ITC 1/52
Information Technology and Control
Vol. 52 / No. 1 / 2023
pp. 53-67
DOI 10.5755/j01.itc.52.1.31535

Community Detection by Node Betweenness Using Optimized Girvan-Newman Cuckoo Search Algorithm

Received 2022/06/01 | Accepted after revision 2022/08/01

https://doi.org/10.5755/j01.itc.52.1.31535

HOW TO CITE: Devi, S., Rajalakshmi, M. (2023). Community Detection by Node Betweenness Using Optimized Girvan-Newman Cuckoo Search Algorithm. *Information Technology and Control*, 52(1), 53-67. https://doi.org/10.5755/j01.itc.52.1.31535

Community Detection by Node Betweenness Using Optimized Girvan-Newman Cuckoo Search Algorithm

S. Devi, M. Rajalakshmi

Department of Information Technology, Coimbatore Institute of Technology, Coimbatore-641 014, India

Corresponding author: devi.s@cit.edu.in

Due to technological development, social media platforms like forums and microblogs allow people to share their experiences, thoughts, and feelings. The organization, shopping groups etc. has major discussions regarding their business advertisements and product reviews. Also, there are certain followers for particular person or group due to their interests. Here the major issue is to know who or which group in social media is more influenced. The social media analysis needs to perform for identifying influenced person in the social media. The influencer node/person detection in a certain community is already done using greedy algorithm, genetic algorithm, ant colony optimization, cuckoo search algorithms. These existing techniques takes more time for diffusion and accuracy in prediction is not satisfied by users. To overcome this issues, in this research influencer node is identified using optimized Girvan Newman Cuckoo Search Algorithm (GNCSA). First Grivan Newman is used to identify the community and perform community detection. Cuckoo search algorithm uses host bird strategy in finding cuckoo eggs in his nest. Based on the centrality measure it decides whether the node is an influencer or not. This paper proposed Influencer detection by forming community first and measures angular centrality using optimized Girvan Newman cuckoo search algorithm. Our proposed work GNCSA gives a better accuracy rate for the data sets of Dolphin 0.89, for Facebook dataset got 0.93, Twitter data set got 0.94 and for YouTube data set 0.92, karate club and football got 0.91. This proposed work increases the intracommunity of the social network and improves its performance accurately by detecting the influencer in the social network.

KEYWORDS: Community detection, Girvan Newman, cuckoo search algorithm, node optimization, swarm intelligence, social networks.

1. Introduction

The influencer in the social media is defined as person/node who have more followers and subscribers in the social medias. Grivan Newman is widely used technique to identify the community in the social media. The cuckoo search is swarm intelligent technique which is preferred to identify inflenced node in the detected community. In the social network analysis, community detection is an active area for the flow of dynamic information. The network structure of community detection is in the form of nodes with high probability, and it links with another group of nodes [10]. In the complex network, analysis had done in various fields like text analysis, identification of the user, and Key attribute of community detection [15]. Many research works study complex structures and the Girvan Newman algorithm based on hierarchical clustering for the complex network structure. The basic concept of this algorithm helps to detect the edges linked with various communities in a graph and discard them to get segmentation [19, 26].

Online social media platforms like Facebook, Instagram, Twitter, Snapchat, etc., are the most popular websites which contain millions or billions of users. They can share dynamic information via highly connected user groups called communities. These communities are a group of users with similar characteristics or interests like cooking, living city, collegemate, and communicating with each other. The structure of social media platform is like a graph network with nodes denoting the social users and edges representing the link between them [23]. The topology network itself describes the transfer of information between two users via a path between users and data flow based on the community's detection parameter.

Community detection has acquired many methods to discover the communities and centrality in complex networks. This research has come across popular techniques like label propagation technique, edge detection methods, and spectral clustering techniques. The problems in the above techniques are discussed one by one. First, although the label propagation method is used to run at good speed, it cannot produce stable results. If we study graph networks, they take more time due to the long network tracing strategy. But the advantage is it gives good accuracy. Clustering was more probably used in community detection. This algorithm is probably used due to its similarity matrix computation. The main limitation of this technique is, that when network complexity is increased then the accuracy was decreased. By considering the above issues, we propose node betweenness integrated with the optimization technique. The main role of node betweenness is to identify nodes influenced in the network. Finally, optimization is performed to identify the centrality of the community.

The main motivation of this research is to bring the solution to many economical and business marketing problems. By finding the influencer person, sensitive news can be reached more users. Community detection plays a vital role in controlling the opinions of users and convincing them to improve any particular business or any social welfare-related news. In a community, all users have influenced by a specific node or person due to their similar thoughts. As previously discussed there are a lot of algorithms for detecting the influencers in social media. The drawback of existing algorithms is inaccurate, cost complex, and unstable in detecting influencers of complex networks. This paper proposed detecting the influencer node in each community by node betweenness to overcome these issues using an optimized GNCSA. The main contribution of this proposed work is:

1 Design a Grivan Newman model for detecting community in less time and determine the influence node using the Cuckoo search algorithm.
2 Results are implemented and evaluated on proposed performance is done on various standard datasets.

The paper has been organized as follows: Section 2 describes the review of the literature, Section 3 describes the community detection using GNCSA, Section 4 discusses experimented results, and Section 5 concludes the paper with future directions.

2. Review of Literature

Community detection helps to detect the complex network with a dynamic flow of information. This complex network can be defined by a graph with nodes and edges between the nodes. A node represents individuals, and the node's relationship is considered an edge for identifying the influence node in the network graph by getting high similarity between nodes. Using an influence node, it cannot wholly grasp the hidden information in the network. For the depth analysis of entire networks composed of nodes and edges, community structure helps to detect the community in the graph [16, 17, 27]. There are many existing algorithms available in the detection of community.

The K-means clustering algorithm is used for handling multiple clusters with low error functions. It is easy implementation, fast clustering, and effective classification of large data sets. CNN is used for managing complex networks in the detection of community [6, 20, 30]. The advantage of community detection is making the graph network unique and different from another. And also, it helps to understand the complex system. There are many community detection algorithms for interpreting the graph structure network and detecting the communities. It uses the centrality measure of edges between nodes to see the neighborhood. To expand the community detection around the seed node in the graph network. The traditional algorithm is based on the concept of heuristics techniques such as minimum spanning trees, page ranks, and input streaming for graphs with seed node expansion in the graph network [18, 25].

Al-Andoli et al. [1] proposed deep learning methods for handling community detection in extensive complex data and parallel metaheuristic data. Chen et al. [7] present that incremental algorithm with the Coherent Neighborhood Propinquity (CNP) technique to detect community in the graph network structure. This technique exploits community detection in real-time by eliminating the nodes and edges and including the nodes and edges in the network. Kamakshi et al. [13] proposed solving the broadcast storm problems in the Vehicular Ad hoc Network (VANET) using clustering techniques to detect community. The social media platform can create a less dynamic or static network architecture using the clustering technique. Table 1 shows a survey on community detection on the social media platform.

Table 1

Survey on community detection on the social media platform

Author	Scope of Research	Classifier Model	Algorithms
Xu et al. [31] (2022)	Discarding the high edge betweenness centrality value	Modularity	Divisive
Cheng et al. [8] (2021)	Multiple attribute decision-making strategy,	Density peak clustering model	TOPSIS (Technique for Order Preference by Similarity to an Ideal Solution)
Ji et al. [11] (2020)	Eliminating intercommunity edges	Partitioned clustering	Divisive
El Kouni et al. [9] (2020)	Detach unused labels and improving the validity of community	Overlapping community detection	Label Propagation Algorithm (LPA)
Chen et al. [12] (2020)	Nodes collected by local communities	Dynamic network	Dynamical membership function
Lu et al. [22] (2020)	Implement a series of operation	density peak clustering	DPCNMF and DPCSNMF
Zeng et al. [32] (2019)	Consensus Community-based Particle Swarm Optimization (CCPSO) is used for dynamic community detection	Dynamic	Particle Swarm Optimization (PSO)
Tian et al. [28] (2019)	Detection of optimal overlapping community	Overlapping community detection	Fuzzy clustering
Li et al. [21] (2019)	Integrate the influence node relationship between multi-layer communities	Dynamic	Clustering

The influencer detection in the complex social network was implemented using Genetic algorithm (GA) [3]. Main issue of GA is, it is difficult to design an objective function for complex designs. GA takes more time to generate operators and detection of influencer. The article [14] uses fuzzy theory for predicting influential community in the social media. Modularity is defined for influential node identification [5]. The centrality measures to identify the most influencer in the community are discussed in article [4].

3. Methodology

In the social network consider a graph $G = (V, E)$ with vertices $V = \{v_1, v_2, v_3, \ldots, v_m\}$ denoted as nodes or participants and edges are represented as links between two nodes and $node_j$ in the network. A community is a group of entities created by individuals who interact within the group and rarely with outside groups. Interaction between entities within the group is called closeness, and the similarity or distance between them measures it. In the social network, communities are called clusters, and each node in the cluster is associated with the same group or various groups in the network. Therefore, in the community, the node common to more than one group is called overlapping communities.

3.1. Node Betweenness

In the global network, node betweenness reflects the influencer of a node. That is higher node betweenness and passing many node pairs through the node in the network, and also it is like center of the community. Let us consider a network with node n and the node betweenness of this n node is evaluated as follows:

$$G_B(v) = \sum_{p \neq v \neq q \in V} \frac{\sigma_{pq}(v)}{\sigma_{pq}}. \tag{1}$$

Here σ_{pq} represent the shortest paths via node v. In the network, the graph is an unweighted graph, and calculating the betweenness of the node is done by assigning the weight of the undirected graph is 1.

3.1.1. Structural Similarity

In the network, it resembles the similarity of two nodes and it is defined by:

$$simil(p,q) = \frac{2 * |N(p) \cap N(q)|}{|N(p)| + |N(q)|}. \tag{2}$$

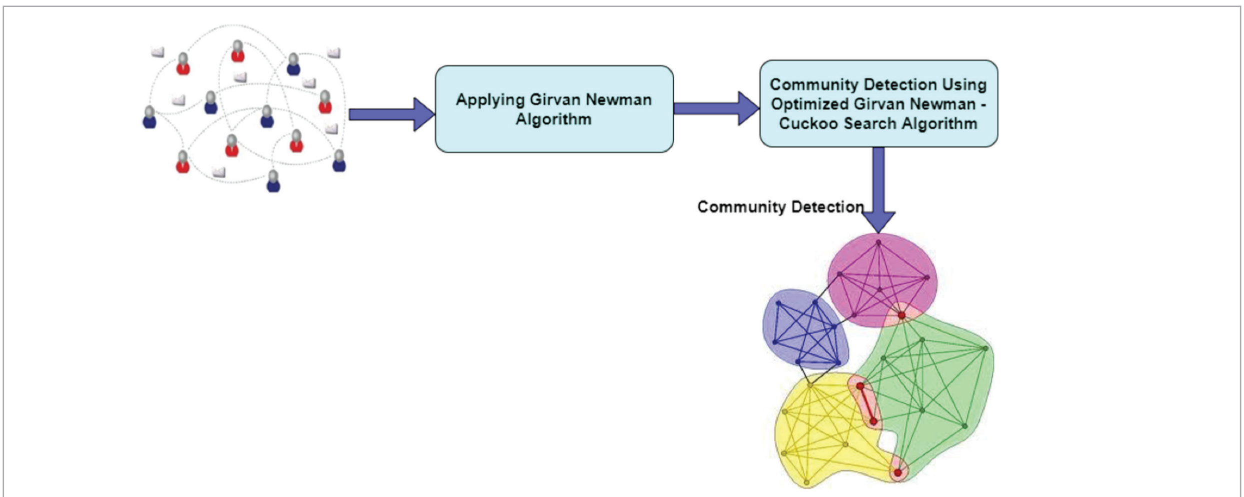
Here p, q are the nodes in the graph network G and $N(p)$ and $N(q)$ are set of nodes that are direct neighborhoods of the nodes p, q.

3.2. Basic Steps in the Detection of Community

The complex network-based community detection algorithm is implemented using the Girvan Newman, with cuckoo search. The outline of the proposed work is given in Figure 1.

Figure 1 uses Girvan Newman with a cuckoo search algorithm for community detection. The compo-

Figure 1
Outline of proposed work



nents of the social network are nodes and edges. These nodes represent an individual participant in the network, and links between two nodes are edges. Based on the input graph, it detects the community structure, and each community may be nested (Overlapped). The implementation of the proposed work uses three concepts:

1. Choosing a neighborhood node as a similar node in the network to expand the community.
2. Based on the similarity of nodes merging and creating communities.
3. Detection of influence node or central node in the network

3.2.1. Choosing a Neighborhood Node as a Similarity Node in the Network for Expanding the Community

If node similarity with its neighborhood node is higher than other communities, then that node is selected for the network. For identifying the similarity between nodes by evaluating the average similarity of nodes and comparing it with their neighbour. For expanding the community, choose the candidate neighbour node using average similarity. In measuring the similarity of two nodes in the network using local structure information by:

$$simi(p,q)=\sum\nolimits_{ct\in\Gamma(p)\cap\Gamma(q)}\frac{1}{\log ck_{ct}}. \tag{3}$$

Here, ct is the common neighborhood node between p and q. The average similarity of a node with its neighbour node is evaluated by:

$$avg(p,q)=\frac{1}{ck_p}\sum\nolimits_{q\in\Gamma(p)}simi(p,q). \tag{4}$$

For expanding the community, comparing the node and then neighbour node is included in the community. The similarity nodes within the same community are larger than other node that is not included in the community.

3.2.2. Based on Similarity of Nodes Merging and Creating Communities

In the community $Commu_i$ with a maximum number of nodes are gathered as a small community. Nodes that are in the same community are closely linked with one another. In this way, small community is merging with highest similarity of nodes in the community. In general modularity maximization and label propagation are two techniques which are used in community similarity measures as well as used in optimization techniques for detecting community in the network structure.

$$merge(Commu_i, Commu_j)= \left(nn_q\left(q\in Commu_j\cap\Gamma(ct, ct\in Commu_i)\right)\right). \tag{5}$$

Here, n_q is the number of nodes in the community $Commu_j$ and which is merged with another community $Commu_i$.

3.2.3. Detection of Influence Node or Central Node in the Network

Influence node in the network can be determined by three ways. They are degree centrality, closeness centrality, and betweenness centrality. In this proposed work, we implement degree centrality.

3.2.3.1. Degree Centrality

The influence node in the network is attracted by its neighborhood node. Therefore, the central node is considered a higher influence in the network. In the network, some node has a greater degree and lesser node will be influenced by their neighbor node. Thus $1/\deg node_q$ is used to represent the influence node p with its neighbor. The influence node p can be evaluated as the sum of the influence with its neighbors:

$$N(p)=\sum\nolimits_{q\in\Gamma(p)}\frac{1}{\deg node_q}. \tag{6}$$

Here, $\Gamma(p)$ is the set of neighbor node p and $\deg node_q$ is the degree of node q.

Sort the nodes in the network with respect to by their influence node and select the first node as the central node in the community of the network.

3.2.3.2. Closeness Centrality

In the detection of the influence node in the network, closeness centrality is used to track the close node to the given input node and other remaining nodes in the network. For the given node $node_i$ is closeness centrality $close_{node}(i)$ and it can be expressed as:

$$close_{node}(i)=\frac{n-1}{\sum_{j=1}^{m}dist_{ij}}. \tag{7}$$

Here $dist_{ij}$ is the distance between node i and node j and n is the total number of nodes in the network.

3.2.3.3. Betweenness Centrality

It is used to measure the betweenness of centrality in the network and capturing the shortest path between the nodes in the network. It is defined as:

$$cent_B(i) = \sum_{i=1}^{n} \frac{\sigma_{sp}(i)}{\sigma_{sp}}. \tag{8}$$

Here, σ_{sp} is the shortest path between nodes in the network from node s and node p.

3.3. Girvan Newman Algorithm

The concept of this algorithm is top-down hierarchical based community detection and also divisive method. It starts with input graph and repeatedly eliminates the edges with highest betweenness at each time. The number of shortest paths that pass through an edge in a network is called as edge betweenness. Similarly, vertex betweenness is the number of shortest paths passing through the vertex. Steps involved in Girvan-Newman algorithm is given below:

Step 1: In the input graph and for every edge evaluate the edge betweenness centrality.

Step 2: Discard the edge with highest value of edge betweenness.

Step 3: Compute edge betweenness centrality for all remaining edges in the network

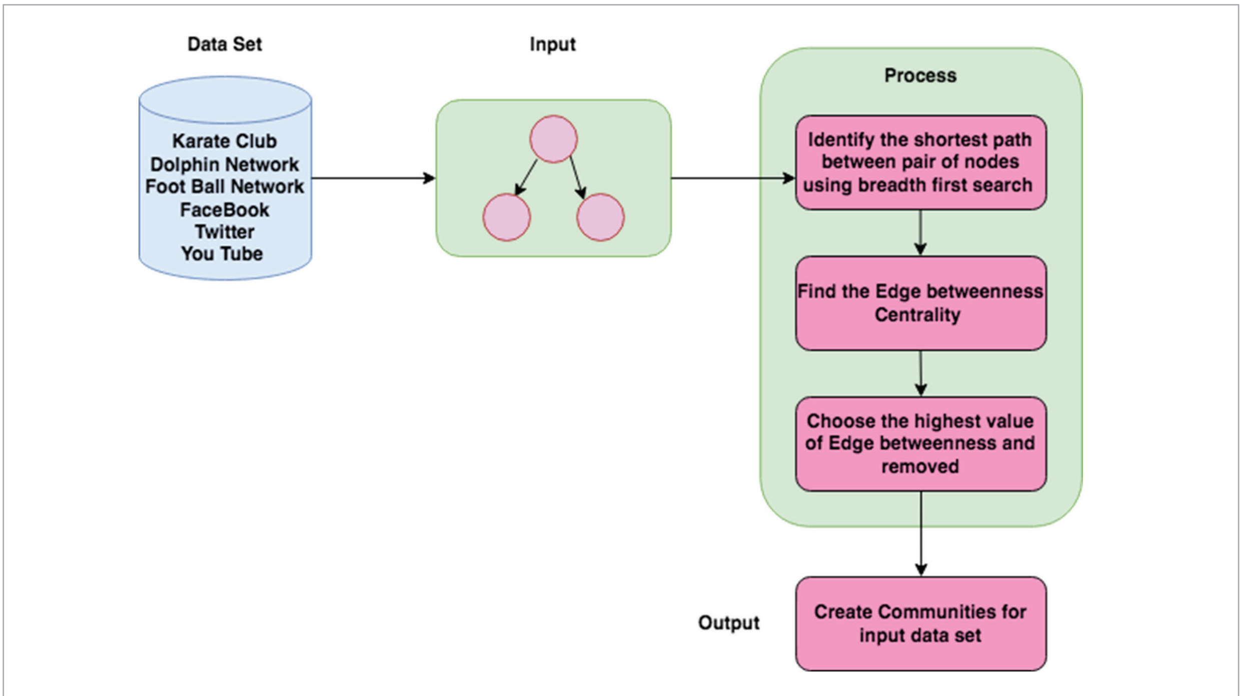
Step 4: Repeat the steps 2-3 until all edges are discarded.

To compute edge betweenness centrality by using the shortest path in the network. It starts from one vertex and calculates edge weights for paths that passes through vertex and repeat it for all vertexes in the input graph. Finally, it sums up the weight of all edges in the graph. To calculate the edge betweenness centrality by using two phases.

Phase 1: Using breadth-first search, find the distance between the source vertex to all vertices in the graph and also identify the number of shortest paths. (Marking Vertex in the graph)

Phase 2: Choose the maximum distance from the source vertex and calculate the number of shortest paths passing via edges.

Figure 2
Architecture of Girvan Newman algorithm



For every vertex $ver_i \in V$ in the graph network calculate the triple ($dist_i$, sw_i, sb_i). Here is the distance from source vertex, sw_i is the number of shortest paths from the source vertex to the remaining vertex ver_i, sb_i is the number of shortest paths between the source to any vertex in a graph that passes through vertex.

Phase 1: Marking the vertex in the graph network

Step 1: Initial vertex $iver_i \in V$; $dist_{si} = 0$, $sw_{si} = 1$ $sb_i = 0$

Step 2: Let $dist_v = \infty$, $sw_v = 1$, $sb_v = 0$

for $ver_i \neq iver_i \in V$.

Step 3: Generate queue *Que, Que* $\longleftarrow \{iver_i\}$. Generate a list *Lst, Lst* $\leftarrow \{iver_i\}$

Step 4: While *Que* is not empty

Step 4.1: Dequeue $i \leftarrow Que$

Step 4.2: For each vertex $verj \in Adja(ver_i)$ //set of all vertices adjacent to V.

Step 4.2.1: If $dist_{sj} = \infty$ then $dist_{sj}$ +1, $sw_{sj} = sw_{si}$. Enqueue $j \rightarrow Que$ Push $j \rightarrow Lst$

Step 4.2.2: If $dist_{sj} \neq \infty$ and $dist_{sj} = dist_{sj} + 1$ then $sw_j = sw_j + sw_i$

Step 4.2.3: If $dist_{sj} \neq \infty$ and $dist_{sj} < dist_{sj}$ +1 then, do nothing.

The above algorithm describes that identify the shortest path from the source vertex to visit all vertices in the graph. In the phase 2 it starts from the vertex which was last visited vertex by Phase 1 and traverse to visit the vertices in the reverse order of the Phase 1 visited order. And it also identifies the shortest path from the source vertex via last visited vertex in Phase 1.

Phase 2: Evaluating Edge betweenness centrality

Step 1: While *Lst* is not empty:

Step 1.1: Pop vertex $ver_i \leftarrow Lst$

Step1.2: For each vertex $ver_j \in Adja(Ver_i)$:

Step 1.2.1: If $dist_{sj} < dist_{sj}$ then $sb_i = 1 + \sum_j \sigma_{ij}$

Step 1.2.2: If $dist_{sj} < dist_{sj}$ then $\sigma_{ij} = \frac{sw_j}{sw_i} * sb_i$

Both Phases are performed for all source vertices and calculate edge betweenness centrality for all edges as a sum of edge betweenness centrality values in each step. This Girvan-Newman algorithm has input data set converted into a number of nodes and edges as an input graph and produces community structure as an output. After calculating the edge betweenness centrality value, the edge with the highest value is eliminated. It repeats until there are no edge remains.

3.4. Cuckoo Search Algorithm

It describes the meta-heuristic Cuckoo Search algorithm. By nature, to lay eggs it randomly selects the nest. In order to implement it parasitic habit follows three rules:

Rule 1: Cuckoo randomly selects another bird's nest and lay one egg at a time.

Rule 2: In order to move over next-generation, a nest with high-quality eggs is selected.

Rule 3: The available host nest is a fixed one and the egg laid by a cuckoo is identified by the host bird and the probability $pb_c \in [0, 1]$. If the alien egg is discovered by the host bird, then either it can throw away the egg or discard the nest and build a new nest. Worst eggs are replaced by the new egg. To represent the D-dimensional optimization problem by $X = x_1, x_2, ..., x_d$.

The new search space X_{newi}^{t+1} at a time t is calculated based on the above said rules as follows:

$$X_{newi}^{t+1} = X_{newi}^{t} + \alpha \otimes levy(step, \lambda), \tag{9}$$

where $\alpha > 0$ is the scaling factor of step size, $levy(step, \lambda)$ is the step-length and its probability distribution is defined as,

$$levy(step, \lambda) = \frac{\lambda \Gamma(\lambda) \sin(\pi\lambda / 2)}{\pi} \frac{1}{step^{1+\lambda}}. \tag{10}$$

In Equation (8) probability distribution has an infinite variance with an infinite mean value.

The implementing the concept of cuckoo search algorithm is given below:

Step 1: Initialization the population of n host nests x_i, $i = 1, 2, 3, ..., n$

Step 2: $c = 1$

Step 3: While $c \leq stop\_criteria$ *or* Max_Iter **do**

Step 4: Evaluate fitness value $fit(x_{new})$ in each nest

Step 5: New solution X_{newi}^{t+1} is generated by Levy flight using Equations (9)-(10).

Step 6: Select the candidate solution x_{oldi}^{t}

Step 7: IF $f(x_{oldi}^{t}) > f(X_{newi}^{t+1})$ **then**

Step 8: Replace x_{oldi}^{t} by new solution X_{newi}^{t+1}

Step 9: End IF

Step 10: Probability $\left(prob_a\right)$ of worst nest is abandoned and built a new one.

Step 11: Ranking solution and choose the current best

Step 12: End While

In the Algorithm 2, describes that the solution is based on egg in the nest and replacing the new solution is defined by cuckoo egg. The main aim of this algorithm is replacing the worst solution by new one. The position of egg is represented by *pos_egg* consist of m elements. Egg laid by a cuckoo is identified by the host bird and the probability $pb_c \in [0, 1]$. The assignment value between and is represented and detect the relation between individual and shows that these two individuals belong to the same community.

3.5. Community Detection by Girvan-Newman Cuckoo Search Algorithm (Proposed)

The detection of community in the network of proposed work contains three stages namely expanding the community, merging the communities, and detecting the influence node. The drawback of using the Girvan-Newman algorithm is time consumption is high, difficult to handle complex networks, and not favour huge volume of data. To overcome these issues, in this work we optimized the community by implementing the cuckoo search algorithm. Therefore, for the optimized community detection based on Girvan-Newman with Cuckoo search algorithm (GNCSA). The implementation steps of the proposed work GNCSA are given below:

Step 1: Read the input graph.

Step 2: Instead of using breadth-first search for finding shortest path in Girvan Newman algorithm, in this proposed work GNCSA, uses fitness function of cuckoo search algorithm to calculate edge betweenness centrality by using best fitness function. (Using Cuckoo Search Algorithm in Section 3.4)

Step 3: Eliminate the highest fitness value of each edge in the network by using Equation (3).

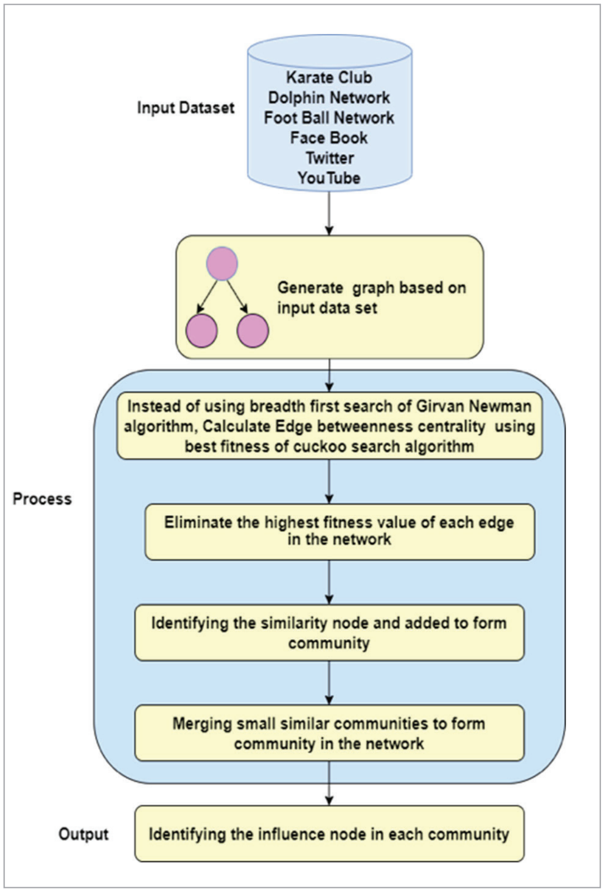
Step 4: Identifying the similarity node to create a community (group).

Step 5: Merging all similar small communities to detect the community in the network by using Equation (5).

Step 6: Identifying the influence node or central node based on Influence Node by using Equation (6).

In Figure 3, it read the input graph; for evaluating the edge betweenness centrality using the cuckoo search algorithm's fitness function. Based on the high best fitness function it eliminates the edge in the network. Repeat the process until all high fitness value edges are discarded.

Figure 3
Workflow of proposed work



4. Result and Discussions

4.1. Data Set

In this section, we use ground-truth datasets to evaluate the performance of the proposed work of the GNCS Algorithm. Spyder Python 3.8 is used for the implementation of this algorithm. To implement our proposed work for the data sets of Facebook, Twitter, and data sets, are available in Stanford University SNAP, Dolphin network, American college football network, karate club, YouTube data set [24]. Table 2 shows the information about the data set for the evaluation of metric measures.

Table 2
Details of Data Set

Data Set	Nodes	Edges	No. of Communities
Facebook	3,959	1,68,486	7,498
Twitter	81,306	17,68,149	58,352
YouTube	11,34,889	29,87,623	8,385
Karate Club	34	78	2
Dolphin Network	62	318	2
American College Football Network	115	1,226	15

4.2. Evaluation of Performance

The community detection in the social network using proposed work GNCSA algorithm uses ground truth communities to evaluate the metric measures like F1-Score, Accuracy, Normalized Mutual Information (NMI), Modularity and it is applied in Girvan Newman algorithm [29] and Cuckoo Search algorithm [2].

F1-Score

In the input graph network $G(V, E)$, the set of ground truth communities $comm^*$ and set of community detection are $\hat{comm}$, each ground-truth community $comm_i \in comm^*$ and each detection of community $\hat{comm}_i \in \hat{comm}$ F1-score of matching of community detection with each ground truth community:

$$F1-score=\frac{1}{2}\left(\begin{array}{l}\frac{1}{|comm^*|}\sum_i \max_j F1\left(comm_i,\hat{comm}_j\right)\\ +\frac{1}{|\hat{comm}|}\sum_i \max_j F1\left(comm_j,\hat{comm}_i\right)\end{array}\right). \quad (11)$$

Here F1($comm_i$, $\hat{comm}_j$), is the harmonic mean of precision and $comm_i$, $\hat{comm}_j$ is recall.

Normalized Mutual Information (NMI)

It is used to measure the similarity of two communities in the social network. It can be evaluated as:

$$NMI=\frac{-2\sum_{i=1}^{comm_x}\sum_{j=1}^{comm_y} NN_{ij}\log\left(\frac{NN_{ij}NN}{NN_i NN_j}\right)}{\sum_{i=1}^{comm_x} NN_i\log\left(\frac{NN_i}{NN}\right)+\sum_{j=1}^{comm_y}\log\left(\frac{NN_j}{NN}\right)}. \quad (12)$$

Here $comm_x$ is the number of original communities, $comm_y$ is the number of communities identified, NN is the number of nodes in the network, NN_{ij} is the number of nodes in the real community that partitions, and found a community that partitions y. NN_i denotes the sum of row matrix of NN_{ij}. NN_j denotes the sum of the column matrix.

Modularity (Q)

It is used to measure the performance of community detection with respect to the unknown community labels in the network. It is evaluated by:

$$MQ=\frac{1}{2n}\sum_{i,j\in Ver}\left(Adj_{i,j}-\frac{d(i)d(j)}{2n}\right)\times \delta\left(label_i,label_j\right) \quad (13)$$

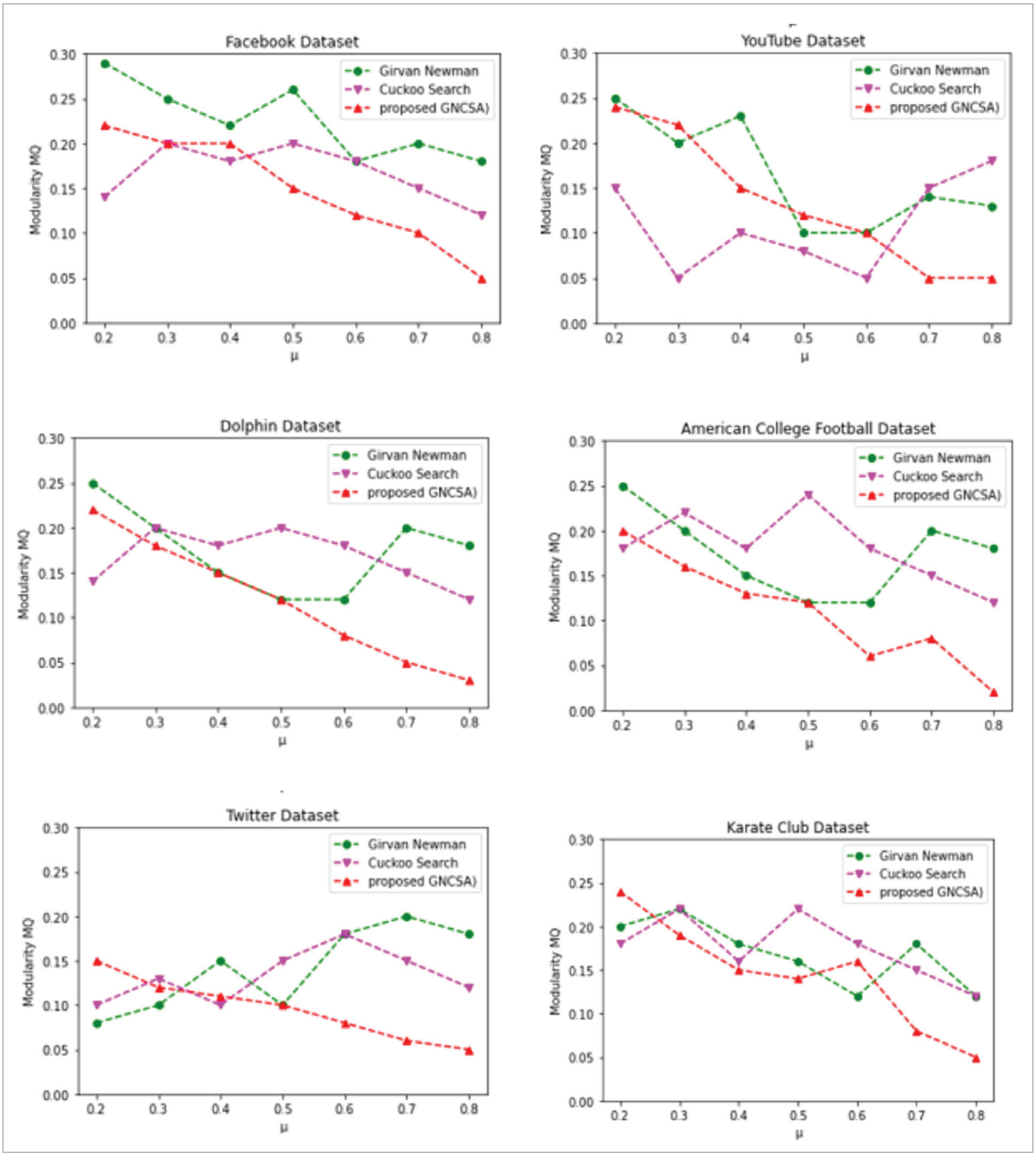
Here denotes the modularity, n is the number of edges in the network, Adj is the adjacency matrix of the network. the vertices and ver_j are connected directly then else $Adj_{i,j}$ = 0. Similarly, $label_i$, $label_j$ are labels of the community of the vertices ver_i and ver_j. If then $\left(label_i,label_j\right)=1$ else $\delta\left(label_i,label_j\right)=0$.

The overlapping modularity parameter is implemented as shown in Figure 4.

As can be seen from Figure 4, μ represents the hybrid parameter in the network and for applying in the overlapping modularity. In general, the range is between 0 to 1. It provides edge connection between nodes inside the community and nodes outside the community. The best community structure in the network is smaller. Our proposed work GNCSA provides best result and for dolphin data set it decreased significantly when $\mu > 0.3$, Facebook data set $\mu > 0.4$, for Twitter data set $\mu > 0.3$ and for YouTube dataset $\mu > 0.4$, American college football data set $\mu > 0.5$ and karate club data set $\mu > 0.6$ .It seems that our proposed algo-

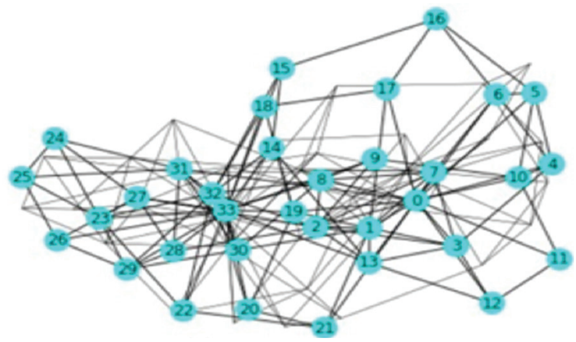
Figure 4

Modularity in hybrid parameter μ

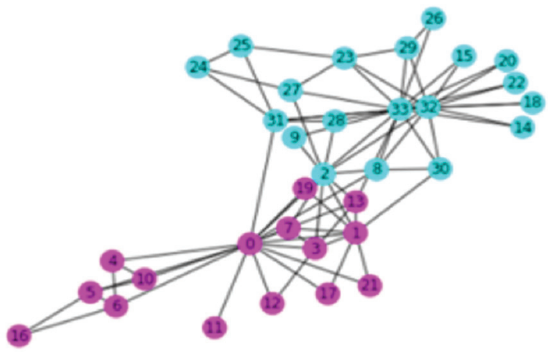


rithm has good adaptability for the detection of community in the complex network architecture. Figure 5 shows that sample community structure of proposed work for the various data set. The usage of μ parameter is a hybrid parameter is increase complexity of network. With the mixed parameters μ increase in network complexity and improving the detection of community. ¡ Figure 5(a)-(b) shows that sample output of karate club network dataset in proposed work GNCSA.

Figure 5



(a) Before applying proposed GNCSA



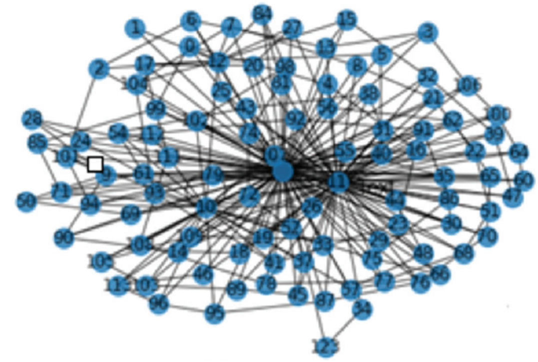
(b) After applying proposed work GNCSA in karate club network data set

In the observation of Figure 5 (a)-(b), Our proposed work gives better result in the detection of community. Figure 5(b) shows that Node 33 and Node 0 is considered as influence node for their community. Figure 6 shows that sample of football data set in proposed work GNCSA

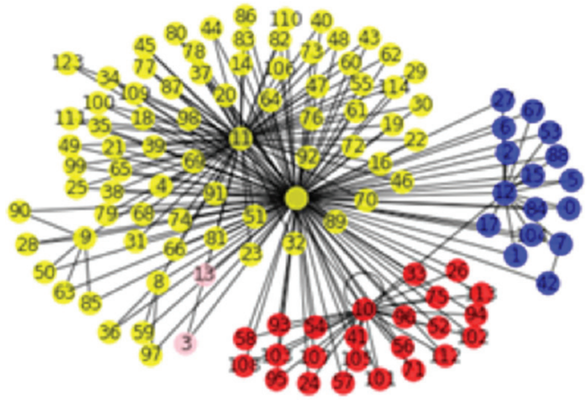
Figure 6(b) shows that Node 11, Node10 and Node 15 are considered as influence node for their community. Figure (a)-(b) shows that sample output of Facebook dataset in proposed work GNCSA.

Figure 7(b) shows that Node 1998, Node 1999 and Node 2000 are considered as influence node for their community of Facebook dataset. Figure 8 shows that computation time for proposed work with existing algorithm.

Figure 6



(a) Before applying proposed GNCSA

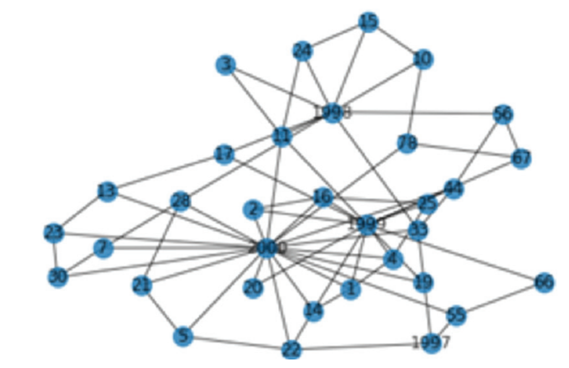


(b) After applying proposed work GNCSA in Football data set

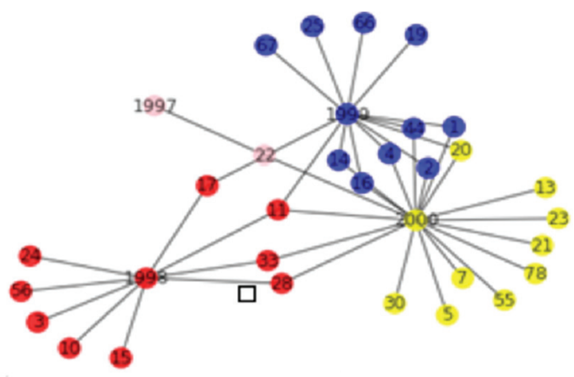
Figure 8 shows that computation time for detecting community in the social network for the various data sets like Dolphin, Facebook, tweet and YouTube, karate club, football data set. It seems that out proposed work GNCSA needs less execution time for detection of community in the social network. Figure 9 shows that accuracy rate of implementing various algorithm for different data sets.

In the observation of Figure 9, our proposed work GNCSA gives better accuracy rate for the data sets of Dolphin0.89, for Facebook dataset got 0.93, Twitter

Figure 7



(a) Before applying proposed GNCSA



(b) After applying proposed work GNCSA in Facebook data set

data set got 0.94 and for YouTube data set 0.92, karate club and football got0.91. Table 3 shows that evaluation criteria of F1-score for the various algorithm with different data sets.

Table 3
F1-Score for different datasets

	F1-Score		
Data Set	Girvan Newman	Cuckoo Search	GNCSA (Proposed)
Dolphin	0.8945	0.9231	0.9552
Facebook	0.9033	0.9381	0.9432
Twitter	0.8898	0.9185	0.9376
YouTube	0.8921	0.9056	0.9288
Karate Club	0.8854	0.9054	0.9133
Foot Ball	0.9087	0.9238	0.9487

Figure 8
Computation time

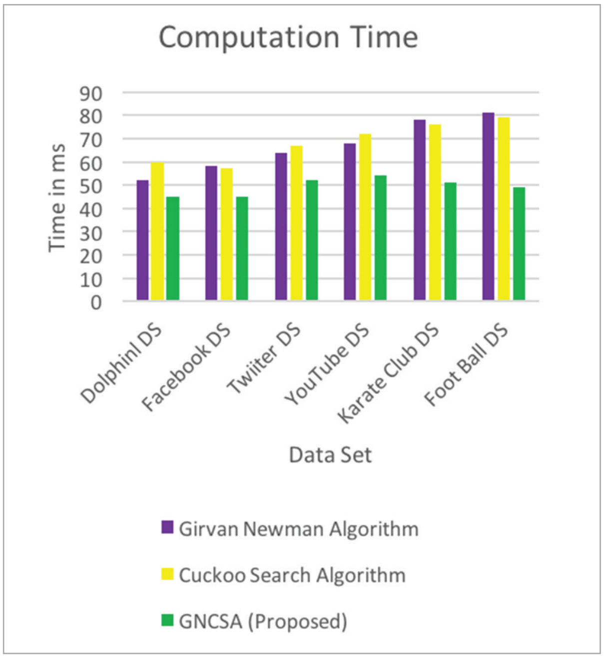
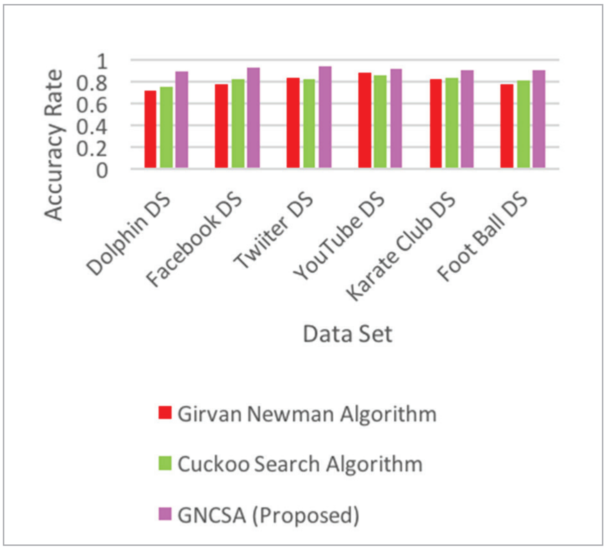


Figure 9
Accuracy Rate



In Table 3, our proposed algorithm provides prominent result in the criteria of F1-score for the Dolphin data set got 0.9522, Facebook got 0.9432, Twitter got 0.9376, YouTube data set got 0.9288, karate club 0.9133and football got 0.9487. Similarly, for the Girvan Newman algorithm produces for the datasets of Dolphin got 0.8945, Facebook got 0.9033, Twitter got 0.8898, YouTube data set got 0.8921, karate club 0.8854and football got 0.9087. For cuckoo search algorithm produces for the various datasets of dolphin got 0.9231, Facebook got 0.9381, Twitter got 0.9185, YouTube data set got 0.9056, karate club 0.9054and football got 0.9238. Table 4 shows the evaluation criteria of the NMI score for the various algorithm with different data sets.

Table 4
NMI-Score for different datasets

	NMI-Score		
Data Set	Girvan Newman	Cuckoo Search	GNCSA (Proposed)
Dolphin	0.9145	0.9464	0.9652
Facebook	0.9286	0.9245	0.9531
Twitter	0.9198	0.9285	0.9323
YouTube	0.9066	0.9164	0.9321
Karate Club	0.9156	0.9187	0.9033
Foot Ball	0.9187	0.9122	0.9101

In Table 4, our proposed algorithm provides prominent results in the criteria of NMI-score for the Dolphin data set got 0.9652, Facebook got 0.9531, Twitter got 0.9323, YouTube data set got 0.9321, Karate Club got 0.9033 and football data set got 0.9101. Similarly, the Girvan Newman algorithm produces for the datasets of Dolphin got 0.9145, Facebook got 0.9286, Twitter got 0.9198, YouTube data set got 0.9066, Karate Club got 0.9156, and football data set got 0.9187. For cuckoo search algorithm produces for the various datasets of Dolphin got 0.9464, Facebook got 0.9245, Twitter got 0.9285, YouTube data set got 0.9164, Karate Club got 0.9187and football data set got 0.9122.

5. Conclusion

This paper proposed an efficient community detection of the social network using Girvan-Newman with the Cuckoo search algorithm. This work collected data from various data sets like Dolphin, Twitter, Facebook, and YouTube. This work selects more nodes for the expansion of community detection. The performance analysis of this work is based on the aspects of accuracy, NMI, modularity, and F1-score. Our proposed work GNCSA gives a better accuracy rate for the data sets of Dolphin 0.89, for Facebook dataset got 0.93, Twitter data set got 0.94 and for YouTube data set 0.92, karate club and football got 0.91. This work is extended to detect community-based on trust-based topic with various similarity measures in the future.

References

1. Al-Andoli, M. N., Tan, S. C., Cheah, W. P., Tan, S. Y. A Review on Community Detection in Large Complex Networks from Conventional to Deep Learning Methods: A Call for the Use of Parallel Meta-Heuristic Algorithms. IEEE Access, 2021, 9.96501-96527.https://doi.org/10.1109/ACCESS.2021.3095335
2. Babers, R., Hassanien, A. E. A Nature-Inspired Metaheuristic Cuckoo Search Algorithm for Community Detection in Social Networks. International Journal of Service Science Management Engineering and Technology, IGI Global, 2017, 8(1), 50-62. https://doi.org/10.4018/IJSSMET.2017010104
3. Behera, R. K., Naik, D., Rath, S. K., Dharavath, R. Genetic Algorithm-Based Community Detection in Large-Scale Social Networks. Neural Computing and Applications, 2020, 32(13), 9649-9665. https://doi.org/10.1007/s00521-019-04487-0
4. Behera, R. K., Naik, D., Sahoo, B., Rath, S. K. Centrality Approach for Community Detection in Large Scale Network. Proceedings of the 9th Annual ACM India Conference, 2016, 115-124. https://doi.org/10.1145/2998476.2998489
5. Behera, R. K., Rath, S. K. An Efficient Modularity Based Algorithm for Community Detection in Social Network. 2016 International Conference on Internet of Things and Applications (IOTA), 2016, 162-167. https://doi.org/10.1109/IOTA.2016.7562715
6. Cai, B., Wang, Y., Zeng, L., Hu, Y., Li, H. Edge Classification Based on Convolutional Neural Networks for Community Detection in Complex Network. Physica A: Statistical Mechanics and Its Applications, 2020, 556. https://doi.org/10.1016/j.physa.2020.124826
7. Chen, N., Hu, B., Rui, Y. Dynamic Network Community Detection with Coherent Neighborhood Propin-

quity. IEEE Access, 2020, 8, 27915-27926. https://doi.org/10.1109/ACCESS.2020.2970483

8. Cheng, J., Wang, X., Gong, W., Li, J., Chen, N., Chen, X. Community Detection Based on Density Peak Clustering Model and Multiple Attribute Decision-Making Strategy TOPSIS. Complexity, 2021, 1-18. https://doi.org/10.1155/2021/1772407
9. El Kouni I. B, Karoui, W, Romdhane, L. B. Node Importance Based Label Propagation Algorithm for Overlapping Community Detection in Networks.Expert Systems with Applications, 2020, 162, 1-13. https://doi.org/10.1016/j.eswa.2019.113020
10. George, R., Shujaee, K., Kerwat, M., Felfli, Z., Gelenbe, D., Ukuwu, K. A Comparative Evaluation of Community Detection Algorithms in Social Networks. Procedia Computer Science, Elsevier, 2020, 171(C), 1157-1165. https://doi.org/10.1016/j.procs.2020.04.124
11. Ji, Q., Li, D., Jin, Z. Divisive Algorithm Based on Node Clustering Coefficient for Community Detection. IEEE Access, 2020, 8, 142337-142347. https://doi.org/10.1109/ACCESS.2020.3013241
12. Jin, D., Zhang, B., Song, Y., He, D., Feng, Z., Chen, S., Musial, K. ModMRF: A Modularity-Based Markov Random Field Method for Community Detection. Neurocomputing, 2020, 405, 218-228. https://doi.org/10.1016/j.neucom.2020.04.067
13. Kamakshi, S., Shankar Sriram, V. S. Modularity Based Mobility Aware Community Detection Algorithm for Broadcast Storm Mitigation in VANETs. Ad Hoc Networks, 2020, 104. https://doi.org/10.1016/j.adhoc.2020.102161
14. Kumari, A., Behera, R. K., Shukla, A.S., Sahoo, S. P., Misra, S., Rath, S. K. Quantifying Influential Communities in Granular Social Networks Using Fuzzy Theory. Computational Science and Its Applications, 2020, 906-917. https://doi.org/10.1007/978-3-030-58811-3_64
15. Li, Y., Su, Z., Yang, J., Gao, C. Exploiting Similarities of User Friendship Networks Across Social Networks for User Identification. Information Sciences: An International Journal, 2020, 506(C), 78-98. https://doi.org/10.1016/j.ins.2019.08.022
16. Li, Z., Ren, T., Ma, X., Liu, S., Zhang, Y., Zhou, T. Identifying Influential Spreaders by Gravity Model. Scientific Reports, 2019, 9. https://doi.org/10.1038/s41598-019-44930-9
17. Li, Z.., Ren, T., Xu, Y., Chang, B., Chen, D., Sun, S. Identifying Influential Spreaders Based on Adaptive Weighted Link Model. IEEE Access, 2020, 8, 66068-66073. https://doi.org/10.1109/ACCESS.2020.2985713
18. Liakos, P., Papakonstantinopoulou, K., Ntoulas, A., Delis, A. Rapid Detection of Local Communities in Graph Streams. IEEE Transactions on Knowledge and Data Engineering, 2022, 34, 2375-2386. https://doi.org/10.1109/TKDE.2020.3012608
19. Liu, C., Du, Y., Lei, J. ASOM-Based Membrane Optimization Algorithm for Community Detection. Entropy, 2019, 21(5). https://doi.org/10.3390/e21050533
20. Liu, Y., Yuan, X., Jiang, X., Wang, P., Kou, J., Wang, H., Liu, M. Dilated Adversarial U-Net Network for Automatic Gross Tumor Volume Segmentation of Nasopharyngeal Carcinoma. Applied Soft Computing, 2021,111. https://doi.org/10.1016/j.asoc.2021.107722
21. Liu, Z., Xiang, B., Guo, W., Chen, Y., Guo, K., Zheng, J. Overlapping Community Detection Algorithm Based on Coarsening and Local Overlapping Modularity. IEEE Access, 2019, 7, 57943-57955. https://doi.org/10.1109/ACCESS.2019.2912182
22. Lu, H., Sang, X., Zhao, Q., Lu. J. Community Detection Algorithm Based on Nonnegative Matrix Factorization and Improved Density Peak Clustering. IEEE Access, 2020, 8, 5749-5759. https://doi.org/10.1109/ACCESS.2019.2963694
23. Mahfoudh, A., Zardi, H., Haddar, M.A. Detection of Dynamic and Overlapping Communities in Social Networks. International Journal of Applied Engineering Research, 2018, 13(11), 9109-9122.
24. Nisa, M.U., Mahmood, D., Ahmed, G., Khan, S., Mohammed, M.A., Damaševičius, R. Optimizing Prediction of YouTube Video Popularity Using XGBoost. Electronics, 2021, 10(23), 2962. https://doi.org/10.3390/electronics10232962
25. Pagourtzis, A., Souliou, D., Potikas, P., Potika, K. Overlapping Community Detection Via Minimum Spanning Tree Computations. 2020 IEEE Sixth International Conference on Big Data Computing Service and Applications, 2020, 62-65. https://doi.org/10.1109/BigDataService49289.2020.00017
26. Rani, S., Mehrotra, M. Community Detection in Social Networks: Literature Review. Journal of Information and Knowledge Management, 2019, 18(2), https://doi.org/10.1142/S0219649219500199
27. Stam, C. Modern Network Science of Neurological Disorders. Nature Reviews Neuroscience, 2014, 15, 683-695. https://doi.org/10.1038/nrn3801

28. Tian, Y., Yang, S., Zhang, X. An Evolutionary Multi-Objective Optimization Based Fuzzy Method for Overlapping Community Detection. IEEE Transactions on Fuzzy Systems, 2020, 28(11), 2841-2855. https://doi.org/10.1109/TFUZZ.2019.2945241

29. Vispute, P., Sane, S. Performance Evaluation of Community Detection Algorithms in Social Networks Analysis. Bioscience Biotech Research Communications, 2020, 13(14), 388-393. https://doi.org/10.21786/bbrc/13.14/90

30. Wang, L., Yuan, X., Zong, M., Ma, Y., Ji, W., Liu, M., Wang, R. Multi-Cue Based Four Stream 3D ResNets for Video-Based Action Recognition. Information Sciences, 2021, 575, 654-665. https://doi.org/10.1016/j.ins.2021.07.079

31. Xu, Y., Ren, T., Sun, S. Community Detection Based on Node Influence and Similarity of Nodes. Mathematics, 2022, 10(6), 970. https://doi.org/10.3390/math10060970

32. Zeng, X., Wang, W., Chen, C., Yen, G. A Consensus Community-Based Particle Swarm Optimization for Dynamic Community Detection. IEEE Transactions on Cybernetics, 2020, 50(6), 2502-2513. https://doi.org/10.1109/TCYB.2019.2938895