A Modified Genetic Algorithm For Community Detection In Complex Networks

Songran LIU
School of Automation, Wuhan University of Technology
Wuhan 430070, China
983290376@qq.com

Zhe LI
School of Automation, Wuhan University of Technology
Wuhan 430070, China

Abstract— Community detection has a very important role in data processing and analysis, which is very hot in recent years. However, traditional algorithms have shortcomings in both time complexity and precision. In this paper, we introduce a Modified Genetic Algorithm (MGA) that with alleles encoding and half uniform crossover to detect community structure. In the algorithm, each allele of the chromosome stands for the community index of the corresponding node. At the same time, half uniform crossover can better prevent the elite individuals from destroying. And we choose modularity function as its fitness function. It does not need to know how many communities the network has. In order to identify our algorithm is effective. We use both artificial random network and real networks to test our algorithm. The experimental results show that the MGA algorithm can be applied to community detection, and its accuracy and time complexity can reach the effect of classical algorithms.

Keywords—Community detection; Data processing; Genetic Algorithm; Modularity function

I. INTRODUCTION

Community detection is a very important branch in the study of complex networks and gets a high degree of attention from home and abroad. Then, there are many classical algorithms in the division of complex networks. Among those algorithms, the best-known is Spectral Analysis Algorithm based on network topology [1]; Potts Model Algorithm based on network dynamics [2]; GN algorithm based on network topology [3]; Fast Newman algorithm [4]; and some other algorithms based on heuristic algorithm [5-7].

However, these traditional algorithms need to be further improved in both accuracy and time complexity. The accuracy for the development of network science plays an important role. What is more, only to improve the accuracy of each, we can clear the relationship between nodes of the community and dig out the potential relationship between nodes. The time complexity is the standard to evaluate weather the community detection algorithm is good. Simultaneously, if the complexity is very high in a particular situation, the algorithm can not be used in engineering practice.

In this paper, we introduce a Modified Genetic Algorithm (MGA) to identify community structure. We have some novel ideals. Firstly, genetic algorithm has main two operators that include crossover operator and mutation operator [8]. As for its crossover operator, two alleles of two chromosomes exchange information. But here we only change one gene of the allele in order to protect good chromosome from destroying. Secondly, we use alleles encoding to assign each node with a community index. If two nodes have the same community index, they will have the same genes. Thirdly, we choose modularity function as its fitness function that to valuate weather each individual stands a good community detection results.

The rest of this paper is organized as follows. We will introduce modified GA algorithm in Section 2. In Section 3, MGA algorithm will be tested in both artificial random network and real networks. In the end, we will conclude MGA algorithm and discuss the future work in Section 4.

II. MODIFIED GENETIC ALGORITHM

Our algorithm is based on the modified genetic algorithm by half uniform crossover and alleles encoding. Specially, we choose modularity function as its fitness function, and use our algorithm to detect community structure.

A Half Uniform Crossover

In this subsection, half uniform crossover will be introduced. We suppose a chromosome $k_i(k_{i1},k_{i2},k_{i3},...,k_{in})$ has n genes and $2 \le i \le 2m$. $2m$ is the number of individuals. One of the chromosomes called k_{sour}, which wait to crossover. And the chromosome we want to get is called k_{dest}.

Here we will present alleles encoding. $k_i(1,3,5,5,3)$ shows that node 3 and 4 are in one community and node 2 and 5 are in other community. In order better to indicate half uniform crossover operator and alleles encoding, we give a simple example.

Table I HALF UNIFORM CROSSOVER WHEN k_{i1} IS SELECTED

n		k_{sour}		k_{dest} (before)		k_{dest} (after)
1	→	1	→	2	→	1
2		3		3		3
3		5		5		5
4		5		3		3
5		3		4		4

The half uniform crossover operator can only change the gene of k_{dest} chromosome.

B Modularity Function

We use $G(V,E)$ to present a complex network, where $V=\{v_i \mid i=1,2,...,N\}$ and $E=\{e_i \mid i=1,2,...,M\}$ are respectively represent the node set and the edge set of the network. More over, N and M represent the nodes number and the edges number, individually.

To quantify the results of community partition, we use the network modularity function Q proposed by S Fortunato, M Barthélemy [9].

$$Q=\sum_{v=1}^{C}\left[\frac{l_v}{L}-\left(\frac{d_v}{2L}\right)^2\right] \quad (1)$$

where l_v and d_v are respectively stand the number of edges within a community and the degree of all nodes within a community; L and C are respectively stand overall edges of the complex network and the community number of the complex network.

III. Experiment Results

The real and the artificial random networks are used to test the performance of the algorithm in this section. Therefore, the algorithm will be compared with some classic algorithms so that the effectiveness of our algorithm can be demonstrated. First, we will choose the appropriate parameters. Second, INBBO algorithm will be used to detect real networks, and be used to compare with other algorithm. Third, the computer-generated network is chosen to test INBBO algorithm.

A GA for real networks

To evaluate the effect of MGA algorithm, apply it to four real networks: Zachary's karate club network, Dolphin sociality network, Books on US politics network, and American college football network. We use four methods to calculate their modularity functions, including the GN algorithm [3]; the fast Newman algorithm proposed by Newman in 2004 [4]; Dong-Xiao He et al presented Clustering Combination Based Genetic Algorithm, (CCGA)[10]; Jin D et al proposed a genetic algorithm with local search (LGA)[11]. The results are shown in table 2

Table II Modularity function value of four real networks

Network name	GN	Fast Newman	CCGA[10]	LGA[11]	MGA
Zachary's karate club	0.4013	0.3807	0.4198	0.4198	0.4198
Dolphin sociality	0.4706	0.4955	0.5273	0.5280	0.5280
Books on US Politics	0.5168	0.4993	0.4445	0.5272	0.5260
American college Football	0.5996	0.5647	0.6054	0.6046	0.6054

Furthermore, to better understand our algorithm, we will draw the topological structure of the Zachary's karate club network in the next.

Zachary's karate club network

In this subsection, we will introduce the Zachary's karate club network. The network is a classic data set in the field of social network analysis. In the early 1970s, sociologist Zachary spent two years observing the social relationships among the 34 members of a US university karate club. Based on the interaction between the members inside and outside the club, he constructed a social network of members. The network contains 34 nodes, and an edge between the two nodes means that the corresponding two members have the frequent contact.

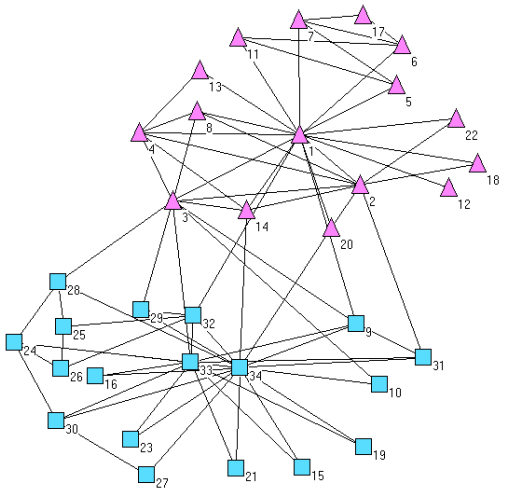

Fig.1. Community partition result of Zachary's karate club network using MGA

In figure 1, the community topological of Zachary's karate club network is consistent with the true situation. Because of the randomness of the algorithm, the result is not always available. In most cases, node 10 is not correctly partitioned.

This is because the modularity is not the maximum when node 10 is correctly partitioned.

B MGA for virtual network

In this subsection the computer-generated network is introduced to test the performance of MGA algorithm. The computer-generated network, named virtual computer network that has 4 communities and 128 nodes. And each node has 16 edges and meets the relationship $Z_{in} + Z_{out} = 16$. In this paper we will choose Zout = 6 and Zout =7 to test MGA algorithm.

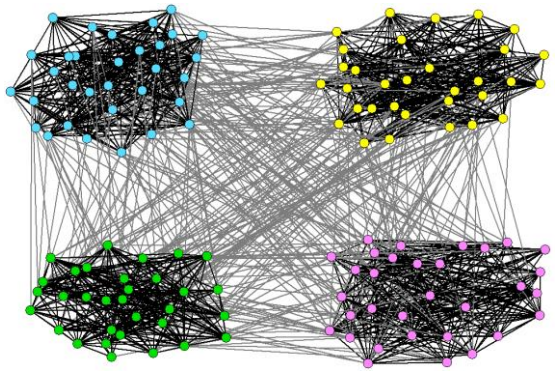

Fig.2. Community partition result of computer-generated network using MGA, Zout = 6.

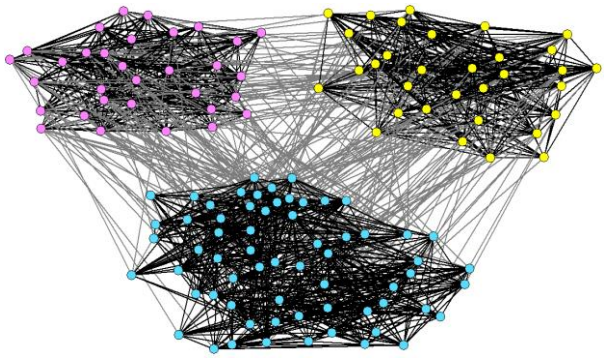

Fig.3. Community partition result of computer-generated network using MGA, Zout = 7.

IV. CONCLUSION

In this section, we make conclusions about our proposed algorithm. Community detection is very important for data processing and analysis. But the general algorithms are inferior in the quality and the speed when they deal with large scale networks. This paper we proposed a Modified Genetic Algorithm that with alleles encoding and half uniform crossover to detect community structure. Furthermore, real and artificial random networks are used to evaluate the effectiveness of MGA. From the experiment results we can know that our algorithm can get comparable results. Especially, the modularity function value can get very high point. Experimental results show that our algorithm is effective. The future work is to use our improved algorithm for other engineering optimization areas.

REFERENCES

[1] Hall, K. M. (1970). An r-dimensional quadratic placement algorithm. Management Science, 17(3), 219-229.

[2] Reichardt J, Bornholdt S. Detecting fuzzy community structures in complex networks with a Potts model.[J]. Physical Review Letters, 2004, 93(21):218701.

[3] Girvan M, Newman M E J. Community structure in social and biological networks[J]. Proceedings of the National Academy of Sciences of the United States of America, 2002, 99(12):7821-7826.

[4] Newman M E J. Detecting community structure in networks[J]. The European Physical Journal B, 2004, 38(2):321-330.

[5] Pizzuti C. GA-Net: A Genetic Algorithm for Community Detection in Social Networks[C]// Parallel Problem Solving From Nature - PPSN X, International Conference Dortmund, Germany, September 13-17, 2008, Proceedings. 2008:1081-1090.

[6] Shi Z, Liu Y, Liang J. PSO-Based Community Detection in Complex Networks[C]// Second International Symposium on Knowledge Acquisition and Modeling. IEEE Computer Society, 2009:114-119.

[7] Hassan E A, Hafez A I, Hassanien A E, et al. Community Detection Algorithm Based on Artificial Fish Swarm Optimization[C]// Intelligent Systems. 2015:509-521.

[8] Goldberg D. Genetic algorithms in search[J]. 1989.

[9] Barthelemy M, Fortunato S. Resolution limit in community detection[J]. Proceedings of the National Academy of Sciences of the United States of America, 2007, 104(1):36-41.

[10] He D X, Xu Z, Zuo W, et al. Community Mining in Complex Networks--Clustering Combination Based Genetic Algorithm[J]. Acta Automatica Sinica, 2010, 36(8):1160-1170.

[11] Jin D, Liu J, Yang B, et al. Genetic algorithm with local search for community detection in large-scale complex networks[J]. Acta Automatica Sinica, 2011, 37(7):873-882