A Modified Particle Swarm Optimization Algorithm for Community Detection in Complex Networks

Alireza Abdollahpouri, Shadi Rahimi, Shahnaz Mohammadi Majd, Chiman Salavati

▶ **To cite this version:**

Alireza Abdollahpouri, Shadi Rahimi, Shahnaz Mohammadi Majd, Chiman Salavati. A Modified Particle Swarm Optimization Algorithm for Community Detection in Complex Networks. 2nd International Cross-Domain Conference for Machine Learning and Knowledge Extraction (CD-MAKE), Aug 2018, Hamburg, Germany. pp.11-27, 10.1007/978-3-319-99740-7_2 . hal-02060038

HAL Id: hal-02060038
https://inria.hal.science/hal-02060038

Submitted on 7 Mar 2019

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Distributed under a Creative Commons Attribution 4.0 International License

A Modified Particle Swarm Optimization Algorithm for Community Detection in Complex Networks

Alireza Abdollahpouri[1], Shadi Rahimi[1],
Shahnaz Mohammadi Majd[2], Chiman Salavati[1]

[1] Department of Computer Engineering, University of Kurdistan, Sanandaj, Iran
[2] Department of Mathematics, Islamic Azad University of Sanandaj

Abstract. Community structure is an interesting feature of complex networks. In recent years, various methods were introduced to extract community structure of networks. In this study, a novel community detection method based on a modified version of particle swarm optimization, named PSO-Net is proposed. PSO-Net selects the modularity Q as the fitness function which is a suitable quality measure. Our innovation in PSO algorithm is changing the moving strategy of particles. Here, the particles take part in crossover operation with their personal bests and the global best. Then, in order to avoid falling into the local optimum, a mutation operation is performed. Experiments on synthetic and real-world networks confirm a significant improvement in terms of convergence speed with higher modularity in comparison with recent similar approaches.

Keywords: Community detection, Complex network, PSO, Modularity

1 Introduction

Most of real-world complex systems can be represented as complex networks. Social networks such as Facebook, collaboration networks such as scientific networks, technological networks such as the Internet and biological networks such as protein interaction networks are only some examples. Networks are modeled as graphs, where vertices represent individual objects and edges indicate relationships among these objects. One of the important properties of complex networks is "community structure"[1]. The term community is considered as a group of nodes within a graph with more internal connections than external connections to the rest of the network [2]. The detection of community structure, is a great important research topic in the study of complex networks, because it can detects the hidden patterns existing in complex systems. Therefore, a significant amount of efforts have been devoted to develop methods that can extract community structures from complex networks [1, 3-6].

Fortunato in [7] studied the community discovery methods in detail and divided them into several categories. Although special strategies adopted are different, most

of the algorithms are mainly divided into two basic categories including: hierarchical clustering methods [1, 3-6, 8-11] and optimization based methods [12-21]. In hierarchical clustering, a network is grouped into a set of clusters in multiple levels, which each level presents a particular partition of the network. Hierarchical clustering methods can be further divided into two groups, depending on how they build the clustering tree: divisive algorithms [1, 4, 6, 9] and agglomerative algorithms [3, 8, 11, 22]. In divisive methods, which is a top-down approach, in each iteration, the graph is divided into two groups. This process is continued until each node is assigned by a distinct cluster label. On the other hand, in agglomerative approaches (i.e., bottom-up methods), clusters are iteratively merged if their similarity is sufficiently high.

In optimization based algorithms, the community detection task is transformed into an optimization problem and the goal is to find an optimal solution with respect to a pre-defined objective function. Network modularity employed in several algorithms [1, 3, 23] and cut criteria adopted by spectral methods [24, 25], are two examples of objective functions. Evolutionary algorithms (EAs) have been successfully applied to identify community structures in complex networks [14, 19, 20]. Genetic algorithm (GA) as a well-known EA, have been frequently used for community detection among the other EA methods [15, 17, 19, 20, 26, 27]. The existing GA-based algorithms have some advantages such as parallel search and some drawbacks such as slow convergence [28]. Also, it has been shown that the GA may stick at local optimal solution and therefore, can hardly find the optimal solution [27]. There are also some challenging problems regarding GA based community detection methods such as discovering reasonable community structure without prior knowledge, and further improvement of the detection accuracy. On the other hand, swarm intelligence-based methods such as particle swarm optimization (PSO) have been successfully used in the literature to solve optimization problems [29]. PSO is a global search method which is originally developed by Kennedy and Eberhart and inspired by the paradigm of birds flocking [29]. PSO initialize the system with a population of random particles. Each particle keeps track of its coordinates in space which are associated with the best solution it has obtained (local optima) and the best solution of the population (global optima). The particles in any movement try to minimize their distances from these two positions. PSO has the advantage of easy implementation and inexpensive computationally for many problems.

In this paper, a novel PSO based approach, called PSO-Net is proposed to discover communities in complex networks. PSO-Net explores the search space without the need to know the number of communities in advance. In the proposed method a specific modularity measure is used to compute the quality of discovered communities, and then a PSO based search process is employed to explore the search space. In PSO-Net two crossover operators are applied to update particle positions and then a mutation operator is used to spread the solutions through the search space. Experiments on a synthetic and several well-known real-world networks such as Zachary's Karate Club network, the Dolphin social network, American College Football and the Books about US politics network, show the capability of the PSO-Net method to correctly detect communities with better or competitive results compare with other approaches.

The rest of this paper is organized as follows. Section 2 describes the description of the problem and related research on community detection. In Section 3, the proposed modified particle swarm optimization algorithm (PSO-Net) for community detection is presented. Section 4 presents the experimental results on synthetic and real world networks with their related analysis, and finally, the conclusion is provided in Section 5.

2 Community Definition and Related Works

2.1 Community definition

Let us consider a network N which is modeled as a graph $G = (V, E)$, where V denotes a set of nodes, and E is a set of edges linking each two nodes. Community is defined as a group of nodes (sub-graph) that has more intra-edges than inter-edges. Most formal definition for community has been introduced in [2]. Suppose that, adjacency matrix of G is A, where the element a_{ij} is 1 if there is an edge between node i and node j, and 0 otherwise. The degree of node i is defined as $k_i = \sum_j a_{ij}$. Suppose, the node i is placed to a sub-graph $S \subset G$, the degree of i with respect to S can be split as $k_i(s) = k_i^{in}(s) + k_i^{out}(s)$, where $k_i^{in}(s)$ is the number of edges connecting i to the nodes of S, and $k_i^{out}(s)$ is the number of edges connecting node i the outside of S (i.e., $G \backslash S$).

2.2 Related works

In recent years, community detection methods have been successfully applied in different research areas such as sociology, physics, biology, and computer science [1-4, 15, 18, 20, 23]. Community detection methods can be divided into two approaches including; hierarchical and optimization-based approaches. As mentioned previously, hierarchical clustering method groups data objects into a tree of clusters to produce multilevel clustering. This type of clustering is further divided to divisive and agglomerative methods. In divisive methods, a given graph is split iteratively into smaller and smaller subgraphs. Up to now, several divisive methods have been proposed in the literature. For example, the Girvan-Newman (GN) algorithm proposed in [1, 4] is a divisive method that extracts the network's communities removing the edges with the highest value of edge betweenness. This process is continued until the graph is divided into two separate subgraphs. The betweenness of an edge is defined as the number of shortest paths which are passing from that edge [30, 31]. A variation of GN algorithm is proposed by Fortunato et al. in [9]. In their method, the concept of information centrality [32] as a way to measure edge centrality, is uses instead of edge betweenness . In their method, communities are discovered by repeatedly identifying and removing the edges with the highest centrality measure. In [6], another divisive algorithm is proposed to find communities based on the principle of GN method. In order to quantify the relevance of each edge

in a network, the authors applied three edge centralities based on network topology, walks and paths, respectively.

Agglomerative hierarchical clustering is a bottom-up clustering method. Till now, several agglomerative graph clustering methods have been proposed in the literature. For example, in [3] an agglomerative clustering algorithm called Fast-Newman (FN) is proposed. In this method, a modularity measure is used to merge clusters iteratively until there is no improvement in modularity. Another example of this type of clustering, is the method proposed in [8]. This algorithm begins with a community division using prior knowledge of the network structure (degrees of the nodes), and then combines the communities as an iterative optimization process for modularity until a clear partition is obtained.

On the other hand, optimization based methods employ an objective function in their processes to evaluate the quality of found clusters. This process is continued until an optimal clustering result is found in the whole solution space. For instance in [4] an objective function called Q-modularity is used in community detection process. In this case, the community detection becomes a modularity optimization problem. In general, the obtained communities are more accurate when the value of Q is larger. Also, Brandes et al. in [33] showed that searching for the optimal modularity value is a NP-complete problem and therefore, it cannot be solved in polynomial time. Thus, many metaheuristic algorithms such as: ant colony optimization [16, 34], genetic algorithm [17, 19, 20, 27] and Extremal Optimization (EO) [13] and other metaheuristic algorithms [12, 14, 21] have been applied to solve community detection problem.

Generally, the metaheuristic methods are defined as an iterative process which employing a learning strategy to effectively explore the search space. Several metaheuristic based methods have been proposed to identify communities in complex networks. For example, taking advantage of genetic algorithm, Pizzuti proposed a new algorithm (named GA-Net), for this purpose [19]. This approach introduced the concept of community score to measure the quality of identified communities. Shang et al. [27] proposed an improved genetic algorithm for community discovery method based on the modularity concept. The computational complexity of this method is very high compare to the traditional modularity-based community detection methods. To overcome this problem, Liu et al. in [34] proposed an ant colony optimization based method for community discovery. The authors employed movement, picking-up and dropping-down operators to perform node clustering in email networks. The authors of [20] proposed a multiobjective approach for community discovery, considering both *community score* and *community fitness* concepts as its objectives. In [21], a hybrid algorithm based on PSO and EO was proposed by employing a special encoding scheme based Ji et al. proposed an ant colony clustering algorithm with an accuracy measure to identify communities in complex networks. Their algorithm focuses on the strategy of ant perception and movements and the method of pheromone diffusion and updating, and searches for an optimal partitioning of the network by ant colony movements [16].

3 Proposed Method

In this section, the proposed community detection method called PSO-Net is described in detail. The proposed method consists of two main steps including; *Initialization* and *Moving*. In initialization step, first a suitable representation for a solution which demonstrates a partitioning of a network is considered. Afterward, the solutions are randomly initialized. Then in the next step, inspired from PSO search strategy, the solutions are moved around the search space to optimize an objective (modularity) function. In the search process of the proposed method, the solutions are moved toward local and global best solutions which are performed by means of a specific crossover operator. Moreover, in order to expand the solution space, a random mutation operation is performed on each particle. The pseudo code of the proposed method is shown in Algorithm 1. Additional details of the steps in proposed method are described in their corresponding sections.

3.1 Initialization

The proposed method exploits the locus-based adjacency representation (LAR) [35]. In LAR scheme, each solution considered as an array of N genes, each of which belongs to a node and each gene takes its values in the range of $1,2,\dots,N$. Each solution represents a new graph which the value of j for the gene i, means that, there is a link between node i and node j in this graph and each connected component represent a cluster. For example, Fig. 1, illustrates LAR scheme for a network with seven nodes. In Fig. 1(a) the graph structure of the network is drawn. Fig. 1(b), shows a solution that was represented by the LAR scheme. As can be seen, for each gene, a value in the range of 1 to N is assigned. According to the Fig. 1(c), the seventh node with position of 6, takes the value of 5, meaning that, in corresponding graph, there is a link from node 6 to node 5. Thereupon, these two nodes are placed in a same cluster, which can be seen in Fig. 1(d).

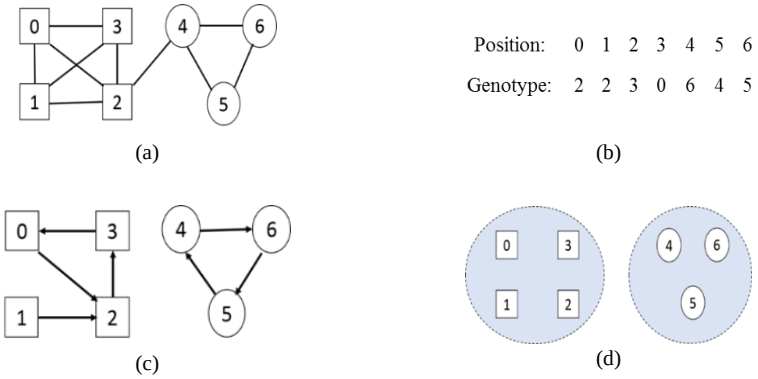


Fig. 1 Locus -based adjacency representation. (a) The topology of the graph. (b) One possible genotype. (c) Translation of (b) to the graph structure. (d) The community structure

Algorithm1. Particle swarm based community detection (PSO-Net)

Input A complex network modeled by $G = (V, E)$
I: Number of iteration that algorithm repeated
p: Number of particles
$@Init$: Function that generates particles with given number by Locus-Based Representation Scheme.
$@Decoder$: Function that decodes a given particle and resulted a partitioning of a given graph.
$@Modularity$: Function that computes modularity of a given clustering.
$@TwoPointCrossover$: Function that performs a two point crossover between two given particles
$@Mutation$: Function that mutates a given particle

Output $C = \{Cluster_1, Cluster_2, \dots, Cluster_k\}$

Begin algorithm
$Population = \emptyset, Clusterings = \emptyset, Fitness = \emptyset, Personalbest = \emptyset, Globalbest = \emptyset$
Run $@Init$ by p parameter to produce p particles and assign to $Population$ array.
Apply $@Decoder$ to obtain community structure, $\forall\ i = 1 \dots p$ and then, assign to $Clusterings$ array.
Apply $@Modularity$ to set the modularity of solutions,$\forall\ i = 1 \dots p$ and then, assign to $Fitness$ array.
for *i=1 to p*
    $Personalbest(i) = Population(i)$
end for
Find the particle with maximum value in $Fitness$ array and assign to $Globalbest$
for *i=1 to I*
    for $j = 1$ *to p*
        $Temp\_Particle = \emptyset$, $Child_1 = \emptyset$, $Child_2 = \emptyset$, $Clus_1 = \emptyset$, $Clus_2 = \emptyset$, $Temp_{Clus} = \emptyset$
        Initialize $Child_1$ and $Child_2$ by applying $@TwoPointCrossover$ on $Population(i)$ and $Personalbest(i)$
        Initialize $Clus_1$ and $Clus_2$ by applying $@Decoder$ on $Child_1$ and $Child_2$
        Apply $@Modularity$ to compute the modularity of $Clus_1$ and $Clus_2$
        Select the Child with maximum modularity among $Clus_1$ and $Clus_2$ and then, assign to $Temp\_Particle$
        Update $Child_1$ and $Child_2$ by applying $@TwoPointCrossover$ on $Temp\_Particle$ and $Globalbest$
        Update $Clus_1$ and $Clus_2$ by applying $@Decoder$ on $Child_1$ and $Child_2$
        Apply $@Modularity$ to compute the modularity of $Clus_1$ and $Clus_2$
        Select the Child with maximum modularity among $Clus_1$ and $Clus_2$ and then, assign to $Temp\_Particle$
        Apply $@Mutation$ on $Temp\_Particle$
        $Population(i) = Temp\_Particle$
        Apply $@Decoder$ on $Population(i)$ and then, assign to $Clustering(i)$
        Apply $@Modularity$ to compute modularity of $Clustering(i)$, and then, update $Fitness(i)$
        if $(Fitness(i) \geq Fitness(Personalbest(i)))$ **then**
            $Personalbest(i) = Population(i)$
        end if
    end for
    Update $Globalbest$ by the particle that has highest maximum $Fitness$ value
end for
Return $Globalbest$
End algorithm

The LAR encoding scheme has some benefits. First, it is dispensable to determine the number of communities in advance, because of automatically determination in the decoding step. Besides, the decoding process can be done in a linear time. Then, standard crossover operators can be easily employed over these types of representation. To initialize the system, a population of random individuals is generated such that for each node i, the value of g_i is randomly chosen among one of its neighboring nodes , which indicates the edge (i, j) in the graph. This type of initialization improves the convergence of the algorithm, due to restriction of the solution space.

3.2 Search strategy

In order to move each solution towards the best positions, we use genetic operators, i.e., crossover and mutation operators as follows.

Moving toward personal best. At first, for each particle a two-point crossover with its personal best is performed and then as a result, two new solutions are obtained. For example, given two parents P_1 and P_2 and two random points i and j, binary string from beginning of chromosome to the crossover point i is copied from parent P_1, the part from crossover point i to the crossover point j is copied from the parent P_2 and the rest is copied from the parent P_1. This action creates the first child. To produce the second child, this action is done in reverse order. (See Fig. 2). Finally, a solution with higher fitness value, i.e., higher modularity, is selected as a temporary position of current particle.

Moving toward global best. To move towards the global best, a two-point crossover is performed between a particle and the global best of population. In this case, two new solutions are obtained. The one with a higher modularity value is selected as temporary state of current particle.

3.3 Enhancing search ability

Finally, to move the solutions around the whole search space, one-point neighbour-based mutation is performed on all particles. Such that, for each particle, a gene $\boldsymbol{i}$ is picked randomly and the possible values for this gene are limited to its neighbours to guarantee that solution space has only possible solutions.

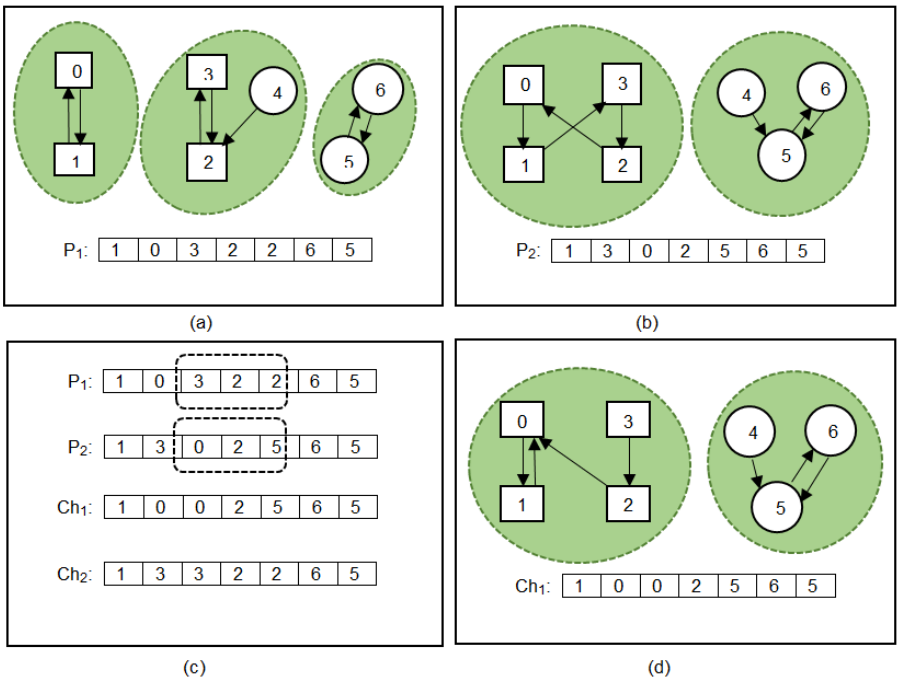


Fig. 2 Two point crossover. (a) P1 and corresponding graph structure. (b) P2 and corresponding graph structure. (c) A random two-point crossover of the genotypes yields the children Ch1 and Ch2. (d) Ch1 and its graph structure.

3.4 Fitness Computation

Modularity of a network [4], measures the goodness of identified communities. A quantitative definition of the modularity can be the fraction of the edges that fall within the clusters minus the anticipated value of this fraction while edges fall at random in a network regardless of the community structure. Let k be the number of clusters found inside a network, the modularity Q is defined as (Eq. 1).

$$Q = \sum_{s=1}^{k} \left[\frac{l_s}{m} - \left(\frac{d_s}{2m} \right)^2 \right] \tag{1}$$

Where, l_s is total number of edges connecting vertices inside the cluster of s, and d_s is the sum of the degrees of nodes of s, and m is the total number of edges in the network. The possible values for this criterion is in the range of [-0.5, 1] and for most real-networks this value is in the range of [0.3, 0.7]. Actually, values larger than 0.3, indicate a meaningful community structure.

4 Experimental Results

In this section, we study the effectiveness of our approach and compare the results obtained by PSO-Net w.r.t. the algorithms of GA-Net, FN and FC on the Girvan-Newman benchmark and then on real-world networks including the Zachary's Karate Club network, the American College Football network, the Bottlenose Dolphin network and the Books about US Politics network. Moreover, the proposed method was compared to three community detection methods which are listed below:

- *Fast Newman* (FN) [3] is an agglomerative hierarchical method which aims to maximizing modularity of obtained communities.

- *GA-Net* [19] is an optimization-based community detection method, which adopts Genetic Algorithm to optimize the community score measure.

- *Fuzzy Clustering* (FC) [5] is a community detection method based on fuzzy transitive rules. This method uses the edge centralities such as edge betweenness centrality to measure the similarity among nodes of a network. Then, by forming a fuzzy relation on the network and applying transitive rules on the relation, when the relation achieve to the stable state, the clusters are discovered. In this study, we report the best results obtained by this method.

4.1 Parameter setting

The PSO-Net algorithm was implemented in visual studio 2010. The experiments have been performed on a computer having Intel® Core™ i5 CPU 2.67 GHz and 4 GB (3.9 GB usable) of memory. The number of generations for all data sets in both PSO-Net and GA-Net was set to 100. The population size is customized according to the size of data sets. In this way, size of population for karate club network is 100, for dolphin network is 200, and for football network, Political Books network and Girvan-Newman benchmark are set to 400. We used the following parameters for implementation of GA-Net: crossover rate of 0.8, mutation rate of 0.2, and tournament selection function. Since PSO-Net and GA-Net algorithms, are the random optimization methods, all the results obtained from these two methods are computed over 10 independent runs.

4.2 Evaluation metrics

In order to compare PSO-Net and other approaches, two measures, the normalized mutual information (NMI) [36] and Modularity [4], mentioned is Section 3.3, are used. NMI criterion is employed to measure the similarity between the real community structure of a network and the structure detected by the proposed method.

Assume two different types of partitioning for a network $A = \{A_1, \dots, A_R\}$ and $B = \{B_1, \dots, B_D\}$, that R and D are the number of communities in the partitioning A and B, respectively. A confusion matrix C is formed first, where an entry C_{ij} is the number of nodes that appear in both communities $A_i \in A$ and $B_j \in B$. Then, normalized mutual information $NMI\ (A, B)$ is defined as (2):

$$NMI\left(A,B\right)=\frac{-2\sum_{i=1}^{R}\sum_{j=1}^{D}C_{ij}\log\left(C_{ij}N\ /C_{i.}C_{.j}\right)}{\sum_{i=1}^{R}C_{i.}\log\left(C_{i.}/N\right)+\sum_{j=1}^{D}C_{.j}\log\left(C_{.j}/N\right)} \quad (2)$$

where $C_{i.}$ ($C_{.j}$) is the sum of the elements of C, over row i (column j), and N is the total number of nodes in the graph. NMI value of 1 indicates that A and B are exactly equal.

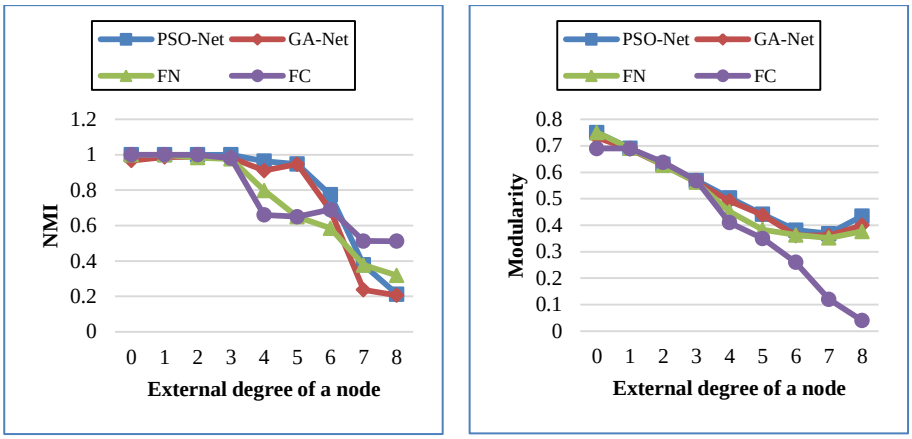
4.3 Experimental results in synthetic datasets

The most famous benchmark for community detection is the Girvan-Newman (GN) networks [1]. Each network has 128 nodes, divided into four communities of 32 nodes. The average degree of this type of networks is equal to 16. The nodes are connected together in a random order, but in such a way, that $k_{in} + k_{out} = 16$, which k_{in} and k_{out} are the internal and external degree of a node, respectively.

Increasing the value of k_{out} leads to more connections between the nodes of different communities, and therefore, the correct detection of communities becomes more difficult. Thereupon, in this case, the resulting graphs pose greater challenges to the community mining methods. Fig. 3(a) shows the average NMI value over 10 independent runs, obtained by each algorithms for different values of k_{out}. As can be seen, for the values of k_{out} less than seven, PSO-Net gets higher NMI value. When k_{out} is 7, performance of PSO-Net is worse than FC. For the k_{out} value of 8, GA-Net and PSO-Net obtain the least NMI value, respectively. It can be concluded that our approach has better performance in detecting communities of networks with more clear clusters.

Another measure that should be investigated is modularity. As can be seen in Fig. 3(b), for all values of k_{out}, the modularity for our proposed method is highest, which means that, the community structure resulted by PSO-Net is more modular than other three approaches. Similarly, in this case, the modularity results are obtained from an average of 10 runs.

In Table 1, the average number of clusters that each of the four algorithms returns over 10 run, is reported. As can be seen, our method for the values of k_{out} in the range of [0-4], divides the GN benchmark into 4 clusters which exactly is equal to true number of communities. For other values of k_{out}, PSO-Net detects the reasonable number of communities in comparison with other methods.



(a) (b)

Fig. 3 Comparison of PSO-Net, GA-Net, FN and FC in terms of (a) NMI and (b) Modularity on the Girvan-Newman benchmark.

Table 1 Number of Communities detected by four methods on GN benchmark, for different values of k_{out}

	0	1	2	3	4	5	6	7	8
PSO-Net	4	4	4	4	4	4.2	3.9	4.7	5.9
GA-Net	4.8	4.3	4.1	4.4	5.5	4.8	6.5	6.6	8.1
FN	4	4	4	4	3.6	3.1	3.2	3.5	3.6
FC	4	4	4	5	2	2	21	70	78

4.4 Experimental results in real-world datasets

We now show the application of PSO-Net on two popular real-world networks, the *Zachary's Karate Club*, and the *American College Football*, and compare our results with GA-Net, FN and FC methods.

Zachary's Karate Club network, studied by Zachary, is a social network of friendships between 34 members of a karate club at a US university in 1970. During the course of Zachary's study, because of disagreements, the club divided in two groups about of the same size. And each of these two groups, are clustered in two subgroups. The community structure of this network is shown in Fig. 4.

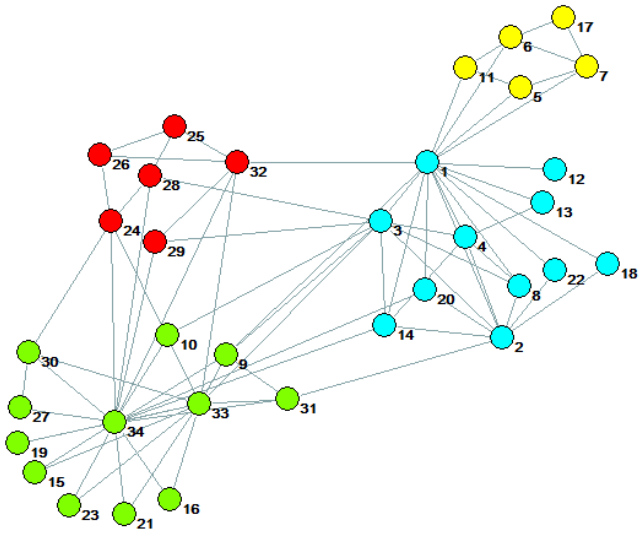


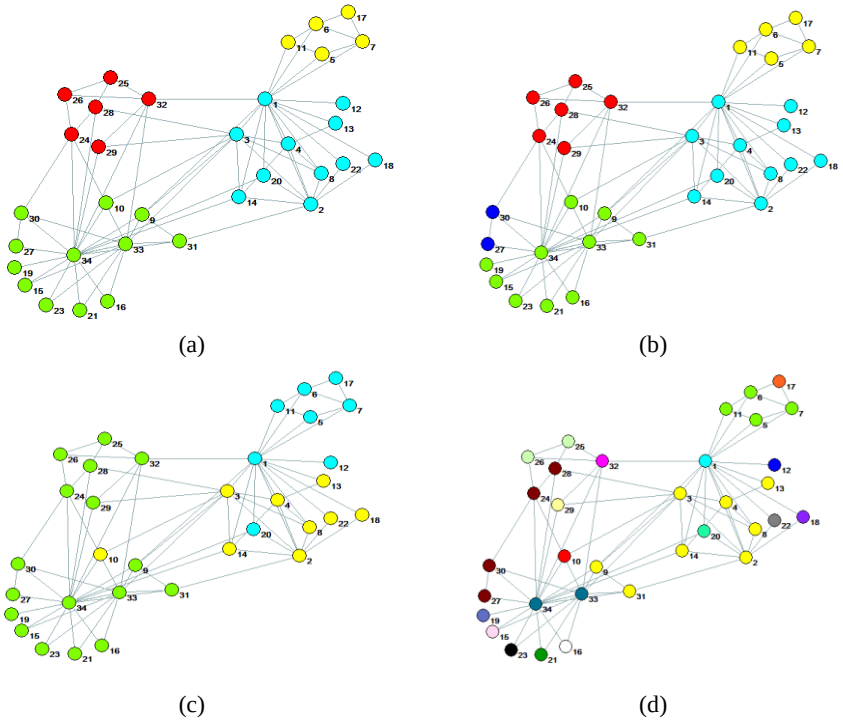
Fig. 4 Zachary's karate club network

Table 2 shows the detailed comparative results of the various algorithms on the Karate network. For each algorithm, we have listed the NMI measure, modularity measure and then the number of communities. As can be seen, the average and best NMI values of PSO-Net are superior to that of other algorithms. GA-Net provides smaller standard deviation than PSO-Net, but the difference between these values, is negligible. Moreover the average and best Modularity values of our method, are higher than other algorithms. Also, standard deviation of our method for modularity is smaller than GA-Net. The column of average number of detected communities, shows that, except FC algorithm, other methods, provide the number of clusters that are near to the real one.

Table 2 Results obtained by four algorithms on Zachary's Karate Club network.

Method	NMI				Modularity				Num. of Com.
	best	average	worst	std	best	average	worst	std	average
PSO-Net	**1**	**0.88**	0.60	0.100	**0.42**	**0.40**	0.37	**0.01**	3.7
GA-Net	0.94	0.80	**0.69**	0.096	0.4،	0.37	0.29	0.03	4.2
FN	0.63	0.63	0.63	-	0.38	0.38	**0.38**	-	3
FC	0.56	0.56	0.56	-	0.13	0.13	0.13	-	19

It can be seen from Fig. 4 and Fig. 5 that, the detected community structure of our method on Zachary's karate club network, is the real community structure. But the detected structure of other methods, are different from the true one.



(a) (b) (c) (d)

Fig. 5 The detected communities of best result of (a) PSO-Net, (b) GA-Net, (c) FN and (d) FC on Zachary's karate network

The American College Football network is a network with 115 nodes and 616 edges that grouped in 12 communities. The vertices represent teams and the edges indicate the season games between nodes in the year. The real communities of this network are shown in Fig. 6.

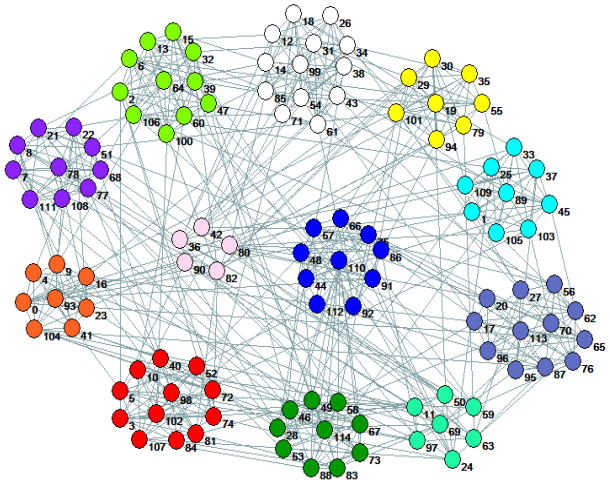


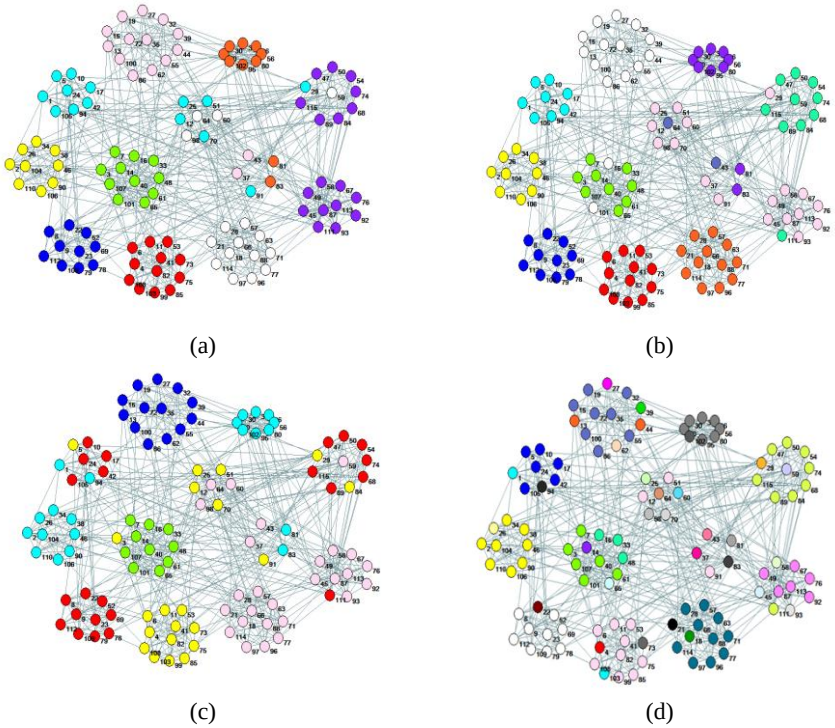
Fig. 6 American college football network

In Table 3, the results of four algorithms on this network are reported. As can be seen, PSO-Net has the highest average and the best NMI values after GA-Net. But standard deviation of our method is smaller than GA-Net. The modularity value for PSO-Net in three cases (best, average and worst) is the highest among all methods and the standard deviation of our method is smaller. GA-Net and PSO-Net, extract the closer number of clusters to real structure, respectively.

Table 3 Results obtained by the four algorithms on American College Football network.

Method	NMI				Modularity				Num. of Com.
	best	average	worst	std	best	average	worst	std	
PSO-Net	**0.89**	**0.82**	**0.80**	**0.033**	**0.52**	**0.52**	**0.51**	**0.003**	4
GA-Net	0.71	0.63	0.59	0.035	0.46	0.44	0.39	0.022	9.6
FN	0.70	0.70	0.70	-	0.49	0.49	0.49	-	4
FC	0.49	0.49	0.49	-	0.30	0.30	0.30	-	17

The best results of the four algorithms on football network are shown in Fig. 7. As can be seen, from Figs. 6 and 7, the community structure discovered by FC method, is very different from the true one. But, other approaches detect similar structure to real community structure on football network.

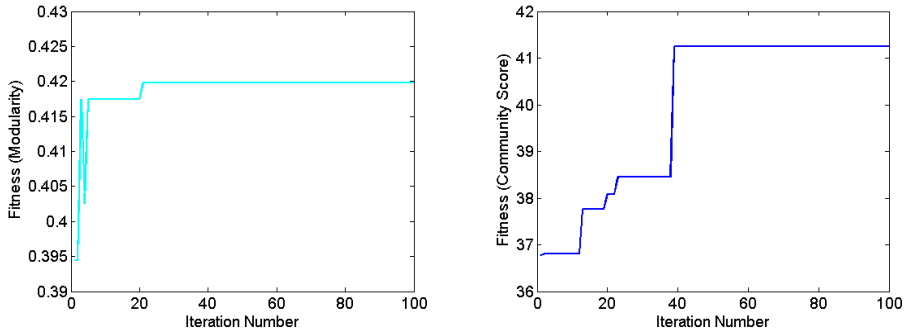


(a) (b) (c) (d)

Fig. 7 Detected communities of best result of (a) PSO-Net, (b) GA-Net, (c) FN and (d) FC on Football network

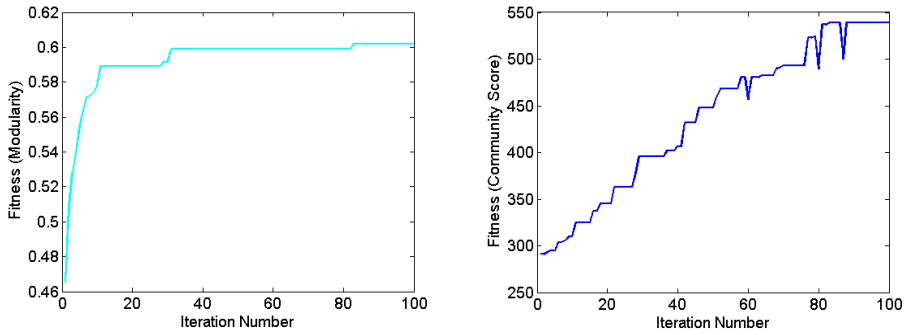
5 Coverage analysis for the proposed algorithm

In this Section, we investigate the convergence rate of our algorithm and another random optimization algorithm, i.e., GA-Net on real-world networks. It is worth noting that the fitness functions of two methods are different, and we just compare the convergence points in these methods. Fig. 8(a) and Fig. 8(b) show the speed of convergence of GA-Net and PSO-Net, for karate club network, respectively. As can be seen, GA-Net in the iteration number of 39, achieves to maximum value of its objective function. However, PSO-Net converges in 21st iteration. That means, convergence rate of our method for karate network is better. It is worth mentioning that the NMI and Modularity of PSO-Net in discovering community structure of this network were largest among all methods. Fig. 9(a) shows convergence rate of GA-Net for football network. As can be seen, this algorithm achieved to maximum value of fitness in 88^{th} iteration. In Fig. 9(b), we can see that, PSO-Net converges in 83rd iteration for football network. Here, the difference is not significant.



(a) (b)

Fig. 8 Comparison of convergence rate of (a) PSO-Net and (b) GA-Net in karate club network



(a) (b)

Fig. 9 Comparison between convergence rate of (a) PSO-Net (b) GA-Net on American College Football network

6 Conclusion

In this paper, a novel community detection method based on particle swarm optimization (PSO) algorithm named PSO-Net has been proposed. We focus on the modification of the PSO. In our method, the particles for approaching to their local and the global best, take part in crossover operation with them. Then, for spreading search space, a mutation operator is performed on each particle. The algorithm takes modularity measure as its fitness function. Experiments on synthetic and real world networks showed that PSO-Net has good results in discovering communities of these networks, especially, in karate club network. Moreover, the convergence rate of PSO-Net in comparison with GA-Net is very faster. In the future, we will aim to applying multi-objective optimization to improve quality results.

References

[1] M. Girvan, M.E. Newman, Community structure in social and biological networks, Proceedings of the national academy of sciences, 99 (2002) 7821-7826.
[2] F. Radicchi, C. Castellano, F. Cecconi, V. Loreto, D. Parisi, Defining and identifying communities in networks, Proceedings of the National Academy of Sciences of the United States of America, 101 (2004) 2658-2663.
[3] M.E. Newman, Fast algorithm for detecting community structure in networks, Physical review E, 69 (2004) 066133.
[4] M.E. Newman, M. Girvan, Finding and evaluating community structure in networks, Physical review E, 69 (2004) 026113.
[5] P.G. Sun, Community detection by fuzzy clustering, Physica A: Statistical Mechanics and its Applications, 419 (2015) 408-416.
[6] P.G. Sun, Y. Yang, Methods to find community based on edge centrality, Physica A: Statistical Mechanics and its Applications, 392 (2013) 1977-1988.
[7] S. Fortunato, Community detection in graphs, Physics reports, 486 (2010) 75-174.
[8] H. Du, M.W. Feldman, S. Li, X. Jin, An algorithm for detecting community structure of social networks based on prior knowledge and modularity, Complexity, 12 (2007) 53-60.
[9] S. Fortunato, V. Latora, M. Marchiori, Method to find community structures based on information centrality, Physical review E, 70 (2004) 056104.
[10] P.G. Sun, Weighting links based on edge centrality for community detection, Physica A: Statistical Mechanics and Its Applications, 394 (2014) 346-357.
[11] J.J.C.D.C. Fain, K.J.L.L. Zhukov, Clustering of bipartite advertiser-keyword graph.
[12] B. Amiri, L. Hossain, J.W. Crawford, R.T. Wigand, Community detection in complex networks: Multi–objective enhanced firefly algorithm, Knowledge-Based Systems, 46 (2013) 1-11.
[13] J. Duch, A. Arenas, Community detection in complex networks using extremal optimization, Physical review E, 72 (2005) 027104.
[14] M. Gong, Q. Cai, X. Chen, L. Ma, Complex network clustering by multiobjective discrete particle swarm optimization based on decomposition, IEEE Transactions on Evolutionary Computation, 18 (2014) 82-97.
[15] M. Gong, L. Ma, Q. Zhang, L. Jiao, Community detection in networks by using multiobjective evolutionary algorithm with decomposition, Physica A: Statistical Mechanics and its Applications, 391 (2012) 4050-4060.
[16] J. Ji, X. Song, C. Liu, X. Zhang, Ant colony clustering with fitness perception and pheromone diffusion for community detection in complex networks, Physica A: Statistical Mechanics and its Applications, 392 (2013) 3260-3272.
[17] S. Li, Y. Chen, H. Du, M.W. Feldman, A genetic algorithm with local search strategy for improved detection of community structure, Complexity, 15 (2010) 53-60.
[18] C. Liu, J. Liu, Z. Jiang, A multiobjective evolutionary algorithm based on similarity for community detection from signed social networks, IEEE transactions on cybernetics, 44 (2014) 2274-2287.

[19] C. Pizzuti, Ga-net: A genetic algorithm for community detection in social networks, Parallel Problem Solving from Nature–PPSN X, (Springer, 2008), pp. 1081-1090.
[20] C. Pizzuti, A multiobjective genetic algorithm to find communities in complex networks, IEEE Transactions on Evolutionary Computation, 16 (2012) 418-430.
[21] J. Qu, A Hybrid Algorithm for Community Detection Using PSO and EO, Advances in Information Sciences and Service Sciences, 5 (2013) 187.
[22] L. Donetti, M.A. Munoz, Detecting network communities: a new systematic and efficient algorithm, Journal of Statistical Mechanics: Theory and Experiment, 2004 (2004) P10012.
[23] A. Clauset, M.E. Newman, C. Moore, Finding community structure in very large networks, Physical review E, 70 (2004) 066111.
[24] J. Shi, J. Malik, Normalized cuts and image segmentation, IEEE Transactions on pattern analysis and machine intelligence, 22 (2000) 888-905.
[25] L. Hagen, A.B. Kahng, New spectral methods for ratio cut partitioning and clustering, IEEE transactions on computer-aided design of integrated circuits and systems, 11 (1992) 1074-1085.
[26] A. Gog, D. Dumitrescu, B. Hirsbrunner, Community detection in complex networks using collaborative evolutionary algorithms, European Conference on Artificial Life, (Springer2007), pp. 886-894.
[27] R. Shang, J. Bai, L. Jiao, C. Jin, Community detection based on modularity and an improved genetic algorithm, Physica A: Statistical Mechanics and its Applications, 392 (2013) 1215-1231.
[28] D. Abramson, J. Abela, A parallel genetic algorithm for solving the school timetabling problem, (1991).
[29] R. Eberhart, J. Kennedy, A new optimizer using particle swarm theory, Micro Machine and Human Science, 1995. MHS'95., Proceedings of the Sixth International Symposium on, (IEEE1995), pp. 39-43.
[30] J. Scott, Social network analysis (Sage, 2012).
[31] L.C. Freeman, A set of measures of centrality based on betweenness, Sociometry, (1977) 35-41.
[32] V. Latora, M. Marchiori, A measure of centrality based on the network efficiency. arxiv: con-math 0402050, (2004).
[33] U. Brandes, D. Delling, M. Gaertler, R. Görke, M. Hoefer, Z. Nikoloski, D. Wagner, On modularity clustering, Knowledge and Data Engineering, IEEE Transactions on, 20 (2008) 172-188.
[34] Y. Liu, J. Luo, H. Yang, L. Liu, Finding closely communicating community based on ant colony clustering model, Artificial Intelligence and Computational Intelligence (AICI), 2010 International Conference on, (IEEE2010), pp. 127-131.
[35] J. Handl, J. Knowles, An evolutionary approach to multiobjective clustering, IEEE transactions on Evolutionary Computation, 11 (2007) 56-76.
[36] L. Danon, A. Diaz-Guilera, J. Duch, A. Arenas, Comparing community structure identification, Journal of Statistical Mechanics: Theory and Experiment, 2005 (2005) P09008.