Bulletin of Electrical Engineering and Informatics
Vol. 8, No. 4, December 2019, pp. 1433~1440
ISSN: 2302-9285, DOI: 10.11591/eei.v8i4.1196

Optimizing community detection in social networks using antlion and K-median

Amany A. Naem[1], Neveen I. Ghali[2]
[1]Al-Azhar University, Faculty of Sience, Cairo, Egypt
[2]Future University in Egypt, Faculty of Computers & Information Technology, Cairo, Egypt

Article Info

Article history:

Received Apr 14, 2018
Revised May 20, 2018
Accepted Mar 17, 2019

Keywords:

Antlion optimization
Community detection
K-median clustering
Modularity
Social network

ABSTRACT

Antlion Optimization (ALO) is one of the latest population based optimization methods that proved its good performance in a variety of applications. The ALO algorithm copies the hunting mechanism of antlions to ants in nature. Community detection in social networks is conclusive to understanding the concepts of the networks. Identifying network communities can be viewed as a problem of clustering a set of nodes into communities. k-median clustering is one of the popular techniques that has been applied in clustering. The problem of clustering network can be formalized as an optimization problem where a qualitatively objective function that captures the intuition of a cluster as a set of nodes with better in ternal connectivity than external connectivity is selected to be optimized. In this paper, a mixture antlion optimization and k-median for solving the community detection problem is proposed and named as K-median Modularity ALO. Experimental results which are applied on real life networks show the ability of the mixture antlion optimization and k-median to detect successfully an optimized community structure based on putting the modularity as an objective function.

Corresponding Author:

Amany A. Naem,
Faculty of Sience,
Al-Azhar University,
Cairo, Egypt.
Email: manoo_basom@yahoo.com

1. INTRODUCTION

Technology and social network have presented significant tools for learning and advance continually. These technologies have enabled the rapid development of life [1]. Persons in the social networks form a relation frame through different connections which produces a large amount of dissemination of information. The relation frame is called by community. Community detection is great important to discover the structure of social networks, analysis the information, and understanding as well as control the public sentiment. Community can be treated as a summary of the whole network thus soft to visualize and understand [2].

Communities are groups of members (nodes) that are connected heavily inside the group but connected sparely with the rest of the network. Community detection in large net-works is potentially very advantageous to member, where nodes belonging to a tight-knit community are more than likely to have other attributes in common. For example on the World Wide Web (WWW) a cluster can be looked at as information or as physical links and paths connecting to each other [3, 4]. Song et al. in [5] applied discrete Bat Algorithm to the community detection of showing networks and achieved good results. Hafez et al. in [6] used Genetic Algorithm (GA) as an effective optimization technique to solve the community detection problem as a single-objective and multi-objective problem, they used the most popular objectives proposed over the past years, and they showed how those objective correlate with each other. Fortunato and Hric in [7]

proposed algorithm to guide tour through the main features of problem. They pointed out strengths and weaknesses of popular methods, and gave directions to their use. Newman in [8] used optimization methods or approximation method such as spectral method to solve community detection problem. Newman and Girvan in [9] introduced a new technique called Modularity. Pizzuti in [10] proposed a genetic based approach to discover communities in social networks. Their algorithm was a simplex but effective fitness function able to identify densely connected groups of nodes with sparse connections between groups. Honghaoet et al. in [11] suggested an ant colony optimization (ACO) based approach to discover communities. They demonstrated that ACO based approach results in a significant enhancement in modularity values as compared to existing heuristics in the literature. Masdarolomoor et al. in [12] proposed a new method for community detection in networks and used to simulated annealing to maximize the modularity. Their algorithm was evaluated by modularity metric and worked better in time and accuracy compared to similar methods. Barawy et al. in [13] presented the idea of using the results of an opt mization algorithms Particle Swarm Optimization (PSO) and Exponential Particle Swarm Optimization EPSO as input to the k-means clustering algorithm in order to have a well community detection for social network data. Naem et al. in [14] proposed a hybrid cat swarm optimization and k-median for solving the problem of community detection. In this paper, a mixture antlion optimization and k-median for solving the community detection problem is proposed and named as K-median Modularity ALO. Where using k-median clustering to detect the community and ALO for optimizing the modularity. By setting the modularity as an objective function in order to have high value for modularity as well community detection for social network data.

The remainder of the paper is organized as follows: Brief introduction on community detection problem, antlion swarm optimization algorithm, and k-median algorithm are introduced in Sections 2. The details of the proposed method are presented in Section 3. Section 4 shows our experimental results on datasets social networks. Finally, we supply conclusions in Section 5.

2. PRELIMINARIES

2.1. Community detection problem

Detecting the hidden community in social networks, which is conclusive to understand the faces of networks, is vary important object in Social Network Analysis. Community detection algorithms aim to find communities based on the network structure to existing groups of nodes that are heavily connected [15]. So, to evaluate the clustering performance, modularity metric is put into use as a measure for the quality of communities of the network [16].

Modularity function was proposed by Girvan and Newman in 2004 [9], modularity measure was designed to measure the strength of division of a network into clusters (modules or communities). The modularity way detects communities by searching over possible divisions of a network for one or more that have particularly best modularity.

Suppose we have a network that includes n vertices, and let the number of edges between vertices i and j be A_{ij}, which will usually be 0 or 1, so the quantities A_{ij} are the elements of the so called adjacency matrix.

Concurrently, the expected number of edges between vertices i and j if edges are placed at random is $k_i * k_j / 2m$, where k_i and k_j represent the degrees of the vertices and m is the total number of edges in the network. So the modularity can be formalized as (1) [16]:

$$Q = \frac{1}{2m}\sum_{j=1}^{n}\left(A_{ij} - \frac{k_i * k_j}{2m}\right)\delta\left(H_i, H_j\right) \quad (1)$$

where: Q represents the modularity of network;

H_i and H_j are the identity of the community which the node i and j belong to in certain iteration respectively. If vertices i and j are in the same community, $\left(H_i, H_j\right)$ =1, else 0.

2.2. Antlion optimization algorithm

Antlion Optimization (ALO) [17] is a new nature inspired algorithm presented by Mirjalili in 2015. Mirjalili depends on the following facts and assumptions in the artificial antlion optimization algorithm:

a. Ants change position around the search space by using different random walks;
b. Random walks are influenced by the snares of antlions;
c. Antlions can make snares proportional to their fitness (the higher the fitness, the greater the hole);
d. Antlions with greater holes have a higher probability of catching ants;
e. Every ant can be caught by an antlion in all iterations;
f. The range of random walks is reduced adaptively to simulate sliding ants across antlions;

g. If ant becomes fitter than antlion, this means that the ant is caught and pulled toward the hole by the antlion;
h. Antlion repositions to the most recently caught prey and builds a hole to update its chance of catching another prey after each hunt.

ALO has very distinguished outcomes in fields of exploitation, local optima avoidance, and convergence. Mathematical modeling of the ALO algorithm can be formulated as in the following items [18]:

Random walks: Random walks of ants when searching for food in nature can be for mulated as follows:

$$X(t)=[0, \text{cumsum}(2r(t1)-1), \text{cumsum}(2r(t2)-1)... \text{cumsum}(2r(tn)-1)] \quad (2)$$

where: n shows the maximum number of iterations;
cumsum computes the cumulative sum;
t denotes to the step of random walk;
r(t) shows a stochastic function and is given by:

$$r(t) = \begin{cases} 1, if\ random\ >\ 0.5 \\ 0, if\ random\ \leq\ 0.5 \end{cases} \quad (3)$$

Where $random$ represents a random number and it falls in [0, 1].
The place of ants is created with this matrix:

$$M_{Ant} = \begin{bmatrix} A_{1,1} & \cdots & A_{1,b} \\ \vdots & \ddots & \vdots \\ A_{n,1} & \cdots & A_{n,b} \end{bmatrix} \quad (4)$$

Here: M_{Ant} is the matrix for utilizing the position of every ant;
$A_{i,j}$ acts the value of the jth variable of ith ant;
d denotes to the number of variables;
n represents to the number of ants.
The place of antlions is created with this matrix:

$$M_{AL} = \begin{bmatrix} AL_{1,1} & \cdots & AL_{1,b} \\ \vdots & \ddots & \vdots \\ AL_{n,1} & \cdots & AL_{n,b} \end{bmatrix} \quad (5)$$

Here: M_{AL} is the matrix for utilizing the position of every antlion;
$AL_{i,j}$ acts the value of the jth variable of ith antlion;
n denotes to the number of antlions;
d denotes to the number of variables.
The random walks of ants inside the search space are normalized using this equation:

$$X_i^t = c_i + \frac{(X_i^t - a_i) * (d_i - c_i^t)}{(b_i^t - a_i)} \quad (6)$$

where: $a_i$shows the minimum of random walk of ith variable;
d_i denotes the maximum of random walk of ith variable;
c_i^t shows the minimum of ith variable at tth iteration;
b_i^t is the maximum of ith variable in tth iteration.

Trapping in pit: Mathematical modeling of ants trapping in antlion's pits is given by these equations:

$$\begin{aligned} c_i^t &= \text{Antliont}_i^t + c^t \\ d_i^t &= \text{Antliont}_i^t + d^t \end{aligned} \quad (7)$$

where: c^t acts the minimum of all variables at tth iteration;
d^t acts the maximum of all variables at tth iteration;
c_i^t shows the minimum of all variables for ith ant;
d_i^t shows the maximum of all variables for ith ant;
Antliont_i^t presents the position of the chosen jth antlion at tth iteration.

Building trap: A roulette wheel is used to get higher probability for catching ants. This technique identifies fittest antlions.

Sliding ants towards antlion: Antlions exit the sands out the center of the hole so any ant trying to escape slide down the trap. The radius of the ant's random walks hypersphere is reduced according to these equations:

$$c^t = \frac{c^t}{U}$$
$$d^t = \frac{d^t}{U} \quad (8)$$

where:U is a ratio and is calculated by the following equation:

$$U = 10^y * \frac{i}{j} \quad (9)$$

where: t indicates to the current iteration;
T presents the maximum number of iterations;
w denotes to a constant and is defined based on T; t where (w=2 when t > 0.1 T, w=3 when t > 0.5 T, w=4 when t > 0.75 T, w=5 when t > 0.9 T, and w=6 when t > 0.95 T).

Catching ant and re-building pit: When the ant arrives to the bottom of the hole and is captured this is the final step of hunting. According to the last position, the antlions update its position by this equation:

$$\text{Antliont}_i^t = \text{Ant}_i^t \quad \text{if } f(\text{Ant}_i^t) > f(\text{Antliont}_i^t) \quad (10)$$

where: Antliont_i^t acts the position of the selected jth antlion at tth iteration;
Ant_i^t is the position of the selected ith ant at tth iteration;
t is the current iteration.

Elitism: It is very important in evolution algorithm where it is to maintain best solution. This can be modeled as the following equation:

$$Ant_i^t = \frac{R_A^t + R_E^t}{2} \quad (11)$$

where: R_A^t indicates to the random walk around the antlion selected by the roulette wheel at tth iteration;
R_E^t indicates to the random walk around the elite antlion at tth iteration.

2.3. K-medians clustering algorithm

Clustering is a standard procedure in multivariate data analysis used for communication and is designed to explore communities structure of the data objects. Clustering has multi methods such as k-means and k-medians [19]. The algorithm of k-medians clustering is symmetric to k-means clustering algorithm, but k-medians updates the cluster center by calculating the median of the same cluster to be the new cluster center. k-medians is sensitive to the initialization points of its k centers, every center having the tendency to remain roughly in the same cluster in which it is first situation [20]. The distance between each point of data and cluster center is evaluated using this equation:

$$d(x, c) = \|x - c\| \quad (12)$$

Where x point in data and c cluster center. The k-medians algorithm attempts to make k disjoint cluster that minimize the following equation:

$$U = \sum_{i=1}^{k} \sum_{x \in D} \|x - c_i\| \quad (13)$$

where: x=member of data D.
c_i=cluster center i.
k=number of clusters.

3. THE PROPOSED METHOD

The proposed method (K-median Modularity ALO) in this research consists of two main parts, clustering the data by using k-median and looking for the best modularity as well community detection for social network dataset by applying ALO algorithm. Steps of proposed method are summarized down:

Step 1: Defining the initial cluster center

Randomly select k points from the data points to be the initial cluster centers.

Step 2: Grouping data into clusters

Place the data into clusters with the closest cluster center by using (13).

Step 3: Calculating the modularity

Modularity is the fitness function in this method. The value of modularity is calculated using (1).

Step 4: Optimization with ALO

We apply the ALO algorithm on cluster centers to get best modularity value and best cluster centers for each data.

Step 5: Repeat steps

Recurrence steps 2-4 until it reaches the stop criteria, after update the clusters centers by calculating the median for each cluster. The steps of proposed method can summarize in Figure 1.

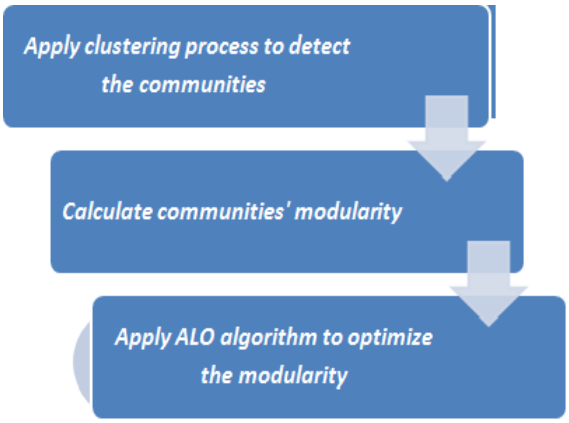

Figure 1. Proposed model

4. EXPERIMENTAL RESULTS

4.1. Datasets

In the section, the proposed method (K-median Modularity ALO) is applied on four real life social network datasets:

The Zachary Karate Club:

It was studied by Wayne W. Zachary from 1970 to 1972 and was observed from the members of a university karate club. The graph of Karate network composes of 34 nodes and 78 edges. Where each member of the club was represented by node and each relation between two members in the club was represented by edge. The problem was often discussed the use of this dataset to find groups of people after a struggle arose between teachers, which led to the divide the karate club into two group [21].

The Bottlenose Dolphins network:

This was named by social network of bottlenose dolphins and was aggregated by Lusseau from 1994 to 2001. The network was built the basis of the behavior of 62 bottlenose dolphins that live in Doubtful Sound, New Zealand. So the nodes represent the bottlenose dolphins and edges represent a frequent associations [22].

American College football network:

It represented American football games between American colleges during a regular season in fall 2000, as reorganized by M. Girvan and M. Newman. Each node in the network refers to team and edge refers to game of regular season between the two teams they connect [23].

The Polbooks network:

It network of books about US politics showed at the time of the 2004 presidential election and marketed by the online bookseller Amazon.com. Edges between books act frequent co purchasing of books by the same shoppers. This network was combined by V. Krebs [24].

Table 1 contains of information for the previously introduced datasets.

Table 1. Comparison of modularity results

Optimization Methods	Data			
	Karate	Dolphins	Football	Polbooks
GN [25]	0.401	0.519	0.599	0.516
FN [25]	0.380	0.489	0.577	0.502
BGLL [25]	0.418	0.518	0.602	0.498
HSCDA [25]	0.419	0.527	0.602	0.527
PSO K-means [14]	0.433	0.445	0.529	0.465
PSO K-median [14]	0.442	0.461	0.566	0.480
BAT K-means [14]	0.449	0.501	0.583	0.50
BAT K-median [14]	0.470	0.472	0.614	0.51
CSO K-means [14]	0.451	0.472	0.593	0.53
CSO K-median [14]	0.489	0.502	0.621	0.559
ALO K-means	0.469	0.519	0.612	0.557
ALO K-median	0.501	0.525	0.628	0.563

4.2. Results and analysis

The results of modularity after applying the proposed method (K-median Modularity ALO) are compared to K-means Modularity PSO, K-means Modularity Bat optimization, K-means Modularity CSO, K-median Modularity PSO, K-median Modularity Bat optimization, K-median Modularity CSO, GN, FN, BGLL, HSCDA [25] and K-means Modularity ALO on the real datasets. Since modularity is a famous community quality measure used vary widely in community detection, and it is used as a quality measure for the result community structure of all other achieves. Also Normalized Mutual Information (NMI) compares the accuracy of the outcome communities where it computes the likeness between two parts. NMI is described in the following equation [25]:

$$NMI(X,Y) = \frac{-2\sum_{i=1}^{C_X}\sum_{j=1}^{C_Y} C_{ij} log\left(C_{ij}N/C_i\,C_j\right)}{\sum_{i=1}^{C_X} C_i \log(C_i/N) + \sum_{j=1}^{C_Y} C_j \log\left(C_j/N\right)}$$

where: $NMI(X,Y)$ Denotes NMI for two parts X and Y;

C_{ij} : Number of nodes assigned to ith community in part X, and jth community inpart Y;

C_i : Number of nodes in part X assigned to ith community;

C_j : Number of nodes in part Y assigned to jth community;

C_X: The Community Number in part X;

C_Y: The Community Number in part Y;

N: Total number of nodes.

If NMI equals 1 this means that the two consequences consistent completely. If NMI equals zero this means that the two consequences inconsistent completely.

Applying the proposed algorithm on previously introduced datasets. On obtaining the results 50 consecutive runs on each datasets the average modularity and NMI are calculated. The results of modularity in Table 1; where it can be concluded that the modularity obtained by K-median Modularity ALO is better than that obtained by another methods when compered to results perviously obtained on [14] and [25] except in case of dolphins network with HSCDA get better result. Table 2 illustrates the experimental results of NMI. It confirms that the proposed method K-median Modularity ALO outperforms over other methods applied on [14] and [25] and K-means Modularity ALO.

Table 2. Comparison of NMI results

Data	Optimization Methods			
	PSO k-median	BAT k-median	CSO k-median	ALO k-median
Karate	0.61	0.68	0.71	0.84
Dolphins	0.50	0.53	0.61	0.68
Football	0.79	0.82	0.86	0.89
Polbooks	0.52	0.56	0.59	0.67

Figure 2-5 discuss the relation between modularity and community number for Karate network, Dolphins network, Football network, and Polbooks network respectively. It is obvious that when the community number equals 4, the Karate network and Dolphins network get the maximum modularity. And

when community number equals 9, the Football network obtains the maximum modularity. Also when community number equals 3, the Polbooks network obtains the maximum modularity.

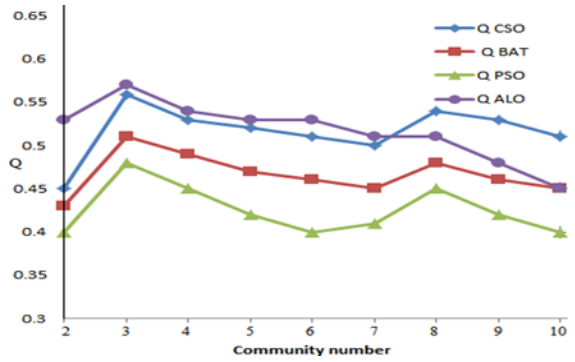

Figure 2. Relation between modularity Q and Community number for Karate network

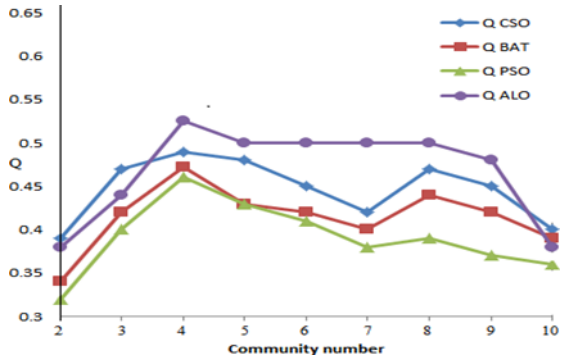

Figure 3. Relation between modularity Q and Community number for Dolphins network

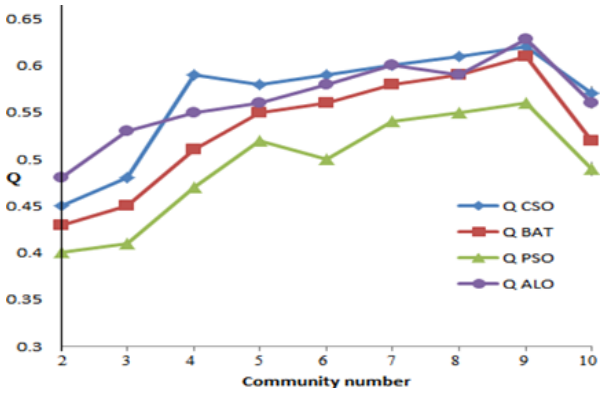

Figure 4. Relation between modularity Q and Community number for Poolbooks network

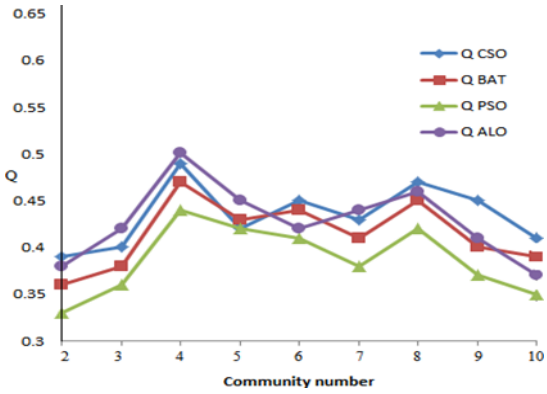

Figure 5. Relation between modularity Q and Community number for Football network

5. CONCLUSION

Community detection is great of important in computer science, biology, physics and sociology to understand of complicated systems. This problem is very challenging and not yet satisf-actorily solved despite many methods have been proposed. The k-median Modularity ALO is successfully implementing of NMI measure on networks confirmed this result when compered to other optimization methods.

REFERENCES

[1] Alsaadat, K., "The impact of social media technologies on adult learning," *International Journal of Electrical and Computer Engineering (IJECE)*, pp. 8(6), 2018.
[2] 2015. Cuijuan Wang, Wenzhong Tang, Bo Sun, Jing Fang and Yanyang Wang, "Review on community detection algorithms in social networks," *2015 IEEE International Conference on Progress in Informatics and Computing (PIC)*, Nanjing, 2015, pp. 551-555.
[3] Danon, l. Duch, J. Diaz-Guilera, A. Arenas, A., "Comparing community structure identification," *Journal of Statistical Mechanics*, vol. 9, pp. 1-10, 2005.
[4] Alzahrani, T and Horadam, K., "Community detection in bipartite networks: algorithms and case studies," *Complex Systems and Networks*, Springer Berlin Heidelberg, pp. 25-50, 2016.
[5] Song, A. Li, M. Ding, X. Cao & W. Pu K., "Community detection using discrete bat algorithm," *IAENG International Journal of Computer Science*, pp. 43, pp. 1-7, 2016.
[6] A. I. Hafez, N. I. Ghali, A. E. Hassanien and A. A. Fahmy, "Genetic Algorithms for community detection in social networks," *2012 12th International Conference on Intelligent Systems Design and Applications (ISDA)*, Kochi, 2012, pp. 460-465.

[7] Fortunato, S and Hric, D, "Community detection in networks: A user guide. Physics Reports," *Elsevier*, 659, pp. 1-44, 2016.
[8] Newman, M., "Spectral methods for network community detection and graph partitioning," *Journal of Physical Review E*, 88: 1-11, 2013.
[9] Newman, M and Girvan M, "Finding and evaluating community structure in networks," *Physical Review* 69, pp. 1-16. 2004.
[10] Pizzuti, C., "GA-net: a genetic algorithm for community detection in social networks," *Springer Berlin Heidelberg*, pp. 1081-1090, 2008.
[11] Chang Honghao, Feng Zuren and Ren Zhigang, "Community detection using Ant Colony Optimization," *2013 IEEE Congress on Evolutionary Computation*, Cancun, 2013, pp. 3072-3078.
[12] Z. Masdarolomoor, R. Azmi, S. Aliakbary and N. Riahi, "Finding Community Structure in Complex Networks Using Parallel Approach," *2011 IFIP 9th International Conference on Embedded and Ubiquitous Computing*, Melbourne, VIC, 2011, pp. 474-479.
[13] EL. Barawy, Y. EL. Bakrawy & L. Ghali. N., "K-means clustering with swarm optimization for social network community detection," *Asian Journal of Mathematics and Computer Research*, vol. 3, pp. 220-230, 2013.
[14] Naem A, EL Bakrawy L, Ghali N. "A hybrid cat optimization and k-median for solving Community Detection Problem". *Asian Journal of Applied Sciences.* 2017; 5: 892-903.
[15] J. Yang, J. McAuley and J. Leskovec, "Community Detection in Networks with Node Attributes," *2013 IEEE 13th International Conference on Data Mining*, Dallas, TX, 2013, pp. 1151-1156.
[16] Yahia, N. Saoud, N & Ghezala, H., "Evaluating community detection using a biobjective optimization," *International Conference on Intelligent Computing, Springer Berlin Heidelberg*, pp. 61-70, 2013.
[17] Mirjalili, S., "The ant lion optimizer," *Advances in Engineering Software*, vol. 83, pp. 80-98, 2015.
[18] Shivani, M. and Meenakshi, M. "Ant lion optimization for optimum power generation with valve point effects. Computer methods in applied machanics and engineering," *International Journal for Research in Applied Science and Engineering Technology (IJRASET)*, vol, 3, pp. 1-6, 2015.
[19] Tamizharasi, A. Selvathai, J. Kavi, A & Maarlin R., "Harinetha M. Energy aware heuristic approach for cluster head selection in wireless sensor networks". *Bulletin of Electrical Engineering and Informatics*, vol. 16(1), pp. 70-75, 2017.
[20] Whelan, C. Harrell, G & Wang, J., "Understanding the k-medians problem. Proceedings of the International Conference on Scientific Computing (CSC)," *The Steering Committee of The World Congress in Computer Science, Computer Engineering and Applied Computing (WorldComp)*, pp. 219-222, 2015.
[21] Available on: http://konect.uni koblenz.de/networks/ucidata-zachary, Accessed Novamber 2016.
[22] Available on:http://konect.unikoblenz.de/networks/dolphins, Accessed Novamber 2016.
[23] Available on: http://www.personal.umich.edu/ mejn/netdata/, Accessed Novamber 2016.
[24] Available on: http://www.casos.cs.cmu.edu/computational tools/datasets/external/polbooks/index11.php, Accessed Novamber 2016.
[25] Bin, X. Jin, Q. Chunxia, Z. Xiaoxuan, H. Bianjia, X & Yanfei, S., "Hybrid self-adaptive algorithm for community detection in complex networks," *Hindawi Publishing Corporation, Mathematical Problems in Engineering*, pp. 1-12, 2015.