Novel Compression Based Community Detection Approach Using Hybrid Honey Badger Bonobo Optimizer for Online Social Networks

K. Sankara Nayaki [1*], M. Sudheep Elayidom [2], R. Rajesh [3]

Abstract

Online social media community detection identifies node connections. Clusters, modules, or groups in different networks define the community. Community detection finds hidden network relationships. Several works have been done to detect network node communities, but performance is often affected by imprecise detection, time complexity, etc. We proposed a Hybrid Honey Badger Bonobo Optimization to detect network node communities (HHBBO)[34]. Before applying HHBBO, networks are compressed to reduce time complexity and identify node communities. Global optimization can be achieved using HBO and BO. Hybrid algorithms optimize global search. This searches nodes globally and detects their relationships. Experimental analyses show that the proposed approach can detect online social media node communities more effectively than other approaches. GA, LSMD, DPCD, ICLA and HHBAVO are used for comparison.

Key Words: Community Detection; Badger; network; clusters; compressed network; and hybrid

DOI Number: 10.14704/nq.2022.20.8.NQ44248 **NeuroQuantology 2022; 20(8):2268-2283**

Introduction

Network models can be applied to a wide variety of real-world systems, including social networks, transportation networks, biological information systems, and information systems. The primary modeling technique for the systems is a graph, which helps to reduce the complexity and make it easier to manage. The community conducts research on the mesoscopic structures of networks with the goals of detecting characteristics and gaining an understanding of how systems work [1]. When detecting communities, one can obtain results that are of great theoretical importance and have wide-ranging applications in the real world. Some examples of these applications include the analysis of epidemics, the prediction of links, the detection of terrorist groups from networks, and recommendation systems [2]. The community detection problem is primarily classified as NP-hard because the intricate configuration of a real-world network cannot be modeled using a straightforward lattice graph form.

The intricate nature of the network structure can be attributed to the tremendous conversations, variability, and quasi topological features that exist between the inherent entities. The astounding improvement in the usability of online social networking platforms that is the result of the humongous information exchange that occurs between users located all over the world [3] In order to gain a better understanding of the development, intricacies, function, and formation of social networks, the information pertaining to the users of Online Social Networks (OSNs), such as their relationships and user behavior, is analyzed. The users spread the information among themselves and their immediate surroundings [4]. The users disseminate information concerning presidential campaigns, hobbies, the sale of goods or services, personal favors, and personal relationships.

Corresponding author: K. Sankara Nayaki
Address: [1]Research Scholar, Division of Computer Science, Cochin University of Science & Technology, Kochi, India, [2]Professor, Division of Computer Science, Cochin University of Science & Technology, Kochi, India, [3]Scientist, NPOL, Kochi, India
E-mail: sankaranayaki@gmail.com

The varied points of view and the dissemination of information both speed up with the assistance of a quick internet connection, which plays an integral role in the process of information dissemination. The diffusion strategy denotes the similarity [5], taking into account the modular design, process optimization, and composition of connections. Learning to compare different pieces of data and taking that into consideration is included in the preferences [6]. When it comes to online social platforms, the tendency toward information preferences and decision-making that are comparable is taken into account. In order to derive expertise from the information that is generated, the process of information diffusion requires the organization of information regarding people's behaviors and patterns of diffusion.

Communities in the social network are developed through the combination of user attributes and the characteristics of the network itself. Friends in a social circle typically share commonalities such as their place of employment, level of education, areas of interest, and frequency of contact [7,8]. The process of locating a community of users who share common characteristics within social networks is known as community detection. A distinct community detection algorithm is developed in order to extract the social actor attributes as well as the topology of the social network. Third parties are the participants in community detection, and some examples of third parties include hospitals, research organizations, business organizations, and other similar groups. The high amount of time and computational complexity associated with information processing brought on by an increasing number of users in a social network are the primary challenges that are encountered when attempting to identify communities [9,10]. The community detection problem that this paper focuses on reduces the high complexity that is normally associated with processing the information provided by big data by selecting a small number of samples that are of a high importance.

As a general rule, community detection is carried out by making use of a variety of machine learning and deep learning strategies, as was recommended in the most recent round of research investigations [7]. The work on modularity is given a new lease of life thanks to the community detection problem, which is primarily regarded as an optimization issue. The multiple objectives are represented by a set of tradeoffs, and each solution that obtains results for community detection serves as a representation of those tradeoffs. The qualities of the community are rated according to a variety of criteria depending on the solution. Several distinct solutions are chosen in response to the many different requirements. Compression is used in this work to solve a problem that the previous system had with community detection in large-scale networks. This problem had been a challenge for the previous system. In this research, we proposed a new algorithm for identifying communities within social networks called the Hybrid Honey Badger based Bonobo Optimization (HHBBO) algorithm. The following is a synopsis of the most important function that this article serves:By exploring network topologies, the multi-objective Hybrid Honey Badger-based Bonobo Optimization (HHBBO) algorithm is proposed to solve the community identification problem in social networks.

The HHBBO algorithm reduces the network's size to a manageable size. In order to create a compressed network, all of the weakly connected nodes are combined into a single node (CN). The primary goal of this initialization step is to improve the algorithm's accuracy and convergence. Ratio Cut, (RC) and Kernel k-means (Kkm) minimization are taken as the multi-objective optimization problem's goals. Comparisons with certain other baseline models depending on multiple massive social network datasets are used to evaluate the proposed framework for efficacy.

This paper is organized as follows: Section 2 discusses related works, and Section 3 presents a formal definition of the problem. Algorithm details are provided in Section 4, work proposal details are provided in Section 5, and experimental findings are discussed in Section 6. Section 7 is the final part of the article.

Literature Survey

Researchers have produced a number of different works in their efforts to identify communities within social networks. The works that are depicted below are those that are the most pertinent and effective. A parameter known as community, which is considered to be a potential characteristic, can be used to identify the behavior of social entities. This behavior can be thought of as the potential characteristic. Behera et al. presented an original genetic algorithm-based (GA) approach in their paper [8]. This method has been utilized in order to identify the communities by taking into consideration the similarity index. A distributed method was utilized to calculate the similarity index across a number of different computing nodes. The topological structure of the network serves as the foundation for the similarity index. The accuracy of the community's detection was quite high; however, the analysis of the network's communities did not take into account the disjoint and overlapping communities.

In addition, Bouyer et al. [9] described a novel approach to the community label allotment that makes use of the information that is specific to the region. The author utilized the label assigning in a multi-level diffusion strategy (LSMD) in order to complete the label assigning process. This approach required significantly less time than the others. The node with one degree is the starting point for the process of assigning labels, which is then followed by the nodes directly adjacent to it and the nodes directly adjacent to it at the second level. From that point forward, the assignments were carried through to nodes with degrees of 2, 3, and so on. Subsequently, the communities that had been assigned were combined with the assistance of a quick and easy strategy that was followed in LSMD. According to the authors, the detection accuracy is significantly improved, and it also eliminates the need for a random search and has parameters that can be adjusted. Nevertheless, the purpose of this work was not to conduct an analysis of the concept of overlapping communities.

Differentially Private Community Detection (DPCD) approach is a novel method that was elucidated by Ji et al. [10] to detect communities at the same time while protecting the user's attributes as well as network topology. This was accomplished by the DPCD approach. The authors explained that the model is a probabilistic generative model that has been broken down into smaller tasks for each individual user to work through and complete. A combination of objective perturbation and differential privacy guarantees is utilized in the process of protecting individuals' confidentiality. The algorithm for the economical recovery of node affiliations has been successful in accomplishing this (NAR). This method improved community detection in addition to privacy policies while only requiring a minimal amount of additional time.

In the meantime, Khomami et al. [11] presented an innovative approach called Irregular cellular Learning Automata (ICLA). The authors made an effort to determine the composition of the communities by mining the social networks with a mechanism known as Significant Community Detection based on Cellular Learning Automata (SIG-CLA). The ICLA that was proposed was utilized in order to detect the community in both the local and the global environments. In the work, the benefits of SIG-CLA in terms of modularity and Normalized mutual information were described (NMI). This ensures a higher level of accuracy in the detection of communities despite the increased computational burden. In addition, Raj et al. [12] proposed a novel method to identify communities within online social networks. This method, which is known as granular computing and is based on rough sets, is described below. In order to evaluate the results obtained from the datasets that were used, the authors described two parameters that are referred to as object community factor and granular community factor. The authors came to the conclusion that the work had an overall performance of 3.98 out of 5. The author did not conduct an analysis of the concept of overlapping communities.

In their paper [13], Wu et al. presented a novel method for the detection of deep communities in online social networks that was based on graph embedding. These shortcomings of the adjacency matrix, such as inadequate spatial proximity, were addressed by the authors through the utilization of this method. In their explanation of this methodology, the authors broke it down into three stages: I matrix reconstruction; (ii) spatial feature extraction; and (iii) community detection. The spatial localization can be extracted by using the opinion leader as well as the neighbors that are closer to you by using the spatial proximity matrix and the acquired matrix. After that, an autoencoder can be used to derive the reconstructed adjacency matrix, which will ultimately result in an increase in modularity. This method improves the accuracy of detection; however, determining the presence of a community in a setting that is constantly changing is a challenging task.

Based on the reviews, it has been determined that the earlier works focused on identifying communities derived from online social networks using a variety of unconventional methods. However, the previous research did not address a number of the concepts that were being discussed, such as community overlapping, time complexity, computational complexity, and disjoining, amongst others. Table 1 contains an illustration of the review's executive summary.

Table 1: Survey on the Traditonal Algorithms

Reference	Method/algorithm	Performance metrics	Limitation
Behera et al. [8]	Genetic Algorithm (GA)	Node Detection & Accuracy	Structural Complexity
Bouyer et al. [9]	Local information for community's label allotment (LSMD)	Structural & Semantic Labelling	Structural Complexity
Ji et al. [10]	Differentially Private Community Detection (DPCD)	Enhanced community detection as well as privacy protection	Time Complexity
Khomami et al. [11]	Irregular cellular Learning Automata approach (ICLA)	Dynamics in community detection accuracy	Computational Complexity
Nayaki et al. [31]	Hybrid Honey Badger African Vulture optimizer (HHBAVO)	Better Compression	Unable to handle larger Dataset

Problem Definition

In order to determine the existence of a community, it is necessary to partition the entire network into subgraphs and ensure that all of the nodes are connected to one another. The community is created by the connections made between the nodes in the subgraphs, and this community is the primary factor that is considered during the classification stage of the community detection algorithm. Graph G (U, F) used to construct the network. Finite sets of the nodes are represented as $U = \{u_1, u_2, ..., u_M\}$. The edges of the set to symbolize the relationships among objects are $F = U \times U \ (|e| = n)$. The linking among the nodes could be denoted as u_j and u_k. The signified set of M_c community(ies) of G is denoted $C = (C_1, C, ..., CM_c)$. In the unweighted and undirected graphs, overlapping community detection is focused [14]. Community detection in networks is an NP-hard problem, and approximation algorithms are used to solve it in order to make it more manageable. The multi-objective optimization model has good performance when it comes to evaluating the qualities of the community [15]. This study formulates multi-objective minimization problems such as RC and Kkm, which are explained as follows:

$$Minimize\ Fit(C) = Rc(C),\ \ Kkm(C) \tag{1}$$

$$Rc(C) = \sum_{j=1}^{M_C} \frac{L(C_j, \overline{C_j})}{|C_j|} \tag{2}$$

$$Kkm(C) = 2(M - M_C) - \sum_{j=1}^{M_C} \frac{L(C_j, \overline{C_j})}{|C_j|} \tag{3}$$

where C_j is the given node-set of the jth community, the M is the count of vertices, MC is the aggregate count of communities, and ρ is the partition for the MC cluster where $C_j \in \rho$. The significance of $\overline{C_j}$ is $\rho - C_j$. The Entirety of the density of the link of intra-communities is denoted using the KKM, parameter providing the densities by Angelini et al. [16]. The count of communities with the reducing function is KKM. By reducing the importance of these two goals, the requirements of communities can be reliably met. As a result, we've decided to use objective functions like RC and Kkm for this investigation.

Background

This section will provide a comprehensive overview of the algorithms that were chosen to implement HBO and BO.

4.1 Bonobo optimizer (BO) algorithm

Bonobos are humankind's closest living relatives, the algorithm mimics bonobo social, reproductive behavior and they're Homininae [37]. Bonobo was discovered in 1929 in a Belgian colonial museum [38]. They share 98% of our genetic profile. Human and bonobo ancestry split 8 million years ago. Later, chimpanzee and bonobo ancestors split. As the common chimpanzee and bonobo are similar, the authors included references to chimpanzees while describing bonobos. In order to solve an optimization problem, the proposed algorithm models bonobos' social behavior and reproductive technique. The proposed BO uses a fixed population size and random initialization, like other metaheuristics. The alpha bonobo (bonobo) has the best rank in a bonobo class's super hierarchy and is the fittest in the population. The alpha bonobo is the best option. In addition, BO parameters (not user-defined) are initialized with their own starting values. FIG is a flowchart of the BO. All random numbers used in the proposed algorithm are in the range (0.0, 1.0).

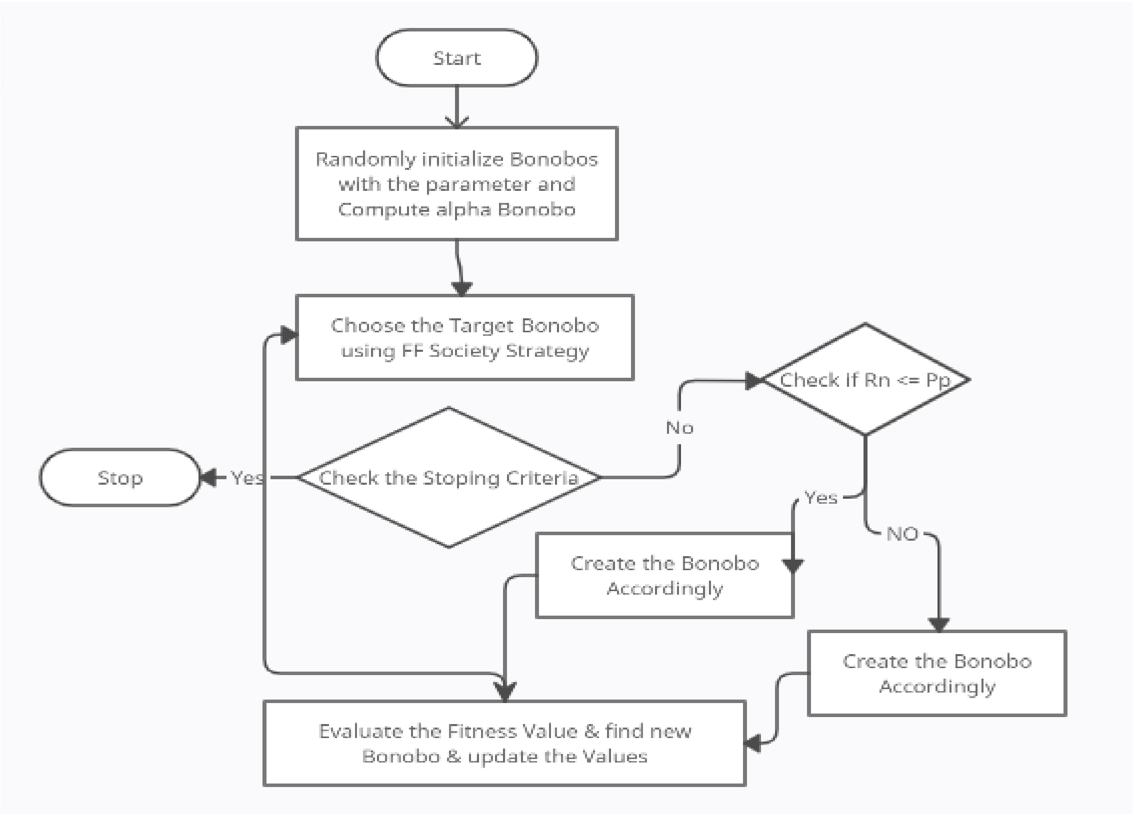

Figure 1: Flowchart of Bonobo Algorithm

4.1.1 Initialization

BO's non-user-defined parameters are initialized as follows: probability of phase (pp), positive phase count (ppc), negative phase count (npc), probability of extra-group mating (pxgm), change in phase (cp), temporary sub-group size factor (tsgsfactor), and directional probability (pd). The algorithm where these parameters are used provides more detail.

$ppc = 0,$ (4)

$npc = 0, cp = 0$ (5)

$p_{xgm} = p_{xgm-initial}$ (6)

$tsgs_{factor} = tsgs_{factor\_initial}, p_p = 0.5, p_d = 0.5,$ (7)

$tsgs_{factor} = tsgs_{factor\_initial}$ (8)

$p_p = 0.5, p_d = 0.5,$ (9)

4.1.2 Positive Phase (PP) and Negative Phase (NP)

The proposed algorithm first considers Positive Phase (PP) and Negative Phase (NP). Positive phases are when adequate food, mating success, protection, etc. are available for peaceful living. Negative phases or situations indicate the absence of peaceful and good living conditions.

In the proposed scheme, PP is a phase where the best solution's fitness improves (i.e., alpha bonobo). If the current-best solution doesn't change, the phase is NP. When iteration progresses, the number is counted for each algorithm phase. When the algorithm passes through PP, ppc is incremented by one in each iteration. If the algorithm passes through NP, negative phase count (npc) is increased by one

eISSN 1303-5150

ppc and npc start at zero. Increasing one parameter sets the other to zero.

In bonobo communities, the alpha bonobo's status can change. Therefore, a bonobo's alpha status may not last until the end of evolution. If a bonobo has more potential than the current-best, it will be named alpha. Different solutions could win alpha bonobo in different BO iterations.exploration phase to the exploitation phase is performed using the following equations. The vultures are satiated or hungry; the rate of being satiated has a declining trend.

4.1.3 Selection of a Bonobo using fission-fusion social strategy

Different updating mechanisms (i.e., mating schemes to generate offspring) are used depending on the current situation or phase (PP or NP). Another bonobo, chosen using the fission-fusion social technique, also mates. Bonobos live in a fission-fusion social community.

A large community of bonobos forms small temporary parties. Unpredictable party sizes. These temporary groups stayed for a few days and then rejoined their larger community. This social strategy motivates the bonobo to mate. First, the maximum size of a temporary sub-group (tsgsmax) is determined based on the total population size (N). As soon as the jth-variable (new bonomj) is found to be higher than Var_maxj's, the new bonomj is assigned the same value as Var_maxj. When the value of new bonoboj is less than the value of Var_minj, a value equal to Var minj is assigned to it.

$$tsgs_{max} = maximum\,(2, (tsg\ \ s_{factor} \times N)) \quad (10)$$

$$\beta_1 = e^{(r_4+r_4-2/r_4)}, \quad \beta_2 = e^{(-r_4+2\times r_4-2/r_4)}, \quad new\_bo\,j = bonobo_j^i + \beta_1 \times (Var\_max_j - bonobo_j^i), \ if\ (\alpha_{bonobo}^j \geq bonobo_j^i\ and\ r_3 \leq p_d), \quad new\_bo\,j = bonobo_j^i - \beta_2 \times (bonobo_j^i - Var\_min_j), \ if\ (\alpha_{bonobo}^j \geq bonobo_j^i\ and\ r_3 > p_d), \quad new\_bo\,j = bonobo_j^i - \beta_1 \times (bonobo_j^i - Var\_min_j),$$

$$new\_bonobo_j = bonobo_j^i + r_1 \times scab \times (\alpha_{bonobo}^j - bonobo_j^i) + (1-r_1) \times scab \times flag \times (bonobo_j^i - bonobo_j^p) \quad (11)$$

To be considered, either the fitness of the newly-created _bonobi must be better than the bonoboi, or pxgm must be less than or equal to the newly-created _bonobo's random number. As a result of this, bonoboi will be phased out and replaced by the new bonobo. New bonobo, on the other hand, can only be identified if its fitness compares favorably to that of the bonobos. Cp is the change in phase probability, and rccp is how quickly that probability changes. The initial value of the phase probability was set at 0.5 to imply that both sets of mating strategies are of equal importance. However, as the iteration progresses, the probability of a particular phase is updated. There is no upper limit to the parameter: temporary social group size factor (tsgsfactor). The algorithm's initial assignment is made by the user. A temporary subgroup size factor (tsgfactor) and extra-group mating probability (pxm) are the initial values of these two variables. pxgm initial is a user-defined parameter in this example.

4.1.4 Parameters' updating

Different updating mechanisms (i.e., mating schemes to generate offspring) are used depending on the current situation or phase (PP or NP). Another bonobo, chosen using the fission-fusion social technique, also mates. Bonobos live in a fission-fusion social community. A large community of bonobos forms small temporary parties. Unpredictable party sizes. These temporary groups stayed for a few days and then rejoined their larger community. This social strategy motivates the bonobo to mate. First, the maximum size of a temporary sub-group (tsgsmax) is determined based on the total population size (N).

$$npc = 0, ppc = ppc + 1, \ \ cp = minimum\,(0.5, ppc \times rcpp) \quad (12)$$

$$p_{xgm} = p_{xgm\_initial}, p_p = 0.5 + cp, p_d = p_p \quad (13)$$

$$tsg\ \ s_{factor} = minimum\,(tsgs_{factor\_max}, (tsgs_{factor\_initial} + ppc \times rcpp^2)) \quad (14)$$

4.2 Honey Badger Algorithm (HBA)

HBA depends heavily on the qualities that honey badgers possess [19]. This one uses a honey assist bird or its sense of smell in conjunction with digging to locate the source of its food. As a result, the behaviors associated with the honey badger's search for food can be divided into two distinct categories:

the honey mode and the digging mode. With in honey mode, this same badger will follow the directions given to it by the honey guide bird. In the latter mode, the badger will use its keen sense of smell to encircle the food and then dig it up to eat it. In both modes, the badger will eventually succeed in eating its prey. The sections that follow will provide an explanation of both of these modes.

4.2.1 Numerical expression

The following table shows the numerical expression of the digging stage as well as the honey stage. The HBA is a type of algorithm that is typically referred to as a global optimization algorithm and it is used to perform both the exploitation and exploration stages. One way to express the candidate solution population, or CSP, is as follows:

$$CSP = \begin{bmatrix} a_{11} & a_{12} & a_{13} & \Lambda & a_{1d} \\ a_{21} & a_{22} & a_{23} & \Lambda & a_{2d} \\ \Lambda & \Lambda & \Lambda & \Lambda & \Lambda \\ a_{m1} & a_{m2} & a_{m3} & \Lambda & a_{md} \end{bmatrix}$$

(15)

Then the jth spot of the honey badger a_j is expressed as,

$$a_j = \left[a_j^1, a_j^2, ..., a_j^d\right]$$

(16)

4.2.2 The Initialization stage:

The Count of honey badgers (number) is primed as follows,

$$a_j = L_j + rand \times \left(U_j - L_j\right)$$

(17)

here, rand: Random Number (0,1). The candidate resolution of the jth honey badger is as a_j for M number of the population. U_j and L_j are Upper and Lower Constraints respectively.

4.2.3 Strength - Intensity (SI)

Intensity can be defined as the strength of the prey as well as the distance between the badger and the prey. The degree to which the prey can be smelled can be defined as SI_j. The law of inverse square states that the motion is influenced by the odor [20], and it can be expressed as,

$$SI_j = rand_2 \times \frac{r}{4\pi D_j^2}$$

(18)

where D_j is the distance between the jth badger and the prey and $rand_2$ describes the random number and rand is the Random Number (0,1). The intensity or concentration of the odor is represented by the symbol r, and its value can be expressed as,

$$r = (a_j - a_{j+1})^2$$

(19)

There is a formula that can be used to calculate the distance between the jth badger and the prey,

$$D_j = a_{prey} - a_j$$

(20)

4.2.4 Density factor (δ)

Controlling the random time-varying function can be done through the use of the Density Factor (δ). As a result, it is essential to bring the density factor up to date, which can be accomplished by,

$$\delta = G \times \exp\left(\frac{-I}{I_{\max}}\right) \qquad (21)$$

here I : current iteration and $I_{\max}$: maximum iterations. G : constant factor $G \geq 1(Default = 2)$.

4.2.5 Escaping Local optimum

It is possible to achieve escapism from the local optimal state by bringing the Agents Position up to date. In order to seal the deal, the directions of the search have been diversified through flag h.

4.2.6 Updating the Agents position

Both the phases have different position updating processes and are elucidated below.

Digging phase:

The digging phase is performed by the cardioid shape as [21],

$$a_{new} = a_{prey} + h \times \gamma \times rand \times a_{prey} + h \times rand_2 \times \delta \times D_j \times [\cos(2\pi ran$$

(22)

a_{prey} is the position of the best prey. All the random numbers are randomly generated hich falls between (0, 1)

$$h = \begin{cases} 1 & if \;\; rand_6 \leq 0.5 \\ -1 & Else \end{cases}$$

(23)

It is possible to determine a better location for prey by allowing the badger to investigate any disturbances that occur while it is digging..

Honey phase:

During this stage, the badger follows the directions provided by the honey birds and arrives at its intended hive of bees. The following is one possible formulation for it:,

$$a_{new} = a_{prey} + h \times rand_7 \times \delta \times D_j$$

(24)

The Updated location is a_{new}. The search behaviour for HB is influenced by the behavior of the search, which varies over time.

5. HHBBO Algorithm

This section elucidates the proposed hybrid HBA-BO approach to detect the community from the compressed network. The proposed method exploits the strong exploration capacity of both the HBA and BO algorithms to attain the optimized exploration capacity. Then the updated location can be given as,

For the digging phase,

$$a(t+1) = a_{prey}(t) + h \times \gamma \times rand_1 \times a_{prey}(t) + h \times rand_2 \times \delta \times D_j(t) \times [cos\,(2\pi rand_3) \times [1 - cos\,cos\,(2\pi rand_4) \; - sin\,sin\,(2\pi rand_5)\,]$$

(25)

For honey phase

$$a(t+1) = a_{prey}(t) + h \times rand_6 \times \delta \times D_j(t)$$

(26)

The procedure followed in the formulation of hybridization of AVO and HBA is illustrated below,

Initialize the total number of badgers, vultures, number of iterations, intensity factor, density factor (δ), vulture starvation rate (T).

Determine the location of the badger $a_{new} = a_{prey} + h \times \gamma \times rand \times a_{prey} + h \times rand_2 \times \delta \times D_j \times [cos\,(2\pi rand_3) \times [1 - cos\,(2\pi rand_4) - sin\,(2\pi rand_5)$ and $a_{new} = a_{prey} + h \times rand_7 \times \delta \times D_j$ and estimate the vector position of vultures as $R(j+1) = R(j) - f + rand_2 \times ((B_U - B_L) \times rand_3 + B_L)$ to determine the distance between the nodes in order to detect the community from the online social network.

5.1 Compression network

Some of the nodes form into inseparable clusters because of the network topologies that allow for their close connection to one another. Every time, the communities of tightly connected nodes serve as the foundation for the formation of the groups. This phenomenon serves as the impetus for our compression strategy. Find the nodes in the groups that have a lot of strong connections in order to investigate the information that is local to the nodes. Detection across the entire community involves each individual group. The relative strengths of each node pair's connectivity are determined with the help of a compression operation. Condensing multiple nodes that are strongly connected into a single node results in the creation of a compressed network known as CN.

When validating the strength of the connectivity between node pairs, the node similarity indexes should be given preference. An abundance of similarity indexes can be generated using the global and local network topology information provided by Katz, Jaccard, and Salton [22].

Adamic et al. [23] proposed an AA index in which the effect is measured by the node's degree, and the index also describes the effect of the node's common neighbors. The contribution of nodes with lower degrees is significantly higher. The AA index can be understood by solving the following equation.

$$AA(u_1,u_2) = \begin{cases} \sum_{u \in CM_{u_1,u_2}} \frac{1}{\log(l_u)}, & if \quad B_{u_1,u_2} = 1 \\ 0, & Otherwise \end{cases}$$

(27)

The node degree u is l_u. Both of these common neighbor nodes (u_1 & u_2)are represented by the symbol, as CM_{u1u2}. Determine the degree to which both nodes are similar when you connect them. Following the identification of commonalities, the same community chooses each node. When the similarity value is too low, there is no point in selecting nodes to investigate further. Please provide the threshold (δ).

The process of compression is outlined in Algorithm 1, which can be found here. At the outset, the AA similarity values of each and every pair of nodes are computed. Following the removal of similarity values, the initial percentage is chosen, and this value is set to be equivalent to zero. It has a degree of one, which indicates that it is directly connected to the single node that is its neighbor. Each component is condensed into one node, and the components that are connected to one another are determined. The use of self-loops is employed in the description of the interior edges of the original components.

Algorithm1: Network compression process (G, δ) [31]

Input: Compression ratio for edges δ and the original complex network G
Compression network CN
For j=1 to mdo
$M_j \leftarrow$ Determine the neighbors node M_j
For j=1 to M_j do
$Nc_{jk} \leftarrow$ Common node neighbors are u_j and u_k
Update equation (6)
End For
End For
$S(\delta) \leftarrow$ Select the node pair similraities
For j=1 to M_j do
If kj==1then
$msM_j \leftarrow M_j$
Else
$msM_j \leftarrow M_j$
End If
Retaining $a(j, msM_j) = 1$ update the adjacency matrix
End For
Determine each linked component of G
Each linked component is compressed into a single node to describe the linked components and the interior edges represent the self-loops
Output: Compressed network CN

A simplified illustration of a network consisting of ten nodes can be found in Figure. 2. The number of interior edges is represented by a loop inside the diagram. After the compression process, the corresponding nodes and are altered from the node sets and respectively. The corresponding nodes $1^{`}$, $2^{`}$, $3^{`}$ and $1^{`}$ are changed from the node sets $\{1,2,3\}$, $\{4,5,6,7\}$, $\{9\}$ and $\{8,10\}$ after the compression process.

5.2 Representation

The proposed method of HHBBO makes use of a locus-based adjacency (LBA) encoding scheme [30] to figure out the answers to questions regarding community detection. In this case, the nodes in the network are denoted by the letter N and are taken to represent the HHBBO that contains N total birds. Because the nodes are connected to one another through other nodes, this represents the community, as shown in figure 3. Both the hypothetical representation of the network with three communities, which is shown in figure 3(a), and the LBA representation, which is shown in figure 3(b), are shown below. The search space can be narrowed down with the help of the community that is represented by this LBA. In addition, the original framework of the networks will not be altered in any way by the network that was designed using LBA representation.

5.3 General Framework of the proposed HHBBO approach

The HHBBO method, as it has been proposed, is comprised of three stages: first, the original network is compressed by adhering to the procedures outlined in algorithm 1. After that, the populations of both counts are first established utilizing the LBA method and the random walk method, respectively. In the meantime, once the construction of the solution has been finished, the pheromone trails are improved in two different efficient ways. By reducing the amount of variability in the pheromone trails to a constant factor, the ineffective solution that finds the community within the condensed network can be ignored. As a consequence of this, considering the quality of the detection provided by the nodes, the procedure for reinforcing is carried out effectively. Pheromone is only deposited on the corners of the compressed network through the use of a single node while this operation is being carried out. At this point in the process, the modularity and NMI are calculated by measuring the distance between the nodes, determining the position of the new node, calculating the intensity factor, and calculating the density factor. In the end, for each individual node, a node Nj will be chosen at random from the population of N nodes. On the basis of the information gathered from the community, these two nodes can be utilized to locate other possible topologies for the network. It is necessary to keep doing this until the maximum number of required iterations has been reached.

6. Result and Discussion

Through the use of a number of different tests, this part of the article demonstrates the usefulness and viability of the suggested procedure. A comparison to the current state of the art is used to determine how successful the proposed method is. The experimental designs are the first topic to be covered. After that, a look is taken at the description of the dataset, the evaluation measures, and the state-of-the-art results..

6.1 Design of the experiment

The comparison between the proposed method and the state of the art is carried out using previously established methodologies such as GA, LSMD, DPCD, and ICLA. An effective community detection algorithm is utilized in order to optimize the community score as well as the community fitness. During the process of community detection, the proposed method effectively optimizes both RC and Kkm. The modularity function Q can be optimized thanks to the high accuracy of the proposed algorithm. The proposed method is a quick algorithm to update the labels, and it takes into account the status of the neighbors. Taking advantage of a network's eigenvector allows for the identification of communities in real-world networks.

Dataset description

Experiments based on the functioning of actual networks were carried out for this study. The commonly employed real-world networks, specifically American college football clubs, Dolphins, and Karate, are investigated here. In addition, we use the networks, such as the Network Scientists' Cooperation Network, Email, and books about politics in the United States. The data set is described in Table 2, which can be found below.

Table 2: Dataset description

DataSet	N	E	C	K	S	A
CNS	250,469	30,230,905	437	690	143.51	0.25
AMAZON	125,120	2,248,406	3,140	33,569	15.54	0.39

Performance metrics

In order to report comparison results for the purpose of the performance evaluation of algorithms, the Normalized Mutual Information (NMI) [24] and modularity (Q) [25] metrics are used. The modularity Q metric is used to represent the quality of the results obtained from the community detection process. The following provides an explanation of these two measures:

$$Q = \sum_{r=1}^{Mc} \left[\frac{L_r}{n} - \left(\frac{D_r}{2n} \right)^2 \right] \tag{26}$$

Count of detected communities are M_c. The aggregate of the node degree and number of detected communities M_c is D_r and L_r.
Another metric that is calculated using the Shannon entropy is is NMI. x and y are the likenesses among the communities with respect to detected and ground-truth communities.

$$N_{MI}(x, y) = \frac{2 \times i(x, y)}{h(x) + h(y)} \tag{27}$$

The mutual information $i(x, y)$ expressed as,

$$i(x, y) = \sum_{jk} R_{jk} \log\left(R_{jk} / (R_j + R_k\right) \tag{28}$$

The Entropy function $h(x)$.

$$h(x) = \sum_{j} R_j + \log\left(R_{jk}\right) \tag{29}$$

The probability of randomly drawn nodes is R_{jk}. The intersections of jth and kth communities are m_{jk}. The NMI falls into the [0, 1] under the interval that was normalized. When there is no ground truth available, the results of the community detection are evaluated based on the modularity value. The NMI and Q are both incorporated into the evaluation metrics. The results of the detection are significantly improved.

6.2 Real-world network experiment

The state-of-art result of final and after network compression is depicted in Fig 4. There are eleven real-world networks with five state-of-art techniques like Genetic Algorithm (GA) [8], label assigning via multi-level diffusion (LSMD) [9], Differentially Private Community Detection (DPCD) [10], Irregular cellular Learning Automata approach (ICLA) [11], and the proposed method are used for the experimental analysis. The community detection results based on the modularity are displayed in Table 3. The state-of-art results based on NMI values with respect to five networks are described in Table 4.

Table 3: Representation of Moduralty based on the Algorithms

Networks	Name of the Techniques									
	GA [8]		LSMD [9]		DPCD [10]		HHBAVO [31]		HHBBO (P)	
	Q	Mc	Q	Mc	Q	Mc	Q	Mc	Q	Mc
Co-authors	0.82	13	0.80	15	0.80	18	0.83	11	0.82	15
Polbooks	0.52	4	0.50	5	0.51	4	0.52	4	0.74	58
Football	0.6	11	0.6	10	0.51	11	0.53	10	0.43	8
Dolphins	0.5	4	0.51	5	0.49	4	0.6	4	0.74	58
Karate	0.42	4	0.37	3	0.42	4	0.38	3	0.5	4

Table 4: Comparison NMI values with the Algorithms

Networks	Name of the Techniques									
	GA [8]		LSMD [9]		DPCD [10]		HHBAVO [31]		HHBBO (P)	
	NMI	Mc	NMI	Mc	NMI	Mc	NMI	Mc	NMI	Mc
Co-authors	0.96	11	0.90	18	0.93	14	0.85	19	0.92	10
Polbooks	0.52	4	0.59	4	0.52	4	0.51	4	0.70	8
Football	0.91	10	0.85	11	0.70	8	0.91	11	0.53	5
Dolphins	1	2	1	2	0.53	5	0.57	5	0.90	18
Karate	1	2	1	2	0.68	4	1	2	0.53	5

Fig 2 describes the state-of-art result of final and after network compression. After network compression, the comparisons of NMI values are plotted. While compression ratio equals 0.4, known ground truth and the final ones on the real-world networks. The compressed network achieves all the NMI values. This helps to obtain more accurate preliminary partitions but the compression procedure not only minimizes the network scale.

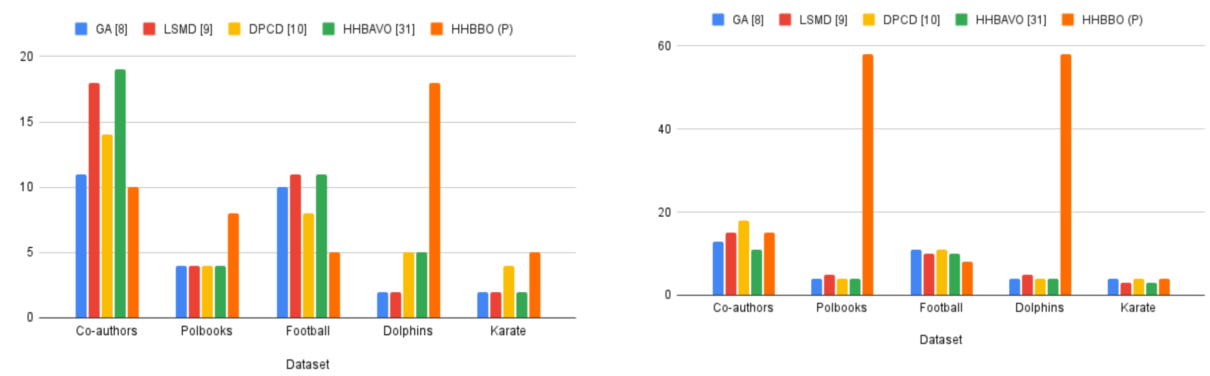

Figure 2. Representation of Moduriality - before and after network compression

6.3 Analyzing the performance of synthetic networks

This type of network possesses power-law distributions for both node degree and community size. The parameter settings of the Lancichinetti–Fortunato–Radicchi (LFR) benchmark which is a synthetic network are shown in table 5. Here, 20 networks are produced for each set of network parameters, and the generated community detection outcomes that are NMI are shown in figure 5. The NMI is inversely proportional to the µ. The higher the µvalue more ambiguous the community framework and hence it is arduous to detect the community more precisely.

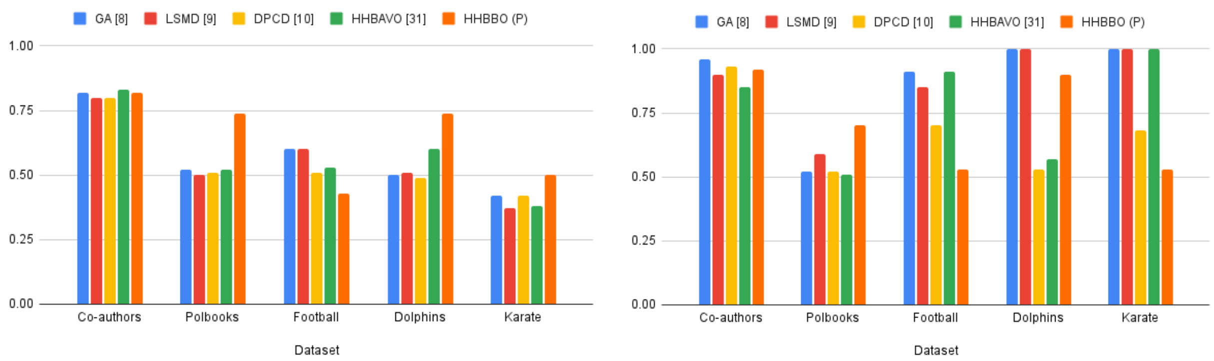

Figure 3. NMI values with the Algorithms and after network compression

Table 3: Representation of Moduriality based on the Algorithms

Networks	Name of the techniques									
	GA [8]		LSMD [9]		DPCD [10]		HHBAVO [31]		HHBBO (P)	
	Q	Mc	Q	Mc	Q	Mc	Q	Mc	Q	Mc
CNS	0.92	129	0.84	131	0.85	122	0.77	121	0.77	126
AMAZON	0.62	57	0.72	59	0.56	56	0.54	54	0.62	56

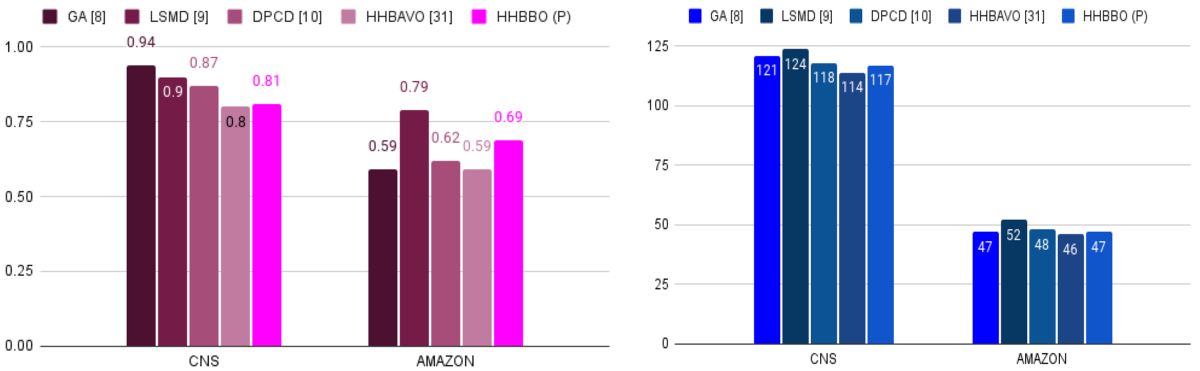

Figure 4. Representation of Moduralty based on the Algorithms

Table 4: Comparison NMI values with the Algorithms

Networks	Name of the Techniques									
	GA [8]		LSMD [9]		DPCD [10]		HHBAVO [31]		HHBBO (P)	
	NMI	Mc	NMI	Mc	NMI	Mc	NMI	Mc	NMI	Mc
CNS	0.94	121	0.90	124	0.87	118	0.8	114	0.81	117
AMAZON	0.59	47	0.79	52	0.62	48	0.59	46	0.69	47

Table 5: Parameter settings

Parameter	Values	Description
Mk	0.1N	Maximum degree
K	15	Mean degree
µ	{0.1, 0.2, …., 0.6}	Mixing parameter
N	1000, 5000, 10,000	Number of nodes
maxC	0.1N	Maximum community size
minC	10, for N=1000 20, for N=5000 25, for N=10000	Minimum community size
τ_2	1, 2	Community size distribution in the minus exponent
τ_1	2	Degree sequence in the minus exponent

The outcomes of LFR networks are shown in Figures 5(a) and 5(b), along with the community distribution $\tau_2 = 1$. Nevertheless, the size of the network is distinct. Based on both of these figures, it is clear that our proposed method is capable of producing results that are superior to those obtained by other existing works, such as GA, LSMD, DPCD, ICLA, and HHBAVO[32, 33, 34]. Despite the fact that the size of the network is different, the NMI values shown by our proposed work are better. In the meantime, the results of LFR networks with different community distributions as well as network sizes are shown in figures 5 (c) and 5 (d), respectively. It can be seen in figure 5 that the proposed approach achieves better values than the other approaches overall.

6.4 Effect of Compression Ratio

During the process of compression, the values of non-similarity that come up first are chosen and given the value. There is a direct proportional relationship between the compression ratio and the reduced scale. The reduced scale is dependent on the ratio. Figure 6 presents the results of the comparative analysis. In the event that the reduced scale value is found to be greater, the compressed network size will be found to be mitigated. This in no way indicates that the output will be optimized when it is finally completed. Figure 7 depicts the relationship between the NMI value and various compression ratios. μ=0.5, the best possible value is obtained.

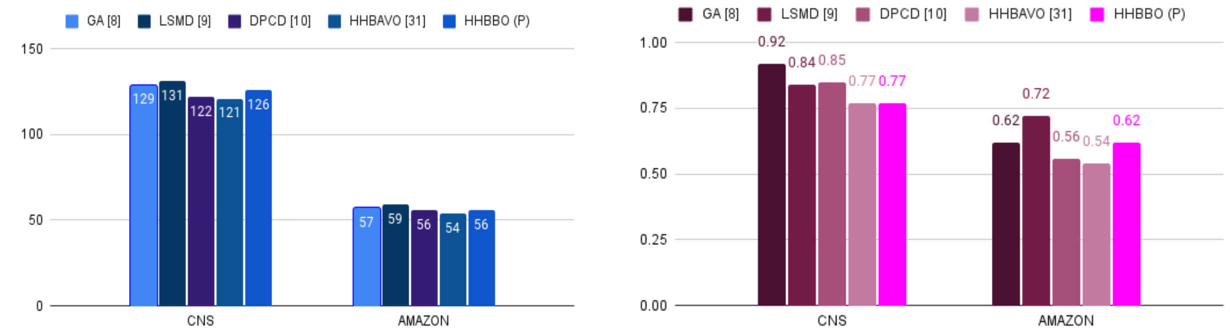

Figure 5. NMI values with the Algorithms – All Methods

6.5 Execution Time analysis

Table 6 displays the amount of time it takes for the various compression network formation algorithms to carry out their tasks. The GA, LSMD, DPCD, ICLA, HHBAVO, and the Proposed method are the algorithms that are being used for the comparison. The proposed method has a lower CPU time of 25 & 32 Seconds for the Synthetic Datases, which is relatively low when compared to other algorithms, and the GSO offers the next best performance.

Table 6. Execution time comparison for different evolutionary algorithms

	Processing Time (sec)	
Algorithm	CNS	AD
GA [8]	41	47
LSMD [9]	34	56
DPCD [10]	30	31
HHBAVO [31]	36	39
HHBBO (P)	25	34

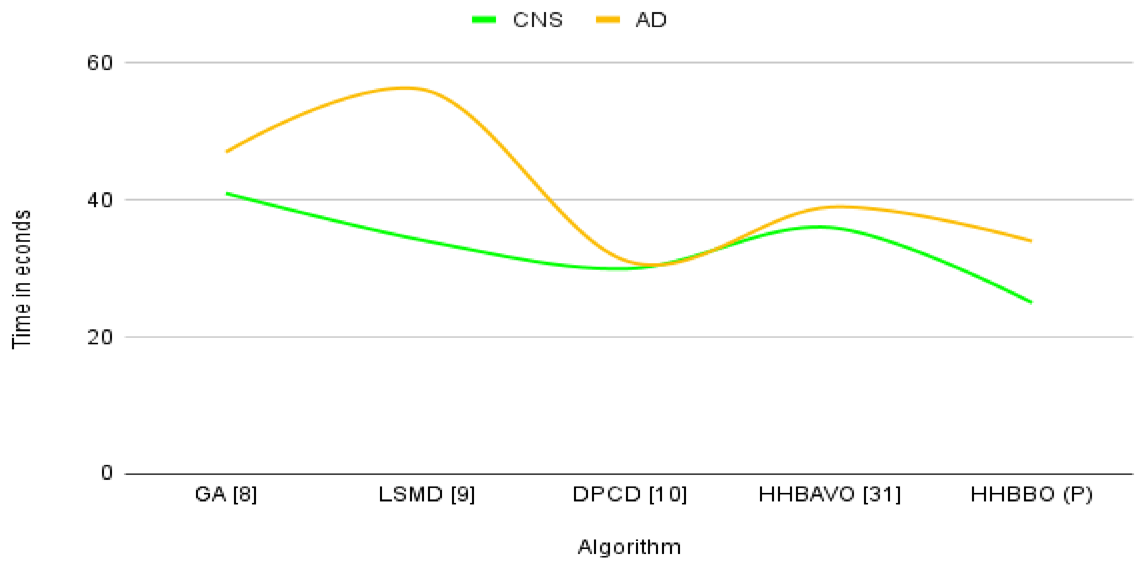

Figure 6: Execution time comparison for different evolutionary algorithms

7. Conclusion

For the purpose of locating communities within social networks, the authors of this study proposed a multi - heuristics optimization method algorithm. Exploring the topologies of the network brings the scale of the network down to a more manageable level. We have utilized the HHBBO algorithm based on the decomposition of the problem. The superiority of the proposed algorithm is demonstrated with the help of examples from the real world. In experimental investigations, methods such as GA, LSMD, DPCD, ICLA,HHBAVO and others that are considered state-of-the-art, as well as the proposed method, are utilized. The performance of the proposed method is validated by performance metrics such as modularity (Q) and Normalized Mutual Information (NMI), which are calculated using the most recent and cutting-edge techniques. When measured against both the artificial network and the actual network, the methodology that was proposed demonstrated superior performance. In the not too distant future, one of our goals is to implement the suggested model into an e-commerce-based recommendation system in order to boost sales and provide higher levels of user satisfaction.

Conflict of Interest

The authors declare that they have no conflict of interest.

Availability of data and material

Data sharing is not applicable to this article as no new data were created or analyzed in this study.

References

Bedi P , Sharma C. Community detection in social networks. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery,2016; 6(3): 115-135.
Pizzuti C. Ga-net: A genetic algorithm for community detection in social networks. In International conference on parallel problem solving from nature, 2008, September (pp. 1081-1090). Springer, Berlin, Heidelberg,.
Xie J , Szymanski BK. Towards linear time overlapping community detection in social networks. In Pacific-Asia Conference on Knowledge Discovery and Data Mining, 2012, May (pp. 25-36). Springer, Berlin, Heidelberg.
Pizzuti C. Community detection in social networks with genetic algorithms. In Proceedings of the 10th annual conference on Genetic and evolutionary computation, 2008, July (pp. 1137-1138).
Wang M, Wang C, Yu JX , Zhang J. Community detection in social networks: an in-depth benchmarking study with a procedure-oriented framework. Proceedings of the VLDB Endowment,2015; 8(10): 998-1009.
Ahajjam S, El Haddad M, Badir H. A new scalable leader-community detection approach for community detection in social networks. Social Networks,2018; 54: 41-49.
Kanavos A, Perikos I, Hatzilygeroudis I, Tsakalidis A . Emotional community detection in social networks. Computers & Electrical Engineering,2018; 65: 449-460.
Behera RK, Naik D, Rath SK, Dharavath R. Genetic algorithm-based community detection in large-scale social networks. Neural Computing and Applications, 2020; 32(13): 9649-9665.
Bouyer A, Roghani H. LSMD: A fast and robust local community detection starting from low degree nodes in social networks. Future Generation Computer Systems, 2020; 113: 41-57.
Ji T, Luo C, Guo Y, Wang Q, Yu L, Li P. Community detection in online social networks: a differentially private and parsimonious approach. IEEE transactions on computational social systems, 2020; 7(1): 151-163.
Khomami MMD, Rezvanian A, Saghiri AM, Meybodi MR. SIG-CLA: a significant community detection based on cellular learning automata. In 2020 8th Iranian Joint Congress on Fuzzy and intelligent Systems, 2020, September (CFIS) (pp. 039-044). IEEE.
Raj ED, Manogaran G, Srivastava G, Wu . Information granulation-based community detection for social networks. IEEE Transactions on Computational Social Systems, 2020; 8(1): 122-133.
Wu L, Zhang Q, Chen CH, Guo K , Wang D. Deep learning techniques for community detection in social networks. IEEE Access,2020; 8 : 96016-96026.
Zhang X, Tian Y, Cheng R , Jin Y. A decision variable clustering-based evolutionary algorithm for large-scale many-objective optimization. IEEE Transactions on Evolutionary Computation,2016; 22(1): 97-112.
Gong M, Cai Q, Chen X , Ma L. Complex network clustering by multiobjective discrete particle swarm optimization based on decomposition. IEEE Transactions on evolutionary computation, 2013; 18(1): 82-97.
Angelini L, Boccaletti S, Marinazzo D, Pellicoro M , Stramaglia S. Identification of network modules by optimization of ratio association. Chaos: An Interdisciplinary Journal of Nonlinear Science, 2007; 17(2): 023114.
Balakrishnan K, Dhanalakshmi R , Mahadeo Khaire U A venture to analyse stable feature selection employing augmented marine predator algorithm based on opposition-based learning. Expert Systems, p.e12816.
Abdollahzadeh B, Gharehchopogh FS , Mirjalili S. African vultures optimization algorithm: A new nature-inspired metaheuristic algorithm for global optimization problems. Computers & Industrial Engineering,2021; 158:107408.
Hashim FA, Houssein EH, Hussain K, Mabrouk M. , Al-Atabany W. Honey badger algorithm: new metaheuristic algorithm for solving optimization problems. Mathematics and Computers in Simulation,2022; 192: 84-110.
Adelberger EG, Heckel BR, Hoedl S, Hoyle CD, Kapner DJ , Upadhye A. Particle-physics implications of a recent test of the gravitational inverse-square law. Physical review letters, 2007; 98(13): 131104.
Hughes DT . Aspects of cardioid processing. SACLANT UNDERSEA RESEARCH CENTRE LA SPEZIA (ITALY),2000.
Song A, Liu Y, Wu Z, Zhai M , Luo J. A local random walk model for complex networks based on discriminative feature combinations. Expert Systems with Applications,2019;

118: 329-339.
Adamic LA, Adar E. Friends and neighbors on the web. Social networks, 2003; 25(3): 211-230.
Danon L, Diaz-Guilera A, Duch J, Arenas A. Comparing community structure identification. Journal of statistical mechanics: Theory and experiment, 2005(09): 09008.
Newman ME. Fast algorithm for detecting community structure in networks. Physical review E, 2004; 69(6): 066133.
Akan T, Agahian S , Dehkharghani R. Binbro: Binary battle royale optimizer algorithm. Expert Systems with Applications, 2022; 195:116599.
Alsaidan I, Shaheen MA, Hasanien HM, Alaraj M, Alnafisah AS. A PEMFC model optimization using the enhanced bald eagle algorithm. Ain Shams Engineering Journal, 2022; 13(6): 101749.
Bernal E, Lagunes ML, Castillo O, Soria J, Valdez F. Optimization of type-2 fuzzy logic controller design using the GSO and FA algorithms. International Journal of Fuzzy Systems, 2021; 23(1): 42-57.
Sharma A, Dasgotra A, Tiwari SK, Sharma A, Jately V, Azzopardi B. Parameter extraction of photovoltaic module using tunicate swarm algorithm. Electronics, 2021; 10(8): 878.
Shi C, Zhong C, Yan Z, Cai Y, Wu, B. A multi-objective approach for community detection in complex network. In IEEE Congress on Evolutionary Computation, 2010, July (pp. 1-8). IEEE.
Sankara Nayaki, Sudeep Elayidom, Rajesh R, A Novel Compression based Community Detection Approach using Hybrid Honey Badger African Vulture optimizer for Online Social Networks, Concurrency & Computation: Practice and Experience, 2022
Kanō T (1992) The last ape: pygmy chimpanzee behavior and ecology, Stanford University Press, Stanford, CA
DeWaal FB (1995) Bonobo sex and society. Sci Am 272(3):82–88
Amit Kumar Das, Dilip Kumar Pratihar, "Bonobo optimizer (BO): an intelligent heuristic with self-adjusting parameters over continuous spaces and its applications to engineering problems", Applied Intelligence, Springe, 2021