A Biogeography-Based Optimization 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*— A new Biogeography-based Optimization (BBO) algorithm for detecting community in complex network science has been proposed. It adopts integral matrix encoding that each element represents the community index of the corresponding node. The migration operator and the mutation operator are the fundamental process to improve the accuracy and the quality of community detection. Each individual in the habitat can exchange information with other individuals by immigrating or emigrating. And the exchange among individuals will change the integral matrix value that closely connects with community index. In this paper, network modularity function is chosen as its suitability function, which to measure the quality of the community structure. The quality and effectiveness of the BBO algorithm are exposed in experimental tests by using artificial random network and real networks. The accuracy of BBO algorithm is batter to that of some classic algorithms, and is comparable to that of some latest algorithms.**

Keywords—Complex networks; Community detection; Biogeography-based Optimization; Integral matrix

I. INTRODUCTION

Community structure detection is burning in complex networks studies and attracts a large number of experts and scholars at home and abroad to carry out research. In complex networks, if the group has a relatively large similarity that the nodes connect closely with each other, the complex networks have a strong community feature [1, 2]. Also, a number of new algorithms are introduced to detect community structure in complex networks. Among those algorithms, the well-known algorithms include GN algorithm[3] proposed by M Girvan and MEJ Newman; the Fast Newman algorithm emerged in 2004 [4]; the GA algorithm is also used to community detection [5]; mapping entropy and evolutionary games are used to community structure detection[6,7].

However, there are several major problems with today's work. First, although the research of community detection algorithm has achieved very outstanding results, the accuracy and complexity of the algorithm still need to be improved. Second, traditional algorithms, such as Fast Newman and GA, are inaccurate and time-consuming, respectively. Third, many algorithms are only used for few networks. Therefore, we introduce a new community detection algorithm to improve accuracy and reduce time complexity.

This paper contributes to the literature in three main aspects. First, we introduces the integral encoding BBO (INBBO) algorithm, which uses the integral encoding scheme to better reflect the community index of each node. Currently, string encoding [8] and graph-based encoding[9] are widely used in the research of community detection. The former probably does not admit traditional migration operator and mutation operator while the latter requires additional decoding. Second, the BBO algorithm is used to detect hierarchical networks. In some complex networks, learning the network hierarchy is more conducive to understanding the internal links between network nodes. Third, each suitability index variable (SIV) [10] of the habitat stands for the community index of the corresponding node.

The rest of this paper is organized as follows. In Section 2, we describe the proposed model in detail, including INBBO algorithm process, INBBO algorithm combined with complex networks. Experimental results and experimental phenomena will be drawn and discussed detailed in Section 3. Finally, we will conclude this study and discuss the future work in Section 4.

II. MODEL

Dan Simon came up with BBO algorithm in 2008 [10]. After that, it has been used to multi-dimensional, time-varying and nonlinear systems, in case time-consuming of traditional algorithm. Simultaneously, the migration operator and the mutation operator are the basic operator of the algorithm. Besides the proposed merits, the BBO algorithm is different from GA[11] algorithms that require the generation of offspring, and only to change the original individual characteristics to achieve the purpose of optimization.

The Biogeography-based Optimization (BBO) model is illuminated by a set of isolated habitats, which can migrate between habitats to find a better living environment [12]. A set of vectors called Suitability Index Variables (SIV) is used to describe the characteristics of each island. In order to describe the level of habitat suitability, a Habitat Suitability Index (HSI) that is directly related to SIV, has been proposed.

A Migration Operator of INBBO Algorithm

Migration is an important means of exchanging habitat information in INBBO algorithm. Migration operators can exchange information of different habitats. Consequently, it

This paper has supplementary downloadable material available at http://ieeexplore.ieee.org, provided by the authors.
This includes one video of the presentation in WMV format.
This material is 129 MB in size.

can change the community index of nodes when each SIV value stands for one node's community index. Immigration operator and emigration operator are the basic operation of the migration operator. Both of them meet the linear law, exponential law and cosine law. Here we use linear model to reveal the migration relationship as the following shows.

$$\lambda = I_{max}\left(1-\frac{S}{S_{max}}\right) \quad (1)$$

$$\mu = E_{max}\bullet\frac{S}{S_{max}} \quad (2)$$

where λ and μ denote the immigration rate and the emigration rate respectively. S is considered the number of species; S_{max} is the maximum species count; I_{max} stands for the maximum immigration rate; E_{max} represents the maximum emigration rate.

B Mutation operator of INBBO algorithm

Mutation is an important supplement of changing habitat information in INBBO algorithm. Mutation operator randomly change the SIV value of each habitat based on habitat species existence probability. More specifically, the higher the species existence probability in one habitat is, the larger mutation rate becomes. Generally speaking, the maximum mutation rate is so small that mutation operator can hardly bring serious damage to all solution set. Especially, the reserved elite individuals will recover the habitat when the serious damage broke. Specially, the proportion of elit*e individuals in all individuals is* ρ *. Form the above discussion,* the mutation rate and species survival probability can be exposed as follow:

$$M(X_i) = M_{max}\left(\frac{1-P_s}{P_{max}}\right) \quad (3)$$

where $M(X_i)$ is the mutation rate of the ith habitat. M_{max} is the maximum mutation rate; P_{max} is the maximum species survival probability.

C Incorporation of INBBO Algorithm and Complex Networks

Previous evolutionary algorithms mainly combine the topology of the network and the distribution of the nodes. Due to the limitation of migration mechanism of INBBO algorithm, only the community information of the nodes can be stored in the habitat. At the same time, the modularity function is used to measure the community features. In this paper, we do some special work to well combine INBBO algorithm with complex networks. The most important one is that we initialize habitats with integral encoding. But traditional algorithms chose string encoding, graph-based encoding and binary encoding. Last but not the least, the modularity function is chosen as the suitability function in INBBO algorithm.

D Modularity function

Let $G(V,E)$ [13] stands for a complex network, where V and E represent the node set and the edge set of the network, respectively. $V=\{v_i \mid i=1,2,\ldots,N\}$ and $E=\{e_i \mid i=1,2,\ldots,M\}$, where N and M represent the nodes number and the edges number, respectively.

To quantify the results of community partition, we use the network modularity function Q:

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

where l_v and d_v represent the number of the edges included in one community and the number of the total degree included in one community. L represents the total edges that included in the complex network; C is the community number of the network. In this community detection algorithm, modularity function is chosen as the suitability function of INBBO algorithm.

E Habitat Initialization

In order to initialize habitat, an integral matrix is made. Supposing the environment has H habitats, and each habitat has N SIV value. And here each SIV value can stand one node community index. So there will be N nodes in one habitat. Simultaneously, a habitat can represent a community detection results. So species habitat represented by a matrix $X_{ij},(1\le i\le H, 1\le j\le N)$, which represents the community index of jth nodes in habitat i.

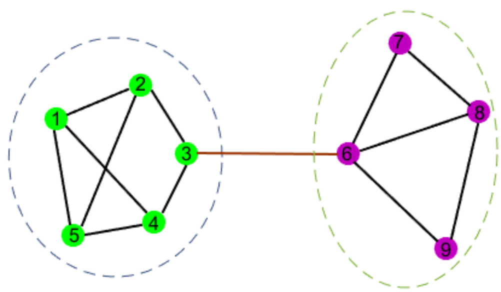

Fig. 1. A artificial network with two communities.

For example, an artificial network in Fig. 1 with two communities can be encoded as $[1,1,1,1,1,2,2,2,2]$, which means node 1, 2, 3, 4, 5 in the same community and node 6, 7, 8, 9 in the same community. The network shown in Fig. 1 represents a habitat that is also called individual.

III. Experimental Analysis

A Zachary's karate club network

The network represents the social relationships of karate clubs in an American university. Its nodes represent club members, and the edges indicate the social communication in the club. The club splits into two clubs due to internal divergence. Then two clubs are respectively leaded by the coach and the director. Fig. 2 is the topological structure of Zachary's karate club network.

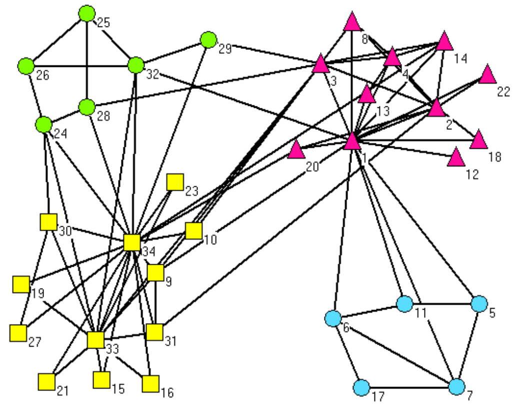

Fig. 2. Community partition result of Zachary's karate club network using INBBO

In Fig. 2, the Zachary's karate club network has been divided to four communities by INBBO algorithm. The four communities can reveal the connections within the network. Running INBBO algorithm 50 times, the average modularity function value is 0.4198, which is the best result. What's more, our algorithm has ability to identify the overlapping networks.

B Dolphin sociality network

The Dolphin social network was inspired by 62 bottlenose dolphins living habits in New Zealand. They found that these dolphins' contact exhibit a specific pattern, and constructed social networks of comprising 62 nodes.

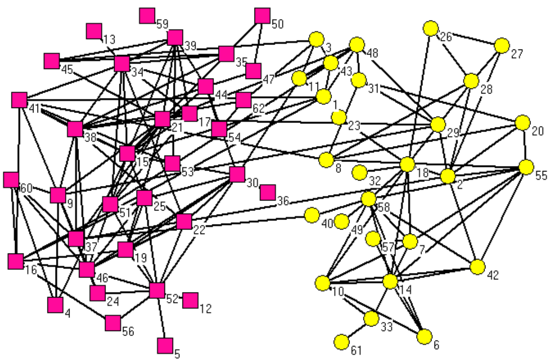

Fig. 3. Community partition result of Dolphin social network using INBBO

In Fig. 3, there is a community topology graph. In the figure, two colors represent the community divided by INBBO. Running INBBO algorithm 50 times, the average modularity function value is 0.5242, which is the better than GN and Fast Newman algorithm.

IV. CONCLUSION

Community detection algorithms have significant impacts on complex networks study. Our results show that the INBBO algorithm with unique encoding has great advantage for detecting complex networks. First, the algorithm with integral encoding can convert complex network problem to optimization problem. Second, the optimization quality is improved. Even some hierarchy networks can be detected. Third, real networks can be detected by INBBO algorithm. Unexpectedly, the maximum modularity value dose not always corresponding the real results. In the next study, we will focus on how to get rid of the limitation of modularity function.

REFERENCES

[1] Rui Xu, X.H., and Donald C. Wunsch II. 2004. Inference of Genetic Regulatory Networks with Recurrent Neural Network Models. Proceedings of the 26th Annual International Conference of the IEEE EMBS 2004.

[2] Clauset, A., M.E. Newman, and C. Moore. 2004. Finding community structure in very large networks. Phys Rev E Stat Nonlin Soft Matter Phy. 70(2004), 066111.

[3] Newman, M.G.M. 2001. Community structure in social and biological networks. Proceedings of the National Academy of Sciences. 99,12(Dec. 2001), 7821-7826.

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

[5] Li, Y., G. Liu, and S.-y. Lao. 2013. A genetic algorithm for community detection in complex networks. Journal of Central South University. 20, 5(Oct. 2013), 1269-1276.

[6] Nie, T., et al. 2016. Using mapping entropy to identify node centrality in complex networks. Physica A: Statistical Mechanics and its Applications. 453(Feb. 2016), 290-297.

[7] Zhen Wang, L.W., Attila Szolnoki. 2015. Evolutionary games on multilayer networks: a colloquium. Physics of Condensed Matter. 88, 5(May. 2015), 1-15.

[8] Liu, D.J.J.L.B.Y.H.D.-X.D. 2011. Genetic Algorithm with Local Search for Community Detection in Large-scale Complex Networks. Acta Automatica Sinica. 37, 7(Jul. 2011), 873-882.

[9] Pizzuti, C. 2009. A Multi-objective Genetic Algorithm for Community Detection in Networks. IEEE Computer Society. (2009), 379-386.

[10] Simon, D. 2008. Biogeography-Based Optimization. IEEE Transactions on Evolutionary Computation. 12, 6(Dec. 2008), 702-713.

[11] Meyarivan, K.D.A.P.S.A.T. 2002. A fast and elitist multiobjective genetic algorithm: NSGA-II. IEEE Transactions on Evolutionary Computation. 6, 2 (Apr. 2002), 182 - 197.

[12] Chattopadhyay, A.B.P. 2010. Biogeography-Based Optimization for Different Economic Load Dispatch Problems.pdf. IEEE Transactions on Power Systems. 25, 2(May. 2010), 1064-1077.

[13] M. E. J. Newman, M.G. 2004. Finding and evaluating community structure in networks. Physical Review E Statistical Nonlinear & Soft. 69, 2 (Aug. 2004), 026113-026113.