Journal of Network Intelligence ©2024 ISSN 2414-8105 (Online)
Taiwan Ubiquitous Information Volume **9**, Number **1**, February **2024**

A Knowledge-aggregated Approach for Identifying Influential Nodes in Dynamic Social Networks

Jing-Dong Wang

School of Computer Science
Northeast Electric Power University
169 Changchun Road, Jilin City, Jilin Province, 132012, China
School of Computer Science
Guangdong Atv Academy For Performing Arts
Huijing Road, Dongguan City, Guangdong Province, 523710, China
707569380@qq.com

Tong Liu

School of Computer Science
Northeast Electric Power University
169 Changchun Road, Jilin City, Jilin Province, 132012, China
18754810372@163.com

Qi-Zi Mu*

School of Computer Science
Northeast Electric Power University
169 Changchun Road, Jilin City, Jilin Province, 132012, China
17855830824@163.com

Fan-Qi Meng

School of Computer Science
Northeast Electric Power University
169 Changchun Road, Jilin City, Jilin Province, 132012, China
mengfanqi@neepu.edu.cn

Guang-Qiang Qu

School of Computer Science
Northeast Electric Power University
169 Changchun Road, Jilin City, Jilin Province, 132012, China
1316349361@qq.com

*Corresponding author: Qi-Zi Mu

Received March 8, 2023, revised May 7, 2023, accepted July 25, 2023.

Abstract. *To improve the efficiency of the nodes' influence evaluation methods in dynamic social networks that are constantly and complex changing over time, we propose a knowledge-aggregate-based method. The method firstly utilizes the time window graph model and time-aggregated graph model to aggregate knowledge of the dynamic network topology, then comprehensively consider the connection strength, the local influence of the third-order and fourth-order neighbors, and the bridging influence of each node to measure nodes' influence in the entire network. Our innovation lies in the use of knowledge aggregation to capture the dynamic changes of complex networks, which overcomes the shortcoming of traditional methods which recalculate influential nodes every time the network changes; and taking into account the multi-dimensional aggregated information of nodes throughout the network dynamic evolution to improve the algorithm accuracy. SIR model is used to evaluate the performance of the proposed approach in information dissemination in real networks, and the results show that this algorithm outperforms current algorithms on both small-scale and large-scale dynamic networks, and performs more prominently on large-scale dynamic social networks. Meanwhile, the algorithm proposed in this article has lower computational complexity than existing algorithms, which can effectively improve the efficiency of identifying key nodes in large-scale social networks.*
Keywords: Dynamic social network, Knowledge aggregation, Local influence, Information dissemination, Bridging influence

---

1. **Introduction.** OCIAL networks [1] are typical large-scale complex networks that dynamically evolve. Influential nodes, as nodes with high influence in social networks, can easily and quickly spread the information throughout the network after receiving the information; and at the same time, if they are effectively controlled, it will be difficult for the information to spread fast across the network. Therefore, identifying influential nodes in social networks plays a vital role in applications such as epidemic prevention and control, public opinion control, word-of mouth marketing, and advertising. To quantitatively analyze the influence of nodes in the network, scientists have proposed a large number of centrality methods, such as degree centrality(DC) [2], betweenness centrality (BC) [3], closeness centrality (CC) [4], Eigenvector centrality(EC) [5], K-shell decomposition [6], PageRank [7], LeaderRank [8], VoteRank [9], CumulativeRank [10], LIR [11], LocalRank [12], etc. Most of these methods use the node's attributes and location as the standard to measure the influence of a node. In recent years, with the deepening of research, new research results continue to emerge. Liu et al. [13] proposed the general tightness index GCC, and approximately found multiple nodes with the highest GCC as the influential nodes in the network through the K-means method. Bian et al. [14] proposed a sorting method based on the analytic hierarchy process, using the multi-attribute decision-making technology AHP application to aggregate several centrality algorithms to evaluate the influence of each node, and the node with the highest influence are regarded as influential nodes in the network.

However, the above-mentioned methods are mainly defined for static networks. With the rapid development become a new feature of social networks, it presents the characteristics of dynamic evolution over time [15]. In this case, it is no longer appropriate to directly apply the methods in static networks for influential nodes identification to a dynamic social network. Specifically, some high centrality nodes may die out quickly, that is, they are only active within a few time steps. In this case, it is difficult to judge whether it is important in the entire dynamic evolution network. Based on this background, temporal networks containing time information are more suitable to describe social networks in the current form [16]. So far, many researchers have begun to pay attention to this problem. Since the research on dynamic networks is still in its infancy, many scholars have extended the influential nodes identification methods of static networks to timeporal networks only.

Kim et al. [17] proposed a time-ordered graph model, which simplified a time-stamped dynamic network into a directional flow network, and extended the time versions of DC, BC, and CC through this model. Jiang et al. [18] proposed an Attenuation-Based Supra-Adjacency Matrix (ASAM) temporal network modeling method based on the attenuation of the inter-layer coupling strength and evaluate the nodes' influence by culating the eigenvector centrality of the nodes in each time layer in the temporal network. Ye et al. [19] used the time window graph model to propose edge-based dynamic K-shell centrality. Taylor et al. [20] proposed an eigenvector-based measure for temporal networks, in which the eigenvectors of the hyper centrality matrix and its components can reflect the influence of nodes. These methods have their unique advantages and disadvantages. The most obvious disadvantage is that the centrality index needs to be recalculated every time the network changes, or the efficiency cannot be effectively guaranteed. Recently, some scholars have further extended the centrality algorithms on the temporal networks on the basis of the predecessors [21,22,23,24]. Tulu et al. [25] proposed the NWI algorithm. They consider the willingness of the neighbors within the two hops of the node to spread information in the network and combine the clustering coefficient to measure the influence of nodes in the dynamic networks. Though the method has low computational complexity, the neighborhood information of nodes is not fully considered, and it cannot guarantee that the selected nodes are critical in the entire dynamic network.

Based on the strong clustering structure characteristics of social networks and the existing research in static networks, the influential nodes in dynamic social networks should not only dominate the local location but also play a good bridging role among clusters. Therefore, we propose a new method for identifying influential nodes in dynamic social networks: on the premise of using the time window graph and time aggregation graph model to aggregate the knowledge of the node characteristics in each tense, the influence of nodes in the network is comprehensively measured by considering the contact times and connection strength between nodes and other nodes, the number of second-order, third-order and fourth-order neighbors, the local influence of topological relations and the bridging influence of nodes; and the nodes with the highest score are the influential nodes in the dynamic social networks. The main contributions of this article are as follows:

(1) Propose a new influential nodes identification method for dynamic social networks based on knowledge aggregation. In the method, the neighbor's information of nodes at each time is fully considered.

(2) The method not only utilize the local information but also considers the bridging influence. So that the obtained nodes have more advantages in the topology.

(3) Proposed an improved temporal version of the network constraint coefficients.

(4) A large number of experiments have been conducted on real data sets. The experimental results show that the correlation between the proposed method and the real dissemination influence is higher than that of other comparison approaches. The accuracy of our method is higher, and it receives a balance between computational complexity and computational accuracy.

2. **Preliminaries.** In this section, the time window graph model and time-aggregated graph model are completely adopted to compose our knowledge-aggregated model. First, we divide the dynamic social network into a series of time windows in a discrete manner. And then, using time-aggregated graph model to summarize all the nodes of the network and the interaction information between nodes to obtain the aggregate connectivity matrix M_aggre of the entire dynamic network.

2.1. **Time Window Graph Model.** The time window graph is a static network relative to the active nodes and edges at a given time in a dynamic network [26,27]. In this paper, the dynamic social network is modeled as a series of continuous time windows $GT = G^1, G^2, ..., G^t$, where the total number of nodes N remains unchanged, and the edges E changes at different time intervals. The size of a window is defined as m, and each window is used to describe the network relationships and network interactions gathered in the m period, as shown in Figure 1. The time window graph in the entire period T can be expressed as $GT = (NT, ET, MT)$, where $NT \in N$ represents all nodes in the network, $ET \in E$ represents the edge set in the dynamic network, and MT represents the connection matrix corresponding to the connected edges of all nodes in the entire T time domain, and the number of time windows in this period is calculated as T/m. The t time window is defined as $G^t = (V, E^t)(t \in 1, 2, \ldots, T)$, where $V = V_1, V_2, \ldots, V_n$ represents nodes, and Et represents the edges with in the time interval $(t, t+1)$.

2.2. **Time Aggregated Graph Model.** The time-aggregated graph model is used to model the changes of nodes and edges in a series of time windows [28]. The edges between each pair of nodes in the network are represented by a three-dimensional array. For example,$(v_1, (v_2, t))$ represents the connection of node v_1 and v_2 at travel time t. By counting the edges of nodes in each time window, the time-aggregated graph of the dynamic network is obtained. Figure 1 shows a continuous time window graph of a simple dynamic network. The time-aggregated graph model obtained from the time window graph model of Figure 1 is shown in Figure 2. In the adjacency matrix, under the condition of record $(v_i, (v_j, t)) = 0$ where node v_i and v_j have no edg at time t, by accumulating the number of occurrences of each node's edges, the aggregate connection matrix M_{aggre} of the entire time domain is obtained, as shown in Figure 3. Take the edges between node 3 and node 7 as examples. In the aggregation connection matrix (3, 7)=3, which indicates that nodes 3 and node 7 (or nodes 7 and node 3) have met 3 times, which is displayed in the time-aggregated graph as it establishes a connection at $t = 2$, $t = 5$, and $t = 7$, and disappears at other times. Meanwhile, it can be seen that in the undirected graph, the aggregate matrix is a symmetric matrix, in which each parameter value is between 0 and T/m.

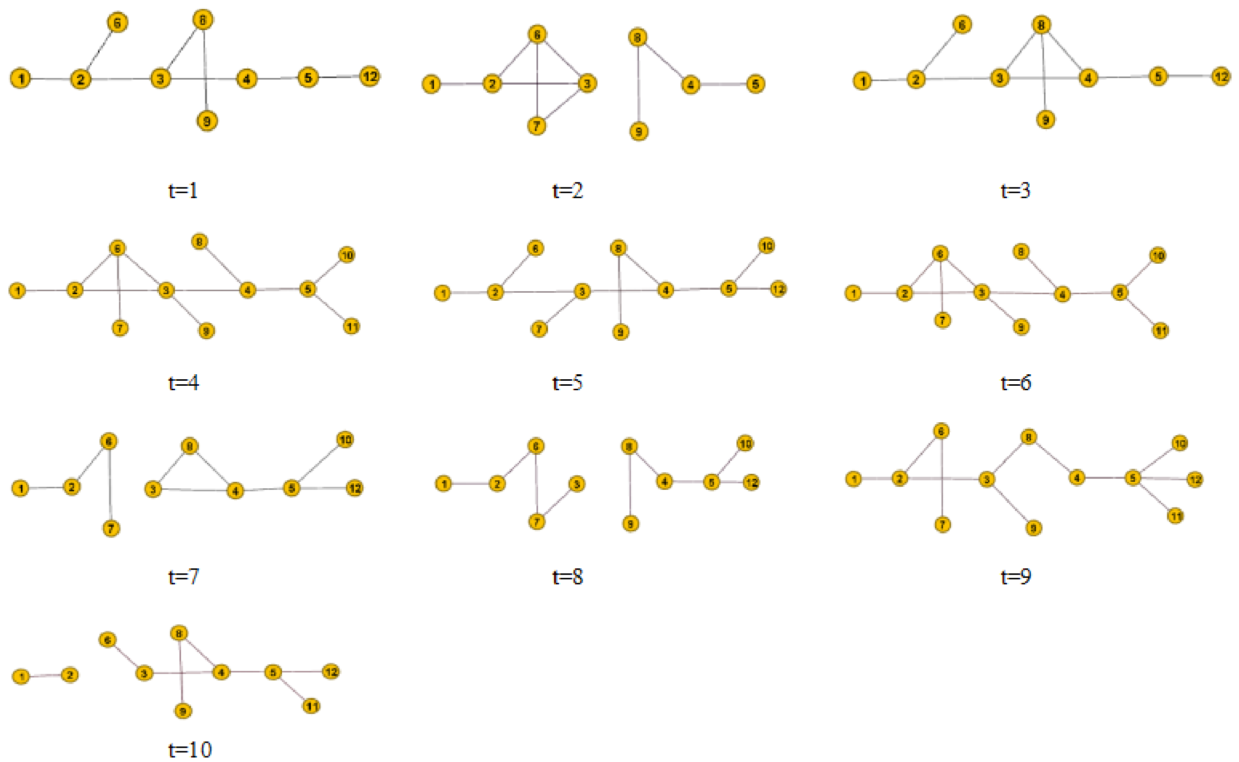

Figure 1. Time window graph model.

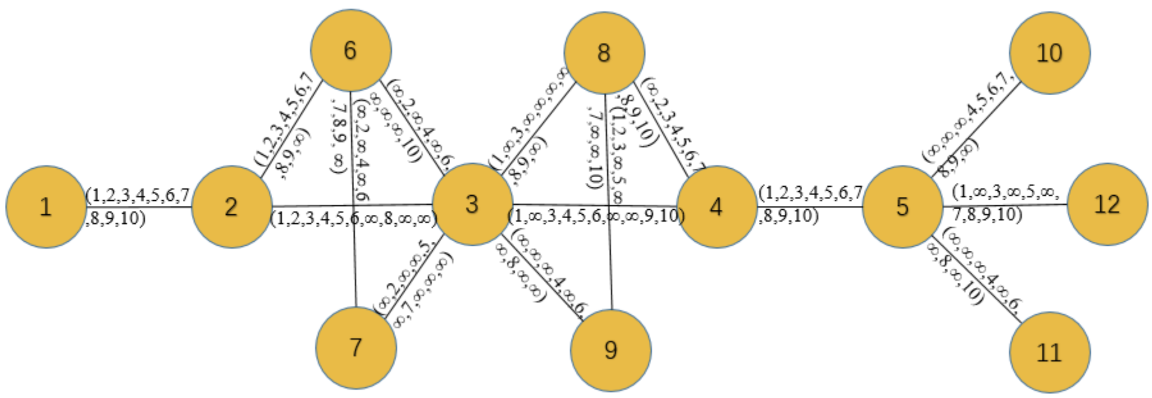

Figure 2. Time aggregated graph model.

	1	2	3	4	5	6	7	8	9	10	11	12
1	0	10	0	0	0	0	0	0	0	0	0	0
2	10	0	7	0	0	9	0	0	0	0	0	0
3	0	7	0	7	0	4	3	4	3	0	0	0
4	0	0	7	0	10	0	0	9	0	0	0	0
5	0	0	0	10	0	0	0	0	0	6	4	7
6	0	9	4	0	0	0	6	0	0	0	0	0
7	0	0	3	0	0	6	0	0	0	0	0	0
8	0	0	4	9	0	0	0	0	6	0	0	0
9	0	0	3	0	0	0	0	6	0	0	0	0
10	0	0	0	0	6	0	0	0	0	0	0	0
11	0	0	0	0	4	0	0	0	0	0	0	0
12	0	0	0	0	7	0	0	0	0	0	0	0

FIGURE 3. Aggregate connection matrix.

3. **A Knowledge Aggregated Approach For Identify Influential Nodes In Dynamic Social Networks.** The relationship between nodes in a dynamic social network is no longer a simple two-tuple, but is constantly evolving. A node that remains active for a long time and has frequent contact with other neighbors must have a higher influence in the network than those who only appear in certain time windows and remain isolated; while in a social network with a strong clustering structure, the most influential nodes should be the center of each cluster and the node that acts as a 'bridge' between each cluster. In addition, the local structure information of the node in the topology plays an important role in the rapid spread of information in the network. Based on the above analysis, we define the influential nodes in the dynamic social network as those who have a relatively long lifespan during the entire time evolution process, frequently contact with other nodes, and occupy local structural advantages and bridging advantages in the topology. Firstly, the weight and connection strength of nodes in the dynamic network are described by the contact times between a node and its direct neighbors in the whole temporal. Secondly, the bridging influence of nodes is described by using the structural hole theory, and the local influence of nodes is measured by using the LIR algorithm. Finally, the influence of each node in the network is obtained by combining all the above indicators. In addition, because LIR has the property that the neighbor degree of the node is smaller than its own degree, the proposed method can also reduce the influence of node radiation range overlap on the propagation, and avoid the rich club effect to a certain extent.

3.1. **Specific Method.** Based on the Knowledge-aggregated model, the evolution of all the edges and nodes of the network is captured to obtain the aggregate connection matrix M_aggre of the entire dynamic network. Assign the edge weight M_{ij} to each edge according to the actual situation, which is equal to the number of contacts between two nodes. M_i represents the total weight of node i in the entire period T. In the real world, it is conceivable that a node with a small total weight will either have a very short lifespan; or it will often appear as a solitary spot, and rarely contact its neighbors in terms of time or space. In either case, it is difficult to say that this node will have an advantage in information dissemination. On the contrary, a node with a large total weight can be inferred that its lifespan is relatively long and has frequent and close contact with surrounding nodes. It must be more effective than other nodes in the process of information dissemination. The normalized weight of node i in a dynamic network is defined as:

$$nwh_i = \frac{M_t}{\sum_{i=1}^{N} M_i} \tag{1}$$

The influence given to a node by the number of node contacts is defined as the node's connection strength, which represents the strength of the connection between the node and other neighbors. The larger the value, the closer the connection between the node and other nodes, and the stronger its inclination to exchange information. The connection strength of node i is defined as:

$$CS_i = \frac{w}{w - M_i} \tag{2}$$

where w is the number of edges of all nodes in the entire dynamic network. The total number of third-order neighbors of node j is calculated by the following formula:

$$Q_j = \sum_{k \in \tau j} R_k \tag{3}$$

Where node k is the direct neighbor of node j. R_k represents the union of the second-order neighbors of node k. In the dynamic network, the third-order local influence of node j is calculated by the following formula:

$$WQ_j = \sum_{k \in \tau j} M_{kj} R_k \tag{4}$$

The fourth-order local influence of node i is calculated by the following formula:

$$WLR_i = \sum_{j \in \tau i} M_{ij} Q_j \tag{5}$$

From a sociological point of view, structural holes are gaps between non-redundant contacts, which measure the ability of a node as a bridge between two unrelated nodes. Burt [29] use the network constraint coefficient to measure the constraints that nodes receive when they form structural holes. Nodes with small constraint coefficients often have a strong influence in information dissemination. Here, by extending the time version of the improved network constraint coefficient $INCC$ in the CumulativeRank algorithm, the bridging influence of the node is used as a measure of the influence of the node in the social network when the neighbor information within the third and fourth order of the node is fully used. The constraint coefficient of node i in the dynamic social network $TINCC_i$ is calculated by the following formula:

$$TINCC_i = \sum_{j \in \tau_i} \left(\frac{WQ_j}{WLR_i} + \sum_{l=1, l \neq i, l \neq j}^{N} \frac{WQ_l}{WLR_i} \frac{WQ_j}{WLR_l}\right)^2 \tag{6}$$

Since the value of $TINCC$ decreases with the increase of node influence, in order to measure the bridging advantage of a node in a dynamic social network, the bridging influence of node i is defined as:

$$TL_i = \frac{1}{1 + e^{TINCC_i}} \tag{7}$$

Since the third-order and fourth-order neighbors' quantity information of the node is used in this indicator, the local structural advantage of the node is considered within a certain range. Moreover, in order to fully consider the topological relationship between node neighbors, the LIR is introduced, and the LI value is also taken as a consideration factor to measure the local influence of the node. The formula for calculating the LI value of node i is as follows:

$$LI_i = \sum_{j \in \tau_i} Q(d_j - d_i) \tag{8}$$

Where d_i represents the degree of node i, for the function $Q(x)$, if $x <= 0$, then $Q(x) = 0$, otherwise, $Q(x) = 1$. For nodes whose LI value is not equal to 0, its nd value is set to 0. By filtering those $LI = 0$, that is, the node whose direct connection neighbors' degree is lower than its own degree, the nodes with the advantage of local information dissemination is obtained. At the same time, this also reduces the possibility of two influence nodes being directly connected, thereby reducing the generation of rich-club effects. Since the LI values of the selected nodes are the same, the inf luence of these nodes is sorted by considering the degree of the nodes, and the calculation formula is as follows:

$$nd_i = \frac{d_t}{\sum_{i=1}^{N} d_i} \tag{9}$$

Since the dimensions of each variable may be different, each index is normalized to obtain the influence of each node of the dynamic network. In this algorithm, the local information of nodes is used in the calculation of each index, so the proposed algorithm is defined as Nodes' Temporal Local-based Importance (NTL):

$$NTL_i = \left(\frac{TL_i}{\sqrt{\sum_{i=1}^{N} TL_i}} + \frac{nd_i}{\sqrt{\sum_{i=1}^{N} nd_i}} + \frac{CS_i}{\sqrt{\sum_{i=1}^{N} CS_i}}\right) \frac{nwh_i}{\sqrt{\sum_{i=1}^{N} nwh_i}} \tag{10}$$

Finally, sort the NTL values of all nodes in the network, and select the top-ranked node as the influential nodes of the entire dynamic social network.The overall algorithm is shown in Algorithm1.

4. **Simulation Results And Analysis.** In this part, firstly, the feasibility of the proposed algorithm in a small-scale network is verified by analyzing the experimental results of a small-scale data set karate club [30]. Secondly, by simulating the information dissemination process and comparing with other algorithms, it shows the performance of NTL algorithm in identifying influential nodes of dynamic social networks. Finally, the running time of all algorithms is compared to verify the effectiveness of the proposed algorithm in largescale networks. The data set used for the experiment is shown in the Table 1. Email-dnc [31] is a mail network in the 2016 Democratic National Committee email leak event, which includes 2039 nodes and 39264 edges that established contacts within

Algorithm 1

Input: symmetric adjacent matrix graph $(M_{aggre}) = (a_{ij})N \times N$ and adjacent table corresponding to weighted complex network

Output: The ranked list and the NTL of each node

for $i = 1$ to N **do**
  M_i = weight of the ith node
  d_i = degree of the ith node
  nwh_i = normalized weight of the ith node
  CS_i = Connection strength of the ith node
  for j in list(G.neighbors (i)) **do**
    Q_j = first-second-and-third neighbors of the ith node's neighbors
    WQ_j = first-second-and-third neighbors of the ith node's neighbors in weighted networks
    LI_i = the LI value of the ith node
  end for
  WLR_i = the fourth-order neighbors of the ith node in weighted networks
  if LI_i !=0 **then**
    $nd_i = 0$
  else
    $nd_i = \frac{d_t}{\sum_{i=1}^{N} d_i}$
  end if
end for
temporal $INCC$ value $TINCC_i = \sum_{j\in\tau_i}(\frac{WQ_j}{WLR_i} + \sum_{l=1,l\neq i,l\neq j}^{N} \frac{WQ_l}{WLR_i}\frac{WQ_j}{WLR_l})^2$
the local bridging influence of every node $TL_i = \frac{1}{1+e^{TINCC_i}}$
the NTL value of the ith node $NTL_i = (\frac{TL_i}{\sqrt{\sum_{i=1}^{N} TL_i}} + \frac{nd_i}{\sqrt{\sum_{i=1}^{N} nd_i}} + \frac{CS_i}{\sqrt{\sum_{i=1}^{N} CS_i}})\frac{nwh_i}{\sqrt{\sum_{i=1}^{N} nwh_i}}$
rank1 = Sort(NTL, 'descend')
return $rank_1$

2 days. Social network [32] is the social connections of 72 teachers and students collected by Bucharest University of Technology in 2012 within 63 days. n and m represent the total number of nodes and edges of the network, respectively. The maximum degree and average degree are expressed by k_{max} and $\langle k \rangle$ respectively.

Table 1. Dataset for experiment

Network	n	m	$\langle k \rangle$	k_{max}
karate	34	78	4.59	17
Email-dnc	2029	39264	40	5.5k
Social network	72	700	9.72	31

4.1. Node dissemination ability evaluation.

4.1.1. *SIR model.* When analyzing the influence of nodes, many scholars use SIR models to simulate the dynamic spread of information and diseases. At the beginning of the spread, the nodes in the SIR model are divided into three states, susceptible state, infected state and recovery state. Frist, the influential nodes obtained through the algorithm are regarded as an infected node, and other nodes in the network are set to a susceptible state.

In each propagation iteration, each infected node infects its randomly selected neighbors with a probability of μ. The threshold of μ is defined as:

$$\mu_{max} = \frac{\langle k \rangle}{\langle k^2 \rangle - \langle k \rangle} \tag{11}$$

At the same time, each infected node will recover with a probability of β and will not be infected again. The infection rate λ is defined as:

$$\lambda = \frac{\mu}{\beta} \tag{12}$$

In order to spread information widely in the network, in the following experiment, the infection rate is set near its threshold. The propagation ability of nodes was compared by using the infection scale at time t. The infection scale at time t is defined as:

$$F(t) = \frac{n_{I(t)} + n_{R(t)}}{n} \tag{13}$$

Where,$n_{I(t)}$and $n_{R(t)}$respectively represent the number of nodes in the infected state and the number of nodes recovered state at time t. N is the number of summary points in the network. At time t, the larger $F(t)$ is, the more nodes are infected by the initial influence nodes. For the same $F(t)$, the smaller the T is, the faster the influence of nodes will spread in the network. $F(t_c)$ denotes the final affected scale. The larger the value, the stronger the propagation capability of the initial node. $F(t_c)$ is defined as:

$$F(t) = \frac{n_{R_{(t_c)}}}{n} \tag{14}$$

Where t_c represents the time for the propagation to reach a steady state, and represents the number of eventually infected nodes in the network.

4.1.2. *Kendall's Tau correlation coefficient.*

The Kendall correlation coefficient is often used to measure the ordered classification data set. It is used here to describe the correlation between the ranking results obtained based on the algorithm and the actual propagation ability of the node, and its value is between $[-1, 1]$. Assuming that X and Y are two different sorted sequences with the same number of elements N, the i-th value randomly obtained from the two sequences X, Y is represented by x_i and y_i respectively, then when the point pair (x_i, y_i) and (x_j, y_j) satisfies $x_i < x_j \wedge y_i < y_j$ *or* $x_i > x_j \wedge y_i > y_j$, this set of point pairs is said to belong to a consistent relationship. And when $x_i < x_j \wedge y_i > y_j$ *or* $x_i > x_j \wedge y_i < y_j$, then this set of point pairs is said to belong to an inconsistent relationship. When $x_i = x_j$ or $y_i = y_j$ occurs, it is neither consistent nor inconsistent. The Kendall correlation coefficient values between the two variables are -1, 1, and 0 indicating complete disagreement, complete agreement, and no relationship, respectively. The Kendall coefficient between two variables is defined as follows:

$$\tau = \frac{2(N_1 - N_2)}{N(N-1)} \tag{15}$$

Where N_1 and N_2 represent the number of consistent point pairs and inconsistent point pairs in sequences X and Y, respectively. In this section, by setting X as the centrality measurement result (such as dynamic betweenness centrality, dynamic degree centrality, dynamic closeness centrality, etc.), set Y as the SIR propagation range of each node to evaluate the Pearson correlation between the influential nodes identification results and the true propagation.

4.2. **Experimental Results and Analysis.**

4.2.1. *Analysis of experimental results in small-scale network.*

First, use the small-scale network karate to verify the feasibility of the proposed algorithm on small-scale networks. The network topology and the influential nodes extracted by each algorithm are shown in Figure 4. The blue nodes in the figure are ordinary nodes, and the red nodes are influential nodes. Table 2 lists the top-6 nodes obtained by the five comparison algorithms TDC (temporal degree), TB (temporal between- ness), TC (temporal closeness), NWI and NTL that proposed in this paper. According to the results, it can be seen that in the karate network, the result of the influential node identified by TC has the lowest repetition rate with other results. TB and TDC have the same performance, which shows that in the small network, TB plays the same role as DC. When the algorithm complexity is higher than TDC, TB loses its utility. The results of NTL and NWI are consistent, and the identified influential nodes include the results of all other algorithms. It shows that NTL can better identify the nodes located in the center of the network.

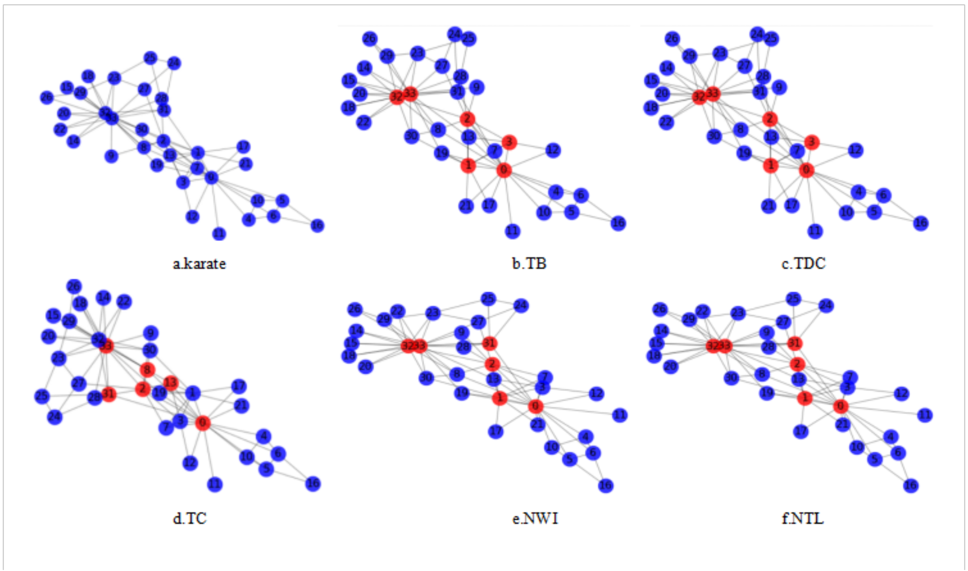

FIGURE 4. Influential nodes obtained by different algorithms in small-scale network karate.

TABLE 2. Ranking of karate network influence nodes

Rank	DC	BC	CC	NWI	NTL
1	33	0	0	33	33
2	0	33	2	0	0
3	32	32	33	32	32
4	2	2	31	2	2
5	1	3	8	1	1
6	3	1	13	31	31

4.2.2. *Analysis of experimental results in large-scale network.*

(1) Difference analysis of algorithms in node influence ranking. The SIR model mentioned above is used to verify the accuracy of the proposed method in large-scale dynamic social networks. Firstly, in order to clearly show the difference between the proposed method and other algorithms in node centrality measurement results, we use Email-dnc and social network to evaluate the effectiveness of the NTL in identifying influential nodes in dynamic networks. Take the top-20 nodes ranked in each algorithm as the initial infection node, and extract a snapshot from the Email-dnc network to analyze the information dissemination based on the SIR model. These nodes, which only appear in the top-20 of each algorithm, just reflect the differences in the centrality of the evaluation of different algorithms. Take $t = 100$ to reach a steady state, perform 1000 independent experiments to get the average value, and use the total number of infected nodes and restored nodes during this period as the node's spread influence. The image of the propagation influence $F(t)$ obtained by each algorithm over time is shown in Figure 5. According to the results of Table 3 and Figure 5, compared with other algorithms, the results of the top 20 nodes in the NTL of the proposed algorithm have certain differences, and according to the analysis of curve changes, it can be seen that these are only in the NTL when acting as the source of infection. Compared with the different nodes in other algorithms, the emerging nodes perform better and have a stronger appeal. It shows that the accuracy of the proposed algorithm in identifying influential nodes is higher than other comparison algorithms. Secondly, a good influential nodes identification algorithm should not only be robust to the initial number of infected nodes, but also the network structure and infection rate. In order to further compare the differences between the algorithms, the performance of NTL and the other four algorithms under different infection rates were compared. By comparing the hit rate of the top-10% nodes of different algorithms with the top-10% nodes with the largest actual infection scale in the real SIR model, the performance of the algorithm is judged. The hit rate HR is defined as:

$$HR = \frac{|S \cap R|}{|R|} \tag{16}$$

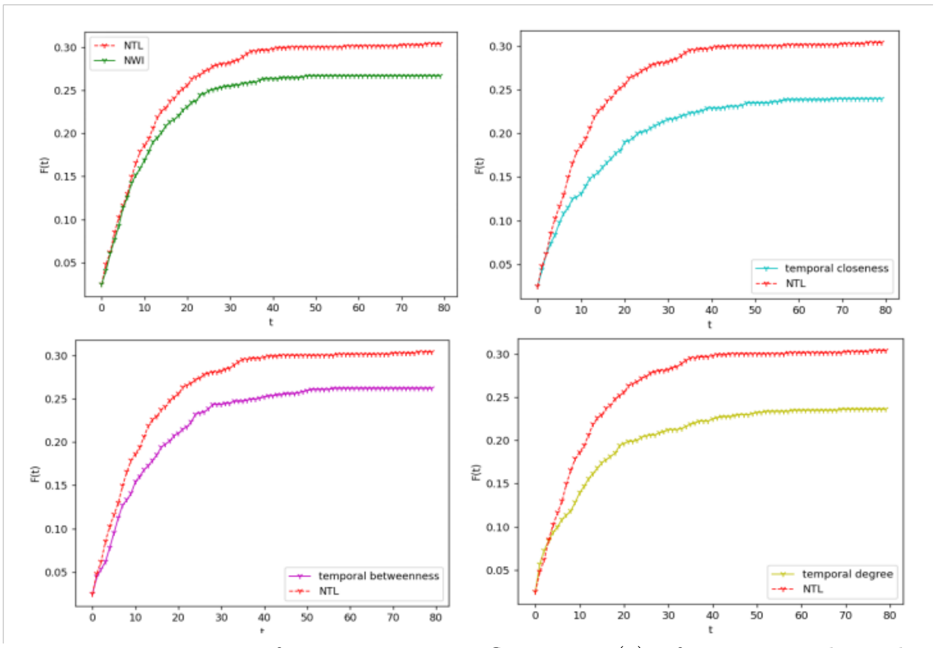

FIGURE 5. Variation of propagation influence $F(t)$ of top-20 nodes selected by each.

TABLE 3. Top-6 nodes in social network ranked by five algorithms

Method	1	2	3	4	5	6
NTL	25	13	15	30	35	8
NWI	25	13	15	30	35	8
TBC	25	31	6	30	27	13
TDC	15	25	13	30	35	34
TCC	25	8	30	49	35	53

Where S is the set of selected top-10% nodes, and R is the set of 10% nodes with the most infection rate obtained by SIR. The higher the HR, the higher the algorithm hit rate and the better the algorithm performance. Due to the variability of the dynamic network, in order to accurately perceive the criticality of the identified nodes in the entire network, randomly extract three time windows from the time window graph models of the two real dynamic social networks and perform information propagation analysis, the HR results obtained by each algorithm are shown in Figure 6. In addition, the top-6 nodes in the ranking of influence obtained by each algorithm in the social network are shown in Table 3.

It can be seen from Table 3 that in a relatively small-scale dynamic network, the NTL and NWI algorithms are consistent in the ranking results of node influence. Combining Figure 6 and Table 3, it can be concluded that in small network- social networks, in most cases, NTL and NWI can achieve consistent results, and the accuracy of identifying the most influential nodes is higher than other algorithms. In a large network with 2000 nodes, it can be seen that the accuracy of the NWI has decreased. Compared with the algorithm NTL proposed in this paper, it cannot well identify the influential nodes that exert the advantages of propagation in each period. Therefore, overall, compared with other algorithms, the influential nodes identified by NTL have more stable results and higher accuracy when they act as the source of infection.

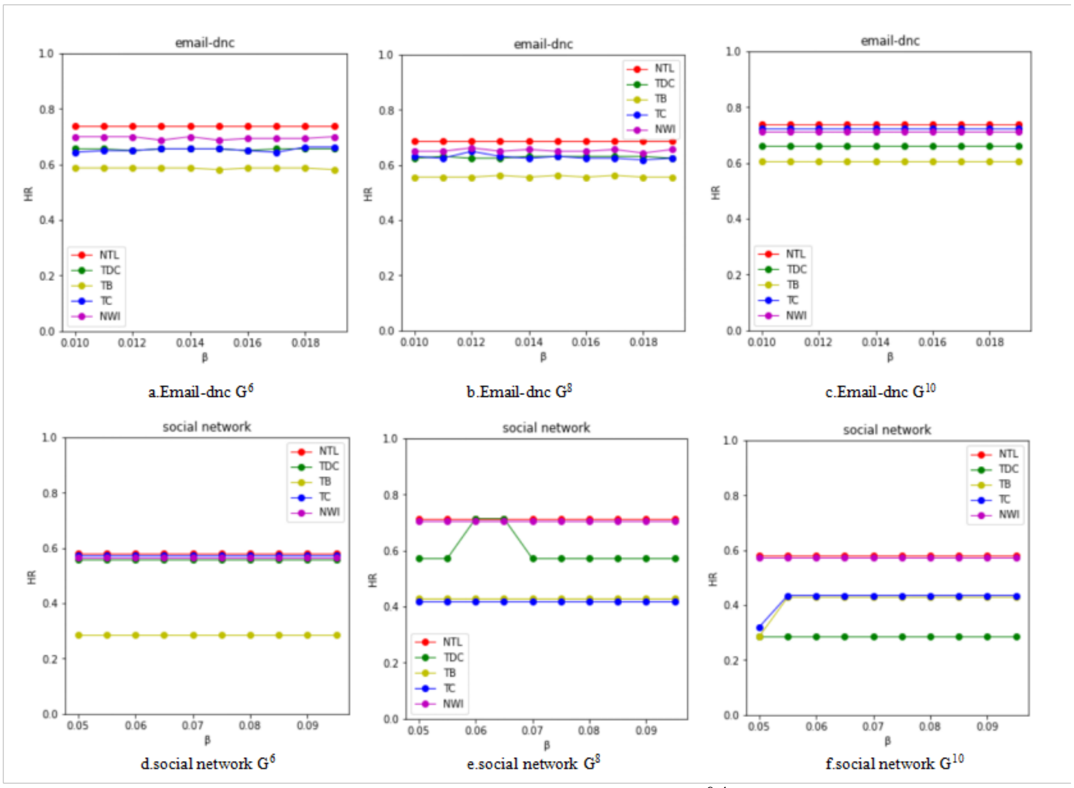

FIGURE 6. Comparison of HR values when top-10% nodes act as the source of infection in five algorithms.

(2)Analysis of the correlation between the sorting results and the true propagation capacity. Use the Kendall's Tau correlation coefficient to verify the effectiveness of the proposed algorithm. Take any node in the network as the initial source of infection, define the total number of infected nodes and recovered nodes in the network after the specified time $t = 10$ as the actual infection capacity of the node, and take the average of 1000 experiments to get the actual ranking of nodes'influence in real networks. The correlation analysis between the proposed method and the ranking results obtained by other methods in Email-dnc and social network can obtain the correlation score between the centrality methods and real communication ability. The higher the correlation score, the more accurate the algorithm is to identify influential nodes in dynamic social networks. Table 4 shows the correlation results between the influential nodes selected by algorithms and the actual influence in different networks. The correlation between the results of each algorithm is shown in Figure 7, and the correlation between the comparison algorithms and the actual influence is shown in Figure 8 (here only show the result in network Email-dnc). Heat map of correlation between methods and actual influence.

The horizontal and vertical axes in Figure 7 represent the comparison algorithms. The color represents the correlation with the true influence obtained by SIR. The lighter the color, the higher the correlation. According to Figure 7, it can be found that in the two networks, TDC, NWI, and NTL have a good correlation, while in the smaller social network, TB has little correlation with other algorithms. However, due to the discovery of the consistency of the concept of 'bridging nodes', the correlation between TB and NTL is still higher than in other algorithms. This is because although bridging nodes have advantages in spreading information between clusters, the number of bridging nodes in the network is relatively small, and nodes occupying local topological advantages play a key role in information dissemination in each cluster. Therefore, compared with other algorithms, the correlation between TB and SIR is low, which is shown in Figure 8 that is, the correlation curve is more divergent. At the same time, the correlation between TC and other algorithms is generally small. Combining with Figure 8, it is found that TC fails in a network with a high degree of aggregation. Therefore, it is not accurate to measure the importance of nodes in the entire dynamic network only by proximity centrality. In addition, there is a positive correlation between all methods and actual influence. The principle that the greater the'degree', the more important the node is still valid in dynamic networks, the more obvious the performance in smaller networks. However, when the community structure is unbalanced, the top-ranked nodes may be concentrated in the same community, leading to the limitation of information dissemination. Compared with NWI, it can be seen that in a large-scale Email-dnc network, NTL has a greater correlation with the actual influence of each node, achieving a result higher than NWI by 7%. Combined with the above analysis, although NWI performs well in each network, NTL performs better than NWI in larger networks and performs in the same way as NWI in smaller networks, which shows that the key nodes identified by NTL It has good information dissemination ability in the dynamic evolution of the entire network. Therefore, the algorithm NTL proposed in this paper can better find the nodes that occupy the advantage of information dissemination in the entire dynamic network structure.

TABLE 4. Comparison of Kendall's Tau correlation coefficient of node ranking obtained by each algorithm and SIR

Methods	TB	TC	TDC	NWI	NTL
Email	0.83	0.53	0.89	0.86	0.93
Social network	0.51	0.68	0.93	0.85	0.84

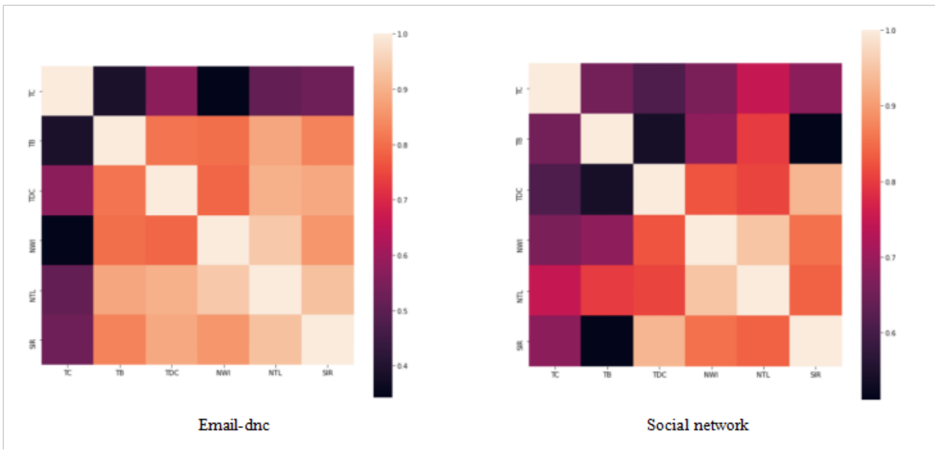

FIGURE 7. Heat map of correlation between methods and actual influence.

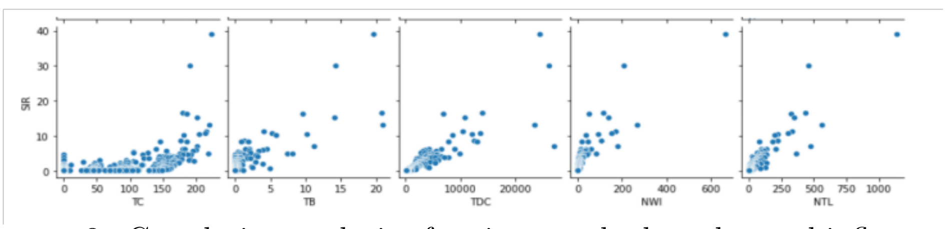

FIGURE 8. Correlation analysis of various methods and actual influence in Email-dnc.

5. **Running Time.** The running time comparison with TB, TC, TDC, and NWI algorithm is shown in Figure 9 (the processor is Intel(R) Core (TM) i7-5500U CPU, the memory is 8G, and the system is win10). It can be seen from the figure that the efficiency of NTL algorithm is second only to TDC algorithm. However, according to the above analysis, the accuracy of the TDC in large-scale social networks is lower than that of the proposed algorithm, and the running time of the NTL increases slowly with the increase of the number of nodes, and the gap with TDC remains within 1 second. Therefore, it achieves a balance between algorithm efficiency and accuracy, and is more suitable for large-scale social networks.

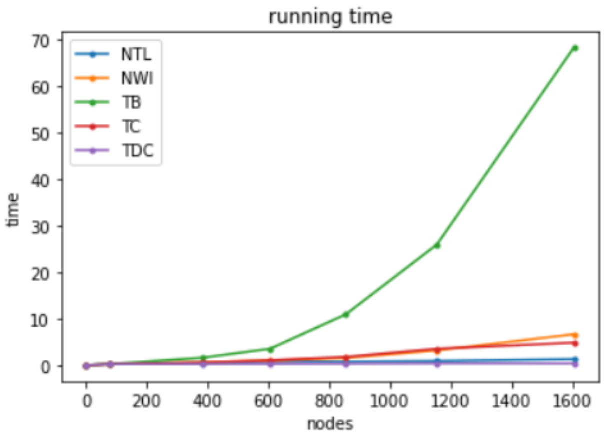

Figure 9. Running time.

6. **Conclusions.** A social network is a large-scale and complex network that continues to dynamically evolve. Traditional social network influential nodes identification methods are mostly based on static network analysis, which is not suitable for real situations. In this paper, we propose a method for identifying influential nodes in dynamic social networks based on knowledge aggregation. It uses time window graphs model and time-aggregated graph models to aggregate dynamic network topology information, which overcomes the shortcomings of the traditional method of recalculating the centrality index of all nodes in the network every time the network changes. Meanwhile, this method fully considers the multi-dimensional aggregated information of the nodes in the network evolution process, so that the identified influential nodes are more in line with the laws of the dynamic evolution of social networks over time, which is of great importance for network public opinion control, epidemic prevention and control, and advertising and marketing. Important role. From the simulation results, the proposed algorithm NTL has achieved a good balance of computational complexity and ranking accuracy, and the results obtained are more accurate and stable than TB, TC, TDC, and NWI, and are suitable for rapidly developing large-scale dynamic social networks.However, the algorithm proposed in this paper has no significant performance improvement on small-scale networks, and its performance is consistent with that of the current optimal algorithm. Therefore, the algorithm proposed in this paper still has a large room for improvement on small-scale networks, and it will try in this direction in the future.

The research results show that the influence of nodes in dynamic social networks not only depends on the number of contacts between node pairs in the tense but also on their local influence in the topology and structural superiority. However, which of these three factors plays a key role in the information transmission in the whole period still needs to be further studied. We will further discuss the influence of these factors on the key node recognition algorithm in the future research process. In addition, when conducting network public opinion and disease control, it is of more practical significance to predict the key nodes in the next time period. The key nodes in the next time period can often bring more valuable information, which is also worthy of further research.

Acknowledgment. This work is one of the research results of the Jilin Provincial Natural Science Foundation's "Research on Large scale Group Decision Optimization Algorithm for Incomplete Hesitation Fuzzy Information" (No.20230101176JC). The authors also gratefully acknowledge the helpful comments and suggestions of the reviewers, which have improved the presentation.

REFERENCES

[1] T.-Y. Wu, X.-L. Guo, L. Yang, Q. Meng, C.-M. Chen, "A lightweight authenticated key agreement protocol using fog nodes in Social Internet of vehicles, " *Mobile Information Systems*, vol. 2021, 3277113, 2021.

[2] L. Candeloro, A. Conte, "A new weighted degree centrality measure: The application in an animal disease epidemic," *PloS One*, vol. 11, no. 11, p. e0165781, 2016.

[3] L. Hoang, M. Pontecorvi, R. Dathathri, G. Gill, V. Ramachandran, "A round-efficient distributed betweenness centrality algorithm," in *Proceedings of the 24th Symposium on Principles and Practice of Parallel Programming*, 2019, pp. 272-286.

[4] C. Salavati, A. Abdollahpouri, Z. Manbari, "Ranking nodes in complex networks based on local structure and improving closeness centrality," *Neurocomputing*, vol. 336, pp. 36-45, 2019.

[5] L. Laporta, J. Afonso, I. Mesquita, "Interaction network analysis of the six game complexes in high-level volleyball through the use of Eigenvector Centrality," *PloS One*, vol. 13, no. 9, p. e0203348, 2018.

[6] K. Angelou, M. Maragakis, P. Argyrakis, "A structural analysis of the patent citation network by the k-shell decomposition method," *Physica A: Statistical Mechanics and its Applications*, vol. 521, pp. 476-483, 2019.

[7] H. Zhao, X. Xu, Y. Song, DL. Lee, Z. Chen, H. Gao, "Ranking users in social networks with motif-based pagerank," *IEEE Transactions on Knowledge and Data Engineering*, vol. 33, no. 5, pp. 2179-2192, 2019.

[8] B. Zhou, Y.-F. Lei, C.-B. Li, B. Fang, Q.-W. Wu, L.-J. Li, Z.-Q. Li, "Electrical LeaderRank method for node im- portance evaluation of power grids considering uncertainties of renewable energy," in *International Journal of Electrical Power and Energy Systems*, vol. 106, pp. 45-55, 2019.

[9] S. Kumar, B.S. Panda, "Identifying influential nodes in Social Networks: Neighborhood Coreness based voting approach," *Physica A: Statistical Mechanics and its Applications*, vol. 553, p. 124215, 2020.

[10] D. Zhang, Y. Wang, Z. Zhang, "Identifying and quantifying potential super-spreaders in social networks," *Scientific Reports*, vol. 9, no. 1, pp. 1-11, 2019.

[11] D. Liu, Y. Jing, J. Zhao, W.-J. Wang, G.-J. Song, "A fast and efficient algorithm for mining top-k nodes in complex networks," *Scientific Reports*, vol. 7, no. 1, pp. 1-8, 2017.

[12] D.-B. Chen, L.-Y. Lü, M.-S. Shang, Y.-C. Zhang, T. Zhou, "Identifying influential nodes in complex networks," *Physica A: Statistical Mechanics and its Applications*, vol. 394, no. 4, pp. 1777-1787, 2012.

[13] H.-L. Liu, C. Ma, B.-B. Xiang, M. Tang, H.-F. Zhang, "Identifying multiple influential spreaders based on generalized closeness centrality," *Physica A: Statistical Mechanics and its Applications*, vol. 492, pp. 2237-2248, 2019.

[14] T. Bian, J. Hu, Y. Deng, "Identifying influential nodes in complex networks based on AHP," *Physica A: Statistical Mechanics and its Applications*, vol. 497, pp. 422-436, 2016.

[15] C.-M. Chen, Y.-H. Lin, Y.-C. Lin, H.-M. Sun, "RCDA: Recoverable concealed data aggregation for data integrity in wireless sensor networks," *IEEE Transactions on parallel and distributed systems*, vol. 23 (4), 727-734, 2011.

[16] T.-T. Wang, W.-D. Dai, P.-P. Jiao, W.-J. Wang, "Identifying influential nodes in dynamic social networks based on degree-corrected stochastic block model," *International Journal of Modern Physics B*, vol. 30, no. 16, p. 1650092, 2016.

[17] H. Kim, R. Anderson, "Temporal node centrality in complex networks, *Physical Review E*, vol. 85, no. 2, p. 026107, 2012.

[18] J.-L. Jiang, H. Fang, S.-Q. Li, W.-M. Li, "Identifying important nodes for temporal networks based on the ASAM model,"*Physica A: Statistical Mechanics and its Applications*, vol. 586, p. 126455, 2022.

[19] Z. Ye, X. Zhan, Y. Zhou, C. Liu, Z.-K. Zhang, "Identifying vital nodes on temporal networks: An edge based k-shell decomposition," in *2017 36th Chinese Control Conference*, pp. 1402-1407, 2017.

[20] D. Taylor, S.-A. Myers, A. Clauset, M.-A. Porter, P.-J. Mucha, "Eigenvector-based centrality measures for temporal networks," *Multiscale Modeling and Simulation*, vol. 15, no. 1, pp. 537-574, 2017.

[21] I. Tsalouchidou, R. Baeza-Yates, F. Bonchi, K. Liao, T. Sellis, "Temporal betweenness centrality in dynamic graphs," *International Journal of Data Science and Analytics*, vol. 9, no. 3, pp. 257-272, 2020.

[22] L. Lv, K. Zhang, T. Zhang, X. Li, W. Xue, "Eigenvector centrality measure based on node similarity for multilayer and temporal networks," *IEEE Access*, vol. 7, pp. 115725-115733, 2019.

[23] L. Lv, K. Zhang, T. Zhang, D. Bardou, J.-H. Zhang, C. Ying, "PageRank centrality for temporal networks," *Physics Letters A*, vol. 383, no. 12, pp. 1215-1222, 2019.

[24] L. Lv, K. Zhang, D. Bardou, X. Li, T. Zhang, W. Xue, "HITS centrality based on inter-layer similarity for multilayer temporal networks," *Neurocomputing*, vol. 423, pp. 220-235, 2021.

[25] M.-M. Tulu, R. Hou, T. Younas, "Vital nodes extracting method based on user's behavior in 5G mobile social networks," *Journal of Network and Computer Applications*, vol. 133, pp. 39-50, 2019.

[26] G. Kossinets, D.Watts, "Empirical analysis of an evolving social network," *Science*, vol. 311, no. 5757, pp. 88-90, 2006.

[27] L. Backstrom, D.Huttenlocher, J. Kleinberg, X. Lan, "Group formation in large social networks: membership, growth, and evolution," in *Proceedings of the 12th ACM SIGKDD international conference on Knowledge discovery and data mining*, vol. 311, no. 5757, pp. 44-54, 2006.

[28] B. George, S. Shekhar, "Time-aggregated graphs for modeling spatiotemporal networks," In *Journal on Data Semantics XI, Springer*, pp. 191- 212, 2008.

[29] R. Burt, "The social structure of competition," *Networks in the Knowledge Economy*, vol. 13, pp. 57-91, 2003.

[30] W.W. Zachary, "An information flow model for conflict and fission in small groups," *Journal of Anthropological Research*, vol. 33, pp. 452-473, 1977.

[31] R. Rossi, N. Ahmed, "The network data repository with interactive graph analytics and visualization," in *Twenty-Ninth AAAI Conference on Artificial Intelligence*, pp.4292-4293, 2015.

[32] A. Khalajmehrabadi, N. Gatsis, D. Akopian, "Modern WLAN finger- printing indoor positioning methods and deployment challenges," *IEEE Communications Surveys and Tutorials*, vol. 19, no. 3, pp. 1974-2002, 2017