

Received February 20, 2020, accepted March 24, 2020, date of publication March 31, 2020, date of current version April 14, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.2984638

A Compression-Based Multi-Objective Evolutionary Algorithm for Community Detection in Social Networks

ZHIYUAN LIU¹, YINGHONG MA¹, AND XIUJUAN WANG²

¹Business School, Shandong Normal University, Jinan 250000, China

²School of Business, Dalian University of Technology, Dalian 116024, China

Corresponding author: Yinghong Ma (yinghongma71@163.com)

This work was supported in part by the National Natural Science Foundation of China under Grant 71471106 and Grant 71701115, in part by the National Nature Science Foundation of Shandong Province under Grant ZR2017MF058, and in part by the Higher School Science and Technology Foundation of Shandong Province under Grant J17KA172.

ABSTRACT Community detection is a key aspect for understanding network structures and uncovers the underlying functions or characteristics of complex systems. A community usually refers to a set of nodes that are densely connected among themselves, but sparsely connected to the remaining nodes of the network. Detecting communities has been proved to be a NP-hard problem. Therefore, evolutionary based optimization approaches can be used to solve it. But a primary challenge for them is the higher computational complexity when dealing with large scale networks. In this respect, a COMpression based Multi-Objective Evolutionary Algorithm with Decomposition (Com-MOEA/D) for community detection is proposed where the network is first compressed to a much more smaller scale by exploring network topologies. After that, a framework of multi-objective evolutionary algorithm based on decomposition is applied, in which a local information based genetic operator is proposed to speed up the convergence and improve the accuracy of the Com-MOEA/D algorithm. Experimental results on both real world and synthetic networks show the superiority of the proposed method over several state-of-the-art community detection algorithms.

INDEX TERMS Network compression, multi-objective optimization, community detection, social networks.

I. INTRODUCTION

Many real-world systems such as information, biological, transportation systems and social networks can be modelled as networks, which appropriately describe system elements and the relationships between them. Exploring network topology structures helps us to better understand system functions and characteristics, and community detection has become an important tool to study the mesoscopic structures of networks. Communities usually refer to a set of nodes that have denser connections inside the set but sparser links outside. For example, in friendship networks, a community may refer to a collection of people that are tightly connected by common interest; in biological networks, a community usually represents one functional unit. Therefore, detecting communities not only has great theoretical importance, but

also has wide applications in real-world such as recommendation systems [1], link prediction [2], epidemic spreading modelling on networks [3], detection of terrorist groups from networks [4] etc.

Recently, a number of work with regard to the design of efficient community detection algorithms have been developed. Detailed surveys describing these methods can be found in [5]–[8].

Community detection has been proved to be a NP-Hard problem [9], so it cannot be strictly solved by applying deterministic methods. Therefore, a number of community detection algorithms including optimization based and heuristic-based algorithms are proposed. For optimization based algorithms, to optimize modularity is a kind of popular method. Newman [10] first put forward a fast algorithm to efficiently find communities in large scale networks by joining communities in pairs with the goal of optimizing modularity function. Inspired by Newman's idea, the

The associate editor coordinating the review of this manuscript and approving it for publication was Murilo S. Baptista.

Clauset *et al.* [11] proposed a modified version that is called fast greedy algorithm. Blondel *et al.* [12] gave a much more faster community detection algorithm by optimizing local modularity, which is usually named as Louvain. As to the heuristic methods, the label propagation algorithm (LPA) [13] is a nearly linear time method even though the results are not always steady. Infomap [14] is also an effective algorithm that makes use of information theory. There also exist some algorithms that apply local information of networks, such as the divide and agglomerate algorithm [15] that employs nodes' similarities and the dynamic expansion algorithm that considers neighbor's power [16].

As a representative of optimization algorithms, evolutionary computation based community detection algorithms have displayed competitive performance in finding communities. They have powerful search and optimization abilities inspired by the process for natural evolution. Such algorithms can be classified into two categories based on the number of objectives, which are single-objective and multi-objective optimization based community detection algorithms. As to the single-objective optimization methods, the community detection problem is only formulated to one objective function, and then this function is solved by applying iterative evolutionary computation process. For example, GACD [17] obtained good community results by optimizing modularity function Q , while Tang *et al.* [18] proposed to optimize Surprise. Meme-net [19] got network partitions by maximizing modularity density. The other category is based on multi-objective evolutionary algorithms (MOEAs), where more than one objective is simultaneously optimized. Pizzuti [20] first proposed to detect communities by optimizing two objectives using evolutionary algorithm, which is called MOGA-Net. It has been shown that its performance is better than single-objective EA algorithms. After that, a number of MOEAs have been proposed. For example, Gong *et al.* [21] developed a community detection algorithm driven by the framework of MOEA based on decomposition (MOEA/D-Net). Recently, several other decomposition based MOEAs, named LMOEA [22], MODBSA/D [23] and RMOEA [24], have been developed. They all adopted the general framework of decomposition based multi-objective evolutionary algorithms to detect communities, and experimental results showed their effectiveness and efficiency.

MOEA based community detection algorithms have advantages in two aspects. At first, the adoption of multi-objectives means the evaluation of communities from different views, which can overcome some limitations that are introduced by single-objective. For example, only optimizing the modularity function can bring about resolution limit [25]. Besides, the MOEA based algorithms return a set of solutions instead of only one solution. Each solution represents a tradeoff among multiple objectives and we get a series of community detection results, thus can choose different solutions according to different requirements.

The MOEAs have demonstrated their competitive performance in coping with the community detection problem,

and in this paper, we also adopt the general framework of decomposition based MOEA (MOEA/D) and propose a network compression based multi-objective evolutionary algorithm, termed Com-MOEA/D, to further improve the quality of detected communities as well as the efficiency. Specifically, the main contributions of the paper are summarized as follows.

- A network compression method based on node pairs' similarities is proposed in the preliminary stage. Nodes with relative higher similarities are pre-partitioned into a group and then generate each group to be a new node in the compressed network. Thus the original network is transformed into a smaller scale one. This is efficient for large scale networks.
- Based on the compressed network, a multi-objective evolutionary algorithm based on decomposition is proposed to solve community detection problems. It adopts the same framework as MOEA/D, and has two contradictory sub-objectives.
- A random walk based population initialization method is adopted, which helps to generate competitive individuals.
- A local information based mutation strategy is proposed in evolutionary process, which aims to improve the accuracy and accelerate the convergency of the algorithm.
- The effectiveness and efficiency are tested on both synthetic and real world networks. Experimental results show that the proposed algorithm performs better than the state-of-the-art MOEAs and some classical algorithms for community detection in social networks.

The remainder of this paper is organized as follows. In section II, the community detection problem is formulated and related work on MOEAs is introduced. Section III gives a detailed description of the Com-MOEA/D algorithm. After that, Section IV describes the efficiency and accuracy of Com-MOEA/D algorithm comparing with state-of-art methods. Not only real-world networks but also synthetic networks are tested. Finally, Section V concludes the paper and discusses the future work.

II. PROBLEM DEFINITION AND RELATED WORK

In this section, the problem definition for community detection based on multi-objective optimization methods is first given, and then the related work on MOEA based community detection algorithms is presented.

A. PROBLEM DEFINITION

The main task of community detection algorithm is to divide the whole network into subgraphs requiring that nodes inside the subgraphs are tightly connected, whereas nodes in different subgraphs should be sparsely linked. The nodes in such subgraphs constitute communities. Therefore, it is natural to formulate community detection problem as a multi-objective optimization problem, where the objectives represent the requirements of communities.

In this paper, the network is represented as a graph $G(V, E)$, where $V = \{v_1, v_2, \dots, v_N\}$ denotes a finite set of nodes that models objects and $E = V \times V (|E| = m)$ represents the set of edges describing relations between objects. $A = (A_{i,j})_{N \times N}$ is the adjacency matrix of G , where $A_{i,j} = 1$ if there is a connection between nodes v_i and v_j , and 0 otherwise. Communities in network G are usually subgraphs of nodes densely connected interior, but sparsely linked with others. Let $\mathcal{C} = (C_1, C_2, \dots, C_{N_c})$ be a set of N_c communities of G . Note that we focus on the non-overlapping community detection in undirected and unweighted graphs, i.e., $C_i \cap C_j = \emptyset, \forall i, j \in \{1, 2, \dots, N_c\}$.

Since community detection in networks is a NP-hard problem, approximation algorithms can be used to solve it. One of them is the multi-objective optimization based algorithms, which evaluate community qualities from various criterions through defining multi-objectives and have good performance. Here we formulate this problem into a bi-objective minimization problem, in which two objectives are termed as Kernel K-Means (*KKM*) and Ratio Cut (*RC*). They are formally defined as follows:

$$\text{minimize } F(\mathcal{C}) = (KKM(\mathcal{C}), RC(\mathcal{C})) \quad (1)$$

$$KKM(\mathcal{C}) = 2(N - N_c) - \sum_{i=1}^{N_c} \frac{L(C_i, C_i)}{|C_i|} \quad (2)$$

$$RC(\mathcal{C}) = \sum_{i=1}^{N_c} \frac{L(C_i, \bar{C}_i)}{|C_i|}. \quad (3)$$

KKM is first proposed by Angelini *et al.* [26]. From the definition, we can see that the right hand of *KKM* represents the sum of the density of the link of intra-communities, and *RC* can be treated as the sum of the density of the link of inter-communities. What's more, it has been shown that *KKM* is a decreasing function of the number of communities. However, *RC* has an opposite trend. That's to say, these two objectives are conflict with each other. Minimizing the two objectives can ensure the requirements for communities. So far, these two objectives have been used in MOEA based community detection algorithms such as [27], [28] and achieve relative good performance. Therefore, we also adopt *KKM* and *RC* as objectives.

B. RELATED WORK ON EVOLUTIONARY BASED COMMUNITY DETECTION ALGORITHMS

In recent years, solving community detection problem by applying evolutionary computation based methods have attracted great interests of researchers. Evolutionary computation that is inspired by the process of natural evolution has shown great advantages in coping with complex optimization problems [29], especially for multi-objective optimization problems. Detecting communities in networks can also be defined as an optimization problem. Therefore, a number of methods adopting evolutionary algorithms to tackle it are developed. These methods are classified into single objective and multi-objective methods.

Single objective optimization algorithms optimize only one objective function. The main differences among them lie in the representation mechanisms, evaluation functions, different optimization methods and specific search operators. Most single-objective evolutionary algorithms detecting communities are based on the genetic algorithm (GA). GA-net [30] is a genetic algorithm for optimizing a community score function, and it adopted the locus-based representation scheme which well preserved the network topology. After that, Li *et al.* [31] proposed a GA based community detection algorithm with the optimization of modularity function Q , which applied the label-based representation mechanism. In [32], a heuristic initialization procedure was proposed to diversify the population along the whole search space. Besides, an effective mutation operator was also designed to improve the efficiency and stability of the algorithm. Gong *et al.* [33] developed a memetic algorithm with optimization of modularity density [34], in which it used the label based representation approach and two-way crossover as well as the neighbor-based mutation. In [35], a multi-agent genetic community detection algorithm was proposed by optimizing modularity function.

Compared with single-objective optimization algorithms, MOEAs have also shown their superiorities in community detection. The first work of formulating community detection as a multi-objective problem was developed by Pizzuti [20], termed MOGA-Net, where two objectives, community score and community fitness, were simultaneously considered. Its optimization method was the fast non-dominated sorting genetic algorithm (NSGA-II) [36]. From the detailed experimental results shown in [37], we can see that MOGA-Net outperformed the single-objective EA based community detection algorithms, and got more accurate community structures. Due to the competitive performance, a number of MOEA based methods have been developed to improve the quality of community detection.

Inspired by [20], Shi *et al.* [38] proposed a multi-objective community detection (MOCD) algorithm based on PESA-II, where two-phase strategy was proposed. At first, MOCD optimized two conflicting objectives (intra-link and inter-link), then returned a set of Pareto optimal solutions. After that, solution selection is performed. Experimental results also showed its effective and efficiency in both real world and synthetic networks. What's more, Gong *et al.* [21] developed a multi-objective evolutionary algorithm based on decomposition (MOEA/D-Net) to solve the community detection problem. It defined two objectives, negative ratio association (*NRA*) and ratio cut (*RC*). Empirical results indicated that the performance of MOEA/D-Net was not only better than the single-objective EAs, but also superior over the non-EA community detection algorithms.

Due to the advantage of MOEA/D-Net, more decomposition based MOEAs for community detection have been developed. Gong *et al.* [27] further developed a MOEA community detection algorithm adopting discrete particle swarm optimization, named MODPSO. The main novelty of this

work lied in the use of a specific solution encoding, as well as the redefinition of the velocity that accelerated the search of the algorithm. Experimental results also displayed its superiority. More recently, two other decomposition based MOEAs, named LMOEA [22] and MODTLBO/D [39], were proposed, where a local information based individual updating strategy and discrete teaching-learning based optimization algorithm are applied to get high quality communities.

There also exist other MOEA based community detection algorithms, which were designed based on different swarm intelligence algorithms. For example, Chen *et al.* [40] proposed an algorithm based on ant colony optimization which applies the associate degree between nodes as a heuristic function. The HSA-CLS [41] is a modified harmony search algorithm with a chaotic local search for community detection. And firefly mechanism is also adopted in community detection problem, such as [42]. Those bio-inspired mechanisms are well suited for network partition.

The MOEAs mentioned above have demonstrated their competitiveness in a variety of complex networks, but they face great challenges when dealing with large scale networks. This is mainly due to the representation difficulty and large search spaces. In order to cope with this problem, a network compression based MOEA/D algorithm, termed Com-MOEA/D is proposed. At first, a compression method is proposed which condensed the original network into a smaller scale one. Based on the compressed network, we continue to employ the general framework of MOEA/D to detect communities. What's more, a local information based mutation operator is designed to accelerate the convergency of population and acquire good community partitions for complex network. The proposed network compression stage can reduce the network scale, therefore suitable for large-scale networks.

III. THE PROPOSED COM-MOEA/D ALGORITHM

In this section, we give a detailed description of the compression based multi-objective evolutionary algorithm for community detection (Com-MOEA/D). At first, the network compression process is introduced, which condenses a network into a smaller scale one. After that, how to apply the general framework of MOEA/D to detect communities in compressed network is presented, in which a local information based mutation operator is proposed to improve the accuracy and accelerate the convergence of the algorithm.

A. NETWORK COMPRESSION

By exploiting network topologies, it is found that some nodes are tightly connected, thus forming into inseparable groups. The tightly connected nodes in such group are always groups kept in the same community, no matter what kind of community detection algorithm is applied. Our compression process is motivated by this phenomenon. Through examining the local information of nodes, the groups that contains strongly connected nodes are detected. Each group can be treated as a whole in community detection. Therefore, the proposed compression operation first calculates each node pairs'

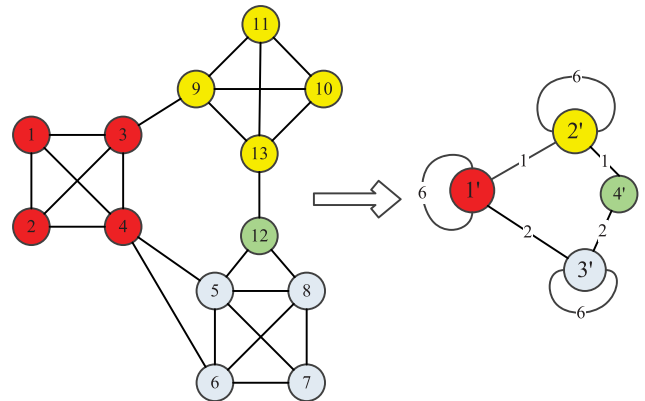


FIGURE 1. A toy example of a network with 13 nodes and its corresponding compression result with threshold $\alpha = 0.7$.

connectivity strength, and then condenses the strongly connected nodes into one node, thus generating a new network, namely, the compressed network G_c .

As to how to evaluate node pairs' connectivity strength, the node similarity indexes are preferred, which calculates to what extent the two nodes are close to each other. There exist an abundance of similarity indexes using local or global network topology information, such as Salton [43], Jaccard [44], Katz [45] etc.

AA index is one of them which is proposed by Adamic and Adar [46]. It considers the common neighbors' effect, and the effect is measured by node's degree. It believes that the contribution of small degree node is greater than that of the big one. The definition of AA index is as

$$Sim(v_1, v_2) = \begin{cases} \sum_{v \in CN_{v_1 v_2}} \frac{1}{\lg(k_v)}, & \text{if } A_{v_1, v_2} = 1; \\ 0, & \text{Otherwise,} \end{cases} \quad (4)$$

where k_v is the degree of node v , and $CN_{v_1 v_2}$ represents the common neighbors of node v_1 and v_2 . Liu *et al.* [15] have shown the effectiveness of applying AA index in community detection, here we also adopt this index. Note that only when two nodes are connected, do we calculate their similarities.

After acquiring similarities, each node chooses to be in the same community as its most similar neighbor. But if the similarity value is too small, it makes no sense for node selection. Therefore, a threshold α is given, which means that the first α percent of node pairs's similarities are considered. Those smaller value are not considered any more.

Algorithm 1 describes the compression process. At first, each node pairs' AA similarity values are calculated, and choose the first α percent of them after eliminating similarity values that equal to zero. After that, each node chooses to connect with its most similar neighbor. Note that, for nodes with degree equal to 1, they are directly linked to its only neighbor. At last, find out the connected components and compress each component into one node with self-loops representing the interior edges of the original components. Figure 1 shows a toy example of a network

with 13 nodes and its corresponding compression result with threshold $\alpha = 0.7$. After the compression process, node sets $\{1, 2, 3, 4\}$, $\{5, 6, 7, 8\}$, $\{9, 10, 11, 13\}$, $\{12\}$, changes to corresponding nodes $1'$, $2'$, $3'$ and $4'$, with loops representing interior number of edges.

Algorithm 1 Network_Compression(G, α)

Input: G : original complex network; α : the compression ratio for edges;
Output: G_c : compressed network;
1: **for** $i = 1$ to n **do**
2: $N_i \leftarrow$ find node v_i 's neighbors;
3: **for** $j = 1$ to N_i **do**
4: $CN_{ij} \leftarrow$ the common neighbors of nodes v_i and v_j ;
5: $Sim(i, j) \leftarrow \sum_{k \in CN_{ij}} \frac{1}{lgK_k}$;
6: **end for**
7: **end for**
8: $\alpha Sim \leftarrow$ Choose the first α percent of node pairs' similarities from Sim ;
9: **for** $i = 1$ to n **do**
10: **if** $K_i == 1$ **then**
11: $MSN_i \leftarrow N_i$;
12: **else**
13: $MSN_i \leftarrow max(Sim(i,))$;
14: **end if**
15: Update adjacency matrix by only retaining $A(i, MSN_i) = 1$;
16: **end for**
17: Find out all connected components (CC) of G ;
18: Compress each component in CC into one node to represent the connected components with self-loops representing interior edges;

B. REPRESENTATION

In this paper, the locus-based adjacency (LAR) encoding schema is employed to denote community detection solutions. It is first proposed by Yang *et al.* and now is widely used in community detection problems. In this representation, a network partition solution with N nodes is denoted by a chromosome that has N genes, i.e. a vector with length N . Each gene corresponds to a specific node of the graph, and its value indicates the index of one of its neighbors. That's to say, if the i^{th} gene takes the value j , then there is an edge between nodes v_i and v_j , and these two nodes are actually in the same community as a result of decoding schema. Figure 2(a) shows a network with 11 nodes and an individual (i.e. chromosome) generated for the network. For decoding a solution, a new graph is generated by connecting each node with the neighbor indicated by the chromosome. Figure 2(b) shows the graph of decoded solution. Each connected subgraph indicates a community that was derived from the individual. Using this kind of encoding method, the number of communities is automatically determined, and there is no need to define it

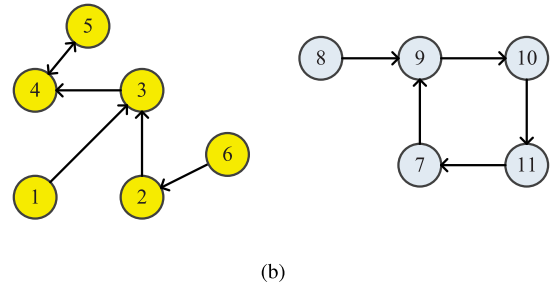
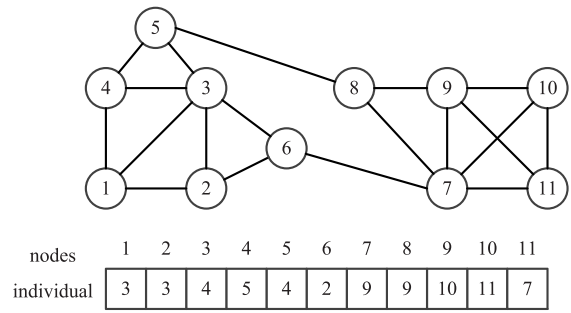


FIGURE 2. (a) Topology of a sample graph with 11 nodes and an example of the locus-based representation; (b) The decoded graph of the corresponding individuals.

in advance. This type of representation results in a reduced search space, and it is well suited for the genetic operators such as crossover and mutation to produce new individuals. One more important thing is that the LAR representation will not destroy the original network's structure.

C. POPULATION INITIALIZATION

In order to get a population as accurate as possible, a random walk based method is proposed. Suppose that an agent starts from an arbitrary node of a network, and then chooses one adjacent node as the next hop according to transition probability. If the agent is located in node v_i , the transition probability to neighboring node v_j is denoted as p_{ij} . Suppose the adjacency matrix of compressed network is $A_c = (w_{ij})_{n' \times n'}$, then

$$p_{ij} = \frac{w_{ij}}{\sum_k w_{ik}}, \tag{5}$$

where n' is the number of nodes in compressed network. If we use matrix to represent transition probability, let $D = diag(d_1, d_2, \dots, d_{n'})$ and $d_i = \sum_k w_{ik}$, then the transition matrix $P = (p_{ij})_{n' \times n'}$ is

$$P = D^{-1}A_c. \tag{6}$$

From the point of random walk, if a network has community structure, and due to the dense interior edges and sparse exterior ones, a random walker would be easy to visit nodes in the same community while difficult to reach to the other communities. That's to say, if an agent starts from an

a local information based mutation operator is proposed. That's to say, we define that the mutation is taken place on the special node, which we call it boundary independent node.

Boundary independent node. If a node v_i is called boundary independent node, it must satisfy the following two conditions:

- 1) node v_i has connections with other nodes in different communities;
- 2) for the LAR representation of individual ind , each gene doesn't equal to i , except for the i th gene.

In the network example in Figure 4, how to find the boundary independent nodes? At first, we need to find out nodes that satisfy the condition 1), and nodes 5, 6, 7 and 8 are the desired nodes. After that, check them whether they satisfy the second condition. Then we find out that nodes 6 and 8 are boundary independent nodes, since for nodes 5 and 7, the mutation for them would induce more than one nodes' movement. We can also find nodes that satisfy condition 2), then check whether they satisfy the first condition. The mutation process is shown in algorithm 2. All the boundary independent nodes perform mutation operations. For each one of them, choose to be in the same community as the node which would induce the greatest modularity increase Q_i . Applying this proposed operator generates effective solutions.

Algorithm 2 Local_Inf_Mutation(G_c, p_i)

Input: p_i : an individual; G_c : compressed network;

Output: p'_i : an individual after mutation;

```

1: for  $i = 1$  to  $n_c$  do
2:   if  $i$  is boundary independent node then
3:      $N_i\_labels \leftarrow$  get the community labels of node  $i$ 
       and its neighbors;
4:      $max \leftarrow -\infty$ ;
5:     for each  $r \in N_i\_labels$  do
6:        $Q_i \leftarrow$  when node  $i$ 's label changes to  $r$ ;
7:       if  $Q_i > max$  then
8:          $max \leftarrow Q_i$ ;
9:          $label_i \leftarrow r$ ;
10:      end if
11:    end for
12:  end if
13: end for
14:  $p'_i \leftarrow$  updated individual  $p_i$ ;

```

F. GENERAL FRAMEWORK OF COM-MOEA/D

The proposed algorithm adopts a similar framework of MOEA/D, which is presented in Algorithm 3. Let pop be the population size, and the weight vectors $\lambda = \{\lambda_1, \lambda_2, \dots, \lambda_{pop}\}$ are a set of weight vectors uniformly distributed on $\lambda_i^1 + \lambda_i^2 = 1$, where $\lambda_i = \langle \lambda_i^1, \lambda_i^2 \rangle, \lambda_i^1, \lambda_i^2 \in [0, 1]$ and $1 \leq i \leq pop$. The weight vectors λ_i are applied to decompose the multi-objective community detection problem into pop single-objective subproblems according to the following

Algorithm 3 General Framework of Com-MOEA/D

Input: G : original complex network; α : the compression ratio for edges; $maxgen$: maximum number of generations; λ : weight vectors $\{\lambda_1, \lambda_2, \dots, \lambda_{pop}\}$; ns : the size of neighborhood; p_c : crossover probability; P : the population $\{p_1, p_2, \dots, p_{pop}\}$ with size pop ;

Output: A set of detected communities;

```

1:  $G_c \leftarrow$  Network_Compression( $G, \alpha$ );
2:  $P \leftarrow$  Pop_Init( $G_c, pop$ );
3: Initialize reference point  $z^*$ ;
4: for  $i = 1$  to  $pop$  do
5:    $N_i \leftarrow$  find the  $ns$  individuals from  $P$  with the nearest
       Euclidean distance to the weight vector  $\lambda_i$ ;
6: end for
7: for  $i = 1$  to  $maxgen$  do
8:   for  $j = 1$  to  $pop$  do
9:     Randomly select one individual  $p_{rand}$  from  $N_j$ ;
10:    if  $rand\_number < p_c$  then
11:      Operate crossover operator on  $p_j$  and  $p_{rand}$ ,
       and generate two children  $chd_1$  and  $chd_2$  by crossover
       operator;
12:    end if
13:     $chd_3 \leftarrow$  Local_Inf_Mutation( $G_c, chd_1$ )
14:     $chd_4 \leftarrow$  Local_Inf_Mutation( $G_c, chd_2$ )
15:    Update the reference point  $z^*$ ;
16:    for  $k = 1$  to  $length(child)$  do
17:      for each index  $m \in NS(i)$  do
18:        if  $g^{te}(chd_k | \lambda, z^*) < g^{te}(chd_m | \lambda, z^*)$  then
19:           $p_m \leftarrow chd_k$ ;
20:        end if
21:      end for
22:    end for
23:    Update individuals in  $N_j$ ;
24:  end for
25: end for
26:  $P_{final} \leftarrow$  MaxModularity( $G_c, P$ );
27: Decode  $P_{final}$ , and return communities;

```

formula:

$$\text{minimize } g^{te}(x | \lambda_i, z^*) = \max_{j=1}^2 \{\lambda_i^j \cdot (|F_j(x) - z_j^*|)\} \quad (7)$$

where $z^* = (z_1^*, z_2^*)$ is the reference point in which each $z_j^*, 1 \leq j \leq 2$, is the minimal value on the j th objective in population.

The proposed Com-MOEA/D algorithm consists of the following three steps. At the very beginning, the original network is compressed by exploiting network topology in Algorithm 1. At the second step, a population with pop individuals is initialized according to the random walk based scheme and locus-based adjacency encoding method, which are effective in evolutionary algorithm based community detection problems. Meanwhile, the reference point z_* is initialized by using the best values of KKM and RC in the initial population. For each weight vector $\lambda_i, 1 \leq i \leq pop$,

the Euclidean distances for all individuals in population P to weight vector λ_i are calculated and ns individuals in P with the nearest Euclidean distances to λ_i are regarded as the neighbors of λ_i , denoted as N_i , where ns is a predefined parameter.

At the last step, for each individual $p_j (1 \leq j \leq pop)$ in population P , randomly selects one individual ind from N_j . Apply these two individuals to generate the children chd by the defined crossover and the local information based mutation operators. If the Tchebycheff value of chd is better than any individual ind' in N_j , then replaces ind' with chd and updates reference point z^* . The Com-MOEA/D keeps going until the maximum number of generations is reached.

The time complexity of the Com-MOEA/D algorithm is mainly composed of two stages. The first stage is to compress networks, and needs at most $O(d_{max} \times m)$ operations for calculating similarities and $O(\alpha m)$ operations for choosing the first α percent edges. The following is the multi-objective optimization stage, in which the population initialization requires $O(pop \times n')$ operations. After that, choosing ns nearest neighbors needs $O(pop^2)$ operations. At each iteration, we need to perform the crossover and mutation operator pop times at the worst case. Each crossover and mutation need $O(1)$ and $O(n')$ at worst. What's more, calculating the modularity cost $O(m)$ operations. According to the operational rules of the symbol O , the worst case's time complexity can be simplified as $O(maxgen \times pop \times m)$.

IV. EXPERIMENTAL RESULTS AND ANALYSIS

In this section, to show the effectiveness and efficiency of the proposed algorithm, a series of experiments are performed. We compare our Com-MOEA/D algorithm with three state-of-the-art evolutionary based algorithms and three classical but effective ones. We first introduce the experimental design, which includes the comparison algorithms, evaluation metrics and data sets. And then the experimental results on both real-world networks and synthetic networks are analysed in detail. Finally, the parameter sensitivity is discussed.

A. EXPERIMENTAL DESIGN

1) COMPARISON ALGORITHMS

In the experiments, the comparison algorithms are among the Com-MOEA/D and popular EA based multi-objective optimization algorithms, as well as classical ones. They are MOGA-Net [20], MODPSO [27], MOEA/D-Net [21], and three representative algorithms including Louvain [12], LPA [13] and LeadingEigen [48].

The compared four EA based community detection algorithms all belong to multi-objective optimization. The parameters of them are set according to recommendation in each reference, which are shown in table 1. MOGA-Net is an efficient community detection algorithm that optimizes the community score and community fitness and adopts the NSGA-II to solve the optimization problem. MOEA/D-Net is a decomposition based multi-objective evolutionary

TABLE 1. The parameter settings for MOEA based comparison algorithms.

Algorithm	pop	maxgen	pc	pm	ns	Reference
MOGA-Net	100	100	0.9	0.1	-	[20]
MODPSO	100	100	-	0.1	40	[27]
MOEA/D-Net	100	100	0.9	0.1	40	[21]
Com-MOEA/D	100	100	0.9	-	40	Ours

algorithm with optimization of two conflicting objectives, termed as Negative Ratio Association (NRA) and Ratio Cut (RC). MODPSO is a multi-objective algorithm based on PSO to optimize KKM and RC . As to the three other classical algorithms, Louvain optimizes the modularity function Q with high accuracy, LPA is a very fast algorithm that update labels according to neighbors' status, and it has near linear complexity. LeadingEigen algorithm adopts networks eigenvector to detect communities that are very efficient especially for real world networks.

2) EVALUATION MEASURES

To evaluate the performance of algorithms, two metrics are adopted to report comparison results. They are modularity (Q) [10] and Normalized Mutual Information (NMI) [49]. Modularity Q is a metric that provides a numerical value that represents the quality of community detection results. It is defined in Equation 8.

$$Q = \sum_{s=1}^{N_c} \left[\frac{l_s}{m} - \left(\frac{d_s}{2m} \right)^2 \right] \quad (8)$$

where N_c is the number of detected communities, l_s is the number of interior edges in community C_s , and d_s is the sum of nodes' degrees in C_s .

NMI is another metric that based on the Shannon entropy and shows the extent of similarity between two communities X and Y . Note that, in this paper, the partition X is referred to the detected communities and Y represents the ground-truth communities. NMI is computed as:

$$NMI(X, Y) = \frac{2 \times I(X, Y)}{H(X) + H(Y)} \quad (9)$$

where $I(X, Y)$ is the mutual information which is defined as:

$$I(X, Y) = \sum_{ij} P_{ij} \log(P_{ij}/(P_{i+}P_{+j})) \quad (10)$$

and $H(X)$ is the entropy function that is:

$$H(X) = - \sum_i P_{i+} \log(P_{i+}) \quad (11)$$

where $P_{i+} = \sum_j P_{ij}, P_{+j} = \sum_i P_{ij}$, and $P_{ij} = \frac{n_{ij}}{N}$ is the probability for a randomly drawn node belonging simultaneously to the i th and j th community, n_{ij} is intersection of two communities. Under this normalization, NMI takes its value in the range of $[0, 1]$.

For networks without ground-truth, the modularity value is used to evaluate the community detection results. However, for those having ground-truth communities, such as synthetic

LFR networks, both the NMI and Q are adopted as the evaluation metrics. No matter for NMI or Q , the larger value signifies the better detection results.

3) DATASETS

The experiments are not only taken on the real-world networks, but also on synthetic networks. For real-world networks, the three well-known, frequently-used real-world networks, including Karate, Dolphins, American college football clubs, are tested in our experiments. Besides, the networks such as books about US politics, Email, Network scientists cooperation network etc are also adopted for experimentation. The description of datasets are shown in Table 2.

TABLE 2. The description of real-world networks.

Networks	nodes	edges	Description
Karate	34	78	Zackary's Karate club [50]
Dolphins	62	159	Dolphin social network [51]
Football	115	613	US college football [52]
Polbooks	105	441	Books about US politics [53]
Co-authors	1033	2554	Co-author's subset in China [15]
Email	1133	5451	Email network [54]
Netscientist	1589	2742	Scientists cooperation network [48]
Facebook	4039	88234	Facebook dataset [55]
GR_QC	4158	13428	Arxiv GR-QC collaborators [56]
GC_Hep_TH	8638	24827	Arxiv HEP-TH collaborators [56]
GC_Hep_PH	11204	117649	Arxiv HEP-PH collaborators [56]

For synthetic datasets, we use the LFR (Lancichinetti-Fortunato-Radicchi) networks [57], which are the most frequently used benchmarks in testing community detection algorithms. That's because LFR networks can simulate some characteristics of real-world networks. For example, the distribution of nodes' degrees and community sizes are power laws with tunable exponents. The detailed parameters of LFR benchmark are displayed in table 3. For each tested network, 20 independent runs are performed for each algorithm, and the mean value is attained. As to the comparison, the solutions with the largest Q or NMI are reported in the results.

TABLE 3. Parameters for the LFR benchmark.

Parameter	Description	Values
N	number of nodes	1000, 5000 and 10000
\bar{k}	average degree	15
$maxk$	maximum degree	$0.1N$
μ	mixing parameter	$\{0.1, 0.2, \dots, 0.6\}$
τ_1	minus exponent for the degree sequence	2
τ_2	minus exponent for the community size distribution	1, 2
$minc$	minimum for the community sizes	10, for $N = 1000$ 20, for $N = 5000$ 25, for $N = 10000$
$maxc$	maximum for the community sizes	$0.1N$

B. EXPERIMENTS ON REAL-WORLD NETWORKS

In this section, the proposed Com-MOEA/D algorithm and six other compared algorithms are tested on eleven real-world

networks with scales from tens to ten thousands. Those networks can be divided into two categories, according to whether the network has ground truth or not. The first five networks in table 2 belong to the category that have ground truth partitions, so both NMI and Q are computed. For others which do not have the ground truth, only Q is considered. The compression ratio is set to 0.4 for all tested networks, which can almost compressed the original network scale into a half. Table 4 and 5 display the community detection results. Note that, here we only show the best modularity value Q and corresponding detected community numbers N_c for each network and each algorithm.

From table 4 and 5, we see that the proposed Com-MOEA/D outperforms other algorithms for six of eleven networks, not only for modularity value Q , but also for NMI . These results demonstrate the effectiveness of the proposed algorithm on the real-world networks. The competitive results are mainly due to the mechanism of network compression at the very beginning, which combine the tightly connected nodes into one inseparable group. Figure 5 shows the comparison of NMI values after network compression and the final ones on the real-world networks with known ground truth when compression ratio equals to 0.4. As can be seen from figure 5, all the NMI values achieved by the compressed network are more than a half of the final results. That's to say, the compression process not only reduces the network scale but also help us get more accurate preliminary partitions.

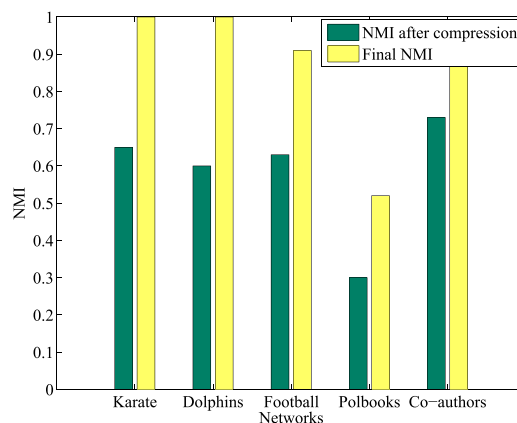


FIGURE 5. The comparison of NMI values after network compression and the final ones.

From the experiments, we know that Zachary's karate club network is a most often used empirical network in community detection, and it was first compiled by Zachary when observing the relationships among karate members in a US university. The network consists of 34 nodes and 78 edges, and is divided into two communities. Although our algorithm only gets a moderate modularity value, but as to the NMI value, our method together with four EA based algorithms and LPA can all get the communities the same as the ground truth. Since the EA based algorithms can get a series of solutions, we can choose one according to different criterions.

TABLE 4. Comparisons of modularity values of Com-MOEA/D algorithm on real-world networks with six other methods.

Networks	Algorithms														
	Com-MOEA/D			MOGA-Net		MODPSO		MOEA/D-Net		Louvain		LPA		LeadingEigen	
	N_c	Q	RS	N_c	Q	N_c	Q	N_c	Q	N_c	Q	N_c	Q	N_c	Q
Karate	3	0.38	0.44	4	0.42	3	0.37	4	0.42	4	0.42	2	0.37	4	0.39
Dolphins	4	0.52	0.55	4	0.49	5	0.51	4	0.5	5	0.52	4	0.52	5	0.49
Football	10	0.6	0.59	11	0.51	10	0.6	11	0.6	10	0.6	11	0.6	8	0.49
Polbooks	4	0.53	0.54	4	0.51	5	0.50	4	0.52	4	0.52	4	0.52	4	0.47
Co-authors	11	0.83	0.57	18	0.80	15	0.80	13	0.82	12	0.81	41	0.79	14	0.82
Email	50	0.52	0.53	8	0.47	15	0.48	10	0.50	11	0.54	4	0.09	7	0.49
Netscientist	428	0.93	0.45	419	0.9	410	0.95	397	0.88	406	0.96	450	0.91	404	0.95
Facebook	11	0.76	0.42	19	0.73	41	0.79	15	0.82	17	0.83	60	0.82	42	0.75
GR_QC	68	0.79	0.47	51	0.68	42	0.76	60	0.82	43	0.85	363	0.78	18	0.8
GC_Hep_TH	58	0.74	0.59	65	0.72	41	0.68	15	0.61	56	0.74	510	0.6	5	0.1
GC_Hep_PH	50	0.65	0.56	49	0.62	43	0.66	15	0.60	47	0.65	366	0.43	8	0.54

N_c represents the number of communities.

RS represents the ratio of reduced number of nodes to the original network scale when compression ratio $\alpha = 0.4$, i.e. $RS = (N - n')/N$

TABLE 5. Comparisons of NMI values of Com-MOEA/D algorithm on real-world networks with six other methods.

Networks	Algorithms													
	Com-MOEA/D		MOGA-Net		MODPSO		MOEA/D-Net		Louvain		LPA		LeadingEigen	
	N_c	NMI	N_c	NMI	N_c	NMI	N_c	NMI	N_c	NMI	N_c	NMI	N_c	NMI
Karate	2	1	2	1	2	1	2	1	4	0.59	2	1	4	0.68
Dolphins	2	1	2	1	2	1	2	1	4	0.55	5	0.57	5	0.53
Football	10	0.91	11	0.85	11	0.92	11	0.94	10	0.89	11	0.91	8	0.70
Polbooks	4	0.52	4	0.59	5	0.50	4	0.51	4	0.51	4	0.51	4	0.52
Co-authors	11	0.96	18	0.90	14	0.91	13	0.91	12	0.91	41	0.85	14	0.93

N_c represents the number of communities.

For Louvain and LeadingEigen algorithms, although they get the higher modularity value, the NMI values are not desirable.

The dolphin network is formed by the frequency of the bottlenose dolphins playing together. The Com-MOEA/D can divide it into 4 communities when choosing the best modularity value. However, when choosing the biggest NMI value, our algorithm can also get the same partitions as the ground truth. The American college football network displays 115 football teams' matches during the fall season in 2000, and it actually has 12 groups. Our Com-MOEA/D can get the modularity value that are almost the same as other algorithms, but as to NMI , our algorithm outperforms others. The political books data was compiled by Krebs, and it can be divided into 3 groups according to the books' attitudes. Our method together with others can get relative good results, but when it comes to NMI , all the algorithms do not perform so good. The email network describes the email interchanges among members of a university. Louvain algorithm gets the best result, but LPA can hardly detect the community structures. The netscientist network's community detection results have little difference.

The Co-authors network is a subset of the cooperators for papers in management science and engineering in China from the year 2000 to 2016. It is an ego-network, and actually has 11 groups. From table 4 and 5, we can see that our method outperforms others. The GR_QC, GC_Hep_TH and GC_Hep_PH are authors' collaboration network compiled by Leskovec etc. The community detection results show that our algorithm performs as good as Louvain. That's maybe due to the fact that coauthor networks has some mode of coauthorship, and such relation can be detected easily at the

compression stage. Other algorithms get much more numbers of communities.

C. EXPERIMENTS ON SYNTHETIC NETWORKS

The synthetic networks are the LFR benchmarks. Those networks have power-law distributions not only for node degree but also for the community size, which are similar to real-world networks. The parameters setting are shown in table 3. For each set of parameters, 20 networks are generated and the average community detection results which are NMI values are shown in figure 6. The uniform trends of the four panels show the the NMI values decrease as the mixing parameter μ increases. This is due to the fact that greater μ means the more ambiguous community structure, therefore, it's more difficult to detect accurate communities as μ increases.

Panel 6(a) and 6(b) show the results for LFR networks with the same communities distribution that $\tau_2 = 1$ but different network sizes respectively. We can see that our proposed Com-MOEA/D and Louvain are the first two algorithms in terms of NMI values, and Com-MOEA/D becomes better than Louvain when the network size exceeds 1000 except for $\mu = 0.6$. MODPSO also performs relatively good, better than the LPA, MOGA-net etc. As to LPA and MOEA/D-net, the NMI decreases when $\mu > 0.5$, which means as the community structure becomes more and more ambiguous, these two algorithms' performances rapidly decreases. As to MOGA-net and LeadEigen, they do not perform well on synthetic networks.

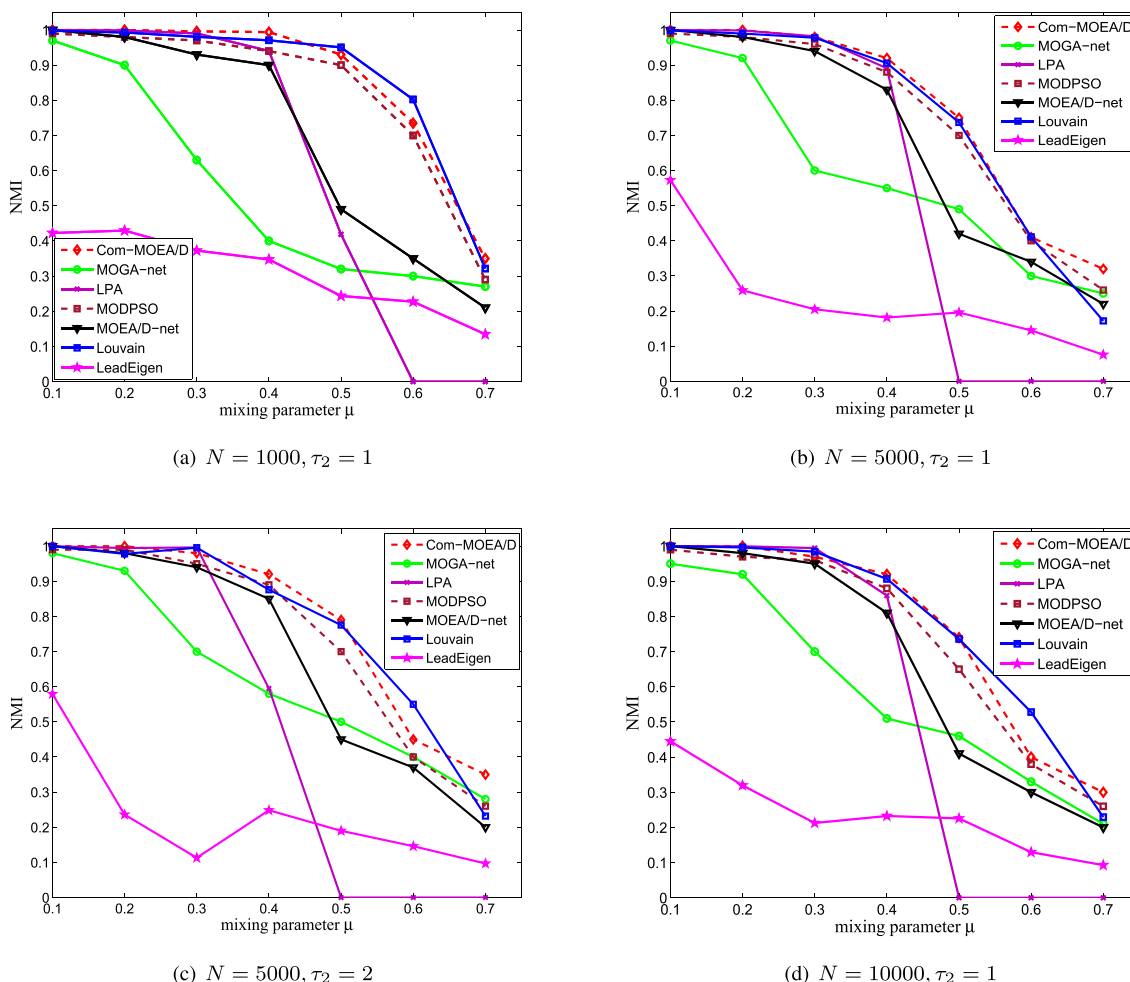


FIGURE 6. Comparisons of the NMI values among our proposed algorithm and six other algorithms based on LFR network.

Panel 6(b) and show the results of increasing community size distribution parameter τ_2 from 1 to 2. In panel 6(c), almost all the $NMIs$ are just a little difference than those in panel 6(b). This means the dissimilarities in community sizes do not affect the accuracy of the algorithm and the blurred modularity structure.

With the size increasing of the network, the results for network scale with $N = 10000$ are shown in panel 6(d). We can see that Louvain and Com-MOEA/D still performs well. LPA is still not good for finding communities when structures are blur, and LeadingEigen and MOGA-net are not so good for synthetic networks.

D. THE COMPRESSION RATIO'S EFFECT

In the compression process, the first α percent of non-zero similarities are chosen. The reduced scale (RS) increases as the compression ratio rises, which is displayed in Figure 7. The bigger value of the reduced scale signifies the smaller scale of the compressed network. But the smaller network doesn't mean the better final result. Figure 8 shows the NMI values with different compression ratios, and we see that

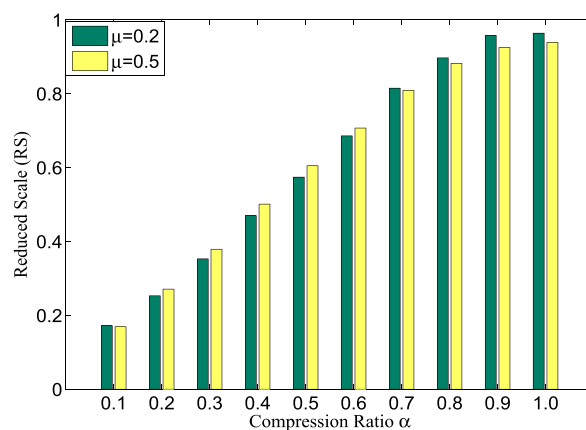


FIGURE 7. The comparison of reduced scales (RS) as compression ratio increases.

when compression ratio equals to 0.4 or 0.5, can we get the best final results. When $\alpha = 1$, the final community detection results are always not the best. This maybe due to the fact that all similarities are considered in this circumstance, even when

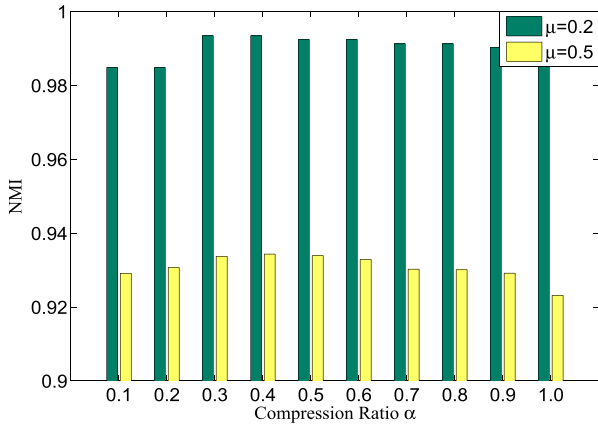


FIGURE 8. The comparison of NMI values after network compression and the final ones when $N = 1000$.

the similarities values are very small, which could induce false selections. Therefore, we choose to set the compression ratio to be 0.4 in all the experiments.

E. SIMILARITY INDEXES SELECTION

The similarity index plays an important role in the compressing stage. How to select an effective and efficient index is crucial to the proposed method. By far, an abundance of similarity indexes using local information are proposed and table 6 displayed 6 well-known indexes. In our experiment, we limit all the indexes to have values bigger than 0 only when two nodes are connected. To determine which index is most suitable for our algorithm, we take the AA, CN (Common neighbors) indexes as well as thoses shown in table 6 into our algorithm on LFR network with scale $N = 1000$. The results are shown in Figure 9.

TABLE 6. Some similarity indexes.

Salton	$\frac{CN_{v_1 v_2}}{\sqrt{k_{v_1} k_{v_2}}}$	Jaccard	$\frac{CN_{v_1 v_2}}{\tau_{v_1} \cap \tau_{v_2}}$
HPI	$\frac{CN_{v_1 v_2}}{\min\{k_{v_1}, k_{v_2}\}}$	HDI	$\frac{CN_{v_1 v_2}}{\min\{k_{v_1}, k_{v_2}\}}$
LHN	$\frac{CN_{v_1 v_2}}{k_{v_1} k_{v_2}}$	AA	$\sum_{v \in CN_{v_1 v_2}} \frac{1}{lg(k_v)}$

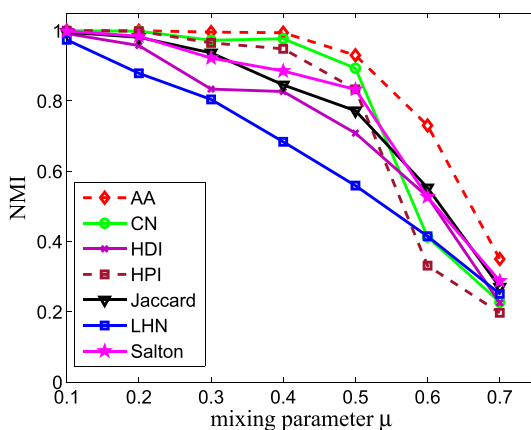


FIGURE 9. The comparison of different similarity indexes.

It is clear to see that AA index outperforms others for any mixing parameter μ . Then AA index is our first choice. What’s more, CN index also displays relative good performance, but when the community structure becomes more ambiguous, i.e. $\mu > 0.5$, its NMI value decreases sharply to 0.4. As to others, they do not exceed AA in our Com-MOEA/D algorithm. Therefore, we choose AA index to evaluate nodes’ similarities in the compression stage.

V. CONCLUSIONS AND FUTURE WORK

In this paper, we have proposed a compression based Multi-Objective Evolutionary Algorithm with Decomposition (Com-MOEA/D) for community detection in social networks. In Com-MOEA/D, the network is first compressed to a much more smaller scale by exploring network topologies. After that, a framework of multi-objective evolutionary algorithm based on decomposition is applied, in which a local information based generic operator is proposed to speed up the convergence and improve the accuracy of the Com-MOEA/D algorithm. Experimental results on both real-world and synthetic networks have shown the superiority of the proposed Com-MOEA/D algorithm.

There still exist some work that deserve to be further investigated. At first, the proposed Com-MOEA/D algorithm is for undirected and unweighted graph, how to apply this framework on signed and dynamic network still needs to be done. What’s more, the proposed algorithm only considers two objectives optimization, what about many-objective optimization for community detection problem? Finally, we would like to consider how to apply the proposed community detection algorithm into practice, such as recommendation systems etc.

REFERENCES

- [1] F. Rezaeimehr, P. Moradi, S. Ahmadian, N. N. Qader, and M. Jalili, “TCARS: Time- and community-aware recommendation system,” *Future Gener. Comput. Syst.*, vol. 78, pp. 419–429, Jan. 2018.
- [2] Z. Wang, Y. Wu, Q. Li, F. Jin, and W. Xiong, “Link prediction based on hyperbolic mapping with community structure for complex networks,” *Phys. A, Stat. Mech. Appl.*, vol. 450, pp. 609–623, May 2016.
- [3] J. Shang, L. Liu, X. Li, F. Xie, and C. Wu, “Epidemic spreading on complex networks with overlapping and non-overlapping community structure,” *Phys. A, Stat. Mech. Appl.*, vol. 419, pp. 171–182, Feb. 2015.
- [4] M. C. Benigni, K. Joseph, and K. M. Carley, “Online extremism and the communities that sustain it: Detecting the ISIS supporting community on Twitter,” *PLoS ONE*, vol. 12, no. 12, 2017, Art. no. e0181405.
- [5] C. Pizzuti, “Evolutionary computation for community detection in networks: A review,” *IEEE Trans. Evol. Comput.*, vol. 22, no. 3, pp. 464–483, Jun. 2018.
- [6] W. Zhang, R. Zhang, R. Shang, J. Li, and L. Jiao, “Application of natural computation inspired method in community detection,” *Phys. A, Stat. Mech. Appl.*, vol. 515, pp. 130–150, Feb. 2019.
- [7] F. D. Malliaros and M. Vazirgiannis, “Clustering and community detection in directed networks: A survey,” *Phys. Rep.*, vol. 533, no. 4, pp. 95–142, Dec. 2013.
- [8] E. Osaba, J. Del Ser, D. Camacho, M. N. Bilbao, and X.-S. Yang, “Community detection in networks using bio-inspired optimization: Latest developments, new results and perspectives with a selection of recent meta-heuristics,” *Appl. Soft Comput.*, vol. 87, Feb. 2020, Art. no. 106010.
- [9] U. Brandes, D. Delling, M. Gaertler, R. Goerke, M. Hofer, Z. Nikoloski, and D. Wagner, “Maximizing modularity is hard,” Aug. 2006, *arXiv:physics/0608255*. Accessed: Dec. 10, 2019. [Online]. Available: <http://arxiv.org/abs/physics/0608255>

- [10] 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.
- [11] 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.
- [12] 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.
- [13] 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.
- [14] M. Rosvall and C. T. Bergstrom, "Maps of information flow reveal community structure in complex networks," *Proc. Nat. Acad. Sci. USA*, pp. 1118–1123, 2007.
- [15] Z. Liu and Y. Ma, "A divide and agglomerate algorithm for community detection in social networks," *Inf. Sci.*, vol. 482, pp. 321–333, May 2019.
- [16] Y. Luo, L. Wang, S. Sun, and C. Xia, "Community detection based on local information and dynamic expansion," *IEEE Access*, vol. 7, pp. 142773–142786, Oct. 2019.
- [17] C. Shi, Z. Yan, Y. Wang, Y. Cai, and B. Wu, "A genetic algorithm for detecting communities in large-scale complex networks," *Adv. Complex Syst.*, vol. 13, no. 01, pp. 3–17, Feb. 2010.
- [18] Y.-N. Tang, Y.-J. Chen, J. Xiang, Y.-Y. Gao, Z.-Z. Wang, H.-J. Li, S. Chen, Y. Zhang, J.-M. Li, and Y.-H. Tang, "An effective algorithm for optimizing surprise in network community detection," *IEEE Access*, vol. 7, pp. 148814–148827, Oct. 2019.
- [19] M. Gong, B. Fu, L. Jiao, and H. Du, "Memetic algorithm for community detection in networks," *Phys. Rev. E, Stat. Phys. Plasmas Fluids Relat. Interdiscip. Top.*, vol. 84, no. 5, Nov. 2011, Art. no. 056101.
- [20] C. Pizzuti, "A multi-objective genetic algorithm for community detection in networks," in *Proc. 21st IEEE Int. Conf. Tools Artif. Intell.*, Nov. 2009, pp. 379–386.
- [21] M. Gong, L. Ma, Q. Zhang, and L. Jiao, "Community detection in networks by using multiobjective evolutionary algorithm with decomposition," *Phys. A, Stat. Mech. Appl.*, vol. 391, no. 15, pp. 4050–4060, Aug. 2012.
- [22] F. Cheng, T. Cui, Y. Su, Y. Niu, and X. Zhang, "A local information based multi-objective evolutionary algorithm for community detection in complex networks," *Appl. Soft Comput.*, vol. 69, pp. 357–367, Aug. 2018.
- [23] F. Zou, D. Chen, S. Li, R. Lu, and M. Lin, "Community detection in complex networks: Multi-objective discrete backtracking search optimization algorithm with decomposition," *Appl. Soft Comput.*, vol. 53, pp. 285–295, Apr. 2017.
- [24] X. Zhang, K. Zhou, H. Pan, L. Zhang, X. Zeng, and Y. Jin, "A network reduction-based multiobjective evolutionary algorithm for community detection in large-scale complex networks," *IEEE Trans. Cybern.*, vol. 50, no. 2, pp. 703–716, Feb. 2020.
- [25] S. Fortunato and M. Barthelemy, "Resolution limit in community detection," *Proc. Nat. Acad. Sci. USA*, vol. 104, no. 1, pp. 36–41, Jan. 2007.
- [26] L. Angelini, S. Boccaletti, D. Marinazzo, M. Pellicoro, and S. Stramaglia, "Identification of network modules by optimization of ratio association," *Chaos, Interdiscipl. J. Nonlinear Sci.*, vol. 17, no. 2, Jun. 2007, Art. no. 023114.
- [27] M. Gong, Q. Cai, X. Chen, and L. Ma, "Complex network clustering by multiobjective discrete particle swarm optimization based on decomposition," *IEEE Trans. Evol. Comput.*, vol. 18, no. 1, pp. 82–97, Feb. 2014.
- [28] X. Zhang, Y. Tian, R. Cheng, and Y. Jin, "A decision variable clustering-based evolutionary algorithm for large-scale many-objective optimization," *IEEE Trans. Evol. Comput.*, vol. 22, no. 1, pp. 97–112, Feb. 2018.
- [29] Q. Zhang and H. Li, "MOEA/D: A multiobjective evolutionary algorithm based on decomposition," *IEEE Trans. Evol. Comput.*, vol. 11, no. 6, pp. 712–731, Dec. 2007.
- [30] C. Pizzuti, "Ga-net: A genetic algorithm for community detection in social networks," in *Proc. 8th Int. Conf. Parallel Problem Solving from Nature (PPSN)*, 2008, pp. 1081–1090.
- [31] S. Li, Y. Chen, H. Du, and M. W. Feldman, "A genetic algorithm with local search strategy for improved detection of community structure," *Complexity*, vol. 15, no. 4, pp. 53–60, Mar. 2010.
- [32] P. Wang and J. Liu, "A multi-agent genetic algorithm for local community detection by extending the tightest nodes," in *Proc. IEEE Congr. Evol. Comput. (CEC)*, Jul. 2016, pp. 3215–3221.
- [33] M. Gong, Q. Cai, Y. Li, and J. Ma, "An improved memetic algorithm for community detection in complex networks," in *Proc. IEEE Congr. Evol. Comput.*, Jun. 2012, pp. 1–8.
- [34] C. Liu, J. Liu, and Z. Jiang, "A multiobjective evolutionary algorithm based on similarity for community detection from signed social networks," *IEEE Trans. Cybern.*, vol. 44, no. 12, pp. 2274–2287, Dec. 2014.
- [35] Z. Li and J. Liu, "A multi-agent genetic algorithm for community detection in complex networks," *Phys. A, Stat. Mech. Appl.*, vol. 449, pp. 336–347, May 2016.
- [36] K. Deb, A. Pratap, S. Agarwal, and T. Meyarivan, "A fast and elitist multiobjective genetic algorithm: NSGA-II," *IEEE Trans. Evol. Comput.*, vol. 6, no. 2, pp. 182–197, Apr. 2002.
- [37] C. Pizzuti, "A multiobjective genetic algorithm to find communities in complex networks," *IEEE Trans. Evol. Comput.*, vol. 16, no. 3, pp. 418–430, Jun. 2012.
- [38] C. Shi, Z. Yan, Y. Cai, and B. Wu, "Multi-objective community detection in complex networks," *Appl. Soft Comput.*, vol. 12, no. 2, pp. 850–859, Feb. 2012.
- [39] D. Chen, F. Zou, R. Lu, L. Yu, Z. Li, and J. Wang, "Multi-objective optimization of community detection using discrete teaching-learning-based optimization with decomposition," *Inf. Sci.*, vol. 369, pp. 402–418, Nov. 2016.
- [40] B. Chen, L. Chen, and Y. Chen, "Detecting community structure in networks based on ant colony optimization," in *Proc. Int. Conf. Inf. Knowl. Eng. (WORLDCOMP)*, 2012, p. 1.
- [41] B. Amiri, L. Hossain, and J. Crawford, "A hybrid evolutionary algorithm based on HSA and CLS for multi-objective community detection in complex networks," in *Proc. IEEE/ACM Int. Conf. Adv. Social Netw. Anal. Mining*, Aug. 2012, pp. 243–247.
- [42] B. Amiri, L. Hossain, J. W. Crawford, and R. T. Wigand, "Community detection in complex networks: Multi-objective enhanced firefly algorithm," *Knowl-Based. Syst.*, vol. 46, pp. 1–11, Jul. 2013.
- [43] G. Salton and M. McGill, *Introduction to Modern Information Retrieval*. New York, NY, USA: McGraw-Hill, 1986.
- [44] J. P. Etude, "Etude comparative de la distribution florale dans une prairie des alpes et des jura," *Bull. de la Société Vaudoise des Science Naturelles*, vol. 37, pp. 547–579, Jan. 1901.
- [45] L. Katz, "A new status index derived from sociometric analysis," *Psychometrika*, vol. 18, no. 1, pp. 39–43, Mar. 1953.
- [46] L. A. Adamic and E. Adar, "Friends and neighbors on the Web," *Social Netw.*, vol. 25, no. 3, pp. 211–230, Jul. 2003.
- [47] R. Guimerà and L. A. N. Amaral, "Functional cartography of complex metabolic networks," *Nature*, vol. 433, no. 7028, pp. 895–900, Feb. 2005.
- [48] M. E. J. Newman, "Finding community structure in networks using the eigenvectors of matrices," *Phys. Rev. E, Stat. Phys. Plasmas Fluids Relat. Interdiscip. Top.*, vol. 74, no. 3, Sep. 2006, Art. no. 036104.
- [49] L. Danon, A. Díaz-Guilera, J. Duch, and A. Arenas, "Comparing community structure identification," *J. Stat. Mech., Theory Exp.*, vol. 2005, no. 9, 2005, Art. no. P09008.
- [50] W. W. Zachary, "An information flow model for conflict and fission in small groups," *J. Anthropol. Res.*, vol. 33, no. 4, pp. 452–473, Dec. 1977.
- [51] D. Lusseau, K. Schneider, O. J. Boisseau, P. Haase, E. Slooten, and S. M. Dawson, "The bottlenose dolphin community of doubtful sound features a large proportion of long-lasting associations," *Behav. Ecol. Sociobiol.*, vol. 54, no. 4, pp. 396–405, Sep. 2003.
- [52] 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.
- [53] V. Krebs. *Uspolbooks*. Accessed: Jan. 5, 2020. [Online]. Available: <http://www.orgnet.com>
- [54] M. Boguñá, R. Pastor-Satorras, A. Díaz-Guilera, and A. Arenas, "Models of social networks based on social distance attachment," *Phys. Rev. E, Stat. Phys. Plasmas Fluids Relat. Interdiscip. Top.*, vol. 70, no. 5, Nov. 2004, Art. no. 056122.
- [55] J. McAuley and J. Leskovec, "Learning to discover social circles in ego networks," in *Proc. 25th Int. Conf. Neural Inf. Process. Syst. (NIPS)*, May 2012, pp. 539–547.
- [56] J. Leskovec, J. Kleinberg, and C. Faloutsos, "Graph evolution: Densification and shrinking diameters," *ACM Trans. Knowl. Discovery Data*, vol. 1, no. 1, p. 2, Mar. 2007.
- [57] 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.



ZHIYUAN LIU was born in Shandong, China. She received the B.S. degree in computer science and technology from the University of Jinan, in 2008, the M.S. degree in computer architecture from the Shandong University, China, in 2011, and the M.S. degree in security and trust from the University of Luxembourg, in 2011.

Her research interests include complex network theory and management decision methods.



XIUJUAN WANG was born in Shandong, China. She received the B.S. degree in management from Shandong Normal University, in 2019. She is currently pursuing the master's degree with the Dalian University of Technology. Her research interest includes complex network analysis and optimization algorithms.

• • •



YINGHONG MA was born in Shandong, China. She received the B.S. degree in mathematics from Shandong Normal University, in 1996, and the M.S. and Ph.D. degrees in mathematics from Shandong University, China, in 1999 and 2002, respectively.

In recent years, she has been engaged in the research of management decision theory and methods, complex network theory, and application in scientific research.