

Article

Overlapping Community Hiding Method Based on Multi-Level Neighborhood Information

Guoliang Yang ¹, **Yanwei Wang** ¹, **Zhengchao Chang** ² and **Dong Liu** ^{1,3,4,*} ¹ College of Computer and Information Engineering, Henan Normal University, Xinxiang 453000, China² College of Computer Science & Technology, Henan Institute of Technology, Xinxiang 453000, China³ Key Laboratory of Artificial Intelligence and Personalized Learning in Education of Henan Province, Xinxiang 453000, China⁴ Big Data Engineering Lab of Teaching Resources & Assessment of Education Quality, Xinxiang 453000, China

* Correspondence: liudong@htu.edu.cn; Tel.: +86-13937322546

Abstract: The overlapping community detection algorithm divides social networks into multiple overlapping parts, and members can belong to multiple communities at the same time. Although the overlapping community detection algorithm can help people understand network topology, it exposes personal privacy. The BIH algorithm is proposed to solve the problem of personal privacy leaks in overlapping areas. However, some specific members in overlapping areas do not want to be discovered to belong to some specific community. To solve this problem, an overlapping community hiding algorithm based on multi level neighborhood information (MLNI) is proposed. The MLNI algorithm defines node probability of community based on multi-layer neighborhood information. By adjusting the probability of the target node belonging to each community, the difference between the probability that the target node belongs to outside and inside the target community is maximized. This process can be regarded as an optimization problem. In addition, the MLNI algorithm uses the genetic algorithm to find the optimal solution, and finally achieves the purpose of moving the target node in the overlapping area out of a specific community. The effectiveness of the MLNI algorithm is demonstrated through extensive experiments and baseline algorithms. The MLNI algorithm effectively realizes the protection of personal privacy in social networks.



Citation: Yang, G.; Wang, Y.; Chang, Z.; Liu, D. Overlapping Community Hiding Method Based on Multi-Level Neighborhood Information.

Symmetry **2022**, *14*, 2328. <https://doi.org/10.3390/sym14112328>

Academic Editors: Jia-Ning Luo and José Carlos R. Alcantud

Received: 7 September 2022

Accepted: 1 November 2022

Published: 5 November 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

There are various complex networks in the real world, which can be abstracted into a graph. For example, social networks [1,2] can be abstracted into an undirected graph. The nodes in the graph represent users, and the links represent the relationship between users. As an important tool for discovering graph structures, the community detection algorithm [3,4] has been applied in many fields, such as social networks [1,2], biological networks [5], power networks [6], and financial networks [7], etc. However, with the research of community detection algorithm, the problem of the privacy leak brought by it has also attracted people's attention [8,9]. For example, by analyzing the topology of the Facebook social network and the attributes of some Facebook users, we can not only understand the social situation of these users, but also mine some private information of these users.

To address these privacy leak issues, the community hiding algorithm has been proposed and studied as a symmetry problem of community detection [10,11]. Community hiding [12–17] uses a symmetry strategy of community detection to modify the network structure as little as possible, and realizing the hiding of target nodes, target communities, or the overall community structure. For the research on community hiding work, the

current community hiding algorithms focus on the non-overlapping communities, and the overlapping community hiding is just the beginning.

In our society, the individual can belong to multiple communities at the same time [18]. For example, one can join the badminton team and the table tennis team at the same time. If this man does not want his social relations to be discovered by detection algorithms to join the badminton team, he needs to use an overlapping hiding algorithm to hide his social relations. Therefore, the research of overlapping community hiding is of great importance. Liu et al. [19] studied overlapping community hiding for the first time, and proposed the overlapping community hiding algorithm BIH, which realized the purpose of moving the target node out of the overlapping area. However, some members in overlapping areas want to hide their identity of specific community. For example, in a social network, a person belongs to multiple communities at the same time, but he does not want it to be discovered that he belongs to a specific community. Therefore, an overlapping community hiding method based on multi-level neighborhood information (MLNI) is proposed. It achieves the purpose of moving specific nodes in overlapping areas out of specific communities by modifying the network structure as little as possible.

There are three difficulties encountered in designing the MLNI algorithm. The first one is how to reconstruct the network topology discovered by the overlapping community detection algorithm. To address it, this paper designs a method based on multi-layer neighborhood information to calculate the probability that a node belongs to a community, and reconstructs the community topology by adding certain constraints. The second is what kind of constraints to add. If a node belongs to a community, the probability of it belonging to this community should be greater than that of other communities, and the probability of nodes in the overlapping area belonging to each community should be similar. This paper uses the above information to establish constraints on the node probability calculation method, and uses gradient descent algorithm to obtain the weights of each layer's neighborhood, that is, the hidden embedding of each layer's neighborhood. The third problem is how to choose the optimal combination of adding and removing links. This is essentially an optimization problem. This paper uses the genetic algorithm [20–22] to find the optimal solution, and then finds the most suitable combination to modify, so as to achieve the purpose of moving the target node out of the target community.

For the study of the hidden problem of overlapping communities, our work makes the following contributions:

- We propose a new hiding algorithm that moves target nodes in overlapping areas out of a specific community.
- We introduce the probability of a node belonging to a community and change the probability by selecting appropriate links to operate.
- We conduct multiple experiments on five real social networks and compare the performance of the proposed hidden algorithm against four well-known overlapping community detection algorithms.

The rest of this paper is as follows: Section 2 describes the related work by community detection algorithms and community hiding algorithms. Section 3 proposes and analyzes the MLNI algorithm. In Section 4, the hidden effect of NPA is experimentally evaluated on several real networks. Section 5 concludes the NPA and outlines the future work.

2. Related Work

2.1. Community Detection

Community detection algorithms divide the network into multiple sub-structures through specific rules, which helps us better understand the community structure. In addition, it can be divided into overlapping community detection algorithms and non-overlapping community detection algorithms.

In the community structure divided by the non-overlapping community detection algorithm [23–28], each node can only belong to one community. Common non-overlapping

community detection algorithms include: Walktrap [29], GN algorithm [30], Infomap [31], GRE [32], Spinglass algorithm [33], etc.

The overlapping community detection [34] algorithm allows a node to belong to multiple communities at the same time. Common overlapping community detection algorithms include: CONGA [17], DEMON [35], CPM [36], LinkCoMM [37], UMSTMO [38], etc. The DEMON algorithm is a simple local-first community detection method that is proposed by Coscia. The LinkCoMM algorithm is a similarity division of links method that is proposed by Ahn et al. With a high degree of time complexity, the CPM algorithm uses the method of searching for the smallest clique in the neighborhood, finding the community structure composed of multiple overlapping and connected communities. The UMSTMO algorithm provides methods to explore the union of all Maximum Spanning Trees (UMST) and model the strength of links between nodes.

2.2. Community Hiding

Community hiding is a symmetry problem of community detection. It refers to hiding the network structure through specific strategies to avoid detection by community detection algorithms. Community hiding algorithms can be divided into non-overlapping community hiding algorithms and overlapping community hiding algorithms.

At present, the research on community hiding algorithms focuses on non-overlapping communities. Nagaraja [39] first introduced the community hiding problem, which did not attract attention at the time. With the in-depth discussion of the hidden scientific and practical significance of the community by Waniek et al. [12], the research had gradually attracted attention, and achieved certain research results, such as Ds algorithm [13], Q-Attack algorithm [14], REM [15], EPA algorithm [16], and so on. Waniek's algorithm and Q-Attack algorithm are based on modularity. Liu et al. [15] constructed information entropy based on the community to depict the community structure problem and identified the link that is most needed to be increased through the entropy residual error minimization (REM) algorithm to realize the global community hiding of the network. Chen et al. [16] constructed a formal representation based on node attacks and proposed the evolutionary perturbation attack (EPA) algorithm to realize the microcommunity hiding target. Chen et al. [40] defined a new community safety evaluation method to realize community hiding, which has a good effect.

However, it is very common for individuals to belong to multiple communities in a social network, so the study of overlapping community hiding has practical significance. Liu et al. [19] proposed BIH to move the target node out of the overlapping area, so as to achieve the purpose of hiding the target node, which is also our research orientation. The BIH algorithm moves the target node from the overlapping area to the target community. However, we study from other perspectives, aiming to move the target node out of a specific community, and the target node may still be in the overlapping area after being hidden.

3. Methods

3.1. Problem Formulation

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 links $E = \{e_1, e_2, e_3, \dots, e_M\}$, where a node represents an individual, and a link represents a social relationship between two connected nodes. There are N nodes and M links in the network G . Assume some overlapping community detection algorithm can partition the network into a set of communities, where each community may have nodes that overlap with other communities. The detected community structure represents $C = \{C_1, C_2, C_3, \dots, C_K\}$, where $C_i \subseteq V, i \in \{1, 2, \dots, K\}$, and at least one pair of i and j between 1 and K satisfies $C_i \cap C_j \neq \emptyset$.

If target node n is simultaneously included in several communities ($C_1, C_2, C_3, \dots, C_P$) after the overlapping community detection, we call node n as being in the overlapping area of $\cap C_{\{1, 2, \dots, P\}}$, which can also be represented by the notation of $n \in \cap C_{\{1, 2, \dots, P\}}$, and C_i is a

specific target community, which represents the community where n is currently located and is to be removed. The purpose of MLNI is to move the target node n out of a specific community C_i , where $i \in \{1, 2, \dots, P\}$. For the link addition operation in the execution process of MLNI, it is represented as E^+ , and the link deletion operation is represented as E^- . Table 1 shows detailed description of some symbols.

Table 1. Symbols used in this paper.

Symbol	Definition
$G = (V, E)$	original network with nodes V , links E .
C	the communities discovered by some community detection algorithms of G .
E^+, E^-	stands for adding and removal links in network.
n	the target node
C_i	the target community
$\cap C_{\{1,2,\dots,P\}}$	the set of nodes in the overlapping area of community C_1, C_2, \dots, C_p
ψ	the propagation function
W_i	the weight of each layer
σ	the aggregation function
$C_i(h_n)$	the probability of node n belonging to community C_i
$N_i(n)$	the neighbors of node n
$\deg(n)$	the degree of node n
$C_{other}(h_n)$	the probability that node n belongs to other communities except the target community

3.2. Overlapping Community Hiding Algorithm Based on Multi Level Neighborhood Information

The MLNI algorithm adjusts the probability that the target node belongs to the target community, so that the detection algorithm detects that the target node belongs to other communities (excluding target communities). Figure 1 shows an example of applying MLNI to move the individual out of the target community. For network G , we can obtain its community structure through the overlapping community detection algorithm (such as CPM [36], LinkCoMM [37], UMSTMO [38], and DEMON [35]). Assume that the red marked target node is in the overlapping area. The purpose of NPA is to move the target node out of a specific target community. The target community removed by the target node is the social relationship that the user does not want to be discovered. According to the combination of links obtained by the algorithm, after deleting or adding some connections, the detection algorithm is used to detect the updated network, and the target community does not contain the target node.

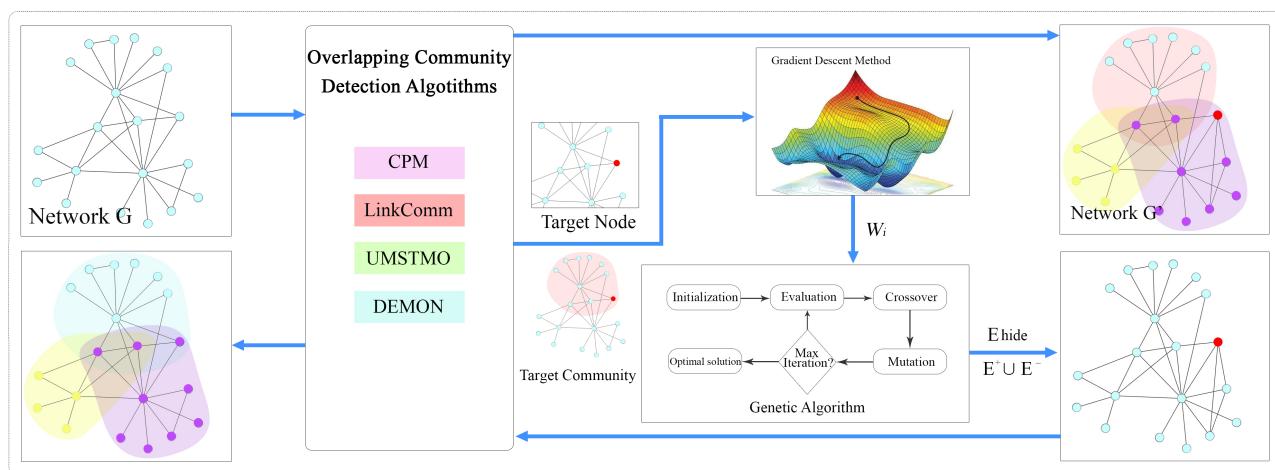


Figure 1. Application of an MLNI algorithm to hide an example of target nodes in overlapping areas.

The realization of the pseudo-code is shown in Algorithm 1.

Algorithm 1 Overlapping community hiding method based on Multi-Level Neighborhood Information (MLNI).

Input: Network G , T , Target node n , $PopSize$, $Generation$;

Output: Updated Network G' ;

```

 $C \leftarrow \text{getCommunities}(G);$ 
 $[n_1, n_2, \dots, n_p] \leftarrow \text{getNodesInOverlappingArea}(C);$ 

if  $n \in [n_1, n_2, \dots, n_p]$  then
     $[C_1, C_2, \dots, C_K] \leftarrow \text{getCommunitiesOfNode}(C, n);$ 
     $C_i \leftarrow \text{ChooseTargetCommunity}();$ 
     $W \leftarrow \text{GetWeightOfCommunity}([C_1, C_2, \dots, C_K])$ 
     $C(h(n)) \leftarrow \text{GetProbability}(n, [C_1, C_2, \dots, C_K])$ 
     $\text{ParentPop} \leftarrow \text{Initialization}(G, PopSize, T);$ 
    while  $i < Generation$  do
         $\text{SelectedPop} \leftarrow \text{Selection}(\text{ParentPop}, PopSize);$ 
         $\text{CrossoverPop} \leftarrow \text{Crossover}(\text{SelectedPop}, PopSize, T);$ 
         $\text{OffspringPop} \leftarrow \text{Elistism}(\text{MutationPop}, \text{ParentPop});$ 
         $\text{ParentPop} \leftarrow \text{OffspringPop};$ 
         $i \leftarrow i + 1;$ 
    end
     $G' \leftarrow \text{Update } G;$ 
end

```

3.3. Probability (the Node Belongs to the Community)

In the MLNI community hiding algorithm, the probability of the node in the community is defined by using multi-layer neighborhood information of the node. Assuming that a target node n is selected from the network, the community C_i to which it belongs can be determined by the overlapping community detection method.

The formula for calculating the probability that node n belongs to the community C_i is as follows:

$$C_i(h_n) = \sigma \left(\sum W_i \sum_{v \in N_i(n)} \psi(C_i(h_v)) \right), \quad (1)$$

where σ is the aggregation function; the function \tanh is selected to aggregate the information transmitted by the nodes in the neighborhood; ψ is the propagation function, which is used to calculate the amount of information transmitted by a single node in the neighborhood. In our algorithm, we define $\psi = C_i(h_v) / \deg(v)$, where $\deg(v)$ is degree of node v , and $C_i(h_v)$ is the probability of node v in the community C_i . W_i is a vector representing the weight of each level. If W_i is less than 0, it is meaningless, so the value must be greater than 0. $N_i(n)$ represents the i th level neighbor node of n .

3.4. Restrictions

The MLNI algorithm obtains the value of W_i by reconstructing the known community structure, and the function needs to satisfy certain constraints during the reconstruction process. If the node n is in the community C_i , the influence of the nodes in the community C_i on n will be greater than that outside the community. For the nodes in the overlapping area, the probability of belonging to each community should be similar, and, in order to accurately describe the degree of similarity, this paper introduces a threshold variable α . From the above discussion, the following constraints can be obtained:

$$\begin{cases} W_i > 0 \\ C_i(h_n) > C_{k/i}(h_n), & C_k \in C \\ |C_i(h_n) - C_j(h_n)| < \alpha, & n \in C_i, C_j \end{cases} \quad (2)$$

In the MLNI algorithm, it is hoped that the probability that the node n in the overlapping area belongs to each community is as close as possible. That is, $|C_i(h_n) - C_j(h_n)|$ is as small as possible, and at the same time, in order to improve the performance of the algorithm, this paper sets the variable α to control the degree of similarity. The gradient descent algorithm is a very widely used optimization algorithm in machine learning, which is used to minimize (or maximize) the objective function. Although the optimal value cannot be obtained in each iteration, the final result is near the global optimal solution. Therefore, the MLNI algorithm chooses the gradient descent method, and at the same time determines the value of W_i according to the constraints of Equation (2).

After determining the value of W , the probability of n belonging to each community can be obtained by using Equation (1). This paper hopes to move the target node n out of the target community C_i by adjusting the probability of n in different communities, that is, to maximize the difference between the probability that the target node n belongs to the target community and other communities. This is an optimization problem. Since the target node may belong to more than two communities at the same time, we define the calculation formula of the probability of n in other communities except the target community C_i as shown in Equation (3):

$$C_{\text{other}}(h_n) = \frac{\sum_{k \in \text{overlap}(n) / C_i} C_k(h_n)}{|\text{overlap}(n)| - 1} \quad (3)$$

where C_k represents the other communities except the target community, and $\text{overlap}(n)$ represents the number of communities to which n belongs at the same time.

3.5. Optimization Algorithm GA

Genetic Algorithm (GA) is a method of searching for the optimal solution, which simulates the natural selection of Darwin's theory of biological evolution and the biological evolution process of genetic mechanism. Due to the good global optimization ability of the genetic algorithm, it adopts a probabilistic optimization method, which can automatically obtain and guide the optimized search space and adjust the search direction adaptively when no rules need to be determined. Therefore, MLNI chooses the genetic algorithm as the optimization algorithm, and its fitness function is as follows:

$$\rho = C_{\text{other}}(h_n) - C_i(h_n) \quad (4)$$

The fitness function ρ is the difference between probability of nodes within and out of the target community. The genetic algorithm is used to maximize ρ , so as to determine the combination of adding and removing links.

The genetic algorithm selects the appropriate link for modification through the following steps, and the specific implementation process is as follows:

Encoding: GA represents the solution data as chromosomes in the genetic space. In this paper, the number of links added and deleted in each chromosome is not limited, while the maximum length of the chromosome is fixed. We hope to use the smallest disturbance, that is, the shortest gene length to achieve the purpose of moving the target node out of the target community.

Initialization: The initial population is randomly generated. Each chromosome is a combination of adding and deleting links, and the GA starts to iterate with these N chromosomes as the initial point.

Selection: Selection reflects Darwin's principle of survival of the fittest, and the next generation of individuals is selected according to the principle of survival of the fittest. The purpose of selection is to select excellent individuals from the current group so that they have the opportunity to reproduce as parents. In real-world situations, individuals with higher fitness have a greater chance of surviving and reproducing. In MLNI, the function of Equation (4) is chosen as the loss function. After evaluating each individual, a method of roulette selection is further used to select offspring. The individual selection probability is as follows:

$$P_{(i)} = \frac{\rho(i)}{\sum_{j=1}^M (\rho(j))} \quad (5)$$

where M is the number of chromosomes.

Crossover: Generally, chromosomes are of the same length and have been fixed throughout evolution. However, in the MLNI, non-equal length crossover is used. During the crossover process, the length of the chromosome can be changed, but the longest cannot exceed the limited length, so that the optimal solution can be found with the smallest budget. The specific steps are as follows: select two chromosomes O_i and O_j to form two exchangeable genomes with additions and deletions. E_{ai} and E_{di} represent the exchangeable genomes added and deleted in O_i , and E_{aj} and E_{dj} represent the exchangeable genomes in O_j by additions and deletions. If the sum of the added and deleted values in chromosomes O_i and O_j is greater than the maximum length of the chromosome, the number of crossover genes is re-determined; otherwise, the crossover operation is performed.

Mutation: Because crossover cannot generate new individuals, which can only be generated by mutation, it is mutation that prevents the solution from falling into local optimum. In addition, the mutation probability is denoted by P_m . In this paper, a fixed P_m is selected, and an appropriate P_m value is selected through testing.

4. Experiments

4.1. Datasets

In this experiment, we conducted experiments on five real data sets. A detailed description of the datasets is as follows:

Football Network [30]: The network represents the 2000 season between American soccer teams. Each node represents a college team, and each link between two nodes represents at least one game between two teams.

Dolphin social network (Dolphins) [41]: It is documented over 7 years by lusseau et al., representing the social relationships of 62 dolphins in the New Zealand Gulf.

Karate Social Network [42]: This is a dataset of karate club memberships recorded by Zachary, consisting of 34 nodes and 78 links.

Political Book Network (Political) [43]: The network consists of 105 nodes and 441 links, in which each node in the network represents a book on a political topic, and the link between the two books indicates that they were purchased by the same consumer.

Facebook [44]: It is a social network among users on Facebook, consisting of a Facebook friend list. The basic attribute statistics of the dataset are shown in Table 2.

Email-Enron: Enron email communication network covers all the email communication within a dataset of around half million emails. These data were originally made public, and posted to the web, by the Federal Energy Regulatory Commission during its investigation.

Table 2. Data set information.

Network	Nodes	Links	Description
Football ¹	115	613	American football teams
Karate ¹	34	78	Zachary Karate's Club
Dolphins ¹	62	159	Dolphins association
Political ¹	105	441	Books about US politics
Facebook ²	4390	88,243	Facebook social network
Email-Enron ²	36,693	183,831	Email communication network from Enron

¹: <http://www-personal.umich.edu/~mejn/netdata> accessed on 10 December 2020. ²: <http://snap.stanford.edu/data> accessed on 10 December 2020.

To verify the effectiveness of the algorithm, we selected four classic overlapping community detection algorithms, which are detailed as follows:

CPM [36]: The CPM algorithm was proposed by Palla in 2005 and was published in the journal Nature. It uses the method of searching for the smallest clique in the neighborhood,

and finds the community structure composed of multiple overlapping and connected communities.

LinkCoMM [37]: It was proposed by Ahn and was published in Nature in 2010. The algorithm uses hierarchical clustering with link (or link) similarity to build a dendrogram, where each leaf is a link from the original network, while branches represent communities. Then, the dendrogram is cut with a density function to obtain a community structure with overlapping and hierarchical structures.

UMSTMO [38]: UMSTMO provides a new method to explore the union of all Maximum Spanning Trees (UMST) and model the strength of links between nodes. Furthermore, each node in UMST is connected to its most similar neighbor. The model extracts local communities for each node, and then combines the generated communities according to the number of shared nodes.

DEMON [35]: This is a simple local-first community detection method capable of revealing community structures in real complex networks. DEMON democratically allows the community around each node to vote by using a label propagation algorithm; and finally merges the local communities into a complete community structure.

SLPA [45]: SLPA is general framework to detect and analyze both individual overlapping nodes and entire communities. In this framework, nodes exchange labels according to dynamic interaction rules.

4.2. Evaluation Metric

The evaluation indicators for overlapping community detection algorithms include ONMI, Omega Index, F-Score, and Precision. These indicators are used to evaluate the network topology obtained by a detection algorithm. The purpose of MLNI algorithm is to hide a node from a community. This has very little impact on the network topology, especially on large datasets. The BIH algorithm proposes an evaluation method to evaluate the hiding efficiency of node. However, the purpose of MLNI algorithm and BIH algorithm is different. The MLNI algorithm is to move the target node out of the target community, while the BIH algorithm is to move the target node from the overlapping area into the target community. To evaluate the effectiveness of MLNI, we design an evaluation index a to evaluate the hidden effect of MLNI, which can be calculated by Equation (6):

$$A(n, C_i) = \left(1 - \frac{\sum_{i \in [C_1, C_2, \dots, C_k]} \frac{|S(C_i)| - 1}{|C_i| - 1}}{|[C_1, C_2, \dots, C_k]|}\right) \times \left(1 - \frac{|O_n|}{|C_i|}\right) \quad (6)$$

where $|S(C_i)|$ is the connected component of the target community C_i after the hidden algorithm is executed; $|C_i|$ is the number of nodes in the target community; $\{C_1, C_2, \dots, C_k\}$ is the community containing the target node n ; and $|O_n|$ is the number of nodes in the target community that are still in the same community as node n after the hidden algorithm is executed. The first part of the formula is to ensure the connectivity of the community after the hiding algorithm is executed, and the second part is used to evaluate the efficiency of hiding.

The value range of $A(n, C_i)$ is $[0, 1]$. The larger the value, the smaller the connection between the target node and the target community. In other words, it means that the number of nodes belonging to the same community in the target node and the target community is less. The larger the value, the better the effect of MLNI, and the less likely it is to be hidden. In particular, when $A(n, C_i)$ is equal to 1, it means that the number of nodes in the same community between the target node n and the target community after MLNI is executed is 0, and it is removed from the target community. At this time, the hiding effect is the best.

4.3. Baseline Algorithms

In this study, two baseline methods were selected, namely random hiding strategy (RH) and node degree-based hiding strategy (DH).

Random Hiding Strategy (RH): Randomly select a node u in the target community C_i (there is a link between u and target node n), delete the link between u and n , and

then randomly add links in other communities containing n . Although the RH is unstable, it does not require prior knowledge of the community structure when operating. The algorithm implementation pseudo code is shown in Algorithm 2.

Algorithm 2 Random hiding strategy.

Input: Network G, T , Target node n ;
Output: Updated Network G' ;

```

 $C \leftarrow \text{getCommunities}(G);$ 
 $[n_1, n_2, \dots, n_p] \leftarrow \text{getNodesInOverlappingArea}(C);$ 

if  $n \in [n_1, n_2, \dots, n_p]$  then
     $[C_1, C_2, \dots, C_K] \leftarrow \text{getCommunitiesOfNode}(C, n);$ 
     $C_i \leftarrow \text{ChooseTargetCommunity}();$ 
    while  $T > 0$  do
         $N_{C_i}^n \leftarrow \text{getNonneighborSet}(C_i, n);$ 
         $u \leftarrow \text{RandomChooseNode}(N_{C_i}^n);$ 
        add link  $e(n, v)$  to  $E$ ;
         $C_j \leftarrow \text{getNeighborSet}(C_j, n);$ 
         $v \leftarrow \text{RandomChooseNode}(N_{C_j}^n);$ 
        remove link  $e(n, v)$  from  $E$ ;
         $T = T - 1$ 
    end
     $G' \leftarrow \text{Update } G;$ 
end

```

Node degree-based hiding strategy (DH): In social activities, everyone has different social relationships. Compared with nodes with relatively small degrees, individuals with larger degrees have connections with individuals in the community. Different from the RH algorithm, DH needs to understand the topology of the network in order to select the node with the largest degree to add or delete links. The specific implementation is shown in Algorithm 3.

Algorithm 3 Base-degree hiding strategy.

Input: Network G, T , Target node n ;
Output: Updated Network G' ;

```

 $C \leftarrow \text{getCommunities}(G);$ 
 $[n_1, n_2, \dots, n_p] \leftarrow \text{getNodesInOverlappingArea}(C);$ 

if  $n \in [n_1, n_2, \dots, n_p]$  then
     $[C_1, C_2, \dots, C_K] \leftarrow \text{getCommunitiesOfNode}(C, n);$ 
     $C_i \leftarrow \text{ChooseTargetCommunity}();$ 
    while  $T > 0$  do
         $N_{C_i}^n \leftarrow \text{getNonneighborSet}(C_i, n);$ 
         $u \leftarrow \text{ChooseNodeBaseDegree}(N_{C_i}^n);$ 
        add link  $e(n, v)$  to  $E$ ;
         $C_j \leftarrow \text{getNeighborSet}(C_j, n);$ 
         $v \leftarrow \text{ChooseNodeBaseDegree}(N_{C_j}^n);$ 
        remove link  $e(n, v)$  from  $E$ ;
         $T = T - 1$ 
    end
     $G' \leftarrow \text{Update } G;$ 
end

```

4.4. Result Analysis

We used five famous social network datasets and four classical overlapping community detection algorithms for experiments. The experimental results show that MLNI is effective in hiding the target node from the target community, and it is detected that the target node does not belong to the target community in most cases after using it. In addition, through the experimental comparison, it can be found that, before and after MLNI is hidden, the more complex communities have less obvious changes in the community structure, and it also shows that MLNI can hide the target node by modifying the community structure as little as possible.

Figure 2 shows the changes of the evaluation index $A(n, C_i)$ on different social networks. In each subgraph, the values of evaluation index $A(n, C_i)$ of three different hidden algorithms are shown under different budgets. Through testing, it can be concluded that MLNI can achieve relatively high $A(n, C_i)$ values in most cases. In summary, it can be seen that the performance of MLNI algorithm is better than other hidden methods.

In addition, we can observe that the CPM algorithm is very sensitive to the community hiding algorithm, and only needs to modify a small number of links to achieve the hiding effect. This is because CPM is an algorithm based on the minimum clique, and only needs to cut off a small number of links to break the minimum clique structure, which in turn affects the performance of the CPM overlapping community detection algorithm. As shown in the figure, the RH algorithm exhibits strong randomness because it randomly selects links to add or delete. The DH algorithm selects nodes according to degree to add or delete links. Its efficiency on Footbook + CPM is better than MLNI, but, in most cases, it is contrary. There are also some special cases where MLNI does not show a good hidden effect, such as Political + LinkComm, Facebook + LinkComm.

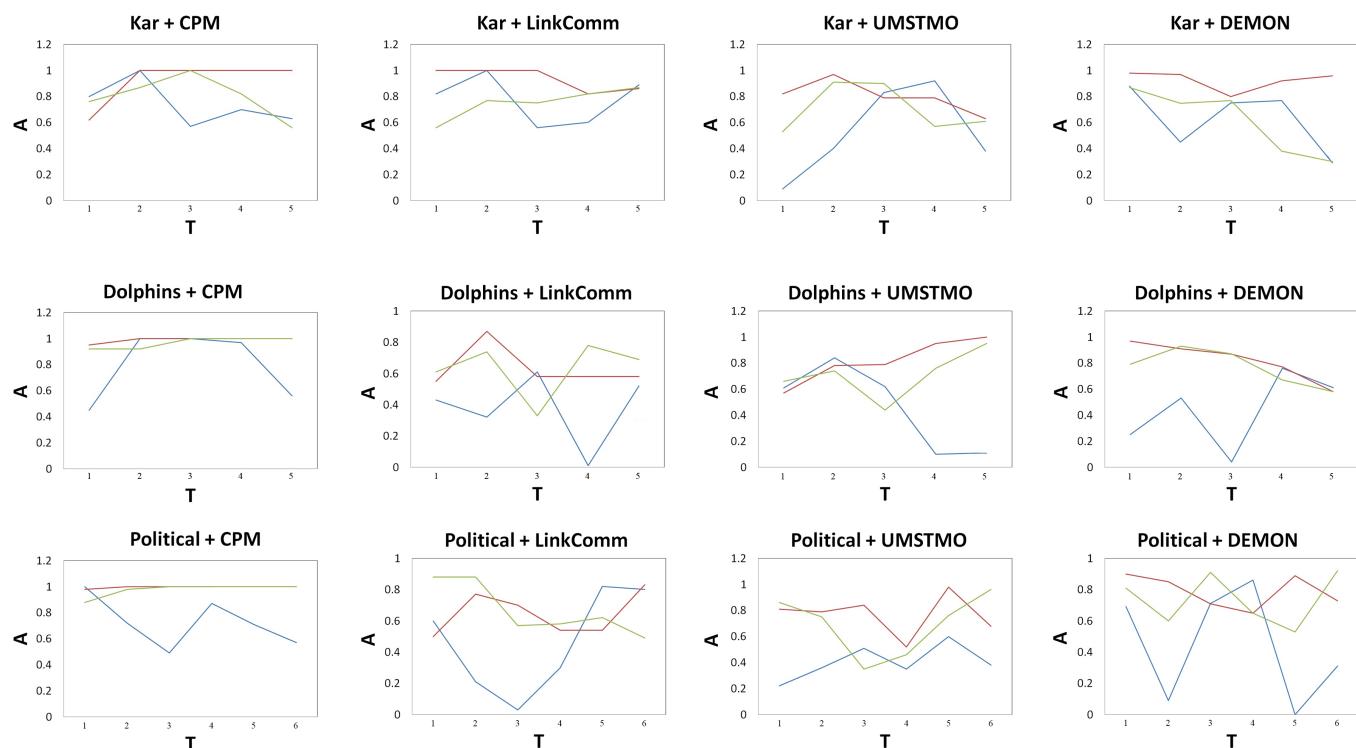


Figure 2. Cont.

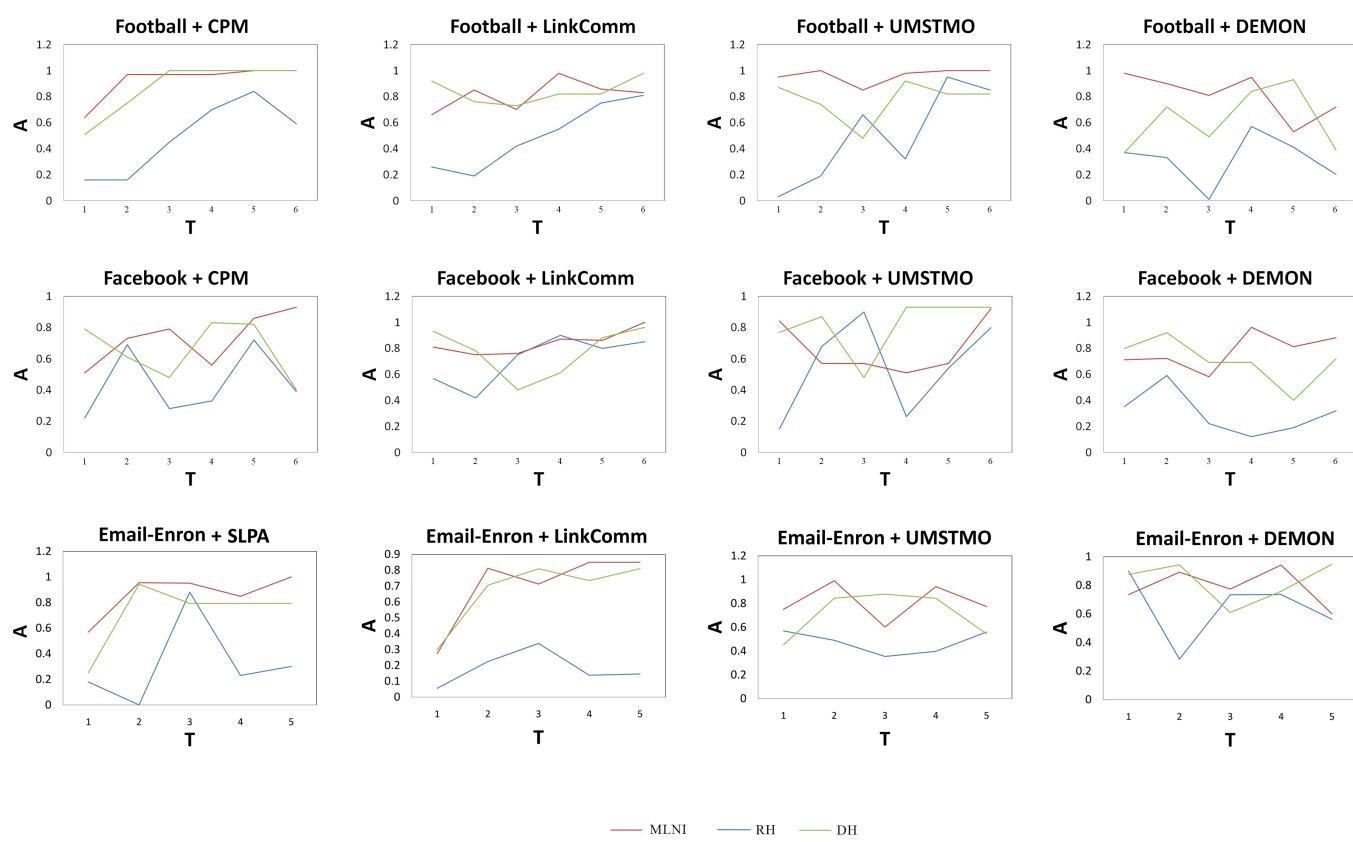


Figure 2. For different community detection algorithms and social networks, the change of $A(n, C_i)$, the T value is the total number of added and deleted links.

5. Conclusions and Future Work

Based on the needs of special populations, we propose an overlapping community hiding algorithm MLNI based on multi-layer neighborhood information. In reality, individuals in overlapping areas need to avoid the discovery of information about their specific communities. In this paper, the node probability is designed to calculate the probability that a node belongs to a community in the network. In addition, the gradient descent method is used to determine the weight vector W_i of each level and the probability that the node belongs to each community. A genetic algorithm is used to find the optimal combination of adding and removing links and generate a new network. In addition, the community overlap detection algorithm to detect that the target node does not belong to the target community under the new network. To evaluate the hidden effect of MLNI, the hidden evaluation index $A(n, C_i)$ is introduced. In addition, through a large number of experiments, the effectiveness of MLNI in removing the target node from the target community is proved. The MLNI algorithm effectively realizes the personal privacy protection for overlapping community detection algorithms in a social network, and avoids the personal privacy information being discovered by detection algorithms.

For the hidden work of overlapping communities, the proposed method is based on the static overlapping community detection algorithm. In future work, we will extend the original algorithm to adapt to the dynamic community detection algorithm. In addition, we will also focus on personal privacy protection of overlapping community detection algorithms. Compared with the hiding of nodes under overlapping communities, the hiding of a single community is more challenging, and it can effectively prevent specific organizations from being discovered, thereby avoiding the exposure of organizational structures.

Author Contributions: Conceptualization, Y.W., D.L. and G.Y.; methodology, G.Y.; software, Y.W., Z.C. and G.Y.; validation, G.Y.; formal analysis, Y.W., D.L. and G.Y.; investigation, Y.W., D.L. and G.Y.; resources, D.L.; data curation, Y.W. and G.Y.; writing—original draft preparation, G.Y.; writing—review and editing, D.L. and G.Y.; visualization, G.Y.; supervision, D.L.; project administration, Y.W., D.L. and G.Y.; funding acquisition, D.L. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the National Natural Science Foundation of China (62072160) and key scientific and technical project of Henan Province (212102310381).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest.

References

- Xuan, Q.; Zhang, Z.; Fu, C.; Hu, H.; Filkov, V. Social Synchrony on Complex Networks. *IEEE Trans. Cybern.* **2018**, *48*, 1420–1431. [[CrossRef](#)]
- Xuan, Q.; Fang, H.; Fu, C.; Filkov, V. Temporal motifs reveal collaboration patterns in online task-oriented networks. *Phys. Rev. E* **2015**, *91*, 052813. [[CrossRef](#)]
- Abbasi, A.A.; Younis, M.F. A survey on clustering algorithms for wireless sensor networks. *Comput. Commun.* **2007**, *30*, 2826–2841. [[CrossRef](#)]
- Vieira, V.d.F.; Xavier, C.R.; Evsukoff, A.G. A comparative study of overlapping community detection methods from the perspective of the structural properties. *Appl. Netw. Sci.* **2020**, *5*, 51. [[CrossRef](#)]
- Garcia, J.O.; Ashourvan, A.; Muldoon, S.; Vettell, J.M.; Bassett, D.S. Applications of Community Detection Techniques to Brain Graphs: Algorithmic Considerations and Implications for Neural Function. *Proc. IEEE* **2018**, *106*, 846–867. [[CrossRef](#)]
- Chen, Z.; Wu, J.; Xia, Y.; Zhang, X. Robustness of Interdependent Power Grids and Communication Networks: A Complex Network Perspective. *IEEE Trans. Circuits Syst. II Express Briefs* **2018**, *65*, 115–119. [[CrossRef](#)]
- Schiavo, S.; Reyes, J.; Fagiolo, G. International trade and financial integration: A weighted network analysis. *Quant. Financ.* **2010**, *10*, 389–399. [[CrossRef](#)]
- Gross, R.; Acquisti, A. Information Revelation and Privacy in Online Social Networks. In Proceedings of the 2005 ACM Workshop on Privacy in the Electronic Society (WPES’05), Alexandria, VA, USA, 7 November 2005; Association for Computing Machinery: New York, NY, USA, 2005; pp. 71–80.
- Zhang, C.; Sun, J.; Zhu, X.; Fang, Y. Privacy and security for online social networks: Challenges and opportunities. *IEEE Netw.* **2010**, *24*, 13–18. [[CrossRef](#)]
- Zhou, B.; Pei, J.; Luk, W. A brief survey on anonymization techniques for privacy preserving publishing of social network data. *ACM Sigkdd Explor. Newsl.* **2008**, *10*, 12–22. [[CrossRef](#)]
- Kearns, M.; Roth, A.; Wu, Z.S.; Yaroslavtsev, G. Private algorithms for the protected in social network search. *Proc. Natl. Acad. Sci. USA* **2016**, *113*, 913–918. [[CrossRef](#)]
- Waniek, M.; Michalak, T.P.; Rahwan, T.; Wooldridge, M.J. Hiding Individuals and Communities in a Social Network. *Nat. Hum. Behav.* **2018**, *2*, 139–147. [[CrossRef](#)]
- Fionda, V.; Pirrò, G. Community Deception or: How to Stop Fearing Community Detection Algorithms. *IEEE Trans. Knowl. Data Eng.* **2018**, *30*, 660–673. [[CrossRef](#)]
- Chen, J.; Chen, L.; Chen, Y.; Zhao, M.; Yu, S.; Xuan, Q.; Yang, X. GA-Based Q-Attack on Community Detection. *IEEE Trans. Comput. Soc. Syst.* **2019**, *6*, 491–503. [[CrossRef](#)]
- Liu, Y.; Liu, J.; Zhang, Z.; Zhu, L.; Li, A. Rem: From structural entropy to community structure deception. *Adv. Neural Inf. Process. Syst.* **2019**, *32*, 12938–12948.
- Chen, J.; Chen, Y.; Chen, L.; Zhao, M.; Xuan, Q. Multiscale Evolutionary Perturbation Attack on Community Detection. *IEEE Trans. Comput. Soc. Syst.* **2021**, *8*, 62–75. [[CrossRef](#)]
- Gregory, S. An Algorithm to Find Overlapping Community Structure in Networks. In Proceedings of the Knowledge Discovery in Databases: PKDD 2007, Warsaw, Poland, 17–21 September 2007; Kok, J.N., Koronacki, J., Lopez de Mantaras, R., Matwin, S., Mladenić, D., Skowron, A., Eds.; Springer: Berlin/Heidelberg, Germany, 2007; pp. 91–102.
- Xie, J.; Kelley, S.; Szymanski, B.K. Overlapping community detection in networks: The state-of-the-art and comparative study. *ACM Comput. Surv.* **2013**, *45*, 43:1–43:35. [[CrossRef](#)]
- Liu, D.; Yang, G.; Wang, Y.; Jin, H.; Chen, E. How to Protect Ourselves From Overlapping Community Detection in Social Networks. *IEEE Trans. Big Data* **2022**, *8*, 894–904. [[CrossRef](#)]

20. Lehnerer, S. Community Detection in Complex Networks using Genetic Algorithms. In Proceedings of the SKILL 2018—Studierendenkonferenz Informatik, Berlin, Germany, 26–27 September 2018; Becker, M., Ed.; Gesellschaft für Informatik e.V.: Bonn, Germany, 2018; pp. 35–46.
21. Liu, H.; Hu, X.B.; Yang, S.; Zhang, K.; Di Paolo, E. Application of complex network theory and genetic algorithm in airline route networks. *Transp. Res. Rec.* **2011**, *2214*, 50–58. [[CrossRef](#)]
22. Wang, S.; Zou, H.; Sun, Q.; Zhu, X.; Yang, F. Community detection via improved genetic algorithm in complex network. *Inf. Technol. J.* **2012**, *11*, 384. [[CrossRef](#)]
23. Liu, D.; Duan, D.; Shikai, S.; Song, G. Effective Semisupervised Community Detection Using Negative Information. *Math. Probl. Eng.* **2015**, *2015*, 109671. [[CrossRef](#)]
24. Liu, D.; Liu, X.; Wang, W.; Bai, H. Semi-supervised community detection based on discrete potential theory. *Phys. A Stat. Mech. Its Appl.* **2014**, *416*, 173–182. [[CrossRef](#)]
25. Liu, D.; Bai, H.Y.; Li, H.J.; Wang, W.J. Semi-supervised community detection using label propagation. *Int. J. Mod. Phys. B* **2014**, *28*, 1450208. [[CrossRef](#)]
26. Fan, L.; Xu, S.; Liu, D.; Ru, Y. Semi-Supervised Community Detection Based on Distance Dynamics. *IEEE Access* **2018**, *6*, 37261–37271. [[CrossRef](#)]
27. Liu, D.; Wang, C.; Jing, Y. Estimating the optimal number of communities by cluster analysis. *Int. J. Mod. Phys. B* **2016**, *30*, 1650037. [[CrossRef](#)]
28. Liu, D.; Chang, Z.; Yang, G.; Chen, E. Community hiding using a graph autoencoder. *Knowl.-Based Syst.* **2022**, *253*, 109495. [[CrossRef](#)]
29. Pons, P.; Latapy, M. Computing Communities in Large Networks Using Random Walks. In Proceedings of the Computer and Information Sciences—ISCIS 2005, Istanbul, Turkey, 26–28 October 2005; Yolum, P., Güngör, T., Gürgen, F., Özeturan, C., Eds.; Springer: Berlin/Heidelberg, Germany, 2005; pp. 284–293.
30. Girvan, M.; Newman, M.E. Community structure in social and biological networks. *Proc. Natl. Acad. Sci. USA* **2002**, *99*, 7821–7826. [[CrossRef](#)]
31. Rosvall, M.; Bergstrom, C.T. Maps of random walks on complex networks reveal community structure. *Proc. Natl. Acad. Sci. USA* **2008**, *105*, 1118–1123. [[CrossRef](#)]
32. Clauset, A.; Newman, M.E.J.; Moore, C. Finding community structure in very large networks. *Phys. Rev. E Stat. Nonlinear Soft Matter Phys.* **2004**, *6*, 66–111. [[CrossRef](#)]
33. Reichardt, J.; Bornholdt, S. Statistical mechanics of community detection. *Phys. Rev. E* **2006**, *74*, 016110. [[CrossRef](#)]
34. Gao, R.; Li, S.; Shi, X.; Liang, Y.; Xu, D. Overlapping Community Detection Based on Membership Degree Propagation. *Entropy* **2021**, *23*, 15. [[CrossRef](#)]
35. Coscia, M.; Rossetti, G.; Giannotti, F.; Pedreschi, D. DEMON: A local-first discovery method for overlapping communities. In Proceedings of the 18th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD’12), Beijing, China, 12–16 August 2012; ACM: New York, NY, USA, 2012; pp. 615–623.
36. Palla, G.; Derényi, I.; Farkas, I.; Vicsek, T. Uncovering the overlapping community structure of complex networks in nature and society. *Nature* **2005**, *435*, 814–818. [[CrossRef](#)] [[PubMed](#)]
37. Ahn, Y.Y.; Bagrow, J.P.; Lehmann, S. Link communities reveal multiscale complexity in networks. *Nature* **2010**, *466*, 761. [[CrossRef](#)] [[PubMed](#)]
38. Asmi, K.; Lotfi, D.; El Marraki, M. Overlapping community detection based on the union of all maximum spanning trees. *Library Hi Tech* **2020**, *38*, 276–292. [[CrossRef](#)]
39. Nagaraja, S. The impact of unlinkability on adversarial community detection: Effects and countermeasures. In Proceedings of the 10th International Conference on Privacy Enhancing Technologies (PETS’10), Berlin, Germany, 21–23 July 2010; pp. 253–272.
40. Chen, X.; Jiang, Z.; Li, H.; Ma, J.; Yu, P.S. Community Hiding by Link Perturbation in Social Networks. *IEEE Trans. Comput. Soc. Syst.* **2021**, *8*, 704–715. [[CrossRef](#)]
41. Lusseau, D.; Schneider, K.; Boisseau, O.J.; Haase, P.; Slooten, E.; Dawson, S.M. The bottlenose dolphin community of Doubtful Sound features a large proportion of long-lasting associations. *Behav. Ecol. Sociobiol.* **2003**, *54*, 396–405. [[CrossRef](#)]
42. Zachary, W.W. An information flow model for conflict and fission in small groups. *J. Anthropol. Res.* **1977**, *33*, 452–473. [[CrossRef](#)]
43. Newman, M. Modularity and community structure in networks. *Proc. Natl. Acad. Sci. USA* **2006**, *103*, 8577–8582. [[CrossRef](#)]
44. Traud, A.L.; Mucha, P.J.; Porter, M.A. Social structure of facebook networks. *Phys. A Stat. Mech. Its Appl.* **2012**, *391*, 4165–4180. [[CrossRef](#)]
45. Xie, J.; Szymanski, B.K.; Liu, X. SLPA: Uncovering Overlapping Communities in Social Networks via A Speaker-listener Interaction Dynamic Process. In Proceedings of the 2011 IEEE 11th International Conference on Data Mining Workshops (ICDMW), Vancouver, BC, Canada, 11 December 2011.