

Received April 11, 2022, accepted April 29, 2022, date of publication May 3, 2022, date of current version May 11, 2022.

Digital Object Identifier 10.1109/ACCESS.2022.3172298

Community Detection Based on Fish School Effect

YANAN SUN¹, ZEJUN SUN^{1,2}, XINFENG CHANG¹, ZHONGQIANG PAN¹, AND LIMIN LUO¹

¹Department of Network Center, Pingdingshan University, Pingdingshan, Henan 467000, China

²School of Information Engineering, Pingdingshan University, Pingdingshan, Henan 467000, China

Corresponding author: Zejun Sun (szj@pdsu.edu.cn)

This work was supported in part by the Science and Technology Research Project of the Science and Technology Department of Henan Province under Grant 222102210129, in part by the “14th Five-Year Plan” Project of Educational Science in Henan Province of China under Grant 2021YB0232, and in part by the Research and Practice Project of “Internet Plus Education” Special Topic in Pingdingshan University under Grant HLW202040.

ABSTRACT Community structure has a more intuitive, physical meaning than how it has traditionally been perceived; this representation plays an important role in networks and is accordingly becoming widely adapted. Consequently, community detection has attracted increasing attention. Although scholars have proposed many community detection methods from different perspectives, due to the complexity, diversity and dynamic characteristics of networks, efficient community detection in many real networks remains a challenge. Inspired by fish school effect in real life, this paper envisions networks as an ecosystem and proposes a novel dynamic model that aims to reveal the communities in a more intuitive way. Relying on the new model, we design a community detection algorithm, known as community detection based on fish school effect (CDFSE). CDFSE has plentiful desirable properties: high-quality community detection, parameters free and notable scalability. To evaluate the performance of CDFSE, this paper employs two widely utilized evaluation metrics and twelve representative algorithms to test the effectiveness of the algorithm in both synthetic and real-world networks. The experimental results show that in most cases, CDFSE is superior to the comparison methods in terms of the quality of community detection.


INDEX TERMS Fish school effect, community detection, cluster, complex network.

I. INTRODUCTION

With the rapid development of computer and information technology, modeling the complex networks found in nature has attracted increasing attention and became a popular research topic. Therefore, complex networks are becoming an important research direction. In the real world, a large number of composite systems involving such fields as biology, computer science, sociology, economics, transportation, and iatrology can be abstractly described as complex networks [1]–[4]. The framework of complex network is mainly contributed by nodes and edges, nodes indicate elements in complex systems and edges represent specific relationships between pair of nodes. Community detection provides a scheme for analyzing the structural characteristics of complex networks to study their organizational functions and explore their potential connections. For example, community detection could elucidate the mechanism behind human migration and social structure formation [5]; constructing a pathophoresis dynamics model to predict and mitigate the spread range

of infectious diseases [6]; investigating brain networks to reveal the potential conjunction between various functions and discovering treatments for major diseases [7]; optimizing and processing massive amounts of data to realize network public sentiment monitoring and improvement of intelligent recommendations and precision marketing [8], [9]; characterizing the interdependence and hidden network behind the economy to predict the occurrence and evolution of financial crises [10], [11]; and detecting roads with frequent traffic accidents and congestion, ensuring the safety and smoothness of transportation networks. In consequence, the study and design of efficient and accurate community structure detection methods is crucial.

Currently, representative community discovery algorithms in the field of complex networks primarily include graph-based segmentation algorithms, hierarchical clustering algorithms and partition-based optimization algorithms [12]. The classical algorithm (KL) is based on bisection. Its basic idea is to introduce a trial function Q for the network. Q represents the difference between the number of edges in two quasi communities minus the number of edges between two quasi communities, and then get the partition method to maximize

The associate editor coordinating the review of this manuscript and approving it for publication was Weisi Guo .

the value of Q . However, this algorithm and many others need to know the number of communities and the number of nodes in each community in advance. Hierarchical clustering algorithms primarily include splitting algorithms and aggregation algorithms. The GN algorithm [13] is one of the most popular splitting algorithms and is based on edge betweenness. In the process of partitioning, the algorithm needs to constantly calculate and eliminate the edge with the highest edge betweenness. The fast greedy algorithm [14] is a representative aggregation algorithm that uses the modularity index as the objective function of the algorithm partition and selects the structure corresponding to the optimal modularity value as the final community partition result [15]. The two algorithms have high time complexity and low effectualness for large-scale network community discovery. The main idea is to constantly adjust the community of each node such that the algorithm partition obtains the best community discovery. Although the k -means algorithm based on partitioning [16] has the advantage of simplicity, low complexity and fast convergence speed, its distinct method for randomly selecting initial community centers affects community partitioning to a certain extent. The label propagation algorithm (LPA) [17], [18] has approximately linear time complexity. On the other hand, outcomes observed from community detection are regularly in volatile. Roy U K proposed a modified local random walk method to catch the fuzzy community based on neighbors' similarity [19]. Xiaodong Li surveys recent theoretical advances in convex optimization approaches for community detection [20]. Zhang, Y proposed the SEAL algorithm to detect communities using generative adversarial networks [21].

With the expanding scale of complex networks, traditional community discovery algorithms appear some shortcomings which limit the efficiency of the algorithm in practice, such as high complexity, the demand to specify the number of communities, community size and other prior conditions, or the need to optimize the predefined objective function.

This paper proposes a community detection method based on fish school effect (CDFSE) to characterize network community structure. The network is regarded as an ecosystem, and its dynamic processes over time are explored. Inspired by this principle, a new dynamic model is designed to simulate fish school effect. Compared with the traditional community detection method, the CDFSE method has better effect. Firstly, it can achieve high-quality community division. Secondly, it does not require parameter setting. Thirdly, it does not claim the prior knowledge of network structure; more importantly, it has the characteristics of low time complexity, and it works efficiently in identifying community structure in large-scale networks. In the next section, we explore the central principle behind this method.

A. BASIC IDEA

In the physical world, common interests or hobbies will attract people to each other. These people may then attract other members to join or be attracted to and join a larger

group. Once a group is large enough, it will have stronger resistance [22], [23]. More importantly, this group is likelier to attract other potential members. In nature, and all aspects of human life, this behavior can be seen; for example, when two people like basketball, they may set up a basketball team to attract other members or join other basketball teams. People are attracted by activities according to their interests. This phenomenon is similar to fish school effect in the ocean [24]. Fish school effect refers to the phenomenon of groups of fish swimming in the sea, both chaotic and orderly. The fish move from east to west in a uniform manner according to the current and food availability; when attacked by hunters, they quickly gather and disperse, similar to an organization with a strict division of labor and cooperation. Fish school effect is mainly used in the field of intelligent transportation. Through the research on the consistency of fish swarm and swimming behavior, a mathematical model of fish school movement is constructed to simulate the vehicle path trajectory, so as to effectively realize the multi vehicle cooperative control in traffic [24], [25]. Community can be regarded as a collection of users with strong common attributes. As time goes on, when an increasing number of members with this common attribute are in the community, they can form a potent organization, similar to fish swarming in the sea, and produce a force far greater than the individual. The individuals can not only resist risks but also attract other members to join to form a larger community. In complex networks, nodes share similarity and attract each other, especially in scale-free networks. Barely nodes occupy larger degree than others while majority nodes own small degree. Majority of edges cluster in the community while only few of them cross between communities within community structure. Based on this feature, we examine whether we can characterize community structure by reproducing fish school effect for nodes in a complex network. We propose a new method to divide community structure based on fish school effect. Regarding the network as an adaptive dynamic system and investigate its dynamics with time-varying becomes a significant notion. Specifically, in an ocean network, there are numerous kinds of fishes, which is the initial state; the same kind of fish attract each other because of their common characteristics (they share great similarity), forming an initial group. The group itself and a small number of fish in the group will attract and gather other fish of the same kind because of their greater attraction, forming a large-scale group that is more attractive, more stable, and even attracts other groups to form a larger group, which is the motion state of fish group. Over time, different kind of fish will be attracted to different groups according to their distinct characteristics, which is the steady state of the fish group, and the whole process is shown in Fig. 1.

Why do all kinds of fish with initial state in the sea move regularly and finally reach a steady state? Because at the beginning, each individual in the network has its own characteristics (this article refers to the characteristics that are different from other individuals, or its charm or resources)

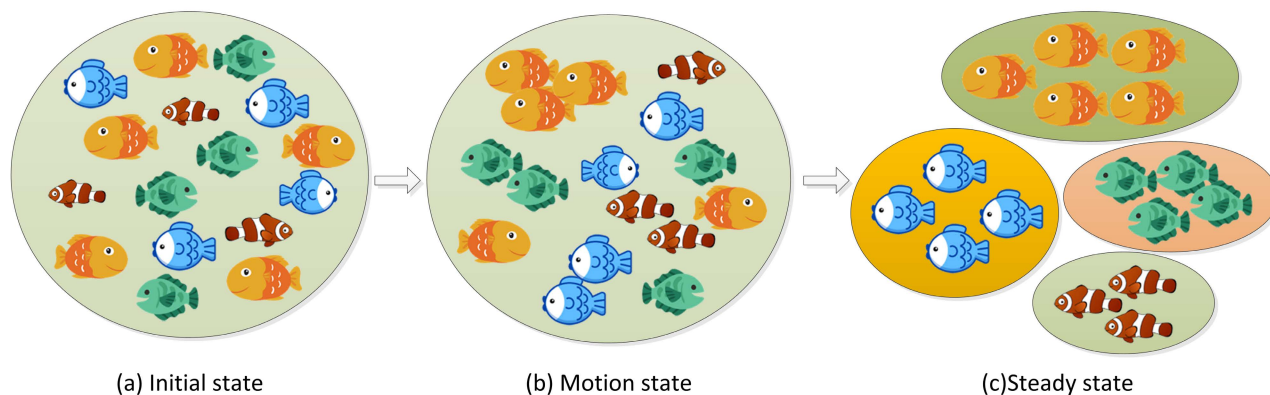


FIGURE 1. The formation process of fish school effect. (a) In the initial stage, different kinds of fishes in the ocean swim disorderly as an individual. (b) after a period of time, the same kind of fish attract each other to form small groups because of their same habits. (c) finally, all kinds of fishes in the ocean form small groups.

and commonness (this paper refers to the common characteristics that distinguish groups from each other, for example, common interests and common goals), it is these features and commonality to push them, over time, each individual is attracted by different groups (communities) to achieve a stable state.

We will formally define the fish school effect model in Section 3. Here, according to the three states of fish school effect, a small fish swarm network in the sea is employed as an example to illustrate this idea. The nodes in Fig. 2 can be regarded as individual fish in the sea. The edge connection between nodes indicates that the two fish have common characteristics, and the value of this connection is their similarity. The procedure of examination in community detection depending on fish school effect splits into three stages: at the beginning, each fish has its own characteristics and commonness, and they are all scattered. The similarity between individuals is calculated by the Adamic-Adar Index [26] (AA) expression. Here we need to introduce the AA similarity index. The index considers the degree value information of the common neighbors of two nodes. The idea is that the contribution of the node with small degree value is greater than that with large degree value. For example, the similarity between individuals 2 and 3 is 0.53, and the similarity between individuals 4 and 6 is 0.32 as illustrated in Fig. 2 (a). The following stage is that, fish with larger characteristics attract each other to form subgroups because of their greater similarity. For example, individuals 2 and 3, individuals 4 and 6, individuals 8, 9 and 10, and individuals 11 and 15 have greater similarity than other individuals, so they are the first four to form subgroups A, B, C and D respectively and presented in Fig. 2 (b). In the third state, an expanding amount of fish are attracted to join the subpopulation in order to form a large population due to fish school effect. For example, the affinity between individual 1 and individual 2 is 0.21, the similarity between individuals 1 and 3 is 0.39, and the similarity between individuals 1 and 7 is 0.22. Therefore, compared with group B, group A is more attractive to individual 1. In the

same way, the similarity between 7 and individual 1 is 0.22, and the similarity between individual 7 and individual 6 is 0.30, so individual 7 is more attracted to group B. In the end, individual 1 is attracted by group A, fish individuals 7, 5 and 12 are attracted by group B, and fish individuals 13 and 14 are attracted by group D. From this prospective, each individual is attracted to different groups or communities in other words. Therefore, the community structure is exposed naturally and this procedure is demonstrated in Fig. 2(c).

B. CONTRIBUTIONS

By simulating fish school effect, the CDFSE algorithm achieves several advantageous community detection characteristics for complex networks. The main advantages are as follows:

- **Novelty:** “Fish school effect” reflects the behavior of fish in the ocean but now is introduced into the complex network for community detection. At the same time, it is also the principle of interlinkage of complex natural laws.
- **High efficiency:** Because the CDFSE algorithm derives from a natural law, it is closer to a real network. Through experimentation, the algorithm is more efficient and has better community detection quality than several other representative community detection methods.
- **Simplicity:** Compared with the traditional classical methods, CDFSE method is simpler and more convenient, that is, CDFSE method does not request parameter settings. Through experiments, we further verify the feasibility of applying fish swarm effect theory to community detection in complex networks.
- **Scalability:** The time complexity of CDFSE is determined as $O(n \cdot k^2)$ due to the only restriction in calculating the attractiveness of adjacent individuals. It is worth noticing that the value of k represents the average degree of nodes and its numerically tiny. For this reason, large-scale networks can be more efficiently processed by applying CDFSE algorithm.

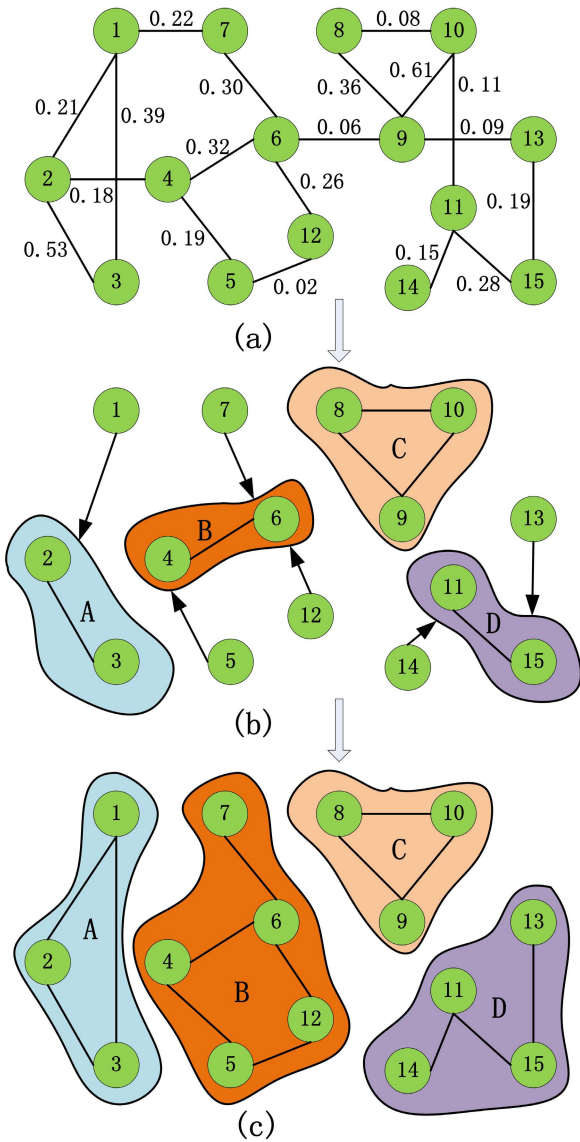


FIGURE 2. Community detection process based on fish school effect. (a) In a fish school network, the edge represents the relationship between individual fish, and the similarity between individuals is different. (b) Different individuals gravitate each other to form subgroups because of their commonness and individuality. (c) As time goes on, the community structure will eventually be revealed naturally.

The first section of this paper introduces the main ideas and innovations of CDFSE. The exposition of CDFSE method and fish school effect model including the corresponding algorithm will be introduced in the following section by a more specific way. In the third section, CDFSE is compared with several typical algorithms on the generation network and the real network. Finally, the conclusion is given in the fourth section.

II. METHODOLOGY

A. RELEVANT DEFINITIONS

Before elaborating the CDFSE method in detail, we firstly formalize some basic notations. TABLE 1 describes all the key symbols used in this paper and gives a brief description.

TABLE 1. Symbol Description.

Symbol	explanation
n_v	The number of nodes in the network ($n_v= V $)
n_e	The number of edges in the network ($n_e= E $)
D_v	Degree of node v
$d_c(v)$	The interior degree of node v in Community c
$N(v)$	Neighborhood of node v
AA_{v1v2}	AA similarity index between nodes $v1$ and $v2$
$NA_{v2 \rightarrow v1}$	Attraction of node $v1$ to node $v2$
$CA_{v \rightarrow c}$	Attraction of Community c to node v

We denote an undirected and an unweighted network as $G = (V, E)$ where V is a set of nodes and E is a set of edges.

Definition 1: (AA similarity index) [26] Given undirected network $G = (V, E)$, the AA similarity index of nodes $v1$ and $v2$ is defined as

$$SAA_{v1v2} = \sum_{v \in CN_{v1v2}} \frac{1}{\log(D_v)} \quad (1)$$

where CN_{v1v2} is the common neighbor node of $v1$ and $v2$. In fact, any two nodes in the community may have similarity. Generally, the similarity of two nodes is relatively large only when they are directly connected. If two nodes are indirectly connected, their similarity is small.

There are many kinds of fishes in the sea. The identical kinds of fish share the same lifestyles and attract each other to form a school of fish. Equivalently, individuals have their own interests and characteristics in reality which caused mutual attractions into the same group among people who shares related interests. For example, both A and B like basketball, and both B and C like music. Compared with basketball, B likes music more. It is obvious that the similarity between A and B is not as great as that between B and C, so B and C will attract each other with same music rather than a basketball team. Because the attraction between individuals is related not only to the similarity but also to the degree of the individuals themselves, the attraction between individuals is defined as the multiplication of the degree of individuals and the similarity between individuals.

Definition 2: (individual attraction) Given the undirected network $G = (V, E)$, the attractiveness of node $v1$ to node $v2$ is defined as

$$NA_{v2 \rightarrow v1} = SAA_{v1v2} * D_{v1} \quad (2)$$

SAA_{v1v2} indicates similarity coefficient of AA between nodes $v1$ and $v2$, and D_{v1} represents the degree of node $v1$.

Once the fish in the sea form a group, it will produce the fish group effect. After joining a fish group, fish have to accept the constraints of the organization, simultaneously, they obtain stronger anti-risk ability and rich food. Thus, they attain a greater ability to attract new members. Correspondingly, the community itself is a collection of users with strong

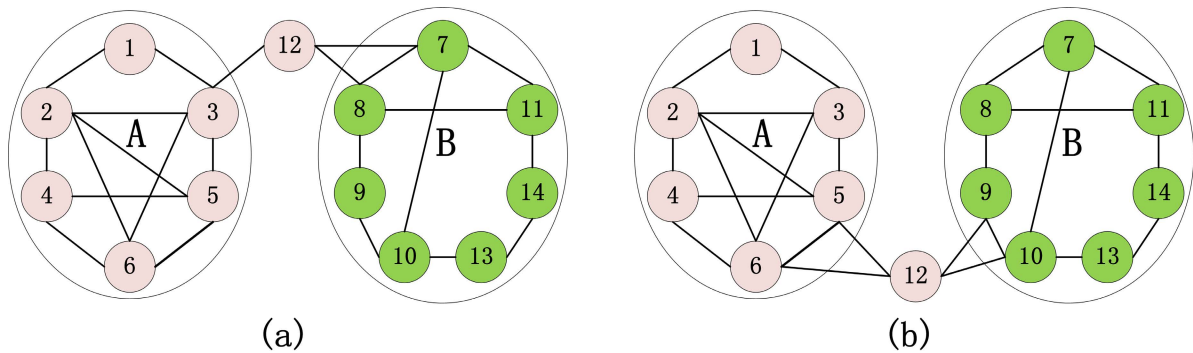


FIGURE 3. Example of group attraction. For node 12, it is ultimately attracted by that community, which is determined by its internal degree and the degree of its adjacent nodes, which can be calculated by formula (3). (a) When the internal degree of node 12 in the two communities is not equal, after calculation, the attraction of community B is greater than that of community A, and it is finally attracted to community B. (b) When node 12 has the same internal degree in the two communities, the attraction of community A is greater than that of community B, and it is finally attracted to community A.

common attributes, which will also have a strong attraction for new members. For example, an excellent research team will attract more researchers to join. In this research, we apply the connection strength of individuals in a group to describe the attraction of groups to individuals.

Definition 3: (cluster attraction) given undirected network $G = (V, E)$, the attractiveness of Community c to node v is defined as

$$CA_{v \rightarrow c} = d_c(v)^2 + \sum_{u \in N(v), u \in c} d_c(u) \quad (3)$$

where $d_c(v)$ denotes the internal degree of node v in Community c . According to the number of edges from node v to different communities, the community attraction $CA_{v \rightarrow c}$ can be divided into two cases: one is $d_c(v)$ inequality, and the other is $d_c(v)$ equality.

To better understand how to use formula (3) to calculate group attractiveness, we use an example network to illustrate. The example network consists of 14 nodes, which are divided into two communities A and B, as shown in Fig. 3. There are also two cases: (1) when a node's internal degree in two communities is not equal. For example, for node 12 in Fig. 3 (a), the degree of internality in community A is $d_A(12) = 1$, while the degree of internality in community B is $d_B(12) = 2$. The internal degree of node 12 in community A is significantly less than that in community B. The attraction of node 12 to community B may be greater than that of community A. To further determine this attraction, we consider the influence of indirect neighbors on node 12. In community A, node 3 is connected with node 12, and the degree of node 3 is 5, $d_A(3) = 4$, so the attraction of community A to node 12 can be calculated as $CA_{12 \rightarrow A} = 1^2 + 4 = 5$ according to formula (3). In community B, nodes 7 and 8 are connected with node 12, and the degree of node 7 is 4, so $d_B(7) = 3$, the degree of node 8 is 4, so $d_B(8) = 3$, so the attraction of community B to node 12 can be calculated as $CA_{12 \rightarrow B} = 2^2 + 3 + 3 = 10$. Since the attraction of community B to node 12 is stronger than that of community A to node 12, node 12 is more likely to join community B, as shown in Fig. 3 (a). (2) When a node is

equal within two communities, for example, for node 12 in Fig. 3 (b), $d_A(12) = d_B(12) = 2$, it is difficult to judge which community, A or B, are more attractive to node 12. Consequently, the impact of indirect neighbors appears in node 12 must be attentively considered. In community A, nodes 5 and 6 are connected with node 12, and the degree of node 5 is 5, so $d_A(5) = 4$, the degree of node 6 is 5, $d_A(6) = 4$, so according to formula (3), the attraction of community A to node 12 can be calculated as $CA_{12 \rightarrow A} = 2^2 + 4 + 4 = 12$. In community B, nodes 9 and 10 are connected with node 12, and the degree of node 9 is 3, so $d_B(9) = 2$, the degree of node 10 is 4, $d_B(10) = 3$, so according to formula (3), the calculation can be given is that the attraction of community B to node 12 is $CA_{12 \rightarrow B} = 2^2 + 2 + 3 = 9$. Because community A is more attractive to node 12, node 12 is more likely to join community A, as shown in Fig. 3 (b).

B. FISH SCHOOL EFFECT MODEL

On the basis of the relevant definitions in Section 2.1, the fish school effect model is constructed. The model consists of three stages: network initialization, subpopulation formation and fish school effect, as shown in Fig. 4.

1) NETWORK INITIALIZATION

In the sea, there are a variety of fishes, where each fish is a single individual with its peculiar living habits and characteristics. Therefore, we can regard the individual fish in the ocean as a node in the network. In this paper, we apply the degree and similarity of nodes to describe their characteristics. Initially, each individual has its own characteristics and is regarded as an independent group or individual, as shown in the first stage of Fig. 4.

2) SUBGROUP FORMATION

In the sea, each fish is an independent individual with its own personality. Therefore, there are certain differences among individuals. Some individuals have greater characteristics and will attract other individuals, while others with smaller characteristics will be attracted by other individuals. Furthermore, distinctive resources have been allocated to

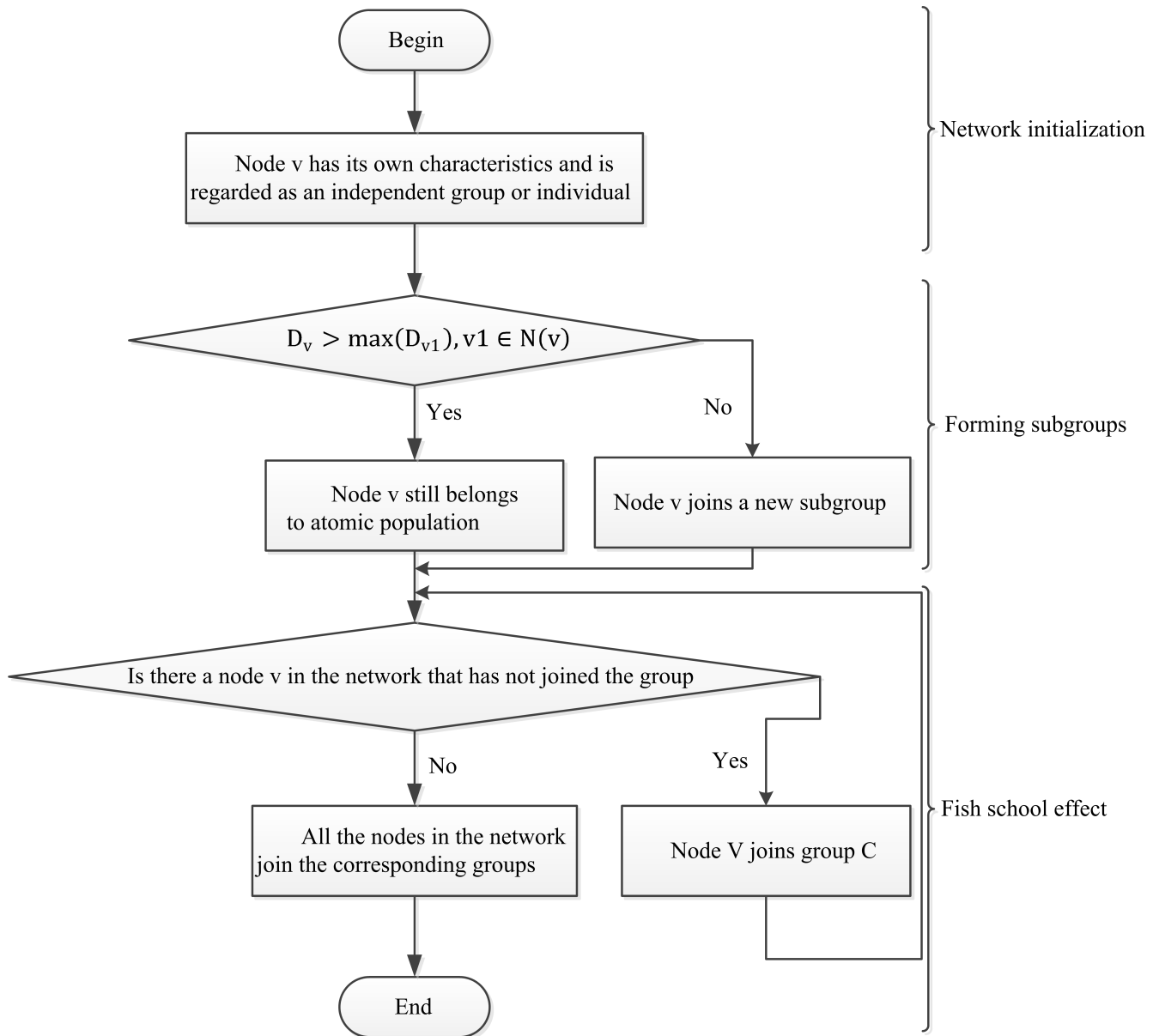


FIGURE 4. Flow chart of the fish school effect model. 1) Network initialization: initially, each individual has its own characteristics and is regarded as an independent group or individual. 2) Subgroup formation: nodes are more likely to attract neighbors to join their communities for which with more resources in order to build a subgroup after processing a series of iteration. 3) Fish school effect: all nodes have joined distinct groups and eventually each group structure tends to reach an equilibrium state after numerous iterations have been operated. Therefore, the community structure of the network is founded naturally.

particular node in the network. The more resources the nodes have, the more attractiveness they obtained to let other nodes to join in. Oppositely, nodes with fewer resources are attracted by remaining nodes In the second stage as shown in Fig. 4, when $D_v > \max(D_{v1}), v1 \in N(v)$, node v still belongs to the atomic group C_v . Otherwise, node v joins a new subgroup C_{v1} where C_v is the community to which node v belongs and D_v represents the degree of node v. By applying a series of iterations, nodes equipped with extra information are presumably to attract more neighbors to join their communities for building a subgroup.

3) FISH SCHOOL EFFECT

An expanding quantity of nodes are appealed by varied subgroups after the formation of subgroups. The structure and attractiveness of the subgroup may change when a node joins it. As long as there are nodes in the network that have not joined the group, a new round of iteration will be carried out. Eventually, driven by the network topology, all nodes will be able to change their attractiveness after processing many iterations. The nodes have joined different groups, the network structure reaches the equilibrium state when the group to which the nodes belong no parameters longer changes, as shown in the third stage of Fig. 4. In this

way, the community structure of the network is formed naturally.

C. CDFSE ALGORITHM

The detailed algorithm named CDFSE will be interpreted in this part.

1) NETWORK INITIALIZATION

Initially, we characterize each node as an independent individual or group. To distinguish different nodes, we label each individual or group with a number for initialization.

2) CORE GROUP CALCULATION

Before calculating the core group, we need to compute the individual attraction first. Individual attractiveness is related to resources and similarities. In the network structure, we apply the degree of nodes to represent the resources owned by nodes and use the AA similarity coefficient to represent the similarity between nodes. According to formula (2), we can measure the attraction between individuals. It is because of the existence of individual attraction that the nodes in the network attract each other and form a subgroup (core group). We can calculate the core group through a round of iteration.

3) FISH SCHOOL EFFECT SIMULATION

According to the fish school effect model proposed in this paper, the interaction between subpopulations constantly attracts other individuals to join in and thus form larger subpopulations. The procedure of fish school effect can be simulated because the subpopulation tends to be stable after repeatedly iterations. In the community, we employ the normalized mutual information index to evaluate the performance of every community partition. With the passage of time and the influence of topology, the network structure achieves a steady state; thus, the best division can be obtained. The CDFSE algorithm is displayed in algorithm 1 CDFSE.

D. ALGORITHM COMPLEXITY ANALYSIS

The CDFSE algorithm is primarily divided into three stages, and the time complexity of each stage is shown in TABLE 2. Therefore, the computational complexity of CDFSE is $O(k^2 \cdot n)$. The time complexity of the CDFSE algorithm is relatively low because the average degree k of the network is generally small, for this reason the algorithm can deal with large-scale networks.

III. EXPERIMENTS

To evaluate the achievement of the CDFSE algorithm, this experiment compares CDFSE with twelve representative community detection algorithms and conducts experiments on the generated network and real network respectively. Firstly, we briefly describe these algorithms.

Ncut (Normalized cuts) [27] is a typical spectral clustering method. Above all, the sample data set is defined as an affinity matrix describing the similarity between the two data points, and then the eigenvalues and eigenvectors of the matrix are

Algorithm 1 CDFSE

```

Input:
G = (V,E)
1://Initialize communities of each node
2: for each node v in V do
3:   Cv=Nv the node number of v
4: end for
5://Compute the core subgroups
6: for each node v in V do
7:   for each node u in N(v) do
8:     compute the AA similarity coefficient using Eq. (1)
9:     compute the individual attraction using Eq. (2)
10:  end for
11:  compute core groups
12: end for
13://Simulate fish school effect
14: Flag = TRUE
15: NMImax = 0
16: while Flag do
17:  for each node v in V do
18:   for each node u in N(v) do
19:    compute the cluster attraction using Eq. (3)
20:   end for
21:  end for
22:  S = NMI(C; true clusters)
23:  if S > NMImax then
24:   NMImax = S
25:  else
26:   Flag = FALSE
27:  end if
28: end while
29:// return communities C
Output: C
    
```

TABLE 2. Complexity analysis of the cdfse algorithm.

stage	Time complexity	explanation
1. Network initialization	$O(n)$	A loop to calculate the initial group
2. Computing core groups	$O(k \cdot n)$	Two loops are needed to calculate the core groups in the network, and K is the average degree of the network
3. Simulate the Fish school effect	$O(L \cdot k^2 \cdot n)$	When l times of iteration are needed, the network reaches a stable state, and l is usually 3-10

determined. Next, the appropriate eigenvectors are selected to cluster different data points. The essence of Ncut is to transform the clustering problem into the optimal partition of graphs.

Infomap [28] is a community detection algorithm combining random walk and information theory. It regards community detection as a coding problem and obtains the optimal community structure according to the principle of minimum description length.

LPA (Label Propagation Algorithm) [29] is a graph-based semi-supervised learning method that uses the relationship between samples to establish a complete graph model. Its essential idea is to use the labeled information of marked nodes to predict the labeled information of unlabeled nodes. The algorithm is simple to implement and shorter execution time cost, but the result obtained from each run is unstable.

WT (WalkTrap) [30] is a random walk algorithm, which divides communities by calculating the similarity among nodes using random walk.

FG (Fast Greedy) [31] is a hierarchical aggregation Approach for community detection. The mechanism behind this approach is to optimize modularity greedily to reveal the community structure.

Louvain [32] is a well-known multilevel modularity optimization community detection algorithm that allows hierarchical community detection. Its time complexity is lower compared with the algorithm proposed by Newman.

MCL (Markov Clustering) [33] refers to the Markov clustering algorithm, which is a fast and scalable clustering algorithm. It is based on the simulation of flow in graphs and is widely used in different fields, mainly in the field of life science.

FluidC (Fluid Communities) [34] is a scalable and diverse community detection algorithm based on propagation. It simulates the expansion and contraction of a fluid until equilibrium is found.

EDCD (Edge-Deleting Community Detection) [35] is an optimized modularity algorithm that operates a restriction strategy to iteratively delete edges to find strongly connected communities.

SCD (Silhouette Community Detection) [36] is an embedded clustering method, which reveals community structure by optimizing the contour measurement, particularly, extracting the real value representation of nodes from its neighborhood.

ASOCCA (Adjacent node Similarity Optimization Combination Connectivity Algorithm) [37] is a combination connectivity algorithm for optimizing the similarity of adjacent nodes to achieve accurate community testing. It uses the local similarity measure based on the clustering coefficient to identify the nearest neighbor of each node. Obtaining multiple groups of connected components by combining different node pairs in order to form the final initial community. Additionally make usage of community merging strategy to further optimize the community structure.

NBCD (Neighbour Based Community Detection) [38] is a neighborhood based community detection algorithm. It is based on two novel similarity measures using a similarity parameter α and a set of ground rules are proposed in this work. The similarity parameter α provides a choice

for the user to select the tightness of the nodes within the communities.

In the following tests, we apply the default value provided by the papers as the parameter value for each algorithm to compare the experimental results. Since the Ncut and FluidC algorithms require prior knowledge, we use the actual number of communities as the parameter. Each algorithm for each network runs independently 20 times in the same experimental environment, and finally, the average value of the results is obtained. The data packets used in the experiment are obtained from the internet, as shown in TABLE 3. We built our experimental environment, as shown in TABLE 4.

TABLE 3. Source of each algorithm.

algorithm	Experimental data sources
Ncut	http://scikit-learn.org/stable/modules/clustering.html
Infomap	http://igraph.org/
LPA	http://igraph.org/
WT	http://igraph.org/
FG	http://igraph.org/
Louvain	http://igraph.org/
MCL	http://micans.org/mcl/
FluidC	https://github.com/FerranPares-/Fluid-Communities
EDCD	https://According to the algorithm to achieve their own
SCD	https://github.com/SkBlaz/SCD
ASOCCA	https://github.com/BasdekD/community-detection-in-socialnetworks
NBCD	https://drive.google.com/drive/folders/1UD7La-t314M5L187KYGO8srDhT9fxTUsJ?usp=sharing

TABLE 4. Experimental environment.

name	configuration
computer	Desktop computer
CPU	Main frequency 3.3 GHz
processor	Intel Core i5
Memory	16.0 GB
Program environment	running Python 3.5.1

A. EVALUATION INDEX

To comprehensively investigate the achievement of various algorithms, this paper uses two popular evaluation methods to evaluate the accuracy of community detection. The following is a brief introduction of these two evaluation indicators.

Normalized mutual information [39] (NMI) is generally used in clustering to measure the similarity of two clustering results. It can objectively evaluate the accuracy of community classification compared with standard division. NMI is defined as follows:

$$\text{NMI}(a, b) = 2 \frac{I(a; b)}{H(a) + H(b)} \quad (4)$$

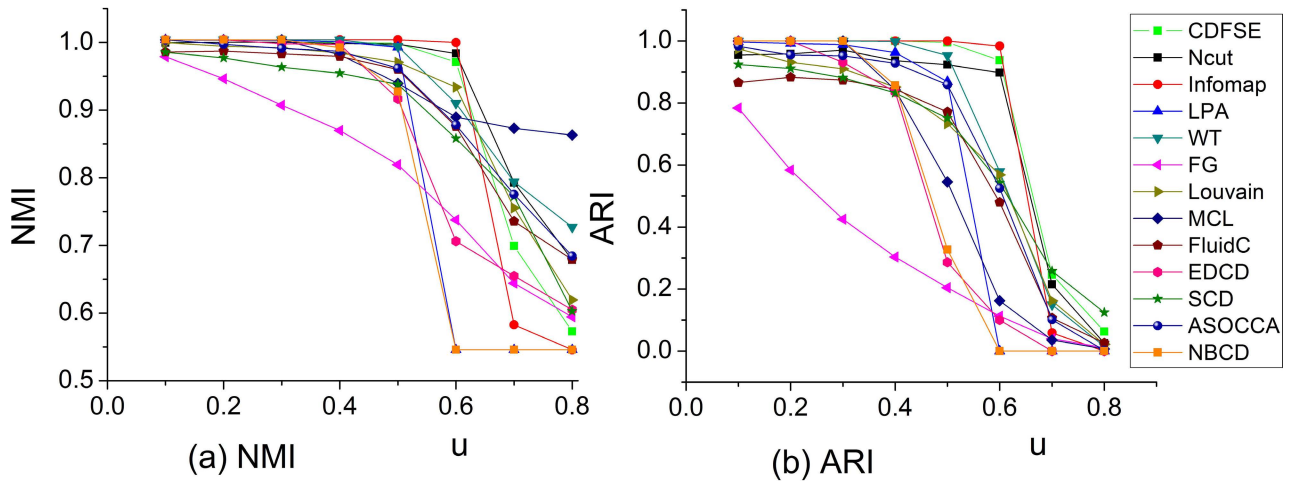


FIGURE 5. Performance comparison of all algorithms on the benchmark network when μ changes from 0.1 to 0.8.

where $I(a;b)$ represents the mutual information between a (true value) and b (Estimated value), and $H(a)$ denotes the entropy of a. The value range of NMI is 0 to 1. The larger the value is, the more accurate the division will be. $NMI = 1$ means that the predicted community is completely consistent with the real community division, while $NMI = 0$ indicates that it is completely different.

Another indicator is the adjusted random index (ARI) [40], which is often used to calculate the similarity between two samples. ARI is given as follows:

$$ARI = x_{11} - \frac{\frac{(x_{11}+x_{01})(x_{11}+x_{10})}{x_{00}}}{\frac{(x_{11}+x_{01})+(x_{11}+x_{10})}{2} - \frac{(x_{11}+x_{01})(x_{11}+x_{10})}{x_{00}}} \quad (5)$$

where x_{11} indicates the logarithm of points belonging to the same community in both the real community partition and the predicted community partition, x_{00} represents the logarithm of points that do not belong to the same community in the real community partition and the predicted community partition, x_{10} represents the logarithm of the points that are not in the same community in the predicted community division but belong to the same community in the real community division, and x_{01} represents the logarithm of points that do not belong to the same community in the real community division but belong to the same community in the predicted community division. The quality of community detection is evaluated by calculating the number of the same and different sample pairs allocated to the predicted community partition and the real community partition.

B. NETWORK GENERATION

In this paper, the well-known LFR model [41] is used to create a synthetic benchmark network. The LFR benchmark network assumes that the distribution in the network and the distribution of community size conform to the power-law distribution: power law index of degree τ_1 sequence, τ_2 Negative index of community size distribution. The edge of each node in the community is $1-\mu$ times its degree, while

the edge connected to the outer community node is μ times its degree, which is called the mixed parameter ($0 < \mu < 1$) and is used to dominate the complexity of community division. In addition, the model has several parameters: the number of nodes in the generated network (N); the average degree of the generated network (k); the minimum number of community nodes in the generated network (C_{min}); and the maximum number of community nodes (C_{max}). The average clustering coefficient (cc) of the network is generated.

To evaluate and compare the performance of various algorithms, the following parameters are set for the generated network: $N=1000$, $K=15$, $C_{min}=10$, $C_{max}=50$, $\tau_1 = 2$, and $\tau_2 = 1$, where μ varies from 0.1 to 0.8 to increase the complexity of network generation. Fig. 5 demonstrates the performance of these several methods on two distinctive indexes. In the NMI metric, when the parameter μ ranges from 0.1 to 0.5, except for the FG algorithm, which seems unsuitable for these networks, other algorithms can divide the generated network perfectly. As the parameter μ changes from 0.5 to 0.6, the performance of the LPA, FG and EDCD algorithms declines significantly, and their NMI values fall below 0.75, while the Infomap, CDFSE and Ncut algorithms are more robust as their NMI values are all above 0.95. The performance of other algorithms, such as Louvain, WT, MCL, ASOCCA, FluidC and SCD, with NMI values are between 0.85 and 0.95 in average. When $\mu > 0.6$, except for the MCL algorithm, which is still strong, the performance of the other algorithms is significantly reduced. In terms of the purity index, when μ ranges from 0.1 to 0.4, except for the FG algorithm, the other algorithms can divide the generated network well. When μ ranges from 0.4 to 0.5, the performance of the EDCD and MCL algorithms begins to decline significantly, and its ARI value drops below 0.6. When μ ranges from 0.5 to 0.6, the performance of the LPA algorithm declines sharply, and its ARI value tends to 0. Infomap, CDFSE and Ncut are still robust, and their ARI values are still above 0.95. The performance of WT, Louvain, SCD, ASOCCA and FluidC algorithms also declines to varying degrees, with NMI

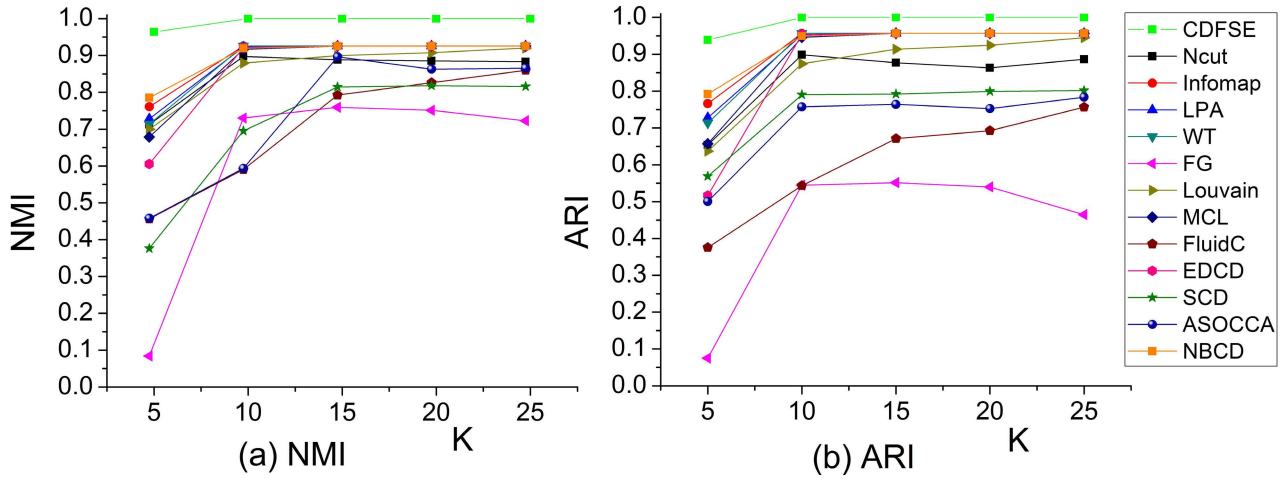


FIGURE 6. Performance of each algorithm and composite benchmark network when the average degree k changes from 5 to 25.

TABLE 5. Nmi performance of each algorithm and composite benchmark network when the number of nodes changes form 1000 to 50000.

NMI	CDFSE	Ncut	Infomap	LPA	WT	FG	Louvain	MCL	FluidC	EDCD	SCD	ASOCCA	NBCD
1000	1	0.990	1	0.999	1	0.945	0.991	1	0.960	1	0.960	0.999	1
2000	1	0.948	1	1	1	0.948	0.990	1	0.959	1	0.952	0.996	1
4000	1	0.902	1	1	1	0.929	0.977	1	0.953	1	0.933	0.986	1
6000	1	0.898	1	1	1	0.899	0.956	1	0.942	1	-	-	1
8000	1	0.884	1	1	1	0.894	0.947	1	0.932	-	-	-	1
10000	1	0.865	1	1	1	0.909	0.960	1	0.922	-	-	-	1
20000	1	0.833	1	1	1	0.873	0.936	1	0.912	-	-	-	1
50000	0.999	0.813	0.999	0.999	0.987	0.884	0.971	0.999	0.902	-	-	-	0.999

values between 0.5 and 0.7. When $\mu > 0.6$, the performance of all algorithms decreases rapidly, but SCD and CDFSE are still better than that of others.

Moreover, for the purpose of evaluating the effectiveness of the comparison algorithms in different community density networks, we apply fixed μ and change the average degree parameter K to generate the network, where $\mu = 0.1$ and k changes from 5 to 25. When k changes from 5 to 25, the performance of each comparison algorithm in different indicators is displayed in Fig. 6. It can be clearly observed that, our CDFSE algorithm achieves the best performance in terms of the NMI index and ARI index compared with other algorithms. When $k=5$, the NMI value of CDFSE is above 0.95, and the ARI value is above 0.92, while the NMI value and ARI value of the other algorithms are not greater than 0.8. When $k \geq 10$, the NMI value and ARI value of the CDFSE algorithm reach 1, and the performance of the other algorithms is not as good. The highest NMI value and ARI value of the CDFSE algorithm do not exceed 0.95. Additionally, the performance of the FG algorithm is worse because the FG algorithm is sensitive to the community density on the generation network. This effect may be caused by the resolution limitation of modularity [13].

Finally, in order to evaluate the effectiveness of the comparison algorithm in large-scale networks, we use the generation network with the increasing number of nodes. When the number of network nodes changes from 1000 to 50000, the performance of each comparison algorithm under different indicators is shown in TABLE 5-6. It can be clearly observed that compared with other algorithms, our CDFSE algorithm has achieved ideal community detection results in NMI index and ARI index.

C. REAL NETWORK

To test the effectiveness of the algorithm on real data sets, this paper also makes comparisons between several representative real networks in different sizes, which all have actual community structure information. The NMI and ARI indexes are still used to evaluate the community detection quality of each algorithm because these networks have known real community partition information. TABLE 7 describes the basic information for these real networks. These websites access real data sets: <http://snap.stanford.edu/data/>, <http://networkrepository.com/networks.php>, and <https://networkdata.ics.uci.edu/index.php>. Firstly, we provide a brief description of these networks.

TABLE 6. Ari performance of each algorithm and composite benchmark network when the number of nodes changes form 1000 to 50000.

ARI	CDFSE	Neut	Infomap	LPA	WT	FG	Louvain	MCL	FluidC	EDCD	SCD	ASOCCA	NBCD
1000	1	0.990	1	0.999	1	0.945	0.991	1	0.960	1	0.960	0.999	1
2000	1	0.938	1	1	1	0.799	0.980	1	0.952	1	0.952	0.997	1
4000	1	0.892	1	1	1	0.694	0.949	1	0.943	1	0.923	0.951	1
6000	1	0.873	1	1	1	0.590	0.871	1	0.932	1	-	-	1
8000	1	0.866	1	1	1	0.541	0.812	1	0.921	-	-	-	1
10000	1	0.821	1	1	1	0.552	0.852	1	0.920	-	-	-	1
20000	1	0.821	1	1	1	0.396	0.717	1	0.910	-	-	-	1
50000	0.999	0.801	0.9975	0.998	0.975	0.542	0.927	0.992	0.895	-	-	-	0.999

TABLE 7. Basic information for each real network.

Dataset	node number	edge number	average degree	clustering coefficient	community number
Football	115	613	10.661	0.40	12
Karate	34	78	4.588	0.57	2
Dolphin	62	159	5.13	0.30	2
Polbooks	105	441	8.4	0.49	3
Amazon	334,863	925,872	5.53	0.40	75,149

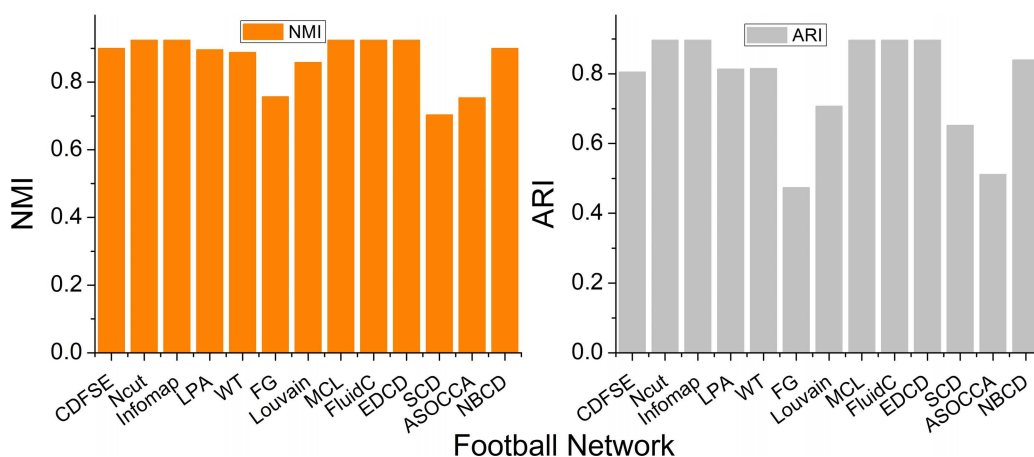


FIGURE 7. Performance of different comparison algorithms on the football network.

Football network: This network originated from American college football games. It consists of 115 nodes and 613 edges. The nodes represent each football team, and the edges represent regular games between the two connected teams. The network consists of 12 sports leagues (i.e., clubs), each of which contain approximately 8-12 teams.

Zachary’s karate network: This is a well-known network of karate clubs in the United States that represents the friendships among the 34 members of the club. The network is divided into two communities due to the different opinions of the leaders.

Dolphin network: A dolphin network constructed by D. Lusseau *et al.* after seven years’ observation of a group of bottlenose dolphins living in Doubtful Sound, New Zealand.

The nodes represent individual dolphins, and the edges represent frequent contact between two dolphins. The network has 62 nodes and 159 edges, which are divided into two communities.

Political book network: This is a network of books about American politics. These books are sold by online book-sellers on Amazon. Its nodes represent the books related to American politics sold in the Amazon online bookstore, as well as the number of readers who have purchased these two books. The nodes are divided into three types: l, n and c, which represent “liberal”, “conservative” and “centrist”, respectively. Therefore, the network is divided into three communities. The network consists of 105 nodes and 441 edges.

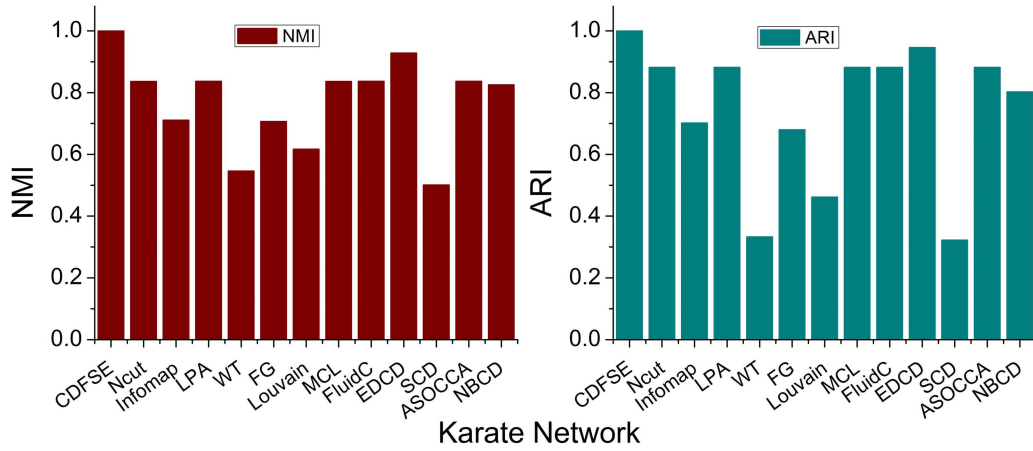


FIGURE 8. Performance of different comparison algorithms on the karate network.

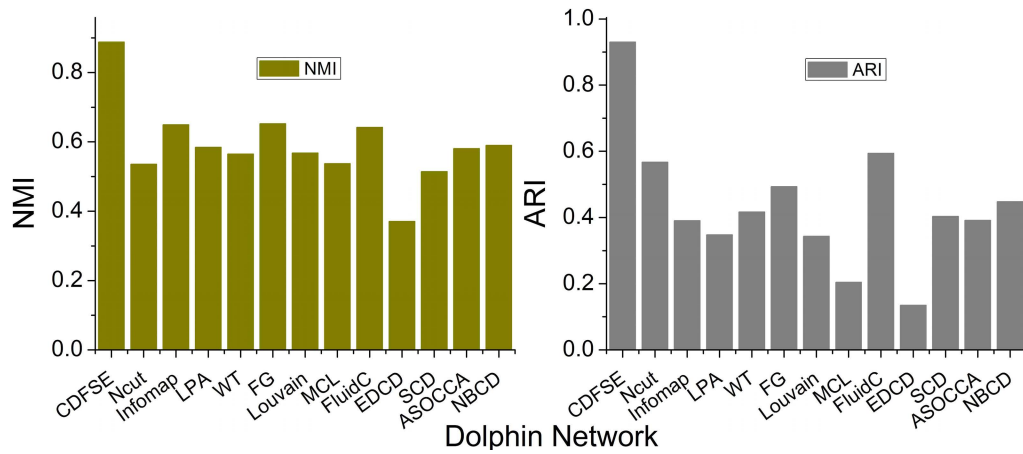


FIGURE 9. Performance of different comparison algorithms on the dolphin network.

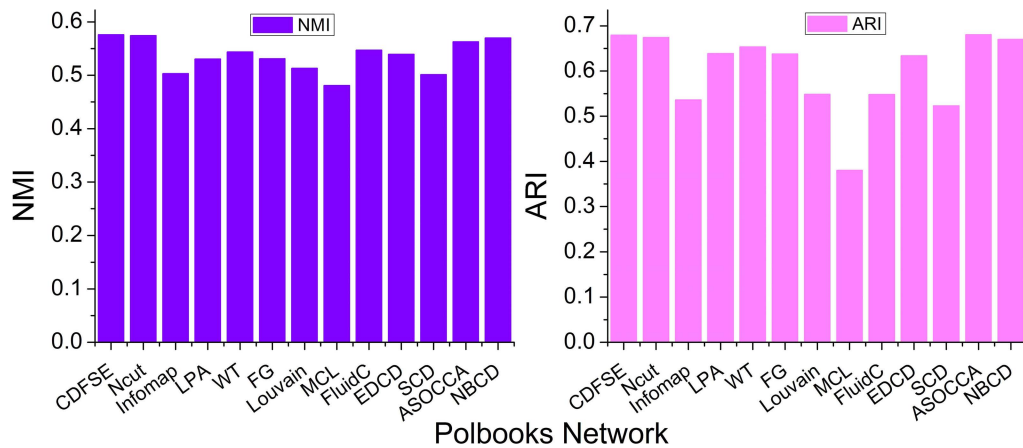


FIGURE 10. Performance of different comparison algorithms on the political book network.

Amazon network: This network is based on the online sales data of products on the Amazon website. The nodes represent products, and the edges represent the connection between frequently purchased goods. Each commodity belongs to a

category that is considered as a real community. The network consists of 925872 edges and 334863 nodes.

Generally, the CDFSE algorithm’s performance on the football network is not outstanding, but it performs very well

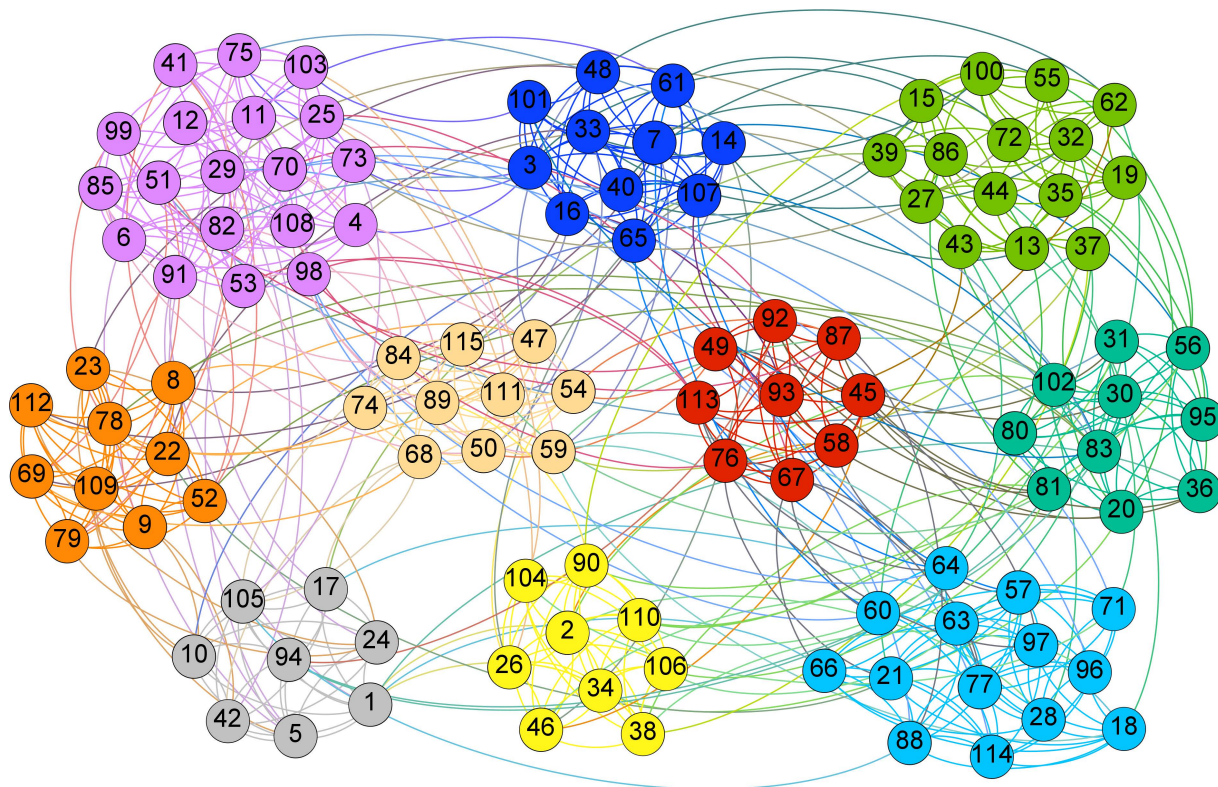


FIGURE 11. Community detection results of CDFSE on the United States football network; a total of 10 communities were detected.

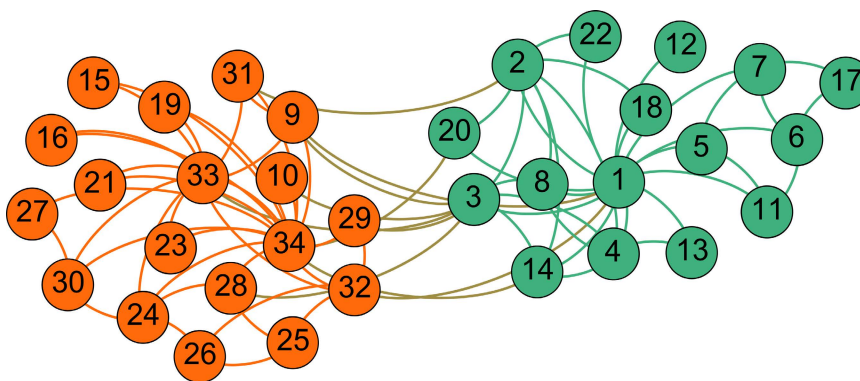


FIGURE 12. Community detection results of CDFSE on the karate network; two communities are detected.

on the karate club network, political book network and dolphin network. The algorithm also achieved good community detection quality.

On the American college football network, most algorithms, such as Infomap, MCL, Ncut, EDCD, FluidC, CDFSE, WT and Louvain, obtained outstanding clustering results due to the high average degree as shown in Fig. 7. The Infomap and MCL methods achieved high NMI and ARI values. These two algorithms also detected 12 high-quality communities. Compared with the above two algorithms, the CDFSE algorithm is slightly inferior (NMI=0.90, ARI=0.81) and detected 10 high-quality communities as shown in Fig. 11. The performance of the FG algorithm on

this network is not ideal and it achieved a low index value (NMI=0.76, ARI=0.47). These results are consistent with the experimental results.

On the karate club network, the CDFSE algorithm achieved the highest index score (NMI=1, ARI=1) with the EDCD algorithm taking second place (NMI=0.93, ARI=0.95). The index scores of the other algorithms are no higher than 0.9 as shown in Fig. 8. Since the NMI and ARI indexes of the CDFSE algorithm are both 1, the community structure detected by the algorithm is completely consistent with the real network partition. In fact, CDFSE successfully detected two communities as shown in Fig. 12, and their community organizations are completely

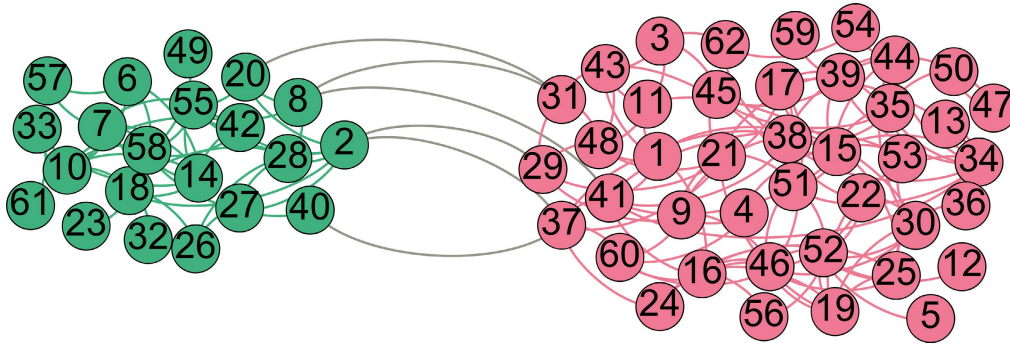


FIGURE 13. Community detection results of CDFSE on the dolphin network; two communities are detected.

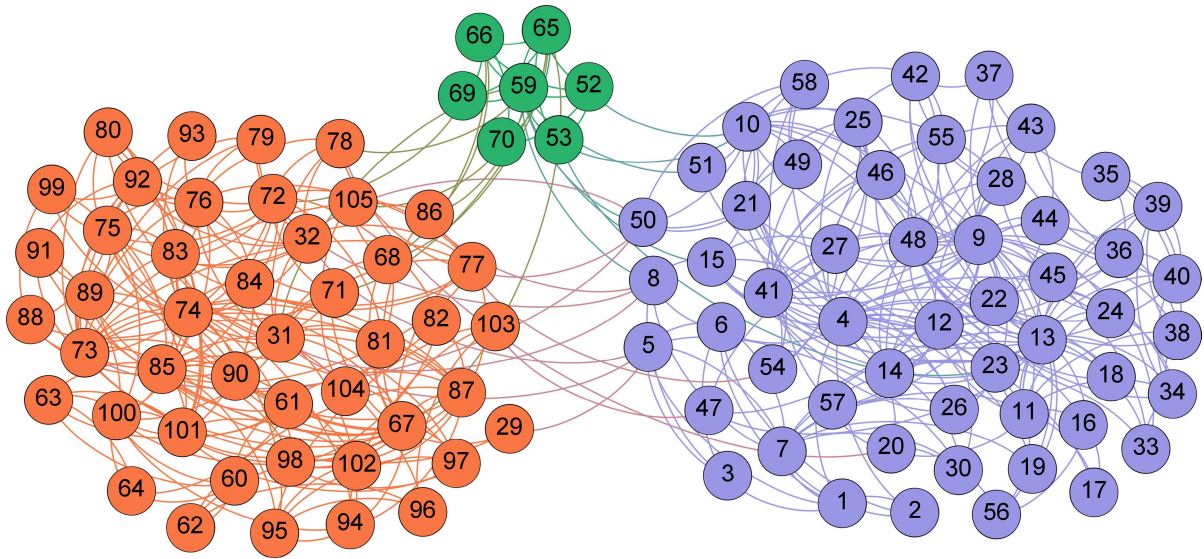


FIGURE 14. Community detection results of CDFSE on the US political book network; three communities are detected.

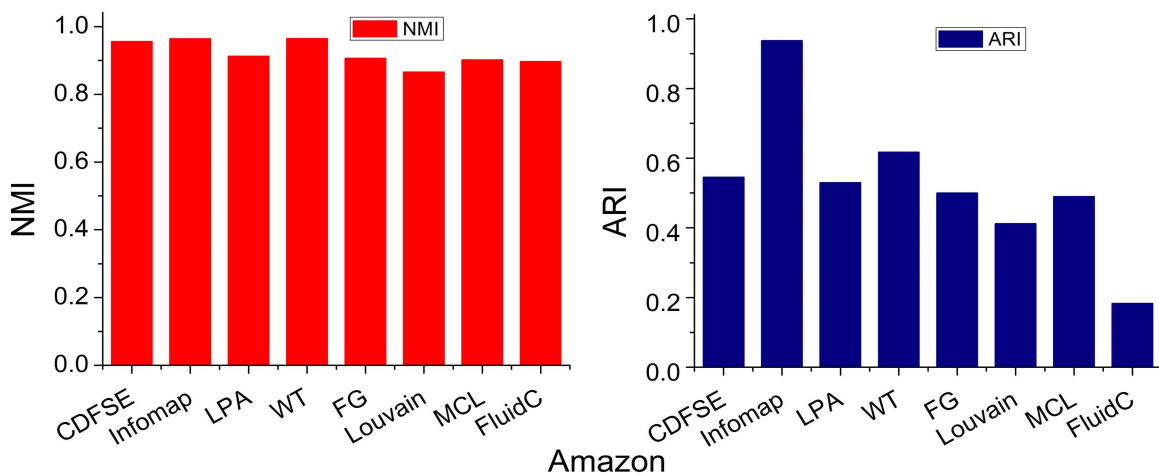


FIGURE 15. Performance of different comparison algorithms on the Amazon dataset.

consistent with the real partitions. In this network, CDFSE attained good detection because the CDFSE method considers not only the attractiveness of individuals but

considers the attractiveness of groups as well. Of course, the MCL, ASOCCA, EDCD, LPA, Ncut, and FluidC methods also achieved good performance. It's worth noticing

that MCL and ASOCCA also successfully detected two communities.

In the dolphin network, the CDFSE algorithm once again outperformed the other algorithms in terms of the NMI index value which reached 0.89, and ARI index value which reached 0.93. Among the other algorithms, the highest NMI index value does not exceed 0.65 and the highest ARI index value does not exceed 0.57, which is more obvious from the column diagram shown in Fig. 9. CDFSE successfully detected two communities, which is consistent with the real network as presented in Fig. 13, the number of communities detected by the other algorithms ranges from 4 to 12. The quality of community detection of some algorithms is not high, and in particular, the NMI value of the EDCD algorithm is the lowest, which indicates that the community division of most nodes is incorrect.

On the political book network, the index score of the CDFSE algorithm achieves the highest again (NMI=0.58, ARI=0.68). With the exception of the MCL algorithm, the index scores of the other algorithms are also exceptional, all of which are above 0.50. Therefore, most algorithms have achieved good clustering as shown in Fig. 10. Here, CDFSE detects three communities as shown in Fig. 14. Other algorithms detect four to six communities.

This paper employs Amazon dataset to make judgement on the performance of the comparison algorithms under large-scale networks. Due to the large number of online communities, this paper selects top-5000 highest-quality actual communities with the largest number of members for comparison. On the Amazon network, the CDFSE, Infomap and WT algorithms achieve the best performance with NMI index values above 0.95; furthermore, the ARI index value of Infomap is as high as 0.94 which is far higher than that of the other algorithms as shown in Fig. 15. There are 20057 communities detected by the CDFSE algorithm and 17296 communities detected by the Infomap algorithm. However, due to the high time complexity, the Ncut, EDCD, SCD and ASOCCA algorithms cannot operate on the Amazon network.

According to the experimental results, both the Infomap and CDFSE algorithms are appropriate for networks of different sizes. They produce better performance when generating benchmark networks and obtain higher index values. However, CDFSE is better than Infomap in real networks, as it obtains the best social division.

D. EXECUTION TIME ANALYSIS

To evaluate the scalability of the CDFSE algorithm at the network scale, this paper takes advantages from the LFR model to generate benchmark networks in different sizes with a fixed average degree of $k = 15$. When the number of nodes changes from 2K to 50K, the CDFSE algorithm apparently has better performance than the other comparison algorithms. The CDFSE algorithm is more advantageous when dealing with small and medium-sized networks because its time complexity is $O(k^2 \cdot n)$ as shown in Fig. 16. When the number of nodes changes from 1K to 100M, the CDFSE algorithm

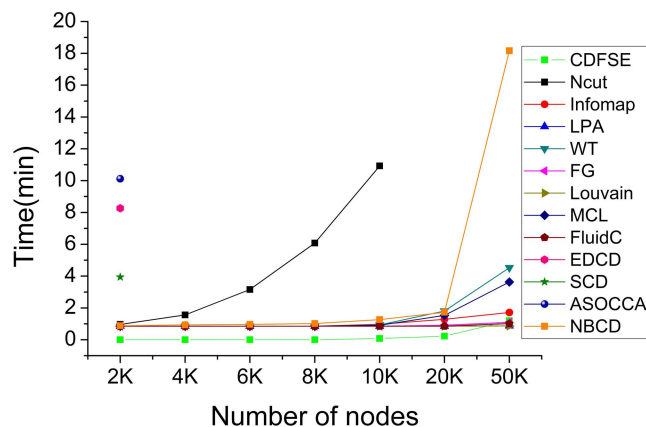


FIGURE 16. Run time comparison of all algorithms when the number of nodes ranges from 2K to 50K.

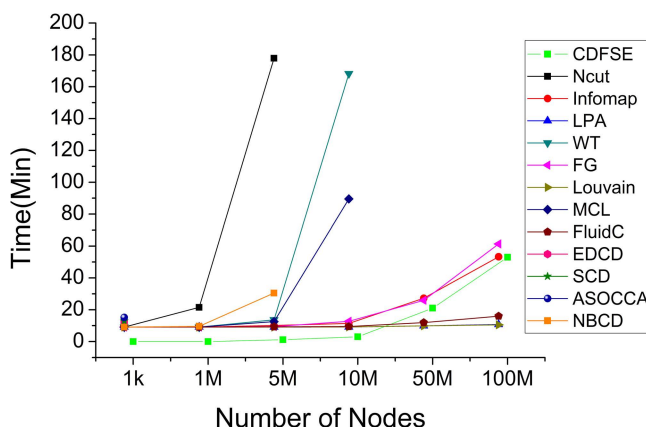


FIGURE 17. Run time comparison of all algorithms when the number of nodes ranges from 1K to 100M.

is faster when compared with the Ncut, EDCD, WT, FG, and MCL algorithms; additionally, for the reason that the value of K is relatively small, the CDFSE algorithm performs well when dealing with large-scale networks. However, the CDFSE algorithm is slower than LPA, Louvain, and FluidC as shown in Fig. 17. In view of the poor quality of community detection by these three algorithms and the stability problems of the LPA and FluidC algorithms, the CDFSE algorithm is preferable.

IV. CONCLUSION

Community detection has always been a argument of interest. Inspired by fish networks found in nature, this paper proposes a community detection model according to fish school effect and presents a community detection algorithm named, CDFSE according to the model. The central idea behind CDFSE is to regard the network as a dynamic system and research its dynamics with time. Individuals in the network form subgroups because of their commonness and characteristics, and subgroups will attract other individuals to form large groups because of their own attraction. As time goes on, all individuals will eventually be attracted to different groups, forming a stable community structure. We make comparisons between CDFSE and several other representative community

detection algorithms by applying them on Football, Karate, Dolphin, Polbooks and Amazon networks. The experimental results show that the CDFSE algorithm performed well on different networks and it achieved better performance than the other comparison algorithms. The main performance is as follows: firstly, this algorithm has higher efficiency and better community detection quality than others which mentioned above, secondly, the algorithm does not need to set parameters and is more simple and convenient; thirdly, the time complexity of the algorithm is small, and it can be applied to large-scale network; next, the algorithm is derived from nature, in line with the laws of nature, closer to the real network, so it can be better applied to community detection. The limitations of this paper is that only time complexity been added to verify its effectiveness when comparing with the classical method on the actual data set, and the space complexity is what we need to consider for further explore.

REFERENCES

- [1] H.-J. Li, L. Wang, Y. Zhang, and M. Perc, "Optimization of identifiability for efficient community detection," *New J. Phys.*, vol. 22, no. 6, Jun. 2020, Art. no. 063035.
- [2] Z. Sun, Y. Sun, X. Chang, Q. Wang, X. Yan, Z. Pan, and Z.-P. Li, "Community detection based on the Matthew effect," *Knowl.-Based Syst.*, vol. 205, Oct. 2020, Art. no. 106256.
- [3] Z. Sun, B. Wang, J. Sheng, Z. Yu, R. Zhou, and J. Shao, "Community detection based on information dynamics," *Neurocomputing*, vol. 359, pp. 341–352, Sep. 2019.
- [4] I. Himelboim, M. A. Smith, L. Rainie, B. Shneiderman, and C. Espina, "Classifying Twitter topic-networks using social network analysis," *Social Media + Soc.*, vol. 3, no. 1, Jan. 2017, Art. no. 205630511769154.
- [5] S. P. Borgatti, A. Mehra, D. J. Brass, and G. Labianca, "Network analysis in the social sciences," *Science*, vol. 323, no. 5916, pp. 892–895, 2009.
- [6] V. Colizza, A. Barrat, M. Barthélemy, and A. Vespignani, "The modeling of global epidemics: Stochastic dynamics and predictability," *Bull. Math. Biol.*, vol. 68, no. 8, pp. 1893–1921, Nov. 2006.
- [7] A.-L. Barabási and Z. N. Oltvai, "Network biology: Understanding the cell's functional organization," *Nature Rev. Genet.*, vol. 5, no. 2, pp. 101–113, Feb. 2004.
- [8] T. Ma, Q. Liu, J. Cao, Y. Tian, A. Al-Dhelaan, and M. Al-Rodhaan, "LGIEM: Global and local node influence based community detection," *Future Gener. Comput. Syst.*, vol. 105, pp. 533–546, Apr. 2020.
- [9] J. Shao, Z. Zhang, Z. Yu, J. Wang, Y. Zhao, and Q. Yang, "Community detection and link prediction via cluster-driven low-rank matrix completion," in *Proc. 28th Int. Joint Conf. Artif. Intell.*, Aug. 2019, pp. 3382–3388.
- [10] F. Schweitzer, G. Fagiolo, D. Sornette, F. Vega-Redondo, A. Vespignani, and D. R. White, "Economic networks: The new challenges," *Science*, vol. 325, no. 5939, pp. 422–425, 2009.
- [11] C. Gao, Y. Zhao, R. Wu, Q. Yang, and J. Shao, "Semantic trajectory compression via multi-resolution synchronization-based clustering," *Knowl.-Based Syst.*, vol. 174, pp. 177–193, Jun. 2019.
- [12] S. Fortunato and D. Hric, "Community detection in networks: A user guide," *Phys. Rep.*, vol. 659, pp. 1–44, Nov. 2016.
- [13] M. Girvan and M. E. J. Newman, "Community structure in social and biological networks," *Proc. Nat. Acad. Sci. USA*, vol. 99, no. 12, pp. 7821–7826, Jun. 2002.
- [14] M. E. J. Newman, "Fast algorithm for detecting community structure in networks," *Phys. Rev. E, Stat. Phys. Plasmas Fluids Relat. Interdiscip. Top.*, vol. 69, no. 6, Jun. 2004, Art. no. 066133.
- [15] S. Bilal and M. Abdelouahab, "Evolutionary algorithm and modularity for detecting communities in networks," *Phys. A, Stat. Mech. Appl.*, vol. 473, pp. 89–96, May 2017.
- [16] T. Kanungo, D. M. Mount, N. S. Netanyahu, C. D. Piatko, R. Silverman, and A. Y. Wu, "A local search approximation algorithm for k -means clustering," in *Proc. 18th Annu. Symp. Comput. Geometry*, 2002, vol. 28, nos. 2–3, pp. 10–18.
- [17] U. N. Raghavan, R. Albert, and S. Kumara, "Near linear time algorithm to detect community structures in large-scale networks," *Phys. Rev. E, Stat. Phys. Plasmas Fluids Relat. Interdiscip. Top.*, vol. 76, no. 3, Sep. 2007, Art. no. 036106.
- [18] R. Hosseini and A. Rezvanian, "AntLP: Ant-based label propagation algorithm for community detection in social networks," *Social Netw.*, vol. 5, no. 1, pp. 34–41, 2020.
- [19] U. K. Roy, P. K. Muhuri, and S. K. Biswas, "NeSiFC: Neighbors' similarity-based fuzzy community detection using modified local random walk," *IEEE Trans. Cybern.*, early access, Jun. 24, 2021, doi: 10.1109/TCYB.2021.3071542.
- [20] X. Li, Y. Chen, and J. Xu, "Convex relaxation methods for community detection," *Stat. Sci.*, vol. 36, no. 1, pp. 2–15, Feb. 2021.
- [21] Y. Zhang, Y. Xiong, Y. Ye, T. Liu, W. Wang, Y. Zhu, and P. S. Yu, "SEAL: Learning heuristics for community detection with generative adversarial networks," in *Proc. 26th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, Aug. 2020, pp. 1103–1113.
- [22] L. Yin, Q. Deng, and H. Chen, "Community integration and community anti risk ability," *Soc. Governance*, no. 6, pp. 65–70, 2020.
- [23] Y. Zhou, "Study on self-organization behavior of fish under hydrodynamic action," M.S. thesis, School Aeronaut. Astronaut., Zhejiang Univ. Zhejiang, Hangzhou, China, 2020.
- [24] G. Lu, Z. Wen, W. Wu, Y. Sun, H. Dong, and Q. Yan, "Research on highway freight car formation algorithm based on fish school effect," *Comprehens. Transp.*, vol. 43, no. 5, pp. 56–62, 2021.
- [25] D. Tian and L. Kang, "Research on unmanned vehicle formation algorithm based on Fish swarm effect," *Unmanned Syst. Technol.*, vol. 1, no. 104, pp. 62–67, 2018.
- [26] L. A. Adamic and E. Adar, "Friends and neighbors on the web," *Social Netw.*, vol. 25, no. 3, pp. 211–230, Jul. 2003.
- [27] J. Shi and J. Malik, "Normalized cuts and image segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 22, no. 8, pp. 888–905, Aug. 2000.
- [28] M. Rosvall and C. T. Bergstrom, "Maps of random walks on complex networks reveal community structure," *Proc. Nat. Acad. Sci. USA*, vol. 105, no. 4, pp. 1118–1123, Jan. 2008.
- [29] U. N. Raghavan, R. Albert, and S. Kumara, "Near linear time algorithm to detect community structures in large-scale networks," *Phys. Rev. E, Stat. Phys. Plasmas Fluids Relat. Interdiscip. Top.*, vol. 76, Sep. 2007, Art. no. 036106.
- [30] P. Pons and M. Latapy, "Computing communities in large networks using random walks," *J. Graph Algorithms Appl.*, vol. 10, no. 2, pp. 191–218, 2006.
- [31] A. Clauset, M. E. J. Newman, and C. Moore, "Finding community structure in very large networks," *Phys. Rev. E, Stat. Phys. Plasmas Fluids Relat. Interdiscip. Top.*, vol. 70, no. 6, Dec. 2004, Art. no. 066111.
- [32] V. D. Blondel, J.-L. Guillaume, R. Lambiotte, and E. Lefebvre, "Fast unfolding of communities in large networks," *J. Stat. Mech., Theory Exp.*, vol. 2008, no. 10, Oct. 2008, Art. no. P10008.
- [33] S. Van Dongen, "Graph clustering by flow simulation," M.S. thesis, Dept. Faculteit Letteren, Univ. Utrecht, Utrecht, The Netherlands, 2000.
- [34] F. Parés, D. G. Gasulla, A. Vilalta, J. Moreno, E. Ayguadé, J. Labarta, U. Cortés, and T. Suzumura, "Fluid communities: A competitive, scalable and diverse community detection algorithm," in *Complex Networks & Their Applications VI*. Cham, Switzerland: Springer, 2018, pp. 229–240.
- [35] X. Chen and J. Li, "Community detection in complex networks using edge-deleting with restrictions," *Phys. A, Stat. Mech. Appl.*, vol. 519, pp. 181–194, Apr. 2019.
- [36] B. Škrlić, J. Kralj, and N. Lavrač, "Embedding-based Silhouette community detection," *Mach. Learn.*, vol. 109, no. 11, pp. 2161–2193, Nov. 2020.
- [37] X. Pan, G. Xu, B. Wang, and T. Zhang, "A novel community detection algorithm based on local similarity of clustering coefficient in social networks," *IEEE Access*, vol. 7, pp. 121586–121598, 2019.
- [38] S. Sahu and T. S. Rani, "A neighbour-similarity based community discovery algorithm," Univ. of Hyderabad, Tech. Rep., 2022, doi: 10.2139/ssrn.4051308.
- [39] A. Strehl and J. Ghosh, "Cluster ensembles—A knowledge reuse framework for combining multiple partitions," *J. Mach. Learn. Res.*, vol. 3, pp. 583–617, Dec. 2002.
- [40] W. M. Rand, "Objective criteria for the evaluation of clustering methods," *J. Amer. Stat. Assoc.*, vol. 66, no. 336, pp. 846–850, Dec. 1971.
- [41] A. Lancichinetti, S. Fortunato, and F. Radicchi, "Benchmark graphs for testing community detection algorithms," *Phys. Rev. E, Stat. Phys. Plasmas Fluids Relat. Interdiscip. Top.*, vol. 78, no. 4, Oct. 2008, Art. no. 046110.

• • •