Received: March 28, 2024. Revised: May 7, 2024.

International Journal of
Intelligent Engineering & Systems

http://www.inass.org/

Community Detection in Complex Network Using Improved Hybrid Harris Hawk and Coot Bird Optimization Algorithm

Yashaswini Doddamane Kenchappa[1]* Mohan Kumar Antharasanahalli Venkataramaiah[1]
Shankara Gowda Solmangowda Rudregowda[2]

[1]*Department of Computer Science and Engineering, PES University, RR Campus, Bangalore, India*
[2]*Department of Information Science and Engineering, Don Bosco Institute of Technology, Bangalore, India*
* Corresponding author's Email: yashaswinidk5@gmail.com

Abstract: Community detection in complex networks plays a pivotal role in identifying the structural composition of nodes, enabling us to pinpoint key topological features critical for a multitude of applications. Usually, the complex networks are in graphical format in which graph nodes denote objects and edges denote a connection between two things. The existing techniques have limitations such as early convergence and local optima issues. To overcome this issue, this manuscript proposed an Improved Hybrid Harris Hawk and Coot Bird Optimization (IHHHCBO) algorithm in community detection. The optimum separate movement of CBO is incorporated in HHO population initialization to strengthen both optimizations. It is employed to initialize the population for enhancing the position of diversity and change with other individuals. The Karate, Dolphin, Football and Political Books datasets are considered for assessing the IHHHCBO performance. The Ensemble Mutation Strategy (EMS) is developed to produce mutant candidate locations which enhances the exploration and population diversity of optimization. The Normalized Mutual Information (NMI), and Modularity (Q) are considered fitness functions in this research. The IHHHCBO performance is estimated through metrics like NMI, Q, f1-score and accuracy. The IHHHCBO reaches better accuracy of 1, 1, 0.998 and 0.883 for Karate, Dolphin, Football and Political Books respectively which is better when compared to existing algorithms like Core Structure Extraction Algorithm (CSEA), and Modified Crossover Opposition-based Genetic Algorithm (MCOBGA).

Keywords: Community detection, Complex networks, Coot bird optimization, Modularity, Normalized mutual information.

1. Introduction

In recent years, the development of social networks has authorized users to transfer data around the world [1]. Community detection in difficult networks presents the group searching according to graphs or networks [2]. In a community, nodes are expected to have mutual properties which becomes a challenging problem in optimization. The social network serves as a dynamic platform facilitating communication and interaction among individuals, leveraging the internet as a medium to connect with diverse communities [3, 4]. The network with huge has become a rapid growth in social networks and users [5]. To search and explore complex data or networks, huge data are produced for determining hidden patterns [6]. Community detection is accomplished by separating complex networks into various groups of nodes [7]. It is one of the basic challenges in communication networks which has a consequence in examining data for utilizing a wide array of application fields like social networks, Machine Learning (ML) and medical science [8]. Community detection has garnered significant attention within communities, defined as groups of nodes closely interconnected with each other. [9].

The network community denotes the network sets through relatively close internal networks and sparse external networks [10]. The community structure is a connection relation created through the community. The community member has high similarity in the same community but there are clear variances between members in various communities [11]. Hence, the community structures demonstrate the variances and connections among network members and mine the structure of the community network which is important for the network [12, 13]. Community detection assists the people in accurately recognizing the network community, improving the perception of network structure, and clasp connection among network nodes [14]. Furthermore, it promotes the expansion of numerous intelligent services like emerging accurate marketing tools, drug target detection, and social network data mining for trend prediction [15]. To achieve desirable outputs on these networks, particular network analysis tasks are required like link prediction, influence maximization and communication detection. The evolutionary-based optimization is utilized as global search abilities which is utilized for community detection [16, 17]. However, it performed with less efficiency when managing complex and huge-dimensional data retrieved from huge communication networks [18]. The paper's contribution is summarized as follows:

- The Improved Hybrid Harris Hawk and Coot Bird Optimization are proposed in this research for detecting communities in complex networks in which modularity and NMI are considered fitness functions.
- The optimum separate movement of CBO is incorporated in HHO population initialization to strengthen both optimizations. It is employed to initialize the population for enhancing the position of diversity and change with other individuals.
- The proposed IHHHCBO algorithm diminishes the probability of local optimum issues and enhances the search space and convergence speed.

This manuscript is arranged as follows: Section 2 provides a literature review, and Section 3 provides a proposed method. Section 4 provides experimental results and discussion and section 5 provides a conclusion and future work of this manuscript.

2. Literature review

Some recent literature work for community detection in complex networks is defined in this section.

Saeid Talebpour Shishavan and Farhad Soleimanian Gharehchopogh [19] implemented an Improved Cuckoo Search Optimization with a Genetic Algorithm (ICSO with GA) for community detection in networks. The GA was used to enhance the speed and accuracy of CSO dynamically which adjusted the populations according to the number of exploration and exploitation. The ICSO with GA was employed to overcome the local optima, premature, and delayed convergence issues. However, by employing ICSO with GA detecting an overlapping community structure was difficult.

Mohammed Nasser Al-Andoli [20] introduced a Distributed Parallel DL with hybrid Backpropagation-Particle Swarm Optimization (BP-PSO) for communication detection. The BP was utilized to locally optimize the DL within every local network and PSO was utilized to enhance the BP in every communication network. The introduced model tackles the issues of scalability and inefficiency which achieves optimum solution through global-local search. However, the performance was affected due to the vanishing gradient and premature convergence issues.

Mahdi Zarezadeh [21] presented a Distance-based Peripheral Nodes Label Propagation (DPNLP) for detecting communities. Primarily, the core nodes and labels are detected and scattered over neighbors which builds primary communities. Then, peripheral labels are employed in the integration of local structures. Lastly, community structures are mined through the degrees of 1 and 2 at the last stage. The presented model enhanced the detection quality however, random nodes affect the accuracy which diminishes the performance.

Farhad Soleimanian Gharehchopogh [22] developed an Improved Harris Hawks Optimization Algorithm with Multi-Strategy (IHHO with MS) for community detection. The MS includes Levy Flight (LF), Opposition-Based Learning (OBL), and Chaotic Map (CM) for balancing the exploration and exploitation of communication networks. The developed model strengthens the exploration and exploitation of optimization techniques. However, it has a huge possibility to stuck in local optima problems.

Rong Fei [23] suggested a CSEA through a variational autoencoder for community detection. Initially, the K-truss algorithm was employed for finding core structure data in the network and a similarity matrix was produced through similarity mapping integrated with local data. Then, the variational autoencoder was employed to reduce and extract the similarity matrix dimension which contains core structure and low-dimensional feature matrix. Lastly, k-means clustering was employed to attain structural information about the community. However, the suggested model suffers from

Received: March 28, 2024. Revised: May 7, 2024.

diminished search space which affects the performance.

Chuanwei Li [24] presented a Density Peak Clustering and Label Propagation (SD-LPA) for detecting communities. Initially, local density estimation was developed to find the community network that enhances the community partition quality. Then, the label update order was defined by calculating the data transmission of nodes, and numerous candidate label solutions were generated to enhance the robustness. However, the presented model enhances the local optimum issues and reduces the convergence speed.

Harish Kumar Shakya [25] developed a MOCBGA for detecting communities in complex networks. The developed model employed enhanced crossover and opposition-based initialization over GA for detecting communities. The enhanced crossover transmits data and population initialization over opposition-based learning ensures community detection. The developed model obtains huge convergence speed however, non-parametric test of soft computing method result validation was not confirmed.

The existing techniques have vanishing gradient and premature convergence issues, and random nodes are affecting the accuracy which diminishes the performance. It has a huge possibility to stuck in local optima problems and suffers from diminished search space which affects the performance. The non-parametric test of soft computing method result validation was not confirmed.

3. Proposed methodology

In this manuscript, an IHHHCBO algorithm is proposed for community detection in complex networks. The Karate, Dolphin, Football, and Political Books datasets are considered for assessing the IHHHCBO performance. The optimum individual movement of CBO is incorporated into the population initialization of HHO to strengthen both optimizations. The EMS is developed to produce mutant candidate locations which enhances the exploration and population diversity of optimization. The Karate, Dolphin, Football, and Political Books datasets are considered for assessing the IHHHCBO performance.

3.1 Dataset

In this manuscript, four types of datasets are employed which are real-life networks like Karate [26], Dolphins [27], Football [28] and Political Books [29]. These four datasets have real communities, number of nodes and edges that are required for community detection. Table 1 presents the dataset description.

Table 1. Dataset description

Network	Karate	Dolphin	Football	Political books
Real community	2	2	12	3
No. of nodes	34	64	115	105
No. of edges	78	159	616	613

3.2 Objective function for community detection

Initially, the in-degree and out-degree of node criteria [30] are defined and show how to use these in a hybrid approach for finding communities which is given in Eq. (1).

$$\mu_i = \frac{1}{|C|} k_i^{in}(C) \tag{1}$$

Where, μ_i is a fraction of edges that integrates node i to other nodes in the community, $|C|$ is a Cardinality of C .The power mean of community order r is signified as $M(C)$ which is given in Eq. (2).

$$M(C) = \frac{\sum_{i \in C} (\mu_i)^r}{|C|} \tag{2}$$

If the r enhances the node weights are connected with other nodes and minimizes the node weights that have few connections within the community. The volume v_C of community is signified as the number of edges that connect vertices in a community. The score of community is signified as $score(C) = M(C) \times v_C$. The score is considered as a fraction and inter-connection between nodes and the number of interconnections contained in the community module. The community score of clustering $\{c_1, \dots, c_k\}$ of the network is given in Eq. (3).

$$CS = \sum_{i=1}^{k} score(C_i) \tag{3}$$

The community detection is formulated as CS maximization issues. The community fitness is given in Eq. (4).

$$P(C) = \sum_{i \in S} \frac{k_i^{in}(C)}{(k_i^{in}(C) + k_i^{out}(C))^{\alpha}} \tag{4}$$

Where, $k_i^{in}(C)$ and $k_i^{out}(C)$ designate the internal and external node degree according to the community, α is an actual parameter that controls community size and is used to find the communities.

3.3 Harris hawks optimization

In community detection, the optimization algorithm is utilized because it finds the distance between maximum and minimum nodes. The HHO is a nature-inspired optimization algorithm that mimics Harris hawk behavior. It contains three phases such as exploration, transmission from exploration to exploitation and exploitation.

3.3.1. Exploration

In this stage, Harris hawks are located at certain spaces randomly and it detects the prey and then selects among two policies through equal possibility of hunting actions. The Harris hawk position update is given in Eq. (5).

$$X(t+1) = \begin{cases} X_{rand} - r_1 \left| X_{rand} - 2r_2X(t) \right| & q \geq 0.5 \\ \left(X_{prey}(t) - X_m(t)\right) - r_3\left(LB + r_4(UB - LB)\right), & q < 0.5 \end{cases} \quad (5)$$

Where, $X(t)$ and $X(t+1)$ are Harris hawk position vectors in the present and next iterations, t is a present iteration, X_{prey} is a prey position, $X_{rand}(t)$ is a position vector of random individuals in the present population. The r_1, r_2, r_3 and r_4 and q are random numbers between $[0, 1]$, UB and LB are upper and lower bound variables, $X_m(t)$ is an average position of the population in every hawk.

3.3.2. The transition from exploration to exploitation

In this stage, the transition from exploration to exploitation is based on the escaping energy of prey (E) that is given in Eq. (6).

$$E = 2E_0\left(1 - \frac{t}{T}\right) \quad (6)$$

Where, E_0 is a random number between $[-1, 1]$, t and T are present and the maximum number of nodes. If $|E| \geq 1$, the hawk continues the prey search in the target area, determined as the exploration stage. If $|E| < 1$, hawk prey hunting is started in the previous step and arrives in the exploitation phase.

3.3.3. Exploitation

In this stage, the search agent preserves to exploit the solution attained through optimum solutions. According to various hunting processes, four probable approaches soft besiege, hard besiege, soft and hard besiege with progressive rapid dives are employed. The r is a random number among $[0, 1]$ if $r < 0.5$, the prey has escaped over the complex node if $r \geq 0.5$ means the prey has failed to escape. According to the r and E values, four strategies are employed that is presented below:

3.3.4. Soft besiege

It is accomplished when $|E| \geq 0.5$ and $r \geq 0.5$, Harris hawk position updating is given in Eq. (7).

$$X(t+1) = \Delta X(t) - E|JX_{prey}(t) - X(t)| \quad (7)$$

Where, $X(t)$ is the distance between position and prey, J is a random jump prey intensity.

3.3.5. Hard besiege

The hawk considered it when $|E| < 0.5$ and $r \geq 0.5$ which is mathematically expressed in Eq. (8).

$$X(t+1) = X_{prey}(t) - E|\Delta X(t)| \quad (8)$$

3.3.6. Soft besiege with progressive rapid dives

The hawk will considered soft besiege with progressive rapid dives when $|E| \geq 0.5$ and $r < 0.5$ which is mathematically expressed in Eqs. (8-9).

$$X(t+1) = \begin{cases} Y, & if\ F(Y) < F(X(t)) \\ Z, & if\ F(Z) < F(X(t)) \end{cases} \quad (9)$$

Where D is a problem dimension, S is a random vector within the size of $1 \times D$, $F(.)$ is an objective function, the selected position among Y and Z is considered as next position.

3.3.7. Hard besiege with progressive rapid dives

The hawk considered it when $|E| < 0.5$ and $r < 0.5$ which is mathematically expressed in Eq. (10).

$$YX(t+1) = \begin{cases} Y, & if\ F(Y) < F(X(t)) \\ Z, & if\ F(Z) < F(X(t)) \end{cases} \quad (10)$$

Where, D is a problem dimension, S is a random vector within the size of $1 \times D$, $F(.)$ is an objective function, $X_m(t)$ is the average position of every hawk in the population.

3.4 Coot bird optimization

The CBO is a population-based bio-inspired optimization algorithm that mimics the behavior of coots on the surface of water. In CBO, four various regular and irregular movements are implemented such as random, chain, and position adjustment based on group leaders and leader movements.

3.4.1. Random movement:

Here, coot followers are stimulated to random positions to explore different hunt areas. The random position Q is given in Eq. (11).

$$Q = rand(1, D) \times (UB - LB) + LB \tag{11}$$

Where D is a problem dimension, UB and LB are upper and lower bounds which produce better searchability and capability to escape from local optimum.

3.4.2. Chain movement

The two individuals average is employed to accomplish the chain movements. The coot followers' new position is estimated in Eq. (12).

$$X_i(t+1) = \frac{1}{2} \times \left(X_{i-1}(t) + X_i(t)\right) \tag{12}$$

Where, $X_{i-1}(t)$ is a $i-1$th follower position in the present iteration t.

3.4.3. Position adjustment based on group leaders

Commonly, the group is managed through group leaders and the rest of the coot followers are required to adjust their position according to leaders and move towards them. However, the issue is found in that every coot needs to update the position based on the leader and the follower's next position is estimated according to the designed leader k as shown in Eq. (13).

$$X_i(t+1) = LeaderX_k(t) + 2 \times r_5 \times \cos(2R\pi) \times LeaderX_k(t) - X_i(t)) \tag{13}$$

Where, $LeaderX_k(t)$ is a position of designated leader, r_5 and R are random numbers among $[0,\ 1]$ and $[-1,\ 1]$.

3.4.4. Leader movement

The group is oriented into the optimum area, therefore in a few cases, leaders need to present the optimum position to hunt the best one. The leader position update formula is given in Eqs. (14) and (15).

$$LeaderX_i(t+1) = \begin{cases} B \times r_6 \times \cos(2R\pi) \times \\ \left(gBest(t) - LeaderX_i(t)\right) \\ +gBest(t), r_7 < 0.5 \\ B \times r_6 \times \cos(2R\pi) \times \\ \left(gBest(t) - LeaderX_i(t)\right) \\ +gBest(t), r_7 \geq 0.5 \end{cases} \tag{14}$$

$$B = 2 - t \times \left(\frac{1}{T}\right) \tag{15}$$

The $gBest$ is a present optimum position, r_6, r_7 and R are random numbers between $[0,\ 1]$ and $[-1,\ 1]$. $B \times r_6$ produces important stochastic movements to assist the model in eliminating the local optimum issues. $\cos(2R\pi)$ is an individual best distance search by various radii to attain a superior position, t and T are present and the maximum number of nodes.

3.5 Ensemble mutation strategy

The set variation is improvised in a hybrid optimization algorithm that produces diverse individuals for enhancing the global search ability of hybrid optimization. The EMS is mathematically expressed in Eqs. (16-18),

$$V_{i1} = \begin{cases} X_{R1} + F_1 \times (X_{R2} - X_{R3}),\ r_8 < C_1 \\ X_i, \qquad r_8 \geq C_1 \end{cases} \tag{16}$$

$$V_{i2} = \begin{cases} X_{R4} + F_2 \times (X_{R5} - X_{R6}) + \\ F_2 \times (X_{R7} - X_{R8})\ r_9 < C_2 \\ X_i, \qquad r_9 \geq C_2 \end{cases} \tag{17}$$

$$V_{i3} = \begin{cases} X_i + F_3 \times (X_{R9} - X_i) + \\ F_3 \times (X_{R10} - X_{R11})\ r_{10} < C_3 \\ X_i, \qquad r_{10} \geq C_3 \end{cases} \tag{18}$$

Where, V_{i1}, V_{i2} and V_{i3} are recently produced mutant candidate locations at ith position of search agent, $R1 \sim R11$ are various exponents within the range of $[1,\ N]$. F_1, F_2 and F_3 are scale factors, C_1, C_2 and C_3 are crossover rate, $r_{10} \sim r_{12}$ are random

Received: March 28, 2024. Revised: May 7, 2024.

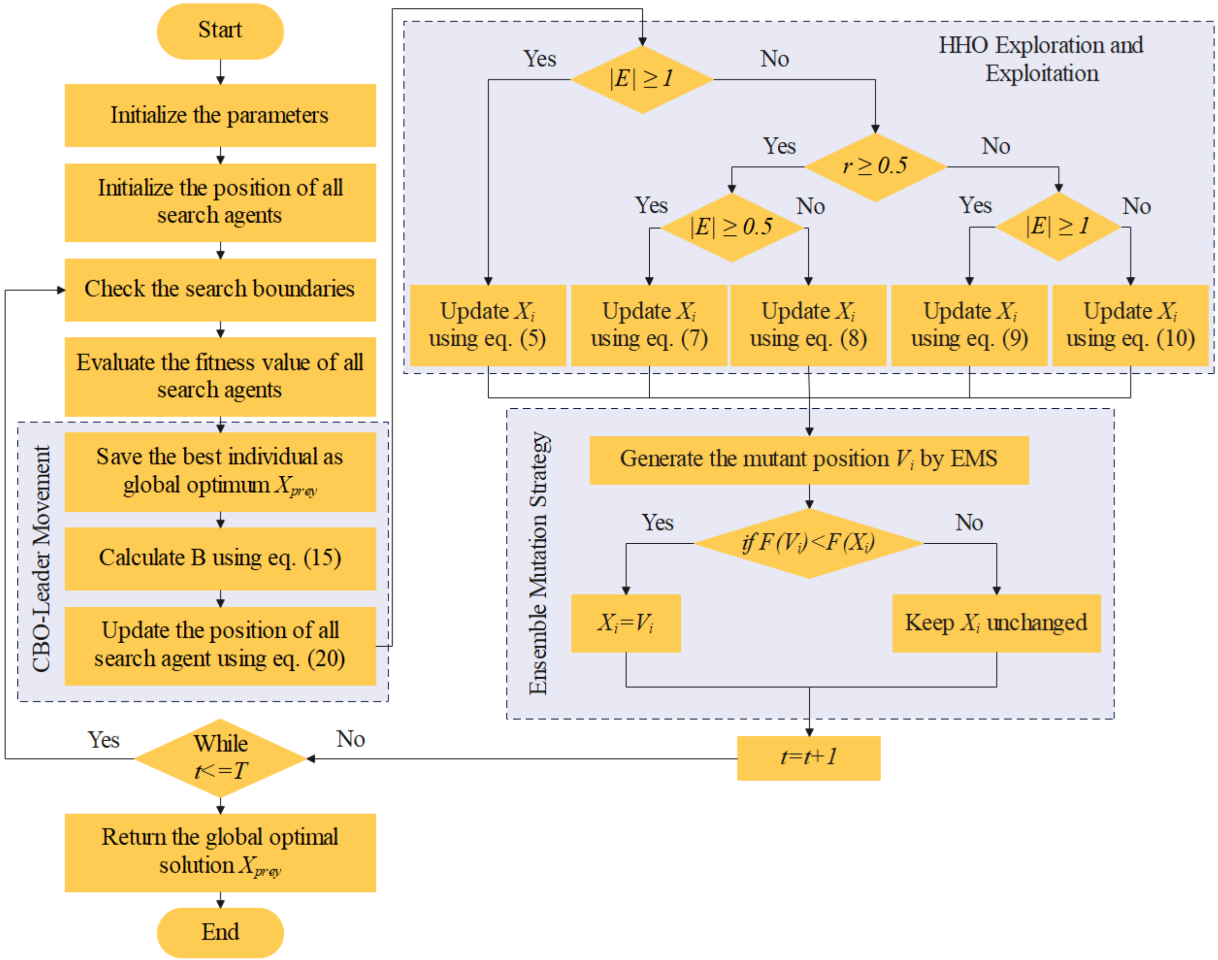

Figure. 1 Flowchart of IHHHCBO algorithm

numbers among $[0,1]$. After V_{i1}, V_{i2} and V_{i3} produced, the best position V_i with the smallest fitness value is designated to compare with the original X_i fitness and then, the best fitness is kept as a new X_i to contribute the iterative estimation which is given in Eq. (19).

$$X_i = \begin{cases} V_i, & if\ F(V_i) < F(X_i) \\ X_i, & otherwise \end{cases} \tag{19}$$

Where, V_i is a mutant candidate location and $F(.)$ is a cost function.

3.6 Improved hybrid HHO and CBO

The traditional HHO and CHO algorithm is integrated with a swarm intelligence algorithm called IHHHCBO. Moreover, the EMS is included in the hybrid optimization algorithm for enhancing the search performance. The HHO contains three phases such as exploration, transmission from exploration to exploitation, and exploitation phase. Then, it moves from global and local search according to prey energy. In exploitation, the energy and probability of the prey are defined in four various techniques such as soft besiege, hard besiege, soft and hard besiege with advanced fast dives. If prey escapes, leader movement is included to make sure the algorithm is jump into local optima. Low population diversity in the exploration phase fails the global search capability of HHO. It takes a long time to attain optimum global solutions and enhance the probability of local optima issues. In the CBO algorithm, the population is randomly separated into coot leaders and followers. The leaders ensure the data in search space and the tendency for exploration which leads followers into the target area. It contains separate random, chain, and optimum individual movements. CBO is selected random position in the search space which assists in overcoming local optima issues. The chain movement enables to enhancement of algorithm accuracy when the position is focused. In the leader movement of the optimum individual, coot followers' position is changed by the corresponding leader position. Fig. 1 shows the flowchart of the IHHHCBO algorithm.

In this manuscript, the optimum separate movement of CBO is incorporated in HHO population initialization to strengthen both optimizations. In CBO, optimum individual leader movement, $B \times r_8$ produces a huge number of randomness. It is employed to initialize the population for enhancing the position of diversity and change with other individuals. From Eq. (14), the population initialization is derivative and given in Eq. (20).

$$Xnew_i(t+1) = \begin{cases} B \times r_6 \times \cos(2R\pi) \times \left(X_{prey}(t) - X_i(t)\right) \\ \quad +X_{prey}(t), r_7 < 0.5 \\ B \times r_6 \times \cos(2R\pi) \times \left(X_{prey}(t) - X_i(t)\right) \\ \quad +X_{prey}(t), r_7 \geq 0.5 \end{cases} \tag{20}$$

Where, $Xnew_i(t+1)$ is a position of ith search agent at iteration $(t+1)$, $X_i(t)$ is a position of iteration, $X_{prey}(t)$ is a present optimum solution. After generating the position of all search agents, the fitness of every new position is matched by the actual position. The candidate position through high fitness is chosen as a position update for HHHCBO. The EMS has numerous mutation operators that produce three various candidate positions based on Eqs. (16)-(18). Then, the position with the smallest fitness is chosen to compare the fitness value of actual positions, and the position with high fitness is chosen as the new position. It diminishes the probability of local optimum issues, and enhances the search space and convergence speed.

3.6.1. Euclidean distance

The Euclidean distance is used to estimate the distance between two nodes in community detection. The formula for calculating distance is given in Eq. (21).

$$d_{ij} = \sqrt{\sum_{k=1}^{n} \left(x_{ik} - y_{jk}\right)^2} \tag{21}$$

Where, d is the distance between node i and j, k is a symbol of every data, n is a number of nodes.

3.6.2. Fitness function

The fitness function denotes the optimization problem which needs to be enhanced or diminished through the IHHHCBO. In community detection, the NMI and modularity are two important optimization functions in metaheuristic algorithms. The NMI is mathematically expressed in Eq. (22).

$$NMI = \frac{-2\sum_{u=1}^{M_x}\sum_{v=1}^{M_y} M_{uv} \cdot \left(\frac{M_{uv} \cdot n}{M_u M_v}\right)}{\sum_{u=1}^{M_x} M_u log\left(\frac{M_u}{n}\right) + \sum_{v=1}^{M_y} M_v log\left(\frac{M_v}{n}\right)} \tag{22}$$

The n is a number of nodes, u and v are dual communities by the network, M is a confusion matrix, M_x and M_y is a number of communities in x and y sections respectively. If the mean is higher, the similarity among x and y is important. If the NMI is equal, the x and y are accurately similar. The modularity is given in Eq. (23).

$$Q = \frac{1}{2Y} \sum \left(m_{ij} - \frac{K_i K_j}{2Y}\right) \delta(i,j) \tag{23}$$

Where, Y is the total number of edges in the network, i and j are indexed to indicate number of nodes, m_{ij} is a value of row i and column j in the matrix and $\delta(i,j)$ is a connection between two nodes.

4. Experimental result

The proposed model is stimulated by python environment with a system configuration of an i7 processor, 16GB RAM, and a Windows 10 operating system. Normalized Mutual Information (NMI), Modularity (Q), f1-score and accuracy. Table 2 shows the initialization of parameters.

NMI: It is employed to estimate the model performance on datasets where real community structures are accessible. Its values range between $[0, 1]$ and the best score is attained whether found communities are equivalent to actual communities which is mathematically expressed in Eq. (22).

Modularity: It is a performance metric employed for checking the main quality in community detection which is mathematically expressed in Eq. (23).

F1-score: It denotes result quality according to harmonic mean among precision and recall which is presented in Eqs. (24)-(26).

$$F1 - \text{score} = 2 \times \frac{Precision \times Recall}{Precision + Recall} \tag{24}$$

$$Precision = \frac{TP}{TP + FP} \tag{25}$$

$$Recall = \frac{TP}{TP + FN} \tag{26}$$

Table 2. Parameter initialization

Parameter	Range
Population size	10-50
Exploration Rate	0.1-0.5
Exploitation Rate	0.1-0.5

Table 3. Performance of IHHHCBO for Karate dataset

Method	NMI	Q	F1-score	Accuracy
HHO	0.679	0.387	0.626	0.652
CBO	0.754	0.416	0.735	0.768
HHHCBO	0.892	0.483	0.818	0.865
IHHHCBO	1	0.521	1	1

Table 4. Performance of IHHHCBO for Dolphin dataset

Method	NMI	Q	F1-score	Accuracy
HHO	0.695	0.437	0.653	0.675
CBO	0.759	0.476	0.715	0.738
HHHCBO	0.836	0.538	0.769	0.817
IHHHCBO	1	0.589	1	1

Where, TP, FP and FN are true positive, false positive and false negative.

Accuracy: It is employed to achieve the predicted cluster allocation from actual community data which ranges between $[0, 1]$. It is mathematically expressed in Eq. (27).

$$Accuracy = \frac{\sum_{i=1}^{N} \delta(R_i.Map(C_i))}{N} \tag{27}$$

Where, $Map(C_i)$ is a mapping function, $\delta(R_i.Map(C_i))$ is a Crocker delta.

4.1 Quantitative and qualitative analysis

The IHHHCBO performance is estimated through the measures of NMI, Q, f1-score and accuracy. The Karate, Dolphin, Football, and Political Books datasets are taken for analyzing the optimization performance. Tables 3, 4, 5 and 6 show the IHHHCBO performance for Karate, Dolphin, Football and Political Books datasets correspondingly.

Table 3 presents the proposed IHHHCBO performance for the Karate dataset through NMI, Q, f1-score and accuracy. The HHO, CBO and hybrid HHCBO performance are compared with IHHHCBO algorithm. The IHHHCBO achieves 1 of NMI, 0.521of Q, 1 of f1-score and 1 of accuracy which is higher when compared with existing algorithms.

Table 5. Performance of IHHHCBO for Football dataset

Method	NMI	Q	F1-score	Accuracy
HHO	0.762	0.436	0.697	0.796
CBO	0.819	0.493	0.738	0.852
HHHCBO	0.856	0.557	0.765	0.906
IHHHCBO	1	0.618	0.852	0.998

Table 6. Performance of IHHHCBO for Political books dataset

Method	NMI	Q	F1-score	Accuracy
HHO	0.815	0.387	0.736	0.719
CBO	0.883	0.492	0.795	0.775
HHHCBO	0.946	0.528	0.847	0.826
IHHHCBO	0.987	0.566	0.893	0.883

Table 4 presents the proposed IHHHCBO performance for the Dolphin dataset through NMI, Q, f1-score and accuracy. The HHO, CBO and hybrid

HHCBO performance are compared with IHHHCBO algorithm. The IHHHCBO achieves 1 of NMI, 0.589 of Q, 1 of f1-score and 1 of accuracy which is higher when compared with existing algorithms.

Table 5 presents the proposed IHHHCBO performance for the Football dataset through NMI, Q, f1-score and accuracy. The HHO, CBO and hybrid HHCBO performance are compared with IHHHCBO algorithm. The IHHHCBO achieves 1 of NMI, 0.618 of Q, 0.852 of f1-score and 0.998 of accuracy which is higher when compared with existing algorithms.

Table 6 presents the proposed IHHHCBO performance for the Political Books dataset through NMI, Q, f1-score and accuracy. The HHO, CBO and hybrid HHCBO performance are compared with the IHHHCBO algorithm. The IHHHCBO achieves 0.987 of NMI, 0.566 of Q, 0.893 of f1-score and 0.883 of accuracy which is higher when compared with existing algorithms.

4.2 Comparative analysis

The IHHHCBO performance is compared with existing algorithms like DPNLP [21], IHHOOBL [22], CSEA [23], DS-LPA [24] and MCOBGA [25]. The NMI, Q, f1-score and accuracy are taken as performance metrics for assessing the algorithm performance. Table 7 shows the comparative analysis of all four datasets.

Table 7. Comparative analysis

Method	Dataset	NMI	Q	F1-score	Accuracy
DPNLP [21]	Karate	1	0.371	1	N/A
	Dolphin	1	0.378	1	N/A
	Football	0.89	0.584	0.80	N/A
	Political books	0.70	0.441	0.87	N/A
IHHOOBL [22]	Karate	1	0.430	N/A	1
	Dolphin	1	0.526	N/A	0.531
	Football	1	0.589	N/A	0.566
	Political books	0.946	0.539	N/A	0.839
CSEA [23]	Karate	1	N/A	N/A	1
	Dolphin	1	N/A	N/A	1
	Football	0.989	N/A	N/A	0.994
	Political books	0.609	N/A	N/A	0.848
DS-LPA [24]	Karate	1	0.371	N/A	N/A
	Dolphin	1	0.379	N/A	N/A
	Football	0.887	0.565	N/A	N/A
	Political books	0.598	0.446	N/A	N/A
MCOBGA [25]	Karate	0.710	0.222	0.203	N/A
	Dolphin	0.511	0.271	0.402	N/A
	Football	0.710	0.301	0.178	N/A
Proposed IHHHCBO	Karate	1	0.521	1	1
	Dolphin	1	0.589	1	1
	Football	1	0.618	0.852	0.998
	Political books	0.987	0.566	0.893	0.883

4.3 Discussion

The advantages of IHHHCBO and the limitations of existing algorithms are designated in this section in detail. Due to the difficult overlapping community structure, the DPNLP [21] reached NMI of 1, 1, 0.89 and 0.70 for Karate, Dolphin, Football and Political Books. The IHHOOBL [22] reached NMI of 1, 1, 1, and 0.946 for Karate, Dolphin, Football and Political Books due to its random nodes affecting the accuracy which diminishes the performance. The CSEA [23] reached NMI of 1, 1, 0.989 and 0.609 for Karate, Dolphin, Football and Political Books due to its limited search space. The DS-LPA [24] reached NMI of 1, 1, 0.887 and 0.5998 for Karate, Dolphin, Football and Political Books due to its huge local optimum issues and less convergence speed. Due to the non-parametric test result validation was not confirmed in soft computing, the MCOBGA [25] reached NMI of 0.710, 0.511, 0.710 for Karate, Dolphin, and Football. The Proposed IHHHCBO reaches NMI of 1, 1, 1, and 0.987 for Karate, Dolphin, Football and Political Books due to its less probability of local optimum issue, enhancing the search space and convergence speed.

5. Conclusion

This manuscript proposed an IHHHCBO algorithm for community detection. The Karate, Dolphin, Football and Political Books datasets are considered for assessing the IHHHCBO performance. The optimum separate movement of CBO is incorporated in HHO population initialization for strengthen both optimizations. It is employed to initialize the population for enhancing the position diversity and change with other individuals. The proposed model diminishes the probability of local optimum issue, enhance the search space and convergence speed. The EMS is developed to produce mutant candidate locations which enhances the exploration and population diversity of optimization. The IHHHCBO performance is estimated through metrics like NMI, Q, f1-score and accuracy. The IHHHCBO reaches better accuracy of 1, 1, 0.998 and 0.883 for Karate, Dolphin, Football and Political Books respectively which is better when compared to existing algorithms. In future, various improvement approaches are utilized to further enhance the IHHHCBO performance.

Notations:

Notation	Description
μ_i	Fraction of edges
$\|C\|$	Cardinality of C
$M(C)$	Power mean of community order
v_C	Community volume
$k_i^{in}(C)$ and $k_i^{out}(C)$	Internal and external node degree
α	Actual parameter
$X(t)$	Position vector of Harris hawk in present iteration
$X(t+1)$	Position vector of Harris hawk in next iteration
X_{prey}	Prey position
$X_{rand}(t)$	Position vector of random individuals in the present population
$r_1 \sim r_{12}$ and q	Random numbers between $[0, 1]$
UB and LB	Upper and lower bound variables
$X_m(t)$	Average position of the population in every hawk
E_0 and R	Random number between $[-1, 1]$
t and T	Present and the maximum number of nodes
J	Random jump prey intensity
D	Problem dimension
S	Random vector within the size of $1 \times D$
$F(.)$	Objective function
$X_{i-1}(t)$	$i-1$th follower position in the present iteration t
$LeaderX_k(t)$	Position of designated leader
$gBest$	Present optimum position
$\cos(2R\pi)$	Individual best distance
V_{i1}, V_{i2} and V_{i3}	Recently produced mutant candidate locations at ith position
$R1 \sim R11$	Various exponents within the range of $[1, N]$
F_1, F_2 and F_3	Scale factors
$Xnew_i(t+1)$	Position of ith search agent at iteration $(t+1)$
$X_i(t)$	Position of iteration
$X_{prey}(t)$	Present optimum solution
d	Distance between node i and j
k	Symbol of every data
n	Number of nodes
u and v	Dual communities by the network
M	Confusion matrix
M_x and M_y	Number of communities in x and y sections
Y	Total number of edges in the network
m_{ij}	Value of row i and column j in the matrix
$\delta(i,j)$	Connection between two nodes
TP	True positive
FP	False positive
FN	False negative
$Map(C_i)$	Mapping function
$\delta(R_i.Map(C_i))$	Crocker delta

Conflicts of Interest

The authors declare no conflict of interest.

Author Contributions

The paper conceptualization, methodology, software, validation, formal analysis, investigation, resources, data curation, writing—original draft preparation, writing—review and editing, visualization, have been done by 1st and 2nd author. The supervision and project administration, have been done by 3rd author.

References

[1] K. Asmi, D. Lotfi, and A. Abarda, "The greedy coupled-seeds expansion method for the overlapping community detection in social networks", *Computing*, Vol. 104, No. 2, pp. 295-313, 2022.

[2] L. Li, F. Gou, and J. Wu, "Modified data delivery strategy based on stochastic block model and community detection in opportunistic social networks", *Wireless Communications and Mobile Computing*, 2022, pp. 5067849, 2022.

[3] M. Naderipour, M.H.F. Zarandi, and S. Bastani, "A fuzzy cluster-validity index based on the topology structure and node attribute in complex networks", *Expert Systems with Applications*, Vol. 187, pp. 115913, 2022.

[4] S. Kumar, D. Lohia, D. Pratap, A. Krishna, and B.S. Panda, "MDER: modified degree with exclusion ratio algorithm for influence maximisation in social networks", *Computing*, Vol. 104, No. 2, pp. 359-382, 2022.

[5] T. Jia, C. Cai, X. Li, X. Luo, Y. Zhang, and X. Yu, "Dynamical community detection and spatiotemporal analysis in multilayer spatial interaction networks using trajectory data", *International Journal of Geographical Information Science*, Vol. 36, No. 9, pp. 1719-1740, 2022.

[6] M. Naderipour, M.H. Fazel Zarandi, and S. Bastani, "Fuzzy community detection on the basis of similarities in structural/attribute in large-scale social networks", *Artificial Intelligence Review*, Vol. 55, pp. 1373-1407, 2022.

[7] P. Pham, L.T. Nguyen, B. Vo, and U. Yun, "Bot2Vec: A general approach of intra-

community oriented representation learning for bot detection in different types of social networks", *Information Systems*, Vol. 103, pp. 101771, 2022.

[8] H. Qie, S. Li, Y. Dou, J. Xu, Y. Xiong, and Z. Gao, "Isolate sets partition benefits community detection of parallel Louvain method", *Scientific Reports*, Vol. 12, No. 1, pp. 8248, 2022.

[9] A. Kumari, R.K. Behera, K.S. Sahoo, A. Nayyar, A. Kumar Luhach, and S. Prakash Sahoo, "Supervised link prediction using structured-based feature extraction in social network", *Concurrency and Computation: Practice and Experience*, Vol. 34, No. 13, pp. e5839, 2022.

[10] F. Kazemzadeh, A.A. Safaei, and M. Mirzarezaee, "Influence maximization in social networks using effective community detection", *Physica A: Statistical Mechanics and its Applications*, Vol. 598, pp. 127314, 2022.

[11] K. Berahmand, M. Mohammadi, F. Saberi-Movahed, Y. Li, and Y. Xu, "Graph regularized nonnegative matrix factorization for community detection in attributed networks", *IEEE Transactions on Network Science and Engineering*, Vol. 10, No. 1, pp. 372-385, 2022.

[12] C. Li, Y. Tang, Z. Tang, J. Cao, and Y. Zhang, "Motif-based embedding label propagation algorithm for community detection", *International Journal of Intelligent Systems*, Vol. 37, No. 3, pp. 1880-1902, 2022.

[13] M. Al-Andoli, S.C. Tan, and W.P. Cheah, "Parallel stacked autoencoder with particle swarm optimization for community detection in complex networks", *Applied Intelligence*, Vol. 52, pp. 3366-3386, 2022.

[14] K. Berahmand, S. Haghani, M. Rostami, and Y. Li, "A new attributed graph clustering by using label propagation in complex networks", *Journal of King Saud University-Computer and Information Sciences*, Vol. 34, No. 5, pp. 1869-1883, 2022.

[15] J. Tang, H. Zhu, J. Lan, S. Song, J. Qu, and Q. Du, "Identifying influential nodes for influence maximization problem in social networks using an improved discrete particle swarm optimization", *Social Network Analysis and Mining*, Vol. 13, No. 1, pp. 94, 2023.

[16] S. Yuan, C. Wang, Q. Jiang, and J. Ma, "Community detection with graph neural network using Markov stability", In: *2022 International Conference on Artificial Intelligence in Information and Communication (ICAIIC)*, pp. 437-442, 2022.

[17] F. Dabaghi-Zarandi, and P. KamaliPour, "Community detection in complex network based on an improved random algorithm using local and global network information", *Journal of Network and Computer Applications*, Vol. 206, pp. 103492, 2022.

[18] L. Zhang, Y. Liu, H. Yang, F. Cheng, Q. Liu, and X. Zhang, "Overlapping community-based particle swarm optimization algorithm for influence maximization in social networks", *CAAI Transactions on Intelligence Technology*, Vol. 8, No. 3, pp. 893-913, 2023.

[19] S.T. Shishavan, and F.S. Gharehchopogh, "An improved cuckoo search optimization algorithm with genetic algorithm for community detection in complex networks", *Multimedia Tools and Applications*, Vol. 81, No. 18, pp. 25205-25231, 2022.

[20] M.N. Al-Andoli, S.C. Tan, and W.P. Cheah, "Distributed parallel deep learning with a hybrid backpropagation-particle swarm optimization for community detection in large complex networks", *Information Sciences*, Vol. 600, pp. 94-117, 2022.

[21] M. Zarezadeh, E. Nourani, and A. Bouyer, "DPNLP: distance based peripheral nodes label propagation algorithm for community detection in social networks", *World Wide Web*, Vol. 25, pp. 73-98, 2022.

[22] F.S. Gharehchopogh, "An improved Harris Hawks optimization algorithm with multi-strategy for community detection in social network", *Journal of Bionic Engineering*, Vol. 20, No. 3, pp. 1175-1197, 2023.

[23] R. Fei, Y. Wan, B. Hu, A. Li, and Q. Li, "A novel network core structure extraction algorithm utilized variational autoencoder for community detection", *Expert Systems with Applications*, Vol. 222, pp. 119775, 2023.

[24] C. Li, H. Chen, T. Li, and X. Yang, "A stable community detection approach for complex network based on density peak clustering and label propagation", *Applied Intelligence*, Vol. 52, No. 2, pp. 1188-1208, 2022.

[25] H.K. Shakya, K. Singh, Y.S. More, and B. Biswas, "Opposition-based genetic algorithm for community detection in social networks. Proceedings of the National Academy of Sciences", *India Section A: Physical Sciences*, Vol. 92, No. 2, pp. 251-263, 2022.

[26] W.W. Zachary, "An information flow model for conflict and fission in small groups",

Received: March 28, 2024. Revised: May 7, 2024.

Journal of anthropological research, Vol. 33, No. 4, pp.452-473, 1977.

[27] D. Lusseau, K. Schneider, O.J. Boisseau, P. Haase, E. Slooten, and S.M. Dawson, "The bottlenose dolphin community of doubtful sound features a large proportion of long-lasting associations: can geographic isolation explain this unique trait?", *Behavioral Ecology and Sociobiology*, Vol. 54, pp. 396-405, 2023.

[28] M. Girvan, and M.E. Newman, "Community structure in social and biological networks", *Proceedings of the National Academy of Sciences*, Vol. 99, No. 12, pp.7821-7826, 2002.

[29] L.A. Adamic, and N. Glance, "The political blogosphere and the 2004 US election: divided they blog", In: *Proc. of the 3rd international workshop on Link discovery*, pp. 36-43, 2005.

[30] A. Lancichinetti, S. Fortunato, and J. Kertész, "Detecting the overlapping and hierarchical community structure in complex networks", *New journal of physics*, Vol. 11, No. 3, pp. 033015, 2009.