How to Protect Ourselves From Overlapping Community Detection in Social Networks

Dong Liu, *Member, IEEE*, Guoliang Yang, Yanwei Wang, Hu Jin, *Senior Member, IEEE*, and Enhong Chen, *Senior Member, IEEE*

Abstract—In recent years, overlapping community detection algorithms have been paid more and more attention, which not only reveal the real social relations, but also expose the possible communication channels between communities. Those individuals (or people) in the overlapping area are very important to the communities that can promote communication between two or more communities. On the other hand, from the privacy perspective, some people may not want to be found out in the overlapping areas. With this in mind, we raise a question "Can individuals modify their relationships to avoid the community discovery algorithms locating them into overlapping areas?" If this problem could be solved, these people may not need to worry about being disturbed. In particular, we first give three heuristic hiding strategies, i.e., Random Hiding(RH), Based Degree Hiding(DH) and Betweenness Hiding(BH), as comparison, utilizing the randomly the node, information of node degree and node betweenness centrality, respectively. And then, we propose a novel hiding algorithm by exploiting the importance degree of nodes in communities based on which the corresponding social connections are added or deleted called name *BIH*. Through extensive experiments, we show the effectiveness of the proposed algorithm in moving out a target node from overlapped areas.

Index Terms—Overlapping community detection, community deception, hiding, privacy protection

1 Introduction

Research on complex networks is a hot field of big data, which involved in many aspects of life, such as social networks, power grids, biological networks, and so on. The research of complex networks includes multiple disciplines, for example computers, physics, social sciences, etc. As a classical research issue in complex network science, community detection has attracted many researchers of various disciplines [1], [2]. Community detection is to find communities in a network whose intra-community connections closer than inter-community ones [3]. Therefore, accurate discovery of the community structure is an effective way to learn complex social networks. Most of the early research works focused on *non-overlapping* community detection algorithm which divides a complex network into several non-intersect communities where a node can only belong to one community [4]. On the other hand, it is natural that *overlapping* community partitions can better reflect the structure of a network in real life [5]. In general, people can be a member of several groups at the same time. For example, a person can be an employee of a company while also being a member of football club. Although these people are important for establishing communication channels between the two communities, in some situations some people may want not to be discovered by any overlapping community detection algorithm if their privacy becomes the first priority. Considering the above practical issues, we raise the following question: Can individuals (or people) artificially manage their social relationships so that overlapping community detection algorithms cannot locate them into overlapping areas? It can help the public protect their privacy from corporate interests [6]. Reversely, it may also help establishment understand how industrial espionage escape from detection, especially given the increasing reliance on social networks.

Community concealment or deception is a kind of hiding algorithms against community detection algorithms, which can reduce the detection accuracy of the community detection algorithms by slightly modifying some network structures [7], [8], [9], [10], [11], [12]. It can be used as a guideline that individuals or organizations can artificially manage their social circles or change the way of communication or channels to protect their privacy from being disclosed [13]. Thus far, most of the existing community hiding algorithms [6], [14] are designed against for *non-overlapping* community

- *Dong Liu is with the College of Computer and Information Engineering, Key Laboratory of Artificial Intelligence and Personalized Learning in Education of Henan Province, Big Data Engineering Lab of Teaching Resources & Assessment of Education Quality, Henan Normal University, Xinxiang 453007, China. E-mail: liudong@htu.edu.cn.*
- *Guoliang Yang and Yanwei Wang are with the College of Computer and Information Engineering, Henan Normal University, Xinxiang 453007, China. E-mail: leo@zzwwjj.com, wangyanwei39@163.com.*
- *Hu Jin is with the Division of Electrical Engineering, Hanyang University, Ansan 15588, Republic of Korea and Henan Normal University, Xinxiang 453007, China. E-mail: hjin@hanyang.ac.kr.*
- *Enhong Chen is with the Anhui Province Key Laboratory of Big Data Analysis and Application, School of Computer Science and Technology, University of Science and Technology of China, Hefei, Anhui 230000, China. E-mail: cheneh@ustc.edu.cn.*

Manuscript received 11 Dec. 2021; revised 12 Jan. 2022; accepted 13 Feb. 2022. Date of publication 18 Feb. 2022; date of current version 9 July 2022.
This work was supported in part by the National Natural Science Foundation of China under Grant 62072160, in part by the Key Scientific and Technical Project of Henan Province under Grant 212102310381 and in part by the Science Foundation of Distinguished Young Scholars of Henan Normal University under Grant 20210347.
(Corresponding authors: Dong Liu and Enhong Chen.)
Recommended for acceptance by D. Jin, G. Song, W. Wang, P. S. Yu, J. Han.
Digital Object Identifier no. 10.1109/TBDATA.2022.3152431

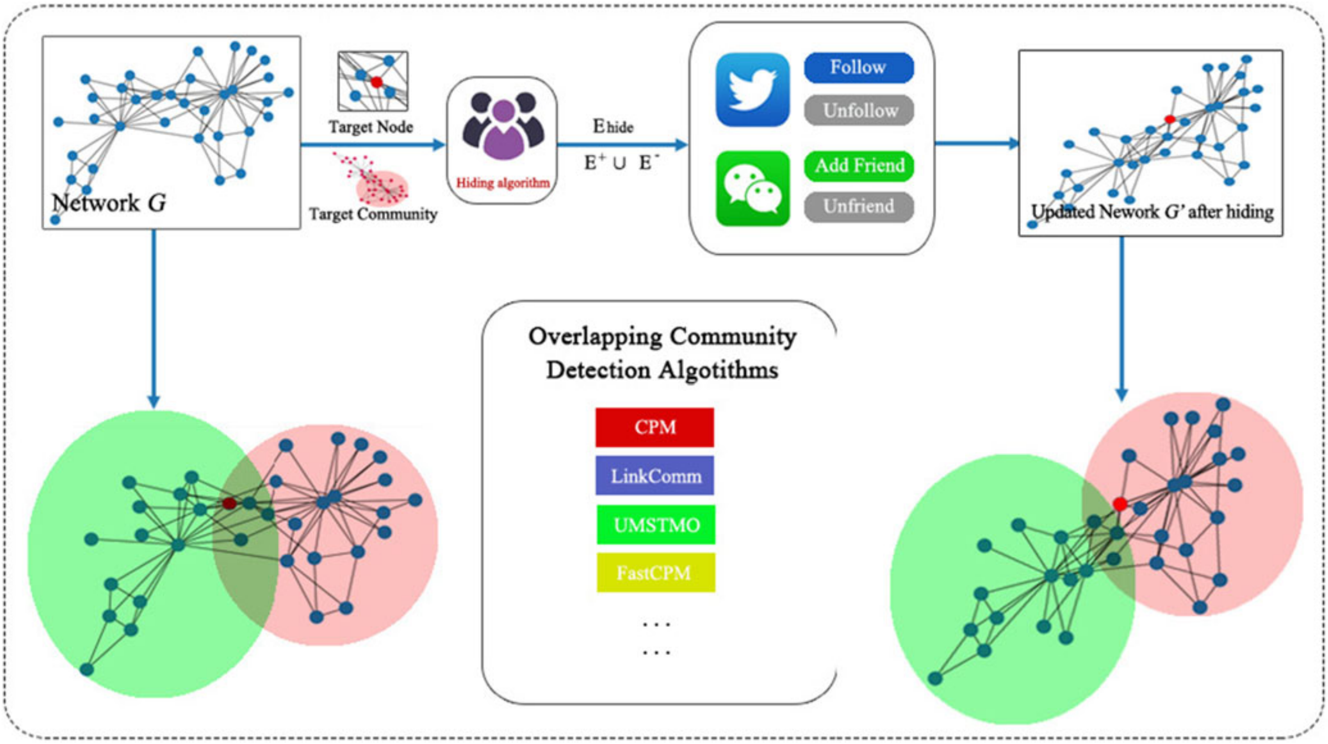


Fig. 1. Overall procedure of applying the proposed hiding algorithm.

detection algorithms [15], and to the best our knowledge, there is no hiding algorithm against for *overlapping* community detection algorithms.

In this paper, we propose a hiding algorithm for nodes (individuals) who are located in the overlapped area found by overlapping community detection algorithms and introduce evaluation methodologies to prove the effectiveness of the proposed algorithm. Our proposed hiding algorithm includes deleting (e.g., unfollow in Twitter, unfriend in Wechat) and adding (e.g., follow in Twitter, befriend in Wechat) some connections (or relationships), which is not difficult to implement in the social networks like Twitter, Wechat, etc. [16]. For example, We suppose that a person has his own social network data and he uses a certain overlapping community detection algorithm to check whether himself is located in some overlapping area. If he wants to hide his social relationships and move himself into a single community, he can apply the hiding algorithm to find the relationships which have the greatest impact on him, and then artificially modify the corresponding relationships. Fig. 1 introduces an example of the overall procedure of applying the hiding algorithm. For a given network G which is constructed from a social network, suppose that a target node tagged by red color is located in an overlapping area after applying some overlapping community detection algorithms like CPM [17], LinkComm [18], UMSTMO [19], FastCPM [20], etc. Our purpose is to make this target node being located in a single target community as possible. This target community represents the social relationships that user want to be discovered by the detection algorithm. After artificially deleting or adding some connections based on their importance through the proposed hiding algorithm, only one of the newly detected communities contains the target node.

In designing our proposed algorithm, the main challenge is how to choose the connections to delete or add. For this, we need to first measure the importance of a node in a community which is detected by overlapping community detection algorithms [21]. One key observation is that as everyone has a different social status in human society [22], the importance of each node in a community is different while its influence [23] on the surrounding nodes is also different [24]. We design our proposed hiding algorithm by quantitatively exploiting this observation.

The main contributions of this paper are summarized as follows.

- We propose a novel hiding algorithm to hide a node from being located in overlapping areas detected by overlapping community detection algorithms.
- We introduce *importance degree* of nodes in a community, which represents the node influence level on other nodes. This importance degree is further utilized by the hiding algorithm when deleting or adding connections between the node of interests (i.e., the target node) and others.
- The purpose of the proposed hiding algorithm is not only to remove a target node from the overlapping area, but also try to find an optimal number of adding and deleting. We propose a *hiding evaluation index* to check if the above purpose is realized.
- We have carried out comprehensive experiments on six real social networks to compare the effects of the proposed hiding algorithm on four well-known overlapping community detection algorithms.

The rest of paper is organized as follows. Section 2 introduces the hiding strategy for nodes in social networks that exploit the importance levels of nodes in communities.

Section 3 presents the experiments we have carried and discusses the corresponding results. Lastly, Section 4 gives the concluding remarks.

1.1 Related Work

1.1.1 Community Detection Algorithm

The purpose of community detection algorithm is to divide the complex network into multiple communities. It plays an important role in studying the characteristics of complex networks. In recent years, community detection algorithms [25], [26], [27], [28] have been studied by many scholars, and a large number of excellent algorithms have been proposed. The existing community detection algorithms mainly can be divided into two categories: overlapping community detection algorithms and non-overlapping community detection algorithms. Here, we will briefly introduce the above two types of algorithms.

Personally, the overlapping community structure is closer to the real social structure. It allows nodes to be in multiple communities at the same time. The classical algorithms of overlapping detection algorithms include cluster percolation method proposed by Palla *et al.* [17], local fitness measure method proposed by Lancichinetti A *et al.* [29], [30], Similarity division of links method proposed by Ahn Y Y *et al.* [18], modularity optimization algorithms [31] and so on.

Non-overlapping community structure means that each node in network can belong to only one community. There are also some mature non-overlapping community detection algorithms. Newman and Girvan propose modularity optimization algorithm that requires global knowledge [32]. M. Rosvall *et al.* propose an algorithm based on random walk method [33]. X. J. Zhu *et al.* propose label propagation algorithms in 2002 [34] and affinity propagation algorithm proposed by Frey *et al.* [35], spectrum analysis algorithm [36], seed spread algorithms are very classic algorithms. Among them, the modularity optimization algorithm is popular and modularity is a measure of the quality of the community division.

1.1.2 Community Hiding Algorithm

Community hiding is an inverse problem of community detection. With continuous breakthroughs in the performance of community detection algorithms, negative effects have gradually emerged, such as privacy leaks. The purpose of community hiding is to hide the target community or some specific member in the community. Community hiding is to slightly modify the network structure or node attributes according to a certain strategy, so as to interfere with the recognition accuracy of the mainstream community detection algorithm and realize the hiding of the target community structure. In 2010, Nagaraja raised the issue of community hiding for the first time, but it did not cause widespread concern among scholars. Until 2018, with Waniek's in-depth discussion of the hidden scientific and practical significance of the community, the research gradually attracted the attention of scholars and achieved certain research results, such as Detection Scoring Model [4], Q-Attack algorithm [37], REM method, EPA algorithm [13], CD-Attack algorithm [38], etc.

2 Proposed Methodology

We first introduce several preliminary definitions used for introducing the proposed hiding problem. An undirected graph $G=(V,E)$ is used to represent a social network which consists of a set of nodes $V=\{v_1,v_2,v_3,\dots,v_N\}$ and a set of edges $E=\{e_1,e_2,e_3,\dots,e_M\}$ [39] where each node represents an individual while each edge represents the social relationship between the two connecting end nodes. Note that N and M indicates the numbers of nodes and edges in the network, respectively. Suppose that a certain overlapping community detection algorithm [40], [41], [42], [43] could divide the network into a group of communities, in which each community may have nodes overlapping with other communities. Let the set of detected communities or equivalently the community structure be denoted by $C=\{C_1,C_2,C_3,\dots,C_K\}$ with $C_i\subseteq V$ for $i\in\{1,2,\dots,K\}$ and suppose that there exists at least one pair of i and j in $\{1,2,\dots,K\}$ that satisfy $C_i\cap C_j\neq\emptyset$ [44]. If a node n is simultaneously included in several communities say $C_1,C_2,\dots C_p$ after the overlapping community detection, then we call that node n is in the overlapping area of $\bigcap C_{\{1,2,\dots,p\}}$ which can also be represented by the notation of $n\in\bigcap C_{\{1,2,\dots,p\}}$.

For a given target node $n\in\bigcap C_{\{1,2,\dots,p\}}$, our purpose is to locate n into a single community C_i for $i\in\{1,2,\dots,p\}$ while escaping from the other communities as much as possible. We name this procedure as *hiding* in this paper. For further description, we define the following two notations:

- $E^+(n,u)$ with $n,u\in C_i, E(n,u)\notin E$: It is the operation of adding an edge between node n and node u in C_i.
- $E^-(n,v)$ with $n\in C_i, v\notin C_i$ and $v\in\bigcup C_{\{1,2,\dots,p\}}/C_i$: It denotes the operation of deleting an edge between node n and node v where v belongs one of $\{C_1,C_2,\dots,C_{i-1},C_{i+1},\dots,C_p\}$.

2.1 Problem Formulation

Given an undirected, unweighted network $G=(V,E)$ and a target node $n\in\cap C_{\{1,2,\dots,p\}}$, suppose that the target node does not want to be located into the overlapping area detected by a certain overlapping community detection algorithm. Let C_i denote the target community that the target node n wants to be located at the end. Our proposal is to artificially modify the social relations of node n through adding (E^+) or deleting (E^-) edges so that it can escape from the overlapping area. Let $E^+_{\text{hide}}=\{E^+_1,E^+_2,\dots,E^+_T\}$ and $E^-_{\text{hide}}=\{E^-_1,E^-_2,\dots,E^-_T\}$ be the hiding solution each of which comprises of T series of edge addition or deletion operations. T is the number of modified edges. When performing such hiding, in order to reduce the impact on original network, it would be better to make T as small as possible. After applying the proposed hiding procedure, it would be desired to have the new network $G'=(V',E')$ and the community structure as follows:

$$\begin{cases} n\in C_i \\ n\notin\bigcup C_{\{1,2,\dots,p\}}/C_i \\ E'=E+E^+_{\text{hide}}-E^-_{\text{hide}} \\ V'=V \end{cases}.$$

Consequently, our purpose is to find out such E^{+}_{hide} and E^{-}_{hide} to remove out the target node n from the overlapping area by modifying some edges in G [37]. Fortunately, the addition and deletion of edges are not so difficult to implement in social networks [45] as we have a high probability of adding friends [46] with people in the same community who may have common interests, hobbies, etc. After hiding, the overlapping community detection algorithm divides node n into a single community instead of the overlapping area.

2.2 Heuristic Attack Strategy

We have introduced that we take hidding algorithm on network to change some relations between nodes. In this paper, our goal is to move the target node from the overlapping area to a single community which is called target community. Obviously, we need to add edges between the target node with other node in target community while delete the edges that already exist in other communities containing n, which called $rewiring+$. In social networks [47], one $rewiring+$ means geting a false friend while cancels the friendship from a real friends in others communities.

While thinking of hiding strategies, we first come up with a Random Hiding (RH) strategy, described in Algorithm 1, for comparison. Here, we add a edge between target node n with a node u that randomly choosed in target community C_i while delete edges one by one in other communities containing n. The neighbor set of the target node n in community is donoted by $N^n_{C_i} = \{v|e(n,v) \in Eandn, v \in C_i\}$, while its nonneighbor set in community is denoted by $\widetilde{N}^n_{C_i}$. The number of $rewiring+$ (or the number of iterations), denoted by T, is another parameter that represents the $hiding$ algorithm cost.

Algorithm 1. Random Hiding (RH)

```
Input: Network G, T, Target node n;
Output: Updated Network G;
C ← getCommunities(G);
[n_1, n_2, ..., n_L] ← getNodesInOverlappingArea(C);
if n ∈ [n_1, n_2, ..., n_L] then
  [C_1, C_2, ..., C_K] ← getCommunities(C, n);
  C_i ← RandomChooseTargetCommunity();
  while T > 0 do
    Ñ^n_{C_i} ← getNonneighborSet(C_i, n);
    u ← RandomChooseNode(N^n_{C_i});
    add edge e(n, v) to E;
    for j ∈ [1, 2, ..., i−1, i+1, ..., K] do
      N^n_{C_j} ← getNeighborSet(C_j, n);
      v ← RandomChooseNode(N^n_{C_j});
      remove the edge e(n, v) from E;
    end
    G' ← Update G;
    T = T − 1
  end
end
```

Obviously, the Randomly Hiding strategy are unstable. In human society, everyone has different social status. If we delete or add edges based on the measurement method of some nodes, this may help the target node to get out of the overlapping area faster. Based on this analysis, We choose degree centrality and betweenness centrality as a measure of the importance [48] of a node in the community and propose the corresponding hiding algorithm, which called name Degree Hiding(DH) and Betweenness Hiding(BH). We execute $rewiring+$ between target node with a node that has height Degree or Betweenness, described in Algorithms 2 and 3.

Algorithm 2. Degree Hiding (DH)

```
Input: Network G, T, Target node n;
Output: Updated Network G;
C ← getCommunities(G);
[n_1, n_2, ..., n_L] ← getNodesInOverlappingArea(C);
if n ∈ [n_1, n_2, ..., n_L] then
  [C_1, C_2, ..., C_K] ← getCommunities(C, n);
  C_i ← RandomChooseTargetCommunity();
  while T > 0 do
    Ñ^n_{C_i} ← getNonneighborSet(C_i, n);
    u ← ChooseNodeBaseDegree(N^n_{C_i});
    add edge e(n, v) to E;
    for j ∈ [1, 2, ..., i−1, i+1, ..., K] do
      N^n_{C_j} ← getNeighborSet(C_j, n);
      v ← ChooseNodeBaseDegree(N^n_{C_j});
      remove the edge e(n, v) from E;
    end
    G' ← Update G;
    T = T − 1
  end
end
```

Algorithm 3. Betweenness Hiding (BH)

```
Input: Network G, T, Target node n;
Output: Updated Network G;
C ← getCommunities(G);
[n_1, n_2, ..., n_L] ← getNodesInOverlappingArea(C);
if n ∈ [n_1, n_2, ..., n_L] then
  [C_1, C_2, ..., C_K] ← getCommunities(C, n);
  C_i ← RandomChooseTargetCommunity();
  while T > 0 do
    Ñ^n_{C_i} ← getNonneighborSet(C_i, n);
    u ← ChooseNodeBaseBetweenness(N^n_{C_i});
    add edge e(n, v) to E;
    for j ∈ [1, 2, ..., i−1, i+1, ..., K] do
      N^n_{C_j} ← getNeighborSet(C_j, n);
      v ← ChooseNodeBaseBetweenness(N^n_{C_j});
      remove the edge e(n, v) from E;
    end
    G' ← Update G;
    T = T − 1
  end
end
```

2.3 Importance Degree for Nodes in a Community

With the above discussion, Our goal is to remove the target node from the overlapping area. Therefore, we need to calculate and adjust the importance of the target node in different communities by addition the edges between the target node and other nodes in the target community or deleting the edges with the target node in other communities.

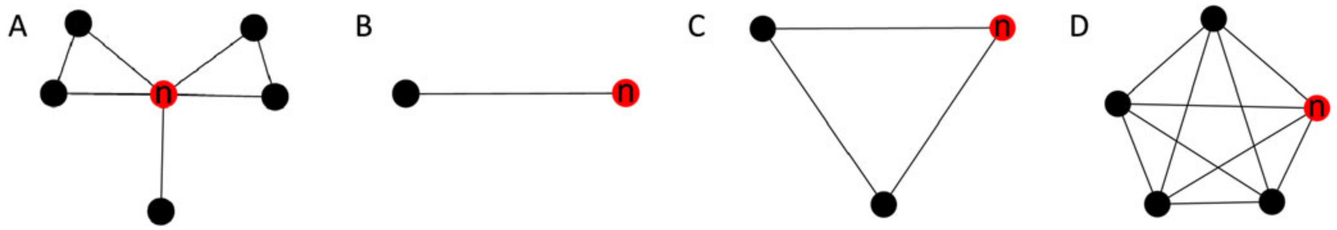


Fig. 2. (A) We assume that node n has five neighbors in a certain community C_i. For this example, $I(n, C_i) = (1+1+0+1+1)/5 = 4/5$. (B) an isolated edge, $I(n, C_i) = 0$. (C) an isolated triangle has $I(n, C_i) = 1/2$. (D) an special case of full connection in which I reaches the maximum, $I(n, C_i) = 3 * 4/4 = 3 = Deg(n) - 1$.

In an undirected, unweighted network, network topology is only available information, so the neighbor of the node will be the key to determine the importance degree for node in the community. It's natural to use neighbor information when we define importance degree for node.

From this, it can be given that the original importance of node n in the community C_i as follows.

$$I^{original}(n, C_i) = \frac{\sum_{v \in N^n_{C_i}} |N^n_{C_i} \bigcap N^v_{C_i}|}{Deg(n)},$$

where $Deg(n)$ is the degree of node n in network, $|N^n_{C_i} \bigcap N^v_{C_i}|$ is the number of elements in the intersection of the neighbors of node n and the neighbors of node $v, v \in N^n_{C_i}$ in community C_i.

An example of the original Importance Degree for node is shown in Fig. 2. It's easy to know that $I^{max}(n, C_i) = Deg(n) - 1$, which happens in the fully connected. By fully connected, we mean that any two vertices in a community have edge. Obviously, the minimum value of I is 0. In order to keep the value of Importence Degree for node at [0,1], we now normalize original I and get the Importence Degree for node as follow:

$$I(n, C_i) = \frac{(\sum_{v \in N^n_{C_i}} |N^n_{C_i} \bigcap N^v_{C_i}|)(Deg(n) - 1)}{Deg(n)},$$

where node $n, v \in C_i$.

In order to remove the target node from the overlapping area and reduce the impact on the original network as much as possible, We have to find a node that has the greatest impact on the Importance Degree for target node to adding(E^+) or deleting (E^-). This is very obvious that we adding or deleting the edges with node v when we get the maximum $|N^n_{C_i} \bigcap N^v_{C_i}|$, which has the greatest impact on the Importance Degree for target node n in the community C_i.

Proof. Edges additions: Suppost node u and target node n be part of a certain community C_i, but u is not n's neighbor.

$$\begin{aligned}\Delta(I(n, C_i)) &= |I_{after}(n, C_i) - I_{before}(n, C_i)| \\ &= \frac{((\sum_{v \in N^n_{C_i}} |N^n_{C_i} \bigcap N^v_{C_i}|) + |N^n_{C_i} \bigcap N^u_{C_i}|)(Deg(n))}{Deg(n) + 1} \\ &\quad - \frac{(\sum_{v \in N^n_{C_i}} |N^n_{C_i} \bigcap N^v_{C_i}|)(Deg(n) - 1)}{Deg(n)} \\ &= \frac{(|N^n_{C_i} \bigcap N^u_{C_i}|)Deg(n) + \sum_{v \in N^n_{C_i}} |N^n_{C_i} \bigcap N^v_{C_i}|}{Deg(n)(Deg(n) + 1)}.\end{aligned}$$

Being $Deg(n) > 0$ always hold, $\Delta(I(n, C_i))$ get max increase when $|N^n_{C_i} \bigcap N^u_{C_i}|$ gets the maxinum. The proof of edge deletion is the acquaintance of edge additions. □

2.4 Proposed Hiding Algorithm

In this section, we introduce the details of the hiding algorithm based on the importance degree. The algorithm adding an edge to the target node with one node in target community that the user want to be discovered social relations by overlapping detection algorithm. After that, its deleting an edge in other commnities simultaneously.

Algorithm 4. Based Importance Hiding (BIH)

Input: Network G, T, Target node n;
Output: Updated Network G;
$C \leftarrow$ getCommunities(G);
$[n_1, n_2, \ldots, n_L] \leftarrow$ getNodesInOverlappingArea(C);
if $n \in [n_1, n_2, \ldots, n_L]$ **then**
  $[C_1, C_2, \ldots, C_K] \leftarrow$ getCommunities(C, n);
  $C_i \leftarrow$ RandomChooseTargetCommunity();
  while $T > 0$ **do**
    $\widetilde{N}^n_{C_i} \leftarrow$ getNonneighborSet(C_i, n);
$u \leftarrow$ getNodewithHighestImportance($N^n_{C_i}$);
add edge $e(n, v)$ to E;
for $j \in [1, 2, \ldots, i-1, i+1, \ldots, K]$ **do**
    $N^n_{C_j} \leftarrow$ getNeighborSet(C_j, n);
    $v \leftarrow$ getNodewithHighestImportance($N^n_{C_j}$);
    remove the edge $e(n, v)$ from E;
  end
  $G' \leftarrow$ Update G;
  $T = T - 1$
  end
end

In the case of one hiding, the importance degrees of two nodes are modified directly and, therefore, the cost of this kind of reconnection hiding is relatively low. In social networks, the operation of addition means that the person selected as the target node makes a new connection to the people with high importance in the target community while the deletion means that the target person cancels the friendship with the people with high importance in other communities. It is worthy to note that these operations are not difficult to implement, as these people are in the same community. The hiding algorithm is described in Algorithm 4.

In Algorithm 4, we select a target node n from an overlapping area which is divided by overlapping community detection algorithm and C_i is the target community in which the node n to be located. The parameter T represents

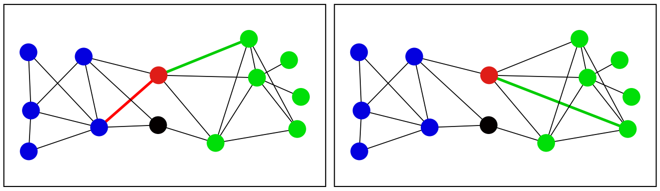


Fig. 3. Example of adding (green) and deleting (red) edges.

the number of hiding iterations. In each iteration, we try to select those nodes with high importance and perform deleting an existing edge and adding a non-existing edge to the target node. Fig. 3 shows a representative example. The target node may be located in several communities simultaneously. In order to remove the target node from the communities other than its target community, it would be better to perform delete operation several times which results in the introduction of iteration T in Algorithm 4. In order to prevent extreme cases, if the target node has only one edge left with the community to be removed, we do not delete the edge.

2.5 Hiding Evaluation Index

When designing the hiding algorithm, we not only want to make the target node move out of the overlapping area, but also want to minimize its impact on the original community and maintain the connectivity of the target node in the the original community. In order to evaluate the effect of the proposed hiding algorithm on hiding a node and changing the community structure, in this section, we introduce a hiding evaluation index.

For this, we first define the hiding accuracy in order to evaluate the efficiency of the target node moving out of the overlapping area. Given a target node $n \in \bigcap C_{\{1,2,\ldots,p\}}$ and its target community C_i for $i \in \{1, 2, \ldots, p\}$, we define the hiding accuracy $R(n)$ of the hiding algorithm as follows:

$$R(n) = \begin{cases} 1, & H_n = 1, \\ 1 - \frac{H_n}{O_n}, & H_n > 1, \\ 0, & H_n > O_n, \end{cases}$$

where O_n is the number of communities that contain the target node n before hiding, and H_n is the number of communities that contain node n after hiding. In the best case, the node is included by only one target community. The worst case is that the number of communities where node n is included does not change or even increase after hiding. According to our definition of $R(n)$, the worse case results the value of $R(n)$ being 0. Moreover, a larger $R(n)$ indicates a more accurate hiding.

Based on the hiding accuracy $R(n)$, we are able to define the hiding evaluation index as follows:

$$A(n) = R(n) \cdot \left(\frac{1}{2} NMI + \frac{1}{2} \frac{|Q_n|}{|O_n|}\right),$$

where NMI is the Normalized Mutual Information [29], which can be used to assess the similarity between the two community structures before and after hiding [49]. Q_n is the number of communities that still maintain edges to the target node n even after hiding. The degree of change in the community structure is captured by the first term in the parenthesis, i.e., NMI. Ideally, except for the target node, it would be better to have the phenomenon that the community structure of all the other nodes remains unchanged. Moreover, the connectivity between the target node and the communities containing it before hiding is captured by the second term in the parenthesis. The best situation is that the target node still has connections to the communities containing it before hiding. Consequently, the hiding evaluation index $A(n)$ considers both the hiding efficiency and the community structure changing level. The value $A(n)$ ranges in [0,1) and a larger $A(n)$ indicates a better performance. While it can infinitely close to 1 but it cannot equal to 1 as moving the target community out of the overlapping area inevitably leads to the change of community structure. The worst case is that the hiding algorithm not only does not move the target node out of the overlapping area, but also brings significant changes to the original community structure.

As an important parameter in $A(n)$, NMI is commonly used criterion to assess the quality of clustering results in analyzing network community structures proposed by Danon *et al.* [50]. NMI is an important measure of community detection which can objectively evaluate the similarity of two community structures. In general, NMI ranges in [0, 1] and the higher the value, the more similar two community structures will be. For the community structure C and the new community structure B after hiding, NMI is defined as follows:

$$NMI = \frac{-2\sum_{i=1}^{L_C}\sum_{j=1}^{L_B} \log\left(\frac{N_{ij} \cdot N}{N_i \cdot N_j}\right)}{\sum_{i=1}^{L_C} N_i \cdot \log \frac{N_i}{N} + \sum_{j=1}^{L_B} N_j \cdot \log \frac{N_j}{N}},$$

where L_C and L_B denote the numbers of communities in C and B, respectively. N is the total number of nodes in the network G. N_{ij} is the number of nodes shared by the ith community in C and the jth community in B. In addition, $N_i = \sum_j N_{ij}$ and $N_j = \sum_i N_{ij}$. Overall, when designing the hiding algorithm, we hope to minimize the impact on the original community structure and maintain the accessibility. Therefore, a larger NMI is preferred.

3 Experiments

In this section, we apply the proposed hiding algorithm to six real-world networks and check its effectiveness in hiding a node from overlapping areas. Experiments have been conducted on a personal computer with i3 CPU and 6 GBs RAM. For each experiment, we run 10 times and record the result. We implement the algorithm using networkX package.

3.1 Evaluation Methods

As there is no previous hiding algorithm in this literature, in this section we try to describe the performance of our proposed hiding algorithm through various perspective to introduce the related phenomena and compared with the heuristic algorithm(RH, DH, BH) mentioned above. As the evaluation methodology, We start with a community structure C found by some overlapping community detection algorithm. In the process of experiment, for the community structure obtained by using the overlapping detection algorithm, we retain the community whose number of members

TABLE 1
Detected Communities From Real Social Networks

Net	Number of nodes (N)	Number of edges (M)	Number of communities (N_c) / Number of nodes (N_n) in overlapping areas							
			CPM		LinkComm		UMSTMO		FastCPM	
			N_c	N_n	N_c	N_n	N_c	N_n	N_c	N_n
Dolphins	62	159	4	6	66	51	7	5	4	6
Football	115	613	4	15	158	115	8	12	4	15
Power	4941	6594	297	120	4928	3398	2967	39	297	120
DBLP	4997	14268	806	180	820	262	639	30	806	180
Geom	6158	11898	929	522	4766	1384	1363	210	929	522
Amazon	16716	48739	1592	293	7820	7002	2165	409	1592	293

is more than two. This is because we don't think the target member needs to move out of a community with only two members. Then, we randomly select one node in an overlapping as the target node, and randomly select a community containing the target node as the target community. For the RH algorithm, we randomly select a nonneighbor node in target commninty to adding edge and delete edges from neightbor node that randomly selected in other communities. For the BH and DH algorithm, we select the node with highest Degree and Betweenness to $adding$ or $deleting$. For the BIH algorithm, we calculate the importance degrees of the nodes in the communities and choose node with highest *Importance Degree* to implement $rewiring+$. And then, we apply the T iterations of hiding operations to the original community structure C and obtain an updated networks G'. Finally, we are able to calculate the hiding evaluation index $A(n)$ and NMI. Through the assessment performed in this section, we try to answer the following questions:

- How to move nodes out of the overlapping area through the hiding algorithm?
- What is the effect of the iteration T?
- How the proposed hiding algorithm changes the original community structure?
- What is the computing load of the hiding algorithm and the community detection algorithms?

3.2 Datasets

In order to verify the effectiveness of our proposed hiding algorithm, we test four overlapping community detection algorithms to detect six different community network datasets. The overlapping community detection algorithms considered are as follows: CPM algorithm [17], LinkComm algorithm [18], UMSTMO algorithm [19] and FastCPM algorithm [20]. The detailed descriptions for the above algorithms are given as follows:

CPM. By applying the method of searching neighboring, CPM found the community structure which consists of segveral overlapping connected communities.

LinkComm. LinkComm uses hierarchical clustering with link (or edge) similarity to build a dendrogram where each leaf is a link from the original network and branches represent link communities. Then, by using the density function the dendrogram is cut to obtain the community structure having overlapping and hierarchical structure.

UMSTMO. The UMSTMO offer a new method that explore the union of all maximum spanning trees (UMST) and models the strength of edges between nodes. Also, each node in the UMST is linked with its most similar neighbor. From this model, the authors extract local community for each node, and then they combine the produced communities according to their number of shared nodes.

FastCPM. FastCPm presents improvements to K-clique percolation, which allow us to perform k-lique percolation on much larger empirical datasets.

In addition, in the following, we briefly describe the six real network datasets as follows.

The USA College Football (Football) [51]. This network represents the matches between American football teams during the season of 2,000. Each node represents a university team and each edge between two nodes indicates that there is at least one game between the two teams.

The US Power Grid Network Dataset (Power). This network is the high-voltage power grid in the Western States of the United States of America. The nodes are transformers, substations, and generators, and the edges are high-voltage transmission lines.

Dolphin Social Network(Dolphins). Dolphin data set is a dolphin social network obtained by D. lusseau *et al*. Who spend seven years to observe the communication of 62 dolphin groups in the doubleful sound channel of New Zealand.

DBLP Dataset(DBLP). Digital Bibliography Project (DBLP) is a computer science bibliography. In this data set, authors are considered as nodes, the paper titles of the authors are the text of nodes and the coauthor relationship forms the edges.

Geom. The network Geom.net is based on the file geombib.bib that contains Computational Geometry Database, version February 2002.

Amazon Data. The data was collected by crawling Amazon website.

Table 1 shows the number of communities detected by each overlapping community detection algorithm and the number of nodes in the overlapping areas. The networks considered have large variations in scale, i.e., different numbers of nodes and edges. Therefore, we are able to check the effectiveness of the proposed hiding algorithm in diverse scenarios. In additions, high density network is not used because CPM algorithm is difficult to divide it.

3.3 Performance Evaluation

Our goal is to find suitable T value and maintaining connectivity between the target nodes and communities while

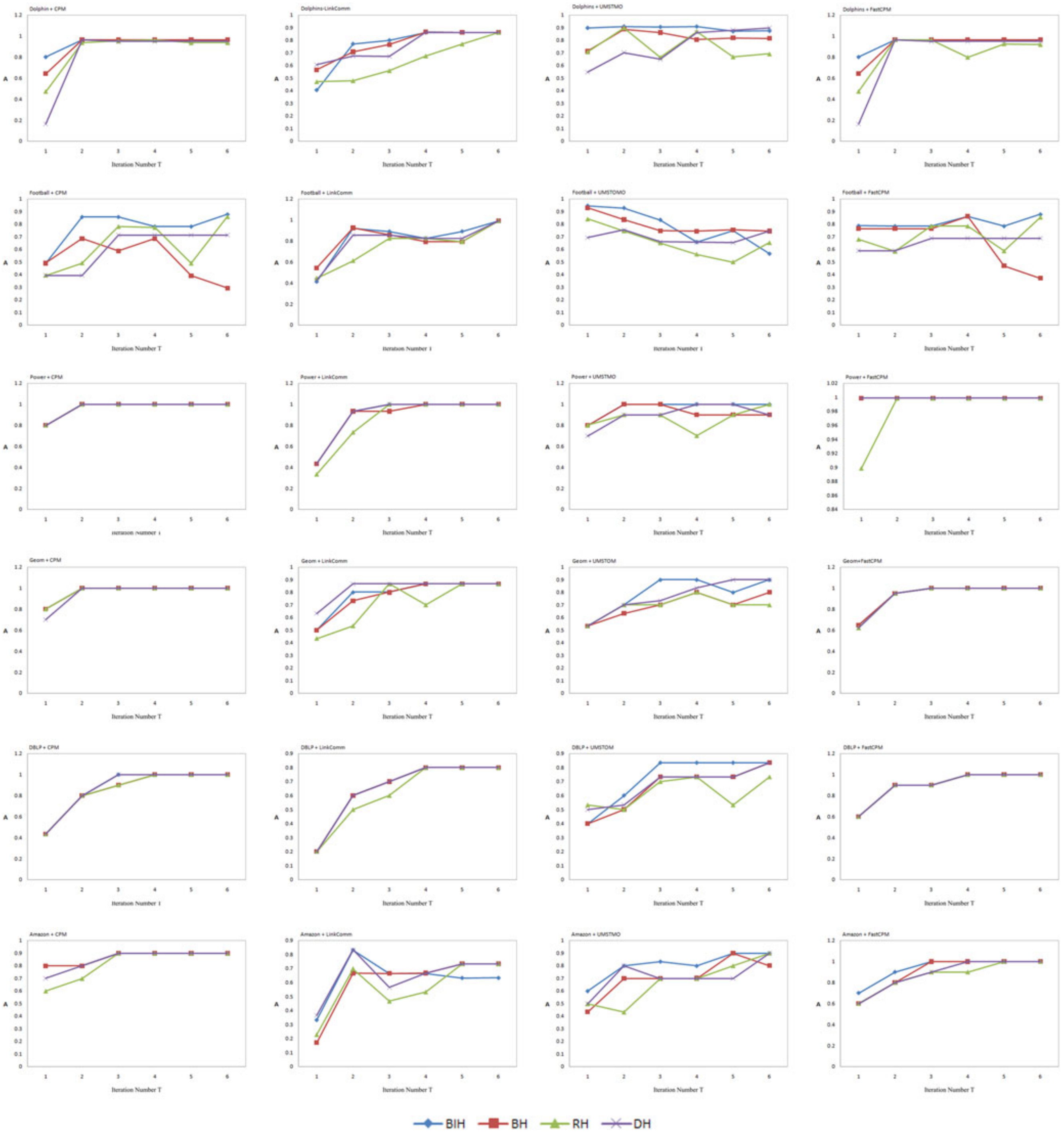


Fig. 4. The performance of $|A(n)|$ over varying the iteration number T for different overlapping community detection algorithms, hiding algorithm and social networks.

moving the target node out of the overlapping area. For this, we first evaluate the hiding evaluation index $A(n)$. For the evaluation, we randomly select 10 target nodes and their target communities and apply the proposed hiding algorithm BIH, RH, BH and DH. In addition, to obtain the optimal value of the iteration number T, we plot $A(n)$ over varying T from 1 to 6. Fig. 4 shows the values of $A(n)$ when those proposed algorithm is applied for the four overlapping detection algorithms and the six networks.

We can observe that as we increase T for the *BIH* hiding algorithm, the value of $A(n)$ up to 0.8 for the overlapping community detection algorithms of CPM, UMSTMO, fastCPM, and LinkComm. RH, DH and BH algorithm also achieved good results, although their performance is not as good as BIH algorithm(except the power and dolphins network for LinkComm). Therefore, we can conclude the proposed *BIH* hiding algorithm works well with the above community detection algorithms. In addition, we can also observe the following phenomena: 1) CPM and FastCPM is sensitive to all the hiding algorithms(execept football network), which means that we just need to modify a few edges to move the target node out of the overlapping area.

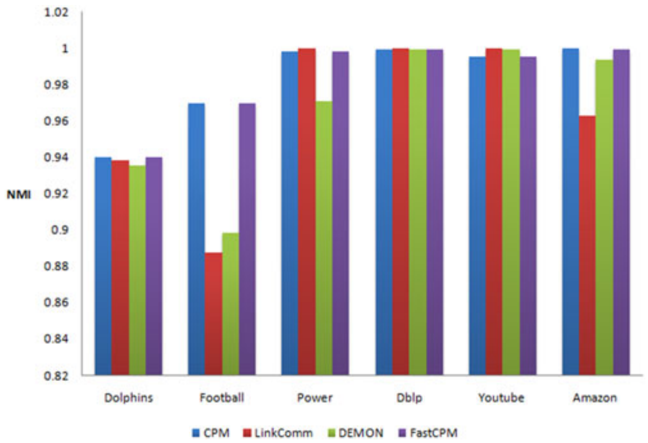


Fig. 5. The evaluated NMI values for different overlapping community detection algorithms and different networks.

2) RH algorithm shows randomness, as we expected. 3) For LinkComm on Dolphins network, the performance of BH algorithm is better than BIH algorithm when the $T = 1$, but when $t > 4$, they get the same $A(n)$ value. 4) For UMSTOM on Football network, the value of $A(n)$ decreased with the increase of T which caused by the great influence of hiding algorithm on community structure. 5) For LinkComm on Geom network, The performance of BH algorithm is better than BIH algorithm.

In order to understand how the hiding algorithm changes the original (pre hiding) community structure, we also observe the NMI [49] value when the BIH hiding evaluation index reaches the maximum value. Fig. 5 shows the NMI values for different community detection algorithms over different networks. We can observe that a smaller network yields a smaller NMI value and with increasing the network scale, NMI gets close to 1. A smaller network usually has a smaller number of communities and if we move the target node from the overlapping area to the target community only, the community structure should be affected. For instance, on *Dolphins* network we observed that the algorithms such as CPM and UMSTMO yield a smaller number of communities as compared to the others and, consequently, introduces lower NMI values. In addition, on the Football network we observed that LinkComm yields a smaller NMI value, which is due to the fact that the community divided by LinkComm have many very small communities.

TABLE 2
Running Time (Unit: Second)

Networks	Algorithms	Load	BIH	Running time for the detection algorithms
Dolphins	CPM	0.0312	2.389	0.0056
	LinkComm		2.215	0.016
	UMSTMO		2.301	0.0820
	FastCPM		2.106	0.0032
Football	CPM	0.109	4.852	0.0156
	LinkComm		2.512	0.078
	UMSTMO		3.105	0.179
	FastCPM		2.028	0.0155
Power	CPM	1.326	3.307	0.1249
	LinkComm		4.467	0.593
	UMSTMO		3.619	64.75
	FastCPM		1.919	0.125
DBLP	CPM	0.249	1.7628	0.1248
	LinkComm		2.1372	3.4788
	UMSTMO		2.0592	7.739
	FastCPM		2.5896	0.5772
Geom	CPM	0.548	3.59	0.15
	LinkComm		3.331	0.659
	UMSTMO		4.508	21.59
	FastCPM		1.965	0.0624
Amazon	CPM	0.78	4.11	1.528
	LinkComm		2.713	11.435
	UMSTMO		3.75	98.36
	FastCPM		10.342	0.95

3.4 Running Time Evaluation

We compared the running time of the proposed BIH hiding algorithm and the overlapping community detection algorithms for different networks and the corresponding values are summarize in Table 2. Load refers to the time duration for which data is loaded into the memory and is transformed into the networkX graph structure. We can observe that the running time of the hiding algorithm has only a little variation over different network sizes. It is mainly because our proposed hiding algorithm operates on the communities containing the target node and does not need to utilize the global knowledge on the whole network. Time complexity of BIH algorithm is $O(|C|^2 + |C| \cdot Z)$ where Z is number of node in maximal community . The first part of it is the time complexity of finding the nodes in the overlapping area. The second part is the time complexity of the hiding algorithm. Obviously, the time complexity of the hiding algorithm is related to the community structure obtained by the overlapping detection algorithm.

3.5 Experiments Summary

From the above extensive evaluations conducted, we can conclude that the proposed BIH algorithm can not only hide the target node appropriately, but also introduces little impact on the network structure of most overlapping community detection algorithms. We observed that there is a correlation between the community distributions and the success of hiding. In general, the hiding effect is better when the number of nodes shared by two or more communities where the target node located is small and the community is relatively large. When it comes to the change of the community structure after hiding, we noticed that the impact can be ignored with the increase of network scale and community number.

4 Conclusion

Thus far, the research on overlapping community detections mainly focused on the discovery algorithms, while the research on the community hiding algorithms mainly focused on the non-overlapping community detection algorithms. However, in some situations, it is required to hide some personal information from the overlapping community. This paper studied the problem of hiding a node from overlapping communities. In particular, we first developed a formula to check the importance degree of nodes in communities based on which we can delete or add connections associated with the target node of interests. As we also want to minimize the changes to the original community structure while applying the hiding algorithm, we introduced the hiding evaluation index to check the corresponding effect. Lastly, through extensive experiments, we investigated the effectiveness of the proposed algorithm in moving out a target node from overlapped areas. For the future work, it would be interesting to extend the hiding algorithm to accommodate the dynamic community detection algorithms.

References

[1] P. M. Gleiser and L. Danon, "Community structure in jazz," *Adv. Complex Syst.*, vol. 6, no. 4, pp. 565–574, Oct. 2003.

[2] A. A. Abbasi and M. F. Younis, "A survey on clustering algorithms for wireless sensor networks," *Comput. Commun.*, vol. 30, no. 14/15, pp. 2826–2841, Oct. 2007.

[3] M. Wang, C. Wang, J. X. Yu, and J. Zhang, "Community detection in social networks: An in-depth benchmarking study with a procedure-oriented framework," *Proc. VLDB Endowment*, vol. 8, no. 10, pp. 998–1009, Jun. 2015.

[4] V. Fionda and G. Pirrò, "Community deception or: How to stop fearing community detection algorithms," *IEEE Trans. Knowl. Data Eng.*, vol. 30, no. 4, pp. 660–673, Apr. 2018.

[5] J. Xie, S. Kelley, and B. K. Szymanski, "Overlapping community detection in networks: The state-of-the-art and comparative study," *ACM Comput. Surv.*, vol. 45, no. 4, pp. 43:1–43:35, Aug. 2013.

[6] M. Waniek, T. P. Michalak, T. Rahwan, and M. J. Wooldridge, "Hiding individuals and communities in a social network," *Nature Hum. Behav.*, vol. 2, no. 2, pp. 139–147, Jan. 2018.

[7] H. Dai *et al.*, "Adversarial attack on graph structured data," in *Proc. Int. Conf. Mach. Learn.*, 2018, pp. 1115–1124.

[8] X. Chen, Z. Jiang, H. Li, J. Ma, and P. S. Yu, "Community hiding by link perturbation in social networks," *IEEE Trans. Comput. Soc. Syst.*, vol. 8, no. 3, pp. 704–715, Jun. 2021. [Online]. Available: https://doi.org/10.1109/TCSS.2021.3054115

[9] J. Jia, B. Wang, X. Cao, and N. Z. Gong, "Certified robustness of community detection against adversarial structural perturbation via randomized smoothing," in *Proc. The Web Conf.*, 2020, pp. 2718–2724.

[10] Y. Chen *et al.*, "Practical attacks against graph-based clustering," in *Proc. ACM SIGSAC Conf. Comput. Commun. Secur.*, 2017, pp. 1125–1142.

[11] K. He, Y. Li, S. Soundarajan, and J. E. Hopcroft, "Hidden community detection in social networks," *Inf. Sci.*, vol. 425, pp. 92–106, Jan. 2018.

[12] F. Bloch, B. Dutta, and M. Dziubinski, "A game of hide and seek in networks," *J. Econ. Theory*, vol. 190, Nov. 2020, Art. no. 105119.

[13] J. Chen, Y. Chen, L. Chen, M. Zhao, and Q. Xuan, "Multiscale evolutionary perturbation attack on community detection," *IEEE Trans. Comput. Soc. Syst.*, vol. 8, no. 1, pp. 62–75, Feb. 2021.

[14] S. Nagaraja, "The impact of unlinkability on adversarial community detection: Effects and countermeasures," in *Proc. 10th Int. Conf. Privacy Enhancing Technol.*, 2010, pp. 253–272.

[15] M. E. J. Newman and M. Girvan, "Finding and evaluating community structure in networks," *Phys. Rev. E*, vol. 69, no. 2, pp. 26 113–26 113, Feb. 2004.

[16] W. Chen and S. Teng, "Interplay between social influence and network centrality: A comparative study on shapley centrality and single-node-influence centrality," in *Proc. 26th Int. Conf. World Wide Web*, 2017, pp. 967–976.

[17] G. Palla, I. Derényi, I. Farkas, and T. Vicsek, "Uncovering the overlapping community structure of complex networks in nature and society," *Nature*, vol. 435, no. 7043, pp. 814–818, Jun. 2005.

[18] Y. Y. Ahn, J. P. Bagrow, and S. Lehmann, "Link communities reveal multiscale complexity in networks," *Nature*, vol. 466, no. 7307, Jun. 2010, Art. no. 761.

[19] K. Asmi, D. Lotfi, and M. E. Marraki, "Overlapping community detection based on the union of all maximum spanning trees," *Library Hi Tech*, vol. 38, no. 2, pp. 276–292, Jun. 2020.

[20] F. Reid, A. F. McDaid, and N. J. Hurley, "Percolation computation in complex networks," in *Proc. Int. Conf. Adv. Soc. Netw. Anal. Mining*, 2012, pp. 274–281.

[21] S. Kelley, "The existence and discovery of overlapping communities in large-scale networks," Ph.D. dissertation, Rensselaer Polytechnic Institute, Troy, NY, Nov. 2009. [Online]. Available: https://www.proquest.com/openview/3f717f51c7865c6d25ea2b236cd7df87/1?pq-origsite=gscholar&cbl=18750

[22] Q. Xuan, Z. Zhang, C. Fu, H. Hu, and V. Filkov, "Social synchrony on complex networks," *IEEE Trans. Cybern.*, vol. 48, no. 5, pp. 1420–1431, May 2018.

[23] I. A. Kovács, R. Palotai, M. S. Szalay, and P. Csermely, "Community landscapes: An integrative approach to determine overlapping network module hierarchy, identify key nodes and predict network dynamics," *PLoS One*, vol. 5, no. 9, pp. 1–14, Sep. 2010.

[24] C. Tai, P. S. Yu, D. Yang, and M. Chen, "Structural diversity for resisting community identification in published social networks," *IEEE Trans. Knowl. Data Eng.*, vol. 26, no. 1, pp. 235–252, Jan. 2014.

[25] H.-J. Li, L. Wang, Y. Zhang, and M. Perc, "Optimization of identifiability for efficient community detection," *New J. Phys.*, vol. 22, no. 6, Jun. 2020, Art. no. 063035.

[26] D. Jin, C. Huo, C. Liang, and L. Yang, "Heterogeneous graph neural network via attribute completion," in *Proc. Web Conf.*, 2021, pp. 391–400.

[27] Z. Yu *et al.*, "AS-GCN: Adaptive semantic architecture of graph convolutional networks for text-rich networks," in *Proc. IEEE Int. Conf. Data Mining*, 2021, pp. 837–846.

[28] D. Jin *et al.*, "A survey of community detection approaches: From statistical modeling to deep learning," *IEEE Trans. Knowl. Data Eng.*, to be published, doi: 10.1109/TKDE.2021.3104155.

[29] A. Lancichinetti, S. Fortunato, and J. Kertész, "Detecting the overlapping and hierarchical community structure in complex networks," *New J. Phys.*, vol. 11, no. 3, Mar. 2009, Art. no. 33015.

[30] J. Baumes, M. K. Goldberg, M. S. Krishnamoorthy, M. Magdon-Ismail, and N. Preston, "Finding communities by clustering a graph into overlapping subgraphs," in *Proc. IADIS Int. Conf. Appl. Comput.*, 2005, pp. 97–104.

[31] S. Zhang, R.-S. Wang, and X.-S. Zhang, "Identification of overlapping community structure in complex networks using fuzzy c-means clustering," *Physica A: Statist. Mechanics Appl.*, vol. 374, no. 1, pp. 483–490, Jan. 2007.

[32] J. Su and T. C. Havens, "Quadratic program-based modularity maximization for fuzzy community detection in social networks," *IEEE Trans. Fuzzy Syst.*, vol. 23, no. 5, pp. 1356–1371, Oct. 2015.

[33] M. Rosvall and C. T. Bergstrom, "Maps of random walks on complex networks reveal community structure," *Proc. Nat. Acad. Sci. USA*, vol. 105, no. 4, pp. 1118–1123, Jan. 2008.

[34] X. Zhu and Z. Ghahramani, "Learning from labeled and unlabeled data with label propagation," Tech Rep., vol. 3175, no. 2004, pp. 237–244, 2002.

[35] B. J. Frey and D. Dueck, "Clustering by passing messages between data points," *Science*, vol. 315, no. 5814, pp. 972–976, Feb. 2007.

[36] M. Mitrović and B. Tadić, "Spectral and dynamical properties in classes of sparse networks with mesoscopic inhomogeneities," *Phys. Rev. E*, vol. 80, no. 2, Aug. 2009, Art. no. 026123.

[37] J. Chen *et al.*, "GA-based Q-attack on community detection," *IEEE Trans. Comput. Soc. Syst.*, vol. 6, no. 3, pp. 491–503, Jun. 2019.

[38] J. Li, H. Zhang, Z. Han, Y. Rong, H. Cheng, and J. Huang, "Adversarial attack on community detection by hiding individuals," in *Proc. The Web Conf.*, 2020, pp. 917–927.

[39] R. Cazabet, F. Amblard, and C. Hanachi, "Detection of overlapping communities in dynamical social networks," in *Proc. IEEE 2nd Int. Conf. Soc. Comput./ IEEE Int. Conf. Privacy Secur., Risk Trust*, 2010, pp. 309–314.

[40] A. F. McDaid and N. J. Hurley, "Detecting highly overlapping communities with model-based overlapping seed expansion," in *Proc. Int. Conf. Adv. Soc. Netw. Anal. Mining*, 2010, pp. 112–119.

[41] J. Xie and B. K. Szymanski, "Community detection using A neighborhood strength driven label propagation algorithm," *IEEE Netw. Sci. Workshop*, pp. 188-195, Jun. 2011.

[42] P. Gopalan and D. M. Blei, "Efficient discovery of overlapping communities in massive networks," *Proc. Nat. Acad. Sci. USA*, vol. 110, no. 36, pp. 14 534–14 539, Sep. 2013.

[43] Y. Li, K. He, D. Bindel, and J. E. Hopcroft, "Uncovering the small community structure in large networks: A local spectral approach," in *Proc. 24th Int. Conf. World Wide Web*, 2015, pp. 658–668.

[44] A. Lancichinetti, F. Radicchi, J. J. Ramasco, and S. Fortunato, "Finding statistically significant communities in networks," *PLoS One*, vol. 6, no. 4, Apr. 2011, Art. no. e18961.

[45] A. Campan, Y. Alufaisan, and T. M. Truta, "Preserving communities in anonymized social networks," *Trans. Data Priv.*, vol. 8, no. 1, pp. 55–87, Dec. 2015.

[46] A. L. Traud, P. J. Mucha, and M. A. Porter, "Social structure of Facebook networks," *Physica A-Statist. Mechanics Appl.*, vol. 391, no. 16, pp. 4165–4180, Aug. 2012.

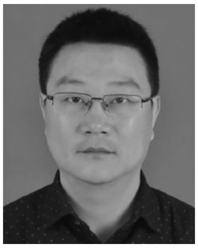
[47] H.-J. Li, W. Xu, S. Song, W.-X. Wang, and M. Perc, "The dynamics of epidemic spreading on signed networks," *Chaos Solitons Fractals*, vol. 151, Oct. 2021, Art. no. 111294.

[48] H.-J. Li, Z. Bu, Z. Wang, and J. Cao, "Dynamical clustering in electronic commerce systems via optimization and leadership expansion," *IEEE Trans. Ind. Informat.*, vol. 16, no. 8, pp. 5327–5334, Aug. 2020.

[49] A. F. McDaid, D. Greene, and N. J. Hurley, "Normalized mutual information to evaluate overlapping community finding algorithms," Oct. 2011, *arXiv:1110.2515*.

[50] L. Danon, A. Díaz-Guilera, J. Duch, and A. Arenas, "Comparing community structure identification," *J. Statist. Mechanics: Theory Exp.*, vol. 2005, no. 9, Sep. 2005, Art. no. 9008.

[51] M. Girvan and M. E. J. Newman, "Community structure in social and biological networks," *Proc. Nat. Acad. Sci. USA*, vol. 99, no. 12, pp. 7821–7826, Jun. 2002.



Dong Liu (Member, IEEE) received the BS and MS degrees in computer science from Zhengzhou University, Zhengzhou, China, in 1998 and 2004, respectively, and the PhD degree in computer science from Tianjin University, Tianjin, China, in 2013. He is currently a professor with the College of Computer and Information Engineering, Henan Normal University, Xinxiang, China. His research interests include educational data mining and complex network analysis.

Guoliang Yang is currently working toward the MS degree with the College of Computer and Information Engineering, Henan Normal University, Xinxiang, China. His research focuses on information hiding in complex networks.

Yanwei Wang is currently working toward the MS degree with the College of Computer and Information Engineering, Henan Normal University, Xinxiang, China. Her research focuses on information hiding in complex networks.

Hu Jin (Senior Member, IEEE) received the BE degree in electronic engineering and information science from the University of Science and Technology of China, Hefei, China, in 2004, and the MS and PhD degrees in electrical engineering from the Korea Advanced Institute of Science and Technology, Daejeon, South Korea, in 2006 and 2011, respectively. From 2011 to 2013, he was a postdoctoral fellow with The University of British Columbia, Vancouver, BC, Canada. From 2013 to 2014, he was a research professor with Gyeongsang National University, Tongyeong, South Korea. Since 2014, he has been with the Division of Electrical Engineering, Hanyang University, Ansan, South Korea, where he is an associate professor. His research interests include machine learning, Big Data, and wireless communications.

Enhong Chen (Senior Member, IEEE) received the PhD degree from the University of Science and Technology of China. He is currently a professor and vice dean with the School of Computer Science, University of Science and Technology of China. His research interests include data mining and machine learning, social network analysis, and recommender systems. His research is supported by the National Science Foundation for Distinguished Young Scholars of China. He was on the program committees of numerous conferences, including KDD, ICDM,and SDM.

▷ **For more information on this or any other computing topic, please visit our Digital Library at** www.computer.org/csdl.