Frontiers in Computing and Intelligent Systems
ISSN: 2832-6024 | Vol. 16, No. 2, 2026

Community Detection Algorithm based on Multi-objective Biogeographical Optimization with Community Belongingness

Wei Zhang, Yangyang Xia *

School of Computer Science and Artificial Intelligence, Wanjiang University of Technology, Ma'anshan, Anhui, China

*** Corresponding author:** Yangyang Xia

Abstract: In community detection algorithms based on optimization, although single-objective optimization methods have been widely applied, they only consider a single objective function and ignore other factors related to community structure, making it difficult to effectively reveal complex networks with multiple potential characteristics. In contrast, multi-objective optimization methods can overcome these limitations and have received increasing attention in the field of community detection in recent years. For the problem of community detection in complex networks, this paper proposes a multi-objective biogeographical optimization-based community detection algorithm (MOBBO-CMD) based on community membership. Firstly, by combining non-dominated sorting and crowding distance, the biogeographical optimization (BBO) mechanism is improved, and a BBO model suitable for multi-objective optimization is established. Simultaneously, the proportion of internal connections within the community (NRA) and the proportion of external connections outside the community (RC) are selected as two optimization objectives, modeling the community detection problem as a multi-objective optimization problem to alleviate the limitation of modularity resolution. Secondly, in the migration operation, a one-way migration method based on community membership is proposed to avoid disrupting the existing relatively correct community division. Furthermore, utilizing node division information, a mutation strategy based on boundary nodes is designed to ensure that mutations can produce better solutions. On real network datasets, MOBBO-CMD is experimentally evaluated compared with four contrasting algorithms. The results show that compared to traditional single-objective community detection algorithms, MOBBO-CMD using multi-objective optimization can evaluate community division results more comprehensively and accurately. Compared to other multi-objective contrasting algorithms, MOBBO-CMD improves community detection accuracy and finds communities closer to the true network structure.

Keywords: Multi-objective; Biogeographical Optimization Algorithm; Community Detection.

1. Introduction

In complex networks, discovering latent community structures is crucial for understanding network functions and characteristics. However, community detection is a discrete optimization problem, and its objective function has been proven to be NP-hard [1], making it extremely challenging to find the optimal partition. Intelligent optimization algorithms, with their powerful search and optimization capabilities, are widely used in community detection problems and have achieved good results by optimizing predefined objective functions. Among them, methods based on single-objective optimization are widely applied, but they only consider a single metric and ignore the diverse characteristics of community structures, making them difficult to apply to complex networks. In contrast, multi-objective optimization has two advantages: firstly, it can simultaneously consider multiple metrics such as intra-community tightness and inter-community connectivity, providing a more comprehensive and accurate evaluation; secondly, it can obtain a Pareto optimal solution set, satisfying different research needs and providing multiple partitioning schemes. Therefore, introducing multi-objective optimization helps to deeply understand the complexity of community structures and uncover latent information.

2. Properties

2.1. Encoding Method Currently

most community detection algorithms primarily adopt two methods: label-based encoding and adjacent node-based encoding. In comparison, the label-based encoding method can more intuitively represent the structural characteristics of communities, and thus, this encoding method is also adopted in this paper. During the encoding process, the division of a network containing N nodes is represented by a chromosome with N gene loci (i.e., a vector of length N). Each gene corresponds to a specific node in the network, and the gene value represents the community label (community number).

2.2. Initialization Method

In the MOBBO-CMD algorithm, an initialization strategy based on neighbor nodes is adopted to enhance the accuracy and diversity of the initial population. This strategy stipulates that the gene value corresponding to each gene i can only be derived from the identifier of a node in the neighbor set of node v_i. In other words, if the value of the ith gene position is j, then nodes v_i and v_j are directly connected in the real network.

2.3. Transfer Model

In single-objective BBO, each habitat represents a solution, and migration behavior is determined by HSI. Multi-objective optimization relies on the Pareto non-dominated set to balance various objectives, and the original HSI calculation method is no longer applicable. This paper draws on the non-dominated sorting and crowding distance of NSGA-II [2] to rank habitats: the lower the rank and the greater the crowding distance, the higher the ranking. According to the order of habitat individuals, the immigration rateλ_i and emigration rate $\mu_i$are defined as follows:

$$\begin{cases} \lambda_i = \dfrac{H_{irank}}{Popsize} \\ \mu_i = 1 - \mu_i \end{cases} \quad (1)$$

In the formula, H_{irank} represents the ranked position of the individual in the habitat, and Popsize denotes the total number of individuals in the habitat.

2.4. Migration Method based on Community Belongingness

The standard BBO migration method directly replaces the feature components of one dimension in both habitats. Since this paper adopts a label-based encoding approach, the gene values (community labels) in the habitats are interrelated. Simple replacement may destroy correctly classified communities or result in invalid migration. To address this, we first introduce the single-way crossover method proposed by Tasgin et al. Randomly select a node v_i , find the community label to which v_i belongs and all its nodes in the outgoing habitat X_j, and then uniformly change the labels of these nodes in the incoming habitat X_i to that community label. This method enables the incoming habitat to fully retain the local community of the outgoing habitat.

The single-way crossover approach enables the complete preservation of local communities from the emigrant habitat in the immigrant habitat. To further inherit the excellent traits of the emigrant habitat while avoiding the destruction of existing reasonable communities in the immigrant habitat, this paper proposes a one-way migration approach based on community membership degree. This strategy involves calculating core nodes, ordinary nodes, and boundary nodes using the community membership degree formula during the migration process. The similarity of local communities is determined by comparing whether the core nodes of the local communities in the immigrant habitat are the same as those in the emigrant habitat. Then, the quality of local community division is measured using Local Community Modularity Density (LCMD). The LCMD values of the local communities in the immigrant habitat and the emigrant habitat are calculated separately, and the migration is determined by comparing the LCMD values of the local communities. The community membership degree formula (CMD) and the calculation method of the local community fitness function (LCMD) are as follows:

$$CMD(v,C) = \frac{C_{in}(v,C)}{C_{in}(v,C) + C_{out}(v,C) + \varepsilon} \quad (2)$$

In the formula, $C_{in}(v,C)$ represents the number of edges connecting node v with nodes inside the local community C, while $C_{out}(v,C)$ denotes the number of edges connecting v with nodes outside the local community C. $\varepsilon = 10^{-9}$ is used to prevent division by zero. Clearly, $0 \leq CMD(v,C) \leq 1$. In this paper, we define: Core node: $CMD(v,C) > \theta_{core} = 0.7$, Boundary node: $CMD(v,C) < \theta_{boundary} = 0.3$, and Internal ordinary node: falls between the two and is not treated specially. When $CMD > 0.5$, it indicates that the node is more inclined to stay in the current community, and $CMD > 0.3$ indicates that the node is at the boundary of the community.

$$LCMD = \frac{C_{cin} - C_{cout}}{|C_i|} \quad (3)$$

In the formula, C_{cin} and C_{cout}represent the number of connecting edges between nodes within the local community$C_i$and other nodes within the community, and the number of connecting edges between nodes within the local communityC_i and nodes outside the community, respectively. $|C_i|$| denotes the number of nodes in the local community.

2.5. Differentiated Mutation Strategy based on Node Types

In traditional community detection algorithms, mutation operations often randomly replace the gene value of a node with the identifier of its neighboring nodes. Although this method can ensure the generation of feasible solutions, the selection of mutated nodes is overly random. To address this, this paper proposes a differentiated mutation strategy based on node types: core nodes should undergo conservative mutation (with low probability), boundary nodes, which are crucial for community division, should undergo active mutation (with high probability), and ordinary nodes should adopt a moderate mutation probability. In addition, for the node to be mutated, we sequentially attempt to add it to the communities to which each neighboring node belongs. We calculate the number of edges connecting the node to both the interior and exterior of each candidate community, and compute the community fitness value using the following formula. The community with the largest CE is selected as the mutation result. The formula for the number of edges connected to a node is defined as follows:

$$CE(v) = \frac{G_{in}}{G_{in} + G_{out}} \quad (4)$$

In the formula, G_{in} represents the number of internal connection edges between the mutation node v and the ith community, while G_{out} denotes the number of external connection edges between the mutation node v and the ith community.

2.6. Objective Function

The MOBBO-CMD algorithm selects the proportion of intra-community connections (NRA) and the proportion of inter-community connections (RC) as the objective functions. The formulas for the NRA and RC objective functions are as follows:

$$\begin{cases} NRA = -RA = -\displaystyle\sum_{i=1}^{m} \frac{L(V_i, V_i)}{|V_i|} \\ RC = \displaystyle\sum_{i=1}^{m} \frac{L(V_i, \overline{V_i})}{|V_i|} \end{cases} \quad (5)$$

Among them, RC represents the sum of the average values of edges connecting nodes within each community and those between other communities. The smaller RC is, the fewer the number of communities will be divided, and the connections between communities will be sparse. RA represents the sum of the average values of edges connecting nodes within each community. The larger RA is, the network will be divided into a small number of community structures with high internal connection density.

3. Condition Tests of Rougher Flotation

3.1. Real Network Experiment Results

In this section, the proposed MOBBO-CMD algorithm is employed to perform community detection in real-world networks, and its results are compared and analyzed with those obtained from five other algorithms: GA-NET[3], MOGA-NET[4], MOCD[5], MOGME[6], and MOCD-ACO. Notably, the GA-NET algorithm is a single-objective community detection algorithm. For both the MOBBO-CMD algorithm and the comparative algorithms, the population size is set to 200, the number of iterations is 100, and each dataset is run 30 times.

To verify the authenticity and effectiveness of the MOBBO-CMD algorithm, four commonly used real-world networks were selected for experimental validation, namely the Karate network, Dolphins network, Polbooks network, and Football network. Since these four real-world networks have known community structures, NMI and Q values can be selected as evaluation metrics. Each algorithm was independently run 20 times on each network, and the maximum value of modularity (Q) was recorded as Q_{max}, and the maximum value of NMI was recorded as NMI_{max} after each run. After the 20 independent runs, their average values were taken, with the average value of modularity (Q) recorded as Q_{avg} and the average value of NMI recorded as NMI_{avg}. The experimental results of each algorithm are shown in Table 1.

Table 1. NMI and Q values of the MOBBO-CMD algorithm and comparative algorithms on real networks

Network		Index	GA-net	MOCD	MOCD-ACO	MOGAME	MOBBO-CMD
Karate		NMI_{max}	0.8990	0.8372	1.000	1.000	**1.000**
		NMI_{avg}	0.8029	0.8372	1.000	1.000	**1.000**
		Q_{max}	0.4112	0.4188	0.4197	0.4696	0.4198
		Q_{avg}	0.3768	0.4188	0.4197	0.4696	0.4198
Dolphins		NMI_{max}	0.6725	0.9872	1.000	1.000	**1.000**
		NMI_{avg}	0.6369	0.9968	1.000	0.8235	**1.000**
		Q_{max}	0.4664	0.5238	0.5258	0.5432	0.5276
		Q_{avg}	0.4441	0.5213	0.5230	0.5346	0.5232
Polbooks		NMI_{max}	0.4648	0.5880	0.6240	0.5548	**0.6252**
		NMI_{avg}	0.4462	0.5861	0.5972	0.4920	**0.6000**
		Q_{max}	0.5005	0.5184	0.5208	0.5235	**0.5269**
		Q_{avg}	0.4895	0.5105	0.5194	0.5214	**0.5267**
Football		NMI_{max}	0.8842	0.8631	0.9374	0.9270	**0.9383**
		NMI_{avg}	0.8556	0.8613	0.9268	0.9268	**0.9270**
		Q_{max}	0.5803	0.5860	0.5999	0.6131	0.6040
		Q_{avg}	0.5619	0.5847	0.5886	0.6092	0.6020

In both the Karate network and the Dolphins network, when comparing NMI_{max} and NMI_{avg} (as shown in Table 1), both the MOBBO-CMD algorithm and the MOCD-ACO algorithm effectively detect results consistent with the true community structure, and can identify the true community structure in each independent run. When comparing Q_{max} and Q_{avg}, the MOBBO-CMD algorithm achieves the maximum module value and average modularity value in all 30 independent runs except for the MOGAME algorithm. Although the modularity value of the MOGAME algorithm is higher than that of the MOBBO-CMD algorithm, the MOGAME algorithm does not always detect the true community structure in the Dolphins network. Therefore, on both the Karate network and the Dolphins network, the MOBBO-CMD algorithm demonstrates good accuracy and stability.

In the Polbooks and Football networks, due to the complex structure, all algorithms failed to detect the true community structure. However, as shown in Table 1, the MOBBO-CMD algorithm achieved a higher NMI_{max} =0.6252 and NMI_{avg}=0.6000 compared to other comparative algorithms on the Polbooks network. Additionally, its Q_{max}=0.5269, Q_{avg} =0.5267 were also superior to all comparative algorithms, indicating that the algorithm achieved a better community division. On the Football network, the MOBBO-CMD algorithm achieved Q_{max}=0.6040 and Q_{avg}=0.6020, which were not significantly different from those of the MOGAME algorithm. It can be seen that the MOBBO-CMD algorithm also achieved a better community division structure.

After the MOBBO-CMD algorithm finishes running, it returns a set of non-dominated solutions, each corresponding to a community division result. To verify the hierarchical division capability of the MOBBO-CMD algorithm. Figure 1 depicts the Pareto optimal frontier obtained by the MOBBO-CMD algorithm during a run on the Dolphins network. In Figure 2, the Dolphins network is divided into two communities, corresponding to NMI=1.000 and Q=0.3735, which aligns with the actual network communities. In Figure 3, the Dolphins network is divided into three communities, corresponding to NMI=0.7922 and Q=0.4514. In Figure 4, the Dolphins network is divided into four communities, corresponding to NMI=0.5714 and Q=0.5268, which represents the maximum modularity value.

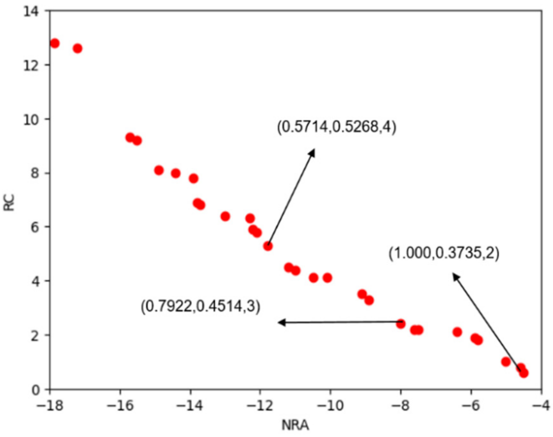

Figure 1. Dolphins network partitioning: Pareto solution set

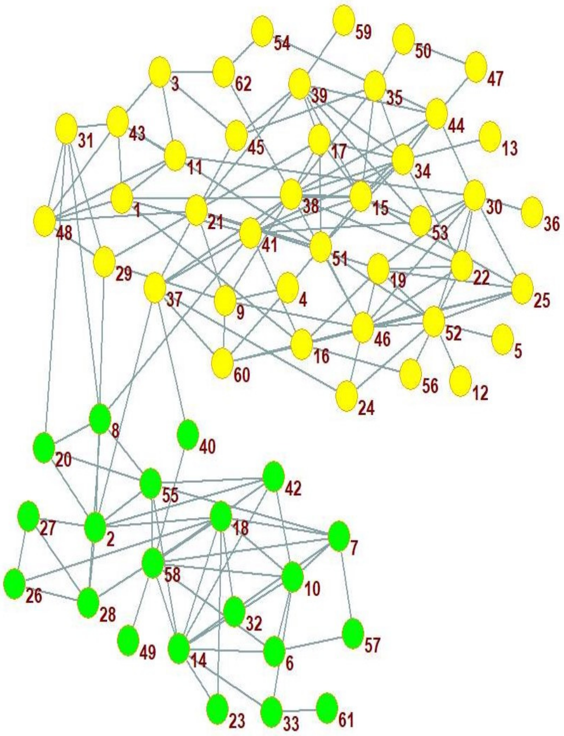

Figure 2. Dolphins network partitioning: two communities

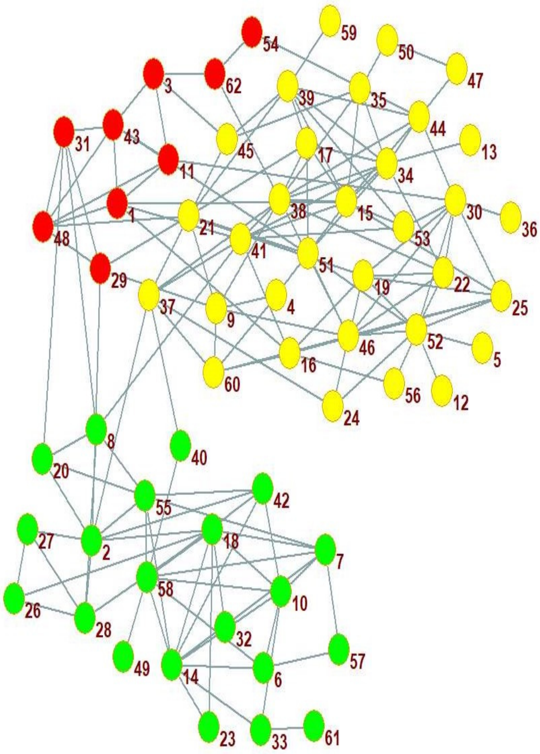

Figure 3. Dolphins network partitioning: three communities

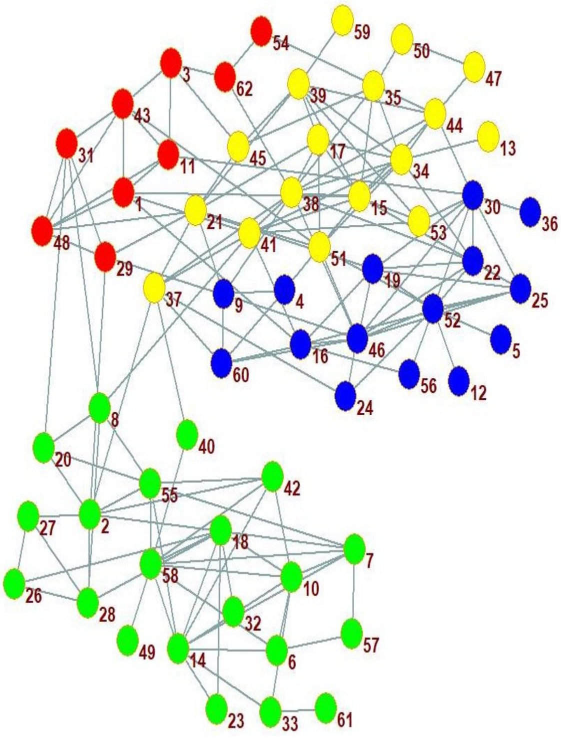

Figure 4. Dolphins network partitioning: four communities

4. Conclusion

Under the framework of biogeographic algorithms and multi-objective optimization, this paper proposes a community detection algorithm based on community membership degree for multi-objective biogeographic optimization (MOBBO-CMD). To effectively address the issue of complex network community detection, MOBBO-CMD makes improvements in the following aspects: considering that the single-objective HSI evaluation method is not suitable for multi-objective optimization, it introduces the non-dominated sorting and crowding distance of NSGA-

II; it improves the traditional single-dimensional migration approach by adopting single-way crossover to preserve the complete community structure, and further determines whether to migrate based on local community modularity density (LCMD) to inherit excellent features; for nodes that may be misclassified, it identifies boundary nodes based on node types and corrects them using node similarity. When applying MOBBO-CMD to real networks, experimental results show that the improvements proposed in this paper are effective. MOBBO-CMD not only discovers superior community structures but also obtains partitions that are closer to real communities, and the solutions obtained exhibit good diversity and hierarchy.

References

[1] ALOISE D, DESHPANDE A, HANSEN P, et al. NP-hardness of Euclidean sum-of-squares clustering [J]. Machine Learning, 2009, 75(2): 245-248.

[2] DEB K, PRATAP A, AGARWAL S, et al. A fast and elitist multiobjective genetic algorithm: NSGA-II [J]. IEEE Transactions on Evolutionary Computation, 2002, 6(2): 182-197.

[3] PIZZUTI C. Ga-net: A genetic algorithm for community detection in social networks[C]; proceedings of the Parallel Problem Solving from Nature–PPSN X: 10th International Conference, Dortmund, Germany, September 13-17, 2008 Proceedings 10. Springer Berlin Heidelberg, 2008 :1081-1090.

[4] PIZZUTI C. A Multiobjective Genetic Algorithm to Find Communities in Complex Networks [J]. Ieee Transactions on Evolutionary Computation, 2012, 16(3): 418-430.

[5] SHI C, YAN Z Y, CAI Y N, et al. Multi-objective community detection in complex networks [J]. Applied Soft Computing, 2012, 12(2): 850-859.

[6] YANG S, LI Q, WEI W, et al. A multi-objective evolutionary algorithm based on mixed encoding for community detection [J]. Multimedia Tools and Applications, 2022: 1-16.