

Research Article

Multiobjective Brain Storm Optimization Community Detection Method Based on Novelty Search

Xiaoying Pan,^{1,2} Jia Wang^{id},¹ Miao Wei,¹ and Hongye Li¹

¹*Xi'an University of Posts and Telecommunications, Xi'an 710121, China*

²*Shaanxi Key Laboratory of Network Data Analysis and Intelligent Processing, Xi'an University of Posts and Telecommunications, Xi'an, Shaanxi 710121, China*

Correspondence should be addressed to Jia Wang; 17691032356@163.com

Received 1 February 2021; Revised 16 March 2021; Accepted 7 April 2021; Published 22 April 2021

Academic Editor: Qingzheng XU

Copyright © 2021 Xiaoying Pan et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

A complex network is characterized by community structure, so it is of great theoretical and practical significance to discover hidden functions by detecting the community structure in complex networks. In this paper, a multiobjective brain storm optimization based on novelty search (MOBSSO-NS) community detection method is proposed to solve the current issue of premature convergence caused by the loss of diversity in complex network community detection based on multiobjective optimization algorithm and improve the accuracy of community discovery. The proposed method designs a novel search strategy where novelty individuals are first constructed to improve the global search ability, thus avoiding falling into local optimal solutions; then, the objective space is divided into 3 clusters: elite cluster, ordinary cluster, and novel cluster, which are mapped to the decision space, and finally, the populations are disrupted and merged. In addition, the introduction of a restarting strategy is introduced to avoid stagnation by premature convergence. Experimental results show that the algorithm with good global searchability can find the Pareto optimal network community structure set with uniform distribution and high convergence and excavate the network community with higher quality.

1. Introduction

Complex networks can be seen everywhere in real life, such as social networks [1], communication information networks [2], biological networks [3], and computer networks [4], which usually have the characteristics of small world [5], scale-free [6], and community structure [7]. The nodes within the community are closely connected, while the nodes between the communities are sparsely connected. The study of complex networks is to abstract networks into mathematical models formed composed of points and edges between points. Detecting community structure in complex network plays an important role in understanding and analyzing the topology structure of the entire network and discovering hidden functions in the network. Related research results have been successfully applied to many fields such as terrorist organization identification [8], protein function prediction, and public opinion analysis and treatment [9].

In recent years, more and more attention has been paid to the problem of complex network community detection. Many scholars have carried out a lot of research work on community detection, regarding community discovery as a single-objective optimization problem and adopting different heuristic algorithms or approximate algorithms to optimize community-related objective functions, thereby obtaining the community structure of the network. To identify the division of sparse connections and dense connections among communities, Tasgin et al. [10] optimized the community modularity criterion Q with genetic algorithm (GA) that has good performances in large networks, such as fast speed and no requirement on the number of communities.

Pizzutiz [11] defined community scores, which are used to judge the quality of network division, and utilized Genetic Algorithm in Networks (GA-Net) to optimize a simple and effective fitness function. The function can identify the connections between groups of nodes and its sparsity and

improve the variational operator, making it consider the actual correlation between nodes, thus significantly reducing the space for studying possible solutions. Lipczak and Milios [12] proposed agglomerative clustering genetic algorithm (ACGA) based on the hypothesis that communities are small enough and the number of communities is limited. Agglomerative clustering genetic algorithm based on the ordered table of neighbors of nodes was used as a particle coding method to explore the community through global search for the cluster.

Although the above single-objective optimization algorithm has the advantage of time efficiency and can mine the network satisfying a certain objective, the network community detection in practical application often needs to take into account multiple objectives, and there may be conflicts among these objectives. In view of the above limitations of the community detection algorithm based on single-objective optimization, the community detection based on multiobjective optimization begins to be concerned.

In solving complex network community detection problems, a large number of multiobjective evolutionary algorithms have emerged. Chen et al. [13] proposed a multiobjective discrete MODTLBO/D method based on teaching-learning-based Optimization (TLBO). The algorithm adopted the multiobjective decomposition mechanism and introduced neighbor-based mutation to solve the community detection problem of complex network, thus maintaining the diversity of population and avoiding local optimization.

Gong et al. [14] proposed multiobjective evolutionary algorithm with decomposition (MOEA/D-Net), making the multiobjective optimization problem transformed into a series of single-objective optimization subproblems and then, with the help of the information of a certain number of adjacent problems, employed evolutionary algorithm to optimize these subproblems at the same time. Li et al. [15] proposed a quantum-behavior discrete multiobjective particle swarm optimization algorithm for complex network clustering, which performed well in large networks. Gong et al. [16] proposed multiobjective discrete particle swarm optimization (MODPSO), which decomposed multiobjective network clustering problems into multiple scalar problems with the decomposition mechanism, and generated different individuals with high clustering efficiency in virtue of neighbor-based turbulence operator to promote diversity. Jiang et al. [17] proposed a community detection method based on a new link prediction strategy. The operation steps of the method were as follows. Firstly, the designed link prediction strategy based on the central node was used to add and remove edges in order to enhance the community structure of the network. Secondly, the community extension strategy was adopted to detect all communities in the network. However, the proposed link prediction strategy needed to calculate the similarity of a large number of node pairs, so the algorithm was very time-consuming. Aiming at the network Gong et al. [18] with fuzzy community structure, a nondominated neighbor immune algorithm-Net (NNIA-NET) was proposed to

optimize the density of both internal link and external link and discovered the community in the network by NNIA. Zhang et al. [19] considered critical node detection based on the cascade model as a biobjective optimization problem (BCVND) proposed an effective multiobjective evolutionary approach termed as MO-BCVND to solve BCVND. In MO-BCVND, a cost-reduced population initialization strategy was raised to increase the population diversity and an adaptive local search strategy was designed to accelerate the population convergence.

Zhou et al. [20] came up with multiobjective local search (MOLS-Net) algorithm, a multiobjective optimization algorithm based on local search, and designed different local search algorithms for different objectives. A hybrid genetic algorithm with link strength-based local search strategy (HGALS) is proposed by Malhotra [21] for solving the community detection problem. The local search method presented in the algorithm is faster than the traditional modularity-based search operations.

The above multiobjective community detection optimization algorithms are all intelligent optimization algorithms which need to solve the balance between global search and local search to achieve the optimal structure division in the community detection problem. Therefore, the difficulty of multiobjective optimization lies in how to maintain population diversity and avoid falling into local optimization. The brainstorm algorithm searches for the local optimal with the clustering idea, obtains the global optimal through comparing the local optimal results, and increases the diversity of the algorithm by adding the mutation operation.

To solve the problem of high complexity and slow speed of Gaussian, we put forward multiobjective brainstorming community detection methods based on novelty search, in which, by constructing novelty individuals, the objective space is divided into elite clusters, ordinary clusters, and novelty clusters and then mapped to the decision space. Novelty clustering took the individual with the smallest *NMI* value as the novelty solution to generate offspring through the cross fusion of the elite solution and the novelty solution. The novelty search mechanism has high search efficiency and excellent global optimization capabilities and adopts the restart strategy to help individuals escape from the local optimum and avoid premature convergence. "Premature" and stagnation form a single optimal solution, which improves the global search ability of the population. An external archive was established to save the Pareto optimal solution. Nondominated sorting was carried out on the population in each iteration, and the nondominated solution stored in the external archive was updated to finally get a set of Pareto optimal solutions. The brainstorming optimization algorithm has not been used to solve the problem of community detection. Therefore, the application of novelty search-based multiobjective brainstorming optimization algorithm to complex network community detection problems can better balance diversity and convergence and search for better community structure division, which is also the research motivation of this paper.

1.1. Section Arrangement. The rest of this paper was arranged as follows. In Section 2, the community detection problem, related work for multiobjective community detection, and brainstorming algorithms are introduced. Section 3 designs a multitarget brainstorming community detection method based on novelty search and gives a detailed algorithm flow and introduction. Section 4 shows experiments conducted on MOBSO-NS on four real network datasets by comparing them with the classic methods and presents detailed information and empirical results of the proposed algorithms. Section 5 summarizes the conclusions and discussed future work.

2. Related Works

2.1. Multiobjective Optimization Problem. The problem of multiobjective optimization is to find a compromise solution among the conflicting objectives through the mutual restriction of decision variables. The solution can simultaneously make multiple subobjective functions as optimal as possible. It can be modeled as the following equation:

$$\begin{aligned} \min F(X) &= [f_1(x), f_2(x), \dots, f_k(x)]^T \\ \text{s.t. } g_i(x) &\leq 0, i \\ h_{i=0}, i &= 1, 2, \dots, q, \end{aligned} \quad (1)$$

where $x = (x_1, x_2, \dots, x_m) \in \Omega$ represents the m dimension decision vector, where m means the dimension of the decision vector and Ω is the feasible region of the decision space. Multiobjective optimization problem is to optimize the function vector $\min F(X) = \min[f_1(x), f_2(x), \dots, f_k(x)]^T$, where k is the number of objective functions. The objective function $F(X)$ is composed of k functions mapped from the feasible region of the decision space to the objective space, where $g_i \leq 0, i = 1, 2, \dots, p$, is p inequality constraints, and $h_i = 0, i = 1, 2, \dots, q$, is q equality constraints.

The following are some important definitions of a domination-based multiobjective evolutionary algorithm.

Definition 1 (Pareto dominance). In the decision vector $X_A, X_B \in \Omega$, if and only if $i = 1, 2, \dots, k$, it is arbitrary, X_A is dominant X_B , and $f(x)$ dominant relation is consistent with x dominant relation;

Definition 2 (Pareto optimal solution). Assuming that there is a decision variable x^* in the feasible decision space Ω , if there is no decision variable $x \in \Omega$ like $x > x^*$, the decision variable x^* is called the Pareto optimal solution.

Definition 3 (Pareto optimal solution set). The set consisting of all noninferior optimal solutions is called the optimal solution set of the multiobjective optimization problem, also known as Pareto optimal solution set.

2.2. Multiobjective Community Detection. A concrete network can be modeled as an undirected graph $G = (V, E)$, where V is a set of nodes and E is a set of edges connecting two elements in V . If the network is composed of N nodes,

the graph can be represented by an $N \times N$ adjacency matrix A , and the undirected graph shown in Figure 1 is represented by a symmetric adjacency matrix. Its adjacency matrix A is a symmetric matrix. When $A_{ij} = 1$, it means that there is an edge connection between node I and node J , and when $A_{ij} = 0$, it means that there is an infinite connection between the two nodes.

Radicchi et al. [22] gave the definition of community in an undirected network. Supposing there is a subgraph $S \subset G$ and node i is a node in the subgraph S , the degree of node i is defined as $K_i(S) = K_i^{\text{in}}(S) + K_i^{\text{out}}(S)$. $K_i^{\text{in}}(S) = \sum_{j \in S} A_{ij}$ represents the internal degree of node i , that is, the number of edges between node i and other nodes in subgraph S . $K_i^{\text{out}}(S) = \sum_{j \notin S} A_{ij}$ represents the external degree of node i , namely, the number of node edges outside node i and subgraph S . Communities are obtained by categorizing the structural information of nodes in the network, so a community is a group of nodes, in which the connections between nodes in the same community are close, while the connections between communities are relatively sparse.

In the MOEA/D-net algorithm [19], the module density function D , as shown in formula (2), is decomposed into two different aspects of a society and are divided into 2 evaluation functions: Ration Association (RA) and Ration Cut (RC). In order to convert them into a minimization problem, the function Ration Association (RA) invert gets Negative Ratio Association (NRA), and the optimal solution is obtained by minimizing the NRA and RC set:

$$D = \sum_{i=1}^k \frac{L(V_i, V_i) - L(V_i, \bar{V}_i)}{|V_i|}, \quad (2)$$

where k is the number of community, $I \in \{1, 2, \dots, m\}$, V_i is the set of all nodes within community i , \bar{V}_i is the set of nodes connected externally by community i , $|V_i|$ is the number of internal nodes within community i , $L(V_i, V_i)$ is the number of edges within community i , and $L(V_i, \bar{V}_i)$ is the number of edges between community i and external nodes.

The ratio of community internal connection NRA and the ratio of community external connection RC are taken as the objective function. RC is the sum of the density of connections between communities, and NRA is the sum of the density of connections between nodes in the community, as shown in formula (3), where $V_i \in P, \bar{V}_i = P - V_i$, $L(V_i, V_i) = \sum_{j \in V_i} A_{ij}$, $L(V_i, \bar{V}_i) = \sum_{j \in V_i, j \in \bar{V}_i} A_{ij}$:

$$\min \left\{ \begin{array}{l} \text{NRA} = -\sum_{i=1}^m \frac{L(V_i, V_i)}{|V_i|}, \\ \text{RC} = \sum_{i=1}^m \frac{L(V_i, \bar{V}_i)}{|V_i|}. \end{array} \right. \quad (3)$$

Through minimizing the NRA and RC functions, communities are found with dense internal connections and sparse internal connections. The optimization of RC function can reduce the number of communities and increase the number of nodes in the community. These two objective functions can balance the effect of reducing or increasing the number of communities.

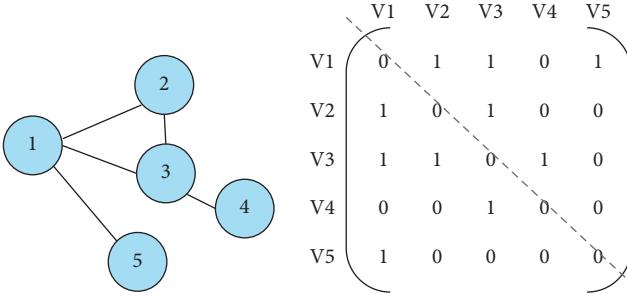


FIGURE 1: Adjacency matrix represents an undirected graph.

2.3. Brain Storm Optimization Algorithm. Brain storm optimization algorithm (BSO) is a random optimization algorithm based on human brain storm process where a group of people with different backgrounds are gathered to brainstorm on the same problem, propose a large number of ideas for the problem to be solved, and finally obtain the optimal solution through mutual communication and thought fusion. It has novel ideas and strong global searching capability. In the algorithm, the search strategy based on grouping operation enhances the search capability, and the individual update method increases the diversity of solutions and makes use of the advantages of human intelligence in dealing with problems. Thus, it can be seen that it is a very potential algorithm, which can produce results beyond the classical intelligent algorithm.

At present, the brain storm optimization algorithm has been successfully applied to solve complex, nonconvex, and NP-difficult problems with highly correlated variables. The brainstorming optimization algorithm clusters information with the clustering method. The clustering center in each class is the optimal value of the class. The cooperation among the classes and the perturbation of the clustering operation make the algorithm jump out of the local optimal and carry out the global search, thus ensuring the convergence performance of the algorithm through the optimization process of the clustering center.

The process of K-means clustering adopted by BSO algorithm is relatively complex. For this reason, Shi [23] proposed a clustering-based brain storm optimization algorithm in objective space (BSO-OS). After the objective space is sorted, several optimal individuals are selected as elite groups and other individuals as ordinary groups. The algorithm has been improved on the clustering operation, which makes the algorithm faster and more time-efficient. However, because the classification method is too simple, the population diversity cannot be guaranteed during the evolution process.

Multiobjective brain storm optimization (MOBSO) saves Pareto optimal solutions by setting the strategy of archive set and eventually obtains a set of uniform solutions close enough to the optimal frontier. MOBSO can solve the optimization problem of the two objectives well and be applied to practice. However, it is found that MOBSO increases the algorithm complexity and reduces the operation efficiency due to clustering and variation, while the increase of objective function has led to the evolution slowdown and

the low efficiency of diversity maintenance strategy. Therefore, it is necessary to redesign the objective space and design the novelty search mechanism to settle the above problems. Section 3 focuses on the novelty search mechanism.

3. Multiobjective Brain Storm Community Detection Method Based on Novelty Search

3.1. Novelty Search Mechanism. In each generation of evolutionary agents, traditional evolutionary algorithms, with a method of a single population search, select the best performing individual to produce offspring according to some metric, but this single selection will cause a loss of diversity. To solve this problem, a multipopulation parallel search based on elite, ordinary, and novelty clusters is used to enhance the diversity of the algorithm. In novelty searching, though the novelty individual is not based on the optimal fitness in the target search, it is the solution far away from the best individual, so it is selected as the starting point for further evolution in this mechanism. This search mechanism can correct for the loss of diversity and stagnation of evolution in a single evolutionary population.

Figure 2 shows the schematic diagram of novelty search mechanism, in which point A represents the local best, point B represents the global best, and point C represents the novelty solution. Obviously, between point C and the global optimal point B is less than that between point C and the local optimal point A. In the iteration process, the novelty searching mechanism searches from different positions and selects the solution opposite to the current optimal fitness as the novelty solution, which ensures the balance of population diversity and convergence to a great extent and has good robustness.

According to the above mechanism, in complex network community detection, traditional evolutionary algorithms often choose the maximum value of *NMI* as the starting point of continuous evolution, while here, the most novel individual, namely, the minimum value of *NMI*, is selected as the novelty solution by maximizing novelty measurement. Under the specific operation, the external archive set (storing elite solutions) and the current generated population are utilized to perform nondominant ranking, the *NMI* value of the solution obtained from the external archive is compared with that of the original population, and the one with the lowest *NMI* value is taken as the novelty individual. In the individual renewal strategy, two new solutions are generated as the elite solution and the novel solution by the two-point cross fusion of the external archived solution.

3.2. The Overall Flow of the Algorithm. In this paper, a multiobjective brain storm optimization based on novelty search (MOBSO-NS) community detection method is proposed to design a novelty search strategy, where the objective space is divided into elite cluster, the common cluster, and novelty cluster, among which, the novelty cluster is the smallest among the individuals with the *NMI*

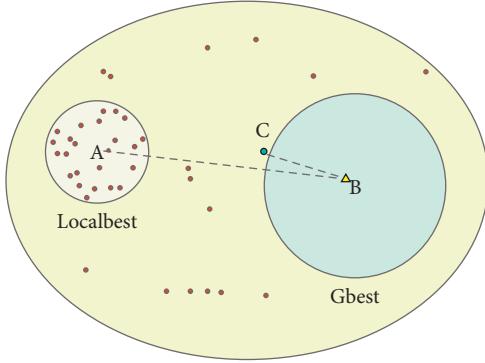


FIGURE 2: Novelty search mechanism.

value as a novelty solution. At present, many swarm intelligence optimization algorithms replace traditional selection models with mutation methods to generate new individuals, which increases the diversity of information, prevents the algorithm from falling into local optimum, and enhances the algorithm's global search ability. In MOBSO-OS, the offspring are generated by the cross-fusion of elite solutions and novel solutions, which effectively maintains the diversity of the population. As the search process progresses, the individual solutions gradually get better, making the gap between each solution smaller. At this point, even if the interaction of individuals in the population is used to reach the optimal solution, the population may fall into the local optimal solution. The strategy of restarting is used to help individuals escape from the local optimal point, avoid "precocity" caused by premature convergence, form a single optimal solution, and improve the global searching ability of the population. The external archive is set to save Pareto optimal solution. In each iteration, the nondominant solution is sorted into the nondominant solution in the external archive and then updated to finally obtain a set of Pareto optimal solutions.

The flowchart of multiobjective brainstorm community detection method based on novelty search is presented in Figure 3.

Steps of MOBSO-NS algorithm for detecting complex network communities:

Step 1: read the input network and initialize the algorithm by means of the neighbor node-based LAR coding, as shown in Figure 4.

Step 2: randomly generate the initial $popNum$ solutions and calculate the NRA and RC values of the initial solutions by formula (3).

Step 3: update the external archive EP including all the solutions in the population.

Step 4: elite individuals are disturbed. An individual $C1$ is randomly selected from the external archive EP to generate new individuals.

Step 5: obtain the novelty solution, calculate the NMI value between the solution in the external archive EP and the solution of the original population, and take the solution with small NMI value as the novelty solution.

Step 6: randomly select individuals $C1$ and $C2$ from the external archive and the current population to generate new individuals by two-point crossover fusion of elite solution and ordinary solution to generate new individuals.

Step 7: randomly select individuals $C1$ and $C2$ from external archives and novelty solution to generate new individuals by the cross fusion of elite individuals and novel individuals.

Step 8: calculate the NRA and RC values for the new population and update the external archive.

Step 9: when the external archive Q is not updated or the number of iterations p is reached, the restart operation is performed, returning to step 2.

Step 10: determine whether the termination condition is met. If so, calculate the Q value and the maximum NMI value of the external archive, otherwise, return to step 4.

Step 11: output a set of Pareto frontiers, that is, a set of partition network structures.

3.3. Coding Method Based on Neighbor Node. The encoding method based on neighbor nodes is adopted in this paper. All nodes in the graph are first numbered, then the neighbors of each node are sorted according to their numbers, and eventually the neighbor ordered table is obtained. In this coding method based on neighbor node, nodes located in the same connected subgraph are divided into a community so that each individual is valid and no illegal solution will occur, thus ensuring the convergence of the algorithm, and there is no need to know the number of partitioned communities in advance, which can be obtained through the decoding process.

The encoding method based on neighbor nodes consists of four steps: determining the neighbor set of each node, selecting the neighbor of each node, constructing the encoding list, and carrying out the decoding process. Taking Figure 4 as an example, the process is as follows.

3.3.1. Determine the Neighbor Set of Each Node. In a network topology, all nodes to which each node is connected via an edge are called the set of neighbors of that node. Figure 4(a) shows the topological structure of 11 nodes. Taking node 7 as an example, the number of the nodes it connects are 4, 6, and 8, respectively, that is, the neighbor set of node 7 is $\{4, 6, 8\}$.

3.3.2. Select the Neighbor of Each Node. Each node will randomly select a neighbor from the neighbor set as the neighbor of the current node. If the corresponding gene value of the No. i node is j , it can be interpreted as there is an edge connection between node i and node j . According to the seventh node with ID 7 in Figure 4(b), the gene value is taken as 8. In the corresponding figure, there is an edge from between nodes 7 and 8.

3.3.3. Construct the Encoding List. The list of each node and its randomly selected neighbors is defined as the encoding list, as shown in Figure 4(b). Position refers to the node

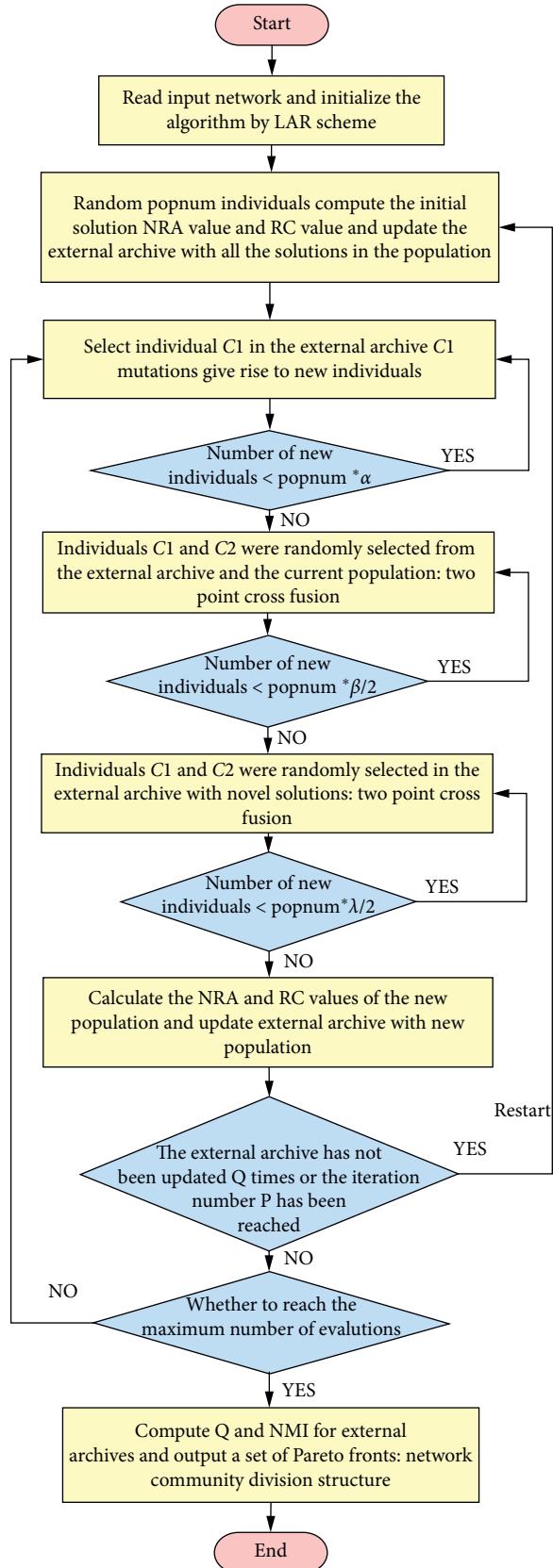


FIGURE 3: Flowchart of complex network community detection.

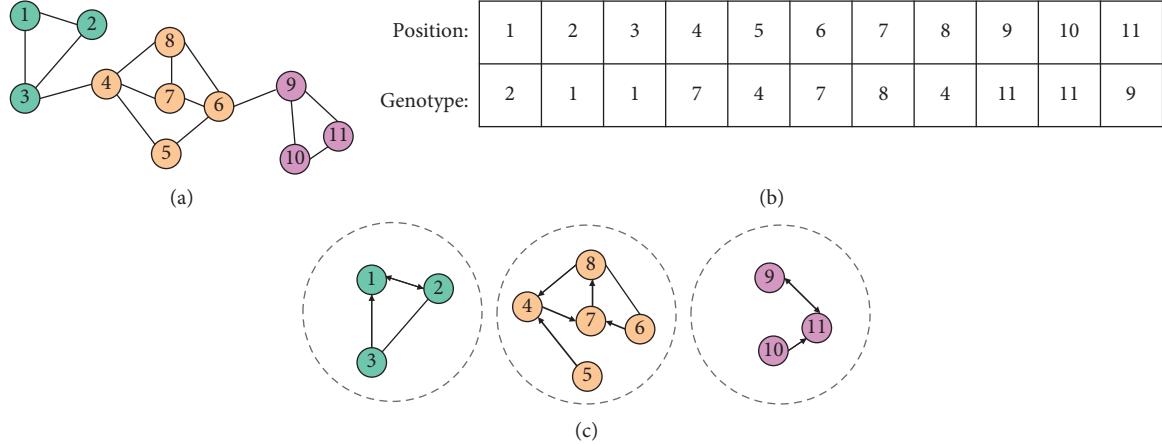


FIGURE 4: Encoding and decoding modes based on neighbor nodes. (a) A network with 11 nodes. (b) Encoding method based on neighbor nodes. (c) Decoded connected subgraph.

position, and genotype means the gene bit formed by randomly selected neighbors of each node.

3.3.4. Decoding Process. Decoding is the process of converting the encoded list into the community structure corresponding to the figure. All community divisions need to be identified during the decoding process, as shown in Figure 4(c). The network is divided into three communities, {1, 2, 3}, {3, 4, 6, 7, 8}, and {9, 10, 11}.

3.4. Update the External Archive. An external archive is set up in the algorithm to save the nondominant solution obtained by the algorithm after searching. After the objective value of the individual is calculated, each individual needs to be compared with other individuals in the population to determine whether it is a nondominant solution, that is, the more solutions stored in the external archive are , the more representative the Pareto frontier becomes.

In this paper, the solutions in the population and the solutions in the external archive are sorted in a nondominant way to update the external archive. In the iteration search phase, if a new nondominant solution is more dominant than the elements in the library, the elements in the library will be removed from the library, and if the nondominant solution in the newly generated population is dominated by some members of the library, the nondominant solution cannot enter the library.

3.5. Restart Policy. In the process of population iteration, the failure of an individual to constantly update its position means that the individual has been in the best position. If the individual is not helped to escape from this point in time, as the group evolves, more and more individuals will gather around the optimal point. If the greatest advantage, for the local optimal point, is "premature" phenomenon in a group, the ability to explore new global optimal solution will be lost. In order to overcome this defect, MOBSO-NS algorithm adopts a restart strategy to help individuals jump

out of the local optimal in time and maintain global search ability of the group.

In Algorithm 1, the restart judgment conditions are divided into the following two points:

- (1) In the iteration process, if the external nondominant population is the same for Q times continuously, the population is considered to converge to the local optimal when Q times are not renewed. In this paper, Q is referred to as the maximum number of times the external archive has not been updated, and the value of Q is determined based on different datasets.
- (2) P is the maximum number of iterations to restart. If the number of iterations reaches P , the restart strategy will be executed.

3.6. Multiobjective Brain Storm Optimization Community Detection Method Based on Novelty Search. The framework of multiobjective brain storm optimization community detection method based on novelty search is shown in the figure below, where $Pops$ is the initial population list, EP is the external archive, $tempRestart$ is the number of times that the external archive has not been updated, and $tempIter$ is the number of iterations. The overall algorithm framework of MOBSO-NS is shown in Algorithm 2.

Individual renewal adopts three strategies: first, elite cluster generates part of the new population through mutation; second, elite cluster and ordinary cluster are fused to generate part of the new population; and third, elite cluster and novelty cluster are fused to generate part of the new population to maintain the population diversity. The specific process is shown in Algorithm 3.

4. Experiment and Analysis

4.1. Dataset. Comparing with other related methods, the MBSO-OS algorithm was experimentally analyzed in four real networks including Zachary's karate club, the Bottlenose dolphins, the American college football, and the books about US Politics by comparing with other related methods.

```
(1) If  $tempRestart == Q$  or  $tempIter == P$  do:  

(2)   go to Algorithm 2 step 3
```

ALGORITHM 1: Restart policy.

Input: the edge set data of complex network, the maximum number of iterations of the algorithm $maxIter$, the population size $popNum$, and the proportion of the three new individuals generating the new population are different $\alpha, \beta, \gamma (\alpha + \beta + \gamma = 1)$; Q is the maximum number of times that the partial archive has not been updated, and P is the maximum number of iterations restarted.

Output: a group of network communities are divided into structures.

- (1) Initialize: $Pops = \emptyset$, $EP = \emptyset$
- (2) **For** $i = 1$ to $popNum$ **do**:
- (3) The LAR code is used to randomly generate an initial solution s , as shown in Figure 4, and the NRA and RC values for s are calculated according to formula (2).
- (4) Add s to $Pops$.
- (5) **End For**
- (6) Iterative search:
- (7) **For** $iter = 1$ to $maxIter$ **do**:
- (8) **For** s in $popNum$ **do**:
- (9) If no solution in EP can dominate s **do**
- (10) Add s to EP , and remove all solutions in EP that can be dominated by s .
- (11) **End If**
- (12) **End For**
- (13) $newPops = \emptyset$.
- (14) Individual update
- (15) Restart
- (16) **End For**
- (17) Calculate the indicator value:
- (18) **For** s in EP **do**:
- (19) Calculate the Q value and NMI value of s .
- (20) **End For**
- (21) Return the two solutions SI and $S2$ a maximum value of 0 and NMI in EP

ALGORITHM 2: MOBSO-NS.

//The elite cluster generates part of the new population through mutation

- (1) **For** $i = 1$ to $popNum * \alpha$
- (2) Randomly select a solution $P1$ from EP , and then perform a mutation operation to generate a new solution $C1$.
- (3) Calculate the NRA and RC values of $C1$ and add $C1$ to $newPops$.
- (4) **End For**
- //Elite clusters and ordinary clusters are fused to generate a new population.
- (5) **For** $i = 1$ to $popNum * \beta / 2$
- (6) Randomly select a solution $P1$ and $P2$ from EP and $Pops$, and perform the fusion operation (two-point crossover) operation
- (7) Generate two new solutions $C1$ and $C2$.
- (8) Calculate the NRA and RC values of $C1$ and $C2$, and add $C1$ and $C2$ to $newPops$.
- (9) **End For**
- //Elite clusters and novelty clusters merge are fused to generate some new populations.
- (10) **For** $i = 1$ to $popNum * \lambda / 2$
- (11) Randomly select a solution $P1$ from the EP , calculate the NMI value of all solutions in $P1$ and $Pops$, and select the $P2$ corresponding to the solution with the smallest NMI value. Perform a fusion operation (two-point crossover) into two new solutions $C1$ and $C2$.
- Calculate the NRA and RC values of $C1$ and $C2$, and add $C1$ and $C2$ to $newPops$.
- (12) **End For**

ALGORITHM 3: Individual update strategy.

The network of Zachary's karate club represents a social relationship between 34 members of a university karate club in the United States, where each node represents a member, and the edge between two nodes means that the corresponding two members are friends in frequent contacts. Due to the differences between the club management and the coaches, the club was finally divided into two: the coach-centered team was split from the club, forming a new club, while the manager-centered team remained at the club. The network has a total of 34 nodes and 78 links.

The dolphin network is an animal social network constructed by Lusseau by observing 62 dolphins of different genders in New Zealand's Doubtful Bay. Each node in the network represents a dolphin. If two dolphins are closely connected, there will be an edge connection between the corresponding vertices of the dolphins that are naturally divided into two communities: male and female groups.

The network of the American college football has a total of 77 nodes with 121 edges. The nodes in the network represent football teams, and the edge between the nodes indicates that a game has already been played between the two teams.

The political books' network edited by Krebs consists of 105 nodes and 441 edges. The nodes represent books on American politics from Amazon, and the edges between the two nodes indicate that the two books are frequently purchased together. Newman divided the books into different parts in light of the political views of the book, with a few exceptions.

Table 1 shows the number of nodes and edges of the four real networks.

4.2. Introduction to Evaluation Indexes. Currently, the most widely used community quality evaluation index is the Q value function developed by Newman and Girva [24]. The modularity of Q is defined as follows:

$$Q = \sum_{i=1}^K \left[\frac{e_i}{m} - \left(\frac{d_i}{2m} \right)^2 \right]. \quad (4)$$

By studying the related problems of complex network clustering based on optimization methods, the algorithm selects the two most common community quality evaluation indicators (Q , NMI) as the objective function.

The standard of modularity is a measure of the degree of goodness of the identified communities in the network. It is considered that the larger the Q value, the stronger the community structure. The modularity is defined as the score of the edges falling into the community minus the expected probability of random allocation of these edges, and the edges are randomly added to the network, independent of the community structure. k is the number of clusters found in the network, e_i is the total number of edges connecting nodes in cluster i , d_i is the sum of nodes in cluster i , and m is the total number of edges in the network. The standard of modularity is generally within the range of $[-0.5, 1]$, whereas most practical networks have a modularity value within the

range of $[0.3, 0.7]$, where a value greater than 0.3 indicates an important community structure.

Normalized mutual information (NMI) measures the similarity between the actual classification of societies and the detected societies. According to information theory, normalized mutual information NMI (A , B) is defined as follows:

$$NMI(A, B) = \frac{-2 \sum_{i=1}^{C_A} \sum_{j=1}^{C_B} C_{ij} \log(C_{ij} n / C_{i\cdot} C_{\cdot j})}{\sum_{i=1}^{C_A} C_{i\cdot} \log(C_{i\cdot} n) + \sum_{j=1}^{C_B} C_{\cdot j} \log(C_{\cdot j} n)}. \quad (5)$$

Set two partitions of the network as A and B , and let C represent the mixed matrix and its element C_{ij} the number of nodes that appear in communities $A_i \in A$, $B_j \in B$.

Among them, C_A and C_B , respectively, signify the number of communities in the partition of A and B , $C_{i\cdot}$ represents the sum of the elements in matrix C , over row i , $C_{\cdot j}$ indicates the sum of elements in matrix C , over column j , and n is the number of nodes in the network. The value of NMI ranges from 0 and 1. If A and B are completely consistent, the maximum value of NMI is 1, while if A and B are completely inconsistent, for example, the whole network is detected as A community, the minimum value of NMI is 0.

4.3. Parameter Setting. In order to verify the detection performance of the algorithm proposed in this paper, python3.7 was used to program the algorithm under the MAC OS X environment. The performance of the algorithm was simulated using real networks, respectively, and the test results were measured with standardized normalized mutual information. The maximum number of iterations of all datasets is set as 160, and the population number of $popNum$ is set as 100 for each dataset. The objective space is divided into elite cluster, novelty cluster, and ordinary cluster. The proportions of new population α generated by mutation of elite cluster, new population β generated by the fusion of elite cluster and ordinary cluster, and new population λ generated by the fusion of elite cluster and novelty cluster are set as 0.5, 0.4, and 0.1, respectively, and the EP (external archive has not been updated) restart parameter Q is 5, and the number of iterations restart parameter P is 20.

4.4. Test Results and Analysis

4.4.1. MOBSO-NS Test Results. For each network, the algorithm runs 30 times. After each run, the best Q value and NMI value are selected for segmentation and recording. After 30 runs, the average is taken between the best Q value and NMI each time. The results are shown in Table 2. From the results of Zachary's karate club network and Bottlenose dolphins' experiments, it can be seen that the MOBSO-NS NMI value reaches the maximum value of 1, which indicates that the results of the community detected by the algorithm are the same as the real community division. The American College Football also has an NMI value of close to 1. For the Q value, the four network division results are between 0.3 and 0.7.

TABLE 1: Network in the experiment.

Networks	Nodes	Edges
Karate	34	78
Dolphins	62	159
Football	77	121
Political books	105	441

TABLE 2: MOBSO-NS experimental results.

Network	Karate	Dolphin	Football	Political books
NMI_{\max}	1	1	0.9367	0.6287
NMI_{avg}	1	1	0.9229	0.6101
Q_{\max}	0.4198	0.5220	0.6044	0.5269
Q_{avg}	0.4198	0.5220	0.6023	0.5269

4.4.2. Parameter Analysis. There are three individual generation operations in the algorithm, involving three parameters α , β , and λ , among which, α is the proportion of the offspring generated by the disturbance of the elite individual, β is the proportion of the new population generated by the fusion of the elite cluster and the ordinary cluster, and λ is the proportion of new populations generated by the fusion of elite clusters and novelty clusters. In order to ensure the validity of the experimental results, a higher weight should be assigned to the elite clusters, with ordinary clusters and novel clusters as an aid to maintain the diversity of the population, which means the value of α should be above 0.5. In this paper, the parameter α starts from 0.5 and gradually increases to 0.8 in steps of 0.1, β corresponds to 0.4, 0.2, 0.2, and 0.1, and λ is 0.1, 0.2, 0.1, and 0.1. The rest parameters remain unchanged. The statistics of NMI_{\max} , NMI_{avg} , Q_{\max} , and Q_{avg} are performed on the test results as shown in Tables 3–5. Figures 5–8 are the NMI_{avg} line graphs of the values of the three parameters α , β , and λ for four datasets: Zachary’s karate club, the Bottlenose dolphins, the American college football, and the political books.

From the experimental results in Tables 3–5, it can be seen that although the values of Q_{\max} and NMI_{\max} are higher than those in Table 2 on individual datasets, the maximum Q values and maximum NMI values that appear due to the randomness of the algorithm have no reference value, and their mean values are both less than those in Table 2. Therefore, the experimental results reach the best when α is 0.5, β is 0.4, and λ is 0.1.

As can be seen from Figures 5–8, among the four sets of different values of α , β , and λ , only when α is 0.5, β is 0.4, and λ is 0.1, NMI_{\max} , NMI_{avg} , Q_{\max} , and Q_{avg} performed the best in four evaluation indexes. From the experimental results of the karate network, it can be found from the analysis in Figure 5 that NMI is 1 regardless of the value of α . This is because the size of the karate network is very small, α is the proportion of offspring generated by elite individual disturbance, and the changes of α , β , and λ have no effect on NMI . For the network, only when α is equal to 0.5, the result is the best, while when α is equal to 0.4, the NMI value decreases, and the scale of the network is larger than karate network, as shown in Figure 6. In Figures 7 and

TABLE 3: The value of parameter α is 0.6, β is 0.2, and λ is 0.2

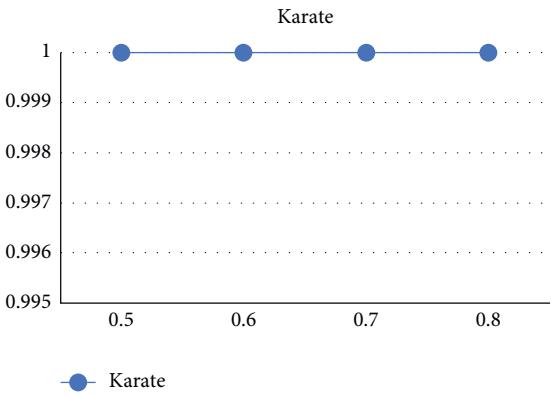
Network	Karate	Dolphins	Football	Political books
NMI_{\max}	1	1	0.9269	0.6310
NMI_{avg}	1	1	0.9058	0.5985
Q_{\max}	0.4198	0.5222	0.6044	0.5269
Q_{avg}	0.4198	0.5195	0.5984	0.5259

TABLE 4: The value of parameter α is 0.7, β is 0.2, and λ is 0.1

Network	Karate	Dolphins	Football	Political books
NMI_{\max}	1	1	0.9269	0.6313
NMI_{avg}	1	0.9974	0.9102	0.6018
Q_{\max}	0.4198	0.5238	0.6044	0.5269
Q_{avg}	0.4195	0.5200	0.6000	0.5264

TABLE 5: The value of parameter α is 0.8, β is 0.1, and λ is 0.1

Network	Karate	Dolphins	Football	Political books
NMI_{\max}	1	1	0.9361	0.6287
NMI_{avg}	1	0.9963	0.9094	0.5972
Q_{\max}	0.4198	0.5268	0.6046	0.5269
Q_{avg}	0.4198	0.5211	0.6001	0.5260

FIGURE 5: Line chart in karate when α , β , and λ are given different values.

8, a similar line graph appears. When α is 0.6, β is 0.2, and λ is 0.2, the NMI value is the smallest, which is caused by the large number of nodes in football network and political books’ network.

Figures 9 and 10, respectively, show the real communities detected on Zachary’s karate club network and dolphins social network through MOBSO-NS. The different colors of the nodes indicate different communities obtained by the algorithm.

Figure 11 shows the partition structure with the highest NMI value and Q value generated by the political book network through the MOBSO-NS algorithm. It can be seen that, for the partition with the highest NMI , the algorithm generates 3 communities, which are exactly equal to the number of correct communities in the political book network. That is to say, the MOBSO-NS algorithm can find a structure that is closer to the real structure.

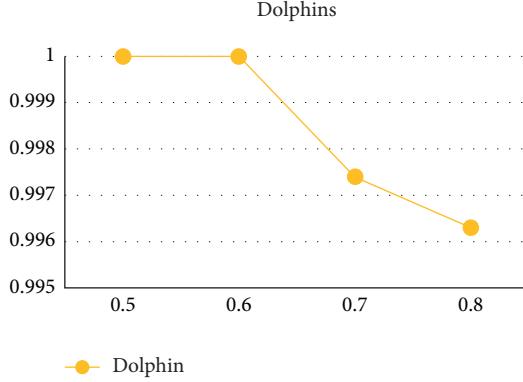


FIGURE 6: Line chart in dolphins when α, β , and λ are given different values.

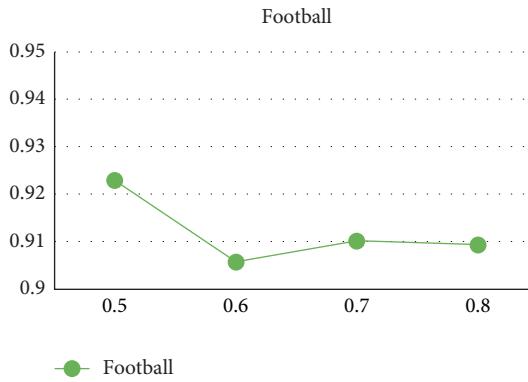


FIGURE 7: Line chart in football when α, β , and λ are given different values.

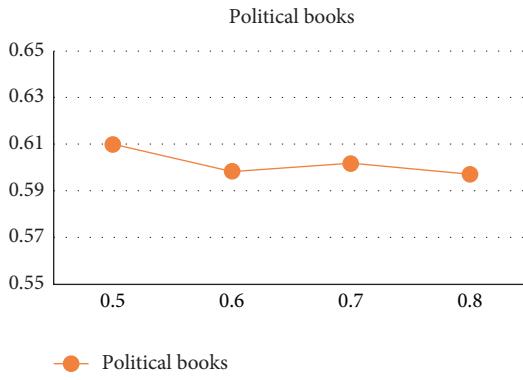


FIGURE 8: Line chart in political books when α, β , and λ are given different values.

Figure 12 shows the communities in which the MOBSONS algorithm detects the maximum NMI on the American college football network. MOBSONS produced 11 communities, but the correct number of communities in the soccer network was 12.

4.5. Comparison of Qualities of Online Communities. In order to demonstrate the advantages of MOBSONS algorithm in community detection from various aspects, the experimental

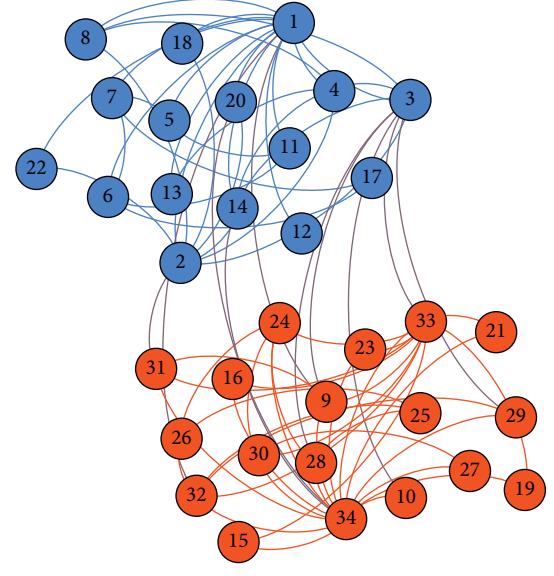


FIGURE 9: Karate.

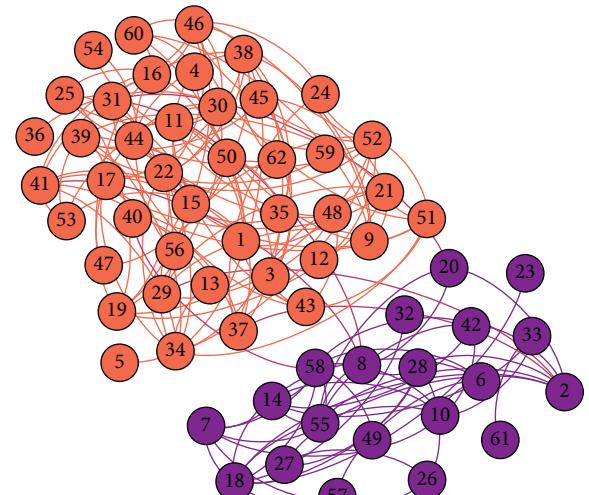


FIGURE 10: Dolphins.

data of MOBSONS algorithm, MOLS-Net, MODPSO, MOEA/D-NET, and BGLL algorithm in NMI_{\max} , NMI_{avg} , Q_{\max} , and Q_{avg} are compared, as shown in Table 6.

Among them, the multiobjective discrete particle swarm optimization (MODPSO) proposed by Gong et al. [15] designed a swarm optimization method specific to the problem of tag propagation, which adopted neighbor-based turbulence operators to produce different individuals and improve diversity, showing high clustering efficiency. Zhou et al. [20] proposed the multiobjective local search (MOLS-NET) algorithm to express the community detection problem as a multiobjective optimization problem and then presented a multiobjective optimization algorithm based on local search. Different target local search methods designed can optimize two targets simultaneously. The BGLL algorithm is an aggregation algorithm proposed by Blondel et al. [25] based on the concept of modularity, which can be used to analyze the

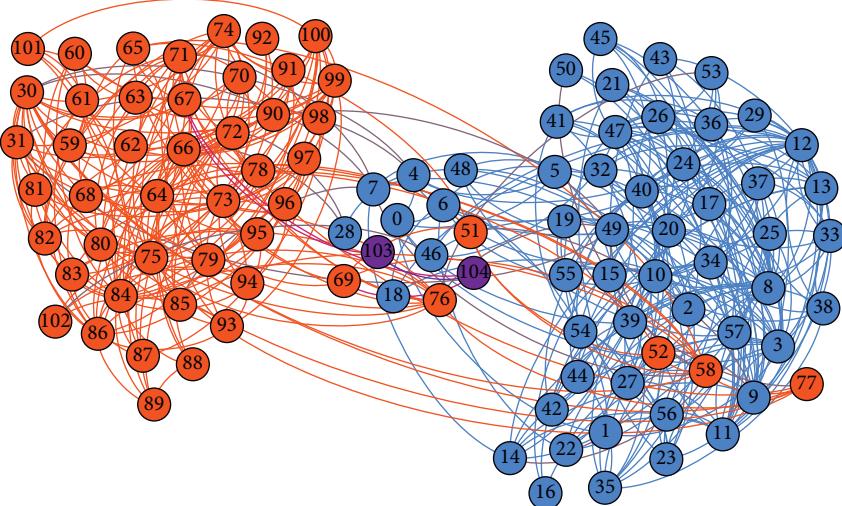


FIGURE 11: Political books.

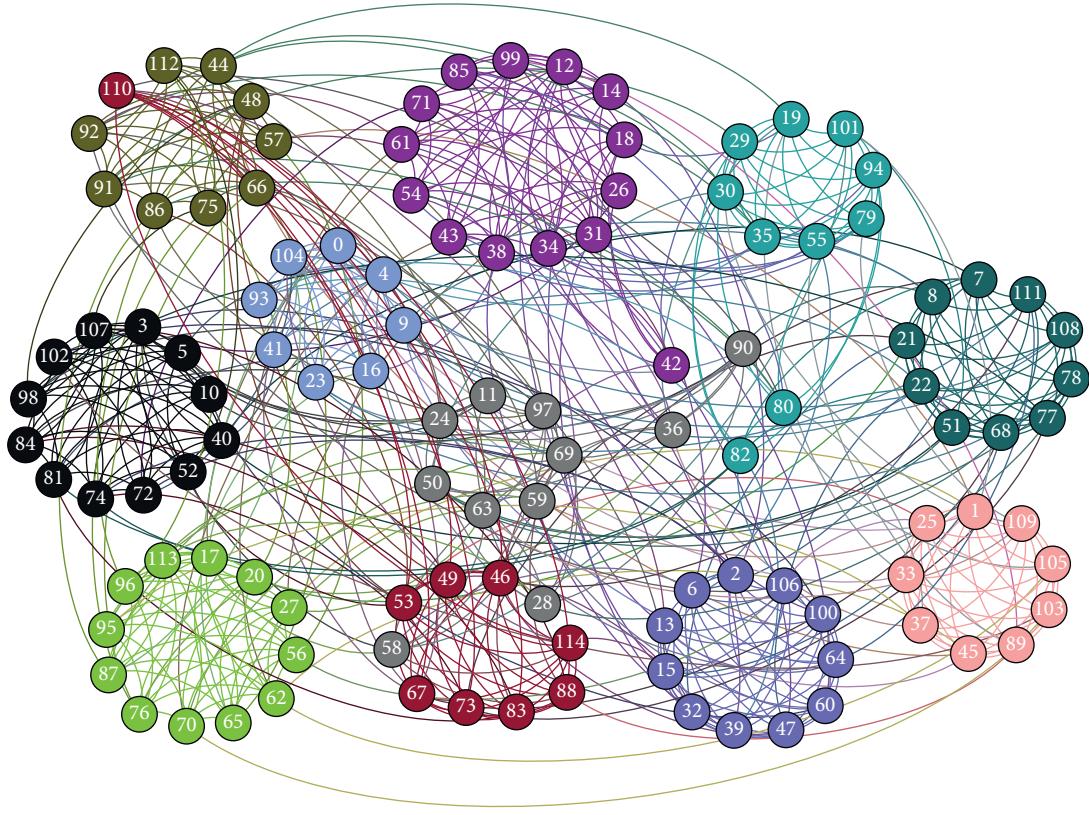


FIGURE 12: Football.

hierarchical structure of weighted networks. The MOEA/D-NET algorithm was designed by Gong et al. [14] to solve the community detection as a multiobjective optimization problem using the decomposition based multiobjective evolutionary algorithm. The proposed algorithm maximized the density of the interior degree and minimized the density of the exterior degree.

Table 6 describes the comparison results between MOBSO-NS proposed in this paper and other four algorithms on the four evaluation indexes. The network of

karate and dolphins is simple because the real network is divided into two communities. The NMI value given by the algorithm in this paper is the same as that of the compared algorithm, both of which are 1. This results in the community structure of the two networks being divided accurately enough. As for the Q value, the MOBSO-NS is not the best in the network, but the NMI value of 1 indicates that the divided network is exactly the same as the real network, which does not affect the experimental results.

TABLE 6: Comparison of network community quality.

Network	Index	MOBSO-NS	BSO-OS	MOLS-Net	MODPSO	MOEA/D-Net	BGLL
Karate	NMI_{\max}	1	1	1	1	1	0.7071
	NMI_{avg}	1	1	1	0.9729	1	0.6280
	Q_{\max}	0.4198	0.4198	0.4198	0.4198	0.3715	0.4198
	Q_{avg}	0.4198	0.4198	0.4198	0.4138	0.3715	0.4118
Dolphin	NMI_{\max}	1	1	1	1	1	0.6363
	NMI_{avg}	1	1	1	1	1	0.5148
	Q_{\max}	0.5220	0.5269	0.5268	0.5268	0.3735	0.5277
	Q_{avg}	0.5220	0.5269	0.5237	0.5196	0.3735	0.5210
Football	NMI_{\max}	0.9367	0.9357	0.9361	0.9289	0.9367	0.8923
	NMI_{avg}	0.9229	0.9229	0.9273	0.9245	0.9334	0.8765
	Q_{\max}	0.6044	0.6044	0.6046	0.6032	0.6005	0.6046
	Q_{avg}	0.6023	0.6032	0.6044	0.5995	0.5993	0.6038
Political books	NMI_{\max}	0.6287	0.6254	—	0.5910	0.5895	0.4420
	NMI_{avg}	0.6101	0.5999	—	0.5910	0.5885	0.4420
	Q_{\max}	0.5269	0.5259	—	0.5263	0.5259	0.5186
	Q_{avg}	0.5269	0.5259	—	0.5263	0.5259	0.5186

For Zachary's karate club network and the Bottlenose dolphins network, because the real network is relatively simple, two communities are divided. The NMI value given by the algorithm in the paper is the same as that given by the compared algorithm, both of which are 1, indicating that these two networks have been accurately divided into network community structures. As for the Q value, although the performance of MOBSO-NS is not the best performance in dolphin network, the NMI value of 1 indicates that the partitioned network is completely consistent with the real network, which has no effect on the experimental results.

The political books' network is complex, and most existing methods fail to detect the true division of the network. MOBSO-NS is the result when the actual number of communities is unknown. Compared with those of MOLS-NET algorithm with the known actual number of communities, the experimental results of MOBSO-NS algorithm are more objective and accurate. In addition, it is clear from the contents recorded in Table 6 that the experimental results in the political books' network and the community quality are much higher than those of the other four known algorithms.

In the American college football network, the real network has 12 community divisions. For the Q value, MOLS-NET is superior to all comparison algorithms. Due to the complexity of the network, no algorithm can realize the real partition structure. Other algorithms perform better than MOBSO-NS in terms of the average Q because the American college football league network is essentially a network with numerous and difficult community divisions. However, it is found in the experimental results that MOBSO-NS algorithm has the maximum NMI value compared with other algorithms, indicating that the community structure divided by the algorithm in this paper is most similar to the real network.

Compared with MOLS-NET, MODPSO, MOEA/D-NET, and BGLL, MOBSO-NS algorithm has significant advantages in terms of average and maximum value of community structure, and the similarity between the community detected by it and the actual community division is also closer. The network of Zachary's karate club and the Bottlenose dolphins,

in particular, has a strong community structure. Because the NMI values of the algorithm are uniformly distributed on each network, the algorithm has stronger robustness.

5. Conclusion

The research of complex network community detection is of great significance to Internet culture security and information personalized service. At present, most of the complex network community detection algorithms based on heuristic optimization are the strength of a single community structure quality evaluation, and the diversity of community quality evaluation indexes makes the network community structure analysis more decision-making. In this paper, a novel multiobjective brain storm community detection method (MOBSO-NS) based on novelty search is proposed to solve the problem of complex network community detection. The novelty search method can effectively avoid premature convergence and enhance the global search ability while maintaining the diversity of the population. Secondly, the restart operation is used to help individuals escape from the local optimal point. The idea of novelty search is integrated into the brainstorming optimization algorithm, and the mechanism of generating new individuals is innovated. Experimental results show that the algorithm in this paper has better optimization ability and can obtain better results of network community division.

There still remains some work related with MOBSO-NS that deserves to be further investigated. The MOBSO-NS suggested in this paper has shown that utilizing local community information is a promising idea to obtain good community partition in the static networks. In the future, we would like to combine this strategy with other frameworks of MOEA, such as NSGA-II, SPEA2, and IBEA, to further explore the local information in other kinds of complex networks, such as overlapping communities and dynamic networks. In addition, as the real data communities tend to be very large, how to improve the performance of MOBSO-NS by identifying the communities with small sizes is also an interesting work.

Data Availability

The data, models, or code generated or used during the study are available at <http://vladowiki.fmf.uni-lj.si/doku.php?id=pajek:data:urls:index>.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This work was supported by the National Natural Science Foundation of China (no. 62001380) and General Special Scientific Research Program of Shaanxi Provincial Education Department (no. 20JK0910).

References

- [1] A. Asikainen, G. Iñiguez, J. Ureña-Carrión et al., “Cumulative effects of triadic closure and homophily in social networks,” *Science Advances*, vol. 6, no. 19, Article ID eaax7310, 2020.
- [2] C. Esposito, “Interoperable, dynamic and privacy-preserving access control for cloud data storage when integrating heterogeneous organizations,” *Journal of Network and Computer Applications*, vol. 108, no. APR, pp. 124–136, 2018.
- [3] K. Devkota, L. J. Murphy, and L. J. Cowen, “GLIDE: combining local methods and diffusion state embeddings to predict missing interactions in biological networks,” *Bioinformatics*, vol. 36, no. Supplement_1, pp. i464–i473, 2020.
- [4] A. Castiglione, R. Pizzolante, A. De Santis, B. Carpentieri, A. Castiglione, and F. Palmieri, “Cloud-based adaptive compression and secure management services for 3D healthcare data,” *Future Generation Computer Systems*, vol. 43–44, pp. 120–134, 2015.
- [5] D. J. Watts and S. H. Strogatz, “Collective dynamics of “small-world” networks,” *Nature*, vol. 393, no. 6684, pp. 440–442, 1998.
- [6] A.-L. Barabási and E. Bonabeau, “Scale-free networks,” *Scientific American*, vol. 288, no. 5, p. 60, 2003.
- [7] A. D. King, N. Przulj, and I. Jurisica, “Protein complex prediction via cost-based clustering,” *Bioinformatics*, vol. 20, no. 17, pp. 3013–3020, 2004.
- [8] T. Sangkaran, N. A. Abdullah, and N. Z. Jhanjhi, “Criminal community detection based on isomorphic subgraph analytics,” *Open Computer Science*, vol. 10, no. 1, pp. 164–174, 2020.
- [9] S. Fortunato, “Community detection in graphs,” *Physics Reports*, vol. 486, no. 3–5, pp. 75–174, 2010.
- [10] M. Tasgin, A. Herdagdelen, and H. Bingol, *Community Detection in Complex Networks Using Genetic Algorithms*, Corrosion -Houston Tx-, Odessa, TX, USA, 2007.
- [11] A. I. Hafez, N. I. Ghali, A. E. Hassanien et al., “Genetic algorithms for community detection in social networks,” in *Proceedings of the International Conference on Intelligent Systems Design & Applications*, IEEE, Vellore, India, September 2013.
- [12] M. Lipczak and E. Milios, “Agglomerative genetic algorithm for clustering in social networks,” in *Proceedings of the 11th Annual Conference on Genetic and Evolutionary Computation*, pp. 1243–1250, New York, NY, USA, July 2009.
- [13] 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,” *Information Sciences*, vol. 369, pp. 402–418, 2016.
- [14] M. Gong, Q. L. Ma, and L. Jiao, “Community detection in networks by using multiobjective evolutionary algorithm with decomposition,” *Physica A: Statistical Mechanics and Its Applications*, vol. 391, no. 15, pp. 4050–4060, 2012.
- [15] L. Li, L. Jiao, J. Zhao et al., “Quantum-behaved discrete multi-objective particle swarm optimization for complex network clustering,” *Pattern Recognition*, vol. 63, pp. 1–14, 2016.
- [16] M. Gong, X. Q. Cai, and L. Ma, “Complex network clustering by multiobjective discrete particle swarm optimization based on decomposition,” *IEEE Transactions on Evolutionary Computation*, vol. 18, no. 1, pp. 82–97, 2014.
- [17] H. Jiang, C. Z. Liu, and X. Y. ZhangSu, “Community detection in complex networks with an ambiguous structure using central node based link prediction,” *Knowledge-Based Systems*, vol. 195, p. 105626, 2020.
- [18] M. Gong, T. Hou, B. Fu et al., “A non-dominated neighbor immune algorithm for Community detection in networks,” in *Proceedings of the 13th annual Conference on Genetic and Evolutionary Computation*, pp. 1627–1634, Dublin, Ireland, July 2011.
- [19] L. Zhang, J. Xia, and C. Fan, J. Qiu and X. Zhang, Multi-objective optimization of critical node detection based on cascade model in complex networks,” *IEEE Transactions on Network Science and Engineering*, vol. 7, no. 3, 2020.
- [20] Y. Zhou, N. J. Wang, and Z. Zhang, “Multiobjective local search for community detection in networks,” *Soft Computing*, vol. 20, no. 8, pp. 3273–3282, 2016.
- [21] D. Malhotra, “Community detection in complex networks using link strength-based hybrid genetic algorithm,” *SN Computer Science*, vol. 2, no. 1, pp. 1–16, 2021.
- [22] F. Radicchi, C. Castellano, F. Cecconi, V. Loreto, and D. Parisi, “Defining and identifying communities in networks,” *Proceedings of the National Academy of Sciences*, vol. 101, no. 9, pp. 2658–2663, 2004.
- [23] Y. Shi, *Brain storm Optimization algorithm in Objective space*, IEEE Congress on Evolutionary Computation, Sendai, China, 2015.
- [24] M. Newman and M. Girvan, “Finding and evaluating community structure in networks,” *Physical Review. E, Statistical, Nonlinear, and Soft Matter Physics*, vol. 69, no. 2, pp. 26113–26120, 2004.
- [25] V. D. Blondel, J. L. Guillaume, R. Lambiotte et al., “Fast unfolding of communities in large networks,” *Journal of Statistical Mechanics Theory & Experiment*, vol. 10, 2008.