



Research article

Improving modularity score of community detection using memetic algorithms

Dongwon Lee¹, Jingeun Kim² and Yourim Yoon^{1,*}

¹ Department of Computer Engineering, Gachon University, 1342 Seongnam-daero, Sujeong-gu, Seongnam-si, Gyeonggi-do, Republic of Korea

² Department of IT Convergence Engineering, Gachon University, 1342 Seongnam-daero, Sujeong-gu, Seongnam-si, Gyeonggi-do, Republic of Korea

* **Correspondence:** Email: yryoon@gachon.ac.kr; Tel: +82317505326.

Abstract: With the growth of online networks, understanding the intricate structure of communities has become vital. Traditional community detection algorithms, while effective to an extent, often fall short in complex systems. This study introduced a meta-heuristic approach for community detection that leveraged a memetic algorithm, combining genetic algorithms (GA) with the stochastic hill climbing (SHC) algorithm as a local optimization method to enhance modularity scores, which was a measure of the strength of community structure within a network. We conducted comprehensive experiments on five social network datasets (Zachary's Karate Club, Dolphin Social Network, Books About U.S. Politics, American College Football, and the Jazz Club Dataset). Also, we executed an ablation study based on modularity and convergence speed to determine the efficiency of local search. Our method outperformed other GA-based community detection methods, delivering higher maximum and average modularity scores, indicative of a superior detection of community structures. The effectiveness of local search was notable in its ability to accelerate convergence toward the global optimum. Our results not only demonstrated the algorithm's robustness across different network complexities but also underscored the significance of local search in achieving consistent and reliable modularity scores in community detection.

Keywords: genetic algorithm; local search; community detection; modularity; memetic algorithm

Mathematics Subject Classification: 68T20

1. Introduction

As the internet has swiftly grown, there's been a significant increase in the online engagement of users. A variety of social networking platforms, such as Facebook and Twitter, have emerged to support these user interactions. Social networks depict and simulate the social connections between individuals [1]. Also, numerous intricate systems found in biology, sociology, and physics can be modeled as complex networks in the real world [2]. In real-life scenarios, the elements of a complex system can be depicted as a network configuration. At the heart of network development are communities, which serve as the fundamental components [3]. Examining the community framework within a network aids in understanding the social dynamics of group interactions among individuals [4]. Additionally, uncovering and studying the community structure is of significant interest due to its economic and marketing consequences [5].

Community structure in networks typically refers to the natural division of nodes into densely linked clusters where the links between different clusters are relatively infrequent compared to the connections within the same cluster [6]. Pothén et al. [7] introduced the spectral bisection technique, a hierarchical clustering-based method for identifying community structures within networks. One quantitative measure for partitioning a network into communities is modularity, proposed by Girvan and Newman [8]. However, the issue of maximizing modularity is recognized as a nondeterministic polynomial time (NP)-hard [9]. The Girvan-Newman (GN) algorithm, proposed by Girvan and Newman [8], utilizes a divisive hierarchical algorithm and has shown good performance in community detection. However, as the authors note in [8], the GN algorithm requires a huge computational cost, which means it's impractical for large-size community detection. Newman and Moore [10] proposed the Clauset, Newman, and Moore (CNM) algorithm, which employs local search to handle very large community detection problems, but the CNM algorithm has a problem that can easily fall into the local optima. To address this problem, many researchers aimed meta-heuristic algorithms to enhance the modularity of community detection [11–15]. In [16–18], the authors introduced a genetic algorithm (GA)-based method for community detection. In particular, [17,18] used hill climbing and simulated annealing, a mathematical optimization technique, as a local search algorithm.

Current GA-based methods are constrained by issues such as prolonged durations to reach the global optimum and susceptibility to premature convergence. When addressing multi-peak issues that contain numerous optimal solutions, GA tends to get trapped in local minima, ceasing their search prematurely. This results in the challenge of premature convergence and the inability to reach global optima [19]. GAs are adept at quickly identifying the vicinity of the global optimum, yet they require a considerably longer period to find the precise local optimum within the convergence area [20,21]. In a hybrid approach, using local search to refine solutions directed by a GA toward the most favorable area can hasten the process of converging on the global optimum. Utilizing local search techniques and area-specific expertise can additionally shorten the duration required to arrive at the global solution [22].

Here, we propose a memetic algorithm that uses a combination of GA and Stochastic-Hill-Climbing (SHC) for community detection. We applied SHC as a local search to improve the performance of GA and conducted experiments on five datasets (Zachary's Karate Club, the Dolphin Social Network, Books About U.S. Politics, the American College Football Club, and the Jazz Musician Club datasets) with ten runs. Modularity Q was used as the community detection quality and compared to the results of [17] using average and the best modularity scores. In summary, the contribution of our methodology is as follows:

- We applied SHC as a local search mechanism within our memetic algorithm, demonstrating its superiority over other methods through comparative performance metrics.
- We address the limitation of existing GAs by explaining the issues such as extended time to the global optimum. Our study contrasts the efficiencies of approaches with and without local search, providing insight into the substantial benefits of local search in enhancing performance. The detailed analysis of convergence speed and modularity scores in Section 6 not only refines the detection of community structures but also adds to the mathematical rigor, providing clear evidence of the algorithm's efficiency and reliability.
- We provided a comparative analysis with 5 social network datasets, which reveals that the use of local search not only refines the detection of community structures but also expedites the convergence of the algorithm, showcasing the practical advantage of our proposed method.
- We introduced a memetic algorithm framework in this paper, particularly highlighting the effective use of local search within the GA. This framework not only enhances the efficiency of GA but also can be adapted and applied to other optimization problems beyond community detection, contributing a versatile tool to the mathematical and computational toolkit for network analysis. This adaptation showcases the practical advantages and broader applicability of our proposed method.

The structure of this paper is as follows. Section 2 provides the background methodology of this study and a description of the dataset used in this study. Section 3 presents the procedure for how we implemented GAs with an addition of local search, a memetic algorithm. Sections 4 and 5 present the simulation of experiments and results analysis, respectively. Section 6 presents the summary.

2. Background

In this study, we introduce a meta-heuristic approach for community detection that leverages a memetic algorithm. This combines GAs with the SHC algorithm as a local optimization method to enhance modularity scores. By integrating SHC with GA, we aim to address the limitations of existing GA-based methods, such as prolonged durations to reach the global optimum and susceptibility to premature convergence.

The memetic algorithm employed in this study is designed to improve the performance of traditional GAs in community detection tasks. The methodology involves using SHC as a local search mechanism within our memetic algorithm, demonstrating its superiority over other methods through comparative performance metrics.

2.1 Modularity Q

For this study, Modularity Q is taken as the performance measure for community detection. Modularity, as described in [23], is a prominent indicator of the robustness of community structures within networks. It has a propensity to bias in favor of larger communities at the expense of smaller ones, particularly when these smaller groups are beneath a certain scale. Nonetheless, this metric is indispensable in evaluating and contrasting different community detection techniques.

In study [17], a parameter λ is introduced within the modularity density D , where adjusting λ is necessary to achieve optimal community detection outcomes. To simplify the process and enhance the accuracy of detecting communities, this paper opts to use modularity Q instead of D as the primary

objective function, as defined in [23]:

$$Q = \frac{1}{2M} \sum_{ij} \left(a_{ij} - \frac{k_i k_j}{2M} \right) \delta(C_i, C_j), \quad (1)$$

where M denotes the total number of edges in the network, and i and j represent two nodes within the network. The terms k_i and k_j refer to the degrees of the i -th, j -th nodes. The matrix element a_{ij} is located at the i -th row and j -th column of the adjacency matrix. The function $\delta(C_i, C_j)$ describes the relationship between nodes i and j . It equals 1 if nodes i and j belong to the same community, and 0 otherwise.

2.2. Memetic algorithm

GA is a well-known adaptive heuristic search algorithm that is modeled after the principles of natural selection and genetics, originally introduced by Holland at the University of Michigan [24]. GA begins with initially randomly generated populations and evolves through successive generations toward a population of better quality [25]. In particular, memetic algorithms (MAs) are a class of meta-heuristic optimization methods that combine a meta-heuristic algorithm such as the GA with a local search strategy. These local search strategies are employed during the generational cycles of the evolutionary process [26]. To be more specific, the process of MA is similar to GA. After a population is initialized with feasible solutions, offsprings are generated by selection, crossover, and mutation according to probability. After employing genetic operators, local search aids in refining the solutions generated by offspring and circumventing local optima.

From an optimization perspective, MAs have demonstrated greater efficiency and effectiveness compared to traditional GAs in certain problem domains [17]. In particular, MAs have been widely accepted for combinatorial optimization problems and various practical applications [27–30]. Li et al. [31] has proven MAs to have better results than evolutionary algorithms (EAs). Furthermore, outperforming existing optimization algorithms proposed in [32–35]. MAs demonstrate robustness across various network complexities, adapting effectively to different topologies. This adaptability is particularly beneficial for dynamic networks where community structures can change over time. The Robust Dynamic Memetic Algorithm (RDMA)_Net algorithm [36], for example, has shown superior performance in maintaining community detection accuracy and quality under changing network conditions. The integration of local search methods enhances the consistency of community detection results. MAs tend to provide more reliable solutions across multiple runs, ensuring that the detected community structures are stable and repeatable [37]. Therefore, we decided to use MA with SHC as the local search method over other optimization methods.

2.3. Local search

When a GA is adeptly crafted, the best candidate within the population has the potential to evolve into an optimal solution for the given issue. Nonetheless, due to its inherent randomness, a GA might experience sluggish progress toward convergence. Additionally, the precision of the optimal solution it identifies can fluctuate depending on the specific problem being addressed. Owing to these shortcomings, numerous research efforts have focused on enhancing GA's performance [38,39].

Local search is an intuitive approach to tackling combinatorial optimization problems. It starts with an initial solution s that is likely suboptimal and seeks to enhance its value through “local

changes”. Such changes might include adding or removing elements to s , reorganizing how elements are grouped within s , or altering their sequence. If these modifications lead to a better solution, the process yields a new solution s . This iterative method is repeated until it reaches a point where no further advancements can be made [40].

Local search techniques have been effectively applied to discover satisfactory solutions for a myriad of intricate issues, with the traveling salesman problem standing out as the most notable example [41]. The practical efficacy of local search methods has been thoroughly analyzed across various domains such as scheduling, very large scale integration (VLSI) design, network configuration, distributed planning, production management, and beyond [42]. The consensus from most of these analyses is that local search offers an efficient way to compute solutions that are close to the optimum for problems of practical scale [40]. In this study, we have applied a SHC method to find the optimum solution.

2.4. Stochastic hill climbing algorithm

The SHC algorithm [43], a modification of the SHC method, represents an incomplete strategy for tackling optimization issues. It operates through a loop that persistently advances toward higher-value points, akin to ascending a hill. The process concludes upon arriving at a “peak” where none of the surrounding points offer an increase in value. Unlike traditional hill climbing, this variant randomly selects from potential uphill moves, with the selection likelihood adjusting based on the ascent's gradient. It essentially modifies an existing assignment into a collection of slightly altered versions. Each modified version is then evaluated based on certain measures intended to draw nearer to a suitable assignment, thereby enhancing the state's evaluation score. The most favorable modification is then selected as the new assignment. This core procedure is carried out repeatedly until a satisfactory solution emerges or a pre-established stopping condition is met [44].

3. Methods

Figure 1 represents a flow chart of the MA. We will provide a detailed description starting from the selection, crossover, mutation, and local search in this section.

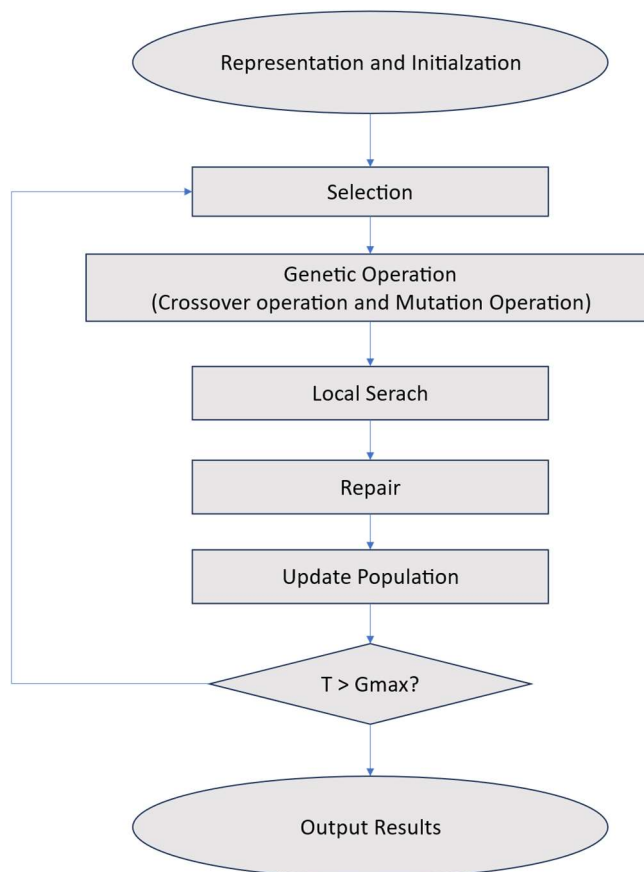


Figure 1. Flowchart of MAs.

3.1. Representation and initialization

The effectiveness of a GA's convergence is often influenced by the initial population's diversity. In the context of actual network applications, the number of communities that need to be established is usually predetermined. In our experiment, we have specified the number of communities before streamlining the search space. We ensure diversity by randomly generating chromosomes for each new generation, while also avoiding infeasible solutions through a careful initialization process. This process ensures that only connected nodes are placed in the same community, while unconnected nodes are segregated at the outset [3].

Each gene in the chromosome holds a value that falls between 0 and $n - 1$, with n being the set number of communities, which is known as label-based representation. We have appraised the algorithm's performance across varying values of n , considering a range of varieties based on the dataset we are using. For instance, Figure 2 illustrates a potential solution for a network of 7 nodes where the number of communities, denoted by n , is fixed at 2. Consequently, the candidate solution's values lie between 0 and 1. As in the scenario depicted in Figure 2, we create chromosome sets for each prescribed value of n . The length of each candidate solution is determined by the total number of nodes in the network.

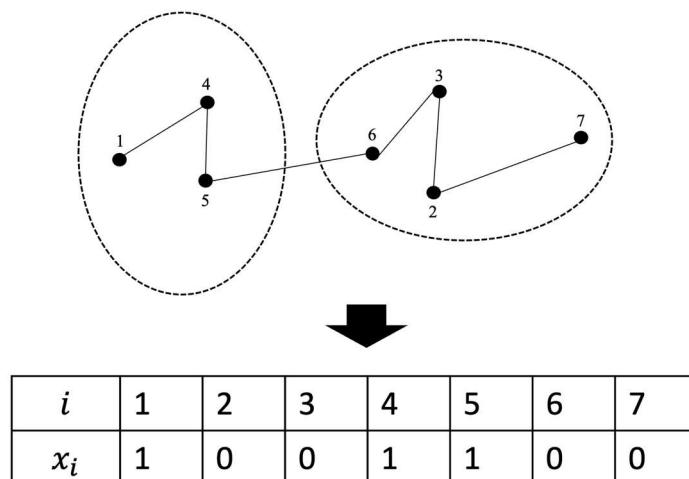


Figure 2. Example of label-based representation.

3.2. Tournament selection

Selection is a key stage in a meta-heuristic algorithm that aids in the extensive search for the optimal solution to a given problem [3]. For this study, we have applied tournament selection as the selection operator. Tournament selection intensifies the search for the best solution by staging a competition among ' S ' participants, where ' S ' denotes the size of the tournament. The competitor with the greatest fitness value from these ' S ' individuals emerges as the winner of the tournament. This selected individual is then placed into the mating pool. Consequently, the mating pool, filled with the winners of these tournaments, possesses an average fitness value that surpasses the general population's average fitness value [45]. In our method, we select a random sample of the tournament size, which is 2. This means that the tournament selection function will conduct a tournament between two competitors at a time from the given population to select individuals for the mating pool. Then, the winner of the tournament in which the chromosome that has the maximum fitness value in the sampled group is selected as the parent chromosome.

3.3. Uniform crossover

In GAs, the creation of the next generation hinges on specific procedures that remix and alter chosen individuals from the existing population. In this study, we used a uniform crossover to generate a new offspring. Two chromosomes are chosen by tournament selection from the current population as parent chromosomes, respectively. For each iteration within every gene of the chromosome, a random number between 0 and 1 is generated. If k (meaning the threshold value) is larger or equal to 0.5, genes from parent 1 are chosen. If k is smaller than 0.5 genes from parent 2 are chosen. Consequently, the diversity of the chromosomes increases and expands the optimal space by appropriately combining the features of both parent chromosomes. For instance, Table 1 illustrates a potential offspring after performing crossover with 2 parents for a network of 7 nodes with a random number r generated as the crossover probability between 0 and 1.

Table 1. Example of uniform crossover.

v	O_x	O_y	k	O_{new}
1	1	3	0.23	1
2	3	2	0.54	2
3	2	3	0.89	3
4	2	2	0.21	2
5	1	2	0.02	3
6	3	1	0.69	1
7	2	3	0.18	2

3.4. Mutation

Beyond recombination mechanisms that generate offspring through the fusion of elements from two parents, another class of operators creates offspring from a parent. Specifically, the mutation operator introduces minor, random alterations to the bit string by randomly selecting a single bit and flipping its value. Typically, the mutation is applied subsequent to the crossover process.

The mutation operator stands as a fundamental component within GAs, inducing random variations in chromosomes [46]. Its role is crucial in augmenting the population's diversity and accelerating convergence. This operator functions by inverting the genetic values at specific loci within chromosomes, doing so with a predefined likelihood. Algorithms based on GA incorporate a variety of crossover operators, including but not limited to boundary, uniform, nonuniform, directional, and Gaussian mutations [47].

In our algorithm, a random vertex is chosen and then assigned a class label in a stochastic manner during the mutation phase. This approach to mutation can streamline the mutation procedure and help in narrowing down the search space, as detailed in Table 2. For instance, in a network with three classes as described in Table 2, we first select a chromosome, say chromosome O . Next, we randomly select a vertex within this chromosome, such as vertex 5. If the class label of vertex 5 is initially 1, the mutation changes it to a different community label, in this case, to 3. Such a mutation enhances the variety of chromosomes, thereby increasing the probability of discovering a higher modularity score Q .

Table 2. Example of mutation.

v		O		O_{new}
1		1		1
2		3		3
3		2		2
4		2		2
5	→	1	→	3
6		3		3
7		2		2

3.5. Repair

In this study, we have a predefined number of communities, but if there are missing communities in an individual's gene, we employed a repair phase. In the repair phase, a two-level nested loop

mechanism is employed. The outer loop iterates over the list of missing labels. For each missing label, the inner loop randomly selects positions in the individual's structure and inserts the missing label into these positions.

This repair process is crucial for preserving genetic diversity while simultaneously respecting the label distribution. The resulting individuals are both diverse and adhere to the problem's constraints, leading to a more effective and robust search for the optimal solution.

3.6. Stochastic hill search

In this study, we have refined our vertex movement heuristic to exploit the neighborhood solution of each solution by employing SHC with an integrated mutation strategy to optimize the neighborhood solutions within a graph. Specifically, we focus on the community detection problem within networks, where our goal is to discover a partitioning of the graph that maximizes modularity—a measure of the strength of the division of a network into communities. Hence, the repositioning of a node from its original community into another, forms an adjacent solution.

In our algorithmic approach, we refine the strategy for community detection in networks by introducing a heuristic based on SHC complemented by genetic operations. The local search procedure works as follows, as shown in Table 3.

Table 3. Pseudo-code of local search.

Algorithm: Stochastic Hill Climbing	
Input: offspring - an array representing a current partition of the graph	
Output: offspring - an improved partition with enhanced modularity	
idx ← random integer between 0 and length(offspring)	<i>// Index Selection</i>
neighbor ← copy(offspring)	<i>// Neighbor Creation</i>
current_modularity ← Q (offspring)	<i>// Fitness Evaluation</i>
change ← \emptyset	
For label index from 0 to num_label do	<i>// Local Change Process</i>
neighbor [idx] ← label index	
Δ modularity ← Q (neighbor) - current_modularity	
change ← Δ modularity	
End For	
max_idx ← index of maximum value in change	<i>// Selection of Best Suboptimal</i>
offspring [idx] ← max_idx	
return offspring	

Table 3 illustrates the SHC local search algorithm, a method developed to refine the partitioning of nodes within a network to maximize modularity, which is indicative of optimal community structure. The input and output of the SHC algorithm is an offspring (a single array), and the main steps are as follows: (1) Select a random index in a single array to change. (2) Compute the current objective function (the modularity Q) of offspring. (3) Enter all possible labels for the randomly selected index to calculate the change from the modularity of offspring. (4) Assign the label with the largest change

to the randomly selected index and return it.

Through local search, GAs evolve into individuals in a better direction and improve diversity. The SHC local search algorithm is thus portrayed as both robust and efficient, capable of effectively navigating the search space to arrive at superior network partitions.

4. Experimental setup

4.1. Dataset

In this study, for the verification of our performance on the proposed GA method to improve modularity scores, we have compared 5 social network datasets from small-scale and simple datasets to large-scale datasets. Our datasets include Zachary's Karate Club [48], the American College Football Club [49], the Dolphin Social Network [50], Books About U.S. Politics [51], and the Jazz Musician Club dataset [52]. These datasets were commonly used in previous studies related to community detection. Shang et al. [18] used a community detection method based on modularity and an improved genetic algorithm (MIGA) on four real-world networks (Zachary's Karate Club, Dolphin Social Network, American College Football, and Books about U.S. Politics) to enhance the normalized mutual information (NMI) score. Pizzuti [53] applied a multi-objective approach to discover communities in networks by employing genetic algorithms named multi-objective genetic algorithm (MOGA-Net) on Zachary's Karate Club, Dolphin Social Network, American College Football, and Books about U.S. Politics. Gurrero et al. [54] applied a new generational genetic algorithm that includes efficient initialization methods and search operators under the guidance of modularity named GGA+ using the Books about U.S. Politics and the American College Football dataset. Shi et al. [55] applied a new genetic algorithm to enhance modularity scores using the Zachary's Karate Club and the American College Football Dataset.

Our selection of these datasets, which have been extensively used and validated in the study of community detection, allows us to benchmark our proposed GA method against established methods. This approach ensures the robustness and reliability of our results, contributing to the existing body of knowledge with a focus on improving modularity scores in community detection.

4.1.1. Zachary's Karate Club

Zachary's Karate Club, often utilized in community detection research, is composed of 34 nodes that represent the club's members. The connections, or edges, between these members, numbering 78 in total, illustrate the interpersonal relationships within the club [18]. Over a two-year observation of a karate club involving 34 members, Zachary witnessed the emergence of a conflict between the club's administrator and its instructor. This discord eventually led to the instructor departing to establish a new club, taking approximately half of the original members with him [8].

4.1.2. Dolphin Social Network

The Dolphin Social Network is a real-world example commonly utilized in testing community detection algorithms. Documented by D. Lusseau, this network maps out the interactions among 62 dolphins off the coast of New Zealand. It is structured into 62 nodes and 159 edges, where each node

signifies an individual dolphin, and the edges represent their interaction frequencies. This network is divided into two main dolphin groups: a larger one with 42 individuals and a smaller group of 20 [56]. The two major communities in the Dolphin Social Network can be divided into five subcommunities [57].

4.1.3. Books about U.S. politics

This social network presents a political book network curated by V. Krebs (through personal communication). This network features nodes that signify 105 recent books on American politics, acquired from the online retailer Amazon.com. Edges connect books often bought by the same customer. The books were categorized based on their expressed or evident political orientation—liberal or conservative—with a few exceptions for books that were distinctly bipartisan, centrist, or lacked a definite political stance [51].

4.1.4. American College Football

The American College Football Dataset maps out the Division I football games for the 2000 season, featuring 115 vertices and 616 edges that symbolize the football teams and the regular season games played between them, respectively. Throughout the season, these 115 teams were organized into 12 conferences, each hosting about 8 to 12 teams. Within this setup, games occurred more frequently among teams within the same conference compared to those from different conferences. This pattern suggests that each conference functions as an individual community within the overall network [58]. The network was proposed by Girvan and Newman. The nodes represent different football teams and the edges represent the matches between two teams. The network consists of 115 nodes and 616 edges. The network consists of 12 communities, which are 12 football teams [18].

4.1.5. Jazz club dataset

Data from The Red Hot Jazz Archive was used to analyze 198 bands active from 1912 to 1940, primarily in the 1920s [52]. The database, listing 1275 musician names within these bands, does not specify the timing of musicians' participation, hindering the study of the network's temporal evolution. Notably, the count of names does not equate to distinct individuals due to aliases and unnamed musicians labeled as "unknown". For instance, Henry Allen is listed under multiple names. The analysis reveals a distribution with most bands having 5 to 10 musicians, and a few larger ones with up to 171 members. In this network, each node represents a jazz musician, and an edge is established between two musicians if they have played together. There are a total of 198 jazz musicians within this network. The average degree of 27.697 indicates that the network's nodes are highly interconnected [59].

4.2. Characteristics and parameters used for datasets

Table 4 describes the characteristics of various network datasets used in the experiments along with the specific parameters set for the MA employed. Table 4 lists five distinct networks: Karate, Dolphin, U.S. Books, Football, and Jazz. For each network, Table 4 specifies the number of nodes, the number of edges, and the number of labels which indicate how many communities each dataset will be divided into, and for the parameters of MA, the population size and the number of generations are

described above. The crossover rate and the mutation rate are identical to all five datasets with rate values of 1.0 and 0.5, respectively. The crossover rate is set to 1.0 to ensure that every gene of the two chromosomes is considered for crossover, aligning with the general guideline to utilize a high crossover rate [60]. It ensures that crossover occurs for every parent chromosome, and all offspring are generated through crossover at every generation. The information of the two parent chromosomes is randomly exchanged through uniform crossover, enhancing the genetic diversity of the offspring.

Table 4. Characteristics of the used dataset and parameters of the MA.

Network	# nodes	# edges	# label	Population Size	Generation
Karate	34	78	2	100	50
Dolphin	62	159	5	400	200
U.S. Books	105	440	3	400	200
Football	115	613	12	500	400
Jazz	798	2742	4	400	400

5. Results

To assess the efficacy of our algorithm's performance, we juxtaposed its outcomes with those from four established algorithms for enhancing modularity: the GN [8], which is a greedy heuristic, CNM [61] which is an improved heuristic algorithm, genetic algorithm using a variation of information (GATHB) [16], which is a GA, and the Meme-Net [17], which is a memetic algorithm.

Table 5 represents the comparative analysis between the maximum modularity of all five different social network datasets with the maximum modularity (Q_{max}) and the average modularity (Q_{avg}) obtained by 50 runs of our method. The best results for each community detection method are shown in boldface. We can imply that our method is superior to the modularity scores achieved by GN, CNM, GATHB, and Meme-Net, although we have slightly inferior results compared to Meme-Net for two datasets.

Table 5. Comparison of modularity scores with various heuristic algorithms.

Network	GN	CNM	GATHB	Meme-Net	Our Method	
					Q_{max}	Q_{avg}
Karate	0.401	0.381	0.402	0.402	0.404	0.404
Dolphin	0.519	0.515	0.522	0.518	0.529	0.518
U.S. Books	0.510	0.502	0.518	0.523	0.522	0.510
Football	0.599	0.565	0.551	0.604	0.603	0.591
Jazz	0.439	0.439	0.445	0.438	0.445	0.444

Our method consistently demonstrates superiority by achieving the highest modularity scores in each dataset, outperforming the GN, CNM, and GATHB methods, thereby asserting its advanced capability in identifying and delineating the community structures within the networks. This is evidenced by the maximum modularity score (Q_{max}) values, which not only equal but in some cases surpass the best results obtained by the existing algorithms, underscoring the precision and efficacy of our algorithm.

The average modularity scores (Q_{max}) obtained from our method also reflect its robust

performance, maintaining an edge over the GN, CNM, and GATHB scores across almost all networks. Particularly, the Dolphin, U.S. Books, and Jazz networks exhibit higher values Q_{avg} when our method is applied, indicating an enhanced reliability in the community detection process. While our method's Q_{avg} for the Football network slightly trails behind that of the Meme-Net, it nonetheless remains competitive, illustrating the adaptability and resilience of our algorithm even in complex network scenarios.

To summarize, the empirical findings are drawn from Table 5, solidifying the standing of our algorithm as a formidable contender in the field of community detection. With its methodological enhancements, it not only stands up to the challenge against established algorithms but also often surpasses them, achieving high-quality community structure detection in diverse networks. These results fortify the assertion that our method is not only a significant contribution to the existing body of research but also a reliable tool for practical applications in community detection.

Figure 3(a)–3(e) showcases the facets of community detection using MAs in the analysis of five different datasets.

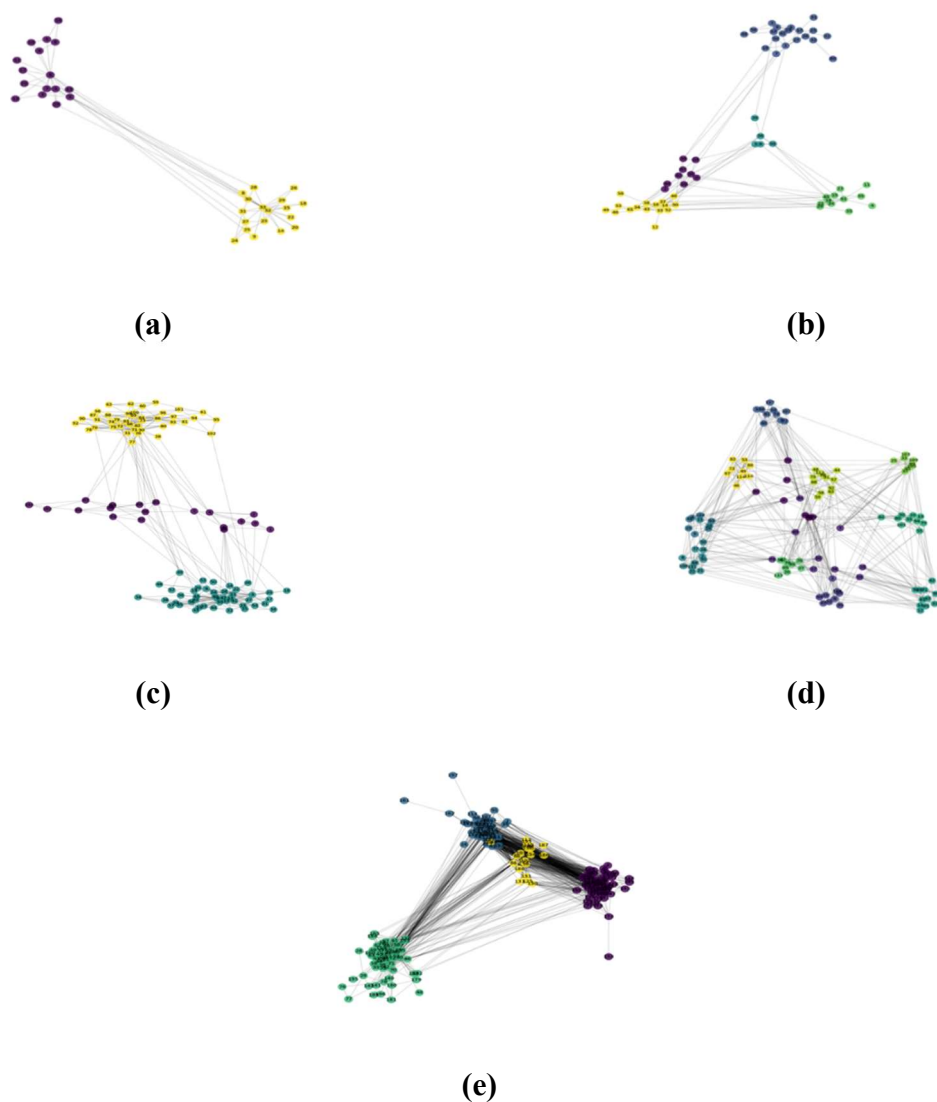


Figure 3. Detected communities after MAs. (a) Karate; (b) Dolphins; (c) Books; (d) Footballs; (e) Jazz.

The figure presents a visual representation of the detected communities within each social network dataset described above. Each node represents a member of the dataset, and the edges indicate the relationships between them. The color coding denotes the algorithm's division into communities, where we can see clear segregation into primary groups based on the number of their actual communities in real life.

6. Discussion

6.1. Effectiveness of local search

Table 6 provides an insightful comparative study on the effectiveness of the local search for community detection across five different social networks where w/ L.S. represents using local search and w/o L.S. represents the absence of local search. The results are segmented into two sets: one that includes local search and another without it. The best results for community detection performance are shown in boldface.

Table 6. Comparison of the presence and absence of local search.

Method	Karate		Dolphin		Books		Football		Jazz	
	Q_{max}	Q_{avg}	Q_{max}	Q_{avg}	Q_{max}	Q_{avg}	Q_{max}	Q_{avg}	Q_{max}	Q_{avg}
w/ L.S.	0.4036	0.4036	0.5285	0.5186	0.5221	0.5085	0.6027	0.5911	0.4451	0.4448
w/o L.S.	0.4036	0.4036	0.5277	0.5131	0.5221	0.5043	0.5946	0.5789	0.4451	0.4442
p -value	1.00		4.9×10^{-1}		6.15×10^{-2}		3.8×10^{-2}		8.1×10^{-2}	

Table 6 delineates both the highest modularity scores (Q_{max}) with average modularity scores (Q_{avg}) for the Karate, Dolphin, Books, Football, and Jazz networks. It reveals that the incorporation of local search does not necessarily enhance the best modularity scores but does offer notable improvements in average modularity scores across most networks. This suggests that the optimization consistency and the algorithm's ability to reliably identify community structures are strengthened when local search techniques are applied.

When we applied the local search, there was a discernible consistency in the Q_{max} across the Karate and Dolphin networks, with values of 0.4036 and 0.5285. When we do not apply the local search, there is a slight decrement in the Dolphin network's Q_{max} to 0.5277, while the Karate network remains unchanged. This pattern suggests that for certain network structures, such as that of the Karate network, which has a comparatively small size, local search may not significantly impact the peak modularity score.

Looking at the Q_{avg} , the enhancement due to local search is more apparent. For instance, in the Dolphin network, the Q_{avg} sees an increase from 0.5131 to 0.5186 with the inclusion of local search. This improvement emphasizes the role of local search in providing a more consistent detection of community structures.

In the Books and Football networks, a similar trend is observed where the Q_{avg} increases by approximately 0.042 and 0.122, respectively, when local search is applied. However, in the Jazz network, the Q_{avg} is only slightly increased by 0.006 with the implementation of local search. This

could suggest that while the approach generally enhances Q_{avg} , suggesting more consistent community detection, the Jazz network's unique complexity may require a custom-tailored local search strategy.

To confirm the significance of the performance improvements, we performed a Mann-Whitney U test between the modularity of GA with and without local search, with the p-value included at the bottom of Table 6. For the Dolphins and Football networks, using local search led to a significant performance improvement ($p < 0.05$). However, for the remaining networks, there was no significant performance enhancement ($p > 0.05$). In networks where a statistically significant performance enhancement was observed, the GA with local search outperformed the one without it in terms of Q_{max} . Conversely, in cases without significant performance enhancement, the GA with local search and the one without yielded equal results.

In summary, the quantitative analysis of Table 6 indicates that the application of local search in a GA framework tends to bolster the average performance in detecting community structures, though its impact on the Q_{max} may vary depending on the specific characteristics of the network under investigation. The consistency and reliability of community detection appear to be enhanced by the use of local search, especially in networks with less complex structures. The improved Q_{avg} underscores the enhanced robustness of the GA due to the local search, suggesting it is an effective strategy for achieving more reliable and stable solutions in the realm of community detection.

6.2. Convergence analysis

Figure 4 presents a figurative comparative analysis of the convergence behavior of GA with and without the implementation of local search across five different datasets: Karate, Dolphin, Books, Football, and Jazz. Each dataset is represented by a pair of plots in a two-row configuration, with the top row showing the convergence of GA with the implementation of local search and the bottom row depicting GA convergence without implementing local search. The red line indicates the point at which the best performance was achieved.

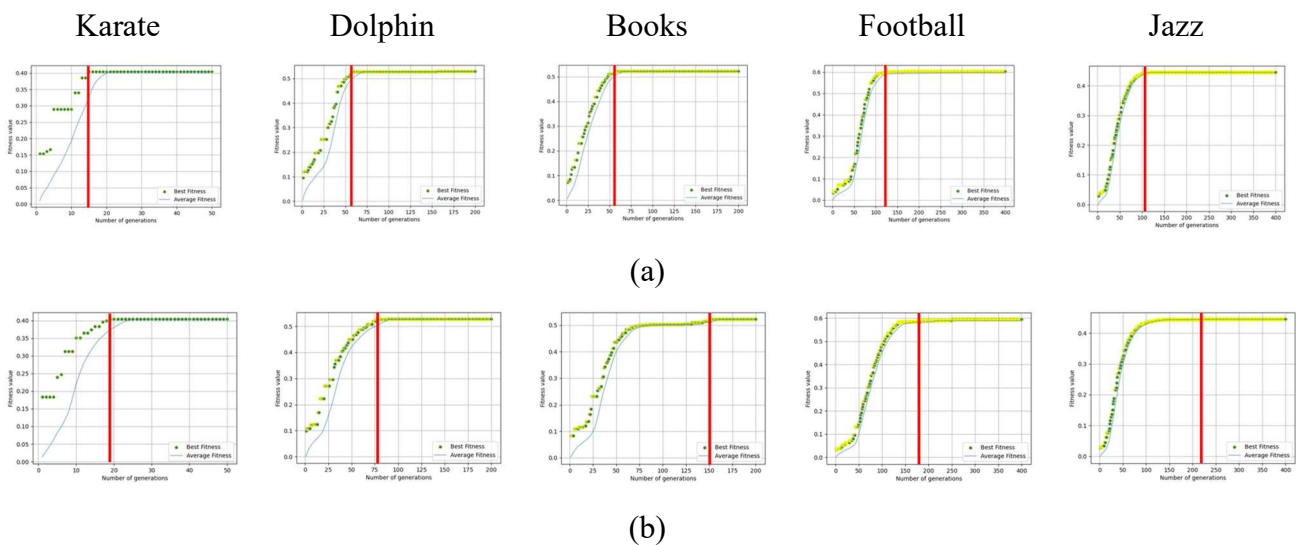


Figure 4. Convergence analysis of local search utilization across five datasets. (a) with local search; (b) without local search.

From the convergence curves, it is evident that the incorporation of local search generally results in a more rapid approach toward higher modularity values, indicative of more optimal community structures. This is characterized by a steeper ascent in the early generations for the GA with local search, suggesting an accelerated discovery of better solutions. In contrast, the GA without local search demonstrates a more gradual increase in modularity, reflecting slower progress toward optimizing community partitions.

Specifically, in the Karate, Dolphin, and Books datasets, the initial slope of the convergence curve with local search is markedly steeper compared to without, illustrating that local search significantly quickens the initial phase of optimization. For the Football and Jazz datasets, which are larger and presumably more complex, the acceleration in convergence is still visible with local search, although the difference in the rate of convergence compared to the GA without local search is less pronounced than in the smaller datasets.

The graphs also reveal that GA with local search reaches a plateau sooner, indicating an earlier stabilization at high modularity values. This suggests that local search not only speeds up convergence but potentially also contributes to achieving more robust solutions.

Overall, the comparative visualization underscores the effectiveness of local search in enhancing the efficiency of GAs in community detection tasks, particularly in terms of convergence speed to optimal or near-optimal solutions.

6.3. Novelty of our study

The novelty of our study lies in its introduction of a MA combining GA with the SHC algorithm to enhance modularity scores in community detection. Traditional community detection algorithms, although effective to some extent, often struggle with complex systems. While current GA-based methods are hindered by issues such as prolonged durations to reach the global optimum and susceptibility to premature convergence, our approach addresses these limitations by integrating SHC as a local optimization method, which significantly enhances both convergence speed and modularity scores. This integration demonstrates superior performance across multiple social network datasets compared to existing GA-based methods, providing a robust solution for detecting community structures in various complex networks.

While community structure is a key property of complex networks and indicates the potential functionality of networked systems, attacks and errors are common in everyday life, making network robustness crucial. In reality, networked systems frequently encounter unpredictable failures or intentional attacks. Therefore, maintaining the functionality of these networks under such disturbances is crucial [62]. In [63], the simulated annealing algorithm is applied to alleviate damage to the robustness of communities under attacks and errors. Recent studies have increasingly focused on the robustness of communities, which should maintain much of their original structure even under structural fluctuations [64]. We have devised a method that remains robust to changes in network structure, achieved through a GA using SHC as a local search. By integrating strategies that account for robustness against attacks and errors, our study offers a comprehensive approach to community detection that maintains functional integrity even in adverse conditions.

6.4. Limitation

While Newman's modularity [51] is effective and is the one of the most accepted measures for the evaluation of community structure in networks, modularity-based methods might introduce biases by favoring specific network structures and often merging smaller communities into larger ones although smaller communities are more appropriate. For example, they often perform well on synthetic networks designed with clear community structures but struggle with real-world networks where community structures overlap each other [65]. To provide a more comprehensive view of community structures, future research should explore alternative metrics such as NMI [66], the Omega index [67], which is the adaptation of the Rand index and is originally designed for partition problems that evaluates the proportion of communities in which pairs of nodes are correctly placed together, and the F1-score, which is frequently used to evaluate the communities identified by the methods that can be understood as a weighted average of precision and recall [65].

Moreover, the performance of the MA in networks with overlapping or hierarchical community structures should be further investigated. To date, numerous methods for detecting overlapping communities have been proposed, with the majority being node-based algorithms [68]. However, the combination of exploring the solution space through global search and exploiting promising regions within that space makes MAs highly effective tools for addressing challenging computational problems, such as detecting overlapping communities [2]. In real networks, communities often exhibit overlapping and hierarchical structures [69]. Thus, further studies should also focus on hierarchical detection methods that can be integrated with MAs. Approaches that evaluate multiple hierarchical levels and apply constraints on community capacity and hierarchical relationships can improve the detection quality. Therefore, research that focuses on adapting the algorithm to handle these complex structures more effectively and exploring the robustness of the algorithm under varying network conditions may be prospective.

7. Conclusions

In this study, we have attempted an approach of GAs implemented with the SHC local search method to increase the modularity scores based on community detection. By providing a comparative analysis with existing GAs such as GN [8], CNM [61], GATHB [16], and Meme-Net [17], our algorithm has consistently outperformed traditional GAs and shown that even without guaranteeing the global optimum, the local search mechanism significantly enhances the average performance metrics, indicating a stable and reliable detection of community structures. Particularly noteworthy is the evidence suggesting that local search prevents the algorithm from becoming ensnared in local optima, thereby facilitating a more effective exploration of the solution space.

The comparative analysis across five datasets: Karate, Dolphin, Books, Football, and Jazz reveals that our memetic algorithm is robust and versatile, adapting effectively to the complexities of each network. This study has rigorously explored the role of SHC as a local search strategy within the framework of a MA for community detection across a diverse set of social network datasets. The experimental results demonstrate a marked improvement in the convergence speed and modularity scores when local search is employed alongside GAs.

While in our study, we explored the effectiveness of using a MA for community detection in small and medium-sized datasets, Jin et al. [70] proposed using the integration of graph convolutional

networks (GCN) and Markov random fields (MRF) to solve the problem of community detection using a large scale dataset. For instance, the Never-Ending Language Learning (NELL) dataset has 65,755 nodes compared to our largest Jazz dataset having only 798 nodes. Shi et al. [71] proposed a community-based variational autoencoder (ComVAE) model in which both community information and deep learning techniques are utilized, where large scale datasets including 2000 to 4000 vertices were used. Therefore, we recognize the need for future research to combine local search strategies with deep learning methods. Integrating our MA with approaches like VAE and GCNs could leverage the strengths of both local optimization and deep learning. This hybrid approach has the potential to enhance detection accuracy and scalability for large community detection problems. Furthermore, exploring the application of our enhanced algorithm on more extensive and diverse datasets could provide deeper insights into its generalizability and robustness. Developing a comprehensive framework that incorporates both local optimization and deep learning could contribute significantly to the advancement of community detection techniques.

In conclusion, this study has affirmed that integrating local search into GA constitutes a significant advancement in the field of network analysis and community detection. It opens avenues for further research into hybrid meta-heuristic methods. It underscores the potential of such approaches in dealing with NP-hard problems that are prevalent in complex systems across various domains.

Author contributions

Dongwon Lee: Conceptualization, methodology, investigation, software, writing - original draft, validation; Jingeun Kim: Conceptualization, methodology, software, formal analysis, writing-review and editing; Yourim Yoon: Supervision, project administration, writing - review and editing, validation, funding acquisition. All authors have read and approved the final version of the manuscript for publication.

Use of AI tools declaration

The authors declare they have not used Artificial Intelligence (AI) tools in the creation of this article.

Acknowledgments

This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT) (No. 2022R1F1A1066017). This work was also supported by the Ministry of Education of the Republic of Korea and the National Research Foundation of Korea (NRF-2022S1A5C2A07090938).

Conflict of interest

The authors declare no conflicts of interest.

References

1. P. Bedi, C. Sharma, *Community detection in social networks*, Wiley interdisciplinary reviews: Data mining and knowledge discovery, **6** (2016), 115–135. <https://doi.org/10.1002/widm.1178>
2. L. M. Naeni, R. Berretta, P. Moscato, *MA-Net: A reliable memetic algorithm for community detection by modularity optimization*, In Proceedings of the 18th Asia Pacific Symposium on Intelligent and Evolutionary Systems, **1** (2015), Springer. https://doi.org/10.1007/978-3-319-13359-1_25
3. R. K. Behera, D. Naik, S. K. Rath, R. Dharavath, Genetic algorithm-based community detection in large-scale social networks, *Neural Comput. Appl.*, **32** (2020), 9649–9665. <https://doi.org/10.1007/s00521-019-04487-0>
4. E. Ferrara, A large-scale community structure analysis in Facebook, *EPJ Data Sci.*, **1** (2012), 1–30. <https://doi.org/10.1140/epjds9>
5. J. Goldenberg, B. Libai, E. Muller, Using complex systems analysis to advance marketing theory development: Modeling heterogeneity effects on new product growth through stochastic cellular automata, *Acad. Mark. Sci. Rev.*, **9** (2001), 1–18.
6. M. E. Newman, M. Girvan, Finding and evaluating community structure in networks, *Phys. Rev. E*, **69** (2004), 026113. <https://doi.org/10.1103/PhysRevE.69.026113>
7. A. Pothen, H. D. Simon, K. P. Liou, Partitioning sparse matrices with eigenvectors of graphs, *SIAM J. Matrix Anal. A*, **11**(1990), 430–452. <https://doi.org/10.1137/0611030>
8. M. Girvan, M. E. Newman, Community structure in social and biological networks, *P. Natl Acad. Sci.*, **99** (2002), 7821–7826. <https://doi.org/10.1073/pnas.122653799>
9. U. Brandes, D. Delling, M. Gaertler, R. Gorke, M. Hoefer, Z. Nikoloski, et al., On modularity clustering, *IEEE T. Knowl. Data En.*, **20** (2007), 172–188. <https://doi.org/10.1109/TKDE.2007.190689>
10. K. Wakita, T. Tsurumi, *Finding community structure in mega-scale social networks*, In Proceedings of the 16th international conference on World Wide Web, 2007. <https://doi.org/10.1145/1242572.1242805>
11. I. Koc, A fast community detection algorithm based on coot bird metaheuristic optimizer in social networks, *Eng. Appl. Artif. Intel.*, **114** (2022), 105202. <https://doi.org/10.1016/j.engappai.2022.105202>
12. Y. Zhang, Y. G. Liu, J. T. Li, J. J. Zhu, C. H. Yang, W. Yang, et al., WOCDA: A whale optimization based community detection algorithm, *Physica A*, **539** (2020), 122937. <https://doi.org/10.1016/j.physa.2019.122937>
13. C. Pizzuti, *Ga-net: A genetic algorithm for community detection in social networks*, In International conference on parallel problem solving from nature, Springer, 2008. https://doi.org/10.1007/978-3-540-87700-4_107
14. X. Zeng, W. Wang; C. Chen, G. G. Yen, A consensus community-based particle swarm optimization for dynamic community detection, *IEEE T. Cybernetics*, **50** (2019), 2502–2513. <https://doi.org/10.1109/TCYB.2019.2938895>
15. C. Honghao, F. Zuren, R. Zhigang, *Community detection using ant colony optimization*, In 2013 IEEE congress on evolutionary computation, 2013.
16. M. Tasgin, A. Herdagdelen, H. Bingol, Community detection in complex networks using genetic algorithms, *arXiv:0711.0491*, 2007.

17. M. Gong, B. Fu, L. C. Jiao, H. F. Du, Memetic algorithm for community detection in networks, *Phys. Rev. E*, **84** (2011), 056101. <https://doi.org/10.1103/PhysRevE.84.056101>
18. R. Shang, J. Bai, L. C. Jiao, C. Jin, Community detection based on modularity and an improved genetic algorithm, *Physica A*, **392** (2013), 1215–1231. <https://doi.org/10.1016/j.physa.2012.11.003>
19. Y. Liang, L. Wang, Applying genetic algorithm and ant colony optimization algorithm into marine investigation path planning model, *Soft Comput.*, **24** (2020), 8199–8210. <https://doi.org/10.1007/s00500-019-04414-4>
20. K. De Jong, *Genetic algorithms: A 10 year perspective*, In Proceedings of the first International Conference on Genetic Algorithms and their Applications, Psychology Press, 2014.
21. P. Preux, E. G. Talbi, Towards hybrid evolutionary algorithms, *Int. T. Oper. Res.*, **6** (1999), 557–570. <https://doi.org/10.1111/j.1475-3995.1999.tb00173.x>
22. T. A. El-Mihoub, A. A. Hopgood, N. Lars, B. Alan, Hybrid genetic algorithms: A review, *Eng. Lett.*, **13** (2006), 124–137.
23. M. E. Newman, Fast algorithm for detecting community structure in networks, *Phys. Rev. E*, **69** (2004), 066133. <https://doi.org/10.1103/PhysRevE.69.066133>
24. J. Holland, *Adaptation in natural and artificial systems*, Applying genetic algorithm to increase the efficiency of a data flow-based test data generation approach, the university of michigan press, Ann. Arbor. 1975, 1–5.
25. L. Haldurai, T. Madhubala, R. Rajalakshmi, A study on genetic algorithm and its applications, *Int. J. Comput. Sci. Eng.*, **4** (2016), 139.
26. W. E. Hart, N. Krasnogor, J. E. Smith, *Memetic evolutionary algorithms*, In Recent Advances in Memetic Algorithms, Springer, 2005, 3–27. https://doi.org/10.1007/3-540-32363-5_1
27. P. Moscato, C. Cotta, *A gentle introduction to memetic algorithms*, In Handbook of metaheuristics, Springer, 2003, 105–144. https://doi.org/10.1007/0-306-48056-5_5
28. N. Krasnogor, J. Smith, A tutorial for competent memetic algorithms: Model, taxonomy, and design issues, *IEEE T. Evolut. Comput.*, **9** (2005), 474–488. <https://doi.org/10.1109/TEVC.2005.850260>
29. G. Acampora, V. Loia, S. Salerno, A. Vitiello, A hybrid evolutionary approach for solving the ontology alignment problem, *Int. J. Intel. Syst.*, **27** (2012), 189–216. <https://doi.org/10.1002/int.20517>
30. R. Cabido, A. S. Montemayor, J. J. Pantrigo, High performance memetic algorithm particle filter for multiple object tracking on modern GPUs, *Soft Comput.*, **16** (2012), 217–230. <https://doi.org/10.1007/s00500-011-0715-2>
31. Y. Li, J. Liu, C. Liu, A comparative analysis of evolutionary and memetic algorithms for community detection from signed social networks, *Soft Comput.*, **18** (2014), 329–348. <https://doi.org/10.1007/s00500-013-1060-4>
32. B. Yang, W. Cheung, J. Liu, Community mining from signed social networks, *IEEE T. Knowl. Data En.*, **19** (2007), 1333–1348. <https://doi.org/10.1109/TKDE.2007.1061>
33. P. Doreian, A multiple indicator approach to blockmodeling signed networks, *Soc. Networks*, **30** (2008), 247–258. <https://doi.org/10.1016/j.socnet.2008.03.005>
34. V. A. Traag, J. Bruggeman, Community detection in networks with positive and negative links, *Phys. Rev. E*, **80** (2009), 036115. <https://doi.org/10.1103/PhysRevE.80.036115>
35. L. Wu, X. Ying, X. Wu, A. Lu, Z. H. Zhou, *Spectral analysis of k-balanced signed graphs*, In Advances in Knowledge Discovery and Data Mining: 15th Pacific-Asia Conference, PAKDD 2011, Shenzhen, China, May 24–27, 2011, Proceedings, Part II 15. 2011. Springer.

36. S. Ranjkesh, B. Masoumi, S. M. Hashemi, A novel robust memetic algorithm for dynamic community structures detection in complex networks, *World Wide Web*, **27** (2024), 3. <https://doi.org/10.1007/s11280-024-01238-7>
37. M. Miao, J. R. Wu, F. J. Cai, Y. G. Wang, A modified memetic algorithm with an application to gene selection in a sheep body weight study, *Animals*, **12** (2022), 201. <https://doi.org/10.3390/ani12020201>
38. J. Andre, P. Siarry, T. Dognon, An improvement of the standard genetic algorithm fighting premature convergence in continuous optimization, *Adv. Eng. Softw.*, **32** (2001), 49–60. [https://doi.org/10.1016/S0965-9978\(00\)00070-3](https://doi.org/10.1016/S0965-9978(00)00070-3)
39. Y. D. Kwon, S. B. Kwon, S. B. Jin, J. Y. Kim, Convergence enhanced genetic algorithm with successive zooming method for solving continuous optimization problems, *Comput. Struct.*, **81** (2003), 1715–1725. [https://doi.org/10.1016/S0045-7949\(03\)00183-4](https://doi.org/10.1016/S0045-7949(03)00183-4)
40. T. F. Gonzalez, *Handbook of approximation algorithms and metaheuristics*, 2007: Chapman and Hall/CRC.
41. C. H. Papadimitriou, K. Steiglitz, *Combinatorial optimization: Algorithms and complexity*, Courier Corporation, 1998.
42. E. Aarts, J. H. Korst, P. J. Laarhoven, *Simulated annealing*, E. Aarts, JK Lenstra, eds., Local Search in Combinatorial Optimization, John Wiley and Sons, New York, NY, **91120** (1997).
43. S. J. Russell, P. Norvig, *Artificial intelligence: A modern approach*, Pearson, 2016.
44. B. Mondal, K. Dasgupta, P. Dutta, Load balancing in cloud computing using stochastic hill climbing-a soft computing approach, *Procedia Technol.*, **4** (2012), 783–789. <https://doi.org/10.1016/j.protcy.2012.05.128>
45. B. L. Miller, D. E. Goldberg, Genetic algorithms, tournament selection, and the effects of noise, *Complex Syst.*, **9** (1995), 193–212.
46. R. Halalai, C. Lemnaru, R. Potolea, *Distributed community detection in social networks with genetic algorithms*, In Proceedings of the 2010 IEEE 6th International Conference on Intelligent Computer Communication and Processing, 2010. <https://doi.org/10.1109/ICCP.2010.5606467>
47. D. E. Goldberg, *Genetic algorithms in search, optimization and machine learning*, Addison-Wesley Longman Publishing Co., Inc. 1989.
48. W. W. Zachary, An information flow model for conflict and fission in small groups, *J. Anthropol. Res.*, **33** (1977), 452–473. <https://doi.org/10.1086/jar.33.4.3629752>
49. J. Q. Jiang, L. J. McQuay, Modularity functions maximization with nonnegative relaxation facilitates community detection in networks, *Physica A*, **391** (2012), 854–865. <https://doi.org/10.1016/j.physa.2011.08.043>
50. D. Lusseau, K. Schneider, O. J. Boisseau, P. Haase, E. Sloaten, S. M. Dawson, The bottlenose dolphin community of doubtful sound features a large proportion of long-lasting associations: Can geographic isolation explain this unique trait? *Behav. Ecol. Sociobiol.*, **54** (2003), 396–405. <https://doi.org/10.1007/s00265-003-0651-y>
51. M. E. Newman, Modularity and community structure in networks, *P. Natl Acad. Sci.*, **103** (2006), 8577–8582. <https://doi.org/10.1073/pnas.0601602103>
52. The red hot Jazz archive. Available from: <http://www.redhotjazz.com>.
53. C. Pizzuti, *A multi-objective genetic algorithm for community detection in networks*, In 2009 21st IEEE International Conference on Tools with Artificial Intelligence, 2009. <https://doi.org/10.1109/ICTAI.2009.58>

54. M. Guerrero, F. G. Montoya, R. Baños, A. Alcayde, C. Gil, Adaptive community detection in complex networks using genetic algorithms, *Neurocomputing*, **266** (2017), 101–113. <https://doi.org/10.1016/j.neucom.2017.05.029>
55. C. Shi, W. Yi, B. Wu, C. Zhong, *A new genetic algorithm for community detection*, In Complex Sciences: First International Conference, Complex 2009, Shanghai, China, February 23-25, 2009, Revised Papers, Part 2, Springer, 2009. https://doi.org/10.1007/978-3-642-02469-6_11
56. R. Zheng, *A fast community detection algorithm based on clustering coefficient*, In 3rd International Conference on Mechatronics Engineering and Information Technology (ICMEIT 2019), Atlantis Press, 2019. <https://doi.org/10.2991/icmeit-19.2019.100>
57. D. Choudhury, S. Bhattacharjee, A. Das, *An empirical study of community and sub-community detection in social networks applying Newman-Girvan algorithm*, In 2013 1st international conference on emerging trends and applications in computer science, IEEE, 2013. <https://doi.org/10.1109/ICETACS.2013.6691399>
58. N. Du, B. Wu, X. Pei, *Community detection in large-scale social networks*, In Proceedings of the 9th WebKDD and 1st SNA-KDD 2007 workshop on Web mining and social network analysis, 2007.
59. A. Said, R. A. Abbasi, O. Maqbool, A. Daud, N. R. Aljohani, CC-GA: A clustering coefficient based genetic algorithm for detecting communities in social networks, *Appl. Soft Comput.*, **63** (2018), 59–70. <https://doi.org/10.1016/j.asoc.2017.11.014>
60. W. Y. Lin, W. Y. Lee, T. P. Hong, Adapting crossover and mutation rates in genetic algorithms, *J. Inf. Sci. Eng.*, **19** (2003), 889–903.
61. A. Clauset, M. E. Newman, C. Moore, Finding community structure in very large networks, *Phys. Rev. E*, **70** (2004), 066111. <https://doi.org/10.1103/PhysRevE.70.066111>
62. S. Wang, J. Liu, Constructing robust community structure against edge-based attacks, *IEEE Syst. J.*, **13** (2018), 582–592. <https://doi.org/10.1109/JSYST.2018.2835642>
63. S. Wang, J. Liu, X. Wang, Mitigation of attacks and errors on community structure in complex networks, *J. Stat. Mech.-Theory E.*, **2017** (2017), 043405. <https://doi.org/10.1088/1742-5468/aa6581>
64. S. Wang, J. Liu, Community robustness and its enhancement in interdependent networks, *Appl. Soft Comput.*, **77** (2019), 665–677. <https://doi.org/10.1016/j.asoc.2019.01.045>
65. V. D. F. Vieira, C. R. Xavier, A. G. Evsukoff, A comparative study of overlapping community detection methods from the perspective of the structural properties, *Appl. Netw. Sci.*, **5** (2020), 1–42. <https://doi.org/10.1007/s41109-020-00289-9>
66. L. Danon, A. Díaz-Guilera, J. Duch, A. Arenas, Comparing community structure identification, *J. Stat. Mech.-Theory E.*, **2005** (2005), P09008. <https://doi.org/10.1088/1742-5468/2005/09/P09008>
67. L. M. Collins, C. W. Dent, Omega: A general formulation of the rand index of cluster recovery suitable for non-disjoint solutions, *Multivar. Behav. Res.*, **23** (1988), 231–242. https://doi.org/10.1207/s15327906mbr2302_6
68. M. Li, J. Liu, A link clustering based memetic algorithm for overlapping community detection, *Physica A*, **503** (2018), 410–423. <https://doi.org/10.1016/j.physa.2018.02.133>
69. H. Shen, X. Q. Cheng, K. Cai, M. B. Hu, Detect overlapping and hierarchical community structure in networks, *Physica A*, **388** (2009), 1706–1712. <https://doi.org/10.1016/j.physa.2008.12.021>
70. D. Jin, Z. Y. Liu, W. H. Li, D. X. He, W. X. Zhang, *Graph convolutional networks meet markov random fields: Semi-supervised community detection in attribute networks*, In Proceedings of the AAAI conference on artificial intelligence, 2019. <https://doi.org/10.1609/aaai.v33i01.3301152>

-
71. W. Shi, Network embedding via community based variational autoencoder, *IEEE Access*, **7** (2019), 25323–25333. <https://doi.org/10.1109/ACCESS.2019.2900662>



AIMS Press

© 2024 the Author(s), licensee AIMS Press. This is an open access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0>)