Influence Maximization Based on Network Motifs in Mobile Social Networks

Xinxin Zhang, Li Xu, and Zhenyu Xu

***Abstract*—A mobile social network (MSN) is a mobile communications system that involves the social relationship of the users, in such a network, mobile users can spread information, opinions, ideas, rumors. Influence Maximization (IM) aims to identify k nodes from a network such that the influence spread generated by the k nodes is maximized, which has been attracting increasing attention in recent years. However, existing methods of influence maximization are heuristic algorithms based on network topology and greedy algorithms based on spreading. Accordingly, in this paper, we focused on Network Motifs (NM) as drivers of influence to impact the spreading process, we proposed IM-NM, a network motifs-based influence maximation scheme for delivering information efficiently. In consideration of the communication relationship and the users' attributes, we first defined Weight Ratio (WR), Degree Density (DD), and Structural Stability Level (SSL). Then we identified the key network motifs by Naive Bayesian machine learning. Finally, we adopted the k key network motifs as the unit structure to reconstruct the network, and select the bridge node with strong communication ability in the key motifs to maximize the information. We implement our proposed methods on a set of real-world networks to evaluate the performance, the experimental results demonstrate that our proposal achieves better performance than other related methods.**

***Index Terms*—Influence maximization, mobile social networks, network motifs, structural stability level.**

I. Introduction

INFLUENCE propagation is a fundamental process in network interaction. When an individual is affected, his behavior, attitude and thought will change in a network. The higher the influence of an individual, the more others in the network can be influenced by the individual. Influence analysis is generally accepted as one of the most significant problems for network mining because the influence degree of a user or a group reveals its impact on communication or information propagation in terms of scale, level, or other factors [1]. Influence analysis has become one of the most interesting research topics in network and graph mining, due to its wide applications in viral marketing [2], [3], recommender systems [4], [5], leaders opinion [6], [7], etc. Indeed, finding the optimal set of influencers is a many-body problem in which the topological interactions between them play a crucial role [8].

Complex network analysis has drawn substantial attention from researchers, influence analysis has many applications in various fields of science such as computational biology, biochemistry, sociology [9]. Nowadays, participating in multiple online social networks to share and communicate has become a new culture in modern life, with the proliferation of personal mobile devices and Internet technologies, *mobile social networks* (MSNs) have a major system for social interaction, message dissemination, and lifestyle share. A mobile social network is a user-driven mobile communication network that includes connections of nodes (or users), since the MSN is a combination of the mobile communication networks and social networks, it acquires and combines some properties of the two systems [10]. Therefore, the research of influence in MSNs is a challenging problem.

The influence of acquaintances' opinions is boosted by explosive exposure to mobile social networks. As a consequence, people in MSNs influence each other in both direct and indirect ways [11]. The top-k problem referred to as Influence Maximization (IM) through the powerful viral marketing, the company hopes that these users (top-k nodes) would like the application and to start influencing their friend to use it, and their friends would influence their friends' friends and so on, which is called the word-of-mouth phenomenon [9]. Thus, influence maximization aims to identify k nodes from a network such that the influence spread generated by the k nodes is maximized.

To the best of our knowledge, most methods of influence maximization are heuristic algorithms based on network topology and greedy algorithms based on spreading. The heuristic algorithm measures the importance of nodes in the network by considering the topology structure of the network and takes the nodes with high influence as top-k nodes. Chen *et al.* designed MIA model considers influence propagated through these local arborescences [12]. Compared with the shortest-path based heuristics, MIA algorithm is always among the best in influence spread. Rui *et al.* [13] proposed a novel influence maximization algorithm in social networks, named Reversed Node Ranking (RNR), and it outperforms the degree discount algorithm [14] under IC model. The greedy algorithm based on propagation pursues the optimal solution approximately through greedy strategy: it initially sets an empty set of

Manuscript received July 10, 2021; revised March 12, 2022; accepted March 26, 2022. Date of publication March 29, 2022; date of current version June 27, 2022. This work was supported by the National Natural Science Foundation of China under Grants U1905211, 61771140, and 62171132. Recommended for acceptance by Dr. Jiangchuan Liu. *(Corresponding author: Li Xu.)*

The authors are with the College of Computer and Cyber Security and Fujian Provincial Key Lab of Network Security and Cryptology, Fujian Normal University, Fuzhou 350007, China (e-mail: 1175045320@qq.com; xuli@fjnu.edu.cn; 1246233565@qq.com).

Digital Object Identifier 10.1109/TNSE.2022.3163203

influential nodes, and constantly adds the most influential nodes in the current network to the set of nodes.

In mobile social networks, the solution to the problem of user interactive communication can be to analyze the local network structure, also known as network motifs. Conventional non-motif-based only makes use of edge-based relations, which represent first-order relations between two nodes, it ignores higher-order relations that may exist between nodes. Motifs analysis stems from bioinformations and theoretical biology, we apply this biologically-inspired method to MSNs, the frequency of occurrence and the distribution of the individual motifs serve as the basis to define the interaction patterns between users. The network motifs are high-frequency low-order subgraphs, such subgraphs appear much more frequently in real networks than in random networks [15], they have been regarded as the basic structural unit in big networks. In MSNs, individuals inevitably exist in various subgraph structures, or are surrounded by internally connected structures because of the communication relationship of others, the network motifs can study the communication relationship between users through interconnection, and it plays key information processing role in mobile social networks. Motif analysis offers low computational overhead and oppotunity to gain an insight into the local structure of huge networks. We use *network motifs* to study the pattern of interactions among the nodes and analyze the 3-node motifs structural properties, we combined it with the attribute characteristics and communication characteristics of nodes in MSNs to study the problem of influence maximization.

These observations motivate us to investigate the IM-NM algorithm, a novel influence maximization scheme based on network motifs in mobile social networks. The aim is to identify the key network motifs with social attributes and enhances the information maximization efficiency.

The main contributions of our study are as follows:

- We develop a network motifs-based influence maximation (IM-NM) scheme, an efficient method from the new perspective of network motifs. We identify the key network motifs by Naive Bayesian machine learning and then adopt the key network motifs as the unit structure to reconstruct the network and maximize the information.
- We design a new indicator, namely, structural stability level (SSL) to classify network motifs. According to the communication relationship and the users' attributes, we define network motifs into four stability levels to improve the efficiency of information propogation.
- We implement our proposed methods with data extracted from real websites and compare them with the classic strategies. Our results show that the key network motifs identified by IM-NM scheme exhibit superior preference information maximization efficiency.

The rest of this paper is organized as follows. We briefly discuss the related work in Section II. We design the structural stability level and elaborate the IM-NM scheme in Section III. We analyze the experiment results in Section IV, and Section V provides the conclusion.

II. Related Work

A. Influence Maximization

Influence maximization [11] is first modeled as an algorithmic problem by Kempe, Kleinberg and Tardos in 2003, they use the Greedy algorithm (GA) to solve it and prove that the optimal solution for influence maximization can be approximated to within a factor of $(1 - \frac{1}{e} - \epsilon)$. There are several classical methods of Influence Maximization. The Degree Centrality (DC) [16] is a simple but inaccurate indicator, and it focuses on the number of links attached to the node. Betweenness Centrality (BC) [17] can identify the hub nodes in a network. Closeness Centrality (CC) [18] can get the best view of the information flow in a network. Based on these works, many scholars have done a lot of the state of the art technologies. Song *et al.* [19] proposed a totally different approach based on Simulated Annealing for the influence maximization problem, and they also proposed a divide-and-conquer method, it has been shown that a Greedy algorithm with provable approximation guarantees can obtain a good approximate result [20]. The goal of the IM problem is to find a k-sized set of users with the maximum influence in graph G. The influence of any seed set is defined based on the information diffusion process among the users. Based on Linear Threshold (LT) and Independent Cascade (IC) propagation models, Li, Wang and Liu [21] attempted to select ordinary grassroots as seeds and showed that grassroots are better choices than elites in the influence maximization problem from the aspects of relationship strengths and polarities. Wang *et al.* [22] proposed efficient distance-aware influence maximization in geo-social networks. Li *et al.* [23] survey and synthesize a wide spectrum of existing studies on IM from an algorithmic perspective. Yin *et al.* [24] designed a novel Signed-PageRank algorithm to achieve influence maximization in the signed social networks. Interpersonal communication plays a significant role in shopping public opinions of large populations. Xue *et al.* [25] investigated how the opinion-forming process evolves over social networks under the media influence. Recent studies show that individuals in a social network can be divided into different groups of densely connected communities, these individuals who bridge different communities are referred to as structural hole spanners, are crucial in many real applications. Xu *et al.* [26] proposed an efficient algorithms for the identification of top-k structural hole spanners. Tulu *et al.* [10] summed up many measures have been measure to identify influential nodes in complex networks.

While most of these studies only focused on individual nodes as a factor of influence in the information spreading process, they ignored the role of network motifs.

B. Network Motifs

Network motifs have been well studied in biological, sociological, physics, and computer science literature. Milo *et al.* [15], [27] found that most of the network structures contain many meaningful motifs and some kinds of motifs appear repeatedly all the time, which can help to reveal and understand the changes of network structure characteristics to a great extent. They defined a network motif as "*a pattern of interconnections occurring in complex networks at numbers that are significantly higher than those in randomized networks*". From this definition, a motif is a subgraph that is represented over a threshold with more than several times accidentally [28]. Conventional non-motif-based only makes use of edge-based relations, which represent first-order relations between two nodes, it ignores higher-order relations that may exist between nodes [29]. In [15], a motif having k nodes is called k-motif. Network motifs are subgraphs that appear significantly more often in a real network than in a randomized network with similar characteristics. Motifs may uncover the building blocks of most networks. Community structure is a network feature, the gathering of nodes into groups such that there is a higher density of edges with groups than between them. In mobile social networks, both network motifs and community structures are mesoscopic structures. The network motif can be regarded as a small community structure, and similar motifs may be found in the community structures. The smallest motifs contain two nodes, and 3-network motifs are the most frequent structures in undirected networks due to their stability. The definition is proposed purely based on the topologies of networks and neglected the attributes of different network motifs. In gene regulatory networks, specifically, the transcription network [27] illustrates the motifs with specific functionality in determining gene expressions. Besides the definitions of motifs mentioned above, Rane [30] proposed maximal motifs, which does not contain other subgraphs as motifs so that it can diminish the total number of discovered motifs.

In the motifs discovery process, some tools and platforms are effective to analyze and visualize the data. MFinder [31], FanMod [32], MODA [33], Kavosh [34], G-tries [35], QuateXelero [36] are all the most currently used methods for discovering network motifs and respectively present representative methods. Braines *et al.* [37] offer suggestions as to how network motif techniques can be applied to intra-group or inter-group behavior. Motif discovery in complex networks can be assisted by data mining methods such as network clustering, Joveini and Keong proposed [38] a novel structure so-called Augmented Multiresolution Networks, and provided an efficient solution for clustering and motif discovery. Underwood, Elliott, and Cucuringu [39] proposed motif-based spectral clustering of weighted directed networks. Sarkar, Guo, and Shakarian [40] use motif patterns to characterize the information diffusion process in social networks. In signed social networks, network motifs are a very useful tool. Liu, Xiao, and Xu [41] proposed a link prediction method based on the number of edge-dependent motifs, and explain it by a Naive Bayes model.

However, these researches focused on motifs from the perspective of topology. In MSNs, each node has many attribute characteristics, in this paper, we give network motifs to attribute features to study influence maximization. The Bayesian model can observe multiple features input by nodes, and obtain the corresponding network motifs of these features, this is an interesting topic.

TABLE I
THE ADOPTED NOTATIONS

Symbol	Description
$G(V,E,w)$	graph G with nodes set V, edges set E, and edge weight w
G'	generated graph
$\lvert N\rvert$	the size of nodes set
v_i	the lable of the i-th node
$\mathrm{M}i$	the lable of the i-th motif
n	the number of nodes in each motif
$WR_{(v_1,v_2)}$	the proportion of motifs communication frequency
$C_{(v_1,v_2)}$	the number of communications between two nodes
$C_{\max}$	the maximum communication times of any two nodes
DD_M	the degree density of network motifs
d_{v_i}	the degree of the node in a motif
d_{max}	the maxization degree of the graph
$d_{out}^{v_i}$	the out-degree of the node v_i

III. INFLUENCE MAXIMIZATION BASED ON NETWORK MOTIFS

Network motifs can be used to describe the basic connection patterns among members of a community. For some special networks, it can even be considered as an extension of many "network motifs". Generally speaking, the larger the network scale is, the more obvious the network motifs structure is. In this section, firstly, we designed stability levels for different network motif classes. Then we combined Weight Ratio and Degree Density, and we train the important and stable motif classes by Bayesian machine learning. After obtaining the key motifs, the network is reconstructed to maximize the influence of communication. To facilitate understanding of our work, all the notations are listed in Table I as follows.

A. Learning-Based Identification of the Network Motifs

Mapping a mobile social network into a graph directed weighted multigraph $G(V,E,w)$, where (1) V is a finite set of nodes, (2) $E \subseteq V \times V$ is a finite set of edges, in which $(v_1,v_2) \in E$ denote an undirected edge between v_1 and v_2, (3) $w_{(v_1,v_2)}$ is the weight that maps the communication frequency between users. Fig. 1 shows an example of digraph G from a real application, where nodes correspond to users of the network and edges correspond to the interaction between them. In this digraph, nodes with only in-degree play the role of "receiver," nodes with only out-degree play the role of "initiator," the nodes with both in-degree and out-degree play these two roles at the same time. The weight of the edge represents the communication frequency. For example, $v_1$and v_2 communicate with each other, but the frequency of sending and receiving information is different. The send frequency of v_1 to v_2 is 0.8 and the send frequency of v_1 to v_3 is 0.9. v_1, v_3, v_4 will send information to v_2, among these, the one with the lowest send frequency is v_4. We define several metrics to select key network motifs as follows:

1) Weight Ratio: In the MSN, because users have different attitudes towards information transmission, the communication relationship and frequency between users are not the same. In this work, we prefer to find the users with high

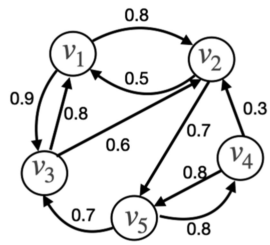


Fig. 1. Example of an interaction graph.

communication frequency to transmit information, thus we denote *Weight Ratio* (WR) to measure the proportion of network motifs communication frequency in the whole network, denote by $WR_{(v_1,v_2)}$ the Weight Ratio is:

$$WR_{(v_1,v_2)} = \frac{C_{(v_1,v_2)}}{C_{\max}}$$

where $C_{(v_1,v_2)}$ is the number of communications between two nodes v_1, v_2, $C_{\max}$ is the maximum communication times of any two nodes in the network. The larger $WR_{(v_1,v_2)}$ is, the more times v_1 and v_2 interact with each other.

2) Degree Density: In the MSN, the more friends a user has, the more likely he is to influence others. Similarly, the nodes in the network motifs which have a large degree mean that they are closely related to the nodes outside the motif. We denote *Degree Density* (DD) to measure the importance of network motifs to other nodes in the network, the more the nodes in the network motifs connect to the outside, the greater the influence on the other nodes. What's more, the more outdegree, the more users will receive information, and the role of the "initiator" is even more important. Denote by DD_M the Degree Density of network motifs is:

$$DD_M = \frac{\sum_{i=1}^{n} d_{v_i}}{n \times d_{max}},$$

where d_{v_i} is the degree of the node in a motif, d_{max} is the maximization degree of the graph, n is the number of nodes in each network motif.

3) Design of Motifs' Structural Stability Level (SSL): As we know the triangle is the most stable in geometry. In this study, we use 3-node network motifs (triangle network motifs) as the minimum structure units to study the MSN. According to the definition of network motifs, Z-score is applied in this situation. We give a threshold α to restrict and evaluate the value of Z-score and enumerate all the network motifs of G in the randomized networks when $\alpha > 0$. The definition of Z-score is described as follows:

$$Z - score = \frac{f_{ori} - f_{rand}}{std(f_{rand})} = \alpha,$$

Herein, f_{ori} is the frequency of a given motif in G, while f_{rand} is the frequency in a randomized network, and the $std(f_{rand})$ is the standard deviation of motif frequency in different randomized networks.

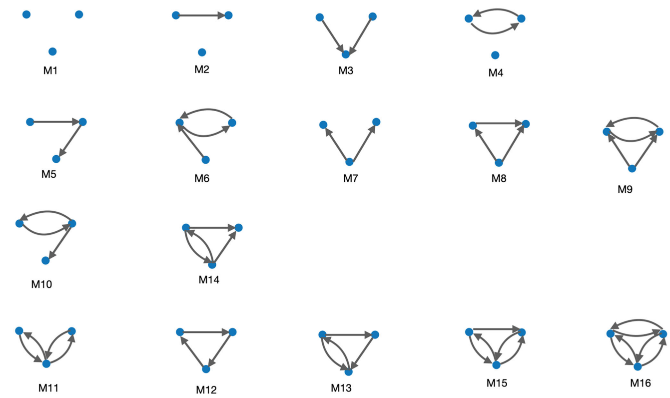


Fig. 2. The displays of triangle network motifs.

Then, as shown in Fig. 2, we define the Structural Stability Level (SSL), there are 16 kinds of combinations of 3-node, M1 to M16, respectively. According to the communication relationship, we divide these network motifs into four levels. The communication at the bottom layer (M11 to M16) is the most stable, and the information of any node can be sent to the other two nodes in various ways. The second layer (M10 and M14) is less stable than the bottom layer since one node can't send information to the other nodes. The third layer (M5 to M9) is a little stable since only one node can send information to the other nodes. The fourth layer (M1 to M4) is the most unstable, because the information cannot be delivered to all nodes in this kind of motifs.

According to the user attributes and the communication relationship established by the users in the mobile social network, we choose M16 as the standard ideal network motif. Since all its nodes in M16 not only communicate with each other but also have the same indegree and outdegree, we denote the SSL of M16 is 1. Since the four motif structures in the top layer can not undertake the task of communicating with each other, we define their SSL as 0. The remaining motifs are scored as follows according to their communication ability:we denote the SSL of network motifs in the third layer as $\frac{1}{4}$ since only one node in these motifs can transmit the received information to the other two nodes. We denoted the SSL of network motifs in the second layer as $\frac{1}{2}$ since there are two nodes in these motifs that can transmit information. In the bottom layer, all the network motifs are communicated with each other. Considering the effectiveness and efficiency of communication, we denote the SSL of M11 as $\frac{3}{5}$ since although any of its three nodes can transfer information to the other two, it is not connected. The M12 is a circle with a single arrow, we denote the SSL of it as $\frac{7}{10}$. M13 and M15 have one and two more reverse arrows than M12, thus, we denote them $\frac{4}{5}$, and $\frac{9}{10}$ respectively. To facilitate understanding, the SSL of network motifs are listed in Table II.

Several network motifs with the stable structures are selected by combining SSL and Z-score.

4) Identification of Key Network Motifs by Naive Bayesian Machine Learning: In statistical machine learning projects, we often have a collection of motifs that may consider for a particular learning task. The three kinds of motifs with the

TABLE II
THE SSL OF NETWORK MOTIFS

Mi	SSL
M1, M2, M3, M4	0
M5, M6, M7, M8, M9	$\frac{1}{4}$
M10, M14	$\frac{1}{2}$
M11	$\frac{3}{5}$
M12	$\frac{7}{10}$
M13	$\frac{4}{5}$
M15	$\frac{9}{10}$
M16	1

Algorithm 1: Identification of the Key Network Motifs.

Require: The three largest sets of motifs M_i, the number of key motifs t
Ensure: The key network motifs set M_k
for the network motifs Mi **do**
  $X_i^1 = WR_{(v_1,v_2)} \times DD_M, X_i^2 = SSL$
  $P_j(Y=1|X=X_j) = \frac{P_j(Y=1)}{P_j(X=X_j)} \prod_{i=1}^{2} P(X^j = X_i^j | Y=1)$
 end for
 for $j = 1 \rightarrow t$ **do**
  When the value of P_{M_i} is biggest, and the M_i join the key network motifs set M_k
end for
return The key network motifs set M_k

largest number in the network are counted as the initial motifs of Bayesian training. The problem of selecting the best motifs in this collection is called motifs selection and is typically based on some theoretical metric. We randomly selected 5% of the different samples extracted from different networks in the data set to construct the training sets. Then, according to SSL, the motifs with good communication relationships and structural stability are identified as key motifs, which are set as label 1, and the labels of other motifs are set as 0. These key motifs are regarded as ground truth.

We define $X_i = (X_i^1, X_i^2)$ as a vector of social characteristics for M_i, X_i^1 represents the Weight Ratio and Degree Density, X_i^2 represents the Structural Stability Level. We use a Naive Bayes model to train classification and identify the key network motifs. A classifier maps input feature vector X to output the probability of labels $Y \in \{0, 1\}$ of the network motifs. We employ Naive Bayesian inference to perform our selection task. And this method can also be essentially described as follows:

$$P_{M_i}(Y=1|X=X_i) = \frac{P(Y=1)}{P(X=X_i)} \prod_{j=1}^{2} P(X^j = X_i^j | Y=1)$$

In Algorithm 1, we try to identify the key network motifs. The model can output the importance of motifs. A model can be trained from training data, which should contain the groundtruth of labels. The algorithm cycles the input all the network motifs data, and computes the social characteristics of the network motifs, determine the key motifs according to the P_{M_i}.

Algorithm 2: Reconstruct the Network Based on the Key Network Motifs.

Require: A original graph $G = (V, E, w)$, key network motifs M_k
Ensure: Generating graph $G' = (V', E', w')$ with the new structure
for Any key motif we've selected M_k **do**
  if There is only one node can send information to the other two nodes in the network motif, such as M5, M6, M7, M8, M9 **then**
   Select this node as the bridge node to establish the connection with other network motifs
  else
   if There are two nodes or three nodes can send information to the other two nodes in the network motif, such as M10, M11, M12, M13, M14, M15, M16 **then**
    Calculate the propagation capacity of these communicable nodes: $ComC = \sum_{j=1}^{n} d_{out}^{v_i} \times W_{(v_i,v_j)}$, selecte the bridge node with high communication capacity and establish the connection
   end if
  end if
 end for
return Generated graph G'

B. Communication Influence of Reconstructed Network

In the last section, we selected the key network motifs, which have a stable structure, frequent communication, many connections with peripheral nodes. In this section, we will reconstruct the local network based on this kind of network motifs. To minimize the damage to the structure of the network, we choose the reconstruction nodes of each network as the bridge nodes to connect the network motifs. We emphasize that the information loss degree plays a very important role in the stability of the whole network, especially at small scales where it percolates to larger scales of interconnection. In MSNs, due to the different attributes of individuals, each node in the same network motif has a different ability to send information, we define the *Communication Capacity* (ComC) to measure the information loss degree as follows:

$$ComC = \sum_{j=1}^{n} d_{out}^{v_i} \times W_{(v_i,v_j)},$$

where $d_{out}^{v_i}$ is the out-degree of the node v_i, $W_{(v_i,v_j)}$ is Weight Ratio to measure the communication frequency between two nodes v_i, v_j.

The interconnection of "motifs as motifs" arises as a network motifs design procedure in which the contraction loss remains minimal at each step of construction of the network. We hope to get the network structure that can maximize the dissemination of information with the minimum loss of information, we select the k nodes with strong communication ability as the bridge nodes in the key motifs to propagate information. If there is only one node can send information to the other two nodes in the key network motif, select this node as the bridge node to establish the connection with other network motifs. If there are two nodes or three nodes that can

TABLE III
STATISTICS OF THE DATASETS USED FOR EVALUATION

Networks	Users	Relations	Weight	Anchor motifs
Freeman's EIES Network	32	440	√	9
Facebook-like Forum Network	899	7089	√	15
Facebook-like Social Network	1899	13838	√	30
Enron Network	6187	50897	√	40

send information to the others in the network motif, calculate the propagation capacity of these communicable nodes, select the bridge node with high communication capacity, and establish the connection. Based on the Section III-A, we designed the following algorithm to reconstruct the network:

IV. Experimental Evaluation

This section is devoted to the assessment of the performance of our scheme. The basics of a pure experiment, comprising the introduction of a dataset and the evaluation criteria, are described in Section IV-A. In Section IV-B we show the case study in a real dataset, and then we evaluate the importance of the key motifs by *Motif Abundance* in Section IV-C. Finally, we perform the information maximization efficiency in Section IV-D.

A. Experimental Settings

Our experimental environment is python 3.8+anaconda, and the software is pycharm. We test all the algorithms on a real-world dataset (https://toreopsahl.com/datasets) that is openly available.

1) Freeman's EIES Network: This dataset is a social network with three networks of researchers working on social network analysis. Node attributes are constructed by a matrix with the number of messages sent among 32 researchers, the frequency matrix is an electronic communication tool.

2) Facebook-like Forum Network: We use this dataset with 899 users and 522 topics which a weight can be assigned to the ties based on the number of messages or characters that a user posted to a topic. The focus in this network is not on the private messages exchanged among users, but on users' activity in the forum. The number of users in this network is smaller than in the online social network as all users that sent or received private messages did not participate in the forum.

3) Facebook-like Social Network: This is the larger of a well-known social network with 1899 nodes with many attributes, which originates from an online community for students at the University of California. A total number of 59835 online messages were set over 20296 directed ties among these users. Although this dataset contains many nodal attributes, these are not made available as it would be possible to reverse engineer the anonymization procedure of users.

4) Enron Email Network: The dataset to build a network of email addresses, it contains 614586 emails sent over the period from 6 January 1998 until 4 February 2004. The emails exchang between 6600 nodes, it has 50897 edges. The node represents the email address, edge means that there is communication between two email addresses, edge weigh is the number of email exchanges.

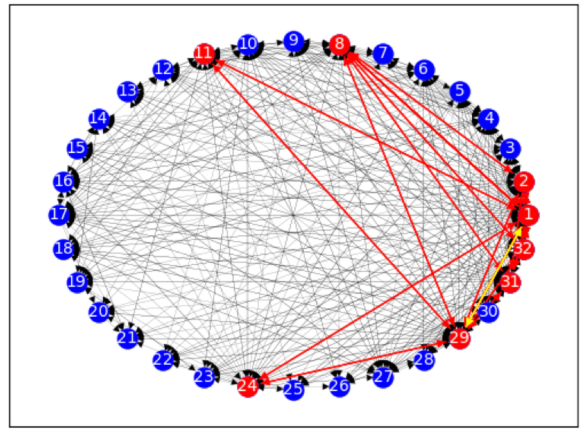


Fig. 3. Training graph of the Freeman's EIES Network.

In summary, the sizes of these datasets are 32, 899, 1899, 6600 respectively. Accordingly, the complexity of connections between users is different. The attribute characteristics of these three datasets are shown in Table III, each dataset is a directed weighted graph, anchor motifs are the key network motifs by Naive Bayesian machine learning in Algorithm 2, according to the size of the data set, we train 9, 15, 30 and 40 anchor motifs in these four datasets respectively.

Concerning an anchor motif M_i, we measure its abundance in a network G by Motif Abundance [43] which is defined as follows, given a key motif M_i, the motif abundance of M_i in a network G is defined as the total number of occurrences of the motif in G. It can be used to describe the importance of the motif to the whole network.

B. Case Study

After running our model on the four datasets, we verify the feasibility of the key network motifs identification scheme and the efficiency of information spreading. Since the number of network motifs in the fourth dataset is too large to be displayed, we only show the results of the first three datasets to illustrate the physical meaning of the identification of the key network motifs and the reconfiguration of the network. We also compare the computational cost of motif discovery for different algorithms on the four networks.

1) Implementation: As Fig. 3 shows, in the Freeman's EIES Network with 32 nodes, there are 2890 triangular motifs in total. Through the training of Bayesian model Algorithm 1, we identify the 9 key motifs (anchor motifs) are shown in the red nodes of the graph, for clarity, we list the 9 key network motifs in Fig. 4. Due to the small scale of the network (32-node network) and the limitation of node attributes, the structural stability level is the most important factor in the selection of key motifs. The 9 key network motifs are all type-M16 since the score of M16 is higher than others. Thus, we take M16 as the basic unit structure to reconstruct the network to

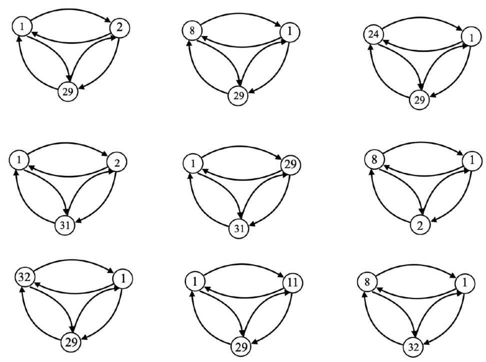


Fig. 4. The 9 key network mitifs of the Freeman's EIES Network.

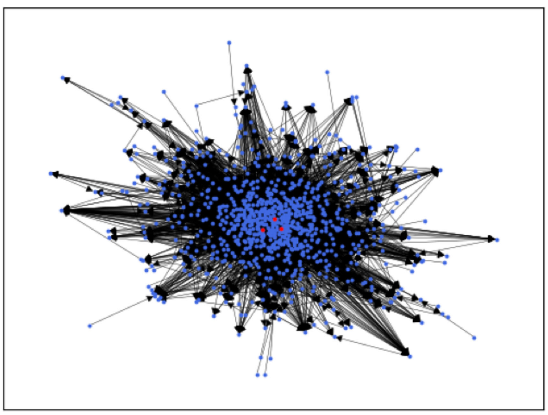


Fig. 5. Training graph of the Facebook-like Forum Network.

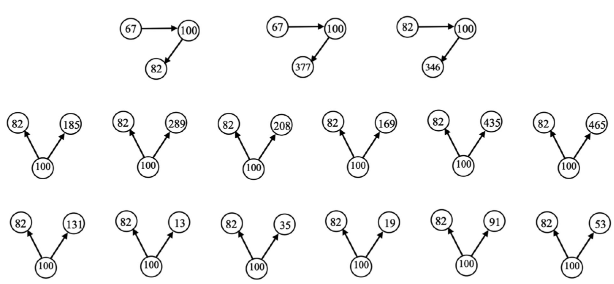


Fig. 6. The 15 key network mitifs of the Facebook-like Forum Network.

better send information. And then, according to Algorithm 2, node 1 and node 29 (the yellow edge) are the bridge nodes to connect the different motifs.

The Facebook-like Forum Network with 899 nodes (in Fig. 5), there are 199466 triangular motifs in total. After the training of the Naive Bayesian model according to Algorithm 1, we identify 15 key motifs. Due to the large scale of the network, we did not mark the key motifs one by one in the original graph, to show the results more clearly, we list the 15 key network motifs in Fig. 6. We can see that the training results of this network are two type motifs, M5 and M7, the influence of Weight Ratio and Degree Density on key motifs is higher than that on structural stability, in other words, the communication frequency and connectivity of the motif nodes in this network have better effect. This reflects the importance of communication relationships and intimate relationships to

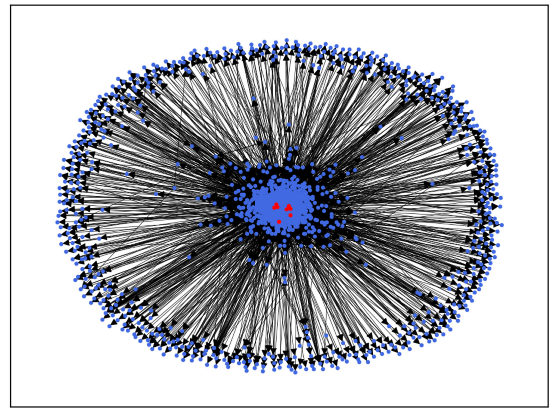


Fig. 7. Training graph of the Facebook-like Social Network.

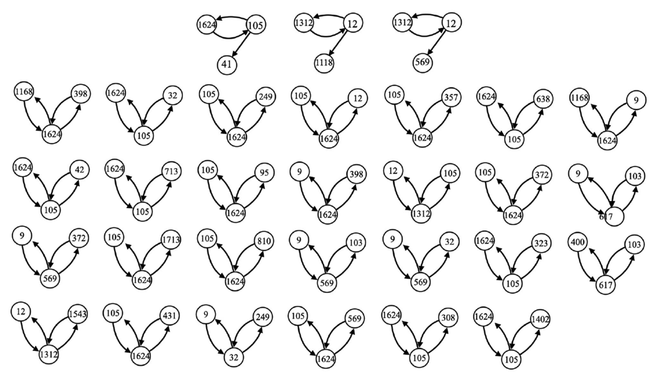


Fig. 8. The 30 key network mitifs of the Facebook-like Social Network.

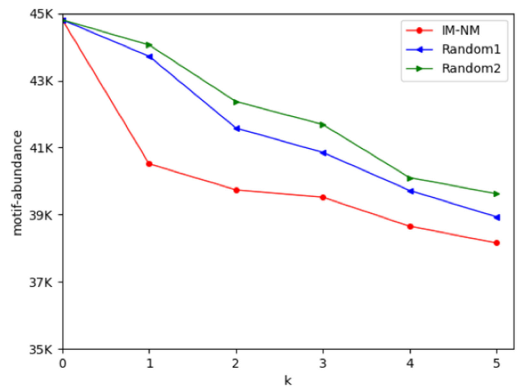
information dissemination. Then, according to Algorithm 2, node 167, node 100, and node 346 (red nodes in Fig. 5) are the bridge nodes to connect the different motifs. We take M5 and M7 as the basic unit structure to reconstruct the network to better send information similar to Fig. 3.

As Fig. 7 shows, in the Facebook-like Social Network with 1899 nodes, there are 727244 triangular motifs in total. Training of Bayesian model Algorithm 1, we identify the 30 key network motifs. Due to the large scale of the network, we did not mark the key motifs one by one in the original graph, for a better explanation, we list the 30 key network motifs in Fig. 8. This figure clearly shows that the training results of this network are two type motifs, M10 and M11, compared with the network structure of the previous two datasets, the importance of network structural stability, weight ratio, and degree density to the selection of key network motifs is equal. That is the communication frequency, connectivity, and relationship stability of the motif nodes play the same role in this network. According to Algorithm 2, node 105, node 12, node 1624, node 398, node 638, node 9, node 323, and node 617 (red nodes in Fig. 5) are the bridge nodes to connect the different motifs. And we take M10 and M11 as the basic unit structure to reconstruct the network to better send information like the connection in Fig. 3.

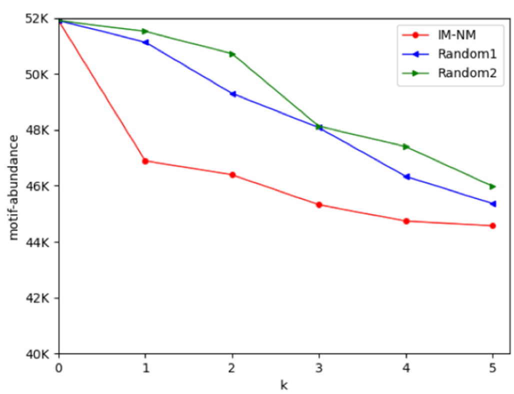
2) Computational Cost: Table IV is the computational cost of motif discovery for different algorithms on different networks, the rows indicated different networks, columns are related to different algorithms, the execution time rated in

TABLE IV
COMPUTATIONAL COST FOR DIFFERENT ALGORITHMS ON DIFFERENT NETWORKS

	Freeman's EIES networks	Forum network	Social network	Enron Network
FANMOD	0.0005	0.0221	0.1200	0.5585
IM-NM	1.604	2.104	3.121	7.569
Mfinder	7.0	12.7	-	-
MAVisto	156	13532.0	-	-



(a) M5-motif



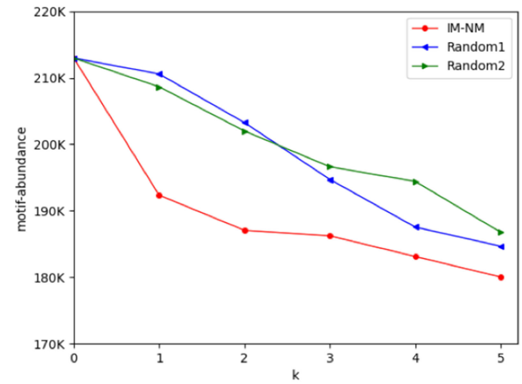
(b) M7-motif

Fig. 9. The change motif abundance after removing M5-motif and M7-motif in the Facebook-like Forum network.

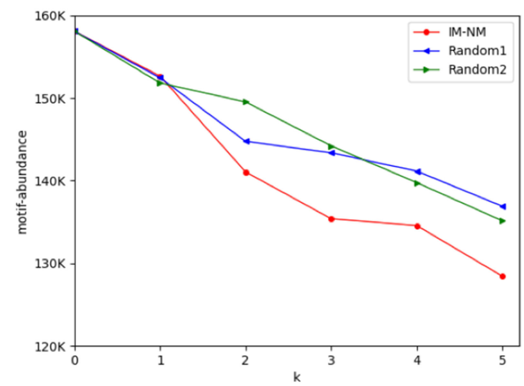
seconds. It can be seen that the maximum time is MAVisto for all works, and the time of Mfinder is larger than that of FANMOD and IM-NM (ours). These existing motif discovery algorithms focused on motifs from the perspective of topology, while the attributes of motifs in the network are input into the Bayesian model, the evaluation of the importance of each motif is output, thus our algorithm computational cost has a few more seconds than FANMOD, it is reasonable.

C. Motif Abundance

After identifying the anchor motifs, we now exam the change of motif abundance after removing the top-k motifs from the corresponding networks to measure the importance of the anchor network motifs. It clear that these networks have different sets of key motifs, which may have different information propagation functionalities, and the key network



(a) M10-motif



(b) M11-motif

Fig. 10. The change motif abundance after removing M10-motif and M11-motif in the Facebook-like Social network.

motifs we selected have stronger communication capability since we consider the communication frequency ratio in Algorithm 2. Fig. 9 and Fig. 10 show the effects of removing key motifs (with communication frequency ratio) and randomly removing motifs (without communication frequency ratio) in different networks on the network structure, the x-axis represents the growth of k in top-k, and the red curve shows the removal of key motifs by IN-NM scheme (we proposed), the blue and green curves are the trend of network abundance with randomly removed motifs, respectively.

In the Facebook-like Forum network, the key notifs include M5, M7. Fig. 9 demonstrates the change of M5 and M7 motif abundance, respectively. Among them, (a) is the variation of motif abundance between removing the top-k M5-key motifs and randomly removing other motifs. We can see that the abundance of motifs decreased significantly after removing the top-5 M5-key motifs, the decrease is much larger than that of random removal. Moreover, when $k = 1$, the decrease is most obvious, this shows that M5 is the key motifs, and its removal is very destructive to the network information structure. Similarly, in comparison of motifs abundance by the top-10 M7-key motifs removal and other random motifs removal in (b), M7-key motifs play a key role in maintaining the stable transmission of network information.

With the increase of network scale, the number of key motifs increases, and the removal of key motifs are more destructive

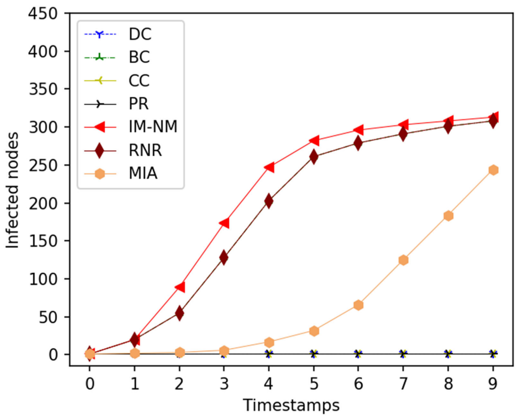


Fig. 11. The information maximization efficiency in Facebook-like Forum Network under SI model.

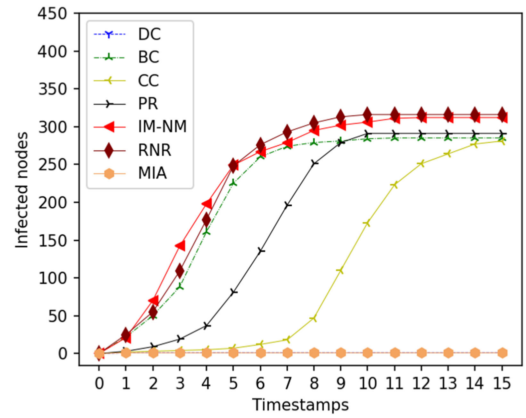


Fig. 12. The information maximization efficiency in Facebook-like Forum Network under IC model.

to the network structure. In the Facebook-like Social network, the key motifs include M10, M11. Fig. 10 is the change of M10 and M11 motif abundance, respectively. Both of the two subfigures show that the abundance decreases significantly with the increase of k value. In (a), after removing the top-5 M10-key motifs, the abundance decreased from 212972 to 169123, this difference is larger than the other two methods of randomly removing motifs. The abundance decreased from158064 to 115493 after removing the top-5 M11-key motifs in (b), it is also the most obvious decline in the three curves, at this time, the network damage rate is as high as 26.93%.

It can be observed that as more key motifs are removed from each network, the value of motif abundance decreases. Consequently, the information propagation functionality of the network will be destroyed. This experiment shows that the key motifs with the communication frequency ratio play a more important role in mobile social networks.

D. Information Maximization Efficiency

The empirical results related to the influence spread on real social networks are illustrated in Figs. 11–16 to show the effectiveness of the algorithms. We compared the influence spread of IM-NM with MIA model [12]), Reversed Node Ranking (RNR) [13], Degree Centrality (DC) [16], Betweenness Centrality (BC) [17], PageRank (PR) [24], Closeness Centrality (CC) [18]. We conducted experiments on several real-world different features under SI model and IC model. SI model [42] is a classical infectious disease transmission model, which is used to simulate the transmission of diseases that cannot be cured after infection in the network. In the model, S (Susceptible) is a susceptible person, refers to a healthy person who lacks immunocompetence and is susceptible to infection after contact with an infected person. I (Infectious) is a diseased person, refers to a patient who is infectious and can transmit to S, and the probability of node changing from S to I is α. In independent cascade (IC) model [12], each edge (u, v) in the graph is associated with a propagation probability $pp(u, v)$, which is the probability that node u independently activates (influences) node v at step $t + 1$ if u is activated at step $t + 1$. The parameter α is the probability of

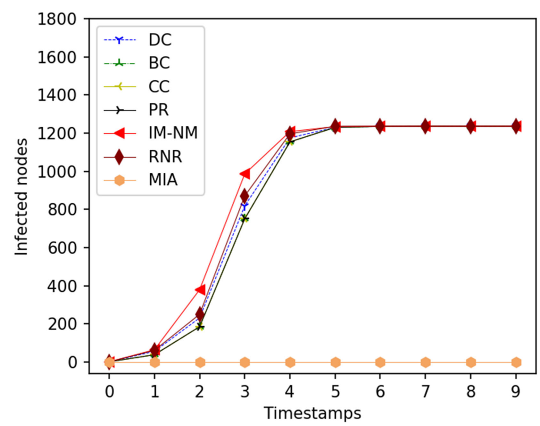


Fig. 13. The information maximization efficiency in Facebook-like Social Network under SI model.

users changing from Susceptible to Infectious. Since the first targeted node, if chosen somewhat judiciously, will activate almost 25% fraction of the social network [11]. We set propagation probability to 0.25 in the implementation of the information maximization experiment in MSNs.

The information maximization efficiency of Facebook-like Forum Network under the SI model shows in Fig. 11, the infected nodes stand for the nodes who receive the information, the curve shows the change of the number of receiving information nodes with time. We compare the efficiency of information maximization between our method (IM-NM) and the other six methods. We can see that the number of people who finally receive information of ours is consistent with RNR, which is the largest, and our advantage in propagation speed is better than the RNR algorithm. Fig. 12 is the information maximization efficiency of Facebook-like Forum Network under IC model, the number of infected nodes of IM-NM is more than most algorithms, and there is little difference from the RNR algorithm, the propagation speed of IM-NM is the fastest before the number of nodes receiving information is stable.

As we can see from Fig. 13 and Fig. 14, the information maximization efficiency of IM-NM under SI model and IC model in Facebook-like Social Network. According to the

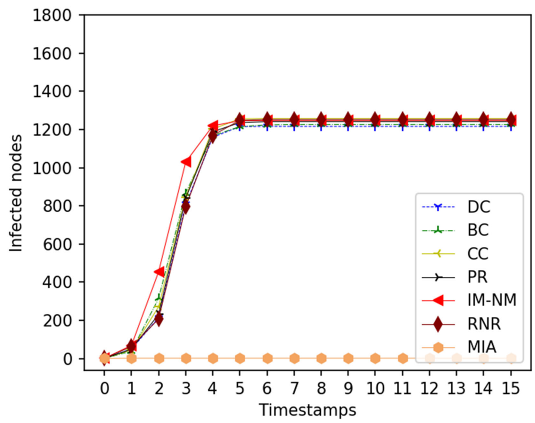


Fig. 14. The information maximization efficiency in Facebook-like Social Network under IC model.

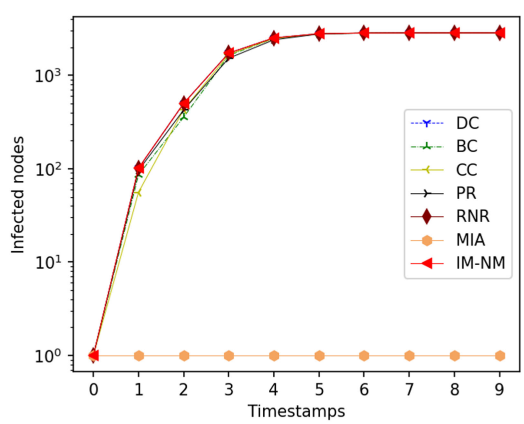


Fig. 15. The information maximization efficiency in Enron Network under SI model.

trend of the curve, we can see that the information maximization efficiency of IM-NM under the SI model has obvious advantages compared with MIA algorithms, the number of people who finally receive information of ours is consistent with other five algorithms, and our advantage in propagation speed is better than these algorithms. The curve trend of each algorithm in Fig. 14 is the same as that in Fig. 13.

Fig. 15 and Fig. 16 show the comparisons of the information maximization efficiency of the IM-NM algorithm with six algorithms under SI model and IC model in Enron Network. It can be seen that under the SI model, the information maximization efficiency of the IM-NM algorithm is better than that of MIA, BC, and CC algorithms, IN-NM algorithm almost overlaps with the other three. Under the IC model, IM-NM has the fastest propagation speed before the number of nodes receiving information is stable and almost overlaps with the RNR algorithm.The experimental results confirmed the reduction in the running time of the IM-NM algorithm while improving the influence spread. Consequently, IM-NM performs better on mobile social networks which has communication relationship and users' attributes.

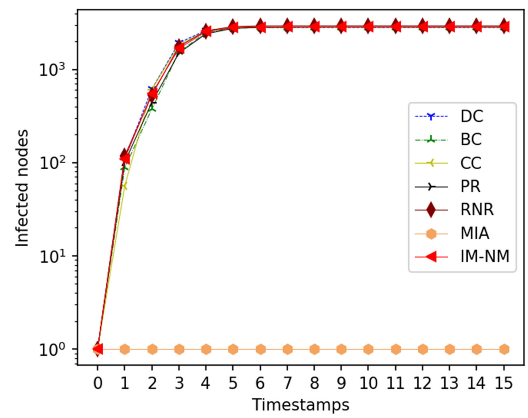


Fig. 16. The information maximization efficiency in Enron Network under IC model.

V. Conclusion

Influence maximization has become an important approach for studying mobile social networks, in particular, for viral marketing, recommender systems, leaders' opinions. Different from a lot of previous researches which focused on individual nodes as a factor of influence in the information diffusion process, we propose an influence maximization based on network motifs in this paper. The Naive Bayesian machine learning is employed to identify the key network motifs. Then we adopted the key network motifs as the unit structure to reconstruct the network and maximize the information. We compared the proposal with the classic strategies. The results show that IM-NM achieves better information maximization efficiency.

With the proliferation of personal mobile devices and Internet technologies, MSNs have a major platform for social interaction, message dissemination, and lifestyle share. It is important to develop an information controllable communication scheme. In our future work, we will integrate more network information, and establishing an information dissemination model in mobile social networks.

References

[1] B. Zhang, L. Zhang, C. Mu, Q. Zhao, Q. Song, and X. Hong, "A most influential node group discovery method for influence maximization in social networks: A trust-based perspective," *Data Knowl. Eng.*, vol. 121, pp. 71–87, 2019.

[2] Z. Yu, D. Zhang, Z. Yu, and D. Yang, "Participant selection for offline event marketing leveraging location-based social networks," *IEEE Trans. Syst., Man, Cybern. Syst.*, vol. 45 no. 6, pp. 853–864. Jun. 2015.

[3] C. Long, R.C.-W. Wong, and V. J. Wei, "Profit maximization with sufficient customer satisfactions," *ACM Trans. Knowl. Discov. Data*, vol. 12, no. 2, pp. 19:1–19:34, 2018.

[4] Y. Pan, F. Cong, K. Chen, and Y. Yu, "Diffusion-aware personalized social update recommendation," in *Proc. 7th ACM Conf. Recommender Syst.*, 2013, pp 69–76.

[5] I. A. Christensen, and S. Schiaffino, "Social influence in group recommender systems," *Online Inf. Rev*., vol. 38, no. 4, pp. 524–542. 2014,

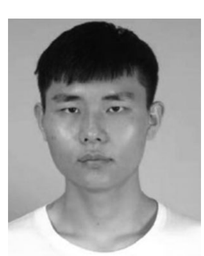
[6] A. Goyal, F. Bonchi, and L.V.-S. Lakshmanan, "Discovering leaders from community actions," in *Proc. 17th ACM Conf. Inf. Knowl. Manage.*, 2008, pp. 499–508.

[7] D. Saez-Trumper, G. Comarela, V. Almeida, R. Baeza-Yates, and F. Benevenuto, "Finding trendsetters in information networks," in *Proc. 18th ACM SIGKDD Int. Conf. Knowl. Discov. Data Mining*, 2012, pp. 1014–1022.

[8] M. Flaviano and H. A. Makse, "Influence maximization in complex networks through optimal percolation," *Nature*, vol. 524, no. 7563, pp. 65–68, 2015, doi: 10.1038/nature14604.
[9] D. Liu, Y. Jing, J. Zhao, W. Wang, and G. Song, "A fast and efficient algorithm for mining top-k nodes in complex networks," *Sci. Rep.*, vol. 7, no. 1, pp. 1–8, 2017, doi: 10.1038/srep43330.
[10] M.-M. Tulu, M.-E. Mkiramweni, R. Hou, S. Feisso, and T. Younas, "Influential nodes selection to enhance data dissemination in mobile social networks: A survey," *J. Netw. Comput. Appl.*, vol. 169, 2020, Art. no. 102768, doi: 10.1016/j.jnca.2020.102768.
[11] D. Kempe, J. Kleinberg, and E. Tardos, "Maximizating the spread of influence through a social network," in *Proc. 9th ACM SIGKDD Int. Conf. Knowl. Discov. Data Mining*, 2003, pp. 137–146.
[12] W. Chen, C. Wang, and Y. Wang, "Scalable influence maximization for prevalent viral marketing in large-scale social networks," in *Proc. 16th ACM SIGKDD Int. Conf. Knowl. Discov. Data Mining*, 2010, pp. 1029–1038.
[13] X. Rui, F. Meng, Z. Wang, G. Yuan, "A reversed node ranking approach for influence maximization in social networks," *Appl. Intell.*, vol. 49, no. 7, pp. 2684–2698, 2019.
[14] W. Chen, Y. Wang, and S. Yang, "Efficient influence maximization in social networks," in *Proc. 15th ACM SIGKDD Int. Conf. Knowl. Discov. Data Mining*, 2009, pp. 199–208.
[15] R. Milo, S. Shen-Orr, S. Itzkovitz, N. Kashtan, D. Chklovskii, and U. Alon, "Network motifs: Simple building blocks of complex networks," *Science*, vol. 298 no. 5594, pp. 824–827, 2002.
[16] L.-C. Freeman, "Centrality in social networks conceptual clarification," *Social Netw.*, vol. 79, no. 1, pp. 215–239, 1978.
[17] K. Nakajima, K. Iwasaki, T. Matsumura, and K. Shudo, "Estimating top-k betweenness centrality nodes in online social networks," in *Proc. IEEE Int. Conf. Parallel Distrib. Process. Appl., Ubiquitous Comput., Commun., Big Data, Cloud Comput., Social Comput., Netw., Sustainable Comput., Commun. (ISPA/IUCC/BDCloud/SocialCom/SustainCom)* 2018, pp. 1128–1135.
[18] P.-H. Du, H.-C. Nguyen, K.-K. Nguyen, and N.-H. Nguyen, "An efficient parallel algorithm for computing the closeness centrality in social networks," in *Proc. 9th Int. Symp. Inf. Commun. Technol.*, 2018, pp. 456–462.
[19] Q. Jiang, G. Song, C. Gao, Y. Wang, W. Si, and K. Xie, "Simulated annealing based influence maximization in social networks," in *Proc. 25th AAAI Conf. Artif. Intell.*. 2011.
[20] G. Song, X. Zhou, Y. Wang, and K. Xie, "Influence maximization on large-scale mobile social network: A divide-and-conquer method," *IEEE Trans. Parallel Distrib. Syst.*, vol. 26, no. 5, pp. 1379–1392, May 2015.
[21] D. Li, W. Wang, and J. M. Liu, "Grassroots VS elites: Which ones are betwwer candidates for influence maximization in social networks?," *Neurocomputing*, vol. 358, no. 17, pp. 321–331, 2019.
[22] X. Wang, Y. Zhang, W. Zhang, and X. Lin, "Efficient distance-aware influence maximization in Geo-social networks," *IEEE Trans. Knowl. Data Eng.*, vol. 29, no. 3, pp. 599–612, Mar. 2017.
[23] Y. Li, J. Fan, Y. Wang, and K.-L. Tan, "Influence maximization on social graphs: A survey," *IEEE Trans. Knowl. Data Eng.*, vol. 30, no. 10, pp. 1852–1872, Oct. 2018.
[24] X. Yin, X. Hu, Y. Chen, X. Yuan, and B. Li, "Signed-PageRank: An efficient influence maximization framework for signed social networks," *IEEE Trans. Knowl. Data Eng.*, vol. 33, no. 5, pp. 2208–2222, May 2021.
[25] D. Xue, S. Hirche, and M. Cao, "Opinion behavior analysis in social networks under the influence of coopetitive media," *IEEE Trans. Netw. Sci. Eng.*, vol. 7, no. 3, pp. 961–974, Mar. 2020.
[26] W. Xu, M. Rezvani, W. Liang, J.-X. Yu, and C. Liu, "Efficient algorithms for the identification of top-k structural hole spanners in large social networks," *IEEE Trans. Knowl. Data Eng.*, vol. 29, no. 5, pp. 1017–1030, May 2017.
[27] S. S. Shen-Orr, R. Milo, S. Mangan, and U. Alon, "Network motifs in the transcriptional regulation network of escherichia coli," *Nature Genet.*, vol. 31, no. 1, pp. 64–68, Jan. 2002.
[28] S. Yu, Y. Feng, D. Zhang, H.-D. Bedru, B. Xu, and F. Xia, "Motif discovery in networks: A survey," *Comput. Sci. Rev.*, vol. 37, 2020, Art. no. 100267, doi: 10.1016/j.cosrev.2020.100267.
[29] H. Zhao, X. Xu, Y. Song, D.-L. Lee, Z. Chen, and H. Gao, "Ranking users in social networks with motif-based PageRank," *IEEE Trans. Knowl. Data Eng.*, vol. 33, no. 5, pp. 2179–2192, May 2021.
[30] L. Parida, "Discovering topological motifs using a compact notation," *J. Comput. Biol.*, vol. 14, no. 3, pp. 300–323, 2007.
[31] R. Rane, "Frequent subpatterns distribution in social network analysis" *Emerging Technologies in Data Mining and Information Security*, Singapore: Springer2019, pp. 393–403.
[32] Y. Chen and Y. Chen, "An efficient sampling algorithm for network motif detection" *J. Comput. Graph. Statist.*, vol. 27, no. 3, pp. 503–515, 2018.
[33] S. Omidi, F. Schreiber, and A. Masoudi-Nejad, "Moda: An efficient algorithm for network motif discovery in biological networks," *Genes Genet. Syst.*, vol. 84, no. 5, pp. 385–395, 2009.
[34] Z.-R.-M. Kashani *et al.*, "Kavosh: A new algorithm for finding network motifs," *BMC Bioinform*, vol. 10, 2009, Art. no. 318.
[35] P.-M.-P. Ribeiro and F.- M.-A. Silva, "G-tries: An efficient data structure for discovering network motifs," in *Proc. ACM Symp. Appl. Comput.*, 2010, pp. 1559–1566.
[36] S. Khakabimamaghani, I. Sharafuddin, N. Dichter, I. Koch, and A. Masoudi-Nejad, "Quatexelero: An accelerated exact network motif detection algorithm," *PLoS One*, vol. 8, no. 7, 2013, Art. no. e68073,
[37] D. Braines, D. Felmlee, D. Towsley, K. Tu, R.-M. Whitaker, and L.-D. Turner, "The role of motifs in understanding behavior in social and engineered networks," *Defense Secur.*, vol. 10653, 2018, doi: 10.1117/12.2309471.
[38] M.-B.-Z. Joveini and K.-C Keong, "An efficient multiresolution clustering for motif discovery in complex networks," *IEEE/ACM Trans. Comput. Biol. Bioinf.*, vol. 19, no. 1, pp. 284–294, 1 Jan./Feb. 2022, doi: 10.1109/TCBB.2020.3003018.
[39] W.-G. Underwood, A. Elliott, and M. Cucuringu, "Motif-based spectral clustering of weighted directed networks," *Appl. Netw. Sci.*, vol. 5, no. 1, pp. 1–41, 2020, doi: 10.1007/s41109-020-00293-z.
[40] S. Sarkar, R. Guo, and P. Shakarian, "Using network motifs to characterize temporal network evolution leading to diffusion inhibition," *Social Netw. Anal. Mining*, vol. 9, no. 1, pp. 1–24, 2019, doi: 10.1007/s13278-019-0556-z.
[41] S.-Y. Liu, J. Xiao, and X.-K. Xu, "Link prediction in signed social networks: From status theory to motif families," *IEEE Trans. Netw. Sci. Eng.*, vol. 7, no. 3, pp. 1724–1735, Mar. 2020.
[42] J. Li, Y. Zhou, J. Wu, and Z. Ma, "Complex dynamics of a simple epidemic model with a nonlinear incidence" *Discrete Continuous Dynamical Syst.*, vol. 8, no. 1, pp. 161–173, 2007.
[43] J. Cao, C. Ding, and B. Shi, "Motif-based functional backbone extraction of complex networks," *Physica A, Stat. Mechanics Appl.*, vol. 526, 2019, Art. no. 121123, doi: 10.1016/j.physa.2019.121123.

Xinxin Zhang received the B.S. degree in applied mathematics from Zhengzhou Normal University, Zhengzhou, China, in 2016, and the M.S. degree in applied mathematics from Fujian Normal University, Fuzhou, China, in 2019. She is currently working toward the doctorate's degree. Her research interest include network and information security.

Li Xu received the Ph.D. degree from the Nanjing University of Posts and Telecommunications, Nanjing, China, in 2004. He is currently a Professor and Doctoral Supervisor with the College of Computer and Cyber Security, Fujian Normal University, Fuzhou, China. He is also the Dean of the College of Computer and Cyber Security, Fujian Normal University. His research interests include network and information security, wireless networks and communication, intelligent information in communication networks. He has authored or coauthored more than 150 papers in international journals and conferences, including IEEE TRANSACTIONS ON COMPUTER, IEEE TRANSACTIONS ON RELIABILITY, IEEE TRANSACTIONS ON PARALLEL AND DISTRIBUTED SYSTEMS, INFORMATION SCIENCE.



Zhenyu Xu received the B.E. degree in network engineering from the Anhui University of Technology, Ma'anshan, China, in 2019. He is currently working toward the master's degree. His research interests include network and information security.