

# Local Heuristic Genetic Algorithm for Discovering Community of Complex Networks

Yuquan Guo<sup>1,a</sup>, Xiongfei Li<sup>1,b</sup>

<sup>1</sup>College of Computer Science and Technology, Jilin University, Changchun, China

<sup>a</sup>guoyuquan@sohu.com, <sup>b</sup>xiongfei@jlu.edu.cn

**Keywords:**Complex network;Heuristic Function; Community Structure; Genetic Algorithm

**Abstract.**Community of complex networks has attracted much attention. For the problem, we propose a local heuristic genetic algorithm (LHGA). Mutation of LHGA is combined with local heuristic function from the perspective of networks dynamics and makes convergence rate of this algorithm expressively improved. Extensive tests on artificial networks and real world networks give excellent results.

## Introduction

In recent years, complex networks have attracted much attention. Many complex systems can be represented in term of networks. For example, the Web is a network of HTML pages interconnected by hyperlinks. Internet is described by service and client as nodes, and the connection between them by TCP/IP. A community is thought of as a set of nodes where nodes are densely interconnected and sparsely linked to other parts of the networks. Discovering community structures in complex networks becomes focus in many research fields. Finding community structure of complex network is important to understand function of complex systems they represent.

In the paper, we propose a novel algorithm to detect community which is local heuristic genetic algorithm (LHGA). The remainder of this paper is organized as follows. In the section 2, related work is introduced. LHGA is presented in section 3. In section 4, we give experimental result. In section 5, we give our conclusions.

## Related work

In recent years, many algorithms<sup>[1-3]</sup> have been proposed to detect communities. For reviews and comparison, one can refer to literature<sup>[4]</sup>. Hierarchical clustering<sup>[5]</sup> is an important approach to detect community structure. Hierarchical clustering algorithms create a dendrogram according the similarity between two groups of nodes, then uncover the community structure through cutting the dendrogram into sub-dendrogram. The modularity optimization algorithms, which uncover the community structure through optimizing a partition of networks according to modularity function<sup>[6]</sup>, are the most popular class methods. As maximizing the modularity has been proven to be a NP-complete problem, the modularity methods are also approximation algorithms. However, some authors have proved that modularity has a resolution limit and the problem raises attention about the precision of the communities uncovered through the optimization of modularity. To address the problem, the conductance of networks was proposed by SHEN-Huawei et al, which overcomes the resolution limit of modularity. The conductance of networks overcomes the resolution limit of modularity. The local heuristic genetic algorithm (LHGA) which employs conductance  $C$  as object function is proposed in this paper. The major innovation in this paper is a mutation method which is based on a local heuristic function. The mutation operation makes convergence rate of this algorithm expressively improved.

## Local heuristic genetic algorithm

A complex network can be defined as a graph  $G=(V,E)$ , where  $V$  is the set of vertices, and  $E$  is the set of edges connecting pairs of vertices. Uncovering community structure is to find the division of network nodes into groups. An unweighted and undirected network with  $n$  nodes is described in terms

of its adjacency matrix  $A$  whose elements  $A_{xy}$  denote the edge connecting the nodes  $x$  and  $y$ . The internal degree volume of community is defined as

$$\text{in\_degree}(V_i) = \sum_{x \in V_i, y \in V_i} A_{xy} \quad (1)$$

The degree volume of community is defined as

$$\text{degree}(V_i) = \sum_{x \in V_i, y} A_{xy} \quad (2)$$

The conductance is denoted by

$$C(P) = \frac{1}{k} \sum_{i=1}^k \left( 1 - \frac{\text{in\_degree}(V_i)}{\text{degree}(V_i)} \right) \quad (3)$$

Where  $P$  is a partition of networks,  $V_i$  is a community of networks,  $k$  is the number of communities of networks, and  $P = \{V_1, \dots, V_k\}$ .  $C(P)$  is object function of LHGA.

LHGA is composed of genetic representation, population initialization, crossover operation and mutation operation. The genetic representation is based on node. An individual  $I$  of the population consists of  $n$  genes  $I = \{g_0, g_1, \dots, g_i, \dots, g_{n-1}\}$ , where  $i \in \{0, \dots, n-1\}$  is the identifier of nodes,  $n$  is the number of nodes, and each  $g_i$  can take one of the nodes of node  $i$ . If  $g_i = g_j$ , node  $i$  and  $j$  are belong same community. Population initialization is based LPA. Initially, for all chromosomes, each vertex is put in a different community. Then we run LPA<sup>[7]</sup>  $m$  times which make individual gets a community structure. After population initialization, a set of individual is created. Crossover operation is an important step of genetic algorithm, and plays a key part in global searching. Many crossover methods are proposed. One-way crossing<sup>[8]</sup> is adopted by LHGA. A example of one-way crossing is given in table 1.

Table 1 One-way crossing over when node 1 is selected

node	Source chromosome	Destination chromosome(before)	Destination chromosome(after)
1	6	4	6
2	3	5	5
3	3	7	7
4	6	2	6
5	6	1	6
6	7	5	5

Mutation operation is embedded into genetic algorithm to reinforce its ability of local search. We adopt the local heuristic function to realize mutation. The local heuristic function  $h$  of mutation is defined as Eq.(4)

$$h(n, l) = \frac{1}{2} \left( \frac{\text{in\_degree}(C^N(n))}{\text{degree}(C^N(n))} + \frac{\text{in\_degree}(C^L(l))}{\text{degree}(C^L(l))} \right) \quad (4)$$

where  $n$  is a node, and  $C^N(n)$  is a community which  $n$  is belong to, and  $C^L(l)$  is a community in which all nodes are with gene  $l$ . A node is randomly picked to be mutated. The node selects a max value of  $h$  in its neighbor genes.

## Experimental result and discussions

In this section, LHGA algorithm is tested on both artificial networks and real networks. The first test focuses on whether the number of communities can be detected in artificial networks. We apply LHGA to the LFR benchmark<sup>[9]</sup> which is a networks with various distributions of vertex degree and community size. There are many parameters to control the created networks: the numbers of nodes  $N$ , the average vertex degree  $k$ , the maximum vertex degree  $\text{max\_}k$ , the mixing ratio, the exponent of the power law distribution of vertexes, the exponent of the power law distribution of community size, the minimum community size  $\text{min\_}c$ , and the maximum community size  $\text{max\_}c$ . In our experiment, the default configuration of LFR benchmark:  $N=1000$ ,  $\langle k \rangle=15$ ,  $\text{max\_}k=50$ ,  $\text{min\_}c=20$ , and  $\text{max\_}c=50$ . We use the recall and precision to evaluate the difference between the community structure uncovered by LHGA method and the true community structure.

The value of recall is between 0 and 1. A larger recall indicates that the detected communities more perfectly match the true communities. In the experiment of artificial networks, we compare several methods, i.e. LPA, COPRA<sup>[10]</sup> and EAGALE<sup>[5]</sup>. We apply LHGA 50 times to each of the artificial networks, and calculate the average recall and precision. Fig.2(a) portrays the recall of the networks of different mixing ratio  $\mu$ , and Fig. 2(b) shows the precision. As shown in Fig.2, when the mix ratio is smaller than 0.3, the community structure of network is clear. The methods compared give better results. LHGA gives the best result. The communities become sparse when the mix ratio is up to 0.5. LHGA exhibit better performance than other methods.

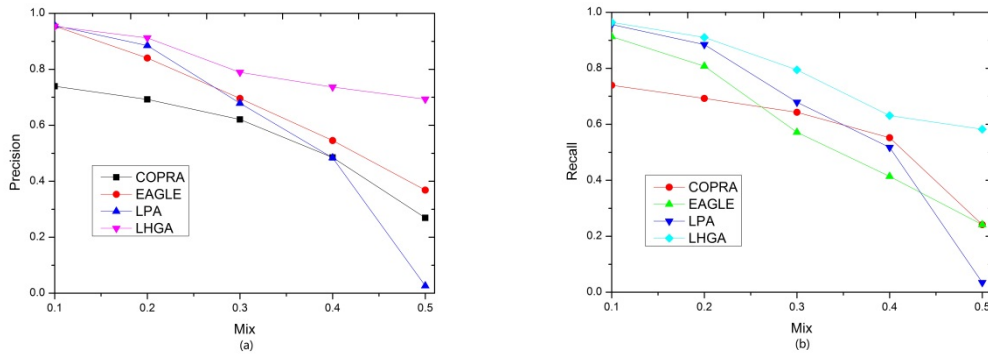


Fig.2 Recall rate and Precision of the networks of different mixing ratio  $\mu$

The second test turns to real networks. We select the network of karate club<sup>[11]</sup> and dolphin<sup>[12]</sup>. Fig.3 shows dividing plan for dolphin networks. As shown in Fig.4, Karate is split into two communities.

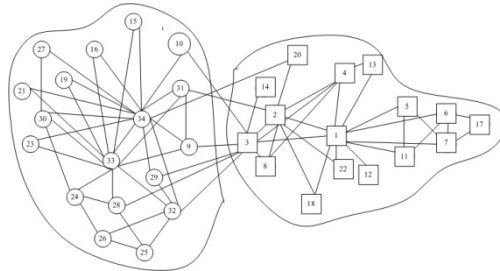


Fig.3 community structure of Karate

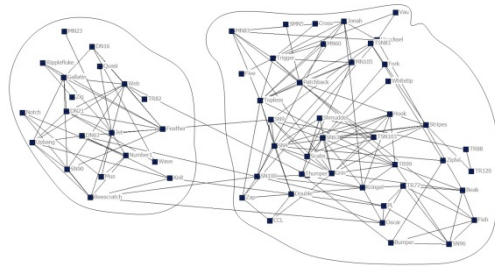


Fig.4 community structure of dolphin

## Conclusion

In this paper, we present an algorithm LHGA to detect community structure in complex networks. In LHGA, a heuristic update strategy, which creates connection between local search and globe object, is proposed to detect community structure. Experiments show that LHGA is effective in term of conductance of the uncovered communities.

## Acknowledgements

This work was financially supported by the Technology Development Program of Jilin Province (No. 20100508 and No. 201105017).

## References

- [1] ORMAN G K, LABATUT V, CHERIFI H. Comparative evaluation of community detection algorithms: a topological approach [J]. Journal of Statistical Mechanics: Theory and Experiment, 2012, 2012(08): P08001.
- [2] FORTUNATO S, CASTELLANO C. Community structure in graphs [M]. Computational Complexity. Springer. 2012: 490-512.
- [3] DANON L, DIAZ-GUILERA A, DUCH J, et al. Comparing community structure identification [J]. Journal of Statistical Mechanics-Theory and Experiment, 2005,
- [4] YANG B, LIU D-Y, LIU J, et al. Complex network clustering algorithms [J]. Journal of Software, 2009, 20(1): 54-66.
- [5] SHEN H W, CHENG X Q, CAI K, et al. Detect overlapping and hierarchical community structure in networks [J]. Physica a-Statistical Mechanics and Its Applications, 2009, 388(8): 1706-12.
- [6] NEWMAN M E J. Detecting community structure in networks [J]. European Physical Journal B, 2004, 38(2): 321-30.
- [7] RAGHAVAN U N, ALBERT R, KUMARA S. Near linear time algorithm to detect community structures in large-scale networks [J]. Physical Review E, 2007, 76(3): 036106.
- [8] TASGIN M, HERDAGDELEN A, BINGOL H. Community detection in complex networks using genetic algorithms [J]. arXiv preprint arXiv:07110491, 2007,
- [9] LANCICHINETTI A, FORTUNATO S, RADICCHI F. Benchmark graphs for testing community detection algorithms [J]. Physical Review E, 2008, 78(4):
- [10] GREGORY S. Finding overlapping communities in networks by label propagation [J]. New Journal of Physics, 2010, 12
- [11] ZACHARY W. An Information Flow Model for Conflict and Fission in Small Groups<sup>1</sup> [J]. Journal of anthropological research, 1977, 33(4): 452-73.
- [12] LUSSEAU D. The emergent properties of a dolphin social network [J]. Proceedings of the Royal Society of London Series B: Biological Sciences, 2003, 270(Suppl 2): S186-S8.