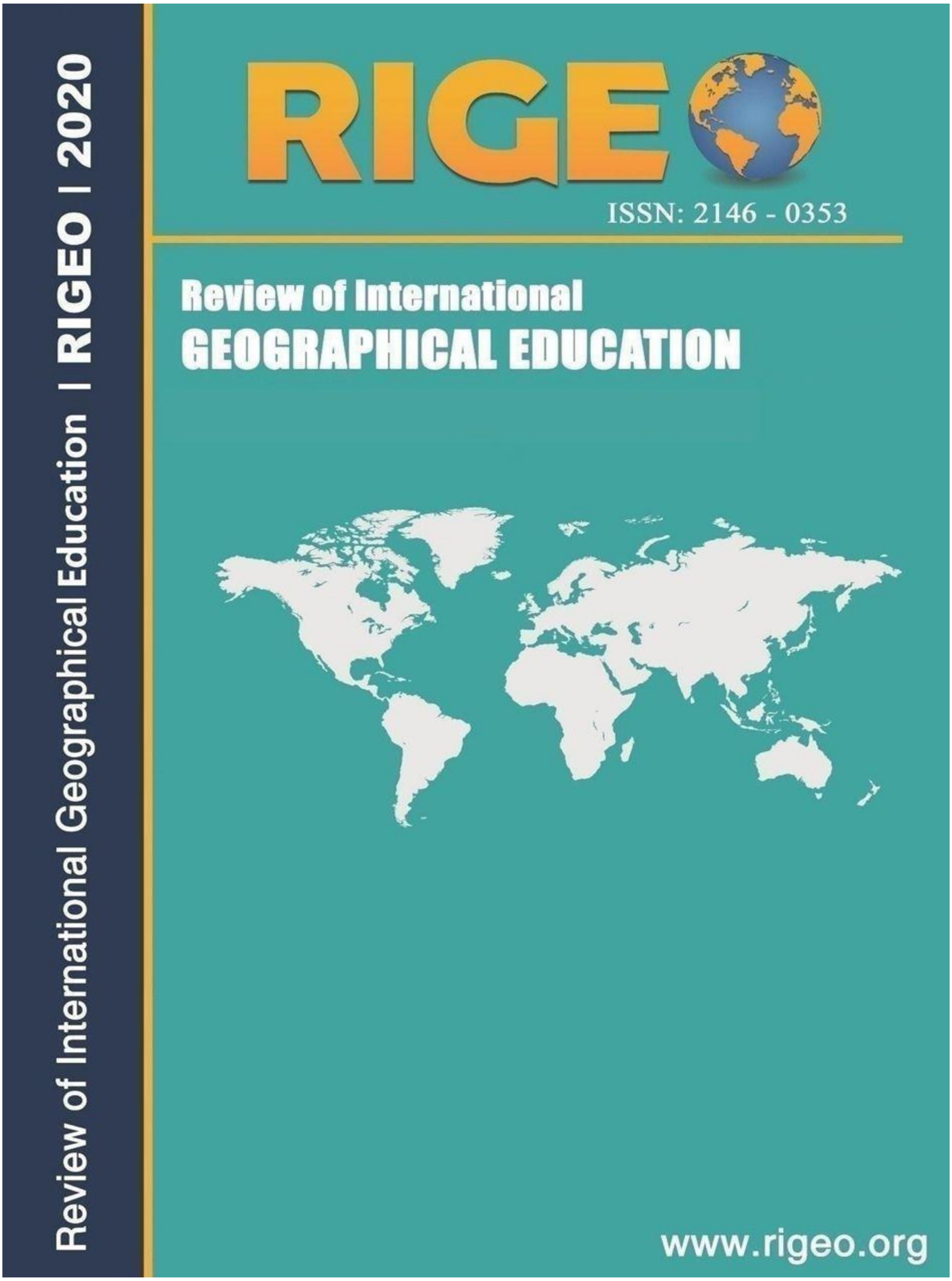
Review of International Geographical Education | RIGEO | 2020
RIGEO
ISSN: 2146 - 0353
Review of International
GEOGRAPHICAL EDUCATION
www.rigeo.org

Retracted: A Rider-Harris Hawks Optimization Algorithm-Based Method for Modular Community Detection in Bipartite Networks

Jayashree D N[1], Dr N Prabhudeva[2], K R Girish[3]

Assoc. Professor[1], Professor & HOD[2], Asst. Professor[3]

jayashreedn12@gmail.com[1], drnprabhudeva@gmail.com[2] , girishhspt@gmail.com[3]

Department of Maths, Proudhadevaraya Institute of Technology, Abheraj Baldota Rd, Indiranagar, Hosapete, Karnataka-583225

Abstract:

Research into the hidden community structure in complex networks has garnered a lot of attention and is now a hot topic in the multifaceted network domain; this is because this work not only sheds light on the hierarchical structure of these networks, but it also helps us understand their fundamental functions and, in turn, how to recommend information to users. In a bipartite network, each set of nodes is distinct from the other, and no edges connect the vertices. This kind of network is known as a multidimensional network. Despite much research on community identification in one-mode networks, bipartite network community detection remains unexplored. Here, we provide a new method for community discovery in bipartite networks using node similarity—the Rider-Harris Hawks Optimisation (RHHO) algorithm. We devised the suggested RHHO by integrating the RO and HHO algorithms, which stand for Rider Optimisation and Harris Hawks Optimisation, respectively. Additionally, a novel assessment measure called the h-Tversky Index (h-TI) is put forth for the purpose of calculating node similarity and fitness, with modularity as a key consideration. By breaking the network into smaller, more manageable pieces, we can test how well the suggested community identification works. This is what modularity is all about. Over the citation networks (cit-HepPh and cit-HepTh) and bipartite networks (Movie Lens 100 K and American Revolution datasets), the suggested method was quantitatively evaluated and compared in terms of robustness and modularity. After 250 simulation tests, the results were analysed. The proposed method outperformed state-of-the-art approaches, including h-index-based link prediction, multiagent genetic algorithms (MAGA), genetic algorithms (GA), memetic algorithm for community detection in bipartite networks (MATMCD-BN), and HHO, according to experimental results. It also showed a maximal fitness of 0.74353 and maximal modularity of 0.77433.

1. Introduction

Naturally, the complex systems are deemed to be divided into multiple communities or modules. Usually, in order to represent the networks, the said communities or modulesare labeled as clusters of compactly linked nodes with scarce

links to the nodes of other clusters. Complex network models have several representations, including one-mode network, bipartite network, and multimode network, but the existence of a bipartite network is very close to a natural phenomenon when especially modeling association relationsbetween two different classes of real-world objects. In a

bipartite network, there are two different kinds of nodes. The existence of the edges between nodes is conditional such that if connecting nodes are associated with other types. Bipartite networks maintain rich information regarding the entire network being modeled and share important statistical properties like clustering coefficient as their single-mode form. Many real-world applications include the P2P net- work, entertainment and audience network, research co- ordination network, and items lending network.

Community detection holds an essentially significantcontribution in many complex networks, especially animportant class, i.e., bipartite networks. Community iden-tification and dichotomy characteristics in a bipartite net-work not only reveal details about the hierarchical structureof a multifaceted network but also assist in better under-standing of the core functions of the network and subse-quently information recommendation. The bipartitenetworks belong to the multifaceted network whose nodescan be divided into a dissimilar node-set so that no edgesassist between the vertices. Even though the discovery ofcommunities in one-mode networks is briefly studied,community detection with bipartite networks is not studied.Community detection is a trending research domain in the field of network science that poses the ability to offervision to the fundamental structure and provides impendingfunctions to the networks. Numerous real-world models likethe Internet, food webs, social relationships, and biologicalsystems are considered complex networks [1, 2]. The com-munity is represented as a complex network which is de-scribed as the collection of nodes which are sturdily linked toone another but sporadically linked to nodes that are presentexternal to the set [3]. The algorithms based on communitydetection are developed for recognizing the nodes, modules,or clusters inside the network that are more likely to inter-relate among themselves than with the other network. Thisprocess is carried out when the nodes belong to the same community whereas it performs differently when the nodebelongs to other communities [4]. The social network usingthe attribute nodes is given by the bipartite graph and theextracted bicommunities are revealed later like other com-munities in bipartite bibliographic network which isemployed for citation recommendation [4, 5]. The bipartitenetwork is also known as two-mode network or the affiliationnetwork wherein the nodes are divided into two differentcollections that involved upper and lower nodes. Here, eachedge is adapted to connect the upper nodes to the lowernodes. Here, no edges exist between the two upper nodes andno edges lie between lower nodes [6]. The bipartite networkoffers an insight exemplification between two disjoint groupsusing the applications that range between the citation net-works, disjoint groups, collaboration networks, and ecologicalnetworks. Here, the bipartite graph contains specific coverage property also termed as maximum matching [7].

Bipartite networks can be epitomized in the real-world scenarios considering two different categories of objects that involved movie-actor relation and paper-reference relation. The dichotomy physiognomies of the target network assist in disclosing more details than the single model networks [5]. Thebipartite networks provide statistical properties in their single-

mode form which helps to define nodes of two parts of single model network and the original bipartite network offers the degree distribution and clustering coefficient to handle more information using the real network being modeled with single model version [4]. The network model is broadly employed in the reality, and the researchers provided bipartite network-related research using real-time application, cooperation net- work, and P2P exchange network [8]. There exist two ways for curving the relation of different object classes which involved projection method and nonprojection method [8]. The pro- jection method projects the two parts of bipartite network considering certain nodes to evaluate further study. Similarity-based strategies are the frequently employed link predictionstrategies wherein each pair of nodes considers proximity score,described based on network structure which implies that twonodes are similar if they pose higher structural similarities [9].The proposed approach [10] is a heuristic method based on modularity optimization and demonstrated high quality of the community detection in terms of modularity in bipartite networks. The efficiency and effectiveness of Louvain algorithmhad been proved by several applications.

Numerous techniques of community detection are designed to recognize the community structure. The near- optimal or optimal values of some criteria are generated by good partition. Moreover, the good separation also reveals the organization using community structure with different reso- lutions [11]. Conventionally, hierarchical clustering and graph partitioning strategies like agglomerative algorithms and dis- ruptive counterparts are used for solving the issues of com- munity detection [12]. In [13], modularity is delineated to measure the quality of partitions. Corrêa et al. [14] followed a complex network approach for word sense disambiguation. Through community detection in input-output bipartite graphs (BGs), Tang and Daoutidis [15] proposed the networkdecomposition for distributed control. In order to discover the necessary patterns in the IP traffic, Viard et al. [16] used cliques in bipartite link streams. Huang et al. [17] presented a novel link prediction for large scale miRNA-lncRNA interaction network in the BG. Rechner et al. [18] introduced the uniformsampling of the BGs with degrees in suggested intervals. Basedon theory of complex network, Guan et al. [19] offered a service-oriented deployment policy of end-to-end network slicing. Bian and Deng [20] carried out the research to identifythe influential nodes in complex networks. Gao et al.

[21] wrote a paper; titled “an adaptive optimal-Kernel time-frequency representation-based complex network method for charac- terizing fatigued behavior using the SSVEP-based BCI system.”Huang et al. [22] carried out a survey on techniques of community detection in multilayer networks. Rostami et al.
[23] presented a genetic algorithm for feature selection that is
based on a novel community detection, Li et al. [24] proposedthe convex relaxation techniques for community detection, andJoo et al. [25] utilized the community detection for studying thestream gauge network grouping. The modularity is employedto reflect the fraction of edges using the communities related to the amount of edges developed using communities. Here, the method devised a null model, which utilizes the nodes degreefor computing the uncertainty of edges that are established between the nodes.

Moreover, the modularity with high value specifies good partition, thereby maximizing the standard using optimi- zation method [26]. However, it is obstinate to determine the precise optimal solution for the issues. Thus, many ap- proximate techniques are designed for community detec- tion. In [13], greedy method was devised for community detection. The algorithm showed effective performance and is considered as an efficient algorithm for detecting the nodes similarities. The Louvain algorithm contained two important processes. In the initial phase, using the com- munity of each node, the modularity is optimized locally. The node with the same communities is aggregated and is considered as supernode and forms the novel coarse-grainednetwork. The process is repeatedly performed using con- centrated networks considering modularity to devise the node movement contained in network. Even though Lou- vain algorithm and modularity-based strategies pose certaindrawbacks, they are still extensively utilized to evaluate real-world issues [27].

The principal purpose of this research is to devise a technique for community detection in bipartite network based on the node similarity. Here, the h-Tversky Index (h-TI) measure is newly designed by modifying h-index based on Tversky index for computing node similarity. This method is based on the similarity measures between nodes that exploit the bipartite networks. The algorithm holds the cycles for connection to maximize the similarity between nodes in order to define the communities. Then, the community detection is performed based on the proposed Rider-Harris Hawks Optimization algorithm (RHHO) and modularity. The RHHO introduced in this study is developed by integrating Rider Optimization (RO) algorithm with the Harris Hawks Optimization (HHO) algorithm. In addition, the fitness is newly devised considering modularity and proposed h-Tversky Index (h-TI) for evaluating the node similarity. The aim of modularity is to enumerate the integrity of specific di- vision of network to evaluate the accuracy of the proposed community detection.
The major contributions are enlisted as follows:

(i) Proposing RHHO for community detection: the proposed RHHO algorithm is a novel derivationwhich is achieved by the combination of RO and HHO algorithms, for community detection

(ii) Proposing h-Tversky Index (h-TI) for node simi- larity: the similarity between the nodes is evaluated by integrating h-index and Tversky similarity index

Rest of the sections of this article is arranged as follows. Section 2 describes the strategies of conventional commu- nity detection that are used in the literature and challenges faced, which motivated developing the proposed technique. The objective model of community detection using bipartite graph is illustrated in Section 3. The method introduced for community detection through proposed RHHO is describedin Section 4. The results of the introduced RHHO are compared with other techniques in Section 5 and finally, Section 6 provides the conclusion.

2. Motivations

The problem of community detection is NP-hard, since people have utilized different techniques to address the optimization problem. Therefore, precise algorithms like swarm evolutionary algorithms (EAs) and intelligence al- gorithms are employed for community detection, but the convergence of the global optimal solution needs more time. The limitation linked with modularity is identified, which is termed as resolution limits. Moreover, the modularity fails to determine the community structure for fewer nodes. The aforementioned limitations stood as the motivation for designing a novel community detection model in bipartite networks.

2.1. Literature Survey. The techniques based on eight existing community detection algorithms using bipartite network are illustrated. Zhou et al. [9] designed two h-index-based link prediction techniques using the citation network. Here, the h-type index was adapted for computing the significant nodes using the citation network. Moreover, the accuracy of prediction was found better but the method was inapplicable with other types of networks to enhance the performance of system. Gmati et al. [4] designed Fast- Bi community detection (FBCD) for detecting the com- munity in social network using the node attributes. The goal of the model is to discover the maximum matching using bipartite graph for minimizing the complication. The method failed to use other kinds of bipartite network like directed, weighted, or dynamic network for determining the community structure. Che et al. [28] devised memetic algorithm, namely, MATMCD-BN for community detec- tion using two-mode networks. The method employed conventional string-driven representation strategy for chromosome representation. Here, population initializa- tion method was devised using bipartite network for en- hancing the convergence rate. Moreover, the density-based bipartite modularity function was devised using the fitness function. However, the method failed to determine more than one node. Chang et al. [7] designed overlapping community detection strategy considering complete bi- partite graph using microbipartite network Bi-EgoNet (CBG and BEN), which combined the benefits of both bipartite graph and the Bi-EgoNet for generating the best community structure. However, the method failed to evaluate associated issues faced by the bipartite network considering Bi-EgoNet. Sun et al. [29] designed BiAttracter for determining the two-mode communities using bipartite networks. The method was computed on the basis of distance dynamics attractor model. Even though the method precisely determined the two-mode communities of bipartite network in less time, it failed to discover community detection considering heterogeneous network, multilevel network, and temporal network. Li et al. [3]designed quantitative function for determining the com-munity structures considering bipartite network. More- over, the Heuristic and Adapted Label Propagation Algorithm (BiLPA) was devised to optimize the quantitative function using huge scale bipartite networks. However, some of the data of bipartite network were missing in the obtained proposed network. Zhou et al. [10] designed a method for community detection using the bipartite network. Here, the expansion of bipartite mod- ularity was designed and Louvain algorithm was devised. The Louvain algorithm adapted indigenous moving heu- ristic to unfold the complete hierarchical structure of the network. In addition, the Laplacian dynamics was con- sidered for analyzing the constancy of community struc- ture but failed to develop community-enabled recommendation model. Xue et al. [30] designed a method for addressing the cold start issue for community detection considering bipartite graph. At first, the decoupled nor- malization strategy was used to extract the inclination patterns considering the ratings. Moreover, two incre- mental community detection methods were devised for capturing the interesting shifts based on missing method of rating. However, the technique was unsuccessful in using the pairwise constraints for semisupervised learning for the enhancement of system’s performance.

2.2. Challenges. The challenges confronted by the conven- tional techniques for developing a method for effective community detection are portrayed as follows:

(i) Another drawback confronted by the community detection method is weighted modularity. Here, the weighted modularity was only effective on networks in which all connections are positive. However, these methods failed to create modules in weightednetworks for devising negative and positive link strengths [26].

(ii) Determining the structure of networks is beneficial for illustrating their formulation function andperformance and is considered a significant issue incommunity detection [1].

(iii) In [30], the Incremental Group-Specific model was designed for community detection. However, the empirical analysis was not performed for offering a reasonable explanation to simplify the grouping method and failed to combine valuable topological information.

(iv) In [9], h-index-based link prediction method was developed using the citation network. Still, it did notconsider the h-index and Tversky similarity indicesand the Salton to improve performance.

(v) Bipartite networks fit in the category of complicated networks, whose vertices are distributed into two alienated collections of vertices, such that there do not exist any edges between vertices of the same collections/set, and edges only subsist between nodes of different collections. Even though, in one-mode networks, the community discovery is widely studied, the community detections in bipartite networks have not been studied due to the fact that the projection loses important information of the original bipartite network.

3. Objective Model for Community Detection

Community detection is a fundamental tool employed for discovering valuable information that is hidden over complex networks. Numerous community detection techniques for bipartite networks are devised considering different viewpoints. However, the efficiency of these techniques worsens when the community structure turns ambiguous. Improving the community structure is a complicated task. The bipartite network is an essential class of complex networks in real-world systems, wherein each node is of different types, and no two nodes are the same type. For example, a bipartite network that has three communities is shown in Figure 1.

As bipartite networks pose community structure and the communities are independent of one another, that helped to expose indefinite functional modules. The analysis and detection of these communities from the bipartite network offer a means for functional classification of the bipartite network. Community detection is a challenging task con- sidering bipartite networks due to the fact that the com- munity detection problem is NP-hard. The algorithm makes the issue of community detection into a combinatorial optimization issue.

Modularity is widely applied for the evaluation of the quality of a specific partition of a network into communities. Moreover, modularity reflects the extent, relative to a null model network, to which edges are formed within commu- nities instead of between them. Further, the bipartite modu- larity measures are proposed, which could be useful in the recognition of communities in bipartite networks. In turn, the model is newly devised that would have the same number of nodes and degree distribution as that of the original network while the edges of the node are replaced.

Assume a bipartite graph which is modeled as an undirected graph G D, I where D represents set of nodes and I indicates a set of edges. The node-set D is expressed as D $X \cup W$ where X and W indicate the types of node X and type W. The set of edges is represented as I. The edges pose the ability to connect different types of nodes, which are modeled as edges $n_{ef} \in I$ $b_e \in X$; $s_f \in W$, and I indicate the number of edges in a bipartite graph. The detection of community in a bipartite graph is expressed as G D, I $X \cup W, I$ which is employed to partition G into subgraphs, modeled as G_e $X_e \cup W_e, I_e$ where e $1, 2, \ldots, o$ and o is a total number of communities.

Bipartite Modularity for Detecting Similarity between Nodes. The modularity [31] is devised to quantify the in- tegrity of a specific part in the provided network and is considered as a widespread benchmark index to compute the accuracy in community detection. The community structure is defined as a model that arranges the edges in a statistically surprising manner. Assume g_j represents the degree of node j and E indicates a total number of edges. The probability of edge being presented between node j and node q is repre-sented as $g_j g_q/2E$. The modularity quantifying the number of edges based on newly devised model can be expressed as

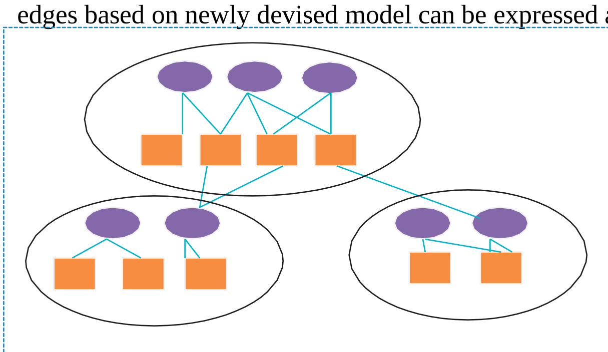

Figure 1: A bipartite network having three communities.

The proposed RHHO is a novel population initialization method, which is useful for accelerating population con- vergence. In addition, the fitness function, namely, h-Tversky similarity index, is newly devised for computing the individuals from the population.

3.1. Bipartite Graph. Consider a bipartite graph which is represented as a graph $G = (D, I)$ where D represents nodes'set and I indicates edges' set. The nodes' set D is expressed as $D = X \cup W$ where X and W indicate the node of type X and type W. The set of edges is represented by I. The edges pose the ability to connect different types of nodes, which are modeled as edges $n_{ef} \in I \{b_e \in X, s_f \in W\}$ and I indicates the number of edges in a bipartite graph. The detection of community in a bipartite graph is expressed as $G(D, I) = (X \cup W, I)$, which is employed to partition G into subgraphs, which are modeled as $G_e = (X_e \cup W_e, I_e)$ where $e \in \{1, 2, \ldots, o\}$ and o is the total number of communities.

3.2. Determination of Node Similarity. The similarity is de- fined as a metric utilized for computing the amount of closeness between two pairs of nodes. Numerous node similarity measures based on the local information are described in the literature, which showed a different per- formance for determining the community structure from the complicated networks. However, the proposed method computes the node similarity in the network using a fitness function derived by the bipartite modularity and h-Tversky similarity index. Here, the fitness function is considered as a maximization function and is expressed as

$$F = \frac{M_a + M_G}{2}, \tag{5}$$

where M_a indicates a bipartite modularity and M_G repre- sents the proposed h-Tversky similarity index considering a specific community structure. The proposed h-Tversky similarity index that considers a community of specific type of node in the network is given as among the exploitation and exploration and helps to boost the exploratory behavior. Moreover, the quality of solutionsis improved during a number of iterations. The HHO al- gorithm is effective in handling the difficulties of search space with local optimal solutions. On the other hand, the RO algorithm is inspired by riders racing to reach a par- ticular destination. Simultaneously, the usual RO algorithm displays good global optimal convergence. Based on the imaginary ideas and thoughts, nothing like the other nature-inspired and artificial computing algorithms, RO algorithm works in the fictional computing platform. In RO algorithm, the optimization behavior depends on four groups of riders, each presenting particular characteristics. The overtaking rider derives the new update rule in the HHO algorithm by using the RO algorithm. The advantages of bypass include the faster convergence with greater global neighborhoods. Hence, in the current RHHO, the optimal global conver- gence is enhanced at the maximal iteration. The algorithmic steps of the introduced RHHO are defined as follows.

4.3.1. Step 1: Initialization. First is initialization of pop- ulation that is denoted as Z with total d rabbits, where $1 \leq c \leq d$,

$$Z = \{Z_1, Z_2, \ldots, Z_c, \ldots, Z_d\}, \tag{7}$$

where d is total solution, and Z_c indicates the cth solution.

4.3.2. Step 2: Determination of Fitness Function. The success rate or fitness of the solution is computed on the basis of bipartite modularity and proposed h-Tversky similarity index, which is elaborated in Section 4.2. Hence, the solu- tion's fitness is depicted in equation (3).

4.3.3. Step 3: Determination of Update Position. The scheme of selection in an HHO [33] algorithm helps to progressively update the position to attain an improved position. More- over, Harris' hawks enclose the anticipated prey by updating their places. In such circumstances, the current place up- dates the solution space as

$$M_G = \frac{1}{E} \sum_{j=1}^{q} \sum_{q=x+1}^{s}$$

3.3. Algorithmic Steps of Proposed Rider-Harris Hawk Opti- mization (RHHO) Algorithm. The HHO [33] is modified using the RO algorithm [32] wherein the update rule of HHO is updated based on the update rule of bypass rider in RO algorithm, thus obtaining the new algorithm, an RHHO, which is used to perform the community detection opti- mally. Basically, HHO is inspired by the chasing behavior ofHarris hawks. The HHO provides a smoother transition

where $Z(v)$ indicates the current position vector. Assuming $Z_{rab}(v)$ as positive, the above equation is represented as

$$Z(v+1) � Z_{rab}(v) - H \; Z_{rab}(v) - Z(v) \; . \qquad (10) \; ($$

Here, the updated position of the bypass rider according to RO algorithm [32] is utilized in the process of update for maximizing the rate of success by finding the position of bypass rider. The bypass riders trail a common route without

stalking the foremost rider. The bypass rider's equation given by riders is represented as

$$Z(v+1) � \mu[Z(v) * c(w) + Z(v) * [1 - c(w)]], \quad (11)$$

where ϑ and ℓ denote the random digits between 0 and 1, inclusive, κ and μ are random digits, and k indicates the iterations. Assume μ � r; the equation is rewritten as

$$Z(v+1) � \mu Z(v) * c(w) + \mu Z(v)[1 - c(w)], \qquad (12)$$

$$Z(v+1) � Z(v)[\mu c(w) + \mu[1 - c(w)]], \qquad (13)$$

$$Z \; v \quad \frac{Z(v+1)}{[\mu c(w) + \mu[1 - c(w)]]} \quad (\;) � \qquad (14)$$

Substituting equation (14) in (10), the update equationderived is

$$Z(v+1) � Z$$

The final equation is given by

$$\mu \frac{}{H} \quad {}_{rab} \quad Z(v+1) � \frac{\mu}{} Z_{rab} \; (v) - H(Z_{rab} \; (v). \qquad (16)$$

4.3.4. Step 4: Determining the Best Solution. If the solution acquired the minimal fitness value, then it is the best so- lution. Furthermore, the parameters of the update of a rider are crucial in order to conclude the best solution.

4.3.5. Step 5: Termination. Repeat the steps in anticipation of the iteration reaching the maximum count.

4. Results and Discussion

The analysis of the community detection model using the proposed RHHO is demonstrated in this section with an effective comparative analysis to prove the effectiveness of the proposed model.

4.1. Experimental Setup. The proposed method is executed in a system running Windows 8 OS with 4 GB of RAM, Intel core i-3 processor, and the implementation is carried out in Python.

4.2. Database Description. The nodes for the experimenta- tion are taken from the datasets, namely, the citation net- works dataset [34] and the bipartite network dataset [35]. The description for each is given below.

5.2.1. Citation Networks Dataset. The experimentation is performed on a citation network dataset wherein the node denotes papers and edges denote citations. The citation network dataset can be employed for clustering the networkand for studying the influence of citation networks to de- termine the most influential papers. Here, cit-HepPh and cit-HepTh are the two datasets used for performing thecommunity detection:

(a) Analysis based on cit-HepPh: the cit-HepPh network is an instance of citation network dataset data that can be temporal, directed, or labeled with 34,546 nodes and 421,578 edges. The cit-HepPh network data is employed in the Arxiv High Energy Physics paper citation network.

(b) Analysis based on cit-HepTh: the cit-HepTh networkdata can be directed, temporal, or labeled with 27,770 nodes and 352,807 edges. The cit-HepTh network data is employed in the Arxiv High Energy Physics paper citation network.

5.2.2. Bipartite Network Dataset. The experimentation is performed on a bipartite network dataset wherein the network consists of two distinct node types, and all edges connect a node of the first type with a node of the second type. Here, Movie Lens 100 K was acquired from the official website (https://grouplens.org/datasets/movielens/), and the American Revolution network was obtained from the website (http://konect.cc/networks/brunson_revolution/) with two public datasets employed for performing the community detection.

(a) Analysis based on Movie Lens 100 K: this bipartite network consists of ten million movie ratings wherein the left nodes are users, and the right nodesare movies. An edge connecting the user and a movierepresents the user who has rated the movie with therating value attached to the edge.

(b) Analysis based on American Revolution: the bi- partite network consists of membership infor- mation of thirty-six people in five organizations considering the period before the American Revolution. The list consists of well-known people like American activist Paul Revere. Here, the left nodes indicate persons, and the right nodes denote organizations. Here, the edge linking the person and the organization reveals that the person is a member of the organization.

4.3. Simulation Results. The simulation results of proposed community detection model considering citation network dataset and bipartite network dataset are illustrated in Figures 2 and 3.

5.3.1. Citation Networks Dataset. In this section, we analyzethe simulation results of community detection based oncitation networks datasets. Figure 2 elaborates the simula- tion results of the proposed community detection model using citation network dataset considering cit-HepPh and cit-HepTh datasets.

Figure 2(a) describes the original network using the cit- HepPh dataset and the communities identified considering the original network with the cit-HepPh dataset described in Figure 2(b). In Figure 2(b), green, red, and blue are the nodes representing different communities. Figure 2(c) elaborates the original network using the cit-HepTh dataset, and the communities detected by the original network using the cit-HepTh dataset are described in Figure 2(d). In Figure 2(d), green, red, and blue are the nodes with different commu- nities present in the original network.

5.3.2. Bipartite Network Dataset. This section explains the simulation results of community detection based on bi- partite networks datasets. Figure 3 elaborates the simulation results of the proposed community detection model using a bipartite network dataset considering Movie Lens 100 K and American Revolution datasets.

Figure 3(a) describes the original network using theMovie Lens 100 K dataset and the communities identified considering the original network with Movie Lens 100 K dataset described in Figure 3(b). In Figure 3(b), green, red, and blue represent different communities. Figure 3(c) elaborates the original network using the American Revo- lution dataset, and the communities detected by the originalnetwork using the American Revolution dataset are de-scribed in Figure 3(d). In Figure 2(d), green, red, and blue are the nodes with different communities' present in the original network.

4.4. Performance Analysis. The performance analysis of the proposed RHHO considering citation network dataset and bipartite network dataset is illustrated considering fitness and modularity measures.

5.4.1. Performance Analysis Based on Citation Networks Dataset Using cit-HepPh. Figure 4 illustrates the perfor- mance analysis of RHHO method using the cit-HepPh based fitness and modularity measures. The analysis of RHHO based on the fitness metric is portrayed in Figure 4(a). When the iteration is 1, the corresponding fitness values computed by the proposed RHHO with population 5, 10, 15, 20, 25, and 30 are 6.628186, 0.000010, 0.000014, 0.000018, 0.000024, and 0.000035. Likewise, when the iteration is 250, then the corresponding fitness values computed by the proposed RHHO with population 5, 10, 15, 20, 25, and 30 are 0.0059, 0.0182, 0.8430, 0.8771, 0.9431, and 0.9958, respectively. The analysis of the RHHO based on the modularity metric is illustrated in Figure 4(b). When the iteration is 1, the corresponding modularity values computed by proposed RHHO with population 5, 10, 15, 20, 25, and 30 are 0.00038, 0.00113, 0.00274, 0.00327, 0.00422, and 0.00455. Likewise, when the iteration is 250, then the corresponding modularity values computed by proposed RHHO method with population 5, 10, 15, 20, 25, and 30 are 6 0.62139, 0.65143, 0.65208, 0.65385, 0.66975, and 0.67397, respectively.

5.4.2. Performance Analysis Based on Citation Networks Dataset Using cit-HepTh. Figure 5 illustrates the perfor- mance analysis of the proposed RHHO using cit-HepTh based fitness and modularity measures. The analysis of the proposed RHHO based on the fitness metric is portrayed in Figure 5(a). When the iteration is 1, the corresponding fitness values computed by proposed RHHO with pop- ulation 5, 10, 15, 20, 25, and 30 are 8.7459, 0.000012, 0.000018, 0.000030, 0.000031, and 0.000068. Likewise, when the iteration is 250, then the corresponding fitness values computed by proposed RHHO with population 5, 10, 15, 20, 25, and 30 are 0.0011, 0.0020, 0.9203, 0.9586, 0.9681, and 0.9932, respectively. The analysis of the proposed RHHO based on the modularity metric is illustrated in Figure 5(b). When the iteration is 1, the corresponding modularity values computed by proposed RHHO with population 5, 10, 15, 20, 25, and 30 are 0.00146, 0.00177, 0.00211, 0.00346, 0.00361, and 0.00390. Likewise, when the iteration is 250, then the corresponding modularity values computed by proposed RHHO with population 5, 10, 15, 20, 25, and 30 are 0.5954, 0.6006, 0.6270, 0.6306, 0.6391, and 0.6855, respectively.

5.4.3. Performance Analysis Based on Bipartite Network Dataset Using Movie Lens 100 K. Figure 6 illustrates the performance analysis of the proposed RHHO using Movie Lens 100 K-based fitness and modularity measures. The analysis of the proposed RHHO based on the fitness metric isportrayed in Figure 6(a). When the iteration is 1, the

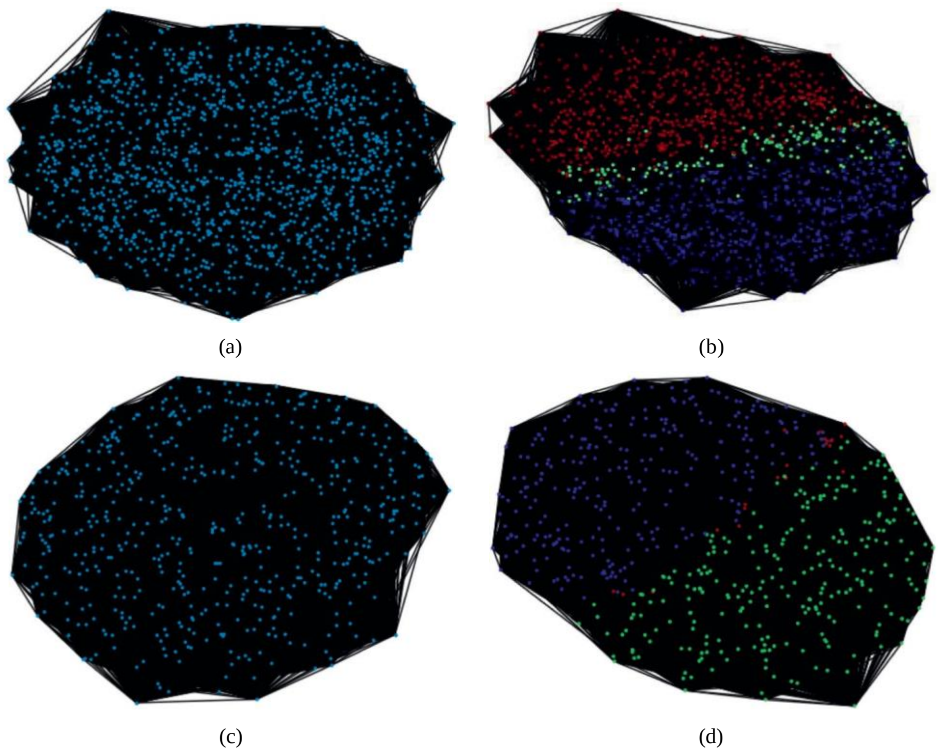

Figure 2: Analysis based on citation networks dataset using (a) original network using cit-HepPh dataset; (b) community detection using cit-HepPh dataset; (c) original network using the cit-HepTh dataset; (d) community detection using the cit-HepTh dataset.

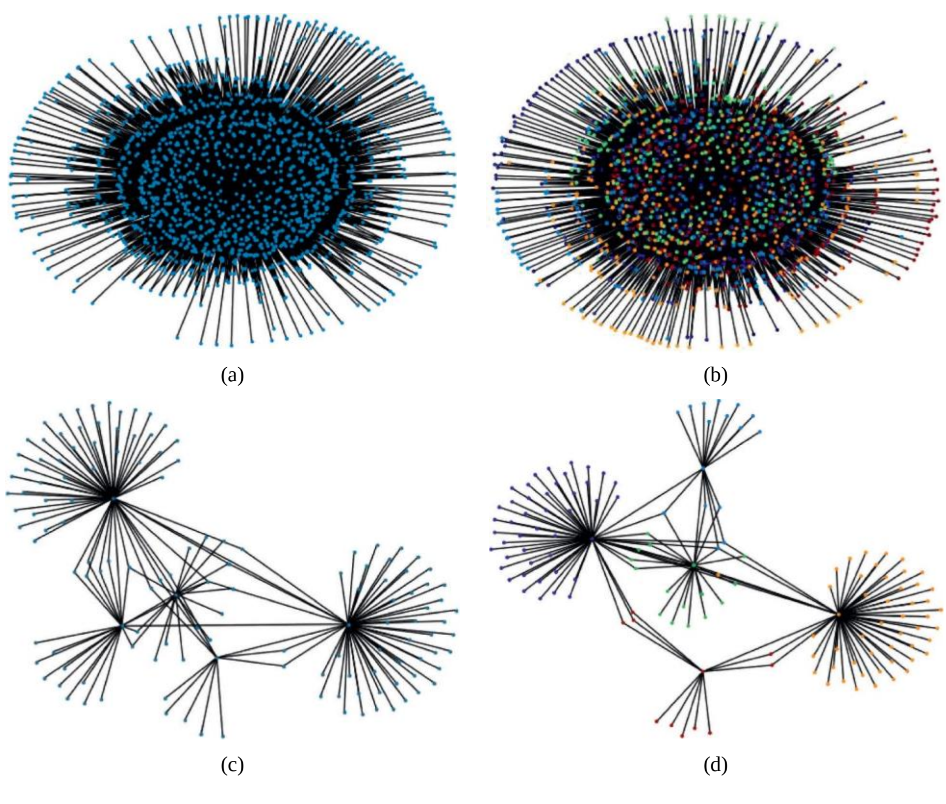

Figure 3: Analysis based on bipartite networks datasets using (a) original network using the Movie Lens 100 K dataset; (b) community detection using the Movie Lens 100 K dataset; (c) original network using the American Revolution dataset; (d) community detection using the American Revolution dataset.

corresponding fitness values computed by proposed RHHO with population 5, 10, 15, 20, 25, and 30 are 0.0000061, 0.000010, 0.000�12, 0.000016, 0.000031, and 0.000063.
Likewise, when the iteration is 250, then the corresponding fitness values computed by proposed RHHO with pop- ulation 5, 10, 15, 20, 25, and 30 are 0.00076, 0.00576, 0.07740� 0.88287, 0.97519, and 0.99201, respectively. The analysis of the proposed RHHO based on the modularity metric is illustrated in Figure 6(b). When the iteration is 1, the corresponding modularity values computed by proposed RHHO with population 5, 10, 15, 20, 25, and 30 are 0.000419, 0.00204, 0.00213, 0.00284, 0.00303, and 0.00376.
Likewise, when the iterati�n is 250, then the corresponding modularity values computed by proposed RHHO withpopulation 5, 10, 1�, 20, 25, and 30 are 0.6069, 0.6205, 0.6398, 0.6404, 0.6542, and 0.6790, respectively.

5.4.4. Performance Analysis Based on Bipartite Network Dataset Using American Revolution. Figure 7 illustrates the performance analysis of the proposed RHHO using American Revolution-based fitness and modularity mea- sures. The analysis of the proposed RHHO based on the fitness metric is portrayed in Figure 7(a). When the iteration is 1, the corresponding fitness values computed by proposedRHHO with population � 5, 10, 15, 20, 25, and 30 are

0.000021, 0.000026, 0.000036, 0.000044, 0.000051, and 0.00040. Likewise, when the iteration is 250, then the cor- responding fitness values computed by proposed RHHO with population 5, 10, 15, 20, 25, and 30 are 0.0016, 0.0017, 0.0069, 0.5085, �.9140, and 0.9962, respectively. The analysis of the proposed RHHO based on the modularity metric is illustrated in Figure 7(b). When the iteration is 1, the corresponding modularity values computed by proposed RHHO population 5, 10, 15, 20, 25, and 30 are 0.00023, 0.00300, 0.00311, 0.00327, 0.00436, and 0.00496. Likewise, when the iteration i�250, then the corresponding modularityvalues computed by proposed RHHO with population 5, 10, 15, 20, 25, and 30 are 0.57925, 0.61982, 0.64765, 0.65844, � 0.66412, and 0.67721, respectively.

4.5. Competing Methods. The methods, such as h-index- based link prediction [9], MAGA [31], GA [36], MATMCD- BN [28], and HHO [33], are employed for the comparison with the proposed RHHO.

4.6. Comparative Analysis. The comparative analysis of the proposed model is performed by evaluating the performance of other techniques based on fitness and modularity. The analysis is conducted by varying the number of iterations.

5.6.1. Analysis Based on Citation Networks Dataset Using cit- HepPh. Figure 8 illustrates the analysis of the methods using cit-HepPh based fitness and modularity measures. The analysis of the methods based on the fitness metric is portrayed in Figure 8(a). When the iteration is 1, the cor- responding fitness values computed by h-index-based link prediction, MAGA, GA, MATMCD-BN, HHO, and pro- posed RHHO are 0.0000513, 0.00183, 0.47324, 0.48179, 0.50100, and 0.51568. Likewise, when the iteration is 250, the

corresponding fitness values computed by h-index-basedlink prediction, MAGA, GA, MATMCD-BN, HHO, and proposed RHHO are 0.130475, 0.574730, 0.59360, 0.60131,
0.64708, and 0.66048, respectively. The analysis of methodsbased on modularity metric is illustrated in Figure 8(b). When the iteration is 1, the corresponding modularity valuescomputed by h-index-based link prediction, MAGA, GA, MATMCD-BN, HHO, and proposed RHHO are 0.000759, 0.00109, 0.00138, 0.00190, 0.00211, and 0.00285. Likewise,
when the iteration is 250, the corresponding modularity values computed by h-index-based link prediction, MAGA, GA, MATMCD-BN, HHO, and proposed RHHO are 0.36686, 0.37422, 0.37475, 0.38952, 0.39476, and 0.77560, respectively.

5.6.2. Analysis Based on Citation Networks Dataset Using cit- HepTh. Figure 9 illustrates the analysis of the methods using cit-HepTh based fitness and modularity measures. The analysis of the methods based on the fitness metric is portrayed in Figure 9(a). When the iteration is 1, the cor- responding fitness values computed by h-index-based link prediction, MAGA, GA, MATMCD-BN, HHO, and pro- posed RHHO are 0.000118, 0.00095, 0.36191, 0.40654,0.43435, and 0.48270. Likewise, when the iteration is 250, the corresponding fitness values computed by h-index-based link prediction, MAGA, GA, MATMCD-BN, HHO, and proposed RHHO are 0.36032, 0.51114, 0.54192, 0.55247, 0.63780, and 0.69778, respectively. The analysis of the methods based on modularity metric is illustrated in Figure 9(b). When the iteration is 1, the corresponding modularity values computed by h-index-based link pre- diction, MAGA, GA, MATMCD-BN, HHO, and proposed RHHO are 0.000375, 0.00083, 0.00101, 0.00120, 0.00134, and 0.00298. Likewise, when the iteration is 250, the *5.6.3. Analysis Based on Bipartite Network Dataset Using Movie Lens 100 K.* Figure 10 illustrates the analysis of the methods using Movie Lens 100 K-based fitness and mod- ularity measures. The analysis of the methods based on the fitness metric is portrayed in Figure 10(a). When the iter- ation is 1, the corresponding fitness values computed by h-index-based link prediction, MAGA, GA, MATMCD-BN, HHO, and proposed RHHO are 0.000035, 0.00069, 0.39608,

0.41400, 0.41713, and 0.49527. Likewise, when the iteration is250, then the corresponding fitness values computed byh-index-based link prediction, MAGA, GA, MATMCD-BN,HHO, and proposed RHHO are 0.04802, 0.31454, 0.52216,
0.58928, 0.64062, and 0.68944, respectively. The analysis ofthe methods based on modularity metric is illustrated in Figure 10(b). When the iteration is 1, the corresponding modularity values computed by h-index-based link pre- diction, MAGA, GA, MATMCD-BN, HHO, and proposed RHHO are 0.00035, 0.00129, 0.00163, 0.00168, 0.00196, and
0.00553. Likewise, when the iteration is 250, then the cor- responding modularity values computed by h-index-based link prediction, MAGA, GA, MATMCD-BN, HHO, and *5.6.4. Analysis Based on Bipartite Network Dataset Using American Revolution.* Figure 11 illustrates the analysis of the methods using Movie Lens 100 K-based fitness and modularity measures. The analysis of the methods based on the fitness metric is portrayed in Figure 11(a). When the iteration is 1, the corresponding fitness values computed by

h-index-based link prediction, MAGA, GA, MATMCD-BN,HHO, and proposed RHHO are 0.00053, 0.00142, 0.23572,
0.26296, 0.3476, and 0.4756. Likewise, when the iteration is250, then the corresponding fitness values computed byh-index-based link prediction, MAGA, GA, MATMCD-BN,HHO, and proposed RHHO are 0.31137, 0.36367, 0.47276,
0.64272, 0.73190, and 0.74353, respectively. The analysis ofthe methods based on modularity metric is illustrated in Figure 11(b). When the iteration is 1, the corresponding modularity values computed by h-index-based link

Table 1: Comparative results of proposed RHHO method with other existing methods in terms of modularity and fitness.

Dataset	Metrics	h-index-based link prediction	MAGA	GA	MATMCD-BN	HHO	Proposed RHHO
cit-HepPh	Fitness	0.130475	0.574730	0.59360	0.60131	0.64708	0.66048
	Modularity	0.36686	0.37422	0.37475	0.38952	0.39476	0.77560
cit-HepTh	Fitness	0.36032	0.51114	0.54192	0.55247	0.63780	0.69778
	Modularity	0.34853	0.37638	0.37948	0.38210	0.38734	0.77807
Movie Lens 100 K	Fitness	0.04802	0.31454	0.52216	0.58928	0.64062	0.68944
	Modularity	0.36727	0.37546	0.37596	0.39952	0.41170	0.77191
American Revolution	Fitness	0.31137	0.36367	0.47276	0.64272	0.73190	0.74353
	Modularity	0.36400	0.38391	0.38523	0.39717	0.39896	0.77433

prediction, MAGA, GA, MATMCD-BN, HHO, and pro- posed RHHO are 0.000207, 0.00116, 0.00190, 0.00217, 0.00224, and 0.00275. Likewise, when the iteration is 250, then the corresponding modularity values computed by h-index-based link prediction, MAGA, GA, MATMCD-BN,HHO, and proposed RHHO are 0.36400, 0.38391, 0.38523, 0.39717, 0.39896, and 0.77433, respectively.

5.7. Comparative Discussion. Table 1 deliberates the com- parative analysis of proposed RHHO with other existing methods in terms of modularity and fitness. The analysis is done by considering citation network and bipartite network dataset. Considering cit-HepPh, the maximal fitness is computed by proposed RHHO with 0.66048 whereas the fitness values of existing h-index-based link prediction, MAGA, GA, MATMCD-BN, and HHO are 0.130475, 0.574730, 0.59360, 0.60131, and 0.64708. The maximal modularity is computed by proposed RHHO with 0.77560 whereas the modularity values of existing h-index-based linkprediction, MAGA, GA, MATMCD-BN, and HHO are 0.36686, 0.37422, 0.37475, 0.38952, and 0.39476, respec- tively. Likewise, considering cit-HepTh, the maximal fitnessis 0.69778, and maximal modularity is 0.77807. Based on Movie Lens 100 K, the maximal fitness is 0.68944, and maximal modularity is 0.77191. Similarly, considering American Revolution, the maximal fitness is 0.74353 and maximal modularity is 0.77433. It is also observed that the proposed RHHO outperformed other methods with maxi- mal fitness of 0.74353 and maximal modularity of 0.77433 considering American Revolution network, respectively.

5. Conclusion

Community discovery is an important yet challenging job in complicated network structures, such as bipartite networks. In a bipartite network, community identification not only helps to comprehend the network's fundamental operations and, by extension, how to propose content, but it also discloses insights about the hierarchical structure of a complex network. An approach to community detection in bipartite networks that takes the node similarity measure into account is presented in this study. By making certain adjustments to the h-index using the Tversky index, a new measure called the h-Tversky may be developed to verify the degree of similarity between nodes. To further enhance the system's convergence rate, a new approach called Rider-Harris Hawks Optimisation (RHHO) was developed for community discovery. This algorithm was created by integrating the RO and HHO algorithms. A newly-designed fitness that takes modularity and the suggested h-Tversky index into account analysing the degree of similarity between nodes. The suggested community detection's accuracy may be calculated by include modularity in the fitness function, which measures the quality of specific network divisions. With a maximum robustness of 0.74353 and maximal modularity of 0.77433, the suggested technique demonstrated effective performance utilising the American Revolution network from the bipartite network dataset. Machine learning researchers may use it to aggregate stock market or social network data into sets with shared attributes, such as a bipartite network, and then extract those sets for various purposes. Eventually, we may use the MapReduce method to find the communities that overlap, meaning they share more than one node.

Data Availability

The Movie Lens 100 K dataset is publicly available at https://grouplens.org/datasets/movielens/ and Ameri- can Revolution network dataset is publicly available at http://konect.cc/networks/brunson_revolution/.

Conflicts of Interest

The authors declare no conflicts of interest about the publication of this research article.

Acknowledgments

This research was supported by the Researchers Supporting Project number (RSP-2021/244), King Saud University, Riyadh, Saudi Arabia.

References

[1] S. Zhang, G. Jin, X.-S. Zhang, and L. Chen, "Discovering functions and revealing mechanisms at molecular level from biological networks," *Proteomics*, vol. 7, no. 16, pp. 2856–2869, 2007.

[2] S. Zhang, J. Zhao, and X.-S. Zhang, "Common community structure in time-varying networks," *Physical Review E*, vol. 85, no. 5, 2012.

[3] Z. Li, R.-S. Wang, S. Zhang, and X.-S. Zhang, "Quantitative function and algorithm for community detection in bipartite networks," *Information Sciences*, vol. 367-368, no. 1, pp. 874–889, 2016.

[4] H. Gmati, A. Mouakher, A. Gonzalez-Pardo, and D. Camacho, "A new algorithm for communities detection in social networks with node attributes," *Journal of Ambient Intelligence and Humanized Computing*, pp. 1–13, 2018.

[5] D. Chen, W. Zhao, D. Wang, and X. Huang, "Similarity-basedlocal community detection for bipartite networks," *Filomat*, vol. 32, no. 5, pp. 1559–1570, 2018.

[6] L. Xiong, G.-Z. Wang, and H.-C. Liu, "New community es- timation method in bipartite networks based on quality of filtering coefficient," *Scientific Programming*, vol. 2019, Article ID 4310561, 12 pages, 2019.

[7] F. Chang, B. Zhang, Y. Zhao, S. Wu, and K. Yoshigoe,"Overlapping community detecting based on complete bi- partite graphs in micro-bipartite network bi-egonet," *IEEE Access*, vol. 7, pp. 91488–91498, 2019.

[8] C. Carusi and G. Bianchi, "Scientific community detection viabipartite scholar/journal graph co-clustering," *Journal of Informetrics*, vol. 13, no. 1, pp. 354–386, 2019.

[9] W. Zhou, J. Gu, and Y. Jia, "h-index-based link prediction methods in citation network," *Scientometrics*, vol. 117, no. 1,pp. 381–390, 2018.

[10] C. Zhou, L. Feng, and Q. Zhao, "A novel community detection method in bipartite networks," *Physica A: Statistical Me- chanics and Its Applications*, vol. 492, pp. 1679–1693, 2018.

[11] L. Feng, C. Zhou, and Q. Zhao, "A spectral method to find communities in bipartite networks," *Physica A: Statistical Mechanics and Its Applications*, vol. 513, pp. 424–437, 2019.

[12] M. Girvan and M. E. J. Newman, "Community structure in social and biological networks," *Proceedings of the National Academy of Sciences*, vol. 99, no. 12, pp. 7821–7826, 2002.

[13] V. D. Blondel, J.-L. Guillaume, R. Lambiotte, and E. Lefebvre, "Fast unfolding of communities in large networks," *Journal ofStatistical Mechanics: Theory and Experiment*, vol. 2008, no. 10, Article ID P10008, 2008.

[14] E. A. Corrêa Jr., A. A. Lopes, and D. R. Amancio, "Word sense disambiguation: a complex network approach," *Information Sciences*, vol. 442-443, pp. 103–113, 2018.

[15] W. Tang and P. Daoutidis, "Network decomposition for distributed control through community detection in input- output bipartite graphs," *Journal of Process Control*, vol. 64,pp. 7–14, 2018.

[16] T. Viard, R. Fournier-S'niehotta, C. Magnien, and M. Latapy,"Discovering patterns of interest in IP traffic using cliques in bipartite link streams," in *Proceedings of the 2018 Interna- tional Workshop on Complex Networks*, Boston, MA, USA, March 2018.

[17] Z.-A. Huang, Y.-A. Huang, Z.-H. You, Z. Zhu, and Y. Sun, "Novel link prediction for large-scale miRNA-lncRNA in- teraction network in a bipartite graph," *BMC Medical Ge- nomics*, vol. 11, no. S6, pp. 17–27, 2018.

[18] S. Rechner, L. Strowick, and M. Müller-Hannemann, "Uni- form sampling of bipartite graphs with degrees in prescribed intervals," *Journal of Complex Networks*, vol. 6, no. 6, pp. 833–858, 2018.

[19] W. Guan, X. Wen, L. Wang, Z. Lu, and Y. Shen, "A service-oriented deployment policy of end-to-end network slicing based on complex network theory," *IEEE Access*, vol. 6,pp. 19691–19701, 2018.

[20] T. Bian and Y. Deng, "Identifying influential nodes in complex networks: a node information dimension approach," *Chaos: An Interdisciplinary Journal of Nonlinear Science*, vol. 28, no. 4, Article ID 043109, 2018.

[21] Z. Gao, K. Zhang, W. Dang et al., "An adaptive optimal- Kernel time-frequency representation-based complex net- work method for characterizing fatigued behavior using the SSVEP-based BCI system," *Knowledge-Based Systems*,vol. 152, pp. 163–171, 2018.

[22] X. Huang, D. Chen, T. Ren, and D. Wang, "A survey of community detection methods in multilayer networks," *Data Mining and Knowledge Discovery*, vol. 35, no. 1, pp. 1–45, 2021.

[23] M. Rostami, K. Berahmand, and S. Forouzandeh, "A novel community detection based genetic algorithm for feature selection," *Journal of Big Data*, vol. 8, no. 1, pp. 1–27, 2021.

[24] X. Li, Y. Chen, and J. Xu, "Convex relaxation methods for community detection," *Statistical Science*, vol. 36, no. 1, pp. 2–15, 2021.

[25] H. Joo, M. Lee, J. Kim, J. Jung, J. Kwak, and H. S. Kim, "Streamgauge network grouping analysis using community

detection," *Stochastic Environmental Research and Risk Assessment*, vol. 35, no. 4, pp. 781–795, 2021.

[26] A. Clauset, M. E. J. Newman, and C. Moore, "Finding community structure in very large networks," *Physical Review E*, vol. 70, no. 6, Article ID 066111, 2005.

[27] R. Rashmi, S. Champawat, G. Varun Teja, and K. Lavanya, "Analysis of road networks using the louvian community detection algorithm," in *Soft Computing for Problem Solving*, pp. 749–757, Springer, Berlin, Germany, 2020.

[28] S. Che, W. Yang, and W. Wang, "A memetic algorithm for community detection in bipartite networks," *IEEE Access*, vol. 7, pp. 126897–126912, 2019.

[29] H.-l. Sun, E. Ch'ng, X. Yong, J. M. Garibaldi, S. See, and D.-b. Chen, "A fast community detection method in bipartite networks by distance dynamics," *Physica A: Statistical Mechanics and Its Applications*, vol. 496, pp. 108–120, 2018.

[30] C. Xue, S. Wu, Q. Zhang, and F. Shao, "An incremental group-specific framework based on community detection for cold start recommendation," *IEEE Access*, vol. 7, pp. 112363–112374, 2019.

[31] W. Zhan, Z. Zhang, J. Guan, and S. Zhou, "Evolutionary method for finding communities in bipartite networks," *Physical review E*, vol. 83, no. 6, Article ID 066120, 2011.

[32] D. Binu and B. S. Kariyappa, "RideNN: a new rider optimization algorithm-based neural network for fault diagnosis in analog circuits," *IEEE Transactions on Instrumentation and Measurement*, vol. 68, no. 1, pp. 2–26, 2019.

[33] A. A. Heidari, S. Mirjalili, H. Faris, I. Aljarah, M. Mafarja, and H. Chen, "Harris hawks optimization: algorithm and applications," *Future Generation Computer Systems*, vol. 97, pp. 849–872, 2019.

[34] "Citation networks dataset," 2019, https://snap.stanford.edu/data/#citnets-cit-HepPh.

[35] "Bipartite network dataset," 2019, http://konect.uni-koblenz.de/-movielens-1m.

[36] M. Tasgin, A. Herdagdelen, and H. Bingol, "Community detection in complex networks using genetic algorithms," 2007, https://arxiv.org/abs/0711.0491.