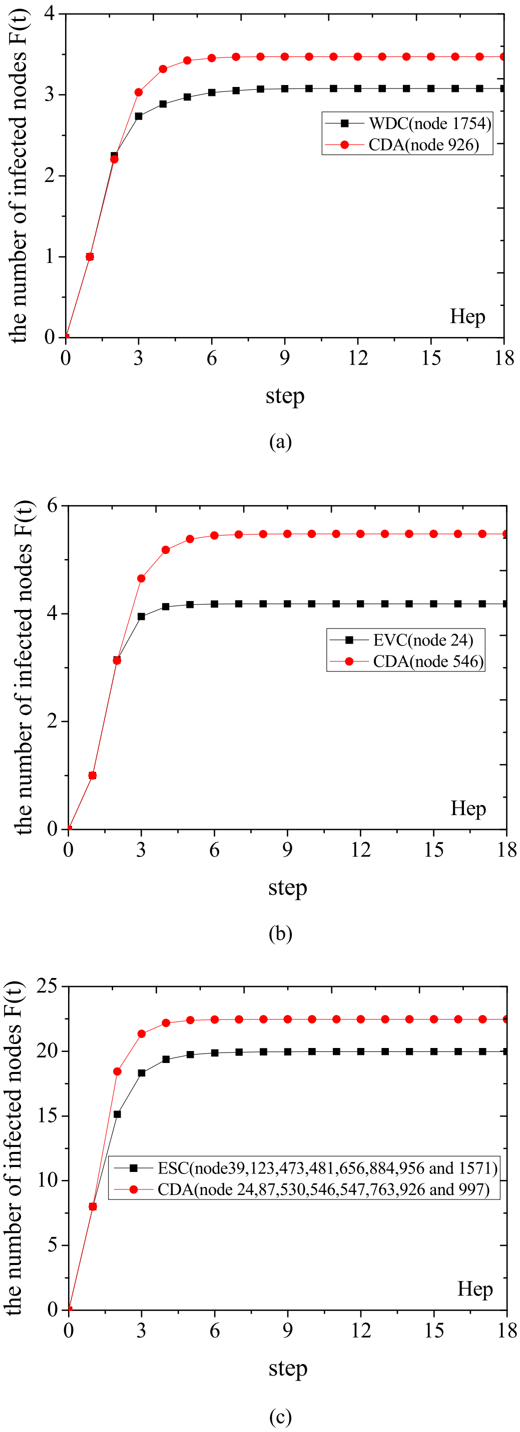


FIGURE 7. The number of infected nodes by initially top-10 nodes in Netscience network. (a) Comparison between WDC and CDA. (b) Comparison between EVC and CDA. (c) Comparison between ESC and CDA.

WCC, EVC and ESC, and provide the most reasonable node rank list. CDA is a weighted degree centrality based algorithm, and further considers the topology of nodes in the network. So, it has good perform especially when there is obvious community structure in the network.

V. CONCLUSION

In this paper, CDA is proposed to identify the influential spreaders in weighted complex networks. By taking into account the node degree, strength, topological position and the neighbor node contribution, results show that CDA algorithm is more effective and produces better list of the most influential nodes. The performance of the CDA has been evaluated by applying the SIR model to simulate the spreading process in real-world networks. Comparing with WDC, WCC, EVC and ESC, the CDA algorithm has achieved the closest node rank to the real propagation ability. In addition, the propagation ability of the top-k nodes based on CDA decreases the most steadily with the increasing of k, it also illustrates that the node rank of CDA is consistent with the ability of node infection, and the node rank is the most accurate among the five algorithms. Meanwhile, the propagation ability of CDA and other algorithms with different top-10 nodes is compared. CDA can identify the of top-10 spreading ability nodes better than WDC, WCC, EVC and ESC.

ACKNOWLEDGMENT

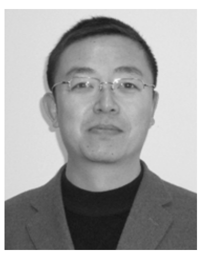
The authors are grateful to valuable comments and suggestions of the reviewers.

REFERENCES

[1] M. Hosseini-Pozveh, K. Zamanifar, and A. R. Naghsh-Nilchi, "A community-based approach to identify the most influential nodes in social networks," *J. Inf. Sci.*, vol. 43, no. 2, pp. 204–220, Feb. 2016, doi: 10.1177/0165551515621005.
[2] J. Liu, Q. Xiong, W. Shi, X. Shi, and K. Wang, "Evaluating the importance of nodes in complex networks," *Phys. A, Statist. Mech. Appl.*, vol. 452, pp. 209–219, Feb. 2016, doi: 10.1016/j.physa.2016.02.049.
[3] Y. Liu, Y. Deng, and B. Wei, "Local immunization strategy based on the scores of nodes," *Chaos*, vol. 26, no. 1, p. 013106, Jan. 2016, doi: 10.1063/1.4940240.
[4] J. Lu and G. W. Wan, "Identification of key nodes in microblog networks," *ETRI J.*, vol. 52, no. 1, pp. 52–61, Oct. 2015, doi: 10.4218/etrij.16.0115.0732.
[5] G. Huang *et al.*, "An algorithm to find critical execution paths of software based on complex network," *Int. J. Modern Phys. C*, vol. 26, no. 9, p. 1550101, Sep. 2015, doi: 10.1142/S0129183115501016.
[6] T. Opsahl, F. Agneessens, and J. Skvoretzc, "Node centrality in weighted networks: Generalizing degree and shortest paths," *Social Netw.*, vol. 32, no. 3, pp. 245–251, 2010, doi: 10.1016/j.socnet.2010.03.006.
[7] L. C. Freeman, "Centrality in social networks conceptual clarification," *Soc. Netw.*, vol. 1, no. 3, pp. 215–239, 1979.
[8] L. C. Freeman, "A set of measures of centrality based on betweenness," *Sociometry*, vol. 40, no. 1, pp. 35–41, Mar. 1977.
[9] G. Sabidussi, "The centrality index of a graph," *Psychometrika*, vol. 31, no. 4, pp. 581–603, Dec. 1966.
[10] D. Wei, X. Deng, X. Zhang, Y. Deng, and S. Mahadevan, "Identifying influential nodes in weighted networks based on evidence theory," *Phys. A, Statist. Mech. Appl.*, vol. 392, no. 10, pp. 2564–2575, 2013, doi: 10.1016/j.physa.2013.01.054.
[11] A. P. Dempster, "Upper and lower probabilities induced by a multivalued mapping," *Ann. Math. Statist.*, vol. 38, no. 2, pp. 325–339, 1967, doi: 10.1214/aoms/1177698950.
[12] J. Inglis, "A mathematical theory of evidence," *Technometrics*, vol. 20, no. 1, p. 106, Jan. 1976.
[13] C. Gao, D. Wei, Y. Hu, S. Mahadevan, and Y. Deng, "A modified evidential methodology of identifying influential nodes in weighted networks," *Phys. A, Stat. Mech. Appl.*, vol. 392, no. 21, pp. 5490–5500, Nov. 2013, doi: 10.1016/j.physa.2013.06.059.
[14] J. Ren, C. Wang, H. He, and J. Dong, "Identifying influential nodes in weighted network based on evidence theory and local structure," *Int. J. Innov. Comput. Inf. Control*, vol. 11, no. 5, pp. 1765–1777, 2015.
[15] P. Bonacich and P. Lloyd, "Eigenvector-like measures of centrality for asymmetric relations," *Social Netw.*, vol. 23, no. 3, pp. 191–201, 2001, doi: 10.1016/S0378-8733(01)00038-7.
[16] M. Kitsak *et al.*, "Identification of influential spreaders in complex networks," *Nature Phys.*, vol. 6, pp. 888–893, Aug. 2010, doi: 10.1038/nphys1746.
[17] S. Brin and L. Page, "The anatomy of a large-scale hypertextual Web search engine," *Comput. Netw.*, vol. 56, no. 18, pp. 3825–3833, 2012.
[18] L. Lü, Y.-C. Zhang, C. H. Yeung, and T. Zhou, "Leaders in social networks, the *delicious* case," *PloS ONE*, vol. 6, no. 6, p. e21202, Jun. 2011, doi: 10.1371/journal.pone.0021202.
[19] T. Harko, F. S. N. Lobo, and M. K. Mak, "Exact analytical solutions of the Susceptible-Infected-Recovered (SIR) epidemic model and of the SIR model with equal death and birth rates," *Appl. Math. Comput.*, vol. 236, pp. 184–194, Mar. 2014. doi: 10.1016/j.amc.2014.03.030.
[20] B. Wei, J. Liu, D. Wei, C. Gao, and Y. Deng, "Weighted k-shell decomposition for complex networks based on potential edge weights," *Phys. A, Statist. Mech. Appl.*, vol. 420, pp. 277–283, Feb. 2015, doi: 10.1016/j.physa.2014.11.012.
[21] A. Barrat, M. Barthélemy, R. Pastor-Satorras, and A. Vespignani, "The architecture of complex weighted networks," *Proc. Nat. Acad. Sci. USA*, vol. 101, no. 11, pp. 3747–3752, Mar. 2004.
[22] W. W. Zachary, "An information flow model for conflict and fission in small groups," *J. Anthropol. Res.*, vol. 33, no. 4, pp. 452–473, 1977, doi: 10.1086/jar.33.4.3629752.
[23] M. E. J. Newman, "Finding community structure in networks using the eigenvectors of matrices," *Phys. Rev. E, Stat. Phys. Plasmas Fluids Relat. Interdiscip. Top.*, vol. 74, no. 3, p. 036104, 2006.
[24] M. E. J. Newman, "The structure of scientific collaboration networks," *Proc. Nat. Acad. Sci. USA*, vol. 98, no. 2, pp. 404–409, 2001.
[25] S. Wang, Y. Du, and Y. Deng, "A new measure of identifying influential nodes: Efficiency centrality," *Commun. Nonlinear Sci. Numer. Simul.*, vol. 47, pp. 151–163, Jun. 2017, doi: 10.1016/j.cnsns.2016.11.008.
[26] P. L. I. Jinan, "Epidemic spreading in weighted scale-free networks," *Math. Pract. Theory*, vol. 40, no. 19, pp. 127–131, 2010.
[27] M. G. Kendall, "A new measure of rank correlation," *Biometrika*, vol. 30, pp. 81–93, Jun. 1938.

QIAN WANG received the B.S., M.S., and Ph.D. degrees from the School of Information Science and Engineering, Yanshan University, China, in 2009, 2012, and 2016, respectively. She has been a Visiting Scholar with the University of Hull from 2015 to 2016.

Since 2016, she has been a Lecture with the School of Information Science and Engineering, Yanshan University, China. Her research interests include data mining, complex network, and software security.



JIADONG REN received the B.S. and M.S. degrees from the Northeast Heavy Machinery Institute in 1989 and 1994, respectively, and the Ph.D. degree from the Harbin Institute of Technology in 1999.

He is a Professor with the School of Information Science and Engineering, Yanshan University, China. His research interests include data mining, complex network, and software security. He is a Senior Member of the Chinese Computer Society and a member of the IEEE SMC Society and the ACM.

YU WANG received the B.S. degree from the School of Modern Science and Technology, Hebei Agricultural University, China, in 2015. She is currently pursuing the master's degree with the School of Information Science and Engineering, Yanshan University, China. Her research interests include data mining and complex network.

BING ZHANG received the B.S. degree in mechanical engineering from the College of Computer and Information Technology, Three Gorges University, China, in 2012, and the Ph.D. degree from the School of Information Science and Engineering, Yanshan University, China, in 2018.

Since 2018, he has been a Lecture with the School of Information Science and Engineering, Yanshan University, China. He is currently with the Norwegian University of Science and Technology as a Visiting Scholar. His research interests include data mining, machine learning, and software security.

YONGQIANG CHENG received the B.S. and M.S. degrees in control theory and control engineering from Tongji University, Shanghai, China, in 2001 and 2004, respectively, and the Ph.D. degree from the School of Engineering, Design and Technology, University of Bradford, U.K., in 2010.

From 2004 to 2007, he was a GSM/WCDMA Product Solution Manager with the ZTE Corporation Shanghai Research and Development Center. He has been a Post-Doctoral Scientific Researcher with the Future Ubiquitous Networking Laboratory, School of Engineering and Informatics, University of Bradford. He is currently a Lecturer with the Department of Computer Science, University of Hull, U.K.

XIAOLIN ZHAO received the B.S., M.S., and Ph.D. degrees from the School of Computer Science, Beijing Institute of Technology, Beijing, China, in 1993 and 2000, respectively.

He is currently an Associate Professor with the School of Software, Beijing Institute of Technology, Beijing, China. His research interests include software security, complex network, and software engineering theory.