International Journal of Engineering Trends and Technology
ISSN: 2231 – 5381 /doi:10.14445/22315381/IJETT-V70I2P212

Volume 70 Issue 2, 86-102, February, 2022

Original Article

Comprehensive Analysis of Single and Multi-Purpose Function-Based Community Detection over Social Media

Rishank Rathore[1], Ravi Kumar Singh Pippal[2]

[1]*PhD Research Scholar, Department of Computer Science, RKDF University, Bhopal, India*
[2]*Professor, Department of Computer Science, RKDF University, Bhopal, India*

Abstract - *In parallel with the development of the Internet, social networks have become attractive as research topics in many different disciplines, and the most accurate systems are expressed in complex networks. The most common feature of complex networks is their community structure, where the connections within the node groups are more closely related to the rest of the network. Identifying major clusters and community structures allows discovering organizational rules of complex networks such as web charts and biological networks. In general, the communities seem to overlap. Overlap is when an individual belongs to more than one social group and is one of the characteristic features of social networks. In recent years, overlapping community discovery has received much attention in the application areas of social networks. Many methods using different tools and techniques have been proposed to solve the overlapping community discovery problem. This paper gives a comparative analysis over heuristic overlap community detection algorithm over social media and presents a comprehensive analysis of single and multi-purpose functions for community detection.*

Keywords - *Community detection, Heuristic Classification, Optimization Algorithm, Single Function, Multi-Purpose Function, Seed Community*.

I. INTRODUCTION

The rapid developments in technology and communication in the 21st century have made the need to access information and use it most effectively an indispensable necessity for people and an inevitable part of their daily lives. The fastest and most practical way to access information in today's world is undoubtedly the Internet. The Internet is not only a network that connects millions of computers around the world but also an environment that connects millions of people and thousands of social groups, but it is constantly growing and developing. Social media is one of the most popular internet applications, which is rapidly advancing to become one of the most important communication tools today; as the frequency of use of the Internet increases, the rate of accessing social media increases within this frequency. It is thought that almost an essential part of internet usage will be provided by social media soon. Social media applications no longer only offer communication but also intend to give nearly every individual's need by using many topics such as games, obtaining information, and searching. Thus, people who find almost everything they are looking for on social media will not need another tool. With the development of computers, network analysis has allowed researchers to obtain and analyse data on large networks as social media.

Complex network analysis is used in many areas today, and the examination of individual and social group structures and behaviours in the press (separation, clustering, determination of relationships), electronic commerce and online advertising (customer profile creation and trend analysis, personalised advertising and offering), analysis of physical structures (transportation, installation, infrastructure) and analysis of large data sets (media tracking, academic publication analysis, genetic research) [1]. The most current issue regarding network analysis is the discovery of communities and communities in networks. Identifying their communities in networks is applied in many fields such as biology, social sciences, physics, chemistry, engineering. For example; With the discovery of biological communities, functional units of proteins can be found, or the functions of proteins can be predicted [2]. In sociology, community structure is an essential topological feature, given the vaccination interventions for infectious diseases in associated networks and understanding the spread of viruses in social networks [3].

An essential point in community discovery is that nodes can be classified by looking at their structures within the group they are located in, and groups can be revealed. If a node-set in a social network structure contains more links than the number of connections outside, this node-set is considered a community. Communities, also called clusters or modules, are groups of nodes that generally share standard features and perform similar tasks in networks [4]. A grid showing the communities schematically is given in Figure 1.

The separation of links between groups is at the heart of most of the approaches used in community detection. The biggest problem encountered in actual network structures is the overlapping situation, called the possibility of nodes belonging to more than one group. However, many algorithms usually include nodes in a group due to the complexity of the operations, ignoring the overlap [5]. This grouping does not allow obtaining accurate information about the structure of complex networks [6].

There are many algorithms for discovering overlapping communities in complex networks. CPM is the most widely used algorithm. However, CPM is not flexible enough for real networks. CPM finds significant clicks when the network is very dense but not when the network is sparse. Therefore, CPM depends heavily on the capabilities of the network.

GA-Net + [7] uses a Genetic Algorithm (GA) to adopt overlapping communities. The method converts the node chart to a line chart. The nodes in the line chart show the edges in the node chart, while the edges show the neighbourhood relationships of the edges in the node chart[8]. The line chart is then given as an introduction to the genetic algorithm, and at each step, the line chart is converted into a node chart to obtain fit [9].

Other common studies for community discovery are; detection of network communities [10], detection of overlapping ensembles in networks [11] and an algorithm for quickly determining overlapping ensembles [12].

Another method that can be used for community discovery in social networks is optimisation algorithms

Optimisation is the process of obtaining the best solution to a problem. Meta-heuristic optimisation algorithms are a decision mechanism that works on heuristic optimisation algorithms [13], which are frequently used in daily life. For example; It is an intuitive approach to act based on a sense of direction while moving from one place to another and not knowing where the road will lead and to make choices at crossroads. When 3 heuristic algorithms are advantageous from different angles for a problem, the structure that decides which methods to choose is meta-heuristic algorithms.

This paper gives a birds-eye over heuristic community detection algorithm over social media. The rest of the paper is organised as follows: Section.2 present an overview of the social media community and algorithm

; Section.3 present Single and Multi-purpose function based metaheuristics community detection Methods; Section.4 covers the comparative analysis community detection algorithm over six different data set, and finally, Sect.5 concludes the paper and outlines the founding and future work.

II. COMMUNITY ON SOCIAL NETWORKS

Most of the complex networks are represented as networks. For example; The World Wide Web (WWW) is a network of interconnected web pages; social networks are networks where people are represented by nodes and connections between them by edges. Similarly, biological networks are networks in which nodes express biochemical molecules, and boundaries define the links between them [14]. In recent years, most of the work has focused on understanding the effects of network topology on system behaviour and dynamics and network organization and development. Finding community structures is another crucial step in understanding complex network structures[15].

Communities in a network are defined as groups of nodes where connections between groups are infrequent, and links within groups are frequent. In another definition, the community is the association of individuals who are in communication frequently. For this reason, communities are primarily groups of nodes that share common characteristics and play similar roles in interaction [16].

Communities within network structures tell us about individuals' common interests, subjects of study, tendencies, similarities, etc. offer a concrete idea. In real networks, the network structure is not homogeneous. The systems that concentrate and cluster in a particular area and we call ensembles are probably clusters of nodes that share the same feature and have a similar role [4].

Communities have many concrete application areas. For example; Clustering of web clients with similar interests or geographically close to each other provides an increase in service performance on the WWW by assigning the same servers to each client cluster. Identifying the community of customers with similar interests enables an effective advisory system to be established between the customer and the seller in online shopping systems[17].

Planning hierarchical organizations in complex real-world networks is possible by identifying communities. Real networks often contain communities of small communities. The human body is the epitome of hierarchical organization. The body consists of organs, tissues from organs, and tissues are made up of cells. Another example of a hierarchical structure is business firms. Business firms consisting of mid-level workgroups can be thought of as a pyramid expanding from workers to the organisation's head.

The purpose of exploring communities in networks is to describe modules and their hierarchical organisation using only the coded knowledge of the network topology. The most frequently used definition in community discovery is the assumption that the number of edges within the group should be more than the number of connections to the outside. The "cut-size" parameter defined from this point is called the number of edges that connects this group to the rest of the line. A good assortment is expected to have a low cut-size value.

Another definition, "vertex similarity", is that when nodes are placed on a spaceplane, the distance between them is considered a similarity criterion. Classical grouping methods often make use of this approach. If nodes cannot be placed on a spaceplane, then an adjacency matrix can be used. If their neighbours are the same, it can be said that they are similar even if they are not neighbours. In addition, similarities between nodes can be determined by measuring the number of independent paths between two nodes, the distance of the shortest route or random walk [18].

While the first studies on discovering community structures suggested that a node can belong to only one community, networks consist of different relationships in which nodes belong to more than one community, and this structure is defined as overlap. For example, there may be family, friendship, and colleague relations between two people in human relations. Therefore, the discovery of overlapping communities is an essential issue for the analysis of real social networks. A network of 3 different communities is given in Figure 2. Within the network, 4 nodes are included in more than one community, indicating the overlapping community structure in networks.

The community and modular structure are considered an essential feature of real-world social networks as it is used to calculate the functionality of the systems. However, many uncertainties about identifying communities, effective and efficient community discovery methods have been developed

A. Traditional Methods

a) Graph Partitioning

It is the division of nodes into k groups of predetermined numbers such that the number of edges between groups is minimal. However, when the number of groups present in social network structures is not known in advance, it is not a suitable approach for social network analysis. Its most essential algorithms are Iterative Bi-sectioning [19] Max-Flow Min-Cut Theorem. Figure 3 shows a dividing line for k = 2 where the number of edges between groups is minimum.

b) Hierarchical Grouping

Social networks generally contain groups that are intertwined in a hierarchical structure. It is a method based on combining similar nodes and making groups and splitting groups by deleting low-affinity nodes. Results will vary depending on the similarity criterion to be determined [4].

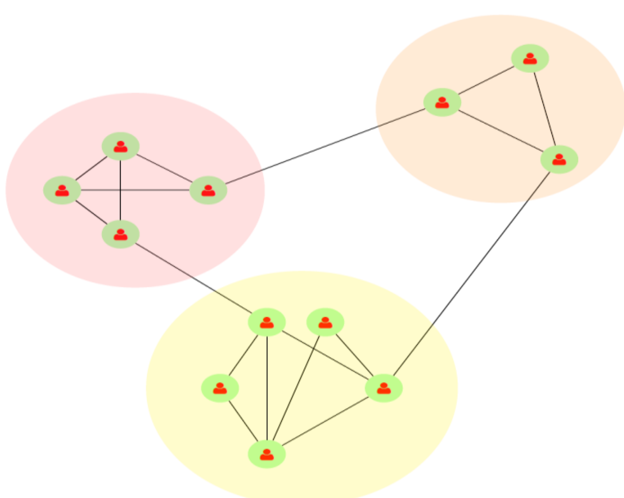

Fig. 1 Network structure consisting of three communities

c) Segmentation Clustering

Here, the group number k is predetermined, and each node is treated as a point in space. According to a given function, the aim is to divide points into k groups according to their distance from the centre, depending on the distance between them. The most used parts are Minimum k-clustering, k-centre, k-median, and k-means [20]. Here, too, the disadvantage is the need to know the number of groups in advance [4].

d) Spectral Clustering

Spectral clustering involves many techniques and methods that partition into sets using an eigenvector such as S or other matrices produced from it[21]. In this method, first, the eigenvectors of the similarity matrix are taken and then divided into groups with a function such as k-means [20]. The most used matrix is the Laplace matrix. Thanks to this approach, it can be learned how many groups are in the line from the components of the eigenvectors.

B. Segmentation Algorithms

It is a method that aims to find and delete the edges that connect groups in the graph and thus discriminate and reveal the groups. The critical point is how to determine the edges connecting these groups. The most popular algorithm is the Girvan-Newman algorithm [22]. Here, edges are selected based on a criterion called edge centrality. The centrality value is calculated for all edges. The edges with the highest centrality value are deleted. The first step is performed again, and the process continues in this way by deleting the edge with the highest value.

Apart from the edge centrality criteria, edge betweenness, random walk edge betweenness and current flow betweenness are also used [4].

C. Modularity Based Methods

Modularity is the most widely known and used quality function in graph analysis. Although not fully proven, a high modularity value is considered to indicate good groups [4]. If a graph has a higher modularity value than a random line of the same size and degree, that line is considered a group structure. However, a high modularity value may not always mean that there is a group structure. Although there is no group structure in some random graphs, high modularity values can be encountered.

Improving the modularity function is an NP-Complete problem, so it has no solution in linear time. However, algorithms that achieve successful results with various convergences have been developed [23, 24]. The change that maximizes the quality function is made from the set of changes made on the chart. This can be a merge, split, or edge deletion.

D. Dynamic Algorithms

Among dynamic algorithms, the most commonly used method to explore communities is the random walker model. In this method, if the connections in the graph are of high density, the random walker stays in the community for a long time; according to the logic in question, the chart consists of strong communities [4].

E. Other Methods

Apart from those mentioned above and frequently used methods, methods based on statistical inference (Bayes etc.) [25, 26], methods that tag nodes and take the tag shared most by their neighbours in each iteration and separate groups in this way [4], click filtering methods [27], strategies to combat overlap and multi-resolution methods are available [28].

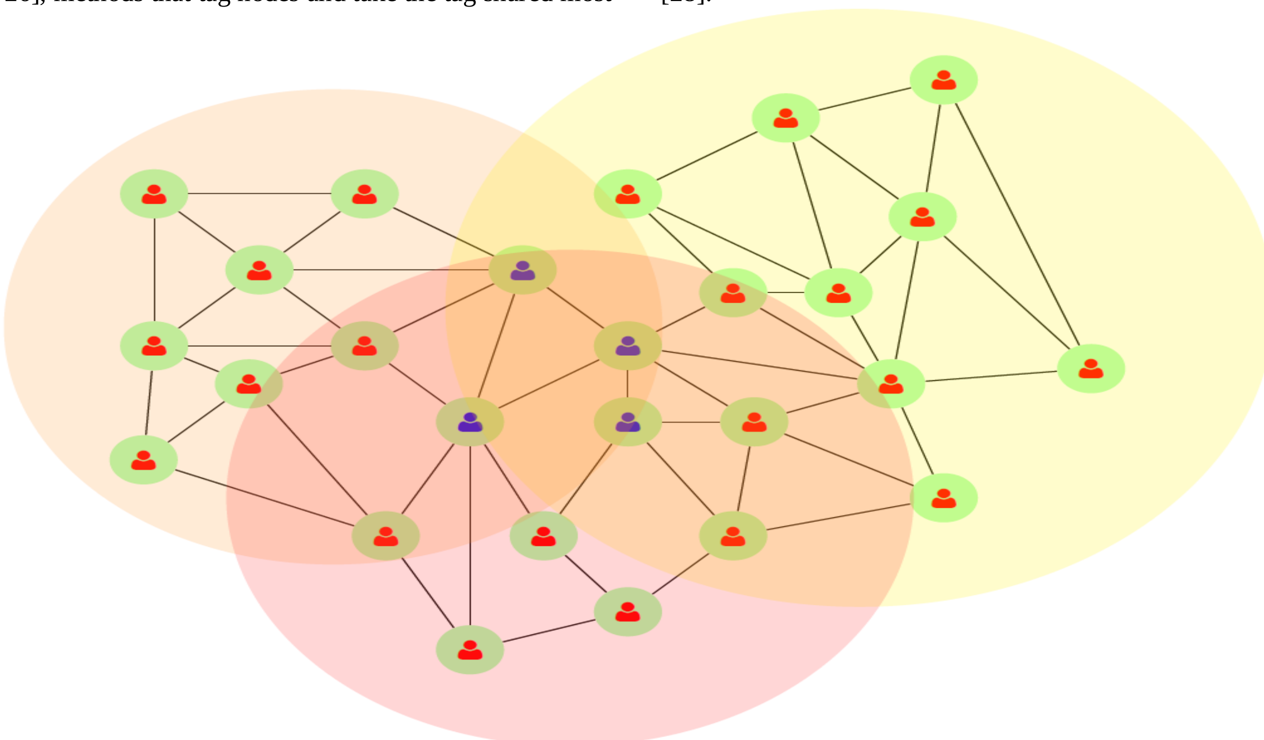

Fig. 2 Overlapping community structure

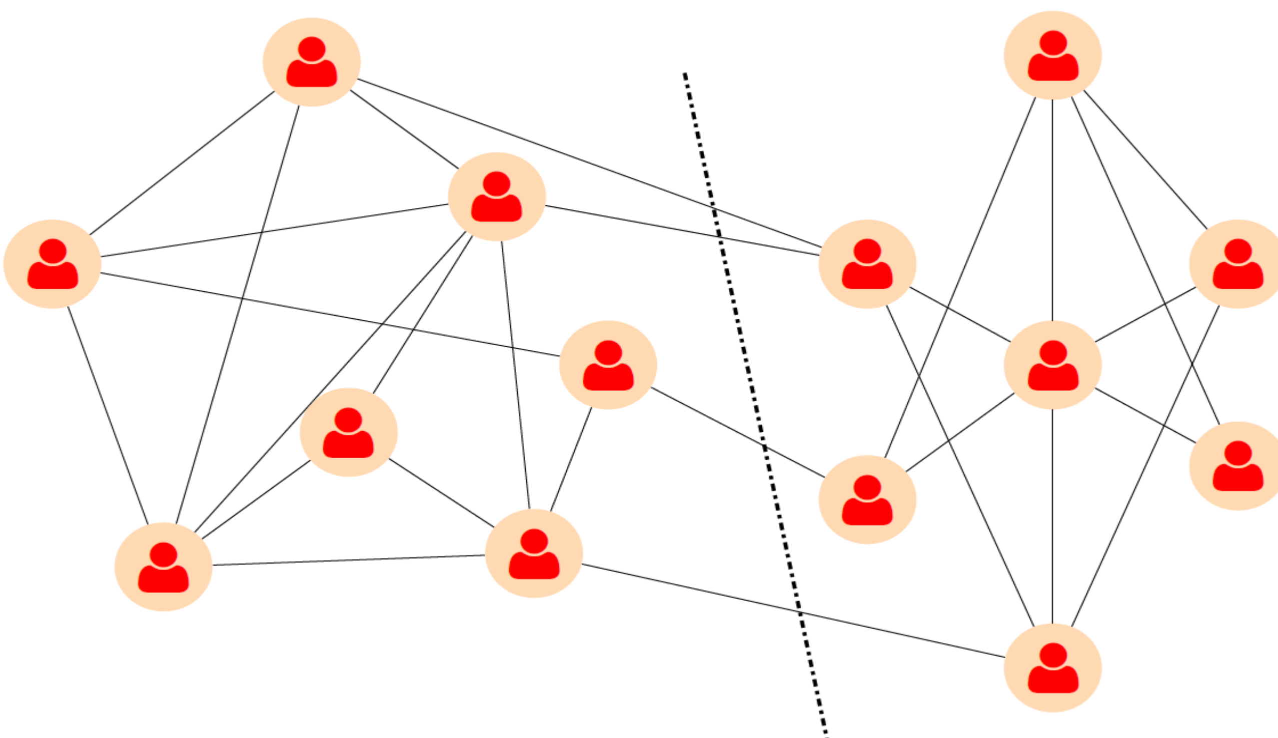

Fig. 3 Graph Partitioning over Social Media Network

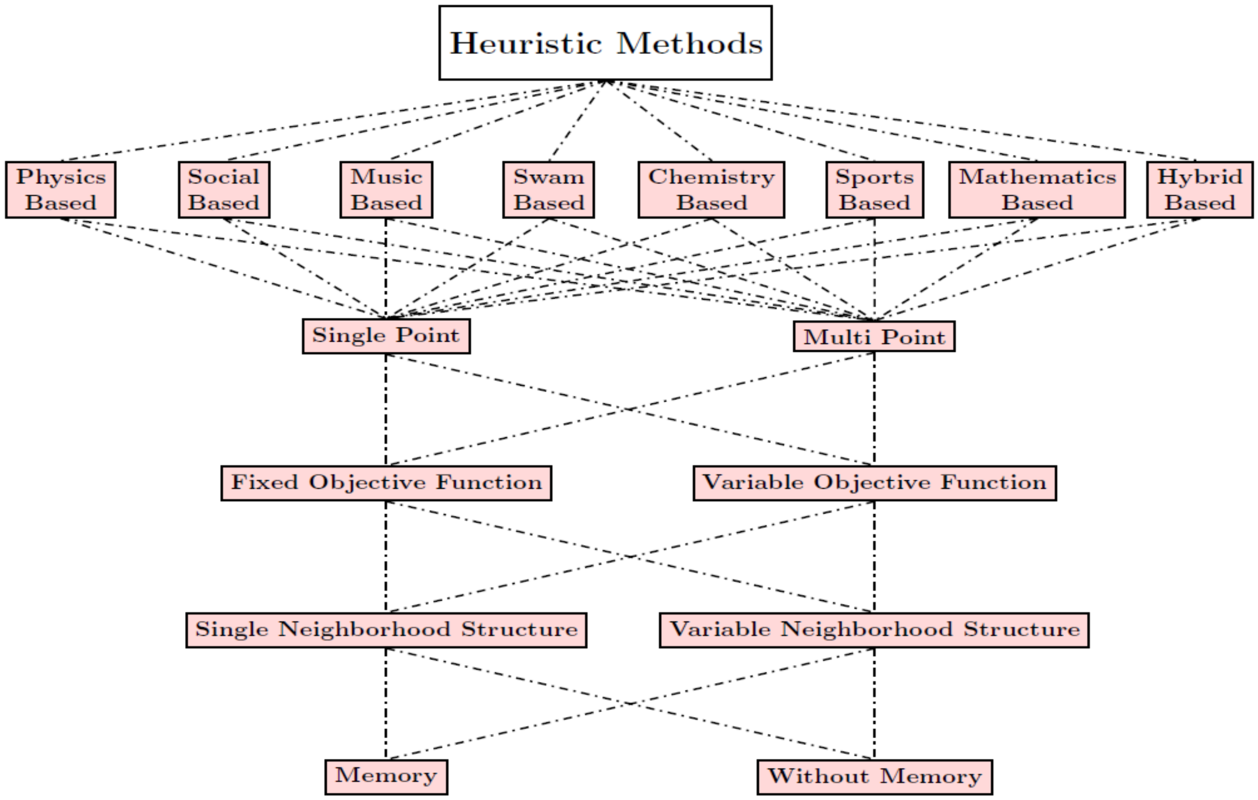

Fig. 4 Meta-Heuristic Methods

III. METHODOLOGICAL OPTIMIZATION METHODS

Any problem involving finding unknown parameter values satisfying certain limitations can be called an optimization problem. Optimization means optimization. It is the job of obtaining the best solution among all solutions for a problem under a given condition. That is, heuristic algorithms can converge but cannot guarantee the exact answer. This situation provides a solution close to the definitive solution [29].

The reason why heuristic algorithms are needed is as follows:

- The optimization problem may have a structure in which finding the exact solution cannot be defined.
- In terms of clarity, heuristic algorithms can be much simpler for the decision-maker.
- Heuristic algorithms can be used for learning purposes and as part of finding the exact solution.
- In definitions made with mathematical formulas, the most challenging aspects of real-world problems (which goals and limitations should be used, which alternatives to be tested, how to collect problem data) are often neglected. Inaccurate data used in the stage of determining model parameters may cause more significant errors than the sub-optimal solution that the heuristic approach can produce [29].

Meta-heuristic optimization algorithms are a decision mechanism that works on heuristic optimization algorithms, which are frequently used in daily life[30]. Meta-heuristic algorithms use a simple approach as a solution technique for search or optimization problems and are getting stronger and more popular in recent years. The reason for these can be summarized as follows:

a. Simultaneously, they present general solution strategies that can be applied to the problem in the presence of different types of decision variables, objective functions, and constraints. Solution strategies do not depend on the type of objective function and limiters and the type of variables used in modelling the problem.
b. It does not depend on the solution space type, the number of decision variables, and delimiters.
c. It does not require very well-defined mathematical models, which are difficult to set up for the model and purpose function of the system and sometimes cannot be used because of the high cost of solution time.
d. They have good computing power, which means they don't need excessive computing time.

e. They are easy to transform and adapt to.
f. It gives effective results in large scale combinational and nonlinear problems.
g. A solution algorithm to a given problem, as in classical algorithms, does not require some assumptions that can be difficult to validate in adaptation.
h. As with classical algorithms, it does not require changes on the problem of interest. They adapt themselves to solve different kinds of issues.

Because of these advantages, meta-heuristic algorithms are used extensively in many fields such as man agreement science, engineering, and computers, and new versions are recommended.

General-purpose meta-heuristic methods as shown in Figure 4, bio-based (evolutionary algorithms, ant colony algorithm, bee colony algorithm, artificial immune algorithms, firefly algorithm, enzyme algorithm, sapling development algorithm, invasive weed optimization, monkey search algorithm, bacterial bait search algorithm), physics-based (multi-point heat treatment algorithm, electromagnetism algorithm, particle collision algorithm, big bang big crash algorithm), swarm-based (particle swarm optimization, ant colony optimization, bee colony optimization), social-based (multipoint taboo Eight different methods: research algorithm, imperialist competitor algorithm, parliamentary optimization algorithm), music-based (harmony search), sports-based (league championship algorithm), chemistry-based methods (artificial chemical reaction optimization algorithm), mathematics-based (meta-heuristic and base algorithm) are evaluated in the group[31]. There are also hybrid methods combining them.

Although very successful algorithms and techniques have been developed in the literature; It is essential to design, develop, and implement new strategies under the philosophy of continuous improvement in the scientific field and always seek the better. In addition, since the algorithm that gives the best results for all problems has not been designed yet, new meta-heuristic algorithms are constantly proposed. The existing ones are offered to work more effectively. With this awareness in recent years, researchers have successfully introduced new meta-heuristic methods to the literature and implemented successful applications.

A. Social Based Meta-heuristic Optimization Algorithms

There are many newly proposed social-based heuristic optimization algorithms in the literature. The most well-known and most applied of these is the tabu search algorithm. More recently, others have been submitted [32].

a) Imperialist Competitor Algorithm

Like similar evolutionary algorithms, the Imperialist Competitor Algorithm (ICA) begins the algorithm by creating an initial population. A few of the top countries in the starting population are chosen to be imperialists, and the remaining individuals become colonies of imperialists. All of the designated territories are distributed among the imperialist states. After the dispersal of the colonies among the imperialist states, the colonies begin to move towards the appropriate imperialists. The power of empires depends on the ability of the imperialist and their territories given to the imperialist. With the race that started between the imperialists, the algorithm process continues. Unable to increase its strength or succeed, the imperialist will be eliminated from the race. During the race, strong empires raise their strength while weak kingdoms diminish and move towards destruction. The race continues until only one empire remains, and as a result of the algorithm, other countries become a colony of the remaining empire. In the ideal world formed at the end of the race, territories and imperialists will have the same position and power [33]. Figure 5 shows the flow chart of the algorithm.

b) Teaching Learning Based Optimization Algorithm

Another recently developed meta-heuristic optimization algorithm is the Teaching Learning Based Optimization (TLBO) Algorithm [34]. TLBO is an algorithm that works according to the effect of a teacher on students in a classroom. The algorithm describes the teaching and learning abilities of teachers and students in a school. Teacher and student are two essential components of this algorithm [35].

The group with students in the algorithm is considered the population, and the different subjects presented to the students are regarded as other design variables of the optimization problem. A student’s result is similar to the fitness value of the optimization problem. The teacher is considered to be the best solution for the whole population. The terms used as design variables are shown as the parameter included in the fitness function of the given optimization problem, and the best solution is the best value of the fitness function. The working process of the TLBO algorithm consists of two situations: The teaching Process and Learning Process[36].

In the Teaching Process, the teacher is generally accepted as the person who shares their knowledge with the students and is very important. The quality of a teacher shows its result on the students. It is observed that there are improvements in the grades and situations of students with good teachers. Therefore, the Teaching Process depends on the relationship between teacher and student. In the Learning Process, the main factor is students [37].

The flow chart given in Figure 6 has been created for the TLBO algorithm to understand the algorithm’s steps better

c) Social-Emotional Optimization Algorithm

Social-Emotional Optimization Algorithm (SEOA) is a new social-based optimization technique that simulates human behaviour [38]. The word social is associated with the human community. People living in the community try to increase their social status.

The operation steps of SEOA are given in Algorithm1.

Stepwise explanation of Social-Emotional Optimization Algorithm

1. Start
2. All individuals are generated sequentially, and their initial positions are randomly allocated to the problem space.
3. The fitness value of each individual is calculated according to the objective function.
4. j. For the individual, behavioural movements are determined according to his emotional index.
5. The location is updated for the entire population.
6. The emotional index is determined.
7. If the termination condition is met, the best solution is accepted. If the condition is not met, step 2 is returned.
8. Stop.

In SEOA, each individual represents a virtual person. At each step, individuals determine their behaviour according to the associated emotional index [39]. The emotional index is divided into three low, medium and high. According to the emotional index, a behaviour is selected. According to the chosen behaviour, the status value is recycled from society depending on whether the desired behaviour is correct. If this choice increases the social status value, the emotional index of the individual increases. Otherwise, the emotional index decreases to decrease the social status value [40].

d) Brainstorming Optimization

Brainstorming is a widely used tool to boost creativity in widely accepted organizations, such as facilitating creative thinking. Brainstorming was first developed in 1939 by Osborn in the advertising firm. In late 1957, he systematized this problem-solving method in Applied Imagination [41, 42]. After that, brainstorming aroused great interest in both academia and industry all over the world. The brainstorming process brings together people of different ethnicities who will collaborate and interact to generate great ideas for a problem solution. The BFOA process steps developed inspired by brainstorming are given in Algorithm 2

Stepwise explanation of Brain Storming Optimization Algorithm:

1. Start
2. n potential solutions (individuals) are generated.
3. n individuals are divided into m clusters.
4. N individuals are evaluated.
5. Individuals in each cluster are ranked, and the best individual is determined as the centre of the cluster.
6. Random a value between 0 and 1 is generated.
   (a). If the value of a produced is less than the predetermined value of P5a
      (i). Choose a cluster centre at random.
      (ii). Generate a random individual to replace with the chosen cluster centre.
7. Produce new individuals.
   (a). A random value between 0 and 1 is generated.
   (b). If the produced value is less than P6b,
      (i). With probability P6i, choose a random set a.
      (ii). Generate a random value of between 0 and 1.
      (iii). If the value is less than the preset value of P6b iii, 1) Select cluster centre and add random value to generate new individuals'. Otherwise, choose a random individual from the cluster and add the randomly generated value to this individual to obtain new individuals.
   (c). Otherwise, randomly select two clusters to generate new individuals.
      (i). Generate a random value.
      (ii). If the generated value is less than the predetermined probability of P6c, select and merge two cluster centres and add the randomly generated value to create new individuals.
      (iii). Otherwise, two individuals are randomly selected to combine from each selected cluster, and the generated value is added to produce new individuals.
8. If n new individuals are generated, go to step 9, go to step 7.
9. The end of the predetermined maximum number of iterations has been reached; otherwise, go to step 3.
10. Stop.

e) Group Leaders Optimization Algorithm

Group Leaders Optimization Algorithm (GLOA) is an evolutionary algorithm developed inspired by the influence of leaders in social groups. The problem space is divided into different groups, and each group's leader is formed [43]. Members of each group need not have similar characters; they can be randomly generated. The best of each group is chosen as the leader. Members of each group try to resemble their leaders in each iteration. In this way, the algorithm creates a solution space between the leader and group members. After some steps, it was observed that group members resemble leaders. To increase the diversity within the group, one of the members is chosen randomly. Some of its variables are replaced by the variables of the other group members. In addition, a crossover operator helps the group to reach the local minimum, and the solution space can be searched again to increase diversity [44]. The algorithm steps in which n groups of P members are formed and group leaders are determined according to their suitability values are given in Figure 7.

f) Hierarchical Social Algorithm

The Hierarchical Social Algorithm (HSA) is inspired by social behaviours observed in various human organizations or biological systems. This meta-heuristic approach has been successfully applied to several problems with unlimited resources, such as DFG timing [45] and critical circuit computation. The basic idea of HSA is the simultaneous optimization of the set of suitable solutions. Each group of society contains a viable solution, and these groups are initially randomly distributed to produce separate solution areas. Using development strategies, each group works to increase their goal function or compete with their neighbours. In this case, a better solution is obtained through relevant social competition and cooperation 20. Thus, the objective solution is optimized. The process ends with a single group containing the best solution found [1].

g) Human Group Formation Algorithm

Human Group Formation Algorithm (HGFA) is an up-to-date social-based meta-heuristic optimization algorithm inspired by the behaviour of in-group members who try to unite with their groups as much as possible, as well as out-group members who are trying to social protection with out-group members[46].

Sociologists have defined the in-group and out-group situations to define the human social category. In-group members are individuals who are accepted by the group to which they belong to the group. When people are described as group members, they adopt their group and perceive them as different from other groups. They consider their group superior to other groups. For this reason, group members try to unite their groups as much as possible even when they are away from the group [47]. It shows how the concepts mentioned in Figure 8 are transformed into applications.

h) Social Based Algorithm

A Social Based Algorithm (SBA) is a new algorithm that combines Evolutionary Algorithm and a socio-political process based on an Imperialist Competitor Algorithm.

People live in different types of communities: Monarchy, Republic, Autocracy, and Multinational. The leadership style in each community is also different. This approach tries to catch a few people in the community development characteristic [48]. Algorithm 3 shows the process steps of the SBA.

Algorithm 3: Stepwise explanation of Social Based Algorithm:

1. Start
2. Loading the parameters
3. follow
    - (a). Defining the optimisation problem,
    - (b). Generating random individuals
    - (c). Random selection of some influential people as leaders,
    - (d). Randomly positioning the remaining individuals in different regions,
    - (e). Starting empires with the imperialist cost function T.Pci,
    - (f). The election of compelling leaders as empires,
4. Ten loops Nd=Nd+1
5. i= 1, 2,..., N
    - (a). Selection
    - (b). Cross
    - (c). Mutation
    - (d). Replacement
6. i=1, 2, ..., N
    - (a). The policy of human assimilation: the relocation of the leaders of each group to their empire,
        - (i). $x ~ U$ (0, internal assimilation x d)
        - (ii). $d:$ distance between leader and imperialist
    - (b). people's revolution
    - (c). The assimilation policy of countries: the leaders of each group move into their empire, and the people of each country move just like their leaders
        - (i). $x ~ U$ (0, coefficient external assimilation x d)
        - (ii). $d:$ distance between leader and imperialist
    - (d). The revolution of the countries
    - (e). Changing the location
    - (f). imperialist race; picking the weak country from the weak empire and giving it to the empire most likely to have it
    - (g). Elimination; powerless principle and elimination of empire
7. Checking the termination condition, repeating steps 4-7 until the termination condition is met.
8. Stop.

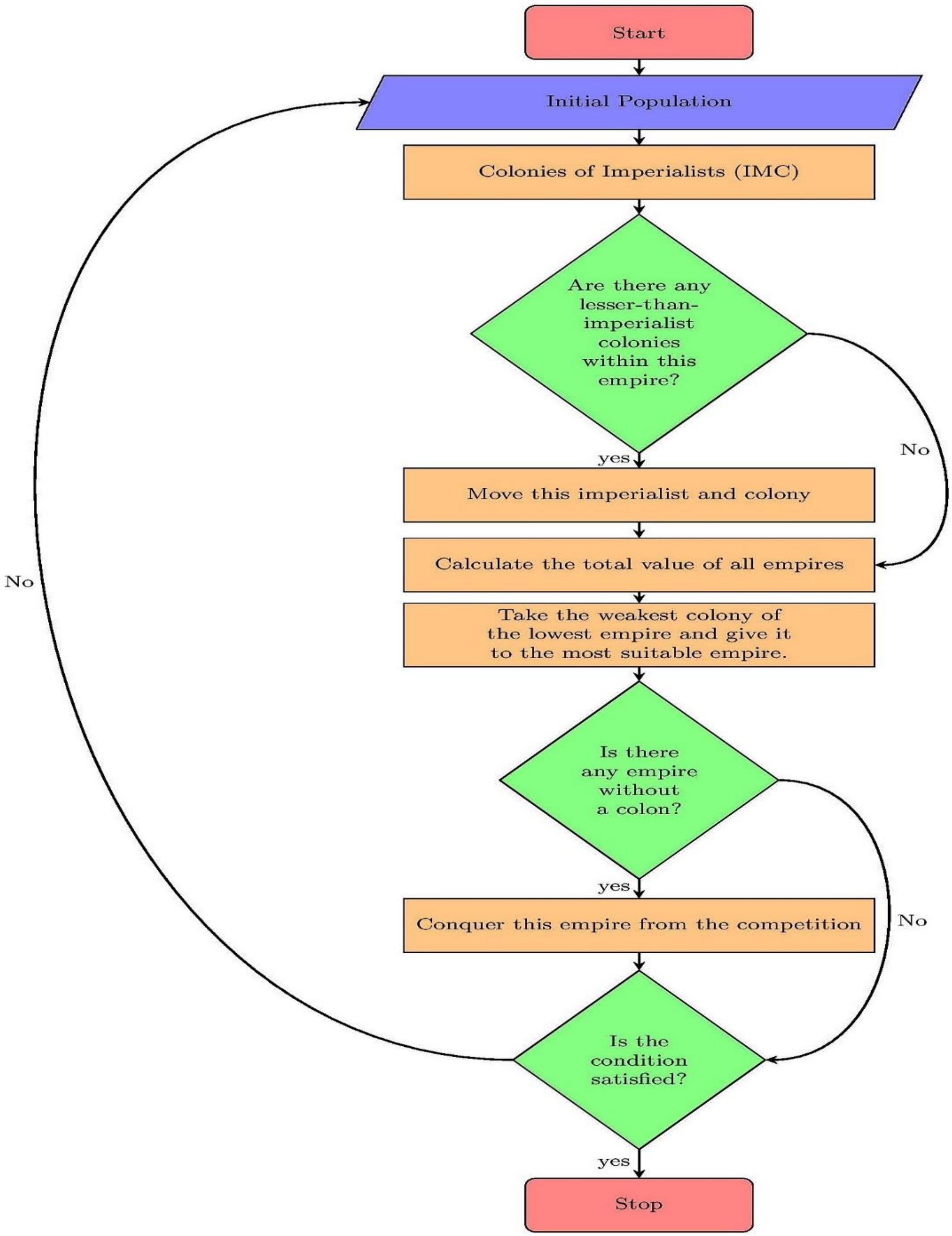

Fig. 5 Flow Chart Of Imperialist Competitive Algorithm

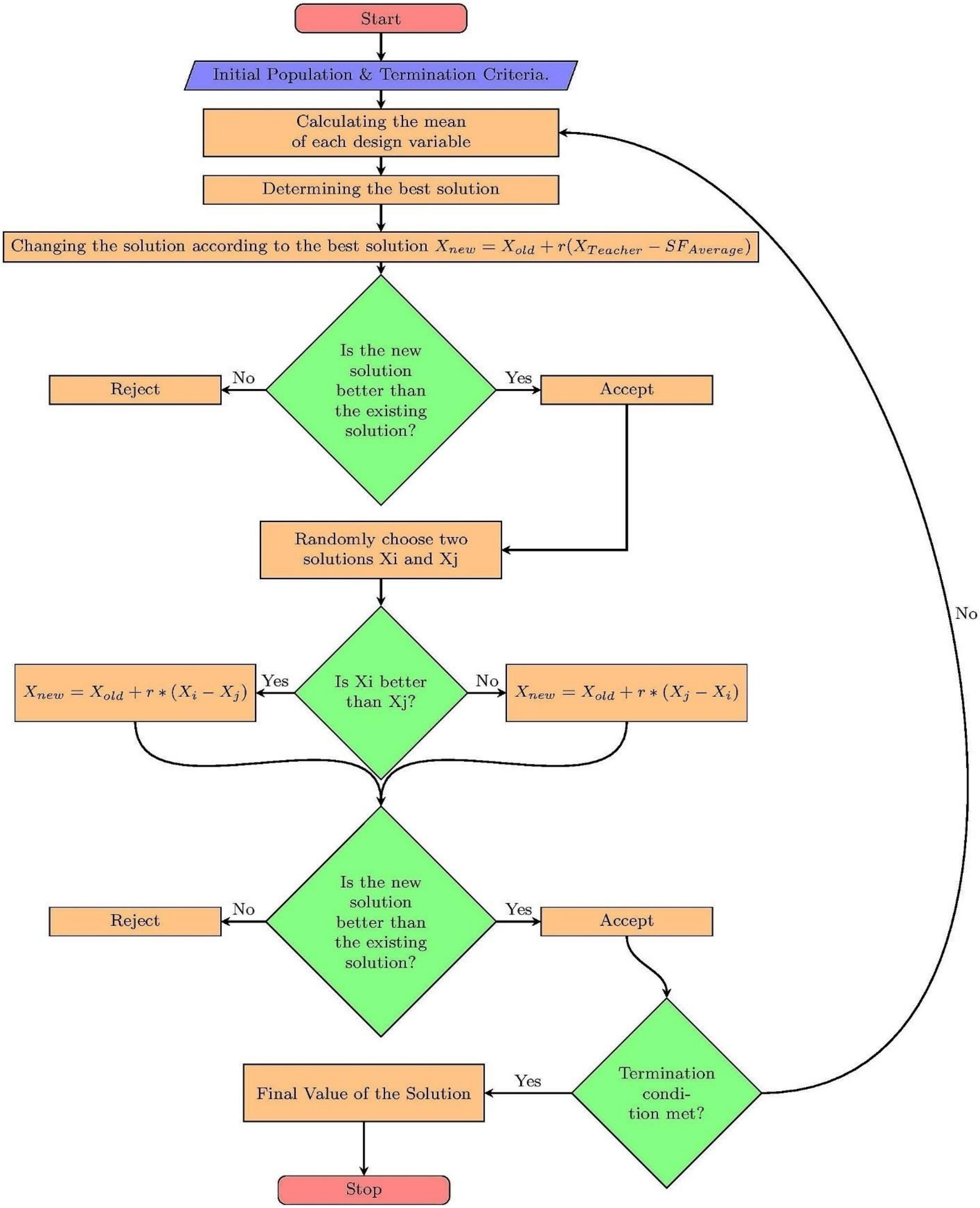

Fig. 6 Flow Chart Of TLBO Algorithm

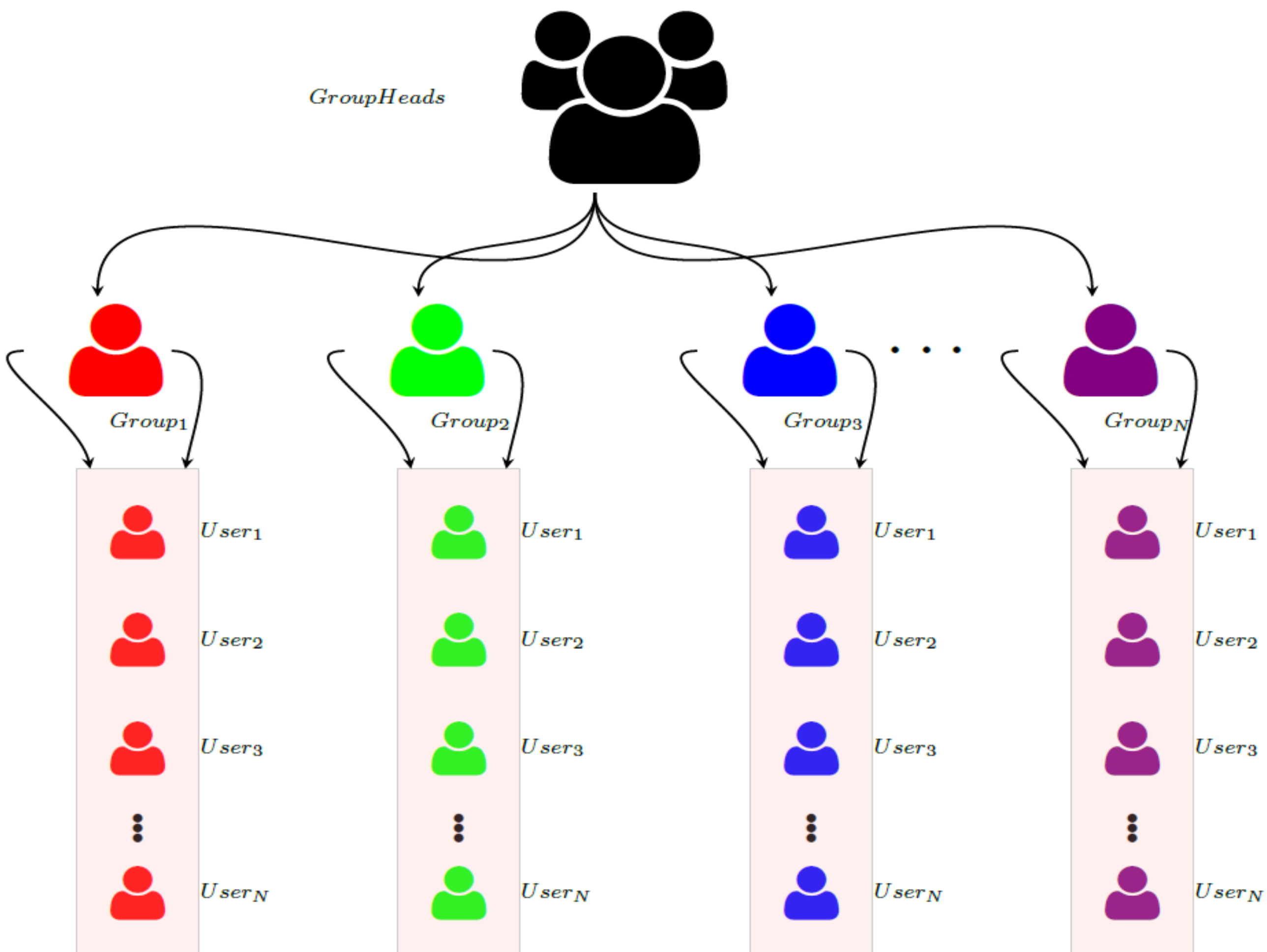

Fig.7 GroupLeadersOptimizationAlgorithm

IV. RESULT ANALYSIS OF BENCHMARK COMMUNITY DETECTION ALGORITHM SOCIAL THEORY

Performance evaluation to detect the impact of single and multi-purpose based heuristic community detection algorithm has been carried out over six different graphical social media data sets, namely Word adjacencies, Zachary karate club [49], Dolphin social network [50], Les Misérables, Books about US politics and American College football [51] over the evaluation parameter modularity and normalized mutual information.

Modularity is network structural measurement that evaluates the strength of subgraph (groups, clusters or communities) in-network for extracting community structure [52]. In a network, groups of a node having higher modularity are relatively dense each other and lead to the appearance of communities in a given network as :

$$M = \frac{1}{2|E|}\sum_{xy}\left[e_{xy} - \frac{\omega_x \omega_y}{2|E|}\right]\delta(c_x, c_y)$$

$$= \sum_{i=1}^{n} f_{ii} - f'^{2}_{i} \dots\dots\dots (1)$$

Where exy represents the edge from node x to node y, Wx representsthe summation of the weights of the edges linked to node x, and cx is the belonging community structure of node x,(cx,cy) is a probabilistic function that equals 1 if both the respective node x and y belong to same community structure, otherwise 0. fii represent the edge in the community i, and F′ is the belonging probability of random edge to the community i that is attached to vertices in the community i. Whereas normalized mutual information is a normalization of intra-community mutual information score to scale the similarity between intracommunity nodes:

$$\mathrm{nmi}(x, c) \begin{cases} 0 & \text{node are totally dissimilar} \\ 1 & \text{node is totally similar} \end{cases} \dots\dots\dots (2)$$

And mutual information is calculated as

$$\mathrm{nmi}(x, c) = \frac{2 * i(x, ci)}{e(x) + e(c)} \dots\dots\dots (3)$$

Table 1. Comparative Analysis of Impact of Social theory on Modularity

Classification Technique	Modularity					
	ZKC	ACF	DCN	BUP	LM	WA
BSO	0.3326	0.6213	0.4844	0.5674	0.5105	0.3611
GLOA	0.3076	0.5947	0.4597	0.5175	0.5098	0.3492
HSA	0.4195	0.5929	0.5981	0.5601	0.5861	0.303
HGFA	0.4305	0.7203	0.5744	0.6768	0.6079	0.3532
SBA	0.4314	0.6014	0.5655	0.6157	0.6071	0.4187
SEOA	0.5313	0.6107	0.661	0.61	0.6121	0.4118
ICA	0.4302	0.7211	0.6179	0.6714	0.6122	0.4192
TLBO	0.5137	0.6204	0.6196	0.7119	0.6217	0.5103

Table 2. Comparative Analysis of Impact of Social theory on Normalized Mutual Information

Classification Technique	Normalized Mutual Information					
	ZKC	ACF	DCN	BUP	LM	WA
BSO	0.7212	0.5025	0.5345	0.4225	0.3107	0.3915
GLOA	0.8611	0.6107	0.6227	0.5209	0.3254	0.3927
HSA	0.8118	0.8628	0.7937	0.5261	0.4255	0.3284
HGFA	0.7153	0.6221	0.5755	0.5253	0.4156	0.4582
SBA	0.8243	0.7431	0.7941	0.5143	0.4173	0.4301
SEOA	0.8102	0.8233	0.7162	0.6955	0.5253	0.5822
ICA	0.8651	0.6324	0.5861	0.5712	0.4128	0.4268

Where exist the class label, c is the community structure, e is the Entropy, and i(x;c) is the information gain for element c_i for class label performance evaluation of benchmark community detection algorithm with and without social theories are shown in tables 1 and 2 as Modularity and Normalized Mutual information, respectively. Both the evaluation parameter is significantly improved after incorporating social theories with community detection algorithm.

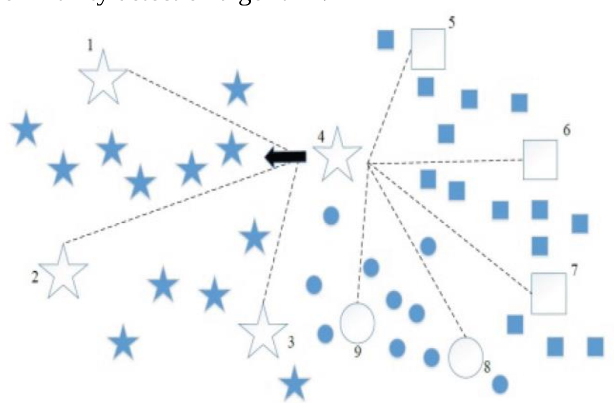

Fig. 8 Human Group Formation Algorithm

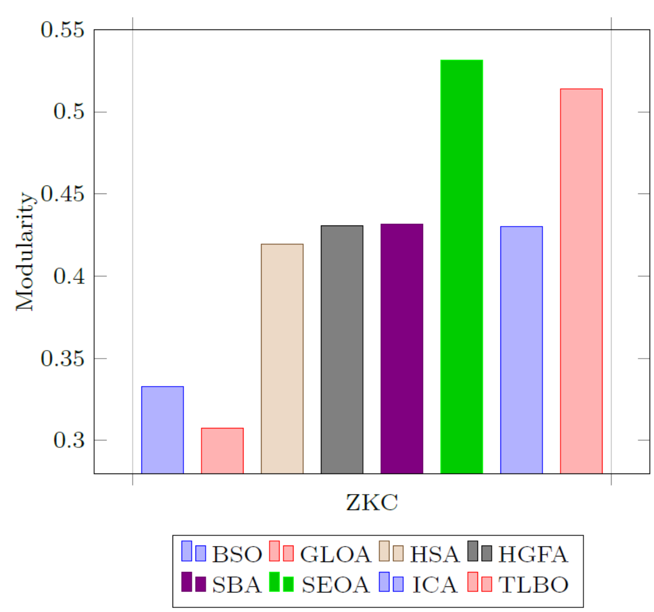

Fig. 9 Modularity of Community Detection Over ZKC Data Set

The community detection algorithm BSO, GLOA, HSA, HGFA, SBA, SEOA, ICA, TLBO gain approximate 33.26%, 30.76%, 41.95%, 43.05%, 43.14%, 53.13%, 43.02%, 51.37% modularity and 72.12%, 86.11%, 81.18%, 71.53%, 82.43%, 81.02%, 86.51%, 86.24% NMI over ZKC datasets respectively, as shown figure 9 and 10. SEOA algorithm leads the modularity, whereas ICA and TLBO algorithm achieves the highest NMI information.

Whereas over AFC dataset, community detection algorithm BSO, GLOA, HSA, HGFA, SBA, SEOA, ICA, TLBO gain approximate 62.13%, 59.47%, 59.29%, 72.03%, 60.14%, 61.07%, 72.11%, 62.04% modularity and 50.25%, 61.07%, 86.28%, 62.21%, 74.31%, 82.33%, 63.24%, 78.23% NMI respectively, as shown figure11 and 12. ICA algorithm leads the modularity, whereas SEOA and HAS algorithm achieves the highest NMI information.

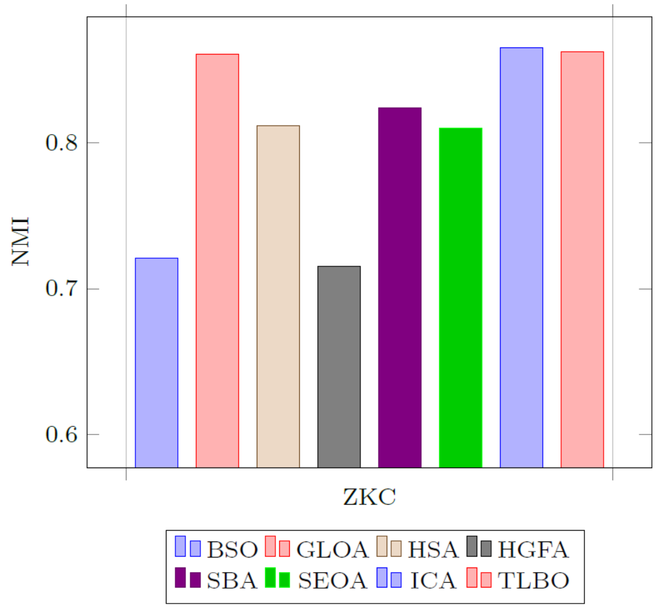

Fig. 10 Normalized Mutual Information of Community Detection over ZKC Data Set

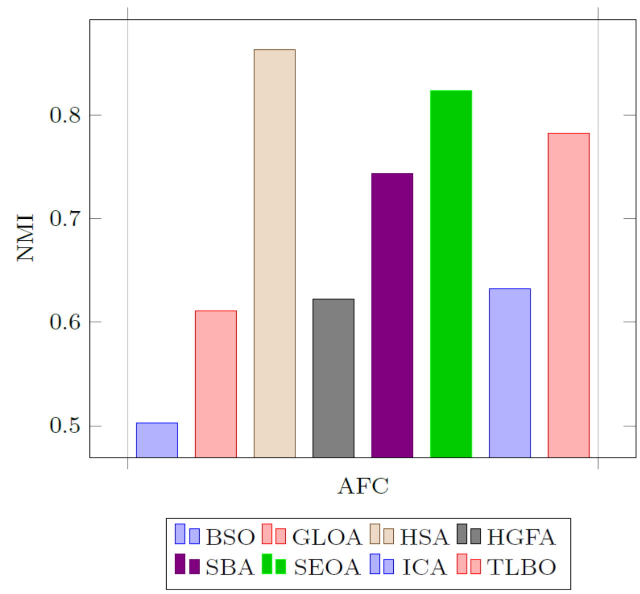

Fig. 12 Normalized Mutual Information of Community Detection Over AFC Data Set

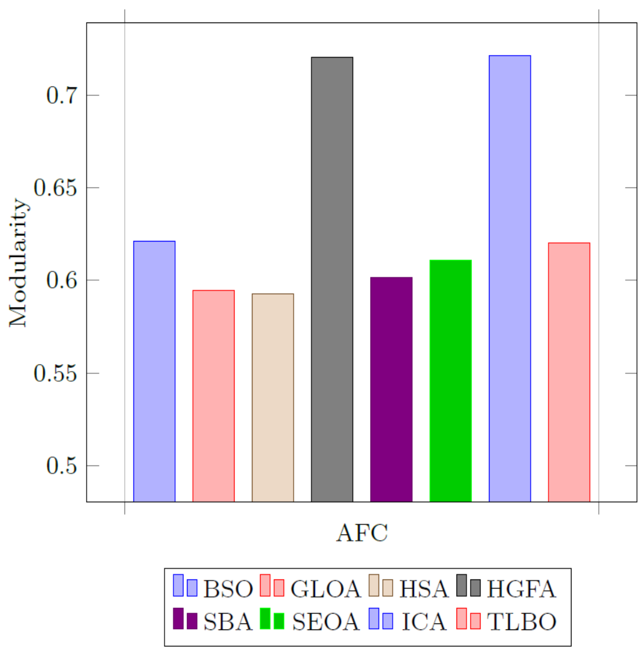

Fig. 11 Modularity of Community Detection Over AFC Data Set

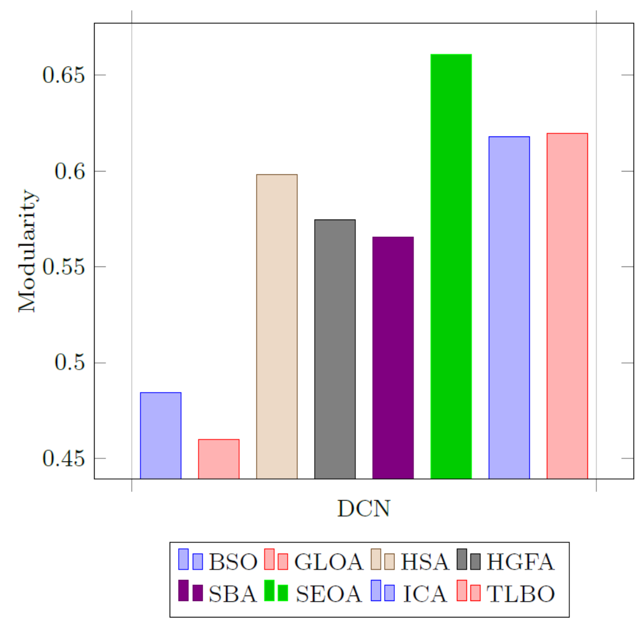

Fig. 13 Modularity of Community Detection Over DCN Data Set

Whereas over BUP dataset, community detection algorithm BSO, GLOA, HSA, HGFA, SBA, SEOA, ICA, TLBO gain approximate 56.74%, 51.75%, 56.01%, 67.68%, 61.57%, 61.51%, 67.14%, 71.19% modularity and 42.25%, 52.09%, 52.61%, 52.53%, 51.43%, 69.55%, 57.12%, 52.51% NMI respectively, as shown figure 15 and 16. TLBO algorithm leads the modularity, whereas SEOA algorithm achieves the highest NMI information.

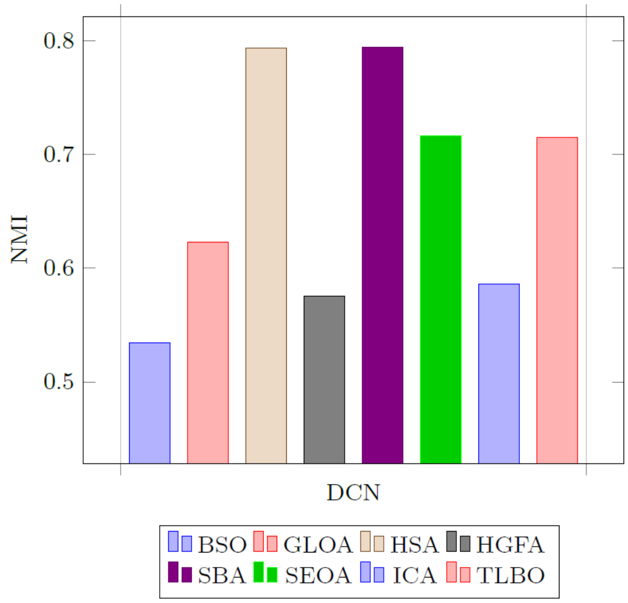

Fig. 14 Normalized Mutual Information of Community Detection over DCN Data Set

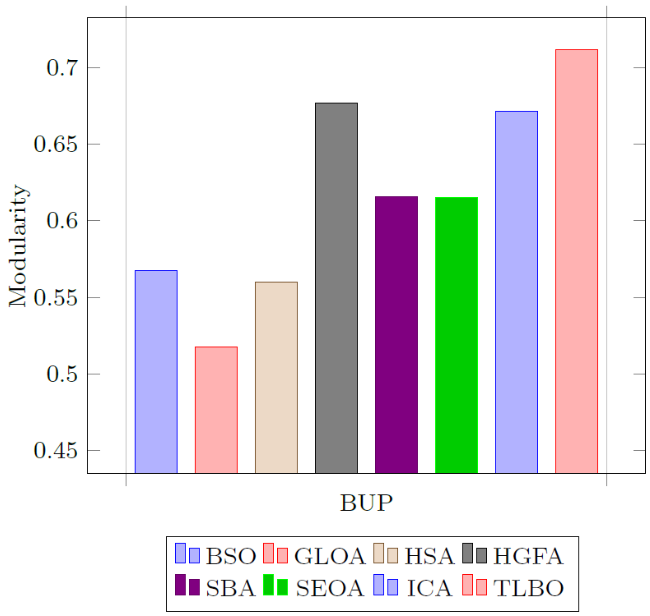

Fig. 15 Modularity of Community Detection Over BUP Data Set

Whereas over LM dataset, community detection algorithm BSO, GLOA, HSA, HGFA, SBA, SEOA, ICA, TLBO gain approximate 51.05%, 50.98%, 58.61%, 60.79%, 60.71%, 61.21%, 61.22%, 62.17% modularity and 42.25%, 52.09%, 52.61%, 52.53%, 51.43%, 69.55%, 57.12%, 52.51% NMI respectively, as shown figure17 and 18. TLBO algorithm leads the modularity, whereas SEOA algorithm achieves the highest NMI information.

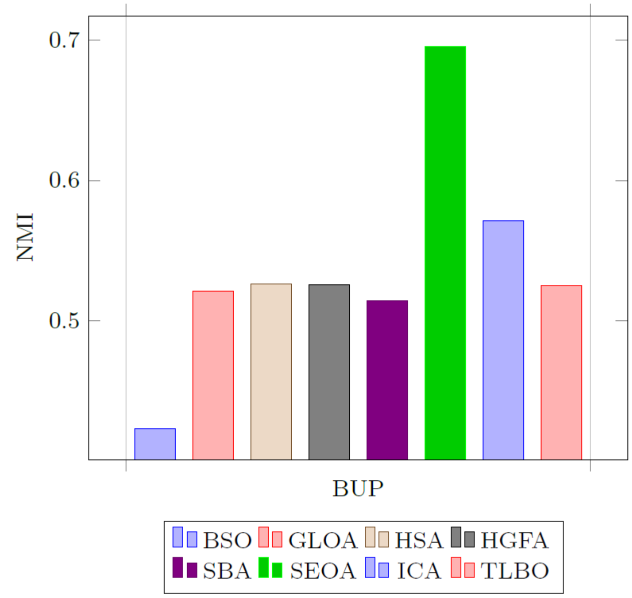

Fig. 16 Normalized Mutual Information of Community Detection over BUP Data Set

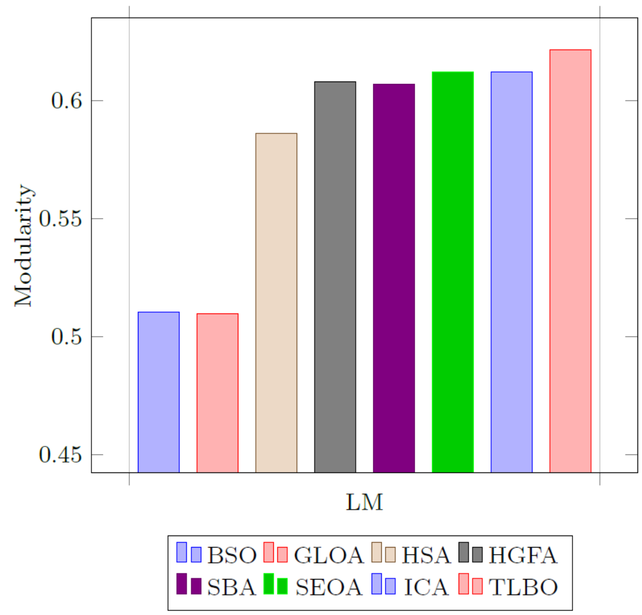

Fig. 17 Modularity of Community Detection Over LM Data Set

Whereas over WA dataset, community detection algorithm BSO, GLOA, HSA, HGFA, SBA, SEOA, ICA, TLBO gain approximate 36.11%, 34.92%, 30.30%, 35.32%, 41.87%, 41.18%, 41.92%, 51.03% modularity and 39.15%, 39.27%, 32.84%, 45.82%, 43.01%, 58.22%, 42.68%, 51.22% NMI respectively, as shown figure19 and 20. TLBO algorithm leads the modularity, whereas SEOA algorithm achieves the highest NMI information.

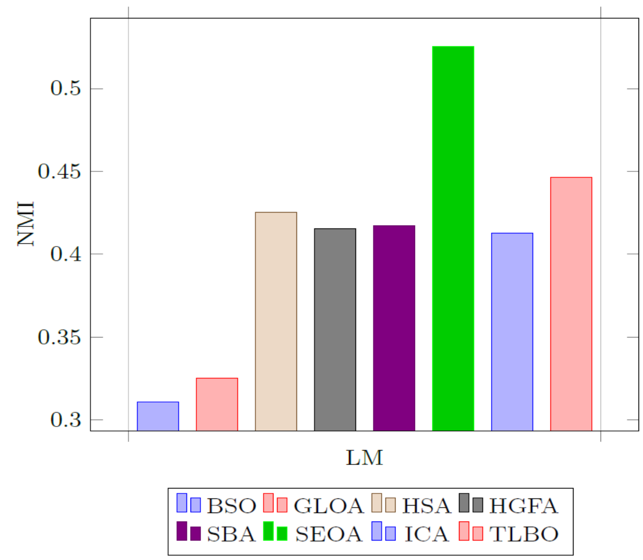

Fig. 18 Normalized Mutual Information of Community Detection over LM Data Set

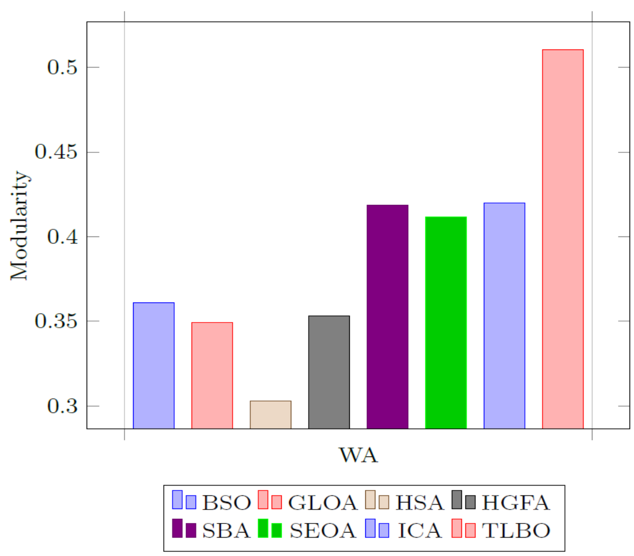

Fig. 19 Modularity of Community Detection over WA Data Set

The performance of community detection algorithm BSO, GLOA, HSA, HGFA, SBA, SEOA, ICA, TLBO over social media datasets varies with network density. It achieves a higher performance rate, higher dense ACF network and relatively lower over lightly dense WA dataset. Heuristic overlapping community detection algorithms over six different social media-based datasets. This paper observed that community detection algorithm BSO, GLOA, HSA, HGFA, SBA, SEOA, ICA, TLBO over social media dataset varying with network density and its achieves higher performance rate higher dense ACF network and relatively lower over lightly dense WA dataset. Moreover, TLBO and ICA achieve extract higher informative community over the higher dense network. At the same time, SEOA and HAS to gain better results with the lower dense networks.

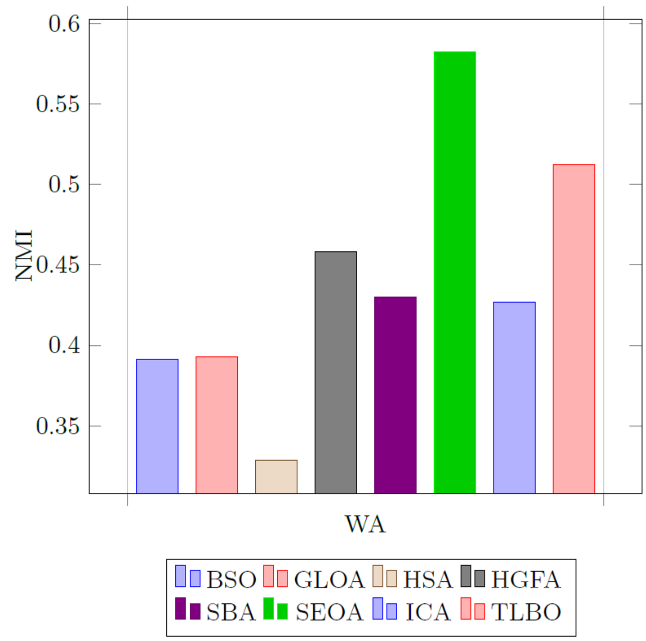

Fig. 20 Normalized Mutual Information of Community Detection over WA Data Set

V. CONCLUSION AND FUTURE WORK

This paper aims to present a comprehensive survey on overlapping community structure on social networks, which is frequently encountered in daily life, and solve community discovery that coincides with a method that has not been applied before. As a result of the research and examinations, it has been observed that the techniques developed for the discovery of overlapping communities in social networks provide solutions to this problem by using a single purpose. At the same time, this paper presents a comparative analysis of meta-heuristic overlapping community detection algorithms over six different social media-based data sets. This paper observed that community detection algorithm BSO, GLOA, HSA, HGFA, SBA, SEOA, ICA, TLBO over social media data set varying with network density and its achieves higher performance rate higher dense ACF network and relatively lower over lightly dense WA data set.

Moreover, TLBO and ICA achieve extract higher informative community over the higher dense network. At the same time, SEOA and HSA gain better results with the lower dense networks.

REFERENCES

[1] Social networks using hierarchical community detection, in 7th Conference on Information and Knowledge Technology (IKT), (2015) 1–5.

[2] Karimi, S.Lotfi, and H. Izadkhah, Communityguided link prediction in multiplex networks, Journal of Informetrics, 15(4) (2021) 101178.

[3] A. Fang, The influence of communication structure on opinion dynamics in social networks with multiple true states, Applied Mathematics and Computation, 406 (2021) 126262.

[4] W. Liu, T. Suzumura, L. Chen, and G. Hu, A generalized incremental bottom-up community detection framework for highly dynamic graphs, in 2017 IEEE International Conference on Big Data (Big Data), (2017) 3342–3351.

[5] Q. Chen and L. Wei, Overlapping community detection of complex network, A survey, in 20th International Conference on Parallel and Distributed Computing, Applications and Technologies (PDCAT), (2019) 513–516, 2019.

[6] D. K. Singh, R. A. Haraty, N. C. Debnath, and P. Choudhury, An analysis of the dynamic community detection algorithms in complex networks, in 2020 IEEE International Conference on Industrial Technology (ICIT), (2020) 989–994.

[7] S. Liu and Z. Li, A modified genetic algorithm for community detection in complex networks, in 2017 International Conference on Algorithms, Methodology, Models and Applications in Emerging Technologies (ICAMMAET), (2017) 1–3.

[8] B. Shen, N. Wang, and H. Qiu, A new genetic algorithm for overlapping community detection, in 2014 Tenth International Conference on Intelligent Information Hiding and Multimedia Signal Processing, (2014) 766–769.

[9] M. Guerrero, F. G. Montoya, R. BaAos, A. Alcayde, and C. Gil, Adaptive community detection in complex networks using genetic algorithms, Neurocomputing, 266 (2017) 101–113.

[10] K. Fatania, D. Joshi, T. Patalia, and Y. Jejani, A comparison of overlapping community detection in a large complex network, in International Conference on Recent Advances in Energy-efficient Computing and Communication (ICRAECC), (2019) 1–6.

[11] C. He, H. Liu, Y. Tang, S. Liu, X. Fei, Q. Cheng, andH. Li, Similarity preserving overlapping community detection in signed networks, Future Generation Computer Systems, 116 (2021) 275–290.

[12] Y. Yan and H. Xie, Detection of hierarchical overlapping communities in complex networks, in International Conference on Intelligent Computing, Automation and Systems (ICICAS), (2019) 893–896.

[13] D. DEMIROL, F. OZTEMIZ, and A. KARCI, Performance comparison of physics-based meta-heuristic optimization algorithms, in International Conference on Artificial Intelligence and Data Processing (IDAP), (2018) 1–5.

[14] G. Rossetti, D. Pedreschi, and F. Giannotti, Nodecentric community discovery, From static to dynamic social network analysis, Online Social Networks and Media, 3(4) (2017) 32–48.

[15] I. Litou and V. Kalogeraki, Pythia, A system for online topic discovery of social media posts, in 2017 IEEE 37th International Conference on Distributed Computing Systems (ICDCS), (2017) 2497–2500.

[16] V. Moscato and G. SperlA, A survey about community detection over online social and heterogeneous information networks, Knowledge-Based Systems, 224 (2021) 107112

[17] M. Li, L. Wen, and F. Chen, A novel collaborative filtering recommendation approach based on soft coclustering, Physica A, Statistical Mechanics and its Applications, 561 (2021) 125140.

[18] R. Vatrapu, R. R. Mukkamala, A. Hussain, and B. Flesch, Social set analysis, A set-theoretical approach to big data analytics, IEEE Access, 4 (2016) 2542–2571.

[19] S. Acer, E. G. Boman, C. A. Glusa, and S. Rajamanickam, Sphynx, A parallel multigpu graph partitioner for distributed-memory systems, Parallel Computing, (2021) 102769.

[20] F. J. BaldA˜ ¡n and J. M. BenAtez, Multivariate time-series classification through an interpretable representation, Information Sciences, 569 (2021) 596–614.

[21] H. Van Lierde, T. W. S. Chow, and G. Chen, Scalable spectral clustering for overlapping community detection in large-scale networks, IEEE Transactions on Knowledge and Data Engineering, 32(4) (2020) 54–767.

[22] H. Jain and G. Harit, Unsupervised temporal segmentation of human action using community detection, in 2018 25th IEEE International Conference on Image Processing (ICIP), (2018) 1892–1896.

[23] N. F. Haq, M. Moradi, and Z. J. Wang, Community structure detection from networks with weighted modularity, Pattern Recognition Letters, 122 (2019) 14–22.

[24] M. Ebrahimi, M. R. Shahmoradi, Z.Heshmati, andM. Salehi, A novel method for overlapping community detection using multiobjective optimization, Physical A, Statistical Mechanics and its Applications, 505 (2018) 825–835.

[25] R. Dong, J. Yang, and Y. Chen, Overlapping community detection in weighted temporal text networks, IEEE Access, 8 (2020) 58118–58129.

[26] A. Karaaslanli and S. Aviyente, Constrained spectral clustering for dynamic community detection, in ICASSP IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), (2020) 8474–8478.

[27] W. Zang, X. Ji, S. Liu, and G. Wang, Percolation on interdependent networks with cliques and weak interdependence, Physica A, Statistical Mechanics and its Applications, 566 (2021) 125612, 2021.

[28] M. A. Javed, M. S. Younis, S. Latif, J. Qadir, and A. Baig, Community detection in networks, A multidisciplinary review, Journal of Network and Computer Applications, 108 (2018) 87–111.

[29] Y. Zhou and F. Xu, Research on application of artificial-intelligence algorithm in a directed graph, in 2017 International Conference on Computing Intelligence and Information System (CIIS), (2017) 116–120.

[30] M. Ehteram, A. Ferdowsi, M. Faramarzpour, A. M. S.Al-Janabi, N. Al-Ansari, N. D. Bokde, and Z. M. Yaseen, Hybridization of artificial intelligence models with nature-inspired optimization algorithms for lake water level prediction and uncertainty analysis, Alexandria Engineering Journal, 60(2) (2021) 2193–2208, 2021.

[31] T. Li, J. Yang, and D. Cui, Artificial-intelligence-based algorithms in multi-access edge computing for the performance optimization control of a benchmark microgrid, Physical Communication, 44 (2021) 101240.

[32] E. Ficarella, L. Lamberti, and S. Degertekin, Comparison of three novel hybrid metaheuristic algorithms for structural optimization problems, Computers and Structures, 244 (2021) 106395.

[33] M. Dehghani, M. Mashayekhi, and M. Sharifi, An efficient imperialist competitive algorithm with likelihood assimilation for topology, shape and sizing optimization of truss structures, Applied Mathematical Modelling, 93 (2021) 1–27.

[34] Y. Ma, X.Zhang, J.Song, and L. Chen, A modified teaching learning-based optimization algorithm for solving an optimization problem, Knowledge-Based Systems, 212 (2021) 106599.

[35] Y. Zhang and Z. Jin, Group teaching optimization algorithm, A novel metaheuristic method for solving global optimization problems, Expert Systems with Applications, 148 (2020) 113246.

[36] A. Ammar and S. ul Islam Ahmad, Teaching-learning based optimization algorithm for core reload pattern optimization of a research reactor, Annals of Nuclear Energy, 133 (2019) 169–177.

[37] A. K. Shukla, P. Singh, and M. Vardhan, An adaptive inertia weight teaching-learning-based optimization algorithm and its applications, Applied Mathematical Modelling, 77 (2020) 309–326.

[38] M. Kumar, A. J. Kulkarni, and S. C. Satapathy, Social evolution learning optimization algorithm, A socioinspired optimization

methodology, Future Generation Computer Systems, 81 (2018) 252–272.

[39] Y. Zhang and P. Zhang, Machine training and parameter settings with a social-emotional optimization algorithm for support vector machine, Pattern Recognition Letters, 54 (2015) 36–42.

[40] P. S. Pal, S. Choudhury, A. Ghosh, S. Kumar, R. Kar, D. Mandal, and S. P. Ghoshal, Social-emotional optimization algorithm-based identification of nonlinear Hammerstein model, in 2016 International Conference on Communication and Signal Processing (ICCSP), (2016) 1633– 1637.

[41] A. Rauniyar, R. Nath, and P. K. Muhuri, Modified brain storm optimization algorithm in objective space for a pollution-routing problem, in 2019 IEEE Congress on Evolutionary Computation (CEC), (2019) 242–247.

[42] Z. Dai, W. Fang, K. Tang, and Q. Li, An optimaidentified framework with brain storm optimization for multimodal optimization problems, Swarm and Evolutionary Computation, 62 (2021) 100827.

[43] L. Jain, R. Katarya, and S. Sachdeva, Opinion leader detection using whale optimization algorithm in online social network, Expert Systems with Applications, 142 (2020) 113016.

[44] X. Feng, Y. Wang, H. Yu, and F. Luo, A novel intelligence algorithm based on the social group optimization behaviours, IEEE Transactions on Systems, Man, and Cybernetics, Systems, 48(1) (2018) 65–76.

[45] A. Ishizaka, B. Lokman, and M. Tasiou, A stochastic multi-criteria divisive hierarchical clustering algorithm, Omega, vol. 103, p. 102370, 2021.

[46] J. Sheng, J. Hu, Z. Sun, B. Wang, A. Ullah, K. Wang, and J. Zhang, Community detection based on human social behaviour, Physica A, Statistical Mechanics and its Applications, 531 (2019) 121765.

[47] J. Ma, J. Zhang, Y. Lin, and Z. Dai, Cost-efficiency trade-offs of the human brain network revealed by a multiobjective evolutionary algorithm, NeuroImage, 236 (2021) 118040.

[48] A. Aylani and N. Goyal, Community detection in social network-based non user as social activities, in 2017 International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud) (I-SMAC), (2017) 625–628.

[49] W. W. Zachary, An information flow model for conflict and fission in small groups, Journal of Anthropological Research, 33(4) (1977) 452–473.

[50] D. Lusseau, The emergent properties of a dolphin social network, Proceedings of the Royal Society of London. Series B, Biological Sciences, 270(2) (2003) S186– S188.

[51] M. Girvan and M. E. J. Newman, Community structure in social and biological networks, Proceedings of the National Academy of Sciences, 99 (2002) 7821–7826.

[52] M. E. J. Newman, Modularity and community structure in networks, Proceedings of the National Academy of Sciences, 103(23) (2006) 8577–8582.