Modified Memetic Algorithm to Solve Partitional Clustering Problem in Complex Network

Feroz Ahmed[1], Sumit Kumar[1,*], and Pradeep Kumar Singh[2]

ARTICLE INFO

Article history:
Received: 3 August 2023
Revised: 20 October 2023
Accepted: 11 December 2023
Online: 20 June 2024

Keywords:
Partitional clustering
Memetic algorithm
Modified memetic algorithm
Recombination operator
Complex network

ABSTRACT

Community detection problem in complex network can be consider as optimization problem of clustering. Clustering is a popular unsupervised learning approach aiming to group similar data objects forming various clusters. The partitional clustering in complex network is a non-deterministic polynomial-time hardness (NP-hard) problem. The metaheuristic algorithms are generally utilized in solving such kind of optimization problems. Many evolutionary algorithms are utilized to solve community detection problem in complex network. In evolutionary algorithms, the most optimal solutions are selected for the recombination phase to generate the new solution. In this paper, a Modified Memetic Algorithm (MMA) is proposed to obtain an optimal set of clusters via partitioned clustering in complex networks. Furthermore, Partition Recombination Crossover Operator (PXRO) is being utilized to solve the problem of partitional clustering. The performance of the proposed algorithm is evaluated on real word problems and results shows that the proposed algorithm is more efficient as compare of other start of the art methods.

1. INTRODUCTION

An unsupervised learning method called as clustering is type of an unsupervised learning method that organizes related data elements into groups. It includes a variety of techniques, such as partitional, hierarchical, grid-based, density-based, and model-based clustering [1]. In contrast, hierarchical clustering uses agglomerative and divisive methods and generates a tree structure. The major drawback of hierarchical clustering is that once the merging of clusters takes place, it cannot be undone. Density-based clustering is based on the density associated with a data object that can be decided by counting the number of data objects in a specified radius. The data objects having a density above a specified threshold are used for the formation of clusters. The density-based clustering technique faces problems in the case of high-dimensional data. The grid-based clustering involves the formation of a finite number of cells that form a grid-like structure by quantizing the data objects. The need for quantization and a large number of parameters are the limitations of grid-based clustering. Model-based clustering involves the concept of a probability distribution, where each data object has a probability of belonging to each cluster. Model-based clustering is not suitable for large databases.

Despite the drawback of each approach, partitional clustering is nevertheless widely used because of its versatility in areas including image processing[2], networking[3], pattern recognition[4], and healthcare[5], especially when dealing with huge datasets. Data is divided into discrete clusters using partitional clustering, which aims to achieve maximum inter-cluster separation and minimal intra-cluster distance for compactness [6]. Meta-heuristic algorithms are chosen for solving complex network problems due to the following reasons: First, complex networks have intricate structures with a vast solution space, making traditional methods impractical. Second, many of these problems are NP-hard, making meta-heuristics a suitable choice for approximating near-optimal solutions. Third, they focus on global optimization, critical for networks with deceptive local optima. Fourth, they are robust and adaptable to handle uncertainties and dynamic changes in network data. Fifth, their parallelization capability accelerates the search process. Sixth, they offer practical solutions balancing quality and computational cost across various domains. Many evolutionary algorithms like Genetic algorithm (GA), Particle Swan optimization (PSO), and Ant colony optimization(ACO) etc. are used to solve community detection problem in complex network. But GA is most widely used algorithm for the same. GA is having its own drawback like slow conversion rate. One local search algorithm is combined with GA to overcome this problem. The combined algorithm of GA and local search is called

[1]*Amity School of Engineering and Technology, Amity University, Noida, Uttar Pradesh.*
[2]*Department of Computer Science and Engineering, Central University of Jammu, J&K, India*.
**Corresponding author:* Sumit Kumar; E-mail: skumar59@amity.edu.

memetic algorithm (MA)[7]. This memetic algorithm is useful to accelerate to find the good solution which is hard to discover.

Detection of community structure in complex networks is a clustering problem. Therefore, it can be considered as a optimization problem. Complex networks are graph-based models that symbolizes the interaction of the underlying graphs in a real networked system [8]. The components of the complex network such as vertex means object of any real system while edge represents the interaction between two objects. Complex networks may have specific clusters of highly interconnected vertices that are loosely coupled with other clusters. In the context of graph theory, clustering can be defined as the partition of the vertices of the underlying graph into disjoint subsets where each subset represents a cluster. The compactness in clustering can be achieved by maximizing intra-cluster density and minimizing inter-cluster density i.e. the cluster must be a cohesive group of vertices that are connected more “densely” to each other than to the vertices in other clusters. Modularity is a measure used to determine the strength of clusters in a complex network in order to evaluate the performance of a given cluster [9].

The objective of work is to perform partitions of the vertices of the underlying graph into disjoint clusters having high modularity. The combinatorial optimization problems are largely solved by metaheuristic algorithms to ensure an optimum solution [10]. The partitional clustering algorithms like K-means, suffer from getting stuck into local optimum solutions. To address the issue, MMA algorithm is proposed in order to solve the partitional clustering in a complex network. Generally, traditional method struggles with population initialization, leading to suboptimal results. MMA performs well in complex network clustering, particularly when starting with unfitted initial populations. Furthermore, MMA leverages memetic algorithm and an exploratory approach, intelligently recombines clusters and improves solutions iteratively using local search. The local search capabilities of the MA are harnessed to fine-tune the non-fitted initial solutions, gradually improving their modularity and overall quality. This iterative refinement process contributes to the algorithm's ability to converge towards high-quality clusters, even when commencing with less-than-optimal initial solutions.

Modified memetic algorithm (MMA) introduces a population updating strategy that is not commonly found in traditional evolutionary algorithms. This strategy involves a distance threshold and the replacement of clusters based on modularity. The method also serves the purpose of maintaining diversity in the population, preventing premature convergence, and ensuring that the best individuals have a chance to survive and contribute to future generations. Thus, the method employs a multifaceted approach, combining memetic algorithm and the innovative PXRO recombination operator, to optimize clustering in complex networks. MMA addresses the partitional clustering challenges by maximizing cohesion within clusters and separation between them. Its unique PXRO operator incorporates domain knowledge to enhance cluster quality through modularity optimization. Furthermore, a population update strategy ensures diversity and prevents premature convergence, resulting in high-quality cluster partitions within complex networks. The proposed method is compared with the state-of-the-art algorithms like BGLL[11] and MA-COM [8] on two benchmark complex network datasets i.e. Zachary Karate Club and Dolphins network against standard objective functions like ERI, NMI and modularity. The parameters chosen for tuning the algorithm are mutation probability, number of generations and population size. The experimental result analysis and observation show that the proposed method is efficient and outperforms the existing state-of-the-art method.

The key contributions can be highlighted as follows: Firstly, MMA, a novel approach for solving complex network partitional clustering problems, leveraging memetic algorithms and the unique PXRO operator has been introduced. Secondly, the algorithm is shown to be highly effective, outperforming existing state-of-the-art methods in modularity, ERI, and NMI on real-world datasets. Thirdly, it exhibits robustness by starting with non-fitted initial populations, increasing adaptability. Additionally, the incorporation of domain knowledge through the PXRO operator enhances cluster quality, while a population update strategy maintains diversity and prevents premature convergence, resulting in a diverse set of solutions.

2. LITERATURE REVIEW

2.1 Partitional Clustering in Complex Network

The analysis of complex networks is useful in understanding the behavior of real-world complex systems in various fields like community detection in social media, fraud detection in banking, opinion mining, sentiment analysis and assessing learning outcomes in education. The inability of deterministic or exact algorithms in partitioning large complex networks leads to the usage of non-deterministic optimization algorithms in partitional clustering over complex networks. The large contribution of evolutionary algorithms (EAs) in partitional clustering over complex networks has proven the supremacy of EAs over other metaheuristic algorithms [12]. Gema Bello-Orgaz et. al [13] introduced a genetic algorithm based on edge encoding to identify overlapping clusters in a complex network. Manuel Guerrero et. al [14] presented a generational genetic algorithm based on an improved initialization technique taking modularity as an objective function.

Maoguo Gong et al. [15] introduced the memetic algorithm in modularity density function for solving clustering problems in a complex network. The memetic algorithm is proven successful in solving the non-deterministic polynomial-time complete problem. The work represents an improved memetic algorithm taking modularity density as an objective function for partitional clustering in complex networks based upon neighborhood mutation and improved simulated annealing combined local search procedure. Cai-Hong Mu et al.[16] introduced a modified memetic algorithm based upon two local search strategies i.e. simulated annealing and tightness greedy optimization (TGO) for solving partitional clustering in complex networks. Mingming Li. et al.[17] presented a link clustering-based memetic algorithm to detect overlapping clusters in complex networks utilizing modularity density. David Chalupa et al.[18] presented a hybrid bridge-based memetic algorithm based upon the hybrid method of the initial population and local optima sampling with a steady-state evolutionary process. This technique is prone to getting stuck in local optimum solution. Thus, the defects of partitional clustering cannot be adequately solved. Therefore, this work introduces an extended modified algorithm to overcome the limitations by proposing a new modified of memetic algorithm for partitional clustering in complex networks.

2.2 Memetic Algorithm

P Moscato et al. [19] proposed a memetic algorithm in 1989 based on population hybrid evolutionary algorithm paired with local refinement techniques. Memetic algorithms are highly effective in a variety of problem domains, including multi-objective optimization [20], non-stationary function optimization [21], and combinatorial optimization [22]. The memetic algorithm employs the mechanism of local search to support the mutability behaviour to improve the optimality of the solution [23]. Hart et al. [24] proposed two ways of implementing the local search mechanism i.e. (1.) FBS (Fitness-Based Strategy) and (2.) DBS (Distribution-Based Strategy). FBS method aims to prioritization of the local search mechanism in terms of population outcomes while the DBS objective is to eliminate local search based upon population data redundancy. Molina et al. [25]suggested the division of population size into three categories, each of which had different local search possibilities and strengths. Nguyen et al. [26] developed a stochastic memetic algorithm to control the strength of local search and calculated a notional constraint for its strength. Nobahari et al. [27] showed the intensity of local search and the percentage of global and local search variation. In this technique, a population with a higher proportion of outstanding results would be assigned a higher local search strength and a lower proportion of results will have lower local search strength. Noman et al. [28] established a local search mechanism that calculates the strength of the local search for an individual's solution improvement based on outcomes in the exploration phase. Liu et al. [29] proposed an accommodation strategy in which the strength of local search is modulated based on exploration for the global optimum solution. Ma et al. [30] proposed an accommodating technique to determine the GLS (Gaussian Local Search) [31] competency stage for different solutions in the population.

The memetic algorithm initiates the solution population, fine-tunes solutions, guides genetic operations, selects high-quality parents, maintains diversity, and adapts to problem characteristics. The memetic algorithm multifaceted role helps in optimizing complex network clustering through various stages. This enables to yield high-quality solutions for the challenging problem of partitional clustering within complex networks, establishing it as a robust and effective approach in the realm of unsupervised learning and network analysis. The variant of memetic algorithm has been introducted to tackle the challenges associated with partitional clustering in complex network. By combining global search capabilities with local optimization and integrating domain knowledge, the algorithm offers an effective and efficient approach to obtain high-quality clusters in complex network data. Thus, this paper aims to provide a valuable finding for researchers and practitioners in various domains, including image processing, networking, pattern recognition, and healthcare, where complex network analysis plays a crucial role.

3. PROPOSED METHODOLOGY

The proposed methodology involves the utilization of a hybrid operator i.e. Partitional Crossover Recombination Operater (PXRO) for the formation of a new complete partition in a complex network by combining two partitions of the same network. The clusters (genes) related to both of the complex networks are then integrated into a single list for traversal. Further, each node in the list categorizes into two clusters formation, one from each of the complex networks. The obtained cluster arranged in a greedy modified of the algorithm so that the cluster with the highest cardinality is chosen at each iteration.

The assessment of the clustering quality for the clusters so obtained is achieved through the modularity function (Q). The algorithm addresses the challenges in partitional clustering problems within complex network. It employs various performance metrics like modularity, ERI, and NMI to quantitatively evaluate the algorithm efficiency. Through rigorous benchmarking against established algorithms, MMA provides competitive solutions on real-world datasets. The paper also explores parameter tuning, examining how different settings affect the algorithm's performance and solution quality. Further, real-world datasets such as the Zachary Karate Club and Dolphins network enhance the algorithm's credibility, while visual representations aid in conveying solution quality. The demonstrated generalizability and likely sensitivity analysis support the algorithm's practical applicability in addressing complex partitioned clustering in networked system. The steps of the proposed methodology are discussed below in detail and are presented in algorithm 1. The flow process of new modified memetic algorithm i.e. MMA algorithm is demonstrated in Figure.1.

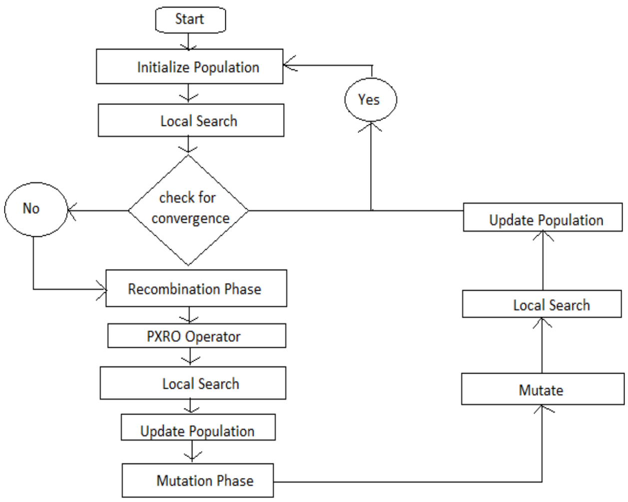

Fig.1 Flow Process of proposed methodology.

A weighted graph is a combination of edges and vertices i.e. $G = (E, V, w)$ where w is a weighting function and $w: V \times V \rightarrow \mathbb{R}$ such that for all $\{u, v\} \in E$, $w(\{u, v\}) \neq 0$, and for all $\{u, v\} \notin E, w(\{u, v\}) = 0$. Let $C \subseteq V$ and $C' \subseteq V$ be two vertex subsets, $W(C, C')$ the weight sum of the edges linking C and C' , i.e., $W(C, C') = \sum_{u \in C, v \in C'} w(\{u, v\})$. The modularity (Q) in partitional clustering with K communities $I = \{C_1, C_2, \dots, C_k\}(\forall i \in \{1, 2, \dots, K\}, C_i \subset V$ and $C_i \neq \emptyset$; $\cup_{i=1}^{K} C_i \subset V$; $\forall i$, $j \in \{1, 2, \dots, K\}$, $C_i \cap C_j = \emptyset)$ is given by Eq. (1) as follows:

$$Q(I) = \sum_{i=1}^{K} \left[\frac{W(C_i, C_i)}{W(V, V)} - \left(\frac{d_i}{W(V, V)} \right)^2 \right] \quad (1)$$

where, d_i is the sum of the degrees of the vertices of the cluster C_i , i.e. $d_i = \sum_{v \in C_i} \deg(v)$ with $\deg(v)$ being the degree of vertex v. Population diversity is a crucial feature in the memetic algorithm to escape premature convergence [19]. To handle this issue, the population updating method is considered. A distance function is needed in order to update the population size. Let $X = \{X_1, X_2, \dots, X_K\}$ and $Y = \{Y_1, Y_2, \dots, Y_{K'}\}$ be two groups of graph $G = (E, V)$. For an edge $e = \{u, v\} \in E$ and a cluster C of X or Y , we use $e \in C$ to justify that the vertices u and v of e are in the same cluster. The rand index [32] is used as a distance function to estimate the distance 'd' (Edge Rand Index)between X and Y is shown in Eq. 2.

$$d(X, Y) = \frac{\sum_{e \in E} d_e(X, Y)}{m} \quad (2)$$

where, $d_e(X, Y)$ of edge $e = \{u, v\}$ is defined by Eq. 3.

$$d_e(X,Y) = \begin{cases} 0 \ if \ \exists X_i \in X, \ \exists Y_j \in Y \ \ s.t. \ \ e \in X_i \ and \ e \in Y_j \\ OR \ if \ \forall X_i \in X, \neg(e \in X_i) \ and \ \forall Y_i \in Y, \neg(e \in Y_j) \\ 1 \ otherwise \end{cases} \quad (3)$$

ERI is a ratio having a range between 0 and 1, whereas Q has a range of (-0.5,1). Given two partitions A and B of a complex network in clusters, let C be the confusion matrix whose element C_{ij} is the number of nodes of cluster i of partition A that are also in the cluster j of partition B. The normalized mutual information $NMI(A, B)$ is defined as shown in Eq. 4.

$$NMI(A, B) = \frac{-2 \sum_{i=1}^{C_A} \sum_{j=1}^{C_B} C_{ij} \cdot \log(C_{ij} N / C_i \cdot C_j)}{\sum_{i=1}^{C_A} C_i \cdot \log(C_i / N) + \sum_{j=1}^{C_B} C_j \cdot \log(C_j / N)} \quad (4)$$

where, C_A (C_B) is the number of groups in the partition A (B), $C_{i.}(C_j)$ is the sum of elements of C in row i (column j), and N is the number of nodes. If $A = B$, then $NMI(A, B) = 1$; if A and B are completely different, then $NMI(A, B) = 0$. The steps of the proposed algorithm are explained as follows:

3.1 Population Initialization

MMA algorithm doesn't require a precisely fitted initial population. The approach offers several key advantages. It reduces sensitivity to initial conditions, making it more adaptable to various scenarios. It avoids getting trapped in suboptimal solutions, thus increasing the likelihood of finding the best global solution. The method enhances exploration in the solution space, which is particularly useful for complex networks. Furthermore, its ability to work with diverse initial populations makes it practical for real-world scenarios where data is noisy and complex, ultimately improving robustness and ensuring consistent, reliable results.

A randomized multilevel algorithm [1] has been used to generate the initial population. The algorithm makes use of the optimum executor [4] for the refinement procedure. Each optimum executor moves a vertex from its current cluster to another cluster if the move ensures an increase in modularity. The optimum executor is applied to a graph G_0 to optimize the modularity of a cluster 'C'of that graph until no further optimization is possible for the cluster. Further, the transformation of G_0 take place into G_1 where each vertex is a member of the cluster C and an edge links two vertices in G_1 if they represent two neighboring members in C. Now the optimum executor is being applied to the new graph G_1in order to obtain another cluster. Then, again the transformation of graph G_1 takes place into G_2. Finally, when the optimization of graph G_2 is not possible then the unfolding of graphs takes place starting from the highest level G_2 until it reaches the lowest level G_0.

3.2 Selection of Parents

In this work, the tournament selection algorithm has been used for the selection of parent individuals in performing the genetic operation. The ability to hold low-fitness individuals in the offspring chromosome set makes it suitable and efficient for both nonparallel and parallel architectures. This capability makes chromosomes in the current population equally probable to function as the parent chromosomes of resulting genetic operations. All of these factors have contributed to the usage of tournament selection as a selection mechanism for the proposed work.

3.3 PXRO Operator for Recombination

The PXRO operator is proposed for the recombination mechanism in the memetic algorithm. The operator ensures that all common edges from the two parents are inherited and offspring are formulated using the edges inherited from two parents [30]. The PXRO operator is being utilized for the recombination in order to generate new offsprings or new solutions. The PXRO operator is to incorporate domain knowledge criteria regarding network structure in the recombination operator. The proposed PXRO operator is shown in Eq.5.

$$R_{\epsilon\alpha\beta} \square \sum_{\ddot{u}} s_t(\ddot{u}) D_{\epsilon\alpha\beta}(\ddot{u}) \qquad (5)$$

Here, α and β refers to two parent chromosomes that serve as an input for the offspring ϵ . The offspring is obtained through recombination of α and β followed by the selection of the best solution ü. The varieties of solutions is represented by $D_{\epsilon\alpha\beta}(\ddot{u})$ along with their probabilities s_t.

The objective is to form groups of complex networks (referred as communities) as genetic material (chromosomes) and try to retain some communities from the parents (clusters). Let (C1, C2) be two parent clusters and *'b'* a priority vector. Let *'u'* and *'v'* be respectively the count of communities of clusters C1 and C2. The vector *'b'*, indexed from 1 to *'u+v'*, is presented by a random permutation of {1, 2, ..., u + v}. The indices between 1 and *'u'* of *'b'* denote the communities of one parent and those between *'u'* + 1 and *'u'* + *'v'* the communities of the other parent. Thus every community of the parents is assigned a distinct number from 1 to *'u'* + *'v'*. For every community G_i, i∈{1, 2, ..., u + v}, the corresponding value in *'b'* (i.e., *Q[i]*) gives the priority of G_i. The crossover procedure generates from (C1, C2) its offspring cluster C0 as follows. While traversing all of the communities in the priority order specified by the vector 'b.'. The selection is being done on the basis of the highest priority community *'G'* according to *'b'* followed by a transfer of all the vertices of the community to form a community of the offspring. The community G' is being selected with the second highest priority followed by the removal of the vertices already in C0 and use the remaining vertices of G' to form a new community of C0 (empty community is discarded). The process is being repaeated until the community with the lowest priority is handled. Finally, the communities of C0 are renamed from one to the number of communities in the offspring.

Algorithm 1. Modified Memetic Algorithm (MMA)

Step 1.	(Initialization) Initialize the population with new solutions.
Step 2.	***for*** each solution *j* in the population, **do**
	j.initialize()
	j. local-search()
	end for
Step 3.	(Recombination Phase) ***for*** each recombination operation k from 1 to the number of recombination, ***do***
	Parents:= *Population*.select(*numparents*)
	Parents ⊆ *Population*
	l := *Parents*.PXRO()
	l.local-search()
	Population.update(l)
	end for
Step 4.	(Mutation Phase) ***for*** each mutation operation k from 1 to the number of mutations, ***do***
	j:= *Population*.select(1);
	j_m:= *j*.mutate()
	j_m.local-search()
	Population.update(j_m)
	end for
Step 5.	(Convergence Check) ***if*** *Population*.converged() ***then***
	Population.restart()
	end if
Step 6.	(Termination) Continue the process until the termination condition is satisfied

PXRO (Partition Recombination Crossover Operator) is crucial in partition clustering due to its role in improving clustering within complex networks. It enhances cluster quality by optimizing modularity, incorporates domain knowledge for more meaningful clusters, explores a diverse solution space, and aids in evolving the population of high-quality solutions. These features collectively make PXRO a valuable component in the MMA, enhancing the effectiveness of complex network clustering.

3.4 Mutation

In this phase, the vertex chosen from the randomly picked cluster is mutated *n* times thereby exploiting the optimal solution. Firstly, the selection of a chromosome is being done for mutation. Secondly, mutation is applied using onepoint mutation on the selected chromosome Thus, this phase contributes to the exploitation and avoids exploration.

3.5 Population Updation Strategy

Let U be the current population and C^0 be the offspring to be considered for inclusion in U. Let $C \in U$ be the closest cluster to C^0 according to the ERI distance and $C_W \in U$ the worst clustering (with the smallest modularity). Let δ_{min} is a fixed distance threshold. We apply the following replacement rule: if $d(C^0, C_C) < \delta_{min}$ and $Q(C^0) \geq Q(C_C)$, then C^0 replaces C_C in U, otherwise, if $Q(C^0) \geq Q(C_w)$ then C^0 replaces C_w in U. The analysis reveals that if the PXRO algorithm is used with two similar individuals as inputs, the obtained result will also be identical. As a result, if an individual already exists in the population, the algorithm will not enable the individuals to be added. This reduces the chance of a premature takeover and helps to maintain population diversity. Therefore, the population size can be kept low. The procedure maintains the concept of offspring tuning in order to improve the individual. Due to the offspring tuning the individuals in the initial population have a chance to survive for a long time, contributing their randomly generated small clusters to diversity and maintaining a diverse population of individuals, thus, enabling exploration and when combined with fit individuals, enabling exploitation. Every time a new individual is discovered, it is added to the population.

4. EXPERIMENTAL RESULTS

The proposed MMA is verified on complex network datasets. The datasets consist of the Zachary Karate Club and Lusseau's network of bottlenose dolphins [12]. The karate club dataset is of 34 nodes that represents the size of the network. The network of 62 bottlenose dolphins represents the number of nodes in dolphins' dataset. The experimental analysis is performed on python 3.6 using Intel core i5 with 2.4 GHz speed, 4GB RAM, and Windows 10 operating system. The parameters that may affect the performance of the developed algorithm are (a.) population Size (b.) number of generations and (c.) mutation probability.

4.1 Parameter Tuning

The individual alteration of parameters is impossible due to their intimate interconnection [33]. The number of generations taken for the experiment are 2000, 4000, 8000, 16000, 32000, 64000, and 128000. The experiment is repeated for population sizes of 10, 20, 30, 50, 100, 150, 200, and 300. The experiment is repeated for probability values of 0, 0.001, 0.01, 0.05, 0.1, 0.2, 0.4, 0.6, and 0.9 for Mutation Probability.

4.2 Evaluation Metrics

The proposed algorithm is tested over 8000 iterations taking NMI, ERI and modularity as evaluation criteria function. NMI ranges between 0 and 1. The '0' value in NMI means 50% of the data objects are correctly clustered while on the other hand '1' value means 100% of the data objects are correctly clustered [34]. ERI also ranges between 0 and 1. The '0' value in ERI means that two clusters do not match on any pair of data objects while on the other hand '1' value means that the data clusters are exactly same [35]. Similarly, the modularity score '1' means that all the edges in a community are connecting nodes within the community. A score of '0' indicates that half of the community's edges connect nodes within the community and other half connect nodes outside the community. The comparative analysis of the MMA is being done with MA-COM [8] and BGLL [11] algorithm over Zachary Karate Club and Dolphins datasets. The number of clusters (K) taken in Zachary Karate Club and Dolphins is 4 and 2 respectively.

Table 1. Benchmark results for the Zachary Karate Club

Metric (Best Case for 8000 iterations)	Zachary Karate Club (K=4)			
	MMA	COMB-MA	BGLL	SAS-LP
NMI	1	0.612	0.512	0.812
ERI	0.02	0.018	0.021	0.021
Modularity	0.419	0.302	0.322	0.312

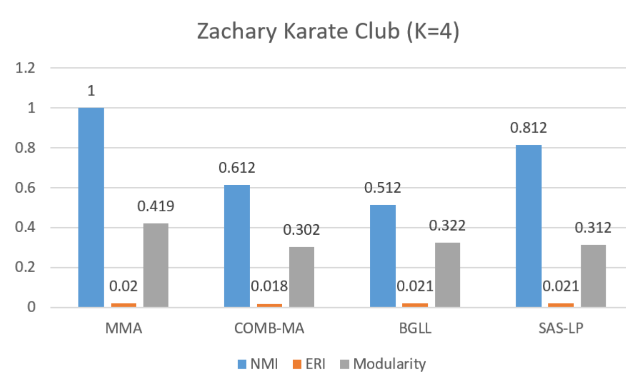

Fig. 2. Best Case Results for Zachary Karate Club (K=4).

Table 1 reflects the results obtained for Zachary Karate Club dataset [36]. The Zachary Karate club dataset consist of 34 nodes. Each node in the dataset represents a member of the club. The network of 34 nodes are connected via edges and the complex network so formed is undirected in nature. The problem associated with Zachary Karate Club dataset is to find the clusters in order to decide the members of each group. The MMA is being applied on Zachary Karate Club dataset. The results obtained in Figure 2 shows the supremacy of the algorithm over two existing algorithms i.e. MA-COM [8] and BGLL [11].

Table 2. Benchmark results for the Dolphins

Metric (Best Case for 8000 iterations)	Dolphins (K=5)			
	MMA	COMB-MA	BGLL	SAS-LP
NMI	0.942	0.818	0.832	0.912
ERI	0.05	0	0	0.021
Modularity	0.527	0.302	0.517	0.483

Table 2 reflects the results obtained for Dolphins dataset [36]. The dolphins' dataset consists of 62 nodes. Each node in the dataset represents a bottlenose dolphin. The network of 62 nodes are undirected in nature and the edge represent frequent communications between dolphins. The MMA is being applied on the dolphin's dataset and the results obtained in Figure 3 are optimal in nature.

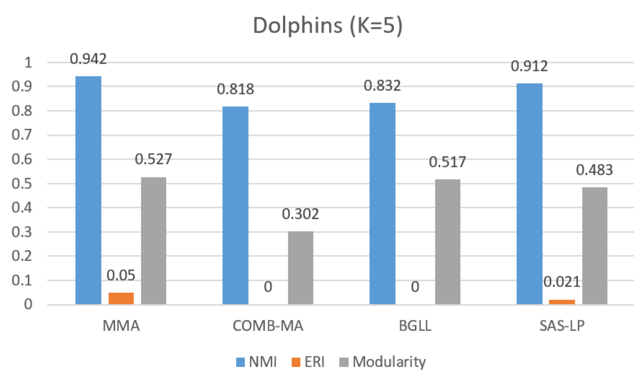

Fig. 3. Best Case Results for Dolphins (K=5).

5. CONCLUSIONS

The partitional clustering in complex network is a non-deterministic polynomial time hard (NP-hard) problem. A modified of memetic algorithm is proposed to solve the problem of partitional clustering in complex network. The proposed algorithm i.e. MMA covers all fundamental aspects of the memetic algorithm. The method has demonstrated some advantages, such as not requiring a fitted initial population, obtaining optimal solutions, and solving the Zachary Karate Club with a large number of generations. If the number of generations is lowered, the capacity to identify optimal solutions remains great. The attempt to fine-tune the genetic algorithm parameters is supported in obtaining optimal solutions. The obtained results are evaluated using relevant field metrics i.e. modularity, ERI, and NMI and thereby confirms the supremacy of the proposed algorithm in solving partitional clustering problem in complex networks. MMA provides promising solution for complex network clustering, but presents specific limitations. MMA heavily relies on parameter settings, making finding the optimal configuration challenging, especially without prior network knowledge. Further, initial population quality affects convergence speed, and scalability is a concern for larger datasets. Generalizability across diverse network types is unexplored, and enhancing convergence speed is crucial for complex networks. The limitations can be address for expanding the algorithm's utility in complex network analysis and clustering. In future, the proposed work can be extended to elevate the local optimum solution and to solve premature convergence for partitional clustering.

REFERENCES

[1] S. Fortunato, "Community detection in graphs," *Phys. Rep.*, vol. 486, no. 3–5, pp. 75–174, 2010, doi: 10.1016/j.physrep.2009.11.002.

[2] S. Das, A. Abraham, and A. Konar, "Automatic clustering using an improved differential evolution algorithm," *IEEE Trans. Syst. Man, Cybern. Part ASystems Humans*, vol. 38, no. 1, pp. 218–237, 2008, doi: 10.1109/TSMCA. 2007.909595.

[3] P. M. Pradhan and G. Panda, "Connectivity constrained wireless sensor deployment using multiobjective evolutionary algorithms and fuzzy decision making," *Ad Hoc Networks*, vol. 10, no. 6, pp. 1134–1145, 2012, doi: 10.1016/j.adhoc.2012.03.001.

[4] S. K. Halgamuge and L. Wang, Eds., *Classification and Clustering for Knowledge Discovery*, 2005th ed. Warsaw: Springer, 2005.

[5] D. Jiang, C. Tang, and A. Zhang, "Cluster analysis for gene expression data: A survey," *IEEE Trans. Knowl. Data Eng.*, vol. 16, no. 11, pp. 1370–1386, 2004, doi: 10.1109/TKDE.2004.68.

[6] M. Q. Duong, L. H. Lam, B. T. M. Tu, G. Q. Huy, and N. H. Hieu, "A combination of K-mean clustering and elbow technique in mitigating losses of distribution network," *GMSARN Int. J.*, vol. 13, no. 4, pp. 153–158, 2019.

[7] Y. Li, L. Jiao, P. Li, and B. Wu, "A hybrid memetic algorithm for global optimization," *Neurocomputing*, vol. 134, pp. 132–139, 2014, doi: 10.1016/j.neucom.2012.12.068.

[8] O. Gach and J. K. Hao, "A memetic algorithm for community detection in complex networks," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2012, vol. 7492 LNCS, no. PART 2, pp. 327–336. doi: 10.1007/978-3-642-32964-7_33.

[9] M. E. J. Newman and M. Girvan, "Finding and evaluating community structure in networks," pp. 1–15, 2004, doi: 10.1103/PhysRevE.69.026113.

[10] H. Swalehe, P. V. Chombo, and B. Marungsri, "Appliance scheduling for optimal load management in smart home integrated with renewable energy by using whale optimization algorithm," *GMSARN Int. J.*, vol. 12, no. 2, pp. 65–75, 2018.

[11] H. Liu, Y. Lu, C. Zhao, L. Guo, L. He, and J. Jian, "BGLL-Based Attribution Overlap Community Discovery Algorithm," *Proc. - 2018 14th Int. Conf. Semant. Knowl. Grids, SKG 2018*, pp. 68–75, 2018, doi: 10.1109/SKG.2018.00017.

[12] B. A. Attea *et al.*, "A review of heuristics and metaheuristics for community detection in complex networks: Current usage, emerging development and future directions," *Swarm*

Evol. Comput., vol. 63, no. April, p. 100885, 2021, doi: 10.1016/j.swevo.2021.100885.

[13] G. Bello-Orgaz, S. Salcedo-Sanz, and D. Camacho, "A Multi-Objective Genetic Algorithm for overlapping community detection based on edge encoding," *Inf. Sci. (Ny).*, vol. 462, pp. 290–314, 2018, doi: 10.1016/j.ins.2018.06.015.

[14] M. Guerrero, F. G. Montoya, R. Baños, A. Alcayde, and C. Gil, "Adaptive community detection in complex networks using genetic algorithms," *Neurocomputing*, vol. 266, pp. 101–113, 2017, doi: 10.1016/j.neucom.2017.05.029.

[15] M. Gong, B. Fu, L. Jiao, and H. Du, "Memetic algorithm for community detection in networks," *Phys. Rev. E - Stat. Nonlinear, Soft Matter Phys.*, vol. 84, no. 5, p. 056101, Nov. 2011, doi: 10.1103/PhysRevE.84.056101.

[16] C. Mu, J. Xie, Y. Liu, F. Chen, Y. Liu, and L. Jiao, "Memetic algorithm with simulated annealing strategy and tightness greedy optimization for community detection in networks," *Appl. Soft Comput. J.*, vol. 34, pp. 485–501, 2015, doi: 10.1016/j.asoc.2015.05.034.

[17] M. Li and J. Liu, "A link clustering based memetic algorithm for overlapping community detection," *Phys. A Stat. Mech. its Appl.*, vol. 503, pp. 410–423, 2018, doi: 10.1016/j.physa.2018.02.133.

[18] D. Chalupa, K. A. Hawick, and J. A. Walker, "Hybrid Bridge-Based Memetic Algorithms for Finding Bottlenecks in Complex Networks," 2018.

[19] F. Neri, C. Cotta, and P. Moscato, *Handbook of Memetic Algorithms*. 2012. doi: 10.1007/978-3-642-23247-3.

[20] "CiNii Articles - Artificial intelligence through a simulation of evolution."

[21] N. Krasnogor and J.E. Smith, "Emergence of Profitable Search Strategies Based on a Simple Inheritance Mechanism," in *Proceedings of the Genetic and Evolutionary Computation Conference (GECCO-2001)*, 2001.

[22] Thomas Back, "Evolutionary Algorithms in Theory and Practice: Evolution Strategies ... - Thomas Back - Google Books," *Oxford University Press*, 1996.

[23] P. Moscato and C. Cotta, "A Gentle Introduction to Memetic Algorithms," in *Handbook of Metaheuristics*, Springer, Boston, MA, 2006, pp. 105–144. doi: 10.1007/0-306-48056-5_5.

[24] W. E. Hart, "Adaptive Global Optimization with Local Search," 1994.

[25] D. Molina, F. Herrera, and M. Lozano, "Adaptive local search parameters for real-coded memetic algorithms," in *2005 IEEE Congress on Evolutionary Computation, IEEE CEC 2005. Proceedings*, 2005, vol. 1, pp. 888–895. doi: 10.1109/cec.2005.1554777.

[26] Q. H. Nguyen, Y. S. Ong, and M. H. Lim, "A probabilistic memetic framework," *IEEE Trans. Evol. Comput.*, vol. 13, no. 3, pp. 604–623, 2009, doi: 10.1109/TEVC.2008.2009460.

[27] H. Nobahari and D. Darabi, "A new adaptive real-coded memetic algorithm," in *2009 International Conference on Artificial Intelligence and Computational Intelligence, AICI 2009*, 2009, vol. 1, pp. 368–372. doi: 10.1109/AICI.2009.259.

[28] N. Noman and H. Iba, "Accelerating differential evolution using an adaptive local search," *IEEE Trans. Evol. Comput.*, vol. 12, no. 1, pp. 107–125, Feb. 2008, doi: 10.1109/TEVC.2007.895272.

[29] C. Liu and B. Li, "Memetic algorithm with adaptive local search depth for large scale global optimization," in *Proceedings of the 2014 IEEE Congress on Evolutionary Computation, CEC 2014*, Sep. 2014, pp. 82–88. doi: 10.1109/CEC.2014.6900456.

[30] A. Ma, Y. Zhong, and L. Zhang, "Adaptive multiobjective memetic fuzzy clustering algorithm for remote sensing imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 8, pp. 4202–4217, Aug. 2015, doi: 10.1109/TGRS.2015.2393357.

[31] Y. Zhong, A. Ma, and L. Zhang, "An adaptive memetic fuzzy clustering algorithm with spatial information for remote sensing imagery," *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.*, vol. 7, no. 4, pp. 1235–1248, 2014, doi: 10.1109/JSTARS.2014.2303634.

[32] W. M. Rand, "Objective Criteria for the Evaluation of Clustering Methods," *J. Am. Stat. Assoc.*, vol. 66, no. December 1971, pp. 37–41, 1971.

[33] B. Yuan and M. Gallagher, "A hybrid approach to parameter tuning in genetic algorithms," in *2005 IEEE Congress on Evolutionary Computation, IEEE CEC 2005. Proceedings*, 2005, vol. 2, pp. 1096–1103. doi: 10.1109/cec.2005. 1554813.

[34] N. N. Kachouie and M. Shutaywi, "Weighted mutual information for aggregated kernel clustering," *Entropy*, vol. 22, no. 3, pp. 1–15, 2020, doi: 10.3390/e22030351.

[35] M. J. Warrens and H. van der Hoef, "Understanding the Adjusted Rand Index and Other Partition Comparison Indices Based on Counting Object Pairs," *J. Classif.*, 2022, doi: 10.1007/s00357-022-09413-z.

[36] K. R. Žalik and B. Žalik, "Memetic algorithm using node entropy and partition entropy for community detection in networks," *Inf. Sci. (Ny).*, vol. 445–446, pp. 38–49, 2018, doi: 10.1016/j.ins.2018.02.063.