

Article

# Uncovering Hidden Community Structure in Multi-Layer Networks

Faiza Riaz Khawaja , Jinfang Sheng, Bin Wang \*  and Yumna Memon

School of Computer Science and Engineering, Central South University, Changsha 410083, Hunan, China; faiza123@csu.edu.cn (F.R.K.); jfsheng@csu.edu.cn (J.S.); yumna.memon@gmail.com (Y.M.)

\* Correspondence: wb\_csut@csu.edu.cn

**Abstract:** Community detection, also known as graph clustering, in multi-layer networks has been extensively studied in the literature. The goal of community detection is to partition vertices in a network into densely connected components so called communities. Networks contain a set of strong, dominant communities, which may interfere with the detection of weak, natural community structure. When most of the members of the weak communities also belong to stronger communities, they are extremely hard to be uncovered. We call the weak communities the hidden or disguised community structure. In this paper, we present a method to uncover weak communities in a network by weakening the strength of the dominant structure. With the aim to detect the weak communities, through experiments, we observe real-world networks to answer the question of whether real-world networks have hidden community structure or not. Results of the hidden community detection (HCD) method showed the great variation in the number of communities detected in multiple layers when compared with the results of other community detection methods.

**Keywords:** community detection; hidden communities; multi-layer network



**Citation:** Khawaja, F.R.; Sheng, J.; Wang, B.; Memon, Y. Uncovering Hidden Community Structure in Multi-Layer Networks. *Appl. Sci.* **2021**, *11*, 2857.

<https://doi.org/10.3390/app11062857>

Academic Editors:

Arcangelo Castiglione and  
Eui-Nam Huh

Received: 3 February 2021

Accepted: 19 March 2021

Published: 23 March 2021

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



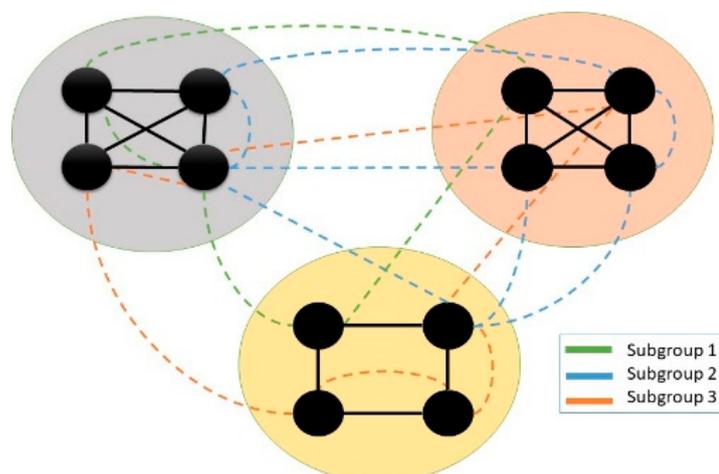
**Copyright:** © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

Uncovering community structures of a complex network can help us to understand how the network functions. On the other hand, finding communities and then attempting to analyze them is a compelling approach to understand all kinds of network organization structures and their functions. These communities often correspond to some functional units. Over the past decades, the problem of finding community has received significant research attention in the field of network analysis, as it can reveal information about the network structure and the flow of information throughout the network [1]. Community detection is substantial for multiple reasons, including node classification which involves homogeneous clusters, group leaders or crucial group connectors. Communities may be associated with groups of pages of the World Wide Web dealing with related topics [2]. Early work focused primarily on identifying disjoint communities that partition the set of nodes within a network [3,4]. More recently, researchers have observed the multiplicity of interwoven memberships of communities and have developed algorithms for finding overlapping communities [5,6]. However, there are several techniques that have been extended to solve other problems in the same field of community detection. One critical issue, which has been extensively studied in the literature since the early analysis of complex networks, is the identification of communities or relationships hidden within the structure of these networks.

Considering the domain of community detection, there is another community structure known as the hidden community structure, which is gradually becoming a hot topic of research. For the hidden communities, there is hidden community detection, which aims to identify the invisible community structure in the network that often goes undetected or ignored by many popular community detection algorithms. In the same way, many real-world networks contain scarce community structures; for example, secret groups

or organizations, which are considered inadequate compared to the dense community structure in particular families, colleagues, or close friends, as calculated by distinguished community scoring metrics. If the members in the less modular community also reside in other, denser communities, then the community is usually omitted. As illustrated in Figure 1, for a small network, the hidden community structure is sparser, and it is perplexed with the structure of the dominant communities making it challenging for algorithms to detect.



**Figure 1.** An example of hidden community structure. The solid black lines show the relationship among the nodes of dominant communities and the colored dotted lines represent the weaker communities.

Moreover, as the networks can be comprised of not only a single layer but several layers, there is a possibility to have some hidden weaker overlapping community structure in them as well. However, multi-layer network models provide a vigorous and more realistic tool for the analysis of many complex systems and networks [7]. For instance, in many social networks, a person can have multiple social accounts across certain social networks and it has also, at present, become foremost to link distributed user profiles which belong to the same person from various platforms. Relationships among those accounts on different social platforms might include online as well as offline ties, such as account following, like/comment interactions, working relationships, hangouts relationships, etc. Such scenarios can effectively be represented using a multi-layer network model. As the concept of this paper is about the detection of hidden communities in networks with a dense structure comprised of multiple layers, there is a possibility that these structures have some weaker communities as well that often go undetected due to the presence of stronger communities.

This paper aims to analyze the hidden community structure through a hidden community detection (HCD) algorithm applied and observed in real-world networks to answer the question of whether the real-world networks have hidden community organization or not. We will see that the hidden community detection algorithm begins its function by using some existing community detection algorithm as its base algorithm in the network, then weakens the structure of the network to detect the hidden communities. The iterative process continues until no further hidden structures are detected.

Through experiments on real-world networks, we observed that the existence of non-redundant or inessential communities having less importance in some cases but some importance in other scenarios, are rarely detected by many other community detection algorithms. A hidden community structure can be contemplated as a special type of overlapping community structure. However, existing overlapping community detection methods mainly focus on communities in which a considerable portion of the members are not “hidden”; that is, they could also belong to other weaker communities, but this

community is clearly the strongest for these members. We believe the intuition and observations obtained on the scheme of hidden community structure will provide valuable guidance for future investigations in the domain of community detection. The main contributions of this paper are as follows:

- Introduction of the concept of hidden community structure in multi-layer networks.
- Introduction of the hidden community detection method by using an overlapping community detection method as a base algorithm and demonstrating the structure weakening method of the layers of communities.
- Application and validation of the proposed method on real-world datasets.
- The most captivating conclusion of the hidden community detection method is that it is one of the works to delineate the presence of hidden/weaker communities in multiple layer networks of several domains. Also, we compare the proposed method with some other multi-layer community detection algorithms.

The structure of the rest of the paper is as follows: in Section 2, we discuss some details about the related work from the proposed work in this paper; in Section 3, we discuss the preliminaries and definitions related to our work; in Section 4, we introduce the hidden community detection method in detail and talk about the steps of its operation; in Sections 5 and 6, we discuss the experiments; and evaluate the results of HCD method with several real-world datasets; later in Section 7, we compare our method with some multi-layer and other hidden community detection approaches developed so far, in terms of some predefined attributes; lastly, we end this paper with the conclusion.

## 2. Related Work

In past years, a tremendous amount of community detection algorithms have been introduced in different categories (overlapping [8,9], disjoint [5,10], clustering [11], etc.) to understand the structure and fundamentals of different kinds of networks. To the best of our knowledge, the examination to detect the hidden communities in the network has nearly begun to gain the attention of various researchers.

In [12], authors have studied fraud detection in networks and modeled an online auction network with fraudsters as a random network with hidden communities (fraudsters and associated accomplices) and proposed a maximum likelihood framework to detect the fraudsters, which is an iterative message-passing algorithm to heuristically solve the maximum likelihood detection problem. Generally, it has been observed that there is a possibility of clustering data in multiple alternative ways, depending on the requirements. The example of this scenario has been explained in [13], where data points of one cluster can belong to different clusters in some other ways.

Hajdu et al. [14] studied the most recent issue going around the world, i.e., the risk from an epidemic outbreak, to reveal the hidden communities of the affected or those to be affected, and analyzed the contact patterns from a public transit assignment model in a major metropolitan city. This work proposed two networks; first, the transfer network, which can reveal passenger groups that travel together on a given day. Second is the community network, which is derived from the transfer network and captures the similarity of travel patterns among passengers. Furthermore, Nath et al. [15] proposed the Hidden Community based on Neighborhood Similarity Computation (HCNC) method based on Neighborhood Similarity Computation, assuming the sub-communities as the hidden or hierarchical communities, which can detect hidden communities irrespective of density variation within the community. Chen et al. [16] remove nodes or edges dependent on the local Fiedler Vector Centrality (LFVC) which is related to the affectability of the algebraic connectivity to node or edge evacuations. Above all, they characterize an idea of profound network as an associated part that must be seen after removing nodes or edges from the remainder of the network.

There have been numerous attempts to address community detection problems in multilayer networks through diverse approaches, e.g., identifying communities in temporal networks by modularity-maximization [17] where it emphasizes the difference between

“null networks” and “null models” in modularity maximization and discusses the effect of interlayer edges on the multilayer modularity-maximization problem. De Bacco et al. [18] proposed a generative model for multilayer networks, which can be used to aggregate layers into clusters or to compress a dataset by identifying especially relevant or redundant layers.

The proposed method clearly defines the technique of its operation in how it breaks the network first and how it reforms it again. The above-mentioned methods used similarity measures to detect the most similar nodes and form communities, but it has been examined that similarity is not always the quick fix. The proposed work is also influenced by the LFVC regarding the removal of either nodes or edges but specifically focuses on edge removal to make it more simplified.

### 3. Preliminaries

In this section, we discuss the basic terms and concepts used in the proposed work. Assume graph  $G = (V, E)$  represents a network with  $n$  number of nodes and  $e$  number of edges. For each graph  $G$ , let  $A$  be its adjacency matrix; each entry of matrix  $A_{ij} \in \{0, 1\}$ , which demonstrates whether there is an edge between the nodes  $i$  and  $j$ . Let  $C = (C_1, C_2, C_3, \dots, C_k)$  be the overlapping communities, and  $C_i = (V_i, E_i)$  be the sub-graph of  $G$ . In the following subsection, we will define the modularity metric and its definition used in the hidden community detection approach.

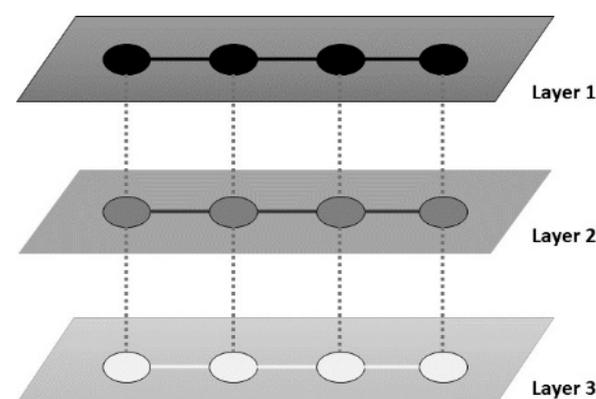
#### 3.1. Layer

A layer is a set of communities that partitions or covers the nodes of the network into several groups. Intuitively, in a social network, a partitioning layer may correspond to a grouping of individuals by their locations, and friend circles could form an overlapping layer where some friends are in interdisciplinary areas.

#### 3.2. Multi-Layer Graphs

As the proposed work is related to the hidden community detection in multi-layer networks, a multi-layer network is a kind of network that consists of multiple graphs sharing the same number of nodes while having different edges. According to the observation, it is obvious that the multi-layer networks reveal more comprehensive information than the single-layer network. Thus, it is possible to have more hidden information as well. The multi-layer network model contains nodes and edges from different layers, which represents the different frequencies (or types) of interactions among them.

Figure 2 illustrates an example of a multi-layer graph. Assume that layer one is the Facebook network, layer two is the Twitter network and layer two is the Instagram network. If the users in the Facebook network also have an account on Twitter and Instagram, then, for example, the Twitter network can be used to represent these users and their relationships. Note that every user can be identified by one account on each layer.



**Figure 2.** A three-layered example network. The solid lined edges in each layer are labeled as intralayer edges, the dotted lines crossed adjacent layers are labeled as interlayer edges.

### 3.3. Modularity

Modularity is used to determine the strength of a set of communities that partition the network. We adopt the popular modularity metric as Newman defines it; the ratio of the number of intracommunity edges to the expected number of edges in the same set of communities if the edges are distributed inconstantly while preserving degree distribution [19]. The modularity score  $Q$  of a partition is calculated by:

$$Q = \sum_{k=1}^k Q_k \quad (1)$$

$$Q_k = \sum_{k=1}^k \left[ \frac{e_k}{e} - \left( \frac{d_k}{2e} \right)^2 \right] \quad (2)$$

where:

- $K$  is the number of communities in the partitioning,
- $e$  is the number of edges in the graph,
- $e_k$  is the number of edges within community  $C_k$ ,
- $d_k$  is the total degree of the nodes in community  $C_k$ .

$Q$  lies in the range  $[0, 1]$ , with larger values illustrating a stronger set of communities.

#### The Modularity of a Community

We call  $Q_k$  the sum modularity contribution of  $C_k$  and the modularity of a single community is defined as the sum modularity contribution of that community divided by the number of nodes in the community, i.e.,  $Q_k / |C_k|$ .

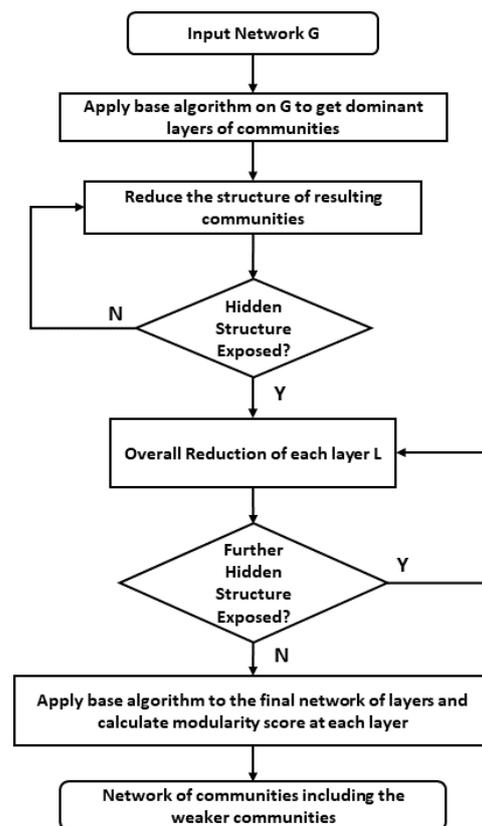
## 4. Hidden Community Detection Method

In this section, we propose a hidden community detection (HCD) method to detect multiple layers of hidden or weak communities in a network. The reason behind introducing this scenario of detecting communities is because for several applications in a large variety of scientific disciplines, the weak, hidden structure is of great interest and deserves to be explored as it may provide some useful information which can go undetected due to its presence beneath the stronger communities in a network.

Let us consider an example of a government agency aiming to detect a group of terrorists' social accounts within a social network. This structure of a community may be much weaker than other community structures, such as the ones based on family and friends. It can be expected that there exist some social accounts of terrorists' groups in a network but weaker to be appeared as a dominant structure, so it can be challenging for algorithms to locate them. Somehow, that structure is a hidden organization, but it is important. Detecting such a structure, especially in a network with multiple layers or overlapping layers, can be challenging.

This method uses an existing overlapping community detection algorithm (Louvain, CNM, Girvan-Newman, etc.) as a baseline algorithm to weaken the structure of layers detected to demonstrate the hidden structure inside the network.

Figure 3 illustrates the flowchart of the whole process of the proposed method. It can be seen that first, it normally detects the communities much like any other community detection algorithm. There are many techniques to weaken the layers. We chose edge removal as it weakens the communities' structures to expose the weaker or hidden communities if any exist. This can be determined by comparing the reduced structure of the resulting communities with the set of communities obtained by applying the baseline algorithm. Later, the iterative process goes to the overall network layers including the weak structure of communities to double check if any structure or layer weaker than the others is left. Finally, the baseline method is applied again to form a better and this time more meaningful community structure of the network.



**Figure 3.** The flowchart illustration of the operation of the hidden community detection (HCD) method.

#### 4.1. Baseline Community Detection Algorithm

The ease of the proposed method is that it can work well with any disjoint community detection algorithm as the base to detect a weaker overlapping community structure in the layers of a network. In this work, for the baseline algorithm, we chose the Louvain method, a heuristic algorithm; because of its robustness and scalability to fit well in large networks [3], we find it more suitable. The Louvain method is an efficient, simple and easy-to-implement method for identifying communities in large and different kinds of networks; also, it has been observed that it unveils hierarchies of communities and allows us to go deeper within communities to discover sub-groups, sub-sub-groups, etc., which is the basic aim of our work, i.e., to discover underlying weaker or hidden structure of communities. The inspiration for this method of community detection is that it is a fast greedy method which has the ability to optimize modularity by iteratively changing the allotment of the node community and collapse the community into meta-nodes. In the Louvain Method of community detection, first, small communities are found by optimizing modularity locally on all nodes, then each small community is grouped into one node and the first step is repeated. There are various overlapping community detection methods that can be used as the baseline algorithm for further evaluation of this work in future.

#### 4.2. HCD Algorithm

For hidden communities in a network, there is hidden community detection, which aims to identify the invisible or weaker community structure in a multi-layer network that often goes undetected or ignored by many popular community detection algorithms. Hence, to make it more precise, the proposed HCD method has three basic steps, and the basic functions of the method are given in Algorithm 1

1. Having a network  $G$  and applying a baseline method to detect the initial set of communities which can be regarded as a stronger community structure.

2. After identification, the weakening of the layers process begins by removing all of the internal edges for the current set of detected communities by using the Remove Edge process by which the effect of their structure on the network is removed to easily detect the hidden structure beneath, and get another set of communities (different than the first set). This process is repeated until no compelling communities can be found.
3. To further improve the quality of these detected layers, there is another iterative process which takes into consideration each layer of the structure and makes improvements in the manner that it first weakens the structures of all other layers in the original network. It then applies the base algorithm to the other resulting network.

---

**Algorithm 1. Hidden Community Detection: Basic Steps**


---

```

1: Step 1, 2:
2: Input:  $G = (V, E)$ 
3: Initialize:  $Gr = G // Gr$  is reduced graph
4: for each Layer in  $R$  do
5:     Partition  $\leftarrow$  baseline method ( $Gr$ )
6:     //Determine whether to store or reduce graph
7:     Hcd_layer  $\leftarrow$  store graph
8:     //Weaken the structure by removing edges
9:     Hcd_layer  $\leftarrow$  reduce graph
10:    Compute modularity  $Q$  of each  $C$  in layers using Equation (1)
11: end for
12: Step 3:
13: for each Clr_i in  $R$  do
14:    //Clr is clarification iteration and R is the range.
15:    Initialize:  $Q = 0$ 
16:    for each  $i$  in  $L_i$  do
17:        for each  $k$  in  $L_k$  do
18:            //where  $L_i$  and  $L_k$  are the layers being computed.
19:            if  $i = k$  then
20:                Clr_itr:  $i = k$ 
21:            end if
22:            //Weaken the structure by removing edges in the same way as previous step.
23:            Hcd_layer  $\leftarrow$  reduce graph
24:        end for
25:        Apply baseline method to get communities
26:         $Q = \text{Compute } |Q|$ 
27:    end for
28: end for

```

---

An important point to note is that there is need to debilitate the effects of both the weaker layers and the stronger ones because any of them can lessen the process of community detection of the current layer. For example, regardless of having the smaller influence on the network, the weaker layer can also hinder the detection process. By exercising it, there is more possibility to get a reliable version of layers. In every iteration of step three, the modularity is calculated from the layers detected, but the concluding outcome will be the one with the maximum average modularity value.

Another main challenge for this approach is to determine whether the number of layers that have been chosen for the operation of the method is correct or not. In our work, we use the policy to determine whether it is correct or not by determining the average modularity score. If the score on the graph increases, then the number of layers chosen is correct; otherwise, the trend on the graph declines.

### 4.3. Technique to Weaken the Layers

Most of the community detection algorithms easily identify the dominant communities with strong relationships among them; however, there can be some nodes inside the network with weak relationships that, together with the dominant communities, can represent some useful information. Hence, there is need to identify those weak communities first, which can be achieved by weakening the structure of the strong communities. Secondly, as is the case for the multiple-layer graphs, there is a possibility among the weak communities to have some hidden overlaps as well. We have identified a method to reduce a single layer in the community structure by removing an edge. This method is inspired by algorithms that find communities by removing edges, such as the classic Girvan–Newman algorithm, which removes edges with high betweenness [1] and follows the scenario that if any network contains communities that are generally associated with a few intergroup edges, then all shortest paths between the different communities must go along one of these few edges. Thus, the edges connecting communities will have high edge betweenness. By removing these edges, we dispartate groups from one another and reveal the underlying community structure of the graph. Another more recent work of Chen and Hero removes edges based on local Fiedler vector centrality [16]. One important point to consider is that, if there is a layer with overlapping community structure, then to avoid duplication in the weakening process, one can categorize communities by their sizes and perform the weakening process first with the larger set of communities, then with the smaller and avoid weakening the overlapping sections again. The remove edge function performs well when there are only a small number of layers. This technique weakens a detected community by removing all intra-community edges, considering it to be a simplified graph. This method also works well when there are small overlaps among several layers.

## 5. Experimental Setup

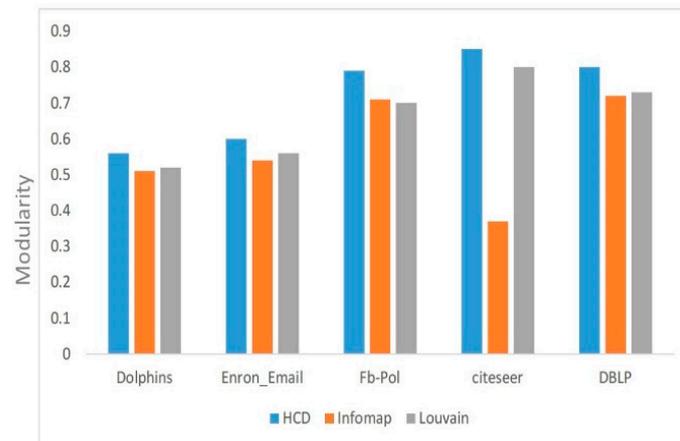
In this section, we will demonstrate the performance of our algorithm through testing on different real-world data sets and then give the evaluation results.

### 5.1. Real-World Datasets

We comprehensively tested our algorithm on the following different real-world datasets, also shown in Table 1. Among the various real-world datasets, the following are some that showed a weak/hidden community structure. More specifically, we chose networks like those that are a part of Facebook’s social network, such as Political pages or the DBLP network where each layer revolts a set of communities grouped by a common attribute, i.e., nodes with common annotations are part of one community, and by combining the set of communities, it becomes an annotated layer. These layers cover all nodes in the relative networks. The graphical representation of the difference in the number of communities is given in Section 6.1 Figure 4.

**Table 1.** Real-world network datasets used in an experiment of the HCD method.

Domain	Dataset	Edges	Vertices
Social	Dolphins	159	62
	FB-Pages-Politician	42 k	6 k
Collaboration	Citeseer DBLP	4.5 k	3.6 k
		1 M	317.1 k
Email	Enron Email	200 k	158



**Figure 4.** Results of modularity comparison of HCD algorithm with Infomap and Louvain algorithms.

- **Dolphins:** An undirected social network of frequent associations between 62 dolphins in a community living off Doubtful Sound, New Zealand, as compiled by Lusseau et al. (2003). This network has 62 nodes with 159 edges. Among them, nodes represent dolphins, and the sides indicate frequent contact between dolphins.
- **Enron Email:** This is an email message dataset between employees of the Enron Corporation [20], mostly senior management of Enron, organized into folders. It contains messaging data belonging to 158 members of the corporation. The relationship among the members can be defined by the existence of their message communications or maybe some other similar attributes.
- **CiteSeer:** This is a citation network of computer science publications [21] containing 3312 vertices and 4536 edges. Here, the relationship attributes can either be defined by citation or by the content similarity.
- **DBLP:** The DBLP [22] computer science bibliography provides a comprehensive list of research papers in computer science with approx. 317 k vertices and 1 M edges. A vertex stands for an author, and an edge is formed if two authors write a research paper together or share the same research interest. Using this dataset, we can make a general multi-layer graph whose layers are more than two, where each layer corresponds to the co-authorships in the different venues (conferences or journals).
- **FB-Pages-Politicians:** Data collected about Facebook pages (November 2017). These datasets represent blue-ticked verified Facebook page networks of different politicians [23]. Nodes represent the pages, and edges are mutual likes among them.

## 5.2. Evaluation Metric

Evaluation (also called benchmarking) is the assessment of how well an algorithm detects the desired set of communities in a network. We used the modularity metric as a performance indicator. Informally, the modularity  $Q$  of each possible partition can be calculated as:

$$Q = (\text{fraction of edges among communities}) - (\text{expected fraction of edges}).$$

Nowadays, the indicators for evaluating the performance of community detection algorithms are numerous. In this paper, we used one of the classical performance indicators, i.e., modularity  $Q$ . The more specific definition of how we calculate  $Q$  in this work is given in Section 3.1. We compare our proposed method with the same algorithm that we used as a baseline method in our work, i.e., Louvain, which is the fast method, popular for greedy modularity optimization and also used as the baseline method in the proposed method. Other metrics like Normalized Mutual Information (NMI) can be used if the focus is on finding the similarity of partitions [24] or the maximum matching ratio (MMR), which is designed specifically for biological networks [25].

## 6. Evaluation of Results and Comparison

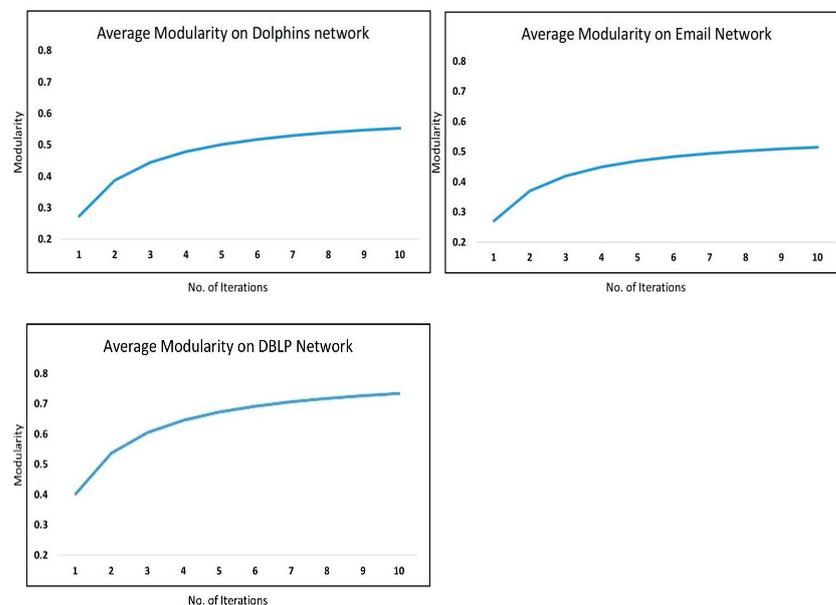
In this section, we discuss the results of the experiments of the HCD method on different real-world datasets and provide an analysis of how the performance of the proposed method gave better results. In the later part of this section, we compare and show the results of the HCD method with different multi-layer and hidden community detection methods.

### 6.1. Experimental Results

For several real-world networks from various domains, HCD identifies multiple layers of high modularity community structure. For our experiment, we applied a HCD method on each of the selected dataset graphs. The proposed HCD method identified multiple layers of communities (2, 3, or 4) in different graphs, and the results of the algorithms are shown below in Table 2. HCD not only identifies the layers with the hidden or weak structure that the other community detection algorithms usually ignore, but also improves the detection accuracy of the dominant layer as it can be observed through the average modularity scores in Figure 5; although there is only a slight difference, in some cases, it does matter.

**Table 2.** Statistics of the detected layers in different graphs with HCD method.

Dataset	Communities in Layers	Modularity in Layers
Dolphins	4, 8	0.56, 0.54
Enron Email	7, 9	0.60, 0.44
Fb-Pages-Politician	29, 74, 32	0.79, 0.74, 0.61
Citeseer	35, 75	0.85, 0.75
DBLP	203, 330, 152	0.80, 0.75, 0.65



**Figure 5.** Results of improvements in modularity scores of different graphs in each iteration. The vertical axis shows the modularity values and the horizontal axis is for the number of iterations.

After getting the above results from experimenting with the HCD algorithm on different real-world network datasets, we compared the results with two algorithms: (1) Infomap [26], which is a community detection algorithm capable of achieving high-quality communities. This algorithm is based on the principle of information theory and defines the community from the coding perspective. It uses Huffman coding and community structure secondary coding to obtain the maximum compression ratio. In this way, the problem of community detection is remodeled into an optimization problem; and (2) the

Louvain algorithm [3], which is an unsupervised modularity optimization algorithm (does not require the input of the number of communities). This method treats each node as a separate community in the initial phase and gradually optimizes the modularity by expanding the size of each community. After repeated iterations, the Louvain algorithm can get a community partition with the largest modularity value. Both algorithms give promising results when it comes to simple community detection but the performance decreases when considered in terms of detecting hidden communities in networks, as both the algorithms are unable to detect hidden communities. The difference is clear in Figure 4; HCD performs better than the other two and shows a high value of modularity.

Additionally, the reason behind implementing the improvement step i.e., step three of the algorithm, in this work is to clarify the structure of the detected communities. From the experiments, it has been observed that communities detected after Layer 1 are more hidden as compared to the ones detected in Layer 1. So the average modularity results of the detected layers after each iteration are improved by rectifying the layers. Moreover, we also figured out that the average modularity scores of the layers detected on the actual network are greatly upgraded by making improvements in the layers. As an example, Figure 5 shows the increasing trend during the iterations when we detect two, three, or four community layers on five real-world networks.

## 6.2. Attribute Based Comparison

In this section, we compare some multi-layer community detection algorithms with the proposed method according to four attributes as shown in Table 3.

**Table 3.** Comparison results of the HCD method with other multi-layer and hidden community detection methods.

Methods	Attribute 1	Attribute 2	Attribute 3	Attribute 4
Li et al. [27]	✗	✗	✗	✓
Qi et al. [20]	✗	✓	✓	✓
Silva et al. [28]	✗	✗	✗	✓
Xu et al. [29]	✗	✓	✗	✗
Zhu et al. [12]	✗	✗	✗	✗
HCD	✓	✓	✓	✓

For the comparison, we choose five multiple layer community detection approaches. The first method (Li et al. [27]) is basically a hierarchical community detection algorithm consisting of four major steps: core probing, core merging, affiliation, and classification. According to the results, this approach usually detects communities in just two-layer networks effectively. The second approach is from Qi et al. [20], and is a matrix factorization category based community detection algorithm. The main contribution of this algorithm is using edge content for the community detection process and it can be a useful source of information when nodes interact with multiple communities. Another method proposed by Silva et al. [28] is based on structural correlation pattern mining, called SCPM, which uncovers the interaction between vertex attributes and dense subgraphs using both frequent item-set mining and quasi-clique mining. Xu et al. [29] proposed a model-based community detection approach based on both structural and attribute aspects of a graph in a multi-layer network. The last method for comparison is the one of an online auction network [12] with fraudsters as a random network with hidden communities.

- **Attribute 1: Applicability to the multi-layer graph.** Defines whether an algorithm is able to detect communities in a multiple layer graph (i.e., should detect more than 2 layers) or not. Undoubtedly, there have been many studies done for community detection in both single layer and multiple layer graphs. The overall phenomenon of each approach might be different from the perspective of hidden community detection, but the general results show that the proposed method is applicable to the multiple

layer graphs and can detect the communities that often go undetected by most of the community detection algorithms.

- **Attribute 2: Equal consideration of each layer.** Since each aspect of the relationships may have a different importance in the real world, considering the importance of each layer differently is more applicable than assigning uniform importance. Thus, it is crucial to automatically find the importance of each layer based on the layer's characteristics. It has been observed that among the selected approaches and the hidden community detection method, some approaches satisfy the property of equally considering each layer during operation; the hidden community detection clearly satisfies this property because it weakens the communities at each layer and in the second phase it also clarifies the whole graph. It also gives better results in the next layer while considering the performance at the prior one.
- **Attribute 3: Insensitivity of the algorithm.** Some approaches are tightly coupled so that they can only be used for a specific kind of network. This tight coupling may limit the freedom of users to choose a network while experimenting and evaluating. It is well-known that certain algorithms tend to perform particularly well or poorly on certain kinds of graphs or networks. From the results given below, it is clear that except one, all other approaches failed to satisfy this property and the HCD method is not a very tightly coupled approach when experimented with in real-world datasets of different domains.
- **Attribute 4: Applicability to the graph with overlapping layers.** The communities can be defined in an overlapping way across layers. That is, a vertex can belong to community C1 on a certain set of layers but to community C2 on another set of layers. Among the selected approaches for comparison, some approaches can detect communities with an overlapping structure of layers but the difference is that the proposed work, i.e., the HCD method, can detect the weak communities in the same overlapping structure of layers while others may ignore them. Table 3 shows the comparison of the HCD algorithm with some previously developed multi-layer community detection algorithms. The sign (✓) shows that an algorithm satisfies the certain property and has advanced features. Some methods are designed for specific environments and they exhibit better results in those specific conditions, but they neither satisfy all the properties required for general environments nor detect hidden community structure. It is clear that the HCD algorithm is not limited to specific network environments and is able to detect weak communities hidden under the strong community structure in the network, which also satisfies all the mentioned attributes.

## 7. Conclusions

A community detection algorithm can fail in many different ways; for example, it may only be able to detect a subset of the desired community structure, or it may detect all of the desired communities but also several spurious communities. To overcome such problems, in this paper, we proposed a community detection method to reveal the weak or the hidden communities in the multi-layer network. The process goes on by weakening the structure of the dominant layers of communities to get the hidden structure inside, which may present some useful information when considering it in a real-world scenario. Through experiments on real-world networks from different domains, we observed that networks have hidden community structure (weak communities) in them. By applying the HCD method, it has become possible to get the hidden structure of communities. HCD has an advantage over modularity-based algorithms in that it simplifies the community structure and gives quality results with a clear picture of all the communities (including weak communities). We further compared the results obtained by the HCD method with two algorithms, i.e., Louvain and Infomap, along with an attribute based comparison with other five multi-layer community detection algorithms; we observed that the HCD method performs better and satisfies the attributes described in the above section.

Moreover, the study of detecting community structures in multi-layer networks is experiencing significance since the last decade. One of the applications of the HCD method could be the detection of suspect or fraud groups on social media; this can lead to confusion because not all hidden communities on social media have the aim of crime. So, a number of extensions or improvements to our method may be possible. This issue can be addressed in future research regarding the same problem to more specifically detect weak structures of communities by using different baseline methods other than the Louvain algorithm.

**Author Contributions:** The authors participated and designed the experiment together. F.R.K. under the supervision of J.S. and B.W. F.R.K. and Y.M. analyzed and evaluated the results. F.R.K. wrote the manuscript. All authors reviewed the manuscript. F.R.K. and J.S. are co-first authors. All authors have read and agreed to the published version of the manuscript.

**Funding:** This work is supported by National Key Research and Development Program of China. under grant No. 2018YFB1003602.

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

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

## References

1. Girvan, M.; Newman, M.E.J. Community structure in social and biological networks. *Proc. Natl. Acad. Sci. USA* **2002**, *99*, 7821–7826. [[CrossRef](#)] [[PubMed](#)]
2. Flake, G.; Lawrence, S.; Giles, C.; Coetzee, F. Self-organization and identification of Web communities. *Computater* **2002**, *35*, 66–70. [[CrossRef](#)]
3. Blondel, V.D.; Guillaume, J.-L.; Lambiotte, R.; Lefebvre, E. Fast unfolding of communities in large networks. *J. Stat. Mech. Theory Exp.* **2008**, *2008*, P10008. [[CrossRef](#)]
4. Pons, P.; Latapy, M. Computing communities in large networks using random walks. *J. Graph Algorithms Appl. Citeseer* **2006**, *10*, 191–218. [[CrossRef](#)]
5. Gregory, S. Finding Overlapping Communities Using Disjoint Community Detection Algorithms. In *Complex Networks*; Springer: Berlin/Heidelberg, Germany, 2009; pp. 47–61.
6. He, K.; Sun, Y.; Bindel, D.; Hopcroft, J.; Li, Y. Detecting overlapping communities from local spectral subspaces. In Proceedings of the 2015 IEEE International Conference on Data Mining, Atlantic City, NJ, USA, 14–17 November 2015; pp. 769–774.
7. Huang, X.; Chen, D.; Ren, T.; Wang, D. A survey of community detection methods in multilayer networks. *Data Min. Knowl. Discov.* **2020**, 1–45. [[CrossRef](#)]
8. Wen, X.; Chen, W.-N.; Lin, Y.; Gu, T.; Zhang, H.; Li, Y.; Yin, Y.; Zhang, J. A maximal clique based multiobjective evolutionary algorithm for overlapping community detection. *IEEE Trans. Evol. Comput.* **2016**, *21*, 1. [[CrossRef](#)]
9. Huang, L.; Wang, C.-D.; Chao, H.-Y. Overlapping community detection in multi-view brain network. In Proceedings of the 2018 IEEE International Conference on Bioinformatics and Biomedicine (BIBM), Madrid, Spain, 3–6 December 2018; pp. 655–658.
10. Xu, Y.; Xu, H.; Zhang, N. A novel disjoint community detection algorithm for social networks based on backbone degree and expansion. *Expert Syst. Appl.* **2015**, *42*, 8349–8360. [[CrossRef](#)]
11. Alandoli, M.; Al-Ayyoub, M.; Al-Smadi, M.; Jararweh, Y.; Benkhelifa, E. Using Dynamic parallelism to speed up clustering-based community detection in social networks. In Proceedings of the 2016 IEEE 4th International Conference on Future Internet of Things and Cloud Workshops (FiCloudW), Vienna, Austria, 22–24 August 2016; pp. 240–245.
12. Zhu, K.; Guan, Y.; Ying, L. Detecting hidden communities in online auction networks. In Proceedings of the 2012 46th Annual Conference on Information Sciences and Systems (CISS), Princeton, NJ, USA, 21–23 March 2012; pp. 1–6.
13. Cui, Y.; Fern, X.Z.; Dy, J.G. Non-redundant multi-view clustering via orthogonalization. In Proceedings of the Seventh IEEE International Conference on Data Mining (ICDM 2007), Omaha, NE, USA, 28–31 October 2007; pp. 133–142.
14. Hajdu, L.; Bóta, A.; Krész, M.; Khani, A.; Gardner, L.M. Discovering the hidden community structure of public transportation networks. *Netw. Spat. Econ.* **2019**, *20*, 1–23. [[CrossRef](#)]
15. Nath, K.; Dhanalakshmi, R.; Vijayakumar, V.; Aremu, B.; Kumar Reddy, K.H.; Gao, X.Z. Uncovering hidden community structures in evolving networks based on neighborhood similarity. *J. Intell. Fuzzy Syst.* **2020**, *39*, 1–10. [[CrossRef](#)]
16. Chen, P.-Y.; Hero, A.O. Deep community detection. *IEEE Trans. Signal Process.* **2015**, *63*, 5706–5719. [[CrossRef](#)]
17. Bazzi, M.; Porter, M.A.; Williams, S.; McDonald, M.; Fenn, D.J.; Howison, S.D. Community detection in temporal multilayer networks, with an application to correlation networks. *Multiscale Modeling Simul.* **2016**, *14*, 1–41. [[CrossRef](#)]
18. De Bacco, C.; Power, E.A.; Larremore, D.B.; Moore, C. Community detection, link prediction, and layer interdependence in multilayer networks. *Phys. Rev. E* **2017**, *95*, 042317. [[CrossRef](#)] [[PubMed](#)]

19. Newman, M.E.; Girvan, M. Finding and evaluating community structure in networks. *Phys. Rev. E* **2004**, *69*, 026113. [[CrossRef](#)] [[PubMed](#)]
20. Qi, G.-J.; Aggarwal, C.C.; Huang, T. Community detection with edge content in social media networks. In Proceedings of the 2012 IEEE 28th International Conference on Data Engineering; Institute of Electrical and Electronics Engineers, Arlington, VA, USA, 1–5 April 2012; pp. 534–545.
21. Bader, D.A.; Meyerhenke, H.; Sanders, P.; Wagner, D. *Graph Partitioning and Graph Clustering*; American Mathematical Society: Providence, RI, USA, 2013; Volume 588.
22. Yang, J.; Leskovec, J. Defining and evaluating network communities based on ground-truth. *Knowl. Inf. Syst.* **2015**, *42*, 181–213. [[CrossRef](#)]
23. Grčić, K.; Babac, M.B.; Podobnik, V. Generating politician profiles based on content analysis of social network datasets. *J. Univers. Comput. Sci.* **2017**, *23*, 236–255.
24. Danon, L.; Díaz-Guilera, A.; Duch, J.; Arenas, A. Comparing community structure identification. *J. Stat. Mech. Theory Exp.* **2005**, *2005*, P09008. [[CrossRef](#)]
25. Nepusz, T.; Yu, H.; Paccanaro, A. Detecting overlapping protein complexes in protein-protein interaction networks. *Nat. Methods* **2012**, *9*, 471. [[CrossRef](#)] [[PubMed](#)]
26. 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](#)] [[PubMed](#)]
27. Li, H.; Nie, Z.; Lee, W.-C.; Giles, L.; Wen, J.-R. Scalable community discovery on textual data with relations. In Proceedings of the 17th ACM Conference on Information and Knowledge Management, Napa Valley, CA, USA, 26–30 October 2008; pp. 1203–1212.
28. Silva, A.; Meira Jr, W.; Zaki, M.J. Mining attribute-structure correlated patterns in large attributed graphs. *arXiv* **2012**, arXiv:1201.6568 2012. [[CrossRef](#)]
29. Xu, Z.; Ke, Y.; Wang, Y.; Cheng, H.; Cheng, J. A model-based approach to attributed graph clustering. In Proceedings of the 2012 ACM SIGMOD International Conference on Management of Data, Scottsdale, AZ, USA, 20–24 May 2012; pp. 505–516.