

Community Detection in Complex Networks Using Improved Artificial Bee Colony Algorithm

Ziyang Wang, Xiaolin Zhao, Peiyuan Wen, Jingfeng Xue and Changzhen Hu

Beijing Key Laboratory of Software Security Engineering Technology, Beijing Institute of Technology, No.5 Yard, Zhong Guan Cun South Street, Haidian District, Beijing, China

Abstract—With more and more various systems in nature and society are proved to be modeled as complex networks, community detection in complex networks as a fundamental problem becomes a hot research topic in a large scale of subjects. Artificial Bee Colony Algorithm (ABC) has high efficiency and does not require any prior knowledge about the number or the original division of the communities. So it is suitable to solve complex clustering problems. We propose an improved ABC algorithm which modifies the number of initial food sources and dynamically adjusts search scope. Experimental results show that our algorithm can discover communities effectively by the classic Zachary Karate Club network. By comparative experiments, the improved artificial bee colony algorithm outperforms the traditional ABC algorithm in complex network.

Keywords—component; complex network; artificial bee colony algorithm; community detection

I. INTRODUCTION

Complex network consists of many nodes and complex links. As it has flexible ability to describe networks, complex network is widely used in personal networks, social network analysis, collaborative networks and reference networks, etc [1]. Complex networks typically have small-world, scale-free, clustering properties [2] [3]. Complex networks usually have community structure of dense connections internally and sparse connections between communities

We can use community detection to analyze complex networks community structure by extracting the hidden information about the complex network topology diagram. The discovery of the community structure helps us to understand the attributes of nodes from the network topology alone and to reveal the structure of networks [4].

Since Girvan and Newman proposed GN algorithm based on edge in 2002, many community detection algorithms have been developed and have been widely used in various areas [5]. Detecting community structure in networks attracts more and more research attention [6] [7].

However, most of those methods require prior knowledge of the entire graph structure. In some cases, the real networks are either too large or too dynamic to be known completely. Some algorithms require prior knowledge of the number or size of communities in the network. These factors limit the community detection in complex networks [8].

Artificial bee colony algorithm is an optimization algorithm based on the intelligent foraging behavior of honey bee swarms. It simulates three groups of bees to search for food sources and

exchange information in order to find out optimal food sources [9] [10].

Since artificial bee colony algorithm is fast and efficient and without prior knowledge, etc., it is appropriate to be used to solve complex clustering problem. But the ABC algorithm is easy to fall into local optimization and has low search speed [11]. To prevent the defects of algorithm, an improved artificial bee colony algorithm is proposed to detect complex networks community.

In this paper, we check the effectiveness and performance of improved ABC by utilizing different test functions to analyze convergence and time complexity of the algorithm.

Experiment results show that the improved ABC algorithm can effectively detect communities, and have fast convergence. In the end, the improved ABC algorithm provides a new way for community detection in complex networks.

II. ARTIFICIAL BEE COLONY ALGORITHM

Artificial bee colony algorithm is a swarm intelligence algorithm, ABC algorithm is a recently proposed optimization technique which simulates the intelligent foraging behavior of honey bees [12-17].

In the ABC algorithm, there are three types of bees: employed bees, onlooker bees, and scout bees and also there are several food sources. The goal is to find the largest food source.

- The employed bees search food around the food source in their memory; meanwhile they share the information of these food sources to the onlooker bees.
- The onlooker bees tend to select good food sources from those found by the employed bees. The food source that has higher quality (fitness) will have a large chance to be selected by the onlooker bees than the one of lower quality.
- The scout bees are translated from a few employed bees, which abandon their food sources and search new ones.

The food resource is abstracted into points in solution space, representing the potential solutions. In the ABC algorithm, the first half of the swarm consists of employed bees, and the second half constitutes the onlooker bees. The number of employed bees or the onlooker bees is equal to the number of solutions in the swarm.

First, the ABC generates a randomly distributed initial population of N solutions (food sources) $\{X_1, X_2, \dots, X_N\}$, where N denotes the swarm size. And each solution X_i ($i=1, 2, \dots, N$) is a vector whose dimension size is D.

Each employed bee X_i generates a new candidate solution V_i in the neighborhood of its present position as equation below:

$$V_{ij} = X_{ij} + \varphi_{ij}(X_{ij} - X_{kj}) \quad (1)$$

Employed bees get a new position V_{ij} using the current position X_{ij} and a random position X_{kj} . If the fitness value of V_i is better than that of its parent X_i , then update X_i with V_i ; otherwise keep X_i unchanged.

Where X_k is a randomly selected candidate solution ($k \neq i$), j is a random dimension index selected from the set $\{1, 2, \dots, D\}$, and φ_{ij} is a random number within $\{-1, 1\}$.

After all employed bees complete the search process; they share the information of their food sources with the onlooker bees through waggle dances. An onlooker bee evaluates the nectar information taken from all employed bees and chooses a food source with a probability related to its nectar amount. This probabilistic selection is described as equation below:

$$P_i = \frac{fit_i}{\sum_{j=1}^N fit_j} \quad (2)$$

where fit_i is the fitness value of the i^{th} solution in the swarm. This probabilistic selection is really a roulette wheel selection mechanism. As seen, the better the solution i , the higher the probability of the i^{th} food source selected.

If a position cannot be improved over a predefined number (called limit) of cycles, then the food source is abandoned. Assume that the abandoned source is X_i , and then the scout bee discovers a new food source to be replaced with i^{th} as equation below:

$$X_{ij} = r(X_j^{max} - X_j^{min}) + X_j^{min} \quad (3)$$

where r is a random number within $[0, 1]$ based on a normal distribution and X_j^{min} , X_j^{max} are lower and upper boundaries of the j^{th} dimension, respectively. j is a random dimension index selected from the set $\{1, 2, \dots, D\}$.

As can be seen from the above steps, artificial bee colony algorithm takes into account the global breadth first search and local depth first search. But there are some problems in initial random selection of population, random steps updating, poor elimination mechanism and the convergence problems and so on. Thus to overcome the limitations of the traditional ABC problem, this paper presents some improved methods, and applies the algorithm to complex networks community detection.

III. IMPROVED ARTIFICIAL BEE COLONY ALGORITHM

A. The Initial Number N of Food Sources

In the traditional ABC algorithm, the number of food source as the initial parameters of the algorithm is given in advance by a priori condition. However it is difficult to set the initial actual number of clusters, which directly affects the accuracy of the algorithm [18-22].

In this paper, initializing cluster centers with the density is proposed to optimize the number of clusters in the solution.

Let N_{fit} represent the best number of clusters. Since it is difficult to directly determine the value of N_{fit} , you need to set a reasonable range which is $N_{min} \leq N_{fit} \leq N_{max}$. For upper limit of clusters with the sample number k , many scholars use the rule \sqrt{k} , so $2 \leq N_{fit} \leq \sqrt{k}$.

The criteria of complex network community detection is close relations in internal community, sparse relations between communities, so the evaluation function F (Equation 6) is used to calculate N_{fit} . Clustering in community takes Euclidean distance.

In the complex network structure diagram, the shortest path between any two nodes, $d(x1, x2)$, is calculated by Dijkstra algorithm. Let $m = N_{max}$, similarity inside the community is described as below:

$$I(G) = \sum_{i=1}^m I(G_i) = \sum_{i=1}^m \sum_{x \in G_i} d(x, G_i) \quad (4)$$

Dispersion between communities is calculated as follows:

$$H(G) = \sum_{1 \leq i \leq j \leq m} d(G_i, G_j) \quad (5)$$

The smaller $I(G)$ is, the closer relations in the internal community are. The bigger $H(G)$ is, the sparser relations between communities are. The smaller $I(G)/H(G)$ is, the better complex network community detection is. So the ratio is described as below:

$$F = \frac{I(G)}{H(G)} \quad (6)$$

B. The Initial Points of Food Sources

Floyd algorithm is used to calculate the distance matrix L of graph formed by complex network. S_{max} is the furthest distance of any two points from L. Let $n = S_{max}/N_0$, the cluster is divided into n intervals. Traverse all nodes and distribute each node to n distance intervals (points in the interval can be repeated), then select maximum degree distribution of points in each interval, ultimately get the initial food source $\{X_1, X_2, \dots, X_N\}$.

C. The Dynamic Adjustment of Search Area

In the traditional bee colony algorithm, employed bees search food source in the vicinity of the location according to formula (1), in which the bigger the value of the control parameter φ_{ij} is, the larger the bees' search range is, otherwise, the smaller the search range is.

Parameter φ_{ij} of formula (1) is randomly generated, and the actual requirements of search range are different in different periods of algorithm.

In the early period of algorithms, search range is large, thereby the position of the best food sources is found accurately. With the increase of the number of iterations, the bees search for small-scale local position.

So in the iteration of algorithm, the parameters φ_{ij} is dynamically adjusted with polynomial decreasing form, which can control search area of the swarm. Modified search formula is:

$$V_{ij} = X_{ij} + k\varphi_{ij}(X_{ij} - X_{kj}) \quad (7)$$

where k is coefficient adjusting parameters, k is calculated as below:

$$k = 1 - t/T \quad (8)$$

where t is the number of iteration and T is the total number of iterations.

D. The Flow of Improved ABC Algorithm

1) Calculate the distance between any two nodes from an adjacency matrix in a complex network, and represent it by the adjacency matrix.

2) Count complex network of nodes n , set the number of employed bees and onlooker bees to k ($k = \sqrt{n}$), initialize the maximum number of iterations and cycle control parameters t . And let the initial number of food source N equal k .

3) According to the distance interval, select maximum degree distribution of points in each interval as the initial cluster centers from k initial intervals.

4) Based on Euclidean distance of nodes, separate the nodes into k different communities.

5) Loop until cycle

Loop a variable from 1 to k , perform the following steps:

a) According to lead the bee formula (7) search for new food sources, according to the evaluation function calculation formula (6) is F , each time retains the latest F , this step cycle several times, the best memory as a food source corrected community center

b) Calculate probability P_t of t food sources according to formula (2).

c) Onlooker bees take a roulette wheel selection mechanism to select the maximum P_t of food source.

d) Onlooker bees also search food sources by formula (7) and store best food source after several iterations.

e) Traverse all nodes and classify nodes into different communities. Save the community structure formed by onlooker bees in order to prepare for output.

f) If the solution to some food source is not updated after the limit cycles, the corresponding onlooker bees are turned into scout bees. Re-initialize the food source according to formula (3).

6) Loop according to the evaluation function F until pick out the optimal modularity of community structure. Merge the nearest two communities in turn. For all computing nodes F , store the community state when the optimal F occurs. Thus merge all communities into more precise number of community G .

7) Reassign the number of employed bees, onlooker bees, food sources to G , then execute the step 2-5. Classify the nodes of complex network to the corresponding G communities of food source, output the effect result of complex network.

IV. EXPERIMENTAL RESULTS AND ANALYSIS

A. Effectiveness of Community Detection in Complex Networks

To check whether our algorithm can detect the complex network communities, we tested the accuracy of the algorithm on well-known data set, namely the Zachary Karate Club. The Zachary Karate Club data contains the community structure in a karate club. The network consists of 34 vertices and 78 edges. We run our algorithm on this dataset for a number of times.

We analyze the Zachary Karate Club with the proposed algorithm. First select $\sqrt{34} \approx 5$ clustering centers and degrees of these centers are relatively large. So node 1, 33, 34, 3, 14 are selected as the initial cluster centers. Ultimately a certain community structure of Zachary Karate Club is shown as Figure I.

The algorithm finds 95%-100% correct community structure. As we can see from Figure II, it places the node 3 and 20 into wrong community.

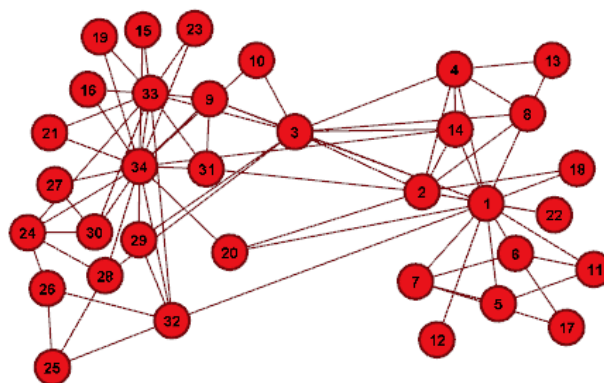


FIGURE I. REAL COMMUNITY STRUCTURE OF ZACHARY KARATE CLUB

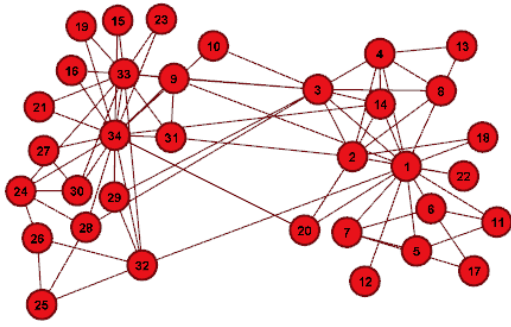


FIGURE II. COMMUNITY STRUCTURE OF ZACHARY KARATE CLUB BY IMPROVED ABC ALGORITHM

B. Performance of Improved ABC Algorithm

TABLE I. SET OF TEST FUNCTIONS

Function	Expression	Variable scope	The optimal value
Sphere	$f_1(x) = \sum_{i=1}^D x_i^2$	[-100 100]	0
abs	$f_2(x) = \prod_{i=1}^D abs(x_i) + \sum_{i=1}^D abs(x_i)$	[-10 10]	0
Generalized	$f_3(x) = 1 + \sum_{i=1}^D \frac{x_i^2}{4000} - \prod_{i=1}^D (\cos \frac{x_i}{\sqrt{i}})$	[-600 600]	0
Rosenbrock	$f_4(x) = \sum_{i=1}^n 100(x_i + 1 - x_i^2)^2 + (1 - x_i)^2$	[-100 100]	0
Rastrigin	$f_5(x) = \sum_{i=1}^D [x_i^2 - 10 \cdot \cos(2\pi x_i) + 10]$	[-5.12 5.12]	0

In order to improve the efficiency of finding the optimal solution, the improved algorithm modifies the search strategy of onlooker bees and scout bees by changing control parameters to make the bees search the food sources in nearby regions of outstanding individuals.

Unimodal function Sphere and multimodal function Rosenbrock are chosen to analyze the performance.

In the following diagram, Q1 represents improving both search strategies of onlooker bees and scout bees. Q2 represents only improving search strategy of onlooker bees, Q3 represents only improving search strategy of scout bees.

We initialize parameters as follows: the swarm consists of 20 employed bees and 20 onlooker bees. Let the dimension of test functions equal 40, iterations are 150 times. We run our algorithm for 5 times and the evolution curve is shown in Figure III and Figure IV.

From the above chart, we can see both Q1 and Q2 can quickly converge to the theoretical optimal value, however Q1 converges two times quicker than Q2 under the same circumstances. So modified search strategy of onlooker bees and scout bees can avoid local optimum and improve the efficiency of calculating the optimal solution. Q3 indicates only modifying the search strategy of onlooker bees can't change the convergence which shows the purpose to increase the diversity of swarm from the experimental point of view.

Representative test functions are selected to verify the performance of the improved ABC algorithm. Table I shows the five common benchmark function for comparison.

To evaluate our improved algorithm, we conducted effectiveness tests and measured performance. All experiments were ran on a DELL computer with 2.53GHz CPU and 4GB memory. We mainly achieve the following objectives:

- Test the effect of the initial value N on test result.
- Calculate parameters of the algorithm and optimize evaluation.
- Compare with known algorithms and analyze the pros and cons.
- Determine algorithm convergence and time complexity

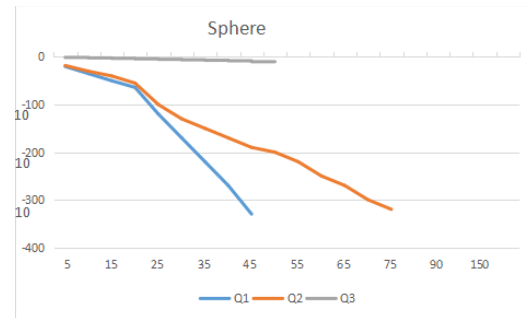


FIGURE III. THE EVOLUTION CURVE OF SPHERE TEST FUNCTION

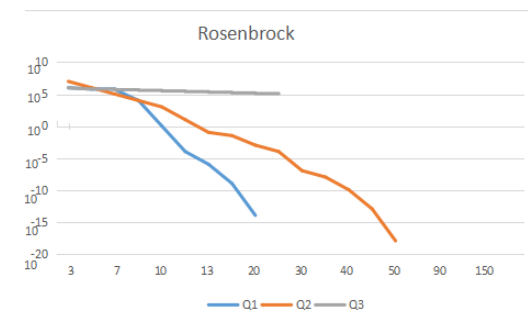


FIGURE IV. THE EVOLUTION CURVE OF ROSENBRACK TEST FUNCTION

C. Effect of Experimental Parameters on the Performance of the Improved ABC Algorithm

For the initial value of parameter N and the number of iterations K, the following two tables have shown different results under situation of different parameters values of N and K with function dimension D=70.

Table II shows the average N of 5 experiments, assuming K = 500 and dimension parameter D = 70.

TABLE II. TEST RESULTS WITH K = 500, D = 70

Function	N=20	N=40	N=60	N=80
Sphere	0	0	0	0
Rosenbrock	0	0	0	0
Generalized	8.4E-6	4.6E-6	7.2E-6	9.5E-6

Table III shows the average K of 5 experiments, assuming N=50 and D=70.

TABLE III. TEST RESULTS WITH N=50, D=70

Function	K=200	K=500	K=800	K=1100
Sphere	5.2E-180	0	0	0
Rosenbrock	7.6E-5	4.8E-5	2E-6	4.4-5
Generalized	0	0	0	0

Table II illustrates the calculation results are in the same scale with parameter K = 500 and different values of parameter N. In Table III, we know that calculation results of all functions except Sphere mainly are in the same scale with parameter N = 50 and different parameter K.

The results show that the proposed algorithm is not sensitive to parameter N and K, it can get satisfactory results with different parameter K and N.

To measure the performance introduced by proposed algorithm, we ran experiments using test functions to compare the performance of improved algorithm with the traditional ABC algorithm. As shown in Table IV and Table V, the experimental data is the average of five experiments and we choose 50 or 100 dimensions of test functions.

Figure V, Figure VI and Figure VII present the evolution curve of two algorithms when the function dimension is 100, in which Q1 stands for the improved algorithm.

TABLE IV. AVERAGE RESULTS OF ABC ALGORITHM

Function	Dimension	Optimal Value	Standard Deviation
Sphere	50	2.06E-6	2s.74E-6
Sphere	100	4.68E-4	5.72E-4
Rastrigin	50	3.34	1.62
Rastrigin	100	42.46	5.58
Rosenbrock	50	4.66E-5	2.43E-6
Rosenbrock	100	5.67E-2	3.24E-4

TABLE V. AVERAGE RESULTS OF IMPROVED ABC ALGORITHM

Function	Dimension	Optimal Value	Standard Deviation
Sphere	50	0	0
Sphere	100	0	0
Rastrigin	50	0	0
Rastrigin	100	0	0
Rosenbrock	50	1.72E-12	1.64E-12
Rosenbrock	100	8.42E-10	6.86E-10

Seen from the data in the tables, for different test functions dimensions, the performance of improved algorithm is better than the performance of traditional ABC algorithm with fast convergence and low standard deviation.

We can obviously see that the iterations of improved algorithm to achieve convergence are significantly less than the ones of traditional ABC from the above diagram.

For multiple local optima multimodal function, ABC algorithm has slow convergence defects. When ABC algorithm updates the food source, it randomly selects neighborhood, since the initial search step which could be unsuitable affects the convergence speed, the search could be stopped as the step size decreases in later period. The proposed algorithm strengthens the colony's ability of exchanging information, improves the global search ability so that algorithm can quickly converge without losing search efficiency.

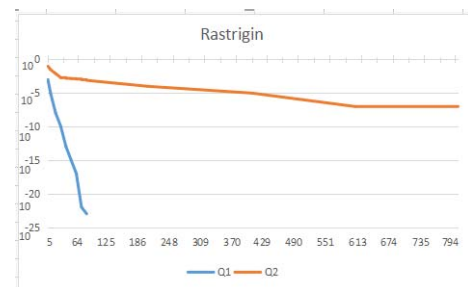


FIGURE V. THE EVOLUTION CURVE OF RASTRIGIN TEST FUNCTION

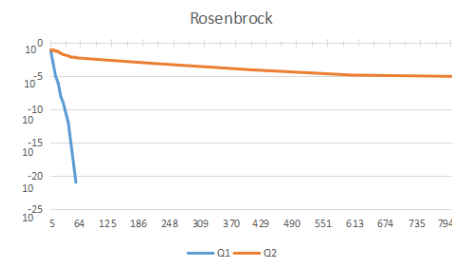


FIGURE VI. THE EVOLUTION CURVE OF ROSENBRUCK TEST FUNCTION

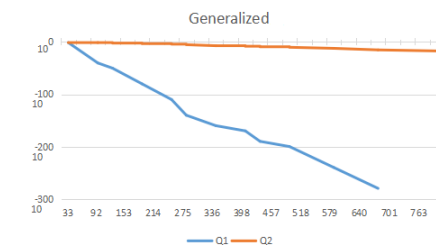


FIGURE VII. THE EVOLUTION CURVE OF GENERALIZED TEST FUNCTION

V. CONCLUSIONS

In this paper, we have presented an improved artificial bee colony algorithm to detect communities of complex network. The proposed algorithm improves by define the initial number of food source, select the initial point of food source and

dynamically adjust the search area. The proposed algorithm's accuracy is tested with the known Zachary Karate Club. Meanwhile, it can speed up the global search and keep high search efficiency and enable the accuracy of the community detection of complex networks.

[22] Q. Chen, and M. Fang. "An efficient algorithm for community detection in complex networks." the 6th Workshop on Social Network Mining and Analysis. 2012.

REFERENCES

[1] Š. Lovro, and M. Bajec. "Software systems through complex networks science: Review, analysis and applications." Proceedings of the First International Workshop on Software Mining. ACM, 2012.

[2] D. J. Watts, and H. S. Steven. "Collective dynamics of 'small-world' networks." *nature* 393.6684 (1998): 440-442.

[3] A. L. Barabási, and A. Réka. "Emergence of scaling in random networks." *science* 286.5439 (1999): 509-512.

[4] F. Ali, and K. A. Samani. "Community detection based on the "clumpiness" matrix in complex networks." *Physica A: Statistical Mechanics and its Applications* 391.7 (2012): 2463-2474.

[5] A. Faqeeh, and S. A. Keivan. "Community detection based on the "clumpiness" matrix in complex networks." *Physica A: Statistical Mechanics and its Applications* 391.7 (2012): 2463-2474.

[6] M. Gong, Q. Cai, Y. Li. "An improved memetic algorithm for community detection in complex networks." 2012 IEEE Congress on Evolutionary Computation. IEEE, 2012.

[7] A. Y. Yeol, J. P. Bagrow, and S. Lehmann. "Link communities reveal multiscale complexity in networks." *Nature* 466.7307 (2010): 761-764.

[8] F. Santo, and C. Castellano. "Community structure in graphs." *Computational Complexity*. Springer New York, 2012. 490-512.

[9] A. S. Shirchorshidi. "Big data clustering: a review." International Conference on Computational Science and Its Applications. Springer International Publishing, 2014.

[10] R. L. Álvaro. "A Comparative study on multiobjective swarm intelligence for the routing and wavelength assignment problem." *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)* 42.6 (2012): 1644-1655.

[11] X. Du, L. Cheng, and L. Liu. "A swarm intelligence algorithm for joint sparse recovery." *IEEE Signal Processing Letters* 20.6 (2013): 611-614.

[12] B. Akay, and K. Dervis. "A modified artificial bee colony algorithm for real-parameter optimization." *Information Sciences* 192 (2012): 120-142.

[13] Y. Zhao, X. Liu, and J. Qu. "The K-medoids Clustering Algorithm by a Class of P System." *JOURNAL OF INFORMATION & COMPUTATIONAL SCIENCE* 9.18: 5777-5790.

[14] K. Dervis. "A comprehensive survey: artificial bee colony (ABC) algorithm and applications." *Artificial Intelligence Review* 42.1 (2014): 21-57.

[15] X. Li, and M. An. "An efficient and robust artificial bee colony algorithm for numerical optimization." *Computers & Operations Research* 40.5 (2013): 1256-1265.

[16] A. M. Rezaei, and A. Karami. "Artificial bee colony algorithm for solving multi-objective optimal power flow problem." *International Journal of Electrical Power & Energy Systems* 53 (2013): 219-230.

[17] X. Yu, and Z. Zhu. "A modified artificial bee colony algorithm with its applications in signal processing." *International Journal of Computer Applications in Technology* 47.2-3 (2013): 297-303.

[18] A. Bahriye, and D. Karaboga. "A modified artificial bee colony algorithm for real-parameter optimization." *Information Sciences* 192 (2012): 120-142.

[19] A. Bilal. "Chaotic bee colony algorithms for global numerical optimization." *Expert Systems with Applications* 37.8 (2010): 5682-5687.

[20] K. Nandakishore J.. "Test case optimization using artificial bee colony algorithm." International Conference on Advances in Computing and Communications. Springer Berlin Heidelberg, 2011.

[21] Tasgin, Mursel, A. Herdagdelen, and H. Bingol. "Community detection in complex networks using genetic algorithms." *arXiv preprint arXiv:0711.0491* (2007).