

Discrete Bat Algorithm and Application in Community Detection

Wang Chunyu* and Pan Yun

School of Science and Technology, Communication University of China, Beijing, 100024, P.R. China

Abstract: With the rapid development of social media, users have more options to form spontaneously reading or watching communities, and detecting such communities has significance both in technological aspects and commercial aspects. Theoretically community detecting problem is NP-hard problem and thus people inclined to choose heuristic algorithms based on objective optimization. Among these algorithms Bat-inspired Algorithm (BA) was originally proposed to solve continuous objective optimization problems. In this paper, authors explore the role of Bat-inspired Algorithm on detecting community structure in networks. This paper firstly introduces the definition of community and Bat-inspired Algorithm, then gives Discrete Bat Algorithm (hereinafter referred to as D-BA), the detailed design of detecting a classic community structure in the Karate club network and the standard of evaluation. Finally analyzes and evaluates the result.

Keywords: Bat-inspired algorithm, community detection, discrete bat algorithm, multi-objective optimization.

1. INTRODUCTION

At the beginning, the study of community structure in networks is a topic coming from the field of sociology research. Now it is gradually extended to physics, mathematics and computer science for studying the structural properties of complex systems. Back to the original meaning of community, people through the common interests or other relationships spontaneously formed communities. Members in the same community have a closer relationship and the relationships among members are always looser in different communities. As shown in Fig. (1), we can find the structure of communities intuitively, but in most cases we cannot easily find the structure of communities in complex networks (Fig. 2).

From the sociological perspective, the result obtained from mathematical formulas does not reflect the reality of the real composition of the community.

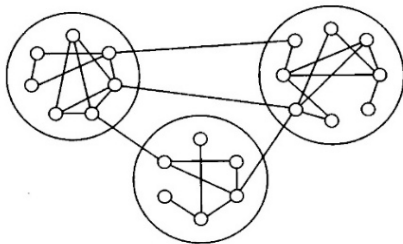


Fig. (1). The diagram of a network with structural community [1].

The result is only an instructive forecast, such as identifying terrorist organizations based Internet, or predicting and analyzing public opinions etc. From the computer science perspective, finding community structure in networks is

a graph partition problem and it is proved to be NP-hard. A typical example is how to allocate n procedures to g processors, so that the number of programs running on each processor is approximately equal, while minimizing the number of edges among the processors.

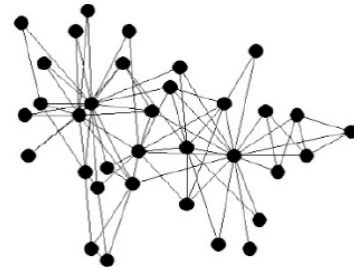


Fig. (2). The Karate club network.

With more and more researchers pay attention to the problem of detecting community structure, many algorithms have been designed. Document [2] classified these algorithms depending on the different community structures. Some researchers divided these algorithms into two categories depending on the applications of these algorithms: graphic segmentation algorithms in the field of computer science and hierarchical clustering algorithms in the field of sociology. The former includes the Kernighan-Lin algorithm, spectral bisection algorithm, factions filtering algorithm etc., while the latter includes splitting algorithms, aggregation algorithm etc. Addition to these two categories, there are other algorithms such as algorithms based on the largest modular of community, the algorithm used in this paper can be seen as the largest modular algorithm. In recent years, many new algorithms are still emerging, Girvan and Newman proposed a 'GN algorithm' [3] in 2003 which iteratively removes edges with the maximal betweenness of edge [4] from the network, thereby breaking it up in communities; Pons and Latapy proposed a 'random walk algorithm' [5] in

2005 which was based on the following intuition: random walks on a graph tend to get ‘trapped’ into densely connected parts corresponding to communities. Using different algorithms on the same network always get different results, for known community structure network, the divided results can be compared with the real community structure of the network get the accuracy of the algorithm. This paper uses the accuracy of division (*i.e.* the number of correct divided nodes divided by the number of total nodes) as the evaluation standard.

This paper presented a discrete variant of the swarm intelligence algorithm-Bat-inspired Algorithm (BA) proposed by Xin-She Yang [6]. In this paper, author adopts ordered adjacency list [7] as the encoded form, and modular Q function as the objective function. Then we take the rate of correct division as the standard of evaluation and simulate D-BA by using MATLAB R2013a. The analysis of simulation results show that the D-BA is effective and efficient in detecting communities in complex networks.

2. THE DEFINITION OF COMMUNITY

There are many definitions of the community from different viewing angles. Coscia M gives the basic definition of community [2]: the community is a collection of elements; within the collection, elements linked with outside elements by same or similar relationships.

The definition of community in this paper is based on the density of edge which means nodes in the same community have more edges than nodes in different communities.

A network can be represented mathematically by an adjacency matrix A , if there is an edge from v_i to v_j , then $A(i,j)=1$; otherwise $A(i,j)=0$. The degree of a node i : $k_i = \sum_j A_{ij}$. A network also can be represented by a graph G , if C is the sub-graph of G and $v_i \in C$ then $k_i(C) = k_i^{in}(C) + k_i^{out}(C)$ where $k_i^{in}(C) = \sum_{j \in C} A_{ij}$ is the number of edges connect the i to the other nodes in C , and $k_i^{out}(C) = \sum_{j \in -C} A_{ij}$ is the number of edges connecting i to the rest of the network. A sub-graph C is a community if $\sum_{i \in C} k_i^{in}(C) > \sum_{i \in C} k_i^{out}(C)$ [8].

3. BAT-INSPIRED ALGORITHM

Bat-inspired algorithm is an intelligent optimization algorithm, which imitate the behavior of bats based on the Genetic Algorithm, Particle Swarm Optimization algorithm, and Simulated Annealing algorithm. First, let's look at the characteristics of the bat, so that we can have a perceptual know about this algorithm.

3.1. The Characterizations of Bat

Bats are the only mammals with wings and they have advanced capability of echolocation they can emit biological wave to prey. The typical range of frequencies are in the region between 25kHz and 150kHz, the wavelength range are in the region between 2mm and 14mm. Typically, bats emit bio-wave at 10-20 times per minute, but when approaching prey, the bio-wave may be issued up to 200 times per minute. Loudness of the bio-wave which is emitted by

bats can reach 110dB, with the distance of the prey getting closer, the loudness gradually reduced.

3.2. Bat-Inspired Algorithm

Xin-She Yang proposed BA by imitating the feeding behavior of bats: bats emit biological waves to detect the position of the prey, thereby change the position to chase their prey. With the distance of the prey getting closer, the loudness gradually reduced and the frequency of emission get faster. The frequency of biological waves and the size of prey are related. Suppose at time t , x indicates the position of a bat, v represents velocity which means the degree of the change of position, the position update procedure of bat can be formulated as:

$$x_i^t = x_i^{t-1} + v_i^t \quad (1)$$

According to the characteristics of bats, bats adjust their flight speed according to the distance to prey, formulated as follows:

$$v_i^t = v_i^{t-1} + (x_i^t - x_*) * f_i \quad (2)$$

where f represents the frequency of biological waves, x_* represents the position of prey.

In order to imitate the ‘irrational’ behavior of bats, such as flight deviated from a known prey with the degree of deviation in the range of loudness, the formula has been proposed as follows:

$$x_{new} = x_{old} + \varepsilon * A^t \quad (3)$$

where ε is a random number between $[-1,1]$ obeys uniform distribution, A^t represents loudness, the loudness update formula is:

$$A_i^{t+1} = \alpha * A_i^t; A_i^t \rightarrow 0 \text{ as } t \rightarrow \infty \quad (4)$$

where α is an empirical value.

Bats judge the distance to the prey by changing the frequency of emission. The frequency of emission update formula is:

$$r_i^{t+1} = r_i^0 [1 - \exp(-\gamma * t)]; r_i^t \rightarrow r_i^0 \text{ as } t \rightarrow \infty \quad (5)$$

where γ is an empirical value.

4. THE DETAILED DESIGN

4.1. Defining the Objective Function

Bat-Inspired algorithm is belong to intelligent optimization algorithm which is based on optimizing the objective function, so defining a quality objective function is an important part of solving the problem. In this paper, we use a modular function Q [9, 10] as the objective function. The definition is as follows:

If the network is divided into n communities, then we defined a $(n*n)$ symmetric matrix F , where f_{ij} indicates the proportion of edges connecting the community C_i and the community C_j in all edges of the network. The trace of Matrix F :

$$\text{Trace}(F) = \sum_i f_{ii}, \quad (6)$$

Which indicates the proportion of edges in the same community in all edges of the network. Defining the sum of a row (or column) as

$$a_i = \sum_j f_{ij} \tag{7}$$

Which indicates the proportion of edges in the community C_i in all edges of the network. In a network in which edges fall between vertices without regard for the communities they belong to, we would have

$$f_{ij} = a_i a_j \tag{8}$$

$$Q = \sum_{i=1}^N (f_{ii} - a_i^2) = \text{Trace}(F) - |F^2| \tag{9}$$

where $|F^2|$ is the modular of matrix F^2 . According to this definition we can calculate the value of Q for each division of the network.

4.2. Encoded Form

Different forms of encoding have a great impact on solving a problem, a better encoded form can narrow the scope of the solutions and improve operational efficiency, so this paper adopts the ordered adjacency list [2] for the form of encoding. All nodes with ID of the network have shown in Fig. (3).

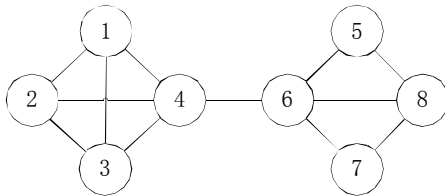


Fig. (3). Example network.

Distribution of nodes is shown in Table 1 below:

Table 1. Distribution of nodes in example network.

Nodes	Adjacent Nodes	The Number of Adjacent Nodes
1	2,3,4	3
2	1,3,4	3
3	1,2,4	3
4	1,2,3,6	4
5	6,8	2
6	4,5,7,8	4
7	6,8	2
8	5,6,7	3

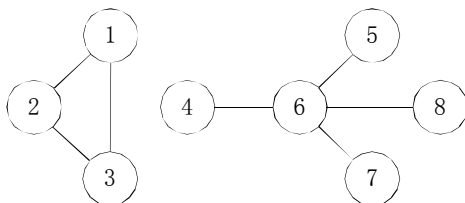


Fig. (4). Community structure example.

There are various kinds of division of the example network (Fig. 3), the community structure corresponding to the sequence {1,1,1,4,1,1,1,2} as shown in Fig. (4).

4.3. D-BA

Because we mainly use the principles of probability and statistics to detect community structures, which is different from the method to solve the problem for continuous mathematical model, so we have to adjust BA to meet the needs of detect community structures in complex network.

Velocity v represents the difference between a new position and an original position in the D-BA. The velocity is equal to zero, when the new position is equal to the original position; otherwise the velocity is equal to the new location. Formulated as follows:

$$v^t = (x_i^t - x_*) = \begin{cases} 0 & \text{if } (x_i^t = x_*) \\ x_i^t & \text{otherwise} \end{cases} \tag{10}$$

In the D-BA, the velocity can be updated by multiplying a frequency of biological wave f . The process is that generate a random number; if the frequency is not less than the random number, the updated velocity is equal to the original speed; otherwise the updated velocity is equal to zero. Formulated as follows:

$$v^{t+1} = v^t * f_i = \begin{cases} v^t & \text{if } \text{rand} \leq f_i \\ 0 & \text{otherwise} \end{cases} \tag{11}$$

In the D-BA, adding two velocities can be represented as a random choice according to equal probability. Formulated as follows:

$$v_i^t = v_i^{t-1} + v^{t+1} = \begin{cases} v_i^{t-1} & \text{rand} > 0.5 \\ v^{t+1} & \text{otherwise} \end{cases} \tag{12}$$

In the D-BA, the position can be updated by adding a velocity to an original position. If the velocity is zero, the updated position is equal to the original position; otherwise, the new position is equal to the velocity, formulated as:

$$x_i^t = x_i^{t-1} + v_i^t = \begin{cases} x_i^{t-1} & \text{if } v_i^t = 0 \\ v_i^t & \text{otherwise} \end{cases} \tag{13}$$

In the D-BA, ‘irrational’ behavior of bats is represented as randomly moving on any number (A^t) of dimensions in the space of position.

In the D-BA, α, γ are constant values between (0,1). By increasing the number of iterations, the loudness gradually becomes weaker and the frequency of emission is gradually increased (see (4) (5)).

4.4. Pseudo-Code Description

Initialize the position and velocity of bats x_i ($i=1, 2, \dots, N$) and v_i

Define frequency of bio-wave f_i

Initialize the frequency of emission r and the loudness A

While ($t < \text{max number of iterations}$)

Generate new solutions by adjusting frequency and update the velocities and position s

```

If (rand > ri)
Generate a local solution around the best solution
Endif
If (rand < Ai & f(xi) < f(x*))
Accept the new solutions, increase ri and reduce Ai
End if
Find the current best solution f(x*).
End while
    
```

5. SIMULATION AND ANALYSIS

5.1. Karate Club Network

Karate club network also known as Zachary network. Zachary had observed members of the karate club of an American university, and then obtained the relationship network [11]. This network later became a classic social network for testing algorithms of detecting community structure. Because of the different ideas on how to spending the tuition, the coach and the president of the club split ultimately this club into two small clubs, and they respectively became the core member of the two small clubs (Fig. 5). The realistic significance of researching this problem is that if we can know the results of split in advance, we will avoid such split or make the split toward to what we want.

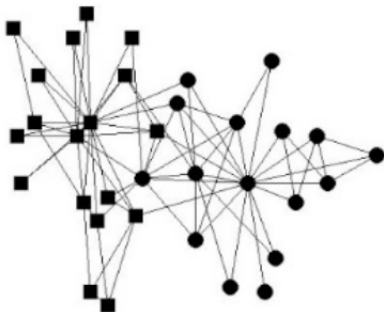


Fig. (5). The Karate club network actual division diagram (roundness and square represent different communities).

Table 2 shows the Distribution of nodes in the karate club network. We use the variable G to store adjacent nodes for each node in the network; the variable B to store the total number of adjacent nodes for each node in the network; the variable A to store the adjacency matrix; N-dimensional vector $x=(x_1, x_2, \dots, x_i \dots x_N)$ (N is the number of nodes in the network; $x_i \leq B_i$) stores the position of one bat, which means the 1st node is connected to the x_1^{st} node in the G{1}, the ist node is connected to the x_i^{st} node in the G{i}...each position of bat corresponds to a value of modular function Q, the optimal Q value is the best solution which we desire to search.

5.2. Experiment Results and Analysis

Using above method directly to calculate has a larger deviation from the actual division (Fig. 5). The objective function (modular function Q) is only the theoretical optimal, which does not reflect the actual network division. To get a higher reliable prediction, we need to do more granular re-

Table 2. Distribution of nodes in the karate club network.

Nodes	Adjacent Nodes	Total Adjacent Nodes
1	2,3,4,5,6,7,8,9,11,12,13,14,18,20,22,32	16
2	1,3,4,8,14,18,20,22,31	9
3	1,2,4,8,9,10,14,28,29,33	10
4	1,2,3,8,13,14	6
5	1,7,11	3
6	1,7,11,17	4
7	1,5,6,17	4
8	1,2,3,4	4
9	1,3,31,33,34	5
10	3,34	2
11	1,5,6	3
12	1	1
13	1,4	2
14	1,2,3,4,34	5
15	33,34	2
16	33,34	2
17	6,7	2
18	1,2	2
19	33,34	2
20	1,2,34	3
21	33,34	2
22	1,2	2
23	33,34	2
24	26,28,30,33,34	5
25	26,28,32	3
26	24,25,32	3
27	30,34	2
28	3,24,25,34	4
29	3,32,34	3
30	24,27,33,34	4
31	2,9,33,34	4
32	1,25,26,29,33,34	6
33	3,9,15,16,19,21,23,24,30,31,32,34	12
34	9,10,14,15,16,19,20,21,23,24,27,28,29,30,31,32,33	17

Table 3. The value of Q for different iterations and community-number.

Iterations	The Number of Community	The Number of Experiments	The Value of Q
100	2	1	0.3718
		2	0.3718
		3	0.3600
	3	1	0.5488
		2	0.5476
		3	0.5581
	4	1	0.6659
		2	0.7165
		3	0.6492
300	2	1	0.3718
		2	0.3718
		3	0.3718
	3	1	0.5611
		2	0.5611
		3	0.5611
	4	1	0.7162
		2	0.7220
		3	0.7202
500	2	1	0.3718
		2	0.3718
		3	0.3718
	3	1	0.5611
		2	0.5611
		3	0.5611
	4	1	0.7244
		2	0.7241
		3	0.7241

search and propose more accurate target function (such as multi-objective optimization [7, 12]) for the actual network. To this end, we will take the number of communities into account; search the maximum value of Q under the conditions which have same number of community. By adjusting the number of iterations, with each iteration do three tests, record values of Q. The number of community respectively set 2, 3, 4, shown in Table 3.

If we only consider the case of 2- divided-community, as shown in the Table 3, Q values mostly equal 0.3718. Only one node inconsistent with the actual division, so the accuracy is about 97% and the Q value equals 0.3718 correspond the best solution we find. The community structure as follows (Fig. 6).

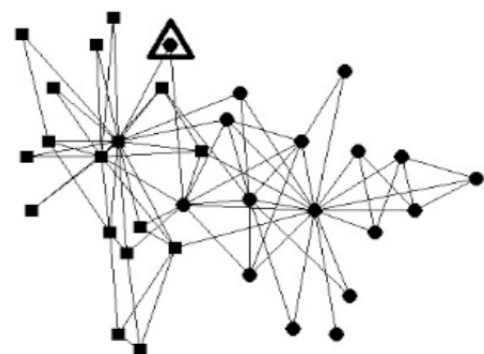


Fig. (6). Q=0.3718 Community structure (Triangle marked the wrong node).

Table 4. Comparison of different algorithms.

	GN	Random Walk	D-BA
The maximum accuracy	0.94	1	0.97
Influence of the parameters on the results	no	big	small
Whether need the number of communities	no	no	yes

The result of experiment shows that D-BA can get a better result on detecting the community structure with known number of communities.

If we compare the D-BA with ‘GN algorithm’ [3] and ‘Random Walk algorithm’ [5], we can get the result as shown in the Table 4.

D-BA can get an approximate result with a reasonable accuracy under the premise of know the number of communities in networks.

CONCLUSION

Detecting special community structure in complex network can help people grasp the overall structure of the network; understand the relationship among nodes in the network. The methods of detecting community have been applied to analyses the relationship of persons in social network, divide species of animal, analyses any correlation of other objects etc.

In this paper, Bat- Inspired algorithm (BA) of Xin-She Yang has been adjusted for mining the karate club network. The Simulation Platform is MATLAB R2013a. The accuracy of the best result is about 97%. Of course, the parameters of the algorithm (D-BA) and the objective function also can be adjusted or optimized in order to get better results continually.

CONFLICT OF INTEREST

The authors confirm that this article content has no conflict of interest.

ACKNOWLEDGMENTS

This work is supported by the National Natural Science Foundation of China (NSFC) (no. 61103198) and Communi-

cate university of China Engineering Planning Project (3132014XNG1445).

REFERENCES

- [1] W. Xiezou, “The research summary of structural analysis of complex network”, *Complex Systems and Complexity Science*, vol. 2, no. 3, July 2005.
- [2] M. Coscia, F. Giannotti, D. Pedreschi, “A classification for community discovery methods in complex networks”, *Statistical Analysis and Data Mining*, vol. 4, pp.512-546, 2011.
- [3] G. Michelle, M.E.J. Newman, “Community structure in social and biological networks”, *Proceedings of the National Academy of Sciences*, USA, vol. 99, pp.7821-7826, 2002
- [4] L. Freeman, “A set of measures of centrality based upon betweenness”, *Sociometry*, vol. 40, pp. 35-41, 1977.
- [5] P. Pons, M. Latapy, “Computing communities in large networks using random walks”, *Lecture Notes in Computer Science*, vol. 3733, pp. 284-293, 2005.
- [6] X.S. Yang, “A new metaheuristic bat-inspired algorithm nature inspired cooperative strategies for optimization”, *Studies in Computational Intelligence*, vol. 284, pp. 65-74, 2010
- [7] F.L. Huang, S.C. Zhang, X.F. Zhu, “Discovering network community based on multi-objective optimization”, *RuanJian Xue Bao/Journal of Software*, vol. 24, no. 9, pp. 2062-2077, 2013
- [8] B. Amiri, L. Hossain, J.W. Crawford, “An Efficient Multiobjective Evolutionary Algorithm for Community Detection in Social Networks”, *Evolutionary Computation*, New Orleans, USA, pp. 2193-2199, 2011
- [9] M.E. Newman, M. Girvan, “Finding and evaluating community structure in networks,” *Physical Review E*, vol. 69, no. 2, p. 026113, 2004.
- [10] M.E. Newman, “Fast algorithm for detecting community structure in networks”, *Physica Review E*, vol. 69, no. 6, pp. 066133, 2004
- [11] W.W. Zachary, “An information flow model for conflict and fission in small groups” *Journal of Anthropological Research*, vol. 33, no. 4, pp. 452-473, 1977
- [12] X.S. Yang, “Bat Algorithm for Multi-objective Optimisation”, *International Journal of Bio-inspired Computation*, vol. 3, no. 5, pp. 267-274, 2011.

Received: June 10, 2015

Revised: July 29, 2015

Accepted: August 15, 2014

© Chunyu and Yun; Licensee *Bentham Open*.

This is an open access article licensed under the terms of the (<https://creativecommons.org/licenses/by/4.0/legalcode>), which permits unrestricted, non-commercial use, distribution and reproduction in any medium, provided the work is properly cited.