Journal of Advances in Computer Research
Quarterly ISSN: 2008-6148
Sari Branch, Islamic Azad University, Sari, I.R.Iran
(Vol. 13, No. 3, August 2022), Pages: 87-103
www.jacr.iausari.ac.ir

# An Improved Symbiotic Organisms Search for Community Detection in Social Networks

**Yahya Ghanbarzadeh Bonab**✉[1]

*1) Department of Computer Engineering, Ajabshir Branch, Islamic Azad University, Ajabshir, Iran*

ghanbarzadeh2019@gmail.com

**Received: 2022/09/21; Accepted: 2022/10/29**

**Abstract**

*Networks are a general language for representing relational information among objects. An effective way to model, reason about, and summarize networks, is to discover sets of nodes with common connectivity patterns. Such sets are commonly referred to as network communities. Research on network Community Detection (CD) has predominantly focused on identifying communities of densely connected nodes in undirected networks. Community structure is an integral part of a social network. Detecting such communities plays a vital role in a wide range of applications, including but not limited to cluster analysis, recommendation systems, and understanding the behavior of complex systems. Researchers have derived many algorithms from discovering the community structures of networks. Finding communities is a challenging task, and there is no single algorithm that produces the best results for all networks. Therefore, despite many elegant solutions, learning communities remain active research areas.*

*CD is a challenging optimization problem that consists in searching for communities that belong to a network or graph under the assumption that the nodes of the same community share properties that enable the detection of new characteristics or functional relationships in the network. Many methods have been proposed to address this problem in many research fields, such as power systems, biology, sociology, and physics. Many of those optimization methods use modularity to identify the optimal network subdivision. This paper proposes a new CD approach based on Symbiotic Organisms Search (SOS) and Lévy Flight (LF). The LF distribution is used to prevent the stagnation of solutions in local minima. Extensive experiments compare the SOS-LF with other state-of-the-art algorithms on real-world social networks. Experimental results show that the SOS-LF is effective and stable.*

***Keywords:*** *Community Detection, Symbiotic Organisms Search, Lévy Flight, Social Network Analysis*

## 1. Introduction

It is common knowledge that complex systems can be represented as complex networks in many different fields, such as social networks, biological networks, collaboration networks, and so on [1]. In these models, nodes represent the objects, and edges represent the relationships between things. Communities, an essential component of networks, do not have a universally accepted definition. Nevertheless, network communities are often considered one of the densely linked nodes compared to the rest of the networks [2]. To put it another way, the connections between nodes that are part of the same community are denser than the connections that are part of separate communities. CD is used to unearth the partitioning of networks and categorizing nodes, both of which are beneficial to understanding and efficiently using networks. For instance, finding communities in social networks may lead to discovering genuine social groups segmented according to interests or backgrounds, while finding. Finding citation networks may lead to finding papers on connected themes [3].

Information can be gained by analyzing data derived from social networks. One of the growing challenges in this field is identifying "communities" within social networks, which means identifying subsets or clusters containing the nodes (people or other entities that form the network) with unusually strong or numerous connections. Social networks are naturally modeled as graphs, where commodities are represented as nodes, and relations are represented as edges between nodes [4]. Different charts can be used to model real social networks depending on the network's characteristics. Social networks are exponentially growing in our society, significantly changing how people react to events and interact with each other. Social network analysis (SNA) is the field of studying these five social behaviors. It has been increasingly popular in the last decade with the ubiquitous use of social networks. Social network analysis relies on the use of networks to investigate social interactions. A set of interacting entities in many domains, like biology, computer science, and economy, leads to complex systems with hidden properties [5]. The categorization of vertices in a network is a common task across a multitude of domains [6]. Specifically, structural divisions into internally well-connected sets are useful in computer science, social science, and biology. In these areas, grouping vertices using structural boundaries helps one understand a network's underlying processes. Identifying such groupings is a non-trivial task and a subject of intense research in recent years [7]. CD is known to identify groups of vertices in a network based on structural properties. Methods to identify such groups take a wide variety of approaches, mirroring the diversity in domains where an accurate view of structural communities is helpful. Depending on the definition of a community, one could discover groups that maximize a global quality function, contain a specific set of substructures, or satisfy a collection of local criteria. Each of these definitions has resulted in several methods which aim to produce the "best" set of communities relative to the purpose chosen [8].

Are they allowed to overlap? In the past, the field of CD has primarily focused on identifying a set of groups such that each vertex in the network is assigned to a single group. Such a requirement results in a set of disjoint groups covering the entire network [9]. However, with the explosion of social networks and online communication data, research has expanded towards methods considering overlapping groups. In the remainder of this text, we will first include a brief discussion on the intuition behind disjoint and overlapping communities and provide the reader with a basic understanding of a small sample of commonly used methods for CD.

Further into the text, we will present the difficulties when detecting overlapping communities and introduce a method for discovering overlapping communities that avoids these common pitfalls. This algorithm will be presented with results on natural and synthetic benchmark networks. Finally, we will show that communities are natural and necessary to capture many associations between vertices in a network in real data.

The process of identifying clusters of linked nodes belonging to the same category inside a network is called CD [10]. The pervasive existence of community structures in complex networks such as online social networks, biological networks, and cooperation networks reveals valuable information about these types of networks and the entities that make them up. CD is one of the most important study subjects in the field of network science and has a wide variety of applications in the actual world. Let us assume that there is a network represented by the graph $G = (V, E)$, where $V = \{v_1, v_2, \dots, v_n\}$ is the set of nodes and $E = \{e_{ij}\}_{ij}^{n} = 1$ is a representation of the vertices in the network. The challenge of community detection (CD) is defined as the task of subdividing a whole graph (G) into a collection of k communities $C = \{C_1, C_2, \dots, C_k\}$. Each node in the graph is a member of one of the communities, and these communities reflect the topological particulars of the community at the local level [11].

The Symbiotic Organisms Search (SOS) [12] method is a robust new metaheuristic search algorithm [13] that has recently gained widespread applicability in resolving complex optimization issues. SOS mirrors the symbiotic connection tactics organisms employ in the environment for survival. When trying to solve optimization problems, all meta-heuristic algorithms, including SOS, run into issues like premature convergence, delayed convergence and becoming caught in the local trap. The difficulty in resolving the CD issue by using the SOS arises from the fact that the SOS solutions need modification whenever the number of problematic nodes in a network grows. If you make the SOS issue and solutions more complex, the system's performance will suffer, and it will have less capacity to find better answers. As a result, Lévy Flight (LF) [14] is required to enhance SOS in performing CD optimization in complex networks. As a result, we will use LF in this study to improve the speed and accuracy of the SOS. The technical contributions of this paper are as follows:

- A novel hybrid algorithm SOS-LF is proposed for community detection.
- Use LF to avoid falling into local optima and discover optimal solutions
- Evaluation of the proposed model on six real datasets
- Solutions by the proposed approach are compared with MFO and SSA.
- The proposed model has better detection accuracy compared to MFO and SSA.

The rest of the paper is organized as follows: Section 2 describes the related works; Section 3 describes the proposed model; Section 4 shows experiments and comparisons with other algorithms; Finally, the work is summarized in Section 5.

## 2. Related Works

The reference [13] suggests using a novel multi-objective optimization method based on the Ant Colony Optimization (ACO) to overcome the problem of CD in complicated networks. The Pareto approach is employed in the suggested model to provide optimum solutions. The goal function and the NMI are optimized to their full potential by the model given as a solution to the issue of community recognition. Using the Pareto idea and Pareto registration in the suggested model resulted in a modification to the pheromone update in ACO. As a result, the only optimization that global updates can achieve is optimizing optimum solutions. When the results of the suggested model were compared to those of other methods, it was clear that it correctly discovered network structures.

In 2015, Li and Liu presented a multi-agent Genetic Algorithm (GA) for CD in complicated networks [15]. To maximize the advantage of modularity in CD, they suggested using a multi-agent GA known as MAGA-Net. A possible answer is represented by a coded agent that was produced by a section of a network. Several operators, such as the split and merging based neighborhood competition operator, the hybrid neighborhood crossover operator, the adaptive mutation operator, and the self-learning operator, have been devised to improve the modularity value. The performance of MAGA-Net is proven in the trials by using well-known real-world benchmark networks and large-scale synthetic LFR networks with 5000 nodes. MAGA-Net surpasses GA-Net and Meme-Net, as shown by the systematic comparisons between the three algorithms. It can discover communities with great speed, accuracy, and stability. MOPSO-Net [16] is a novel multi-objective technique developed for CD. It is based on a modified version of the Particle Swarm Optimization algorithm. Objective criteria such as kernel k-means and cut rates are used for depreciation. Experiments conducted on artificial neural networks, as well as those shown in the real world, demonstrate a considerable improvement in regular mutual information and modulation in comparison to more recent techniques that are comparable.

A discrete cuckoo search optimization in dynamic networks was made available for CD use [17]. A discrete CSO was suggested as a method for detecting communities in dynamic networks to avoid the limits imposed by the parameters and improve CD accuracy. In the beginning, the nest position for the commencement of the first population is encoded using a technique called "neighbor." A discrete CSO algorithm with a method for updating the nest location that has been modified and the abandoned operator is used in the second stage of the process. In addition to this, a multi-objective CSO that was based on the discrete algorithm that was provided was developed by integrating the sorting and metering approaches. According to the findings of the experiments, the CSO algorithm is superior to other comparable algorithms in terms of its validity and precision in natural and artificial networks.

A unique CD technique, CD-SACS, has been presented [18]. This algorithm uses stacked autoencoders and a Crowd Search Algorithm-based k-means clustering algorithm. To prevent the k-means algorithm from prematurely converging into a local optimum and to enable the algorithm to explore a significantly larger global search space to obtain a solution significantly closer to the global optimum, Crowd-Search optimization is used to generate the initial centroids for the k-means. The outcomes of the suggested method were evaluated and contrasted with the results of various conventional and cutting-edge CD algorithms. The achieved findings demonstrated that the proposed model is superior to the CD algorithms used by other organizations. The fact that the modified crow search algorithm, similar to previous meta-heuristics, may also experience the issue of early convergence and falling into a local optimum is a significant drawback of the work that has been suggested. This is one of the primary limitations of the study. In addition, the approach that has been suggested does not apply to multi-layer or complex weighted networks.

The CD Algorithm based on Structural Similarity (CDASS) can identify communities in two distinct stages [19]. The first step involves randomly removing a few edges from the network graph that is not very significant to construct several linked subgraphs that are unrelated to one another. After that, the subgraphs considered central communities are united to generate the best possible community structure that is as near as possible to actual communities. In the second phase, the community structures formed in the first phase of merging are compared and contrasted to determine which community structure is the superior one.

It is argued that the TL-GSO CD method might benefit from a discrete adaptation to achieve quicker convergence of the optimization function [20]. This method utilizes a combination of the exploration tactics of Teachers Learners (I-TLBO) and GSO algorithms to identify communities inside intricate networks. It alters the new search that GSO uses to a step search, and it changes the real-time crossover to a single-point crossover. The adjustments have the effect of reducing the number of parameters that need to be set manually. The runtime may be reduced using optimized search space, and communities can develop without outside intervention. Compared to a variety of other state-of-the-art CD algorithms, the experimental findings on actual and synthetic networks demonstrated that the proposed method converges quicker and develops into accurate communities with greater fitness than those other algorithms.

PSO-Net is a revolutionary PSO-based technique developed to find communities in complicated networks [21]. PSO stands for population spectral optimization. PSO-Net investigates the search space without first requiring users to provide an estimate of the total number of communities. In the technique that has been provided, a particular modularity measure is used to calculate the quality of the communities that have been identified. After that, a PSO-based search process is used to investigate the search space. Particle locations are updated in PSO-Net by applying two crossover operators, and then a mutation operator is employed to disperse the solutions over the search space. Experiments were carried out on a synthetic network and several well-known networks from the real world, such as the network for Zachary's Karate Club, the Dolphin network, the American College Football network, and the Books on US Politics network.

In the reference [22], the authors provided a unique nature-inspired algorithmic technique based on Ant Lion Optimizer to efficiently identify communities in huge networks. The approach that has been presented can maximize the modularity function and identifies tightly connected clusters of nodes that have sparse interconnects between them. The work is evaluated using benchmarks such as Zachary's Karate Club, the Dolphins, books on politics in the United States, and the American college football network. The results are compared to the Ant Colony Optimization (ACO) and Enhanced Firefly algorithm (EFF) techniques. The strategy that has been presented outperforms EFF and ACO for Zachary and Books on US politics, and it generates outcomes that are superior to those produced by ACO for Dolphins and EFF for American Football Club.

To find more optimum and stable solutions, a unique modularity-based discrete state transition technique known as MDSTA has been presented [23]. In addition, to do a global search, the vertex substitute transformation operator and the community substitute transformation operator have been suggested as possible solutions based on the heuristic information provided by the network. After that, each person that has been initiated will progress by way of these two alternative processes. After that, an elite population of people with high fitness values is chosen from among these developed individuals to serve as the basis for the selection. In conclusion, a two-way crossover operation is carried out among the top members of the community to complete the local search. The MDSTA has a very straightforward structure that is not difficult to put into action. Several cutting-edge CD algorithms are used to evaluate MDSTA on both artificial and real-world networks. These algorithms are tested on both kinds of networks. The findings of the experiments showed that MDSTA is an efficient and reliable technique for detecting CD in networks.

A novel model for CD on social media is offered [24], based on the Multi-objective Cuckoo Search Algorithm (MOCSA). The MOCSA model implements a novel approach to the objective function predicated on identifying near neighbors to improve the CD's accuracy and speed while operating on social networks. Eight different datasets will be used for the assessment, including Karate, Dolphins, Polbooks, Football, Email, $Geom$, $NetScience$, and Power Grid. According to the findings, the Normalized Mutual Information (NMI) value for the Karate dataset is 1.00, for the Dolphin dataset, it is 0.9984, for the Football dataset, it is 0.9486, and for the Polbooks dataset, it is 0.7455. The modularity value for the Karate dataset is 0.4192, whereas the Dolphin, Football, and Polbooks datasets are 0.5262, 0.6025, and 0.5264. The modularity of the Email dataset is 0.5362, whereas that of the $Geom$ dataset is 0.7025, that of the $NetScience$ dataset is 0.9497, and that of the Power dataset is 0.8382.

A model that is offered for CD [25] is based on the Harris Hawk Optimization (HHO) algorithm and Opposition-based Learning (OBL). An OBL is used in the model that has been presented to strike a balance between exploration and exploitation. Optimizing solutions requires a healthy equilibrium between exploratory efforts and productive production. Based on modularity criteria and NMI, the proposed model is assessed via its application to four distinct datasets (Normalized Mutual Information). The findings demonstrated that, in comparison to previous approaches, the model suggested had greater levels of modularity, NMI, and generalizability.

The Three Stage algorithm (TS) considers local and global information in its search for communities [26]. It then proceeds through three steps in rapid succession, beginning with identifying core nodes and ending with the merger of communities. Central nodes are chosen in the TS method from among the nodes whose distances from one another are more significant than the average distance between all nodes in the network. After that, in the second stage, a label is issued to every node of the network, and the label of the nodes that are most similar to one another is deemed to have the same color. In the last step, called "combination of communities," specific communities are merged to increase the modularity measure [27].

It has been suggested that the Cuckoo Search Optimization (CSO) method for CD might benefit from adding a GA [28]. The CSO method has several flaws, including premature convergence, delayed convergence, and being stuck in the local trap. In terms of CD in complex networks, GA has been relatively effective, contributing to increased exploration and exploitation. To improve the CSO's speed and accuracy, we have been using GA operators in a dynamic manner. The total number of populations undergo a continuous adjustment process in response to the level of exploration and exploitation. As an optimization function, the modularity objective function, also known as Q, and the normalized mutual information, also known as NMI, are utilized. The research was conducted on actual complicated networks of six different sorts. The suggested method was put through its paces using GA, Artificial Bee Colony (ABC), Grey Wolf Optimizer (GWO), and CSO, each with varying iterations regarding the modularity and NMI criterion. The findings indicate that the suggested algorithm has been more effective than the fundamental comparative algorithms and has shown its superiority in modularity and NMI. The results show that this is the case in the majority of the comparisons. In terms of modularity, the suggested algorithm performed an average of 54 percent better than previous algorithms, while in terms of NMI, it performed an average of 88 percent better.

Combining the Firefly Algorithm (FA) with Learning Automata (LA) is the basis for an approach that has been presented [29]. The efficiency of the FA is improved with the help of LA in the model that has been suggested. Choosing the ideal neighbors for the FA agents is done using the LA. The findings from the four datasets of Karate, Dolphins, Polbooks, and Football demonstrated that the suggested model had a higher level of Normalized Mutual Information (NMI) when compared to previous models.

Mirzadeh and Meybodi came up with the idea for the CD of using a Michigan memetic algorithm called MLAMA-Net [30]. Every node is equipped with a chromosome and a learning automaton (LA) in the newly developed approach. The chromosome stores the histories of exploration and represents the community associated with the respective node. The learning automaton represents a meme and preserves the history of exploitative practices. Although single-objective optimization algorithms have seen widespread use for CD in complex networks, these algorithms continue to have a few obvious drawbacks, such as the resolution limit of modularity objectives or other evaluation objectives. Despite this, single-objective optimization algorithms continue to be widely used.

## 3. Method

The suggested model for CD's steps is broken down and discussed in this portion of the article. First, the formulation of the issue to identify the communities is described, and then, using LF as a basis, the stages of the proposed model to alter the location of the vectors and discover optimum solutions are presented.

### *3.1. Formulation of Problem*

Finding the community structures inside complex networks has been approached from several angles, using various methodologies and algorithms. The primary goal of the algorithms is to make clear the framework within which they place members of a community. To assess the vectors of the proposed model, this work makes use of Eq. (1). The effectiveness of the model that has been developed is determined by analyzing how well a network can be segmented into several communities. Potential communities and the nodes on the relevant network that make up those communities are optimized with the assistance of Eq. (1) to determine which community structure will result in the greatest amount of cost. Modularity-based methods try to maximize modularity value, the definition of which can be seen in Eq. (1).

$$V_i = \frac{\sum_{i=1}^{n}(e_{ij} - \max(e_{ij}))}{\min(e_{ij})} \tag{1}$$

In Eq. (1), the degree of the vertices is denoted by $e_{ij}$, and the number of nodes in a vector is represented by n. $Index,$ and $IDNode$ are the names of the two different bits of information that are included in each element of the vector. Figure (1) represents a network with eight nodes (a) in this investigation. In addition, an example of a vector generated following the presented network is shown in Fig. 1(b). The sub-graphs produced by the supplied vector are shown in Fig. 1(c). In this particular illustration, there are three separate sub-graphs, each of which is represented by a distinct hue. Nodes should have the greatest number of connections or characteristics in common with other nodes in their community. In contrast, nodes in the opposite community should have the fewest number of links or the fewest possible features in common with them.

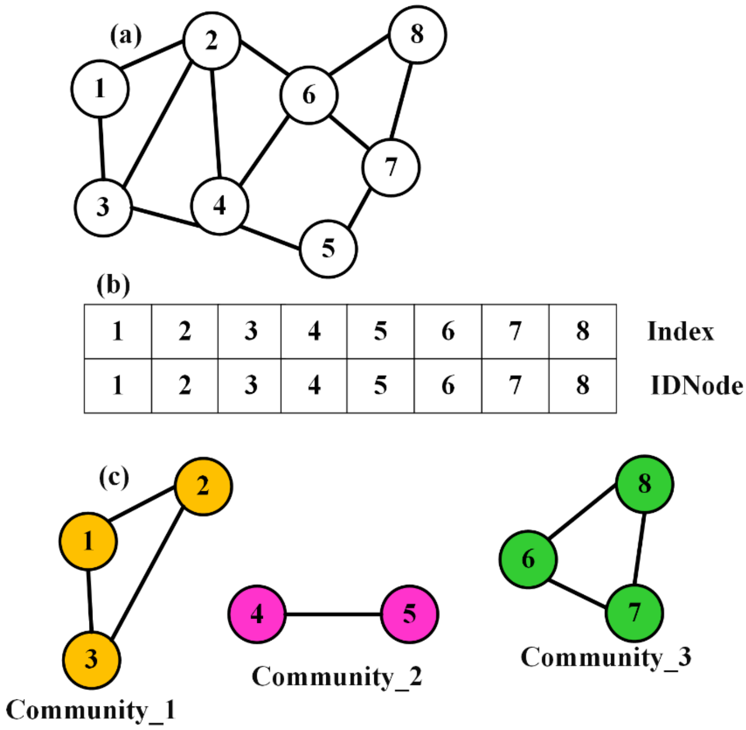


***Figure 1. Representation of a sample network and the acquired subgraphs.***

### *3.2. Proposed Model*

The SOS forms an initial population of organisms at the beginning of the process. Each organism in this population is randomly generated within the specified search area. The biological connection between different species served as the basis for the design of the SOS algorithm, which was modeled after the three types of symbiotic relationships: mutualism, commensalism, and parasitism. The SOS searches for the best possible global solution via mutualism, commensalism, and parasitism with each new generation. Because it is a population-based algorithm, SOS starts its search with an initial population known as the ecosystem, which is often produced at random. This population is what the algorithm uses to begin its search. Each creature stands in for a potential answer to the related issue, and it is given a fitness value within the ecosystem that indicates the degree to which it has adapted to the goal being pursued. The suggested model uses the Lévy Flight (LF) approach to resolve SOS issues. In SOS, the LF distribution is used so that the solution does not get mired in a single optimal solution at any given time. The figure presents a flowchart of the suggested model for your Figure (2).

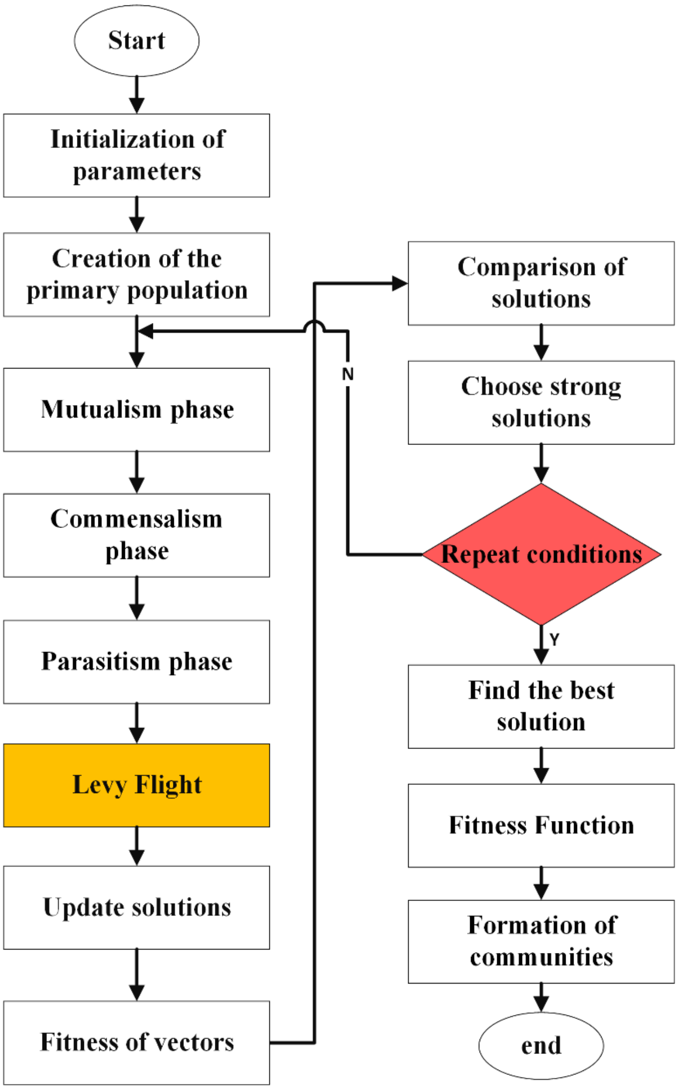


**Fig. 2** flowchart of the proposed model

**Mutualism phase:** The term "mutualism" refers to the symbiotic connection between two species in which both organisms benefit from their interactions. The population is sampled in such a way that two separate creatures, $x_i$ and $x_{R_1}$, are selected at random. Through their interaction, $x_i$ and $x_{R_1}$ give rise to two new creatures. Eq. (2) and Eq. (3) were defined for this phase [12].

$$x_{inew} = x_i + rand[0,1] \times (x_{best} - BF1 \times Mutual\_Vector) \tag{2}$$

$$x_{R_1new} = x_{R_1} + rand[0,1] \times (x_{best} - BF2 \times Mutual\_Vector) \tag{3}$$

$$Mutual_{Vector} = \frac{x_i + x_{R_1}}{2} \tag{4}$$

In Eq. (2) and Eq. (3), rand[0,1] is a random number vector that falls within the range [0,1]; $x_{best}$ is the best individual organism in the current population; $BF1$ and $BF2$ stand for benefit factors of each organism, and they are randomly chosen as either 1 or 2 about partially or fully benefiting from the interaction; $x_{best}$ is the best individual organism in the current population; rand[0,1] is a random number vector that falls within the As a direct result of this, $x_{inew}$ and $x_{R_1 new}$ are employed to evaluate the goal functions. The organisms are then considered against those that have been provided by $x_i$ and $x_{R_1}$ to determine which of the two sets of creatures is superior.

**Commensalism phase:** The symbiotic connection known as commensalism is one in which only one organism benefits from the partnership while the other does not. The population is sampled to choose at random two distinct creatures, denoted by the names $x_i$ and $x_{R_1}$. $x_i$ will be eligible for benefits, whereas $x_{R_1}$ will not be. After then, a brand-new organism will be produced. Following the Eq. (5), this phase is defined [12].

$$x_{inew} = x_i + rand[-1,1] \times (x_{best} - x_{R_1}) \tag{5}$$

In Eq. (5), rand [-1,1] is a random number vector in [-1,1].

The simulated phenomena that occur throughout the phase of commensalism is that natural biological beings improve their viability by developing a commensalism interaction with other individuals. It was important for LF to be conducted on Xi to boost the capability of exploring the global space. After that, a partial symbiosis was established based on the newly discovered $x_i$ and $x_{R_1}$, and $x_{best}$ was used to direct the search for the ideal value. As a result, the algorithm's exploration and exploitation capabilities during the commensalism phase may be more balanced. The production of the random response x(t+1) occurs as follows: first, random numbers are created using the Levy probability distribution; next, in the space of potential replies, the answers are formed using random numbers as Eq. (6); finally, the random answer x(t+1) is produced.

$$x_i^{t+1} = x_i^t + \mu \times sign[rand - 0.5] \oplus Levy(\beta) \tag{6}$$

In Eq. (6) $x_i^t$ is the position of the i-th agent in the t-th iteration, a is a random number with a uniform distribution between 0 and 1, the variable $sign[rand - 0.5]$ can take on any one of the three possible values of -1, 0 or 1, and is the product of the element. These terms can be found in Eq. (6). The sign function will define the direction of the random movement, and the levy distribution function will decide how long the training will last. The value of the levy parameter is specified by Eq. (7).

$$Levy(\beta) \sim \frac{\emptyset \times \mu}{|v|^{\frac{1}{\beta}}} \tag{7}$$

As μ and $v$ are random numbers drawn from a standard normal distribution, β is a parameter fixed at 1.5, while $\emptyset$ is defined as Eq. (8) which is the standard gamma function.

$$\emptyset = \left[\frac{\Gamma(1+\beta) \times \sin(\frac{\pi \times \beta}{2})}{\Gamma\left((\frac{1+\beta}{2}) \times \beta \times 2^{\frac{\beta-1}{2}}\right)}\right]^{\frac{1}{\beta}} \tag{8}$$

**Parasitism phase:** Parasitism is a symbiotic interaction in which one organism reaps the advantages of the partnership while the other suffers the negative effects. Within the existing ecosystem, a choice is made at random between two distinct creatures, denoted by the names Xi and $x_{R_1}$. A duplicate of the organism $x_i$, which stands in for the parasite organism, is denoted by the symbol Parasite-Vector. $x_{R_1}$ signifies the organism that is being parasitized by the other creature. Some of the components in Parasite Vector will attempt to cause damage to the host by making a random alteration. If the goal function value of Parasite-Vector is higher than that of $x_{R_1}$, then $x_{R_1}$ will be replaced in the existing ecosystem by Parasite-Vector.

### *3.3. Evaluation Criteria*

The NMI and modularity criteria have been used throughout the assessment process. The NMI and Modularity criteria are two popular assessment criteria used to evaluate the quality of the produced partitions. The NMI is an external metric that estimates the similarity between the true and the partitions that have been discovered. The modularity-criterion is an internal metric used to assess the degree to which detected communities have outliers in their structural makeup.

**NMI:** The NMI criterion is defined according to Eq. (9) [31]. In Eq. (9), parameters A and B are two separated parts of a network, C is the confusion matrix, $C_{ij}$ is equal to the number of common nodes between community i in part $A$ and community j in part B. The NMI index shows the overall connection between nodes. This index represents the degree of network correlation. In a high-density interconnected network, there are many relationships between nodes. The amount of NMI index is a number between zero and one. $A$ value of 1 means that all the nodes in the network are connected, and a density equal to zero indicates that there are no links between the nodes in the network.

$$NMI = \frac{-2\sum_{i=1}^{C_A}\sum_{j=1}^{C_B} C_{ij}\log(C_{ij}.\frac{n}{C_i.C_j})}{\sum_{i=1}^{C_A} C_i\log(C_i/n) + \sum_{j=1}^{C_B} C_j\log(C_j/N)} \tag{9}$$

**Modularity:** A measure called modularity is presented to evaluate the performance of CD algorithms in social networks. Calculating this criterion is shown in Eq. (10) [31].

$$Q = \sum_{s=1}^{k}\left[\frac{l_s}{m} - \left(\frac{d_s}{2m}\right)^2\right] \tag{10}$$

In Eq. (10), ls is the total number of edges connecting the vertices inside the cluster s, parameter ds is the total degree of nodes in s, and m is the total number of edges of the network. Possible values of the modularity criterion are in the range [1 to -0.5].

**ACC:** Accuracy (ACC) [32] is an index that describes the accuracy of the clustering results. Let $C_i$ be the actual classification label and Ri be the prediction label.

$$ACC = \frac{\sum_{i=1}^{N}\delta(R_i, Map(C_i))}{N} \tag{11}$$

In Eq. (11), $\delta(R_i, map(C_i))$ is Kronecker's delta and $map(Ci)$ is the mapping function that maps each community label Ci to the equivalent label from the real sample. The range of ACC is [0,1]. The higher the accuracy, the better the clustering results.

## 4. Evaluation and Results

all of the assessments were carried out on MATLAB R2021b, which was installed on Windows 10 and powered by an Intel (R) Core (TM) i7-4790 processor operating at 2.4 GHz, with a total of 6 GB of random-access memory (RAM). Six separate datasets, including varying numbers of communities, each served as the basis for the assessment. The average of 15 separate tests is used to determine all of the provided findings. The initial population and the number of iterations in the SOS algorithm are 50 and 200, respectively. A comparison is made between the SOS-LF model and the MFO [33] and SSA [34] algorithms. There are six actual networks throughout the globe shown in Table (1). There is potential for hundreds of thousands, at most, of edges to be present. The number of communities can be represented by c, the number of edges can be represented by m, and the number of nodes may be represented by n. "High school friendship6" and "High school friendship7" are connected to the same network, however, their respective community divisions are unique.

***Table 1. description of the real datasets***

| # | Datasets | n | m | c | Descriptions |
|---|---|---|---|---|---|
| 1 | Karate | 34 | 78 | 2 | Zachary's Karate club |
| 2 | School6 | 69 | 220 | 6 | High school friendship6 |
| 3 | School7 | 69 | 220 | 7 | High school friendship7 |
| 4 | Polbooks | 105 | 441 | 3 | Political books |
| 5 | Dolphin | 62 | 159 | 2 | Dolphin social network |
| 6 | Football | 115 | 613 | 12 | American college football |

### *4.1. Results based on Iterations*

Table (2) shows the results of the proposed model based on three main criteria with 150 iterations on different data sets. According to the obtained results, it is clear that the SOS-LF model has better detection accuracy compared to MFO and SSA. According to the results, the SOS-LF outperformed other methods in Karate, School6, School7, Polbooks, Dolphin, and Football. SOS-LF yielded higher values of accuracy and NMI than the other existing methods.

***Table 2. The results of the proposed model on different datasets with 150 iterations***

| Datasets | Models | NMI | | | Modularity | | | AC | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | Best | Mean | Worst | Best | Mean | Worst | Best | Mean | Worst |
| Karate | SOS-LF | **1.0000** | 1.0000 | 1.0000 | **0.4275** | 0.4252 | 0.4242 | **1.0000** | 1.0000 | 1.0000 |
| | MFO | 0.9862 | 0.9879 | 0.9863 | 0.4259 | 0.4245 | 0.4231 | 0.9986 | 0.9981 | 0.9974 |
| | SSA | 0.9795 | 0.9785 | 0.9775 | 0.4221 | 0.4236 | 0.4224 | 0.9964 | 0.9975 | 0.9962 |
| School6 | SOS-LF | **0.9742** | 0.9736 | 0.9723 | **0.6184** | 0.6159 | 0.6145 | **0.9451** | 0.9446 | 0.9437 |
| | MFO | 0.9665 | 0.9664 | 0.9652 | 0.6076 | 0.6058 | 0.6042 | 0.9437 | 0.9425 | 0.9416 |
| | SSA | 0.9621 | 0.9663 | 0.9658 | 0.6049 | 0.6032 | 0.6019 | 0.9447 | 0.9431 | 0.9421 |
| School7 | SOS-LF | **0.9365** | 0.9371 | 0.9368 | **0.6151** | 0.6142 | 0.6123 | **0.8832** | 0.8819 | 0.8805 |
| | MFO | 0.9348 | 0.9352 | 0.9338 | 0.6112 | 0.6122 | 0.6103 | 0.8825 | 0.8816 | 0.8788 |
| | SSA | 0.931 | 0.932 | 0.932 | 0.612 | 0.612 | 0.611 | 0.880 | 0.879 | 0.878 |

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | 5 | 7 | 2 | 2 | 5 | 2 | 5 | 1 | 3 |
| Polbooks | SOS-LF | **0.9419** | 0.9399 | 0.9369 | **0.5372** | 0.5357 | 0.5335 | **0.8362** | 0.8354 | 0.8337 |
| | MFO | 0.9355 | 0.9358 | 0.9358 | 0.5338 | 0.5333 | 0.5320 | 0.8351 | 0.8342 | 0.8325 |
| | SSA | 0.9376 | 0.9335 | 0.9370 | 0.5312 | 0.5293 | 0.5279 | 0.8327 | 0.8314 | 0.8302 |
| Dolphin | SOS-LF | **0.9989** | 0.9989 | 0.9975 | **0.5237** | 0.5238 | 0.5225 | **0.5273** | 0.5261 | 0.5250 |
| | MFO | 0.9897 | 0.9991 | 0.9987 | 0.5225 | 0.5222 | 0.5222 | 0.5260 | 0.5241 | 0.5231 |
| | SSA | 0.9864 | 0.9869 | 0.9964 | 0.5194 | 0.5189 | 0.5187 | 0.5252 | 0.5237 | 0.5217 |
| Football | SOS-LF | **0.9456** | 0.9461 | 0.9452 | **0.5874** | 0.5874 | 0.5852 | **0.5633** | 0.5619 | 0.5608 |
| | MFO | 0.9417 | 0.9415 | 0.9415 | 0.5843 | 0.5841 | 0.5832 | 0.5624 | 0.5612 | 0.5587 |
| | SSA | 0.9397 | 0.9387 | 0.9381 | 0.5834 | 0.5825 | 0.5675 | 0.5630 | 0.5609 | 0.5895 |

The outcomes of the suggested model, which was evaluated according to three primary criteria and was repeated 300 times using a variety of datasets, are shown in Table (3). Compared to 150 iterations, the acquired findings make it abundantly evident that the identification accuracy of the models has significantly improved after 300 iterations have been carried out. After looking at Table (3), we can see that the NMI value that corresponds to the outcomes of the SSA approach is not optimal. This is because the modularity function Q, which serves as the evaluation index for CD findings, may represent the internal compactness of community structure; nevertheless, it cannot reflect the genuine community structure of the network in the most exact manner. The SOS-LF can achieve the maximum Q value of modularity on six different networks. At the same time, it is possible to observe that although the SSA does not get the most excellent modularity value on the Karate network, the Dolphin network, or the Football network, it does earn the second highest value. Based on getting closer to the actual network partition result, SOS-LF can get the detection result with closer internal connections of community structure.

***Table 3. The results of the proposed model on different datasets with 300 iterations***

| **Datasets** | **Models** | **NMI** | | | **Modularity** | | | **AC** | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | Best | Mean | Worst | Best | Mean | Worst | Best | Mean | Worst |
| Karate | SOS-LF | **1.0000** | 1.0000 | 1.0000 | **0.4298** | 0.4285 | 0.4273 | **1.0000** | 1.0000 | 1.0000 |
| | MFO | 0.9986 | 0.9974 | 0.9966 | 0.4294 | 0.4254 | 0.4240 | 1.0000 | 0.9995 | 0.9980 |
| | SSA | 0.9971 | 0.9962 | 0.9945 | 0.4222 | 0.4232 | 0.4224 | 0.9986 | 0.9973 | 0.9962 |
| School6 | SOS-LF | **0.9811** | 0.9784 | 0.9752 | **0.6194** | 0.6161 | 0.6157 | **0.9569** | 0.9532 | 0.9512 |
| | MFO | 0.9712 | 0.9682 | 0.9671 | 0.6086 | 0.6063 | 0.6054 | 0.9497 | 0.9455 | 0.9447 |
| | SSA | 0.9688 | 0.9688 | 0.9676 | 0.6052 | 0.6032 | 0.6019 | 0.9532 | 0.9526 | 0.9515 |
| School7 | SOS-LF | **0.9386** | 0.9381 | 0.9373 | **0.6183** | 0.6152 | 0.6132 | **0.8879** | 0.8826 | 0.8812 |

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | MFO | 0.9371 | 0.9360 | 0.9345 | 0.6177 | 0.6135 | 0.6128 | 0.8854 | 0.8838 | 0.8796 |
| | SSA | 0.9343 | 0.9338 | 0.9331 | 0.6132 | 0.6125 | 0.6112 | 0.8829 | 0.8805 | 0.8796 |
| Polbooks | SOS-LF | **0.9451** | 0.9426 | 0.9410 | **0.5389** | 0.5375 | 0.5349 | **0.8394** | 0.8376 | 0.8348 |
| | MFO | 0.9430 | 0.9425 | 0.9417 | 0.5384 | 0.5357 | 0.5341 | 0.8387 | 0.8361 | 0.8339 |
| | SSA | 0.9415 | 0.9406 | 0.9388 | 0.5311 | 0.5295 | 0.5281 | 0.8355 | 0.8355 | 0.8355 |
| Dolphin | SOS-LF | **0.9995** | 0.9991 | 0.9986 | **0.5267** | 0.5249 | 0.5231 | **0.5298** | 0.5290 | 0.5265 |
| | MFO | 0.9921 | 0.9923 | 0.9902 | 0.5252 | 0.5226 | 0.5234 | 0.5305 | 0.5285 | 0.5276 |
| | SSA | 0.9981 | 0.9973 | 0.9964 | 0.5197 | 0.5190 | 0.5190 | 0.5252 | 0.5237 | 0.5215 |
| Football | SOS-LF | **0.9956** | 0.9968 | 0.9902 | **0.5894** | 0.5886 | 0.5872 | **0.5669** | 0.5638 | 0.5621 |
| | MFO | 0.9863 | 0.9840 | 0.9819 | 0.5882 | 0.5852 | 0.5849 | 0.5712 | 0.5704 | 0.5694 |
| | SSA | 0.9775 | 0.9756 | 0.9756 | 0.5874 | 0.5844 | 0.5226 | 0.5662 | 0.5651 | 0.5637 |

### *4.2. Fitness of Models*

Figures (3 to 5) show the fitness of models based on NMI. Figures (3-5) show the SOS-LF's improvement rates in terms of NMI metric compared with other methods in each network. In Figure (3), The maximum increases in NMI on Karate and School6 are 1.0000 and 0.9811, respectively. In Figure (4), The maximum increases in NMI on School7 and Polbooks are 0.9386 and 0.9451, respectively. In Figure (5), The maximum increases in NMI on dolphins and Football are 0.9995 and 0.9956, respectively. The SSA algorithm shows poor stability in the above experiments with a relatively large distribution interval of NMI, indicating the significant possibility of producing a local optimum. However, the SOS-LF using the LF strategy can significantly improve the algorithm's stability, proving that the LF strategy can help the algorithm avoid local optimum and get a community division more closely related to the real one.

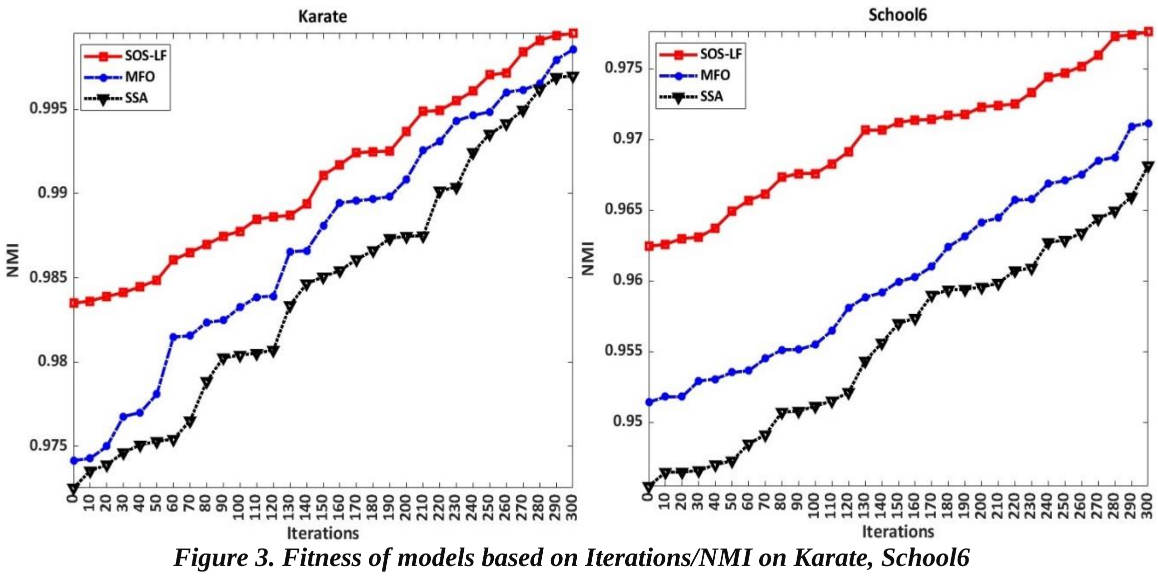


***Figure 3. Fitness of models based on Iterations/NMI on Karate, School6***

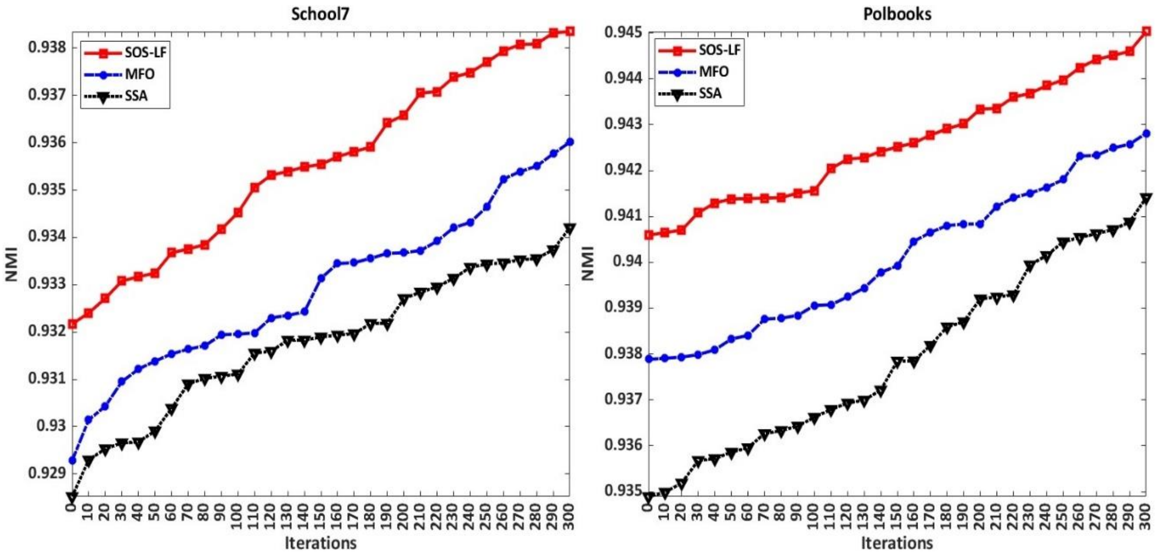


***Figure 4. Fitness of models based on Iterations/NMI on School7, Polbooks***

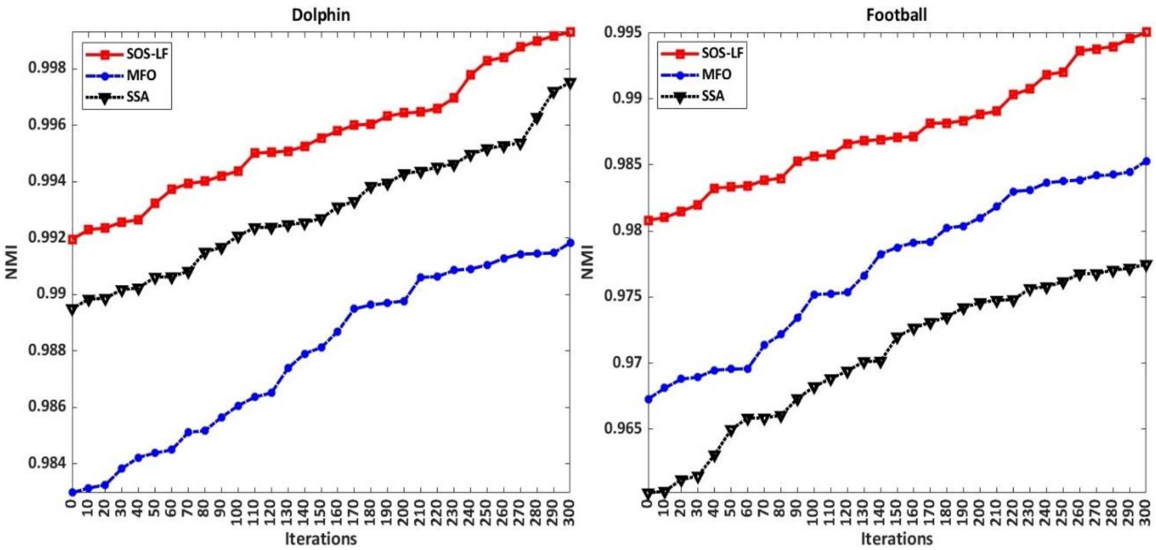


***Figure 5. Fitness of models based on Iterations/NMI on Dolphin, Football***

## 5. Conclusion and Future Works

Advanced web structure in the realm of network data mining, the CD is a significant research path that is being pursued. Some currently available approaches for detecting the community need users to have previous knowledge of the structure of the community. At the same time, other methods make assumptions that are not realistic about the structure of the community. Consequently, due to these factors, the CD technique cannot adequately show the underlying community structure of the actual complex network. This study presented a CD technique in complex networks utilizing SOS-LF to optimize modularity. This paper was written as a reaction to the challenges described above. The author offers a novel CD method based on the SOS and the FL. The suggested model was shown to have a greater detection accuracy than other algorithms based on the experiments conducted on actual networks. The NMI values for the datasets of Karate, Dolphin, Football, and Polbooks were respectively 1.00, 0.9995, 0.9956, and 0.9451. In addition, the modularity values for the Karate, Dolphin, Football, and Polbooks datasets were respectively 0.4298, 0.5267, 0.5894, and 0.5389. We want to apply optimization of experimental design and analysis of complex network community structure discoveries in the work that we do in the future. The experimental section may incorporate African Vultures Optimization Algorithm (AVOA) and Differential Evolution (DE) methods. On the other hand, chaotic maps may be used to enhance the algorithm's premature convergence when applied to large-scale datasets.

## References

[1] F. Dabaghi-Zarandi, P. KamaliPour, Community detection in complex network based on an improved random algorithm using local and global network information. *Journal of Network and Computer Applications.* Vol. 206, No. pp. 103492, 2022.

[2] E. Pourabbasi, V. Majidnezhad, S. Taghavi Afshord, Y. Jafari, A new single-chromosome evolutionary algorithm for community detection in complex networks by combining content and structural information. *Expert Systems with Applications.* Vol. 186, No. pp. 115854, 2021.

[3] A. Mahmood, M. Small, Subspace Based Network Community Detection Using Sparse Linear Coding. *IEEE Transactions on Knowledge and Data Engineering.* Vol. 28, No. 3, pp. 801-812, 2016.

[4] J. Ma, J. Fan, Local Optimization for Clique-Based Overlapping Community Detection in Complex Networks. *IEEE Access.* Vol. 8, No. pp. 5091-5103, 2020.

[5] Y. Su, C. Liu, Y. Niu, F. Cheng, X. Zhang, A Community Structure Enhancement-Based Community Detection Algorithm for Complex Networks. *IEEE Transactions on Systems, Man, and Cybernetics: Systems.* Vol. 51, No. 5, pp. 2833-2846, 2021.

[6] M. Kiani Sarkaleh, F. Khoshnood, Designing a Trust-Based Recommender System in Social Rating Networks. *Journal of Advances in Computer Research.* Vol. 10, No. 2, pp. 75-89, 2019.

[7] X. Zhang, K. Zhou, H. Pan, L. Zhang, X. Zeng, Y. Jin, A Network Reduction-Based Multiobjective Evolutionary Algorithm for Community Detection in Large-Scale Complex Networks. *IEEE Transactions on Cybernetics.* Vol. 50, No. 2, pp. 703-716, 2020.

[8] F. Cheng, C. Wang, X. Zhang, Y. Yang, A Local-Neighborhood Information Based Overlapping Community Detection Algorithm for Large-Scale Complex Networks. *IEEE/ACM Transactions on Networking.* Vol. 29, No. 2, pp. 543-556, 2021.

[9] L. Sun, T. Ye, J. Sun, X. Duan, Y. Luo, Density-Peak-Based Overlapping Community Detection Algorithm. *IEEE Transactions on Computational Social Systems.* Vol. 9, No. 4, pp. 1211-1223, 2022.

[10] X. Teng, J. Liu, M. Li, Overlapping Community Detection in Directed and Undirected Attributed Networks Using a Multiobjective Evolutionary Algorithm. *IEEE Transactions on Cybernetics.* Vol. 51, No. 1, pp. 138-150, 2021.

[11] X. Liu, Y. Du, M. Jiang, X. Zeng, Multiobjective Particle Swarm Optimization Based on Network Embedding for Complex Network Community Detection. *IEEE Transactions on Computational Social Systems.* Vol. 7, No. 2, pp. 437-449, 2020.

[12] M.-Y. Cheng, D. Prayogo, Symbiotic Organisms Search: A new metaheuristic optimization algorithm. *Computers & Structures.* Vol. 139, No. pp. 98-112, 2014.

[13] F.S. Gharehchopogh, M.H. Nadimi-Shahraki, S. Barshandeh, B. Abdollahzadeh, H. Zamani, CQFFA: A Chaotic Quasi-oppositional Farmland Fertility Algorithm for Solving Engineering Optimization Problems. *Journal of Bionic Engineering.* Vol. No. pp. 2022.

[14] E.H. Houssein, M.R. Saad, F.A. Hashim, H. Shaban, M. Hassaballah, Lévy flight distribution: A new metaheuristic algorithm for solving engineering optimization problems. *Engineering Applications of Artificial Intelligence.* Vol. 94, No. pp. 103731, 2020.

[15] Z. Li, J. Liu, A multi-agent genetic algorithm for community detection in complex networks. *Physica A: Statistical Mechanics and its Applications.* Vol. 449, No. pp. 336-347, 2016.

[16] S. Rahimi, A. Abdollahpouri, P. Moradi, A multi-objective particle swarm optimization algorithm for community detection in complex networks. *Swarm and Evolutionary Computation.* Vol. 39, No. pp. 297-309, 2018.

[17] X. Zhou, Y. Liu, B. Li, H. Li, A multiobjective discrete cuckoo search algorithm for community detection in dynamic networks. *Soft Computing.* Vol. 21, No. 22, pp. 6641-6652, 2017.
[18] S. Kumar, A. Mallik, S.S. Sengar, Community detection in complex networks using stacked autoencoders and crow search algorithm. *The Journal of Supercomputing.* Vol. No. pp. 2022.
[19] F. Dabaghi Zarandi, M. Kuchaki Rafsanjani, Community detection in complex networks using structural similarity. *Physica A: Statistical Mechanics and its Applications.* Vol. 503, No. pp. 882-891, 2018.
[20] H. Banati, N. Arora, Detecting communities in complex networks-A discrete hybrid evolutionary approach. *International Journal of Computers and Applications.* Vol. 38, No. 1, pp. 29-40, 2016.
[21] A. Abdollahpouri, S. Rahimi, S.M. Majd, C. Salavati. A Modified Particle Swarm Optimization Algorithm for Community Detection in Complex Networks. in *Machine Learning and Knowledge Extraction*. 2018. Cham: Springer International Publishing.
[22] M. Kaur, A. Mahajan. Community Detection in Complex Networks: A Novel Approach Based on Ant Lion Optimizer. in *Proceedings of Sixth International Conference on Soft Computing for Problem Solving*. 2017. Singapore: Springer Singapore.
[23] X. Zhou, K. Yang, Y. Xie, C. Yang, T. Huang, A novel modularity-based discrete state transition algorithm for community detection in networks. *Neurocomputing.* Vol. 334, No. pp. 89-99, 2019.
[24] S. Ghafori, F.S. Gharehchopogh, 9 - A multiobjective Cuckoo Search Algorithm for community detection in social networks, in *Multi-Objective Combinatorial Optimization Problems and Solution Methods*, M. Toloo, S. Talatahari, and I. Rahimi, Editors. 2022, Academic Press. p. 177-193.
[25] Z. narimani, F. Soleimanian Gharehchopogh, Community Detection on Social Networks Using the Improved Harris Hawk Optimization Algorithm. *Journal of Advances in Computer Engineering and Technology.* Vol. No. pp. -, 2022.
[26] X. You, Y. Ma, Z. Liu, A three-stage algorithm on community detection in social networks. *Knowledge-Based Systems.* Vol. 187, No. pp. 104822, 2020.
[27] N. Rafizadeh Shotorbani Fard, H. Hosseinzadeh, M. Bekravi, Accessibility Evaluation in Biometric Hybrid Architecture for Protecting Social Networks Using Colored Petri Nets. *Journal of Advances in Computer Research.* Vol. 10, No. 1, pp. 43-57, 2019.
[28] S.T. Shishavan, F.S. Gharehchopogh, An improved cuckoo search optimization algorithm with genetic algorithm for community detection in complex networks. *Multimedia Tools and Applications.* Vol. 81, No. 18, pp. 25205-25231, 2022.
[29] H. Rabani, F. Soleimanian Gharehchopogh, An Optimized Firefly Algorithm based on Cellular Learning Automata for Community Detection in Social Networks. *Journal of Advances in Computer Research.* Vol. 10, No. 3, pp. 13-30, 2019.
[30] M. Rezapoor Mirsaleh, M. Reza Meybodi, A Michigan memetic algorithm for solving the community detection problem in complex network. *Neurocomputing.* Vol. 214, No. pp. 535-545, 2016.
[31] K. Steinhaeuser, N.V. Chawla, Identifying and evaluating community structure in complex networks. *Pattern Recognition Letters.* Vol. 31, No. 5, pp. 413-421, 2010.
[32] M. Qin, D. Jin, K. Lei, B. Gabrys, K. Musial-Gabrys, Adaptive community detection incorporating topology and content in social networks. *Knowledge-Based Systems.* Vol. 161, No. pp. 342-356, 2018.
[33] S. Mirjalili, Moth-flame optimization algorithm: A novel nature-inspired heuristic paradigm. *Knowledge-Based Systems.* Vol. 89, No. pp. 228-249, 2015.
[34] S. Mirjalili, A.H. Gandomi, S.Z. Mirjalili, S. Saremi, H. Faris, S.M. Mirjalili, Salp Swarm Algorithm: A bio-inspired optimizer for engineering design problems. *Advances in Engineering Software.* Vol. 114, No. pp. 163-191, 2017.