Evolutionary Multi-Objective Optimization Algorithm for Community Detection in Complex Social Networks

Tanveerulhuq Shaik¹,² · Vadlamani Ravi¹,³ · Kalyanmoy Deb³

Received: 14 August 2020 / Accepted: 21 October 2020 / Published online: 7 November 2020
© Springer Nature Singapore Pte Ltd 2020

Abstract
Most optimization-based community detection approaches formulate the problem in a single- or bi-objective framework. In this paper, we propose two variants of a three-objective function optimization formulation using a customized non-dominated sorting genetic algorithm III (NSGA-III) to find community structures in a network. The first variant, named NSGA-III-KRM, considered Kernel $k$-means, ratio cut, and modularity, as three objective functions; whereas the second variant, named NSGA-III-CCM, considers community score, community fitness and modularity, as three objective functions. Experiments are conducted on four benchmark network datasets. Comparison with state-of-the-art and baseline methods along with decomposition-based multi-objective evolutionary algorithm variants (MOEA/D-KRM and MOEA/D-CCM) indicates that the proposed variants yield comparable or better results. This is particularly significant because the addition of the third objective does not worsen the results of the other two objectives. We also propose a simple method to rank the Pareto solutions obtained by proposing a new measure—the ratio of the hyper-volume and inverted generational distance. The higher the ratio, the better is the Pareto set. This strategy is particularly useful in the absence of empirical attainment function in the multi-objective framework, where the number of objectives is more than two.

Keywords Community detection · Community fitness · Community score · Kernel $k$ means · Multi-objective optimization · NSGA-III · Modularity · NMI · Ratio cut

Introduction
A complex network can be considered as a graph, having a set of nodes and edges between them. Examples of such networks are: The world wide web, academic collaboration networks, online social networks, food web, biological networks, etc.

Analysis of these complex networks provides us better insights into the quality of interconnections among the nodes such as the identification of important nodes and the structure of underlying communities present in it. Community detection is paramount having numerous applications in e-commerce, communication networks, social networks, biological systems, health care, economics, academia, fraud detection, etc. [1].

The aim of detecting communities is to find the sets of nodes such that each set has nodes that are thickly connected with one another and are loosely connected with the nodes present in the remaining sets. This problem is NP hard [1]. In the last decade, numerous approaches have been propounded to find communities in networks. Some of the techniques include hierarchical clustering algorithms, graph partitioning methods and evolutionary algorithms.

In this paper, community detection in a given undirected and unweighted network is formulated as a multi-objective optimization problem with three objectives and is solved using NSGA-III proposed in [2]. Throughout this paper, the words community and cluster are used interchangeably.
In what follows, “Literature survey” presents the related work, “Motivation” presents the motivation, “Contributions” describes the contribution of the present study, Sect. “Basic definitions” presents basic definitions, “Proposed methodology” presents proposed methodology, “Dataset description” describes the datasets analyzed, “Experiment analysis, results and discussion” presents the results obtained and discussion thereof and finally, “Conclusion” concludes the paper.

**Literature Survey**

In the last decade, several metaheuristic algorithms have been suggested to solve community detection problem in complex networks. In 2003, Newman [3] introduced a classical algorithm which optimizes Modularity in a greedy manner. It uses agglomerative hierarchal clustering method to iteratively maximize Modularity. Later in 2008, Blondel et al. [4] designed another classical two-phase algorithm, which also optimizes Modularity. In the first phase, nodes in one community are shifted to another community one at a time iteratively, if Modularity increases and in second phase communities are merged to get larger communities. In the same year, Pizzuti [5] proposed GA-NET by introducing community score as a metric which describes how well a network is partitioned and optimized this metric using genetic algorithm to find the dense communities in a network. It uses locus-based representation to represent a community structure and it does not need to supply the number of partitions in advance. Thereafter, in 2011, Gong et al. [6] developed MEME-NET. It is observed that Modularity suffers from resolution limit problem [7]. Therefore, they optimized Modularity Density instead of Modularity using genetic algorithm (GA) and included hill climbing for local search to find communities in a network. Later, in 2012, Shang et al. [8] proposed MIGA, which also optimizes Modularity using GA and included simulated annealing to perform local search to find communities in a given network. Then, Pizzuti [9] introduced MOGA-NET, which optimizes two objective functions viz., Community score and Community fitness using GA to detect communities in a network. Then, in 2014, Gong et al. [10] developed MODPSO which optimizes two objective functions viz., Kernel $k$ means and ratio cut using discrete particle swarm optimization algorithm to find communities in a network. This approach can be used for both signed and unsigned networks. Later, in 2017, Abdollahpouri et al. [11] proposed MOPSO-Net, a customized version of particle swarm optimization by altering the moving technique of particles. While moving from one iteration to another, this method uses normalized mutual information (NMI). The disadvantage of using NMI is that it needs the ground truth cluster structure of the graph as input. Hence, this method is not helpful if we do not know the ground truth community structure of the network in advance. In 2018, Yuanyuan et al. [12] proposed two quantum inspired evolutionary algorithms viz., QIEA-net and iQIEA-net to find community structures. QIEA-net detects the communities by optimizing Modularity, and in iQIEA-net, it takes the help of the classical partitioning algorithm. Most recently, Tahmasebi et al. [13] in 2019 proposed a many-objective community detection algorithm which takes five objectives. Out of the five, two objectives cannot be calculated if the ground truth community structure is unknown which, indeed, is the case in the real world. In such cases, those methods cannot be used because the very task here is to find communities, when the ground truth is conspicuously absent.

To sum up, single-objective community detection algorithms lead to some difficulties such as limiting to particular community structure properties. Then, bi-objective formulations did indeed leave out some important measures, which could potentially be used as objective functions. We noticed that some of the measures are indeed non-overlapping conceptually. They describe different aspects of a community. Hence, a different approach is proposed in the current paper, which is a multi-objective (three-objective) optimization framework in two variants to search for communities in complex social networks. This is a clear departure from all the works appeared in the literature so far.

**Motivation**

To the best of our knowledge, except for one of the latest papers, all the works in the literature, formulated community detection in networks as an optimization problem in either single-objective or two-objective optimization environment. Frameworks involving single objective have considered mostly modularity as the objective function, while those with two objectives considered two objectives as follows: Kernel $k$ means and ratio cut, or community fitness and community score, or ratio cut and ratio association, or modularity (by dividing the modularity into two parts and considered each part as one objective). In bi-objective optimization frameworks, one objective maximizes the density of communities and the other minimizes the fraction of interlinks present between communities in the network. For instance, Kernel $k$ means tries to find the solution with maximum community density and Ratio cut tries to find the solutions with minimum fraction of interlinks between communities. For evaluating the effectiveness, they employed modularity and NMI (for networks with known ground truth communities) as external measures outside the optimization process.

If we consider only two objectives, we may get solutions having high community density and less interlinks between communities. However, these solutions may or may not have good community structure. For example, in a network N, if
we consider a solution with only one community consisting of all the nodes in the network, that solution has maximum intra-links and zero interlinks but it is not the best structure because the Modularity value becomes zero for that solution and it does not satisfy the goal of the problem namely to find distinct, non-overlapping communities.

Most recently, Tahmasebi et al. [13] also proposed a many-objective community detection algorithm which takes five objectives. Out of the five, two objectives cannot be calculated if the ground truth community structure of the given network is unknown. Thus, in effect, it reduces to a three-objective function formulation. Further, they used another objective function coverage and mentioned that coverage is the proportion of edges inside the community to the total edges in network. Thus, it refers to the density of a given cluster.

In this paper, we propose a multi-objective optimization framework using three objectives, which try to find solutions with good community densities, less fraction of interlinks and good community structures as well. Our approach is more generic enough as it does not need to know the ground truth community structure in advance. Toward this end, we employed customized NSGA-III as the optimizer. We also studied the effectiveness of the MOEA/D for the problem.

**Contributions**

- Some studies [11] performed the selection of solutions after every generation based on NMI. But, it should be noted that computation of NMI requires the ground truth community structure. These methods are not helpful if we do not know the ground truth community structure of the network in advance. Therefore, we developed a framework, which is generic enough and applicable to all the networks where the ground truth is not necessarily known. In essence, we neither included NMI as the objective function nor used it for evaluating the solutions in progressing from one generation to another generation. This is a radical and well-thought-out departure from the state-of-the-art making our approach extremely useful in real-life situations.

- We formulated community detection problem as a multi-objective optimization problem with three objectives.

- We proposed two variants: (1) NSGA-III-KRM, where we considered Kernel k means, Ratio cut and Modularity as the three objective functions, (2) NSGA-III-CCM, where we considered Community fitness, Community score and Modularity as the objective functions. We also conducted experiments on two variants of MOEA/D [14] (using the penalty-based boundary intersection method), i.e., MOEA/D-KRM and MOEA/D-CCM with the same parameters combinations and with 20 neighbors.

- We used locus-based representation of community structure to represent a solution. In this, an array of size equal to number of vertices present in the network is used to represent a community structure. It is noteworthy that a single solution can be represented in its various permutations. However, technically all of them are one and the same. Hence, we customized NSGA-III to solve this problem by adding a filter, which checks for the presence of duplicate (permutation) solutions in a generated population at the end of each iteration and if present, they are replaced by a randomly generated solution.

**Basic Definitions**

**Community**

Community in a network can be described as a subset of nodes that are thickly connected with one another and loosely connected with the remaining nodes present in that network. Intra-links of a given community are represented as the set of edges present inside the community, whereas, interlinks of a given community are represented by the set of edges connecting the vertices of community c to the vertices not present in community c.

**Multi-Objective Optimization Problem**

Multi-objective optimization problems optimize two or more objective functions simultaneously. Let us consider a problem where we need to maximize nob number of objective functions simultaneously as follows:

\[
\max (f_1(x)), \max (f_2(x)), \ldots, \max (f_{nob}(x)).
\]

where \(x = (x_1, x_2, \ldots, x_{noi})\) is the vector or solutions and \(f_1(x), f_2(x), \ldots, f_{nob}(x)\) are the objective functions that need to be optimized and noi is the dimension of the solution vector. We say that a solution \(x\) dominates another solution \(y\), if all the objective functions values with the solution \(x\) as input are better or equal to the respective values of the objective functions with the solution \(y\) as input and at least one objective function value with solution \(x\) as input should be strictly better than the respective objective function value with solution \(y\) as input [15]. Else, we say that the solution \(x\) does not dominate solution \(y\). We call a solution set \(S\) as non-dominated solution set, if any of the pairs of solutions present in \(S\) does not dominate each other.

More than one solution often exists for these types of problems. If we were given a set \(S\) with all possible solutions, then the subset of the solution set \(S\), i.e., \(T_1\) is called Pareto-set with respect to solution set \(S\) if it contains all the solutions which do not dominate each other and dominate
the rest of the solutions $S - T_1$. Similarly, second Pareto front $T_2$ is the set of solutions, which is subset to set $S - T_1$ which contains all the solutions which do not dominate each other and dominates the rest of the solutions $S - T_1 - T_2$. Similarly, third Pareto front, fourth Pareto front, etc. are defined.

**Proposed Methodology**

**Problem Formulation**

**First Variant**: Kernel $k$ means, ratio cut and modularity as the objective functions.

\[
\begin{align*}
\text{Min } f_1(x) &= \text{Kernel } k \text{ means,} \\
\text{Min } f_2(x) &= \text{Ratio cut,} \\
\text{Max } f_3(x) &= \text{Modularity.}
\end{align*}
\]

Subject to $x \in X$, 
Here, vector $x$ is a community structure of a network encoded using locus-based representation explained in the “Representation of solution” and $X$ is the set of all possible community structures in a network.

**Second Variant**: Community fitness, community score and modularity as the objective functions.

\[
\begin{align*}
\text{Max } f_1(x) &= \text{Community Fitness,} \\
\text{Max } f_2(x) &= \text{Community Score,} \\
\text{Max } f_3(x) &= \text{Modularity.}
\end{align*}
\]

Subject to $x \in X$, 
Here, vector $x$ is a community structure of a network encoded using locus-based representation explained in the “Representation of solution” and $X$ is the set of all possible community structures in a network.

**Objective Functions Considered and Justification**

*Kernel $k$ means (KKM)* [10] is used to find dense communities in a network. KKM is computed as follows:

\[
\text{KKM} = 2(n - m) - \sum_{i=1}^{m} \frac{L(V_i, V_i)}{|V_i|},
\]

where $n$ is the number of vertices in a network, $m$ is the number of communities in a network, $|V_i|$ is the number of vertices in community $i$, $L(V_i, V_i) = \sum_{i,j \in S} A_{ij}$, where $A$ is the adjacency matrix of the network. KKM should be minimized to get structures having denser communities.

*Ratio Cut (RC)* [16] is used to find the clusters in a network such that each cluster present in it is sparsely connected to the remaining other clusters. The formula for computing the ratio cut is as follows:

\[
\text{RC} = \frac{m}{|V_i|} \sum_{i=1}^{m} L(V_i, V_i),
\]

where $m$ is the number of communities in a network, $L(V_i, V_i) = \sum_{i,j \in S} A_{ij}$, where $A$ is adjacency matrix of the network. Here, $V_i$ is the set of vertices in the graph but not present in the set $V_j$. Ratio cut needs to be minimized to get the community structures with less interlinks.

*Community Fitness (CF)* [17] is another measure used to find dense communities in a network. When it reaches its highest value, the number of external links is minimized. The formula for computing the CF is as follows:

\[
\text{CF} = \sum_{i=1}^{k} p(s_i),
\]

where $p(s) = \frac{K_i^{\text{in}}(s)}{[K_i^{\text{in}}(s) + K_i^{\text{out}}(s)]^a}$, $s$ is the community in a network, $K_i^{\text{in}}(s)$ and $K_i^{\text{out}}(s)$ are the internal and external degrees of nodes present in the community $s$, and $a$ is the positive real-valued parameter controlling the community size. The higher the value of the parameter, the smaller is the size of the communities found.

*Community Score (CS)* [5] measures the quality of the division in communities of a network. The higher the CS, the denser the clusters obtained. The formula for computing the CS is as follows:

\[
\text{CS}(s) = \sum_{i=1}^{k} \text{score}(s_i),
\]

\[
\text{score}(s) = M(s) \times V_s,
\]

\[
M(s) = \frac{\sum_{i \in s} (\mu_i)^r}{|s|},
\]

\[
V_s = \sum_{i,j \in S} A_{ij},
\]

\[
\mu_i = \frac{1}{|s|} K_i^{\text{in}}(s),
\]

where $\mu_i$ denotes the fraction of edges connecting node $i$ to the other nodes in $s$, $|s|$ denotes the cardinality of $s$, $S$ is the set of communities, the exponent $r$ increases the weight of nodes having a few connections inside the community $s$. Score of a community $s$, i.e., $\text{score}(s)$ is the product of power mean of $s$ of order $r$, i.e., $M(s)$ and $V_s$, which is the volume of the community $s$; $A$ is the adjacency matrix of the network.
Modularity [18] is defined as the fraction of the edges that fall within the given groups minus the expected fraction if the edges were distributed at random. The modularity is computed as follows:

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

where \(l_s\) is the number of intra-links present in community \(s\), \(d_s\) is the sum of degrees of nodes in community \(s\), \(m\) is the total number of edges in a network, \(k\) is the number of communities found inside a network. Greater modularity value is desirable for a community structure.

**Representation of Solution**

The community detection problem formulated as a multi-objective optimization problem, turns out to be a combinatorial optimization problem. Therefore, we need to suitably represent a community, which becomes a solution in the optimization parlance. Toward this end, we used locus-based representation taking cue from [19, 20]. Here, we consider an \(n\)-dimensional array to represent a solution, where \(n\) is the number of nodes in the network. Each cell index in the array represents a node in the network. In any solution, cell of index \(i\) with value \(j\) means that node \(i\) and node \(j\) in the network belongs to same community. A cell with index \(i\), which represents node \(i\) in the network, can have value \(i\) itself or the indices of nodes which are connected to the node \(i\) with an edge in the network. It is to be noted that a single solution can be represented in its various permutations. However, technically all of them are one and the same.

**NSGA-III Algorithm**

Non-dominated sorting genetic algorithm III (NSGA-III) [2], a multi- and many-objective optimization algorithm, is capable of optimizing 3–15 objective functions simultaneously. This algorithm yields well-diversified and converged solutions. It uses a reference-based framework to select a set of solutions from a substantial number of non-dominated solutions to look for diversity. For more details, the reader is referred to [2].

**Customizations Performed**

In this paper, we performed two customizations on the NSGA-III-based approach: (1) As a single solution can be represented in various ways (meaning its permutations), in a population for any iteration, if a solution is repeated more than once, then we replace it with a randomly generated solution, (2) another customization is that we excluded a solution in which entire network is considered as a single community. These customizations are specific to this problem.

**Evaluation Functions**

Normalized mutual information (NMI) and modularity are widely used to figure out the performance of various evolutionary algorithms invoked to detect clusters in any network. NMI [21] is used to measure the likeness between two cluster structures. NMI can help us calculate how close the clusters detected by an algorithm and the ground truth cluster structure are. The maximum and minimum values possible for NMI are 0 and 1, respectively. The higher the NMI value between two cluster structures, the higher is their likeness. If the NMI value is 1 then it means that both the cluster structures are one and the same. The formula for computing the NMI is as follows:

\[
\text{NMI}(A, B) = \frac{-2 \sum_{i=1}^{R} \sum_{j=1}^{D} C_i \log(C_i/N) / C_j \log(C_j/N)}{\sum_{i=1}^{R} C_i \log(C_i/N) + \sum_{j=1}^{D} C_j \log(C_j/N)},
\]

where \(C_i\) is the number of nodes appeared in both clusters \(i\) and \(j\), present in cluster structures \(A\) and \(B\), respectively. \(C_i(C_j)\) is the number of the elements in cluster \(i\) (cluster \(j\)) present in cluster structure \(A(B)\), \(N\) is the total number of nodes in the network and \(R(D)\) is the number of clusters’ present in the cluster structure \(A(B)\). To make our framework more generic, we have not considered NMI of the network or any other evaluation function which requires the knowledge of ground truth community structure as in most of the real-world networks, the ground truth community structure is unknown.

**Measures of Convergence and Diversity**

To measure the extent of diversity and the state of convergence of the solutions found by multi and many objective optimization algorithms such as NSGA-III, at the end of a run (in other words, after convergence), two widely used criteria include inverted generational distance (IGD) [2, 22] and hyper-volume (HV) [23]. IGD is computed as follows:

\[
\text{IGD}(A, Z_{\text{eff}}) = \frac{1}{|Z_{\text{eff}}|} \sum_{i=1}^{|Z_{\text{eff}}|} \min_{a_j} d(z_i, a_j),
\]

where \(d(z_i, a_j) = \| z_i - a_j \|_2 \), \(A\) is the set of solutions obtained by the algorithm, \(Z_{\text{eff}}\) is the set of points present in Pareto optimal surface. \(a_j\) is a solution present in set \(A\). \(z_i\) is a solution in the Pareto optimal surface which is near to \(a_j\).

The IGD measure indicates how close the obtained solutions are to the solutions present in the true Pareto front or Pareto optimal surface. In cases where the true Pareto front is unknown, we run the algorithm by considering a large population size for a large number of generations. Then, the first Pareto front solutions obtained at the end of the execution are considered as approximation to the Pareto optimal...
solutions [24]. In our case, we considered population size as 500 and number of generations as 500 to approximate the true Pareto optimal surface. The hyper-volume [23] of set $X$ is the volume of space formed by non-dominated points present in the set $X$ with any reference point. Here, the reference point is the “worst possible” point or solution (any point that is dominated by all the points present in solution set $X$) in the objective space. For a maximization (minimization) problem with positive (negative)-valued objectives, we consider origin as the reference point. If a set $X$ has a higher hyper-volume than that of a set $Y$, then we say that $X$ is better than $Y$.

Dataset Description

Four benchmark datasets analyzed in this paper are (1) Zachary’s Karate Club [25] having 34 nodes and 78 edges with two ground truth communities (Fig. 1), (2) Bottlenose Dolphin [26] with 62 nodes, 159 edges and two ground truth communities (Fig. 2), (3) American College Football [27] having 115 nodes, 616 edges with twelve ground truth communities (Fig. 3) and finally, (4) Books about US Politics [28] with 105 nodes, 441 edges and three ground truth communities (Fig. 4). Henceforth, we refer the datasets Zachary’s Karate Club, Bottlenose Dolphin, American College Football and Books about US Politics to as D1, D2, D3 and D4, respectively, for the sake of brevity.

Experiment Analysis, Results and Discussion

Parameter Setting

We performed sensitivity analysis with the parameter combinations presented in Table 3 on all datasets using our proposed variants. We conducted 10 runs for each parameter combination. We computed the product of the highest modularity and the highest NMI obtained towards the finish of each run and then computed the mean of those products (over 10 runs) for each parameter combination. Any parameter combination producing the highest average product of $NMI$ and modularity is considered the best combination. The best parameter combinations obtained for all datasets are presented as follows. It may be mentioned that in problems where the ground truth is unknown, it is impossible to compute NMI. Therefore, we recommend decision-making-based solely on modularity taking cue from several works in literature.

For the variant NSGA-III-KRM, the population sizes with values 100, 200, 500 and 400; crossover probabilities with values 0.8, 0.85, 0.9 and 0.9 and mutation probabilities with values 1/34, 1/124, 1/230 and 2/105 turned out to be the best parameter combinations obtained for the datasets D1, D2, D3 and D4, respectively, after performing sensitivity analysis. For the variant NSGA-III-CCM, the population sizes 200, 200, 450 and 500; crossover probabilities 0.8, 0.85, 0.9 and 0.9 and mutation probabilities 1/68, 1/62, 1/230 and 2/105 turned out to be the best parameter combinations obtained for the datasets D1, D2, D3 and D4, respectively, after performing sensitivity analysis. In literature [9], it is found out that considering community size controlling parameter $\alpha$ of the objective function community fitness as 1 yielded relevant partitions in most of the cases. Further, the parameter $r$ used in the objective function community score is set to 1 to calculate the score of the community by giving equal weight/importance to all the nodes present in it. Hence, the parameters of Community fitness and Community score are kept fixed, i.e., community size controlling parameter $\alpha = 1$ and $r = 1$, respectively.

Results and Discussion

This experiments were conducted on a standalone desktop computer having Intel Xeon(R) CPU E5-2640 v4 2.4 GHz, with 8 cores and 32 GB RAM in Ubuntu 16.04 operating system. For visualizing the optimal communities, we employed Circle Pack layout plugin in Gephi tool (https
The codes for NSGA-III [2] and MOEA/D [14] are adapted from the website https://github.com/msu-coinlab/pymoo and customized to suit our requirement. The ground truth communities of D1, D2, D3 and D4 networks are depicted in Figs. 3, 4, 5 and 6, respectively. Figures 5 and 6, respectively, depict the structures of D1 obtained by the variant NSGA-III-CCM with the highest Modularity and the highest NMI obtained for the best parameter combination (mentioned in the “Parameter setting”). Similarly, Figs. 7 and 8, respectively, depict the structures of D1 obtained by NSGA-III-KRM with the highest modularity and the highest NMI for the best parameter combination (mentioned in the “Parameter setting”). The community structure with the highest modularity obtained using NSGA-III-CCM and that with the highest modularity obtained using NSGA-III-KRM are one and the same. Further, these structures have four communities in them. Out of these four, two are sub-communities of the community present in the ground truth community and other two are sub-communities of another community present in the ground truth community structure. Furthermore, the community structure with the highest NMI

---

**Fig. 3** D3 network (the ground truth)

**Fig. 4** The D4 network (the ground truth)
The optimal community structure obtained for the D2 network depicted in Figs. 9 and 11, where the highest modularity values are yielded by the best parameter combination (mentioned in the “Parameter setting”), respectively, for the two variants turned out to be one and the same. This community structure has five communities. Out of these five, one turned to be the same present in the ground truth and other four are the sub-communities of another community present in the ground truth community structure.

The optimal community structure obtained for D2 network is depicted in Fig. 10 where the highest NMI is yielded by the best parameter combination (mentioned in the “Parameter setting”) using NSGA-III-CCM. Here, one community turned out to be the same present in the ground truth and other three communities are the sub-communities of another community present in the ground truth (Fig. 11).

The optimal community structure obtained for D3 network depicted in Figs. 13 and 15 where the highest Modularity is yielded by the best parameter combination (mentioned in the “Parameter setting”), respectively, for the two variants turned out to be the same. It has ten communities. Out of these, four turned out to be identical to that in the ground truth, three are similar to those in the ground truth but with two or three extra nodes, while the remaining three...
are similar to those in the ground truth with two or three less nodes (Fig. 14).

The optimal community structure obtained for D3 network is depicted in Fig. 14 where the highest NMI is yielded by the best parameter combination (mentioned in the “Parameter setting”) using NSGA-III-CCM. It has 13 communities in it. Out of these, nine turned out to be identical to those in the ground truth, two are similar to those in the ground truth but with one or two less nodes, while the rest...
The optimal community structure obtained for D3 network is depicted in Fig. 16 where the highest NMI is yielded by the best parameter combination (mentioned in the “Parameter setting”) using NSGA-III-KRM. It contains 11 communities in it. Out of these, six turned out to be identical to the ones in the ground truth, three are similar to those in the ground truth but with two or three extra nodes, while the remaining two are similar to those in the ground truth but with 1 or 2 less nodes.

The optimal community structure obtained for D4 network is depicted in Fig. 18 where the highest NMI is yielded by the best parameter combination (mentioned in the “Parameter setting”), respectively, using both variants turned out to be identical. It has five communities in it. Out of these, two are sub-communities of two communities present in the ground truth having two extra nodes belonging to another communities. Other three contain nodes belonging to third community in the ground truth and nodes left out in above two communities (Fig. 18).

The optimal community structure obtained for D4 network is depicted in Fig. 19 where the highest NMI is yielded by the best parameter combination (mentioned in the “Parameter setting”) using NSGA-III-CCM. This community structure has four communities in it. Out of these, two are sub-communities of two communities present in the ground truth but with two extra nodes belonging to other communities. Other three contain nodes belonging to the third community in the ground truth and nodes left out in above two communities (Fig. 19).
The optimal community structure obtained for D4 network is depicted in Fig. 20 where the highest NMI is yielded by the best parameter combination (mentioned in the “Parameter setting”) using NSGA-III-KRM. This community structure has three communities in it. Out of these, two are the sub-communities of two communities present in the ground truth having two extra nodes belonging to another communities. Remaining one contains nodes belonging to the third community in the ground truth and nodes left out in above two communities.

As modularity is widely used for comparison in the literature, we too compared the modularity values yielded by different state-of-the-art approaches in the recently published paper [12] with the optimal Modularity obtained by our methods. This is despite the fact that Modularity is an objective function in both the proposed formulations. This is done for the purpose of comparison only.

Accordingly, in Table 1, we compared the average modularity and maximum modularity obtained by the proposed variant i.e. NSGA-III-KRM and NSGA-III-CCM with that of 9 state-of-the-art/baseline approaches, namely, FN [3], BGLL [4], MIGA [8], MEME-net [6], GA-Net [5], MOGA-net [9], MODPSO [10], QIEA-net and IQIEA-net [12] and also with MOEA/D variants, i.e., MOEA/D-KRM and MOEA/D-CCM.

For the D1 dataset, our proposed variants of NSGA-III, MOGA-net [9], QIEA-net and IQIEA-net [12] yielded the same modularity values. For D2 and D4 datasets, both variants of NSGA-III obtained the highest modularity compared to that of all algorithms. For D3 dataset, BGLL [4], proposed NSGA-III variants and MOEA/D-KRM obtained the highest modularity; the mean modularity values obtained by them are close to each other and higher compared to that of the other algorithms. It can be seen from Table 1 that proposed NSGA-III variants achieved the best or equal Modularity value compared to the remaining approaches. The average NMI for all the datasets obtained by both variants using the best parameter combination are presented in Table 2. The communities with the highest modularity obtained by both proposed variants are one and the same, when compared with the ground truth communities. The plots of the sensitivity analysis are depicted in Figs. 21, 22, 23, 24, 25, 26, 27, 28 in “Appendix” (Table 3).

We observed from Table 1 that NSGA-III-KRM outperformed NSGA-III-CCM on two datasets D2 and D3 in terms of average modularity, and producing same optimal results as most of the other baseline approaches on a small dataset D1. This is attributed to the more information contained in NSGA-III-KRM vis-à-vis NSGA-III-CCM in that the former obtained communities closer to the ground truth. This is because unlike other objective functions, Kernel $k$-means derives its strength from the traditional $k$-means clustering method. Further, we point out that at a high level, these measures (objective functions) are designed to measure the density within the communities and separation of communities,

---

**Fig. 18** The obtained community structure by NSGA-III – CCM with highest NMI on D4 network

**Fig. 19** The obtained community structure by NSGA-III -KRM with highest Modularity on D4 network

**Fig. 20** The obtained community structure by NSGA-III-KRM with highest NMI on D4 network
from one another. However, we notice that their performance is dataset dependent.

However, both variants of NSGA-III outperformed MOEA/D variants i.e. MOEA/D-III-KRM and MOEA/D-CCM on all datasets with respect to average modularity. This is because of the superiority of NSGA-III over MOEA/D in obtaining more diverse and better convergent solutions. In cases where our proposed variants could not yield better results than those of the baseline algorithms, we notice that the datasets are too small to pose any challenge to any reasonable baseline algorithm.

Further, to know the diversity and convergence aspects of the solutions obtained by the proposed methods and to see how close the obtained Pareto front is to the true Pareto front or Pareto optimal surface, we computed the ratio of HV and IGD values of solution set obtained at the end of each run. Then, we computed the average HV/IGD ratios for each parameter combination. The results obtained are presented in the Tables 4, 5, 6, 7, 8, 9, 10, 11 in “Appendix”. The ratio HV/IGD is indeed proposed for the first time as a proxy for the empirical attainment function plots used in the bi-objective optimization setting because a similar kind of plot is not yet proposed in the literature for multi/many objective optimization algorithms. This is another significant contribution of the study.

### Conclusion

A novel multi-objective community detection algorithm with two variants, i.e., NSGA-III-KRM, NSGA-III-CCM is proposed in this paper. In the first variant, i.e., NSGA-III-KRM, Kernel $k$ means, ratio cut and modularity are used as the objective functions. In the second variant, i.e., NSGA-III-CCM, community fitness, community score and modularity are used as the objective functions. Customization is performed on the NSGA-III algorithm that checks...
and removes the redundant solutions present in the population at the end of each iteration. The product of modularity and NMI is suggested to find the best parameter combination. Both NSGA-III-KRM and NSGA-III-CCM, are compared with nine state-of-the-art baseline algorithms and MOEA/D variants (MOEA/D-KRM and MOEA/D-CCM). It turned out that our proposed variants yielded the best or identical results in terms of high modularity. Hence, we conclude that our proposed variants have found community structures where the nodes of a community are thickly connected with one another and nodes in different communities are well separated, which is a hallmark of this study.

We also proposed a new measure as an alternative to the empirical attainment function plot available in bi-objective optimization framework.

### Compliance with Ethical Standards

**Conflict of Interest** On behalf of all authors, the corresponding author states that there is no conflict of interest.

### Appendix

See Tables 4, 5, 6, 7, 8, 9, 10, 11 and Figs. 21, 22, 23, 24, 25, 26, 27, 28.

---

**Table 3** Parameter combination considered for datasets when doing sensitivity analysis

| Dataset | Population size | #Generations | Crossover probabilities | Mutation probabilities |
|---------|-----------------|--------------|-------------------------|------------------------|
| D1      | 100, 150, 200   | 100          | 0.8, 0.85, 0.9          | 1/34, 2/34, 1/(2*34)   |
| D2      | 200, 250, 300   | 100          | 0.8, 0.85, 0.9          | 1/62, 2/62, 1/(2*62)   |
| D3      | 400, 450, 500   | 100          | 0.8, 0.85, 0.9          | 1/115, 2/115, 1/(2*115) |
| D4      | 400, 450, 500   | 100          | 0.8, 0.85, 0.9          | 1/105, 2/105, 1/(2*105) |

**Table 4** Average IGD and HV values for each parameter combination obtained for Zachary’s karate club dataset using NSGA III-KRM

| Population size | Generations | Crossover Probability | Mutation Probability | HV/IGD mean | HV/IGD max |
|-----------------|-------------|-----------------------|----------------------|-------------|------------|
| 100             | 100         | 0.8                   | 0.0147               | 130,245.6   | 195,525.4  |
| 100             | 100         | 0.8                   | 0.0294               | 176,007     | 530,593.9  |
| 100             | 100         | 0.8                   | 0.0588               | 119,389.5   | 209,730    |
| 100             | 100         | 0.85                  | 0.0147               | 130,076.6   | 199,905.9  |
| 100             | 100         | 0.85                  | 0.0294               | 152,105.7   | 346,405.8  |
| 100             | 100         | 0.85                  | 0.0588               | 129,156.5   | 278,482.2  |
| 100             | 100         | 0.9                   | 0.0147               | 121,333.8   | 281,488.7  |
| 100             | 100         | 0.9                   | 0.0294               | 125,364.7   | 240,377.2  |
| 100             | 100         | 0.9                   | 0.0588               | 124,365.1   | 229,433.6  |
| 150             | 100         | 0.8                   | 0.0147               | 130,076.6   | 195,525.4  |
| 150             | 100         | 0.8                   | 0.0294               | 176,007     | 530,593.9  |
| 150             | 100         | 0.8                   | 0.0588               | 124,365.1   | 195,525.4  |
| 150             | 100         | 0.85                  | 0.0147               | 121,333.8   | 281,488.7  |
| 150             | 100         | 0.85                  | 0.0294               | 125,364.7   | 240,377.2  |
| 150             | 100         | 0.85                  | 0.0588               | 124,365.1   | 229,433.6  |
| 200             | 100         | 0.8                   | 0.0147               | 130,245.6   | 195,525.4  |
| 200             | 100         | 0.8                   | 0.0294               | 176,007     | 530,593.9  |
| 200             | 100         | 0.8                   | 0.0588               | 119,389.5   | 209,730    |
| 200             | 100         | 0.85                  | 0.0147               | 130,076.6   | 199,905.9  |
| 200             | 100         | 0.85                  | 0.0294               | 152,105.7   | 346,405.8  |
| 200             | 100         | 0.85                  | 0.0588               | 129,156.5   | 278,482.2  |
| 200             | 100         | 0.9                   | 0.0147               | 121,333.8   | 281,488.7  |
| 200             | 100         | 0.9                   | 0.0294               | 125,364.7   | 240,377.2  |
| 200             | 100         | 0.9                   | 0.0588               | 124,365.1   | 229,433.6  |
| Population size | Generations | Crossover probability | Mutation probability | HV/IGD mean | HV/IGD max |
|----------------|-------------|-----------------------|----------------------|-------------|-----------|
| 100            | 100         | 0.8                   | 0.0147               | 7135.85     | 18,741.74 |
| 100            | 100         | 0.8                   | 0.0294               | 8183        | 11,461.89 |
| 100            | 100         | 0.8                   | 0.0588               | 10,041.27   | 38,391.9  |
| 100            | 100         | 0.85                  | 0.0147               | 10,191.73   | 21,908.42 |
| 100            | 100         | 0.85                  | 0.0294               | 7795.31     | 12,362.8  |
| 100            | 100         | 0.85                  | 0.0588               | 8205.73     | 11,545.57 |
| 100            | 100         | 0.9                   | 0.0147               | 7741.37     | 15,742.72 |
| 100            | 100         | 0.9                   | 0.0294               | 11,628.64   | 31,756.23 |
| 100            | 100         | 0.9                   | 0.0588               | 7901.19     | 15,167.73 |
| 150            | 100         | 0.8                   | 0.0147               | 7135.85     | 18,741.74 |
| 150            | 100         | 0.8                   | 0.0294               | 8183        | 11,461.89 |
| 150            | 100         | 0.8                   | 0.0588               | 10,041.27   | 38,391.9  |
| 150            | 100         | 0.85                  | 0.0147               | 10,191.73   | 21,908.42 |
| 150            | 100         | 0.85                  | 0.0294               | 7795.31     | 12,362.8  |
| 150            | 100         | 0.85                  | 0.0588               | 8205.73     | 11,545.57 |
| 150            | 100         | 0.9                   | 0.0147               | 7741.37     | 15,742.72 |
| 150            | 100         | 0.9                   | 0.0294               | 11,628.64   | 31,756.23 |
| 150            | 100         | 0.9                   | 0.0588               | 7901.19     | 15,167.73 |
| 200            | 100         | 0.8                   | 0.0147               | 7135.85     | 18,741.74 |
| 200            | 100         | 0.8                   | 0.0294               | 8183        | 11,461.89 |
| 200            | 100         | 0.8                   | 0.0588               | 10,041.27   | 38,391.9  |
| 200            | 100         | 0.85                  | 0.0147               | 10,191.73   | 21,908.42 |
| 200            | 100         | 0.85                  | 0.0294               | 7795.31     | 12,362.8  |
| 200            | 100         | 0.85                  | 0.0588               | 8205.73     | 11,545.57 |
| 200            | 100         | 0.9                   | 0.0147               | 7741.37     | 15,742.72 |
| 200            | 100         | 0.9                   | 0.0294               | 11,628.64   | 31,756.23 |
| 200            | 100         | 0.9                   | 0.0588               | 7901.19     | 15,167.73 |
| Population size | Generations | Crossover probability | Mutation probability | HV/IGD mean | HV/IGD max |
|-----------------|-------------|-----------------------|----------------------|-------------|------------|
| 200             | 100         | 0.8                   | 0.0081               | 199,009.3   | 316,596.1  |
| 200             | 100         | 0.8                   | 0.0161               | 229,998.6   | 375,624.5  |
| 200             | 100         | 0.8                   | 0.0322               | 199,603     | 302,670.6  |
| 200             | 100         | 0.85                  | 0.0081               | 192,377.3   | 339,173.8  |
| 200             | 100         | 0.85                  | 0.0161               | 212,951     | 348,086.3  |
| 200             | 100         | 0.85                  | 0.0322               | 188,602.5   | 255,757.9  |
| 200             | 100         | 0.9                   | 0.0081               | 195,061.4   | 324,119.9  |
| 200             | 100         | 0.9                   | 0.0161               | 164,325.8   | 248,247.1  |
| 200             | 100         | 0.9                   | 0.0322               | 186,296     | 230,404    |
| 250             | 100         | 0.8                   | 0.0081               | 199,009.3   | 316,596.1  |
| 250             | 100         | 0.8                   | 0.0161               | 229,998.6   | 375,624.5  |
| 250             | 100         | 0.8                   | 0.0322               | 199,603     | 302,670.6  |
| 250             | 100         | 0.85                  | 0.0081               | 192,377.3   | 339,173.8  |
| 250             | 100         | 0.85                  | 0.0161               | 212,951     | 348,086.3  |
| 250             | 100         | 0.85                  | 0.0322               | 188,602.5   | 255,757.9  |
| 250             | 100         | 0.9                   | 0.0081               | 164,325.8   | 248,247.1  |
| 250             | 100         | 0.9                   | 0.0161               | 164,325.8   | 248,247.1  |
| 250             | 100         | 0.9                   | 0.0322               | 186,296     | 230,404    |
| 300             | 100         | 0.8                   | 0.0081               | 199,009.3   | 316,596.1  |
| 300             | 100         | 0.8                   | 0.0161               | 229,998.6   | 375,624.5  |
| 300             | 100         | 0.8                   | 0.0322               | 199,603     | 302,670.6  |
| 300             | 100         | 0.85                  | 0.0081               | 192,377.3   | 339,173.8  |
| 300             | 100         | 0.85                  | 0.0161               | 212,951     | 348,086.3  |
| 300             | 100         | 0.85                  | 0.0322               | 188,602.5   | 255,757.9  |
| 300             | 100         | 0.9                   | 0.0081               | 195,061.4   | 324,119.9  |
| 300             | 100         | 0.9                   | 0.0161               | 164,325.8   | 248,247.1  |
| 300             | 100         | 0.9                   | 0.0322               | 186,296     | 230,404    |
Table 7  Average IGD and HV values for each parameter combination obtained for Bottlenose Dolphin club dataset using NSGA III-CCM

| Population size | Generations | Crossover probability | Mutation probability | HV/IGD mean | HV/IGD max |
|-----------------|-------------|-----------------------|----------------------|-------------|------------|
| 200             | 100         | 0.8                   | 0.0081               | 14,481.02   | 27,483.88  |
| 200             | 100         | 0.8                   | 0.0161               | 17,151.4    | 24,975.29  |
| 200             | 100         | 0.8                   | 0.0322               | 11,640.36   | 20,000.16  |
| 200             | 100         | 0.85                  | 0.0081               | 17,426.11   | 39,490.12  |
| 200             | 100         | 0.85                  | 0.0161               | 14,118.57   | 25,740.14  |
| 200             | 100         | 0.85                  | 0.0322               | 17,530.78   | 28,096.77  |
| 200             | 100         | 0.9                   | 0.0081               | 14,590.83   | 27,520.53  |
| 200             | 100         | 0.9                   | 0.0161               | 16,103.27   | 28,807.43  |
| 200             | 100         | 0.9                   | 0.0322               | 18,114.06   | 28,079.01  |
| 250             | 100         | 0.8                   | 0.0081               | 14,481.02   | 27,483.88  |
| 250             | 100         | 0.8                   | 0.0161               | 17,151.4    | 24,975.29  |
| 250             | 100         | 0.8                   | 0.0322               | 11,640.36   | 20,000.16  |
| 250             | 100         | 0.85                  | 0.0081               | 17,426.11   | 39,490.12  |
| 250             | 100         | 0.85                  | 0.0161               | 14,118.57   | 25,740.14  |
| 250             | 100         | 0.85                  | 0.0322               | 17,530.78   | 28,096.77  |
| 250             | 100         | 0.9                   | 0.0081               | 14,590.83   | 27,520.53  |
| 250             | 100         | 0.9                   | 0.0161               | 16,103.27   | 28,807.43  |
| 250             | 100         | 0.9                   | 0.0322               | 18,114.06   | 28,079.01  |
| 300             | 100         | 0.8                   | 0.0081               | 14,481.02   | 27,483.88  |
| 300             | 100         | 0.8                   | 0.0161               | 17,151.4    | 24,975.29  |
| 300             | 100         | 0.8                   | 0.0322               | 11,640.36   | 20,000.16  |
| 300             | 100         | 0.85                  | 0.0081               | 17,426.11   | 39,490.12  |
| 300             | 100         | 0.85                  | 0.0161               | 14,118.57   | 25,740.14  |
| 300             | 100         | 0.85                  | 0.0322               | 17,530.78   | 28,096.77  |
| 300             | 100         | 0.9                   | 0.0081               | 14,590.83   | 27,520.53  |
| 300             | 100         | 0.9                   | 0.0161               | 16,103.27   | 28,807.43  |
| 300             | 100         | 0.9                   | 0.0322               | 18,114.06   | 28,079.01  |
| Population size | Generations | Crossover probability | Mutation probability | HV/IGD mean | HV/IGD max |
|-----------------|-------------|-----------------------|----------------------|-------------|------------|
| 400             | 100         | 0.8                   | 0.0043               | 439,752.6   | 721,342    |
| 400             | 100         | 0.8                   | 0.0087               | 515,214     | 958,453.7  |
| 400             | 100         | 0.8                   | 0.0174               | 467,021     | 633,551    |
| 400             | 100         | 0.85                  | 0.0043               | 611,048.6   | 1,389,110  |
| 400             | 100         | 0.85                  | 0.0087               | 586,472.2   | 1,130,415  |
| 400             | 100         | 0.85                  | 0.0174               | 580,567.2   | 806,541.8  |
| 400             | 100         | 0.9                   | 0.0043               | 811,402.8   | 1,072,857  |
| 400             | 100         | 0.9                   | 0.0087               | 515,076.5   | 1,003,225  |
| 400             | 100         | 0.9                   | 0.0174               | 596,012.6   | 1,022,361  |
| 450             | 100         | 0.8                   | 0.0043               | 439,752.6   | 721,342    |
| 450             | 100         | 0.8                   | 0.0087               | 515,214     | 958,453.7  |
| 450             | 100         | 0.8                   | 0.0174               | 467,021     | 633,551    |
| 450             | 100         | 0.85                  | 0.0043               | 611,048.6   | 1,389,110  |
| 450             | 100         | 0.85                  | 0.0087               | 586,472.2   | 1,130,415  |
| 450             | 100         | 0.85                  | 0.0174               | 580,567.2   | 806,541.8  |
| 450             | 100         | 0.9                   | 0.0043               | 811,402.8   | 1,072,857  |
| 450             | 100         | 0.9                   | 0.0087               | 515,076.5   | 1,003,225  |
| 450             | 100         | 0.9                   | 0.0174               | 596,012.6   | 1,022,361  |
| 500             | 100         | 0.8                   | 0.0043               | 439,752.6   | 721,342    |
| 500             | 100         | 0.8                   | 0.0087               | 515,214     | 958,453.7  |
| 500             | 100         | 0.8                   | 0.0174               | 467,021     | 633,551    |
| 500             | 100         | 0.85                  | 0.0043               | 611,048.6   | 1,389,110  |
| 500             | 100         | 0.85                  | 0.0087               | 586,472.2   | 1,130,415  |
| 500             | 100         | 0.85                  | 0.0174               | 580,567.2   | 806,541.8  |
| 500             | 100         | 0.9                   | 0.0043               | 811,402.8   | 1,072,857  |
| 500             | 100         | 0.9                   | 0.0087               | 515,076.5   | 1,003,225  |
| 500             | 100         | 0.9                   | 0.0174               | 596,012.6   | 1,022,361  |
Table 9  Average IGD and HV values for each parameter combination obtained for American College Football Club dataset using NSGA III-CCM

| Population size | Generations | Crossover probability | Mutation probability | HV/IGD mean | HV/IGD max |
|----------------|-------------|-----------------------|----------------------|--------------|------------|
| 400            | 100         | 0.8                   | 0.0043               | 60,184.29    | 160,903.4  |
| 400            | 100         | 0.8                   | 0.0087               | 59,089.5     | 123,946.4  |
| 400            | 100         | 0.8                   | 0.0174               | 40,253.66    | 85,587.95  |
| 400            | 100         | 0.85                  | 0.0043               | 46,844.43    | 118,434.7  |
| 400            | 100         | 0.85                  | 0.0087               | 95,767.54    | 425,939.8  |
| 400            | 100         | 0.85                  | 0.0174               | 56,949.12    | 133,113.8  |
| 400            | 100         | 0.9                   | 0.0043               | 44,035.57    | 88,556.5   |
| 400            | 100         | 0.9                   | 0.0087               | 57,628.37    | 103,291.5  |
| 400            | 100         | 0.9                   | 0.0174               | 50,456.42    | 132,127.6  |
| 450            | 100         | 0.8                   | 0.0043               | 60,184.29    | 160,903.4  |
| 450            | 100         | 0.8                   | 0.0087               | 59,089.5     | 123,946.4  |
| 450            | 100         | 0.8                   | 0.0174               | 40,253.66    | 85,587.95  |
| 450            | 100         | 0.85                  | 0.0043               | 46,844.43    | 118,434.7  |
| 450            | 100         | 0.85                  | 0.0087               | 95,767.54    | 425,939.8  |
| 450            | 100         | 0.85                  | 0.0174               | 56,949.12    | 133,113.8  |
| 450            | 100         | 0.9                   | 0.0043               | 44,035.57    | 88,556.5   |
| 450            | 100         | 0.9                   | 0.0087               | 57,628.37    | 103,291.5  |
| 450            | 100         | 0.9                   | 0.0174               | 50,456.42    | 132,127.6  |
| 500            | 100         | 0.8                   | 0.0043               | 60,184.29    | 160,903.4  |
| 500            | 100         | 0.8                   | 0.0087               | 59,089.5     | 123,946.4  |
| 500            | 100         | 0.8                   | 0.0174               | 40,253.66    | 85,587.95  |
| 500            | 100         | 0.85                  | 0.0043               | 46,844.43    | 118,434.7  |
| 500            | 100         | 0.85                  | 0.0087               | 95,767.54    | 425,939.8  |
| 500            | 100         | 0.85                  | 0.0174               | 56,949.12    | 133,113.8  |
| 500            | 100         | 0.9                   | 0.0043               | 44,035.57    | 88,556.5   |
| 500            | 100         | 0.9                   | 0.0087               | 57,628.37    | 103,291.5  |
| 500            | 100         | 0.9                   | 0.0174               | 50,456.42    | 132,127.6  |
Table 10  Average IGD and HV values for each parameter combination obtained for books about us politics dataset using NSGA III-KRM

| Population size | Generations | Crossover probability | Mutation probability | HV/IGD mean | HV/IGD max |
|-----------------|-------------|-----------------------|----------------------|-------------|------------|
| 400             | 100         | 0.8                   | 0.0048               | 273,578.6   | 519,785.3  |
| 400             | 100         | 0.8                   | 0.0095               | 336,682.7   | 568,148.3  |
| 400             | 100         | 0.8                   | 0.0191               | 317,074     | 606,453.4  |
| 400             | 100         | 0.85                  | 0.0048               | 315,524.4   | 579,135.5  |
| 400             | 100         | 0.85                  | 0.0095               | 286,304.3   | 564,227.6  |
| 400             | 100         | 0.85                  | 0.0191               | 382,656.3   | 529,668.8  |
| 400             | 100         | 0.9                   | 0.0048               | 321,926     | 460,325    |
| 400             | 100         | 0.9                   | 0.0095               | 288,575.4   | 461,571.9  |
| 400             | 100         | 0.9                   | 0.0191               | 280,119.6   | 383,115.5  |
| 450             | 100         | 0.8                   | 0.0048               | 273,578.6   | 519,785.3  |
| 450             | 100         | 0.8                   | 0.0095               | 336,682.7   | 568,148.3  |
| 450             | 100         | 0.8                   | 0.0191               | 317,074     | 606,453.4  |
| 450             | 100         | 0.85                  | 0.0048               | 315,524.4   | 579,135.5  |
| 450             | 100         | 0.85                  | 0.0095               | 286,304.3   | 564,227.6  |
| 450             | 100         | 0.85                  | 0.0191               | 382,656.3   | 529,668.8  |
| 450             | 100         | 0.9                   | 0.0048               | 321,926     | 460,325    |
| 450             | 100         | 0.9                   | 0.0095               | 288,575.4   | 461,571.9  |
| 450             | 100         | 0.9                   | 0.0191               | 280,119.6   | 383,115.5  |
| 500             | 100         | 0.8                   | 0.0048               | 273,578.6   | 519,785.3  |
| 500             | 100         | 0.8                   | 0.0095               | 336,682.7   | 568,148.3  |
| 500             | 100         | 0.8                   | 0.0191               | 317,074     | 606,453.4  |
| 500             | 100         | 0.85                  | 0.0048               | 315,524.4   | 579,135.5  |
| 500             | 100         | 0.85                  | 0.0095               | 286,304.3   | 564,227.6  |
| 500             | 100         | 0.85                  | 0.0191               | 382,656.3   | 529,668.8  |
| 500             | 100         | 0.9                   | 0.0048               | 321,926     | 460,325    |
| 500             | 100         | 0.9                   | 0.0095               | 288,575.4   | 461,571.9  |
| 500             | 100         | 0.9                   | 0.0191               | 280,119.6   | 383,115.5  |
Table 11  Average IGD and HV values for each parameter combination obtained for books about us politics dataset using NSGA III-CCM

| Population size | Generations | Crossover probability | Mutation probability | HV/IGD mean | HV/IGD max |
|-----------------|-------------|-----------------------|----------------------|-------------|------------|
| 400             | 100         | 0.8                   | 0.0048               | 16,709.43   | 30,704.67 |
| 400             | 100         | 0.8                   | 0.0095               | 17,396.99   | 32,342.02 |
| 400             | 100         | 0.8                   | 0.0191               | 10,238.8    | 19,612.07 |
| 400             | 100         | 0.85                  | 0.0048               | 14,768.05   | 34,681.02 |
| 400             | 100         | 0.85                  | 0.0095               | 16,828.09   | 32,870.13 |
| 400             | 100         | 0.85                  | 0.0191               | 12,010.52   | 21,758.29 |
| 400             | 100         | 0.9                   | 0.0048               | 15,860.12   | 33,398.29 |
| 400             | 100         | 0.9                   | 0.0095               | 12,166.83   | 24,472.24 |
| 400             | 100         | 0.9                   | 0.0191               | 12,474.38   | 32,580.1  |
| 450             | 100         | 0.8                   | 0.0048               | 16,709.43   | 30,704.67 |
| 450             | 100         | 0.8                   | 0.0095               | 17,396.99   | 32,342.02 |
| 450             | 100         | 0.8                   | 0.0191               | 10,238.8    | 19,612.07 |
| 450             | 100         | 0.85                  | 0.0048               | 14,768.05   | 34,681.02 |
| 450             | 100         | 0.85                  | 0.0095               | 16,828.09   | 32,870.13 |
| 450             | 100         | 0.85                  | 0.0191               | 12,010.52   | 21,758.29 |
| 450             | 100         | 0.9                   | 0.0048               | 15,860.12   | 33,398.29 |
| 450             | 100         | 0.9                   | 0.0095               | 12,166.83   | 24,472.24 |
| 450             | 100         | 0.9                   | 0.0191               | 12,474.38   | 32,580.1  |
| 500             | 100         | 0.8                   | 0.0048               | 16,709.43   | 30,704.67 |
| 500             | 100         | 0.8                   | 0.0095               | 17,396.99   | 32,342.02 |
| 500             | 100         | 0.8                   | 0.0191               | 10,238.8    | 19,612.07 |
| 500             | 100         | 0.85                  | 0.0048               | 14,768.05   | 34,681.02 |
| 500             | 100         | 0.85                  | 0.0095               | 16,828.09   | 32,870.13 |
| 500             | 100         | 0.85                  | 0.0191               | 12,010.52   | 21,758.29 |
| 500             | 100         | 0.9                   | 0.0048               | 15,860.12   | 33,398.29 |
| 500             | 100         | 0.9                   | 0.0095               | 12,166.83   | 24,472.24 |
| 500             | 100         | 0.9                   | 0.0191               | 12,474.38   | 32,580.1  |

Fig. 21  Modularity*NMI vs population size of karate club dataset using NSGA3-KKM
Fig. 22  Modularity*NMI vs population size of Zachary’s Karate Club dataset using NSGA3-CCM

Fig. 23  Modularity*NMI vs population size of Bottlenose Dolphin dataset using NSGA3-KKM
Fig. 24  Modularity*NMI vs population size of Bottlenose Dolphin dataset using NSGA3-CCM

Fig. 25  Modularity*NMI vs population size of American College Football dataset using NSGA3-KKM
Fig. 26  Modularity*NMI vs population size of American College Football dataset using NSGA3-CCM

Fig. 27  Modularity*NMI vs population size of Political dataset using NSGA3-KKM
References

1. Javed MA, Younis MS, Latif S, Qadir J, Baig A. Community detection in networks: a multidisciplinary review. J Netw Comput Appl. 2018;108:87–111. https://doi.org/10.1016/j.jnca.2018.02.011.

2. Deb K, Jain H. An evolutionary many-objective optimization algorithm using reference-point-based nondominated sorting approach, part I: solving problems with box constraints. IEEE Trans Evol Comput. 2014;18:577–601. https://doi.org/10.1109/TEVC.2013.2281535.

3. Newman MEJ. Fast algorithm for detecting community structure in networks. Phys Rev E Stat Phys Plasmas Fluids Relat Interdiscip Top. 2004. https://doi.org/10.1103/PhysRevE.69.066133.

4. Blondel VD, Guillaume JL, Lambiotte R, Lefebvre E. Fast unfolding of communities in large networks. J Stat Mech Theory Exp. 2008. https://doi.org/10.1088/1742-5468/2008/10/P10008.

5. Pizzuti C. GA-Net: a genetic algorithm for community detection in social networks. In: Lect. Notes Comput. Sci. (Including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics); 2008. https://doi.org/10.1007/978-3-540-87700-4_107.

6. Gong M, Fu B, Jiao L, Du H. Memetic algorithm for community detection in networks. Phys Rev E. 2011;84:056101. https://doi.org/10.1103/PhysRevE.84.056101.

7. Li Z, Zhang S, Wang RS, Zhang XS, Chen L. Quantitative function for community detection. Phys Rev E Stat Nonlinear Soft Matter Phys. 2008. https://doi.org/10.1103/PhysRevE.77.036109.

8. Shang R, Bai J, Jiao L, Jin C. Community detection based on modularity and an improved genetic algorithm. Phys A Stat Mech Appl. 2013. https://doi.org/10.1016/j.physa.2012.11.003.

9. Pizzuti C. A multiobjective genetic algorithm to find communities in complex networks. IEEE Trans Evol Comput. 2012;16:418–30. https://doi.org/10.1109/TEVC.2011.2161090.

10. Gong M, Cai Q, Chen X, Ma L. Complex network clustering by multiobjective discrete particle swarm optimization based on decomposition. IEEE Trans Evol Comput. 2014;18:82–97. https://doi.org/10.1109/TEVC.2013.2260862.

11. Abdollahpouri A, Rahimi S, Majd SM, Salavati C. A modified particle swarm optimization algorithm for community detection in complex networks. Lect Notes Comput Sci. 2018;11015:11–27. https://doi.org/10.1007/978-3-319-99740-7_2.

12. Yuanyyan M, Xiyu L. Quantum inspired evolutionary algorithm for community detection in complex networks. Phys Lett A. 2018;382:2305–12. https://doi.org/10.1016/j.physleta.2018.05.044.

13. Tahmasebi S, Moradi P, Ghodsi S, Abdollahpouri A. An ideal point based many-objective optimization for community detection of complex networks. Inf Sci. 2019. https://doi.org/10.1016/j.ins.2019.06.010.

14. Zhang Q, Li H. MOEA/D: a multiobjective evolutionary algorithm based on decomposition. IEEE Trans Evol Comput. 2007. https://doi.org/10.1109/TEVC.2007.892759.

15. Deb K (2001) Multi-objective optimization using evolutionary Algorithms—Kalyanmoy Deb. https://doi.org/10.1109/TEVC.2002.804322.

16. Wei Y-C, Cheng C-K. Ratio cut partitioning for hierarchical designs. IEEE Trans Comput Des Integr Circuits Syst. 1991;10:911–21. https://doi.org/10.1109/43.87601.

17. Lancichinetti A, Fortunato S, Kertész J. Detecting the overlapping and hierarchical community structure in complex networks. New J Phys. 2009;11:033015. https://doi.org/10.1088/1367-2630/11/3/033015.

18. Newman MEJ, Girvan M. Finding and evaluating community structure in networks. Phys Rev E. 2004;69:026113. https://doi.org/10.1103/PhysRevE.69.026113.

19. Handl J, Knowles J. An evolutionary approach to multiobjective clustering. IEEE Trans Evol Comput. 2007;11:56–76. https://doi.org/10.1109/TEVC.2006.877146.
20. Pizzuti C. Evolutionary computation for community detection in networks: a review. IEEE Trans Evol Comput. 2018;22:464–83. https://doi.org/10.1109/TEVC.2017.2737600.

21. Danon L, Diaz-Guilera A, Duch J, Arenas A. Comparing community structure identification. J Stat Mech Theory Exp. 2005;09008:219–28. https://doi.org/10.1088/1742-5468/2005/09/P09008.

22. Van Veldhuizen DA, Lamont GB (1998) Multiobjective evolutionary algorithm research: a history and analysis

23. Zitzler E, Thiele L. Multiobjective evolutionary algorithms: a comparative case study and the strength Pareto approach. IEEE Trans Evol Comput DOI. 1999;10(1109/4235):797969.

24. Ishibuchi H, Masuda H, Tanigaki Y, Nojima Y. Difficulties in specifying reference points to calculate the inverted generational distance for many-objective optimization problems. In: IEEE SSCI 2014—2014 IEEE Symp. Ser. Comput. Intell.—MCDM 2014 2014 IEEE Symp. Comput. Intell. Multi-Criteria Decis. Proc.; 2015. https://doi.org/10.1109/MCDM.2014.7007204.

25. Zachary WW. An information flow model for conflict and fission in small groups. J Anthropol Res. 1977;33:452–73. https://doi.org/10.1086/jar.33.4.3629752.

26. Lusseau D, Schneider K, Boisseau OJ, Haase P, Slooten E, Dawson SM. The bottlenose dolphin community of Doubtful Sound features a large proportion of long-lasting associations. Behav Ecol Sociobiol. 2003;54:396–405. https://doi.org/10.1007/s00265-003-0651-y.

27. Girvan M, Newman MEJ. Community structure in social and biological networks. Proc Natl Acad Sci. 2002;99:7821–6. https://doi.org/10.1073/pnas.122653799.

28. Newman MEJ. Modularity and community structure in networks. Proc Natl Acad Sci. 2006;103:8577–82. https://doi.org/10.1073/pnas.0601602103.

Publisher’s Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.