ISSN: 0067-2904

Parallel Particle Swarm Optimization Algorithm for Identifying Complex Communities in Biological Networks

Dhuha Abdulhadi Abduljabbar*
Computer Science Department, College of Science, University of Baghdad, Baghdad, Iraq

Received: 15/11/2022 Accepted: 12/2/2023 Published: 30/1/2023

Abstract

Identification of complex communities in biological networks is a critical and ongoing challenge since lots of network-related problems correspond to the subgraph isomorphism problem known in the literature as NP-hard. Several optimization algorithms have been dedicated and applied to solve this problem. The main challenge regarding the application of optimization algorithms, specifically to handle large-scale complex networks, is their relatively long execution time. Thus, this paper proposes a parallel extension of the PSO algorithm to detect communities in complex biological networks. The main contribution of this study is summarized in three- fold; Firstly, a modified PSO algorithm with a local search operator is proposed to detect complex biological communities with high quality. Secondly, the variability in the capability of PSO to extract community structure in biological networks is studied when different types of crossover operators are used. Finally, to reduce the computational time needed to solve this problem, especially when detecting complex communities in large-scale biological networks, we have implemented parallel computing to execute the algorithm. The performance of the proposed algorithm was tested and evaluated on two real biological networks. The experimental results showed the effective performance of the proposed algorithm when using single-point crossover operator, and its superiority over other counterpart algorithms. Moreover, the use of parallel computing in the proposed algorithm representation has greatly reduced the computational time required for its execution.

Keywords: Complex community structure, Biological networks, Particle swarm optimization algorithm, Local search operator, Computational parallelization.

خوارزمية تحسين سرب الجسيمات المتوازية لتحديد المجتمعات المعقدة في الشبكات البيولوجية

ضحى عبدالهادي عبد الجبار
قسم علوم الحاسبات, كلية العلوم, جامعة بغداد, بغداد, العراق

الخلاصة

يعد تحديد المجتمعات المعقدة في الشبكات البيولوجية تحديًا حاسمًا ومستمرًا نظرًا لأن الكثير من المشكلات المتعلقة بالشبكات تتوافق مع مشكلة تماثل الرسم البياني الفرعي المعروفة في الأدبيات باسم NP–hard. تم تخصيص العديد من خوارزميات التحسين وتطبيقها لحل هذه المشكلة. يتمثل التحدي الرئيسي بتطبيق خوارزميات التحسين، خاصة عند التعامل مع الشبكات المعقدة واسعة النطاق، هو وقت التنفيذ الطويل نسبيًا. في هذا الصدد، تقترح هذه الورقة امتدادًا متوازيا لخوارزمية PSO لاكتشاف المجتمعات في الشبكات البيولوجية المعقدة. يتم

*Email: dhuha.abd@sc.uobaghdad.edu.iq

تلخيص مساهمة هذه الدراسة في ثلاثة نقاط رئيسية. أولاً، تم اقتراح خوارزمية تحسين سرب الجسيمات المحسنة باستخدام مشغل بحث محلي لاكتشاف المجتمعات البيولوجية المعقدة بجودة عالية. ثانيًا، تمت دراسة الاختلافات في قدرة PSO لاستخراج بنية المجتمع في الشبكات البيولوجية عند استخدام أنواع مختلفة من مشغلي التقاطع. أخيرًا، لتقليل الوقت الحسابي اللازم لحل هذه المشكلة، خاصةً عند اكتشاف المجتمعات المعقدة في الشبكات البيولوجية واسعة النطاق، قمنا بتطبيق الحوسبة المتوازية لتنفيذ الخوارزمية. تم اختبار وتقييم أداء الخوارزمية المقترحة عند تطبيقها على شبكتين بيولوجيتين حقيقيتين. أظهرت النتائج التجريبية الأداء الفعال للخوارزمية المقترحة عند استخدام عامل التقاطع أحادي النقطة وتفوقها على الخوارزميات النظيرة الأخرى. علاوة على ذلك، أدى استخدام الحوسبة المتوازية في تمثيل الخوارزمية المقترحة إلى تقليل الوقت الحسابي المطلوب لتنفيذها بشكل كبير.

1. Introduction

Modelling and investigating complex systems containing biological information is considered a highly challenging process. Since the systems that are used to represent real-world biological data include very important information that help in understanding cell activity, detection a set of proteins that physically or functionally interact to achieve a specific function, as well as identifying relations between the molecular structures in the body. Complex computational systems are generally represented using essential network concepts and modelled using basic graph theories [1]. Where the nodes of the graph denote biological units of interest such as proteins, genes, species, or individuals; while edges of the graph indicate the interaction between nodes such as gene flow, regulatory interaction, or infectious contacts [2]. When utilizing simple analyzing tools, some, if not all, topological features are seen as insignificant information at these structures, and significant data related to patterns are not revealed. Patterns represent important information revealed when analyzing networks, and they help in comprehending dynamical processes in the complex networks [3-6].

In general, significant information could be attained from complex networks that represent complex systems in the real world when identifying the subgroups hidden within them, which are known in the literature as communities. In particular, the detected subgroups have their features, like the maximum relations, common tasks, and structural/ positional similarities. While the objects of these subgroups have the largest number of relationships and many features in common with the objects of their subgroups, they have the fewest number of relationships and few features in common with the objects of the other subgroups [1, 4, 6, 7]. On the other hand, densely connected communities or groups are prevalent in biological networks and may be related to specific molecular, cellular, or tissue functions. Therefore, the discovery of biological communities is a key first step to analyzing and studying the features of related complex biological systems. Moreover, it has become one of the fundamental problems in the field of computer science.

In the literature, the problem is modelled as an NP-hard problem, as several studies based on optimization algorithms have been directed toward detecting biological community structure [6, 8-12]. This paper proposes an extension of a PSO algorithm called MP-PSO (Modified Parallel Particle Swarm Optimization Algorithm) to detect the hidden topological structures of communities in biological networks. The following are the main contributions of our research paper:

- A modified particle swarm optimization algorithm with a local search operator is proposed to identify the community structure in biological networks by optimizing the general modularity density function.

- Different evolutionary operators were employed to examine the differences in the ability of the PSO algorithm when extracting community structure in biological networks.
- Parallel computing implementation was used to execute the PSO algorithm to reduce the computation time needed to solve the problem.

The rest of the paper is structured as follows: Section 2 describes the proposed MP-PSO algorithm in detail. Section 3 presents the settings for the experiments, including the dataset used to evaluate the proposed algorithm, the evaluation metrics, and the experimental results obtained. The last section summarizes the paper with concluding remarks.

2. Literature Review

Many research efforts have been made to adopt optimization algorithms to identify complex communities in different types of biological networks [6, 8-12]. Pizzuti and Rombo [8, 9] solved the issue of discovering protein complexes in protein interaction networks by employing diverse quality functions as fitness functions, namely modularity (Q), Community Score (CS), Conductance (CO), Normalized Cut (NC), Internal Density (ID), Expansion (EX), Cut Ratio (CR) and Scaled Cost Function (SCF). The experimental campaign showed that predictions obtained by the single quality functions, defined above, under the GA framework were often more precise than those found by the counterpart methods from the other categories.

In 2017, six metaheuristic optimization algorithms were proposed by Atay et al. [6] to analyze complex networks and discover important communities. The employed algorithms were modified to address the modularity problem that represents a discrete optimization problem. A comparison analysis among the suggested algorithms was carried out on 4 biological networks and 5 social networks. According to the experimental results gained, it was observed that HDSA (Hybrid Differential Search Algorithm) was more efficient and competitive than the other algorithms.

M'barek et al. [10] proposed an approach based on Genetic Algorithm for discovering communities in a gene interaction network. They introduced special representation for solution coding and applying mutation operators. In addition, they proposed a particular fitness function that depends on the value of the interaction between genes and the measure of similarity. Experiments with real data demonstrated the ability of the proposed approach to successfully discover existing or even new communities.

Recently, Abduljabbar et al. [11] solved the issue of detecting protein complexes in PPI (protein-protein interaction) networks by enhancing the performance of the evolutionary algorithm with a local biological operator based on gene ontology annotations, which helped in guiding the search process towards discovering hyper-connected and biologically related complexes, as well as allowing for more effective exploration of the state space of all possible solutions. The authors employed eight quality functions as fitness functions of [8, 9] which are Q, CS, CO, NC, ID, EX, CR, and SCF, the results of the systematic experiments showed the positive effect of applying the heuristic- biological operator, which considerably enhanced the reliability of the adopted optimization models.

Wang et. al. [12] proposed a novel method called IMA (Improved Memetic Algorithm) for discovering complexes in PPI networks. Firstly, IMA combined the biological and topological properties to construct a weighted protein interaction network and to reduce network noise. Secondly, IMA integrated different clustering results to generate initial populations. Additionally, a fitness function was developed depending on 5 topological properties of protein complexes. Lastly, they described the main steps of the proposed IMA method, which are:

selection and recombination operators, local optimization strategy, and population updating operator. The empirical results demonstrated that IMA performed much better than the existing state-of-the-art methods and the base techniques.
Despite the strength of these approaches, handling large-scale complex networks, is the main challenge regarding the application of evolutionary approaches, due to their relatively long execution time. Another limitation is the slow convergence of the evolutionary algorithms and getting stuck at different local optima points when applied with traditional perturbation operators to solve a biological community detection problem. Thus, this paper proposes an extension of a PSO algorithm called MP-PSO (Modified Parallel Particle Swarm Optimization Algorithm) to detect the hidden topological structures of communities in biological networks. This new extension is supplied with a local search strategy to improve the convergence rate of the PSO algorithm and implemented using parallel computing implementation to reduce the required computational time.

3.Proposed Method

Among the evolutionary algorithms, the PSO algorithm has been customized and developed more due to its simplicity, presence of a variety of particle swarm optimization methods, ease of implementation, and not requiring to know the number of communities as input [13, 14]. In general, PSO relies on the animals' behavior that live together in groups and have some ability to interact with each other and the environment in which they are inserted. The particles update the information to correct their positions and velocities using the information received [15].

The proposed MP-PSO algorithm to detect hidden topological structures of communities in biological networks is described in detail in this section. The MP-PSO framework can be summarized in two key steps which include: initialization (in terms of solution representation scheme and fitness function computation), and particle movement strategy (i.e., search strategy) [16,17]. Figure 1 shows a flowchart of the proposed MP-PSO algorithm, which is based on the framework of the PSO algorithm proposed by [16] and [17], more details are provided in the following sub-sections.

Generally, community detection problems have been solved using metaheuristic optimization methods due to their intractable computations and complexity where they have been categorized as highly combinatorial optimization NP-hard problems [18,19], and because solving it takes a long time form the CPU. most of the current metaheuristic methods suffer from slow convergence when applied to solve large-scale problems [20]. Therefore, this work improves the performance of the PSO in terms of speed by exploiting the inherently parallel nature of the metaheuristic algorithm and implementing parallel computing to execute the proposed MP-PSO algorithm. In this study, the parallel computing method available in MATLAB was used [21] as a measure to reduce the computational time needed to run the proposed algorithm.

3.1Individual Representation and Population Initialization

The MP-PSO algorithm adopts the representation based on community labels as a scheme for particles' representation. A position vector p of a given particle indicates a biological network partition. Given a particle $p_i = \{p_i^1, p_i^2, \ldots., p_i^n\}$, where n represents the total number of network nodes and $p_i^x \in [1, n]$. If $p_i^x == p_i^y$, then the nodes x and y are allocated to the same community. This coding scheme is characterized by ease of implementation, and automatic identification of the number of communities, and reduces computational complexity [22].

To initialize the population, random individuals (i.e., $P = \{p_1, p_2, \ldots, p_N\}$) are generated. Where the community label for each node in each particle is randomly chosen from 1 to n. In order to enhance the proposed algorithm convergence, a node (x) was randomly selected and its community label (p_i^x) was assigned to all of its neighbors, this process was applied for each individual $(0.2 * n)$ times [23].

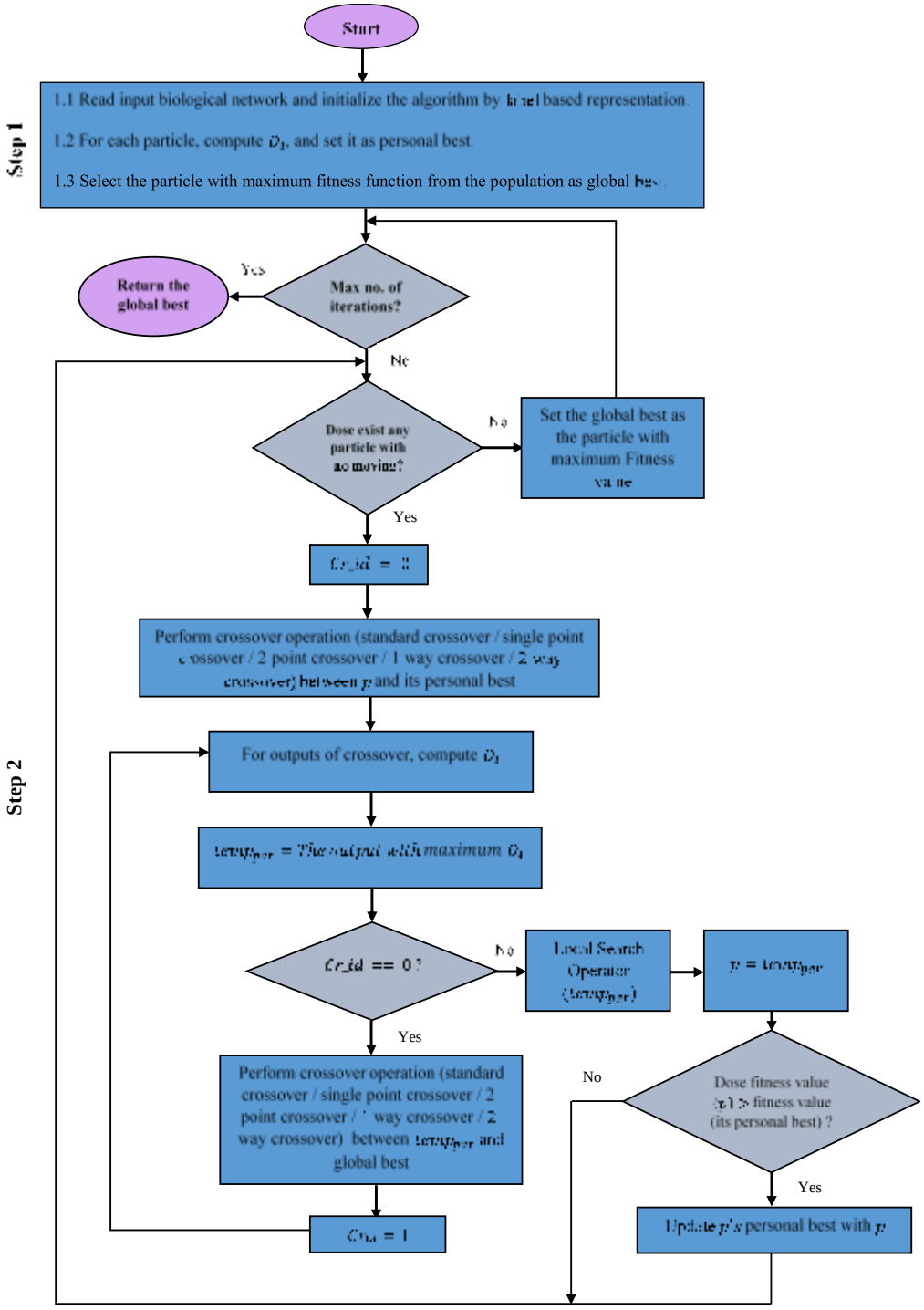

Figure 1: The flowchart of the proposed MP-PSO algorithm

3.2Fitness Computation

In 2008, Li et al. [24] developed a modularity density metric (D) to avoid the issue of resolution limit in the modularity (Q) function. After that, they designed the general modularity

density (D_λ) to discover community structure at diverse resolution levels, by inserting λ, a new parameter, into D function [25]. Accordingly, in this study, we have employed the general modularity density (D_λ) as a quality metric for the proposed MP-PSO algorithm to discover complex communities in biological networks.

Mathematically, let $\mathcal{N}$ represents a given complex biological network consisting of a set of vertices/nodes (V) and edges (E). To represent $\mathcal{N}$, we have defined $n \times n$ binary matrix (A), where n represents the number of biological units or vertices (V) in the network, such that $A_{ab} = 1$ if $(v_a, v_b) \in E$, otherwise $A_{ab} = 0$. Let suppose that V_x and V_y represent the set of vertices of the sub-networks C_x and C_y, respectively, then $E(V_x, V_y) = \sum_{v_i \in V_x, v_j \in V_y} A_{ij}$ indicates the edges number among C_x and C_y. $E(V_x, V_x) = \sum_{v_i, v_j \in V_x} A_{ij}$ indicates the inner degree of C_x, and $E(V_x, \bar{V}_x) = \sum_{v_i \in V_x, v_j \in \overline{V_x}} A_{ij}$ indicates the outer degree of C_x wherein $\bar{V}_x = V - V_x$. Given k clusters C_1 (V_1, E_1), ……, C_k (V_k, E_k) of a certain biological network ($\mathcal{N}$) that captured from a particle (p) , then D_λ (general modularity density) metric can be expressed as follows:

$$D_\lambda = \sum_{x=1}^{k} d(C_x) = \sum_{x=1}^{k} \frac{2\lambda E(V_x, V_x) - 2(1-\lambda) E(V_x, \bar{V}_x)}{|V_x|} \quad (1)$$

wherein $d(C_x)$ represents the difference between two terms $\frac{E(V_x, V_x)}{|V_x|}$ and $\frac{E(V_x, \bar{V}_x)}{|V_x|}$ of C_x [25]. Each particle in the swarm is evaluated using D_λ function and the particle with the highest D_λ value is selected as the best global particle in the swarm. To explore topological structure of a complex biological network at various resolution levels, the λ parameter is utilized. The network is often grouped by D_λ into large-sized communities when the value of λ is close to zero, or into small- sized communities when the value of λ is close to one. Whereas when $\lambda = 0.5$, the D_λ function will group the network similarly to D [24-26].

3.3Search Strategy

The PSO algorithm is based on a distinct search strategy that focuses on directing each particle toward its best local location as well as towards the best global location in the whole swarm. Perturbation operators such as crossover are employed to guide each particle towards the best locations [16, 17]. Furthermore, a local search strategy was applied to improve the accuracy of the PSO algorithm. Following is an explanation of the steps of the search strategy adopted in the application of MP-PSO.

3.3.1 Crossover

Initially, each particle carries out a crossover operation with its personal best location. Generally, crossover is a perturbation operator that takes two parents (p^1, p^2) and generates new solutions by mixing sub-space units of the parents. There are many different methods for combining the parent's sub-space [27]. In this study, we employed five different common types of crossover operators [28], namely standard crossover, single point crossover, 2-point crossover, 1-way crossover, and 2-way crossover in order to study their effect on the PSO detection ability. In sum, the main aim behind employing different crossover operations was to study and analyze the effect of crossover operation on the ability of PSO detection and extract the complex clusters in biological networks. The crossover operators we used to move particles towards personal best are.

1) Standard crossover: consider the two particles p^1 and p^2 as the participating parents in the crossover. A child $child^1$ can be produced from two parents by mixing their elements based on generating a random number $rand$ between 0 and 1 where the community label is copied from

the first parent when $rand \leq 0.5$. On the other hand, the community label is copied from the second parent when $rand > 0.5$. Table 1 illustrates the standard crossover operator based on generating a random number $rand$. This crossover guarantees the generation of offspring that fully exploits the genetic information coming from the parents.

Table 1: Example of standard crossover operator

index	p^1	p^2	$rand \sim [0,1]$	$child^1$
1	1	2	0.4 ($rand \leq 0.5$)	1
2	1	5	0.6 ($rand > 0.5$)	5
3	3	5	0.7 ($rand > 0.5$)	5
4	2	5	0.1 ($rand \leq 0.5$)	2
5	2	1	0.5 ($rand \leq 0.5$)	2
6	4	1	0.3 ($rand \leq 0.5$)	4
7	4	2	0.9 ($rand > 0.5$)	2
8	4	4	0.6 ($rand > 0.5$)	4
9	3	4	0.8 ($rand > 0.5$)	4
10	5	3	0.2 ($rand \leq 0.5$)	5

2) Single point crossover: given the two selected parents (p^1, p^2), first, one arbitrary point a is identified. In order to produce the 1st child, its community label is copied from the beginning of the first parent p^1 to point a, and from point a to the end of the second parent p^2. While the 2nd child is produced by doing this action in reverse order [27]. Table 2 shows how the single point crossover operator is applied to the community detection task when $a = 4$.

Table 2: Example of single point crossover operator when $a = 4$

index	p^1	p^2	$child^1$	$child^2$	p^1	p^2	index
1	1	2	1	2	1	2	1
2	1	5	1	5	1	5	2
3	3	5	3	5	3	5	3
4	2	5	2	5	2	5	4
5	2	1	1	2	2	1	5
6	4	1	1	4	4	1	**6**
7	4	2	2	4	4	2	7
8	4	4	4	4	4	4	8
9	3	4	4	3	3	4	9
10	5	3	3	5	5	3	10

Table 3: Example of 2-point crossover operator (where $a = 7$ and $b = 9$)

index	p^2	p^1	$child^1$	$child^2$	p^2	p^1	index
1	2	1	1	2	2	1	1
2	5	1	1	5	5	1	2
3	5	3	3	5	5	3	3
4	5	2	2	5	5	2	4
5	1	2	2	1	1	2	5
6	1	4	4	1	1	4	6
7	**2**	4	**2**	**4**	2	**4**	7
8	**4**	4	**4**	**4**	4	**4**	8
9	**4**	3	**4**	**3**	4	**3**	9
10	3	5	5	3	3	5	10

3) 2-point crossover: to apply the 2-point crossover, first, two arbitrary points a and b are identified. In order to produce the 1st child, its community label is copied as follows: from the beginning of the first parent p^1 to point a, then from point a to point b from the second parent p^2 and the remainder is copied from the first parent p^1. To produce the second child, this action is done in reverse order [16, 17]. Table 3 illustrates the 2-point crossover operator when two arbitrary points are selected, $a = 7$ and $b = 9$. Accordingly, two new children are produced by mixing elements of $p^1$and p^2 in order to explore the search landscape more broadly.

4) 1-way crossover: given the two selected parents (p^1, p^2), a random node a in the p^1 particle is selected to create a new solution $(child^1)$ by transferring in the p^2 particle the community label of the point a to the node a, and to all the nodes having the same community label of point a in the p^1 [28]. Table 4 illustrates the 1-way crossover operator when $a = 6$ and its $label\ community = 4$. Accordingly, the nodes of the new solution $(child^1)$ have the same community label of the p^2 particle nodes, excluding the locations $\{6;\ 7;\ 8\}$. Since node 7 and node 8 have the same community label of the node 6. The community label of nodes $\{6;\ 7;\ 8\}$ is altered to 4.

Table 4: Example of 1-way crossover operator (where $a = 6$)

index	p^1	p^2	$child^1$
1	1	2	2
2	1	5	5
3	3	5	5
4	2	5	5
5	2	1	1
6	**4**	1	4
7	**4**	2	4
8	**4**	4	4
9	3	4	4
10	5	3	3

5) 2-way crossover: generally, the 2-way crossover operator is a modified version of the 1-way crossover. The procedure for the 2-way crossover operator is defined as follows given the two selected parents (p^1, p^2),a random node a in the p^1 particle is selected, and then its community label is determined (i.e.,p_i^1) to ensure that all the nodes in this community of p^1 are also clustered into the same community in the particle p^2 (i.e., $p_k^2 \leftarrow p_i^1, \forall k \in \{k|p_k^1 = p_i^1\}$). Simultaneously, the community label of the node a in the p^2 particle is also determined and ensuring that all the nodes in this community of p^2 are also clustered in the same community in the particle p^1 (i.e., $p_k^1 \leftarrow p_i^2, \forall k \in \{k|p_k^2 = p_i^2\}$). Accordingly, two new solutions $child^1$ and $child^2$ are returned after applying the above procedure [29]. Table 5 illustrates the 2-way crossover operator when $a = 6$. As shown in Table 5, one of the resulting solutions is the $child^1$, whose nodes {6, 7, 8} was placed in the same detected community with node 9.

Table 5: Example of 2-way crossover operator (where $a = 6$)

index	p^1	p^2	$child^1$	$child^2$	p^1	p^2	index
1	1	2	2	1	1	2	1
2	1	5	5	1	1	5	2
3	3	5	5	3	3	5	3
4	2	5	5	2	2	5	4
5	2	1	1	**1**	2	**1**	5
6	**4**	1	**4**	**1**	4	**1**	**6**
7	**4**	2	**4**	4	4	2	7
8	**4**	4	**4**	4	4	4	8
9	3	4	4	3	3	4	9
10	5	3	3	5	5	3	10

After applying crossover operation (it can be any of the above types depending on what was previously defined) between each particle and its personal best location, new particles are obtained. The obtained output is compared, and the particle (or solution) with the highest fitness value is chosen to be a temporary position of the current particle.
Next, each particle is moved towards the global best in the swarm. Based on that, a crossover operator between particle temporary position and the global best particle is carried out. In this step, new solutions are also produced and compared to identify the temporary state of the current particle.

3.3.2 Local search operator

Finally, a local search operator is applied to all particles to enhance moving them over the whole search space.

Based on community score $CS\,(p) = \sum_{x=1}^{k} \left(\frac{2E(C_x)}{|C_x|}\right)^2$ [9, 11], a local search operator was developed. A node v from community C will change its value to a potential new community label $\acute{C}$ where CS of $\acute{C}$ with node v_i is larger than CS of C with node v_i. More clearly, if and only if CS of community C of node v_i is smaller than CS of other communities when containing node v $(i.e.\ \acute{C} \cup v_i)$, then, there will be a migration of node v_i by the local search operator to another community, which will achieve with its nodes a highest CS value.. Algorithm 1 represents the major steps of the proposed local search operator based on CS. It accepts as input the genotypic representation of a given child individual (i.e., p), the number of nodes in the network (n), and the adjacency matrix (A).

Algorithm 1: Local Search Operator (p, n, A)

for $i = 1$ ***to*** n **do**
  // change the community label of node v_i with control
  set $C_i \leftarrow Community\_Label(v_i)$
  set $K \leftarrow max\ (Community\_Label(p))$
  set $k\_v_i\_CS \leftarrow CS(C_i)_{v_i \in C_i}$
  set $k\_v_i\_\acute{C}S\ _{\acute{C}_j \in p \wedge i \neq j} \leftarrow CS(\acute{C}_j \cup\ v_i)$
  if $(k\_v_i\_\acute{C}S\ >\ k\_v_i\_CS)$
    set $\acute{C} \leftarrow argmax_{\acute{C}_j \in p, \acute{C}_j \neq C_i}(CS(\acute{C}_j \cup\ v_i))$
    set $Community\_Label(v_i) \leftarrow \acute{C}$
  end if
end for

The results of the particle after applying local search operator are compared to its own personal best, and the new particle will take the place of the personal best if it performs better. Otherwise the personal best is still the same. Fitness values are then once more computed using the D_λ measure after all the particles have moved and their personal best values have updated in order to choose the global best for the entire swarm which is the particle with the highest fitness value. This operation is repeated until the predetermined maximum number of iterations is reached.

4. Experiments and Evaluation

4.1Dataset

Two biological networks, namely C. elegans metabolic reaction [30, 6], and Cattle protein-protein interaction [6] were used in this study. The information about each biological network including the relevant references, the number of nodes(n), and the number of edges (m) are summarized in Table 6. It is worth noting, that each biological network was handled as an unweighted and undirected network in all experiments.

Table 6: Characteristics of the complex biological networks used in this study

Networks	Referred to as	# *Nodes* (n)	# *Edges* (m)	References
C. Elegans Metabolic Reaction	C. Elegans MRN	453	2025	[30, 6]
Cattle Protein-Protein Interaction	Cattle PPI	268	303	[6]

4.2Evaluation Measures

In this paper, we have used a modularity (Q) measure to assess the quality of the obtained complex community structures. Modularity (Q) is the most common-internal quality measure proposed by Newman and Girvan in 2004 [31] and has been used primarily to assess predicted solutions when the true solution is unknown. Q scales the difference between the numerical quantity of edges that fall within communities and what would be expected if the edges were randomly placed. Generally, the value of Q measure approaches from zero, if the number of inner edges does not differ from the random distribution of edges. Otherwise, the value of Q measure approaches from one when all the captured communities have dense inner connections. In sum, a high value in terms of Q can be obtained when the network includes strong community structures. The modularity measure is defined as:

$$Q(p) = \sum_{x=1}^{K} \left[\frac{2|E(C_x)|}{|E(p)|} - \left(\frac{\sum_{v \in C_x} deg\ (v)}{2|E(p)|} \right)^2 \right] \quad (2)$$

4.3Results of MP-PSO on Biological Networks with Different Versions

All the experimental studies have been performed on the MATLAB platform. Parameters and features were pre-determined, such as the maximum number of iterations to be equal to 100, and the number of particles in the population to be equal to 100. Moreover, the parameter λ was set in the range from 0 to 1(i.e., $\lambda \in [0,1]$) with step size 0.1. In order to compare the test results of the proposed algorithm at five different versions: $MP\text{-}PSO\text{-}SX$ (represents the PSO version with standard crossover), $MP\text{-}PSO$ -$1PX$ (represents the PSO version with single point crossover), $MP\text{-}PSO$ -$2PX$ (represents the PSO version with 2-point crossover), $MP\text{-}PSO$ -$1WX$ (represents the PSO version with 1-way crossover), and $MP\text{-}PSO$ -$2WX$ (represents the PSO version with 2-way crossover). Each proposed version of MP-PSO was run 30 times on each biological network.

To analyze the test results, first, the results obtained by the MP-PSO algorithm at five different versions on the two complex biological networks when $\lambda = \{0.0, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1\}$ are discussed in terms of average modularity (Q_{avg}),maximum modularity (Q_{max}), and worst modularity (Q_{worst}) as shown in Figure 2. The (Q_{worst}) values obtained at the end of all runs represent the lowest values of Q.

The results presented in Figure 2 reveals several notable points. Firstly, it was seen that all versions of $MP\text{-}PSO$ showed good performance when $\lambda = \{0.4,0.5,0.6,0.7\}$ on C. Elegans MRN network, and when $\lambda = \{0.2,0.3,0.4,0.5\}$ on cattle PPI network. Secondly, it was noted that the detected biological community structure using $MP\text{-}PSO$ -$1PX$ (i.e., PSO with single point crossover) and $MP\text{-}PSO$ -$2PX$ (i.e., PSO with 2- point crossover) achieved the best results in terms of Q_{avg} , Q_{max} and Q_{Worst} over the C. Elegans MRN network. While in the face of Cattle PPI network, $MP\text{-}PSO$ -$1PX$ (i.e., PSO with single point crossover) and $MP\text{-}PSO$ -$1WX$ (i.e., PSO with 1-way crossover) have achieved the best results. On the other hand, the two versions $MP\text{-}PSO$ -$2WX$ and $MP\text{-}PSO$ -SX showed satisfactory performance and recorded good results in most cases. Lastly, over the C.Elegans MRN network, the highest results in terms of $Q_{avg}, Q_{max},$ and Q_{worst} were 0.4066, 0.4192, and 0.3936, respectively, obtained by $MP\text{-}PSO$ -$1PX$ at $\lambda = 0.5$. While over the Cattle PPI network, the highest results in terms of Q_{avg} was 0.7196 obtained by $MP\text{-}PSO$ -$1WX$ at $\lambda = 0.4$ ($with$ $Q_{max} = 0.7407$, $Q_{Worst} = 0.7068$) and $MP\text{-}PSO$ -$1PX$ at $\lambda = 0.5$ (with $Q_{max} = 0.7348$, $Q_{Worst} = 0.7131$).

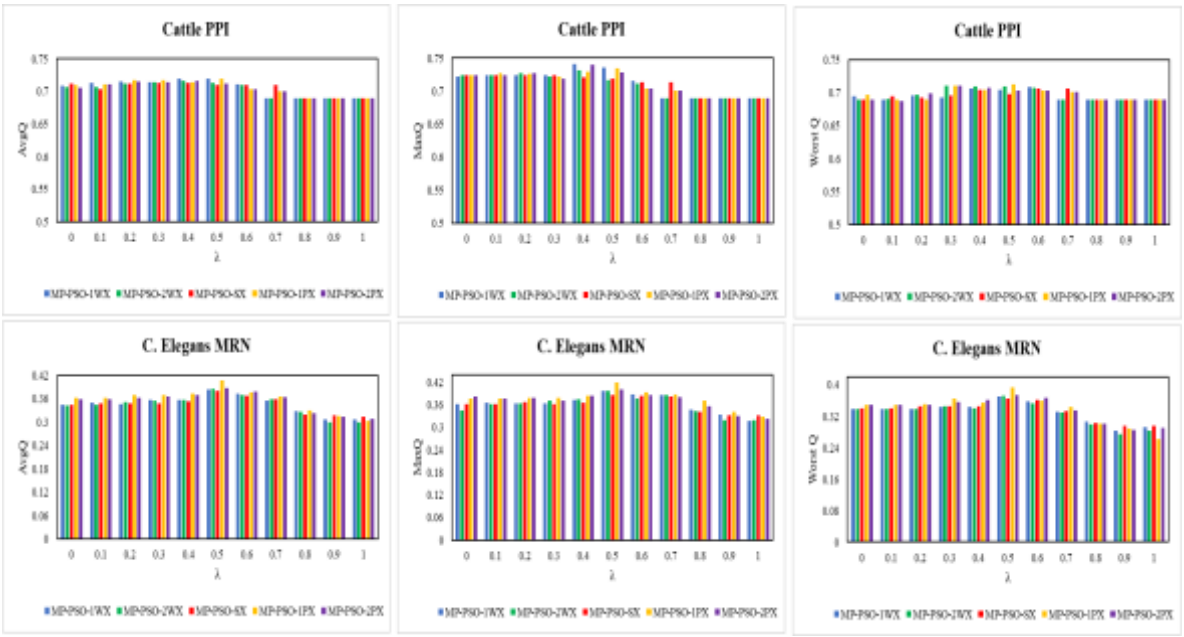

Figure 2: Results of MP-PSO on Cattle PPI and C.Elegans MRN networks with different versions when $\lambda \in [0,1]$ $(left: Q_{avg}, middle: Q_{max}, right: Q_{worst})$

In sum, we can conclude from Figure 2 that the best results in terms of Q_{avg} , Q_{max} and Q_{Worst} were obtained when $\lambda = 0.5$, and employing single point crossover inside PSO framework was more effective in clustering biological networks compared with other crossover types. Accordingly, the effectiveness of the proposed method at $MP\text{-}PSO$ -$1PX$ version and $\lambda = 0.5$ was compared against other state-of-the-art methods. Table 7 reports the performance of the proposed method against six metaheuristic optimization algorithms that were proposed by Atay *et al*. [6]. The results of these metaheuristic algorithms (HDSA: Hyper-heuristic Differential Search Algorithm, BADE: Bat Algorithm based on Differential Evolutionary algorithm, SSGA: Scatter Search algorithm based on Genetic Algorithm, BB–BC: Big Bang-Big Crunch algorithm, BA: Bat Algorithm, and GSA: Gravitational Search Algorithm), in Table 7, are taken from the work of Atay *et al.* [6].

Table 7: Comparison of the proposed method performance (MP-PSO -1PX) with six different optimization algorithms

Networks/Algorithms		HDSA	BADE	SSGA	BB–BC	BA	GSA	MP-PSO -1PX
C.Elegans MRN	Mean	**0.4074**	0.3385	0.3220	0.3266	0.3438	0.3039	0.4066
	Worst	**0.3962**	0.3335	0.3124	0.3194	0.3356	0.2974	0.3936
	Best	0.4185	0.3473	0.3336	0.3374	0.3514	0.3063	**0.4192**
Cattle PPI	Mean	0.7195	0.7138	0.7079	0.7084	0.71	0.6983	**0.7196**
	Worst	**0.7194**	0.7059	0.7052	0.7079	0.7063	0.6949	0.7131
	Best	0.7195	0.7183	0.7118	0.7095	0.7143	0.7053	**0.7348**

The results presented in Table 7 highlighted several positive points. Firstly, due to employing local search based on community score measure within PSO algorithm framework, the proposed algorithm ($MP\text{-}PSO$ -$1PX$) achieved the highest performance and outperformed all the works presented in Table 7 (HDSA, BADE, SSGA, BB–BC, BA, and GSA) in terms of best modularity ($i.e.,\ Q_{max}$) over the two tested biological networks. The extraction of the strong community structure was more emphasized, when applying the local search operator inside PSO algorithm framework.

Secondly, the findings also indicated that the proposed $MP\text{-}PSO$ -$1PX$ algorithm outperformed all counterpart algorithms in terms of mean modularity ($i.e.,\ Q_{avg}$) over Cattle PPI network. While over C.Elegans MRN network, the proposed algorithm outperformed all counterpart algorithms except HDSA algorithm where it has obtained the same value when looking at only two decimal places. On the other hand, when examining all the experimental results in terms of worst modularity ($i.e.,\ Q_{worst}$), the second-best results were usually obtained by our proposed algorithm.

4.4Parallel Computing

Because of its evolutionary nature, the PSO algorithm can be amenable to parallelization, over multiple iterations. The implementation of the proposed MP-PSO algorithm shows a clear acceleration in performance when employing more than one processor, called 'lab' or 'worker'. Therefore, this study exploited the total number of workers available when implementing the algorithm. The acceleration in the PSO performance is computed by measuring the elapsed time required (in seconds) to perform the algorithm in different cases, that is, when 1 to 8 workers are used. The five versions of MP-PSO algorithm complete their run when implemented serially (with No Parallel Implementation (NoPI)) up to the time to complete the run with parallel implementation using x workers (i.e., $x = \{1,2,3,4,5,6,7,8\}$). Figures 3 and 4 illustrate the elapsed time (in second) when the five versions of MP-PSO were run on the biological network (C.Elegans MRN and Cattle PPI) using workers up to 8.

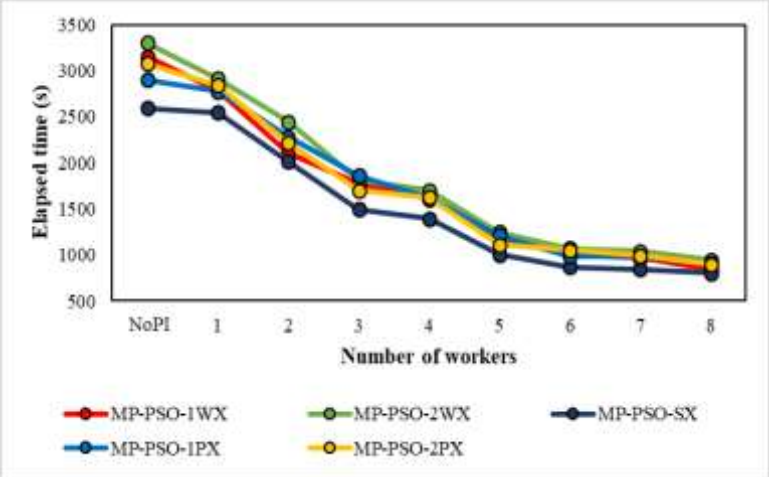

Figure 3: The elapsed time required by x workers to perform the MP-PSO algorithm at different versions on the biological network (C.Elegans MRN)

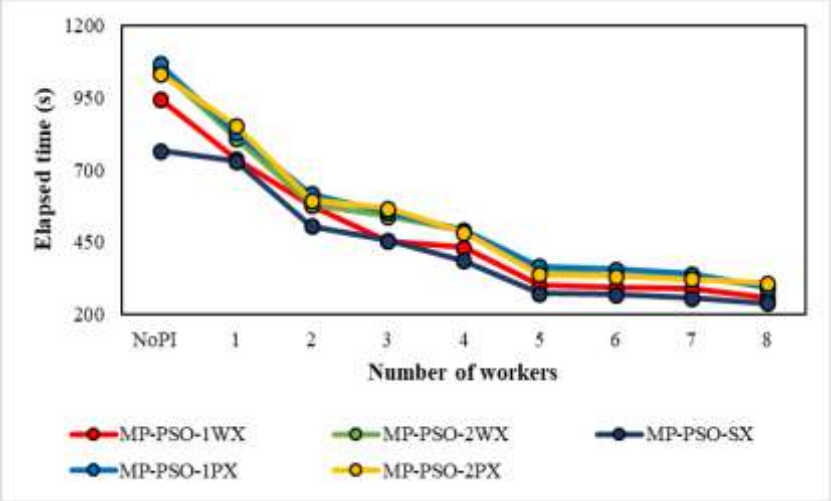

Figure 4: The elapsed time required by x workers to perform the MP-PSO algorithm at different versions on the biological network (Cattle PPI)

From Figures 3 and 4, we can see that the computational time required to implement all versions of the MP-PSO algorithm was reduced by exploiting the inherently parallel nature of metaheuristic algorithms and implementing parallel computing in both objective evaluation and local search operator. Most interestingly, it was observed that the algorithm at MP-PSO-SX version was faster than all the other four versions.

5. Conclusion

In this paper, a modified parallel particle swarm optimization algorithm with a local search operator, named MP-PSO, is proposed to identify complex biological communities by optimizing the general modularity density metric as a fitness function. Five different evolutionary operators were employed to study the variation in the capability of PSO to extract complex biological communities. Moreover, the proposed algorithm was implemented using parallel computing method to reduce the computational time required for its execution. To verify the performance of the proposed algorithm, two real biological networks were examined. The experimental results showed that MP-PSO with single point crossover (MP-PSO -1PX) detected the best results compared to the other employed crossover types, while the MP-PSO with standard crossover was the fastest among the other four versions.

Additionally, the proposed MP-PSO -1PX algorithm was compared with six competent state-of-the-art algorithms on the two biological networks. Experimental results showed that MP-PSO -1PX has promising performance in identifying complex communities in biological networks. In the future, however, the PSO's performance can be enhanced using an improvement operator based on biological information (i.e., gene expression and gene ontology), and adopt other known community detection models as a fitness function.

References

[1] M. E. J. Newman, "*Networks: An Introduction", Oxford University Press, Oxford*, 2010.

[2] P. Sah, L. O. Singh, A. Clauset and S. Bansal, "Exploring community structure in biological networks with random graphs", *BMC bioinformatics*, vol. 15, no. 1, pp. 1-14, 2014.

[3] O. Gach and J. K. Hao, "A memetic algorithm for community detection in complex networks", In *International conference on Parallel Problem Solving from Nature*-PPSN XII, Springer Berlin Heidelberg, vol. 7492, pp. 327-336, 2012.

[4] S. Fortunato, "Community detection in graphs," *Physics reports*, vol. 486, no. 3-5, pp. 75-174, 2010.

[5] M. E. J. Newman, "The structure of scientific collaboration networks", *Proceedings of the National Academy of Sciences*, vol. 98, no. 2, pp. 404-409, 2001.

[6] Y. Atay, I. Koc, I. Babaoglu and H. Kodaz , " Community detection from biological and social networks: A comparative analysis of metaheuristic algorithms", *Applied Soft Computing*, vol. *50*, pp. 194-211, 2017.

[7] D. A. Abduljabbar, S. Z. M. Hashim and R. Sallehuddin, "An improved multi-objective evolutionary algorithm for detecting communities in complex networks with graphlet measure," *Computer Networks*, vol. 169, pp. (107070-1)-(107070-24), 2020.

[8] C. Pizzuti and S. Rombo, " Experimental evaluation of topological-based fitness functions to detect complexes in PPI networks", *In Proceedings of the 14th annual conference on Genetic and evolutionary computation*, pp. 193-200, 2012.

[9] C. Pizzuti and S. E. Rombo, "Algorithms and tools for protein–protein interaction networks clustering, with a special focus on population-based stochastic methods", *Bioinformatics*, vol. 30, no. 10, pp. 1343-1352, 2014.

[10] M. B. M'barek, A. Borgi, W. Bedhiafi and S. B. Hmida, " Genetic algorithm for community detection in biological networks", *Procedia Computer Science*, vol. *126*, pp. 195-204, 2018.

[11] D. A. Abduljabbar, S. Z. M. Hashim and R. Sallehuddin, "An enhanced evolutionary algorithm for detecting complexes in protein interaction networks with heuristic biological operator," In *International Conference on Soft Computing and Data Mining*, pp. 334-345, January 2020, Springer, Cham.

[12] R. Wang, H. Ma and C. Wang, "An Improved Memetic Algorithm for Detecting Protein Complexes in Protein Interaction Networks", *Frontiers in genetics*, vol. *12*, pp. 794354-794354, 2021.

[13] M. Pourkazemi and M. R. Keyvanpour, "Community detection in social network by using a multi-objective evolutionary algorithm," *Intelligent Data Analysis*, vol. 21, no. 2, pp. 385–409, 2017.

[14] D. A. Abduljabbar, "Community Detection in Modular Complex Networks Using an Improved Particle Swarm Optimization Algorithm", *Iraqi Journal of Science*, vol. 64, no. 8, pp 4228-4243, 2023.

[15] R. N. Jawad, "Proposed Hybrid Technique in Cryptanalysis of Cryptosystem Based on PSO and SA," *Iraqi Journal of Science*, vol. 63, no. 10, pp 4547-4558, 2022.

[16] S. Rahimi, A. Abdollahpouri and P. Moradi, "A multi-objective particle swarm optimization algorithm for community detection in complex networks," *Swarm and Evolutionary Computation*, vol. 39, pp. 297-309, 2018.

[17] A. Abdollahpouri, S. Rahimi, S. M. Majd, and C. Salavati, "A modified particle swarm optimization algorithm for community detection in complex networks," In *International Cross-Domain Conference for Machine Learning and Knowledge Extraction*, pp. 11-27, Springer, Cham, 2018, doi: 10.1007/978-3-319-99740-7_2.

[18] D. A. Abduljabbar, S. Z. M. Hashim, and R. Sallehuddin, "An enhanced evolutionary algorithm with local heuristic approach for detecting community in complex networks," *Journal of Theoretical and Applied Information Technology*, vol. 97, no. 20, pp 2452-2466, 2019.

[19] D. A. Abduljabbar, S. Z. M. Hashim, and R. Sallehuddin, "An Evolutionary Algorithm for Community Detection Using an Improved Mutation Operator," In *2019 International Conference on Electrical Engineering and Computer Science (ICECOS)*, pp. 406-410, 2019. IEEE.

[20] L. Samandari Masooleh, J. E. Arbogast, W. D. Seider, U. Oktem and M. Soroush, " An efficient algorithm for community detection in complex weighted networks", *AIChE Journal*, vol. 67, no. 7, pp. e17205, 2021.

[21] Y. Her, R. Cibin and I. Chaubey, "Application of parallel computing methods for improving efficiency of optimization in hydrologic and water quality modeling", *Applied Engineering in Agriculture*,vol. 31, no. 3, pp. 455-468, 2015.

[22] C. Gao, Z. Chen, X. Li, Z. Tian, S. Li and Z. Wang, "Multiobjective discrete particle swarm optimization for community detection in dynamic networks", *EPL (Europhysics Letters)*,vol. 122, no. 2, pp. 28001, 2018

[23] S. Kaur, S. Singh, S. Kaushal and A. K. Sangaiah, "Comparative analysis of quality metrics for community detection in social networks using genetic algorithm," *Neural Network World*, vol. 26, no. 6, pp. 625-641, 2016.

[24] Z. Li, S. Zhang, R. S. Wang, X. S. Zhang and L. Chen, "Quantitative function for community detection," *Physical review E*, vol. 77, no. 3, pp. (036109-1) (036109-9), 2008.

[25] Y. Zhang, Y. Liu, J. Li, J. Zhu, C. Yang, W. Yang and C. Wen, "WOCDA: A whale optimization based community detection algorithm," *Physica A: Statistical Mechanics and its Applications*, vol. 539, pp. (122937-1) (122937-15), 2020.

[26] I. S. Dhillon, Y. Guan and B. Kulis, "Kernel k-means: spectral clustering and normalized cuts," In *Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining*, pp. 551-556, August 2004.

[27] M. Moradi and S. Parsa, "An evolutionary method for community detection using a novel local search strategy," *Physica A: Statistical Mechanics and its Applications*, vol. 523, pp. 457–475, 2019.

[28] C. Pizzuti, "Evolutionary computation for community detection in networks: A review," *IEEE Transactions on Evolutionary Computation*, vol. 22, no. 3, pp. 464–483, 2018.

[29] M. Gong, B. Fu, L. Jiao and H. Du, "Memetic algorithm for community detection in networks", *Physical Review E*, vol. 84, no. 5, pp. 056101, 2011.

[30] J. Duch and A. Arenas, "Community detection in complex networks using extremal optimization", *Physical review E*, vol. 72, no. 2, pp.027104, 2005.

[31] M. E. Newman and M. Girvan, "Finding and evaluating community structure in networks," *Physical review E*, vol. 69, no. 2, pp. 026113, 2004.