Continuous Encoding for Community Detection in Complex Networks

Wei Zheng, Yiqing Zhang, Jianyong Sun*
weizheng@stu.xjtu.edu.cn,zhangyiqing@stu.xjtu.edu.cn,jy.sun@xjtu.edu.cn
School of Mathematics and Statistics, Xi'an Jiaotong University
Xi'an, Shannxi, China

ABSTRACT

In this paper, a novel continuous encoding method developed for multiobjective evolutionary algorithm to solve the community detection problem in complex networks is proposed. Each edge in the considered network is associated with a continuous component of an individual's genotype. Through non-linear operations, each continuous-valued genotype is transformed into a solution to the community detection problem, i.e. a partition of the network nodes. This encoding method is embedded within the algorithmic framework of the multiobjective genetic algorithm for networks (MOGA-Net). Experimental results on synthetic and real-world networks demonstrate that the developed method can improve the performance of MOGA-Net significantly.

CCS CONCEPTS

• **Theory of computation** → Bio-inspired optimization.

KEYWORDS

continuous encoding, community detection, MOGA-Net

ACM Reference Format:

Wei Zheng, Yiqing Zhang, Jianyong Sun*. 2021. Continuous Encoding for Community Detection in Complex Networks. In *2021 Genetic and Evolutionary Computation Conference Companion (GECCO '21 Companion), July 10–14, 2021, Lille, France.* ACM, New York, NY, USA, 2 pages. https://doi.org/10.1145/3449726.3459409

1 INTRODUCTION

In real life, many systems can be modeled as complex networks, such as World Wide Web, which can be represented as a set of vertices (nodes) and a set of edges connecting with these nodes. Community detection in complex networks is a meaningful but difficult task in the data mining research area [4].

The purpose of community detection in complex networks is to find a set of nodes in which edges in communities are as dense as possible while edges between communities are as sparse as possible. In this paper, we consider disjoint community detection in undirected complex networks. That is, there are no nodes which belongs to more than one community.

By defining metrics which can measure the quality of the set of communities, the community detection problem can be modeled as a single/multiobjective optimization problem. As a promising paradigm of solving optimization problems, (multiobjective) evolutionary algorithms have been used to solve these problems since last decade [7], which has become a popular research avenue.

In [6], Pizzuti modeled the problems as discrete multiobjective optimization problems (MOPs) and proposed MOGA-Net based on Non-dominated Sorting Genetic Algorithm-II (NSGA-II) [2]. Specially, MOGA-Net takes locus-based encoding method to represent individual solutions and two discrete evolutionary operators are used to generate offsprings. In their paper, they proposed two objectives, including the community score and the sum of the community fitness of each detected community.

In this paper, we propose a new encoding method to encode individuals in MOGA-Net. Combined with MOGA-Net, continuous encoding MOGA-Net, dubbed as CEMOGA-Net, is proposed. Different from locus-based encoding method, our encoding method is continuous based rather than discrete. That is, individuals' genotype are continuous. By performing certain non-linear operations, each individual can be transferred to a network partition solution (i.e. a set of communities). Except these, CEMOGA-Net uses the same objective functions and optimization framework (i.e., NSGA-II) as MOGA-Net. In the experiments, CEMOGA-Net and MOGA-Net are tested on widely used synthetic and real-world networks. The results show that our new encoding method is very effective for improving the performance of MOGA-Net.

2 CONTINUOUS ENCODING METHOD

Assume that $\mathbf{x}$ is a continuous valued genotype in a population P. Here, $\mathbf{x} = [\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n] \in [0,1]^L$ where n and L stand for the number of nodes and edges in a network, respectively. The sub-vector $\mathbf{x}_i$ is the encoding for node V_i. The length of $\mathbf{x}_i$ is equal to the number of edges linking with node V_i and the set of linking nodes is defined as $\mathcal{N}(V_i)$.

Our encoding method is performed as follows. At first, an edge set $\mathcal{E}$ is set to empty. For each node V_i in the network, we perform the sigmoid function on its corresponding sub-vector $\mathbf{x}_i$ element by element. The softmax function and argmax function are then operated step by step to select a node V_{p_i} within $\mathcal{N}(V_i)$. The node V_{p_i} is considered as in the same community as node V_i. The node pair (V_i, V_{p_i}) will be stored in $\mathcal{E}$.

The above non-linear operations are performed for all nodes in the network one by one. In the end, a discrete permutation solution will be obtained, which is a locus-based encoding solution, i.e. a node's genotype is one of its connected nodes. After decoding, a discrete network partition solution will be obtained.

GECCO '21 Companion, July 10–14, 2021, Lille, France

ACM ISBN 978-1-4503-8351-6/21/07.
https://doi.org/10.1145/3449726.3459409

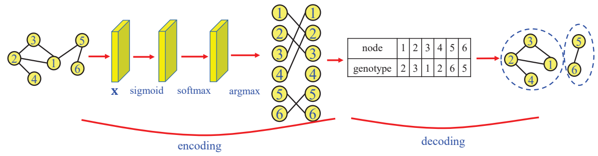

Figure 1: The encoding and decoding processes of an individual x on a network.

The encoding and decoding processes of an individual $\mathbf{x}$ on a simple network is visualized in Fig. 1. As shown in the figure, sigmoid, softmax and argmax operations are performed on individual $\mathbf{x}$ step by step. Consequently, a locus-based encoding solution is obtained. After decoding, two communities $\{1, 2, 3, 4\}$ and $\{5, 6\}$ are obtained as the result.

In CEMOGA-Net, the objective functions are the same as in MOGA-Net, including the community score and the sum of community fitness of each detected community. Evolutionary operators: differential evolution (DE) and polynomial mutation used in [3] are employed directly due to the convenience of continuous encoding.

Note that our encoding method implements the transformation of continuous value-based individual and discrete partition solution of complex network. To the best of our knowledge, our work is the first continuous transformation method in this research area. Moreover, this encoding method is easy to be embedded in any continous evolutionary algorithm based method.

3 EXPERIMENTAL STUDY

In this section, we present the experimental results obtained by performing CEMOGA-Net and MOGA-Net on synthetic and real-world networks. The population size of CEMOGA-Net is set to 100. The number of generations is 200. The parameters of DE are set as $F = 0.7$ and $CR = 0.5$. The mutation probability is set to 0.02 and distribution index of mutation is set to 20. For MOGA-Net, the parameters are set the same as in the original work [6].

The commonly used normalized mutual information (NMI) [1] metric is employed to evaluate the performance of the two MOEAs. The larger the NMI value, the better performance of a method. In the following tables, the mean and standard deviation NMI values summarized from the 30 runs obtained by the two algorithms are reported and the best results are marked in bold.

Ten different LFR [5][1] synthetic networks with 500 nodes are generated. Their main control parameter μ ranges from 0.1 to 1.0 with step size 0.1 and we mark them as LFR1 to LFR10 correspondingly. The NMI results for these LFR networks are shown in Table 1. We can find that CEMOGA-Net achieves higher NMIs than MOGA-Net, which reveals that the proposed encoding method can certainly enhance the performance of MOGA-Net.

Four real-world networks, including Zackary's Karate Club network (abbr. Karate), Bottlenose Dolphins network (abbr. Dolphins), American College Football network (abbr. Football) and Kreb's books on American politics network (abbr. Books), are also used to test CEMOGA-Net and MOGA-Net. The obtained results are summarized in Table 2. From the table, we see that the NMI obtained by

Table 1: The NMI values obtained by MOGA-Net and CEMOGA-Net on ten synthetic networks.

Methods	LFR1	LFR2	LFR3	LFR4	LFR5
MOGA-Net	0.782(0.016)	0.516(0.027)	0.448(0.023)	0.379(0.011)	0.324(0.022)
CEMOGA-Net	**0.793(0.012)**	**0.585(0.012)**	**0.534(0.011)**	**0.433(0.006)**	**0.375(0.008)**
Methods	LFR6	LFR7	LFR8	LFR9	LFR10
MOGA-Net	0.281(0.014)	0.245(0.013)	0.253(0.009)	0.239(0.005)	0.217(0.007)
CEMOGA-Net	**0.311(0.008)**	**0.262(0.011)**	**0.270(0.004)**	**0.250(0.009)**	**0.223(0.007)**

Table 2: The NMI values obtained by the compared algorithms on four real-world networks.

Methods	Karate	Dolphins	Football	Books
MOGA-Net	**1(0)**	**1(0)**	0.795(0.016)	0.597(0.014)
CEMOGA-Net	**1(0)**	**1(0)**	**0.822(0.014)**	**0.614(0.009)**

CEMOGA-Net are higher than MOGA-Net in general, though both methods have the same results of '1' on 'Karate' and 'Dolphins'. Moreover, in both tables, the standard deviation values obtained by CEMOGA-Net are roughly all smaller than those by MOGA-Net, which indicates that CEMOGA-Net is more stable.

4 CONCLUSION

In this paper, we proposed a novel continuous encoding method for the community detection problem in complex networks. With this encoding method, we realized the transformation from a discrete MOP to a continuous MOP. By embedding it in MOGA-Net, CEMOGA-Net was proposed. The experimental studies showed that CEMOGA-Net performs significantly better than MOGA-Net, which indicated that the continuous encoding method is effective.

ACKNOWLEDGMENTS

This work was supported by the National Key Research and Development Program of China (grant no. 2018YFC0809800), the National Natural Science Foundation of China (grant no. 11991023, 62076197), the Major Project of National Science Foundation of China (grant no. U1811461), Key Project of National Science Foundation of China (grant no. 11690011), the Project of National Science Foundation of China (grant no. 61721002) and the Fundamental Research Funds for the Central Universities (No.xzy022019074).

REFERENCES

[1] Leon Danon, Albert Díaz-Guilera, Jordi Duch, and Alex Arenas. 2005. Comparing community structure identification. *Journal of Statistical Mechanics: Theory and Experiment* 2005, 09 (sep 2005), P09008.
[2] K. Deb, A. Pratap, S. Agarwal, and T. Meyarivan. 2002. A fast and elitist multiobjective genetic algorithm: NSGA-II. *IEEE Transactions on Evolutionary Computation* 6, 2 (Apr 2002), 182–197.
[3] J. Sun et al. 2019. Learning From a Stream of Nonstationary and Dependent Data in Multiobjective Evolutionary Optimization. *IEEE Transactions on Evolutionary Computation* 23, 4 (Aug 2019), 541–555.
[4] Santo Fortunato. 2010. Community detection in graphs. *Physics Reports* 486, 3 (2010), 75–174.
[5] Andrea Lancichinetti, Santo Fortunato, and Filippo Radicchi. 2008. Benchmark graphs for testing community detection algorithms. *Phys. Rev. E* 78 (Oct 2008), 046110. Issue 4.
[6] C. Pizzuti. 2012. A Multiobjective Genetic Algorithm to Find Communities in Complex Networks. *IEEE Transactions on Evolutionary Computation* 16, 3 (June 2012), 418–430.
[7] C. Pizzuti. 2018. Evolutionary Computation for Community Detection in Networks: A Review. *IEEE Transactions on Evolutionary Computation* 22, 3 (2018), 464–483.

[1] LFR is a network generator proposed by Lancichinetti, Fortunato and Radicchi in [5] which has been widely used for complex network studies.