Acta Informatica Pragensia
2022, Volume 11, Issue 3, pp. 380–395
Special Issue: Sustainable Solutions for Internet of Things Using Artificial Intelligence and Blockchain in Future Networks
https://doi.org/10.18267/j.aip.201

Article Open Access

Evaluation of Community Detection by Improving Influence Nodes in Complex Networks Using InfoMap with Sigmoid Fish Swarm Optimization Algorithm

Devi Selvaraj, Rajalakshmi Murugasamy

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

Corresponding author: Devi Selvaraj (devi.s@cit.edu.in)

Abstract

In recent years, community detection is important because members of the same community share the same concepts. For efficient community detection in a social network, the influence node plays a vital role. A node in the social network or a user that has great influence and power would have a close relationship with a core of the group, termed a community. Therefore, the status of a person is determined by the user's influence strength. That is, a user who has greater influence and strength plays a vital role in the social media community and also acts as a core in the community of the social network. Therefore, a community is a group of nodes in the complex network structure which are interlinked with one another. Effective community detection in a complex structure is a challenging task. Many studies have been done based on topological networks. The approaches are ineffective, inefficient and require more time to process. To overcome these issues, this paper proposes improving the influence nodes in complex networks by using the InfoMap with sigmoid fish swarm optimization algorithm (I-SFSO). Our proposed I-SFSO gives better accuracy rates for the data sets: 92% for Dolphin, 95% for the Facebook dataset, 96% for the Twitter data set, 94% for the YouTube data set, 93% for a karate club and 94% for football.

Keywords

Community detection; InfoMap; Influencer; Social media; Network; SFSO.

Citation: Selvaraj, D., & Murugasamy, R. (2022). Evaluation of Community Detection by Improving Influence Nodes in Complex Networks Using InfoMap with Sigmoid Fish Swarm Optimization Algorithm. *Acta Informatica Pragensia*, 11(3), 380–395. https://doi.org/10.18267/j.aip.201

Special Issue Editors: Venkatachalam Kandasamy, University of Hradec Kralove, Czech Republic
Mohamed Abouhawwash, Mansoura University, Egypt
Nebojsa Bacanin, Singidunum University, Serbia

Academic Editor: Zdenek Smutny, Prague University of Economics and Business, Czech Republic

1 Introduction

In communities with complex network architecture, detection is a great challenge because it can be applied in different application areas such as linguistics, biology, chemistry, etc. (Liu et al., 2020). The complex system can be represented as a complex graph in the form of connections between vertices and interactions between the vertices (Wajahat et al., 2020). The structure of a complex network consists of several nodes or vertices and links from one node to another. These links are referred to as edges. The link between two nodes is defined using a mathematical model and the computer science concept. Handling the complex network structure requires graphical data which represents the complex world system (Doreian et al., 2020). Detection of the community is based on the concept of data clustering. Data clustering combines the similarity of information and creates a group. The analysis of the cluster is based on the similarity pattern set which is represented as a vector or in a multidimensional data space (Doreian et al., 2020; Adolfsson, et al., 2019).

The social media structure consists of group followers and individual followers under the same concept of community. In the same community, users can share opinions and information, promoting products. In recent years, community detection is important because members of the same community share the same concepts. For efficient community detection in a social network, the influence node plays a vital role. A node in the social network or a user that has great influence and power would have a close relationship with a core of the group, termed a community. Identifying the influence node in social network activities and spreading information through the influence node is a great challenge. In the Twitter data set, information is based on hashtags. Through the hashtag, the same information is clubbed together and forms a group. In this way, the community can generate content for the same topic (Kowald et al., 2017; Li et al., 2016; Kou et al., 2018). Many datasets are used in the detection of communities. Fortunato et al (2010) proposed that the community detection strategy be classified into two categories, namely using the optimization method and the hierarchical clustering method. In the optimization-based community, detection is based on the concept of finding a possible solution for the complex structure data format. For more accurate community detection, optimization is implemented and at the same time, the evolutionary concept is also used in the construction of a community. In hierarchical clustering, the system is categorized into various hierarchical representations of various network formats at each level. A hierarchical clustering technique contains two types of classes, namely agglomerative algorithms and divisive algorithms (Fortunato et al., 2010; Ahmad et al., 2019; Yamada et al., 2020; Babichev, et al., 2019; Al-Sahaf et al., 2019; Elbes et al., 2019).

Many studies have been done based on topological networks (e.g., Raghavan et al., 2007; Wang et al., 2021). More straightforward approaches are often ineffective, inefficient and require more processing time. To overcome these issues, this paper proposes improving the influence nodes in complex networks by using the InfoMap with a sigmoid fish swarm optimization algorithm (I-SFSO). The main contribution of this work is as follows:

1. It generates communities with influence nodes using the InfoMap algorithm for various datasets.
2. It applies the optimization technique of sigmoid-based fish swarm optimization for merging two similar communities and to produce more accurate detection of influence nodes in the community effectively.
3. It evaluates the results of the analysis based on the concept of normalized mutual information (NMI) and the accuracy rate in the social network.

The remainder of this study is structured as follows: Section 2 reviews related work. Section 3 describes the proposed methodologies. In Section 4, we describe the analysis result evaluation methods. Section 5 presents the conclusion and future work.

2 Related Work

In recent years, the development of technology and the use of social networks, forums and blogs have become a part of human activity in which they can share their feelings, opinions and experience with products and discuss trending topics. For that, they need a structure called a community. Detection of the community and predicting the structure of the community for categorizing the same topic is essential. The community structure is in the form of a collection of nodes and sharing of information between the neighbouring nodes of the social network within the same community. Community detection and influence detection also play a vital role in the research areas of network science (Blondel et al., 2008). Many community detection algorithms are used. Influence spreaders in the network are essential for sharing information among various communities. Various centrality measures are implemented for different topological features of the network. However, the majority of information is ignored by its community structure. The centrality of nodes in the network structure comprises a non-overlapping community (Ghalmane et al., 2019). This non-overlapping community depends on two features: a local influence node and a global influence node. The local influence node structure depends on the community and the global influence node structure depends on other communities in the network structure. One of the stepping stones of community detection is the Girvan Newman algorithm. It is based on the concept of module-based network architecture. The modularity score is used to detect the strength of a community in the network (Chobe & Zhan, 2019).

A nature-inspired optimization algorithm for the detection of communities in social networks was proposed by Abduljabbar et al. (2020). Metaheuristic community detection algorithms were reviewed by Attea et al. (2021) and are based on evolutionary algorithms such as the genetic algorithm. Tran-Ngoc et al. (2021) described an ANN-based genetic algorithm for the detection of communities. It is an evolutionary algorithmic concept. It can be used in the field of solving simple data structures. Pan et al. (2021) presented an auto-encoder-based genetic algorithm with an evolutionary method. Lee et al. (2021), Sahan et al. (2021) described how a deep convolutional network (CNN) for community detection and influence maximization produces high performance in community classification.

3 Proposed Methodology

3.1 Preliminaries

In a social media platform, all users are influenced by others and at the same time they influence others. In this work, we represent being influenced as $BInf$, which is used to measure how much the user usr_i is influenced by its neighbours. We represent influence strength as $InfS$, which is used to measure how much the user usr_i can influence others. The main objective of this work is to identify the strength of influence nodes in a community detection system. It is represented by pure influence strength of the usr_i, which is considered the centrality of the user usr_i in the community. In this paper, the social media network is a composition of nodes and edges. Each node denotes an individual and edges represent interactions between nodes. The input pass on this algorithm is considered an undirected graph = $(ver, edges)$. Here, ver is the set of nodes and $edges$ is a set of edges. The link between two vertices in the undirected graph is denoted as:

$$edge = \{ver_1, ver_2\} \in UG. \tag{1}$$

Node neighbourhood in undirected graph

In the undirected graph $UG = (ver, edges)$, the neighbourhood of the node $node_i \in ver$ and the set of nodes in the UG are linked with one another.

Node circle in social network

The undirected graph is $UG = (ver, edges)$ and the circle of the node $node_i$ in the social network is denoted as $node_i \in ver$, which is the set of links containing both the node $node_i$ and its neighbours and is represented as:

$$s(node_i) = N(node_i) \cup \{node_i\}. \quad (2)$$

Intimate degree of community detection

It is defined as the link between two nodes such as $node_i$ and $node_j$ as follows:

$$\omega_{node_i\ node_j} = jaccard\_similarity(node_i\ , node_j) = \frac{|s(node_i) \cap s(node_j)|}{(node_i) \cup s(node_j)} \quad (3)$$

In Equation (1), $\omega_{node_i node_j}$ can be understood as the ratio of common user friends of the users $node_i$ and $node_j$ in their social platform.

3.2 Community detection by improving influence node

A complex network-based community detection algorithm is implemented using the InfoMap with sigmoid fish swarm optimization (I=SFSO). The architecture of the proposed work model is shown in Figure 1.

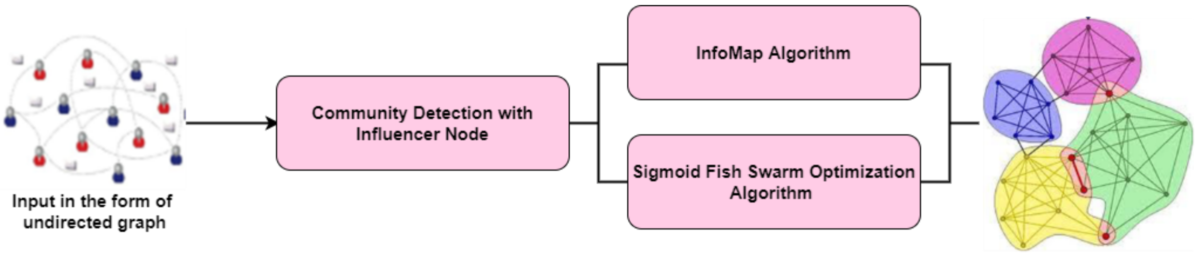

Figure 1. *Architecture of proposed work model.*

In Figure 1, community detection in a complex network requires input values in the form of an undirected graph for high quality of community detection based on influence nodes by using the InfoMap algorithm. From the output of the InfoMap algorithm, identifying the best influence node by optimizing it using the sigmoid fish swarm algorithm. The implementation of the proposed work (I-SFSO) is based on three concepts:

1. selection of the influence node,
2. initialization and expansion of the seed community,
3. community optimization for merging the community.

Figure 2 shows a flow diagram of the proposed work.

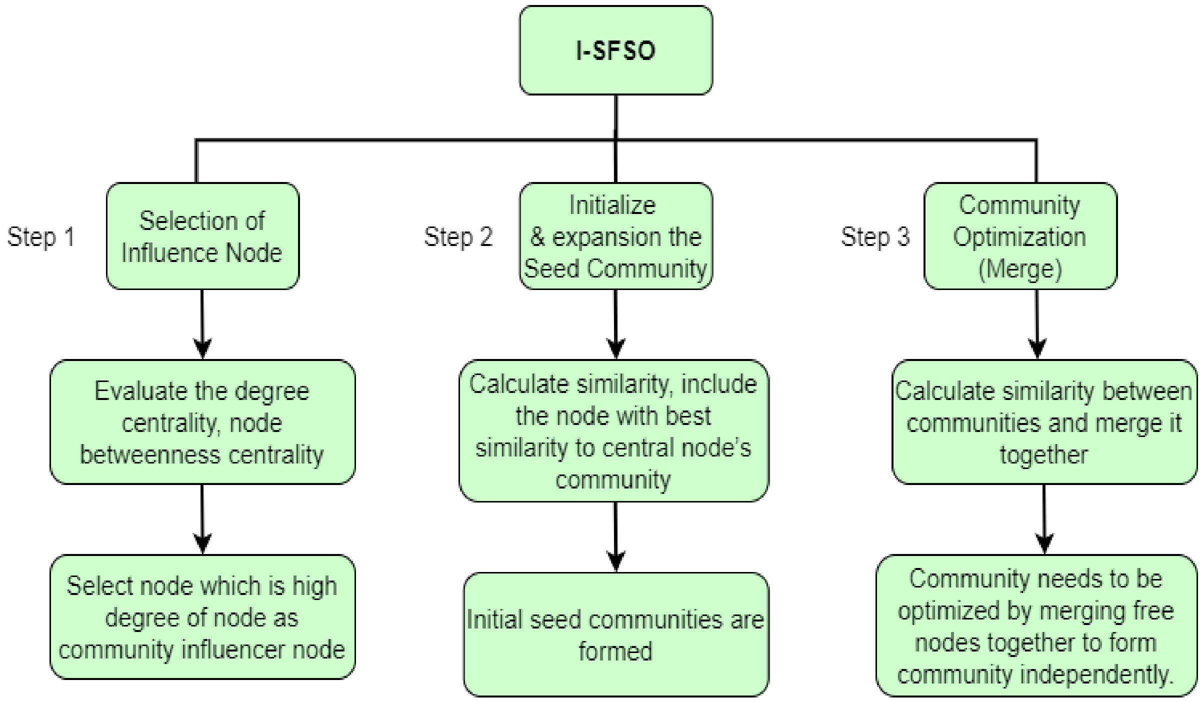

Figure 2. *Flow diagram of proposed work.*

3.2.1 Selection of influence node

The selection of the influence node is based on a node which has a high capability of spreading messages in the complex network and also has a great impact on its neighbours. Conversely, if any node has a lesser influence, it is affected by its neighbour nodes which have greater influence. In this work, the selection of the influence node is based on the degree of nodes and node betweenness centrality in the undirected complex network structure.

Degree of centrality

In an undirected network of dynamic data with a complex structure, the degree of a node is considered as the number of links of the particular node. If any node has a higher degree, it is considered an influence node or centrality node in the network. It can be evaluated by:

$$Degree\ of\ centrality(p) = \frac{degree(p)}{n-1} \tag{4}$$

Here, $degree(p)$ is the degree of the node p and n is the total number of nodes in the network.

Betweenness centrality

Betweenness centrality is used to detect the amount of influence at a node which has spread information in the graph. It is also used to identify nodes that act as bridges from one part of a community to another. In an unweighted graph, we calculate the shortest path between all pairs of nodes in the graph. Each node receives a score; based on that score, the number of shortest paths that pass through the node is detected. Nodes which more frequently lie on the shortest paths between other nodes will have higher betweenness centrality scores; it is calculated by:

$$bet_{cent(n)} = \sum_{st} \sigma_{st}{}^{(node)}/_{\sigma_{st}} \tag{5}$$

Here, σ_{st} is the shortest path between other nodes in a graph based on their scores.

Closeness centrality

The ancestor node, which is as an influence node with its locality, is a node that is close to many nodes and has a higher value of closeness centrality. This measure acts as a good heuristic for selecting the closeness centrality. Therefore, the centre node of a community has a higher closeness centrality.

$$dist_{m,n} = cf_1 p_{mn1} + cf_2 p_{mn2} + \cdots + cf_k p_{mnk} \tag{6}$$

Here, $dist_{m,n}$ is the closeness node between the node m and the node n. p_{mn1} is the total number of paths with the length k which connects the node m and the node n. k is the total number of lengths or hops. A higher value of closeness centrality of a node is determined closer between the node m and the node n. This can be rewritten as:

$$clc(i) = \frac{1}{\sum_x Dist(x,y)} \tag{7}$$

Here, $clc(i)$ is the closeness centrality of the node i and $\sum_x Dist(x,y)$ is the sum of distance from the node x to all other nodes.

Katz centrality

$$KC(i) = \left[\left(\sum_{j=0}^{\infty} \alpha^j B^j\right) C\right]_i \tag{8}$$

Here, $KC(i)$ is the Katz centrality of the node i. B^j is the power of the j^{th} adjacency matrix, α is the attenuation factor and C is the column vector; all values are 1.

Impact of influence centrality

The impact score is calculated as the number of Twitter followers nf_i divided by the total number of tweets nt_i. Based on Kendall's tau value, the centrality of the influence node is identified by the node impact score ranking 95 percent and above.

$$Ic_{impa} = \frac{nf_i}{nt_{i(0.95)}} \tag{9}$$

Followers centrality ($\boldsymbol{fc}$)

Centrality of the node is identified by tweet followers $nfoll$ having a score of 95 percent and above. It is defined by:

$$fc = nfoll_{0.95} \tag{10}$$

3.2.2 Initialization and expansion of seed community

In this work, we initialize the community and expand it based on similarity between nodes by using the InfoMap algorithm.

3.2.2.1 InfoMap

In an input undirected graph $G = (ver, edg)$, ver is a set of vertices and edg is a set of edges. The link between two vertices $(v1, v2)$ and its weight between two vertices are denoted as $w(v1, v2)$. In the undirected graph, a link is represented by 1. The InfoMap algorithm is used to identify the community, that is, identify the set of vertices. It contains spreading of information between high intracommunity and low intercommunity. Now, the non-overlapping community noc of an undirected graph is represented as:

$$\cup\, nc_i = V, \forall nc_i \in noc; nc_i \cap nc_j = \emptyset, \forall nc_i, nc_j \ \in noc \quad (11)$$

In Equation (8), nc_i is the communities which is used to identify the set of vertices V. noc is the non-overlapping community. Here, $\cup\, nc_i = V$ to find the best community, merging all communities. That is based on the number of intra-community and inter-community structure. When two communities are merged together, it has more intra-community structure. The objective function of the InfoMap algorithm is a map equation in which information is a flow and it is used to identify the set of vertices in a random walk of a graph. The clustering of sets of vertices in a random walk is done based on the community detection. To produce a high quality of community detection, the InfoMap algorithm uses minimum description length (MDL).
The objective map function is defined as:

$$LB(N) = bH(BP) + \sum_{n\in M} a^n H(a^n) \quad (12)$$

Here, N is the set of communities; in each community, b is the sum of exit probability in the graph, $H(B)$ is the average code length of movements between the communities, a is the stay probability of random walks in the community n. $H(a^n)$ is the average code length of the community for n. In the community structure M, $LB(N)$ is the lower bound of code length. The expanded map function of Equation (9) is given by:

$$LB(N) = \left(\sum_{n\in M} b_n\right)\log\left(\sum_{n\in M} b_n\right) - 2\sum_{n\in M} b_n \log(b_n) - \sum_{\alpha\in ver} a_\alpha \log(a_\alpha) + \sum_{n\in M}\left(b_n + \sum_{\alpha\in n} a_\alpha\right)\log\left(b_n + \sum_{\alpha\in n} a_\alpha\right) \quad (13)$$

Here, b_n is the exit probability of a community n, a_α is the visit probability of a vertex α in the random walk, and ver is the set of vertices in the graph.
In the undirected graph, vertexa_α and its relative weight wt_α are computed and the sum of the total weight of links which are connected to the vertex α is divided by twice the total weight of all links in the undirected graph. Now, the visit probability of a community n is defined as a_n; it is the relative weight of n. It is calculated by: $\sum_{\alpha\in n} a_\alpha$. The exit probability of n, b_n is the total relative weight between the communities.

Algorithm 1. InfoMap algorithm.

Input: undirected Graph$G = (ver, edg)$.
Output: Detection of communities in the network G.
Require: Initial undirected graph is $G^0 = (ver^0, edg^0)$.
Initial community of undirected graph is M^0.
Threshold value is θ.
Max_Iter is the maximum iteration.

Step 1: $Iter \leftarrow 0$ // Initialize the iteration
Step 2: For all $usr \in ver^0$do
Step 3: $a_v = \frac{degree(v)}{|E|}$
Step 4: End For
Step 5: Repeat
Step 6: For all $v \in ver^k$do

Step 7: Evaluate $N_v^k = \{v\}$
Step 8: $a^{N_v^k} = \sum_{\alpha \in N_v^k} a_\alpha$
Step 9: $b^{N_v^k} = \sum w_{u,usr}(u, usr) \in edg^k, u \in C_u^k andusr \notin C_u^k$
Step 10: End For
Step 11: Compute $LB = LB(N)$ using Equation (7)
Step 12: Repeat
Step 13: for all $usr \in V^k$do
Step 15: IF $N'^k_v = argmin\left(\delta LB_{N_v^k \to N'^k_v}\right) < 0$then
Step 15: $N_v^k = N_v^k - \{usr\}$; $N'^k_v = N'^k_v \cup \{usr\}$
Step 16: $aN_v^k = \sum a_\alpha - a_u a^{N'^k_v} = \sum a_\alpha + a_u$
Step 17: Update $b^{N_v^k}$, and $b^{N'^k_v}$
Step 18: End if
Step 19: End For
Step 20: $Ver^{k+1} \leftarrow C^k$
Step 21: $E^{k+1} \leftarrow e(c_u^k, c_v^k)$
Step 22: $G^{k+1} = \left(Ver^{k+1}, Edg^{k+1}\right)$
Step 23: $iter \leftarrow iter + 1$
Step 24: until $iter \leq$ max $\_iter$ and $LB - LB_{new} < \theta$

In Algorithm 1, the undirected graph network is G^k with its iteration k and N_v^k is the community for a vertex v of the graph G^k. The output of this Algorithm 1 detects the community based on MDL using Equation (6) for the vertex in a graph.

3.2.2.2 Sigmoid fish swarm optimization algorithm (SFSO)

Even though the InfoMap algorithm is used to detect the community in the undirected graph, a sigmoid fish swarm optimization (SFSO) is implemented to improve accuracy and efficiency. It is an improved version of the fish swarm optimization algorithm. It consists of two main phases:

Phase 1: Initialisation;

Phase 2: Movement of fish.

In the initialization phase, it sets up the parameters and maximum iterations of the undirected graph. Phase 2 describes the object function of fish movement, which is used to search for and detect communities in the input graph. The SFSO algorithm uses the basic idea of social media activities in an optimized manner. In a water body (environment), fish can search for food, using group or individual movement. To improve the movement when searching for food in water bodies, fish use the sigmoid function. The sigmoid function includes movement of fish, searching for prey, follow movement, swarming, smooth turning and free movement. The purpose of using SFSO is to improve the food quality level.

Sigmoid function

It is a non-linear function used to map a vast area of information into a small space region between 0 and 1. In this paper, the sigmoid function is used to evaluate the movement of fish in turn position. The sigmoid function is evaluated by:

$$sigmoid(y) = \frac{C}{1+e^{-y}} \quad (14)$$

Here, e denotes the logarithm, C denotes the maximum value of curve movement of fish, y is between $-\infty$ and $+\infty$.

Density of food view

The density varies from 0 and 1. The value of 1 means high density and 0 means low density. Density represents the number of fish within a range. Density can be defined as follows:

$$density = \frac{Number\ of\ fish\ within\ range}{Total\ number\ of\ fish} \tag{15}$$

Prey movement of fish

When searching for food, fish continuously move in the water and finding food is called prey movement of fish. In searching for food, initially a fish analyses the range of food by using Equation (10) and it starts moving to hunt its prey based on food density by using Equation (11).

$$Fish_i = Fish_i + vis\_range \ \times rnd(-1,1) \tag{16}$$

$$Fish_i(t+1) = Fish_i(t) + \left[\frac{Fish_j - Fish_i(t)}{dist(i,j)}\right] \times step\_move \times sigmoid(0,1) \tag{17}$$

Here $Fish_i$ is the current position of a fish at the time t. $step\_move$ is the increment of next move of the fish. $dist$ refers to the distance between the current position and the next position and it is calculated by Euclidean distance. The new direction of the fish to move is calculated using the sigmoid function ranging between -1 and 1.

Free movement of fish

In the fish swarm optimization, when searching for prey, a fish can randomly move in any direction. If it reaches the boundary or gets closer to the boundary, it cannot find food there. It can turn in any direction to continue searching for food. This direction is evaluated using the sigmoid function. It is represented as:

$$Fish(t+1) = Fish(t) + step\_move \ \times \ sigmoid(-1,1) \tag{18}$$

Here, $Fish(t+1)$ is the time of the current position of the fish, $step\_move$ is the movement of the fish for calculating the direction using the sigmoid function; it ranges between -1 and 1.

Swarm movement of fish

One of the features of fish is swarming, aimed to move individually or in a group to reach its goal, that is, search for food. The movement of a group of fish resembles the movement of a swarm, guiding them to reach their target without diversity and achieve the goal quickly. This movement is called swarm movement of fish. Like a swarm, in the group movement of fish searching for food, the fish evaluates its central position first and stays there, trying to achieve the target by using:

$$Fish_{centre} = \frac{1}{M}\sum_{i=0}^{M} Fish_i \tag{19}$$

The swarm movement when searching for food is as follows:

$$Fish_i(t+1) = Fish_i(t) + \frac{Fish_{centre} - Fish_i(t)}{dist(i,centre)} \times step\_move \times sigmoid(0,1) \tag{20}$$

Here, $Fish_i$ shows the current position of the fish at the time t, $step\_move$ is the next movement of the fish, $Fish_n$ is the number of neighbouring fish in the water body, $Fish_{center}$ is the centre of the swarm, $dist$ refers to the distance between the current position and the next position and is calculated by Euclidean distance.

Follow movement of fish

When swarm movement occurs and one fish identifies food, it can change its direction. In that situation, some of the neighbouring fish follow it to get food. This movement is called the follow movement of fish. In the follow movement, the fish searches the range of best food available and compares it with the present state of food. This can be implemented by using:

$$Fish_i(t+1) = Fish_i(t) + \frac{Fish_n - Fish_i(t)}{dist(i,n)} \times step\_move \times sigmoid(0,1) \qquad (21)$$

Here, $Fish_i$ shows the current position of the fish at the time t, $step\_move$ is the next movement of the fish, $Fish_n$ is the number of neighbouring fish in the water body, $dist$ refers to the distance between the current position and the next position and is calculated by Euclidean distance. For calculating the new direction of the fish to move, the sigmoid function ranging between -1 and 1 is used.
The SFSO algorithm is given below:

Algorithm 2. Sigmoid fish swarm optimization algorithm (SFSO).

Input: Range, max iteration number, step move, neighbour fish (community formation)
Output: Optimization of a community in the network

Step 1: For $iter \longleftarrow 1\ to$ max $_iter$do
Step 2: For $FishNo \longleftarrow 1\ totot\_fish$ do
Step 3: $Present\_fish\_neighbor \leftarrow 0$
Step 4: $Present\_fish\_neighbor \leftarrow fish\_range$
Step 5: If $neighbour == 0$
Step 6: $next\_move \leftarrow sigmoid(free\_move\_fish)$ using Equation (16)
Step 7: Break and Goto Step 1
Step 8: Else
Step 9: If $density\_food > Group\_fishandfood\_availabiltiy$
Step 10: $Next\_move \longleftarrow Sigmoid\ (Prey\_Move\_fish)$ by using Equation (14)
Step 11: Else
Step12: $Next_Move \longleftarrow Rnd\ (Sigmoid\ (Swarm_Move_fishorFollow_Move_fish))$by using Equation (14) or Equation (15).
Step 13: End If
Step 14: End for
Step 15: End for
Step 16: Output⟶ Q_ modularity //Apply parametric measure

In this proposed work for phase 1 of I-SFSO, we use the InfoMap algorithm. The output of Algorithm 1 is passed on to phase 2 of SFSO. The output of the InfoMap algorithm generates communities with influence nodes. Phase 2 of SFSO merges two similar communities together and produces more accurate detection of the influence node in the community.

4 Results and Discussion

4.1 Datasets and used algorithms

In this section, we use ground-truth datasets to evaluate the performance of the proposed I-SFSO algorithm. Spyder Python 3.8 is used for the implementation of this algorithm. We implement our proposed work on the data sets of Facebook, Twitter and other data sets, including Stanford University

SNAP (2022), Dolphin network (2022), the American college football network (2012), a Karate club (2022) and YouTube (2018).

These network types are applied in the InfoMap algorithm (Zeng & Yu, 2018), the sigmoid fish swarm optimization algorithm (Ahmad et al., 2020), Louvain (Traag et al., 2019), the label propagation algorithm (LPA) (Malhotra & Chug, 2021), and our previous work Girvan Newman cuckoo search algorithm (GNCSA), which is available on request from the corresponding author. Table 1 shows information on data sets for the evaluation of metrics.

***Table 1**. Information on data sets for evaluation of metrics.*

Data set	Nodes	Edges	No. of communities
Facebook	3,959	168,486	7,498
Twitter	81,306	1,768,149	58,352
Karate Club	34	78	2
Dolphin Network	62	318	2
YouTube	1,134,889	2,987,623	8,385
American College Football Network	115	1226	15

4.2 Evaluation of performance

The community detection in social networks using the proposed I-SFSO algorithm uses ground-truth communities to evaluate metrics such as F1-score, accuracy, normalized mutual information (NMI) and modularity.

F1-score

In the input undirected graph network $G = (ver, edg)$, the set of ground-truth communities cm^* and the set of community detection are $\widehat{cm}$, each ground-truth community $cm_i \in cm^*$ and each detection of community $\widehat{cm}_i \in \widehat{cm}$. The F1-score of matching of community detection with each ground-truth community:

$$F1score = \frac{1}{2}\left(\frac{1}{|cm^*|}\sum_i max_j\, F1\left(cm_i, \widehat{cm}_j\right) + \frac{1}{|\widehat{cm}|}\sum_i max_j\, F1\left(cm_j, \widehat{cm}_i\right)\right) \quad (22)$$

Here, $F1\left(cm_i, \widehat{cm}_j\right)$ is the harmonic mean of precision and $cm_i, \widehat{cm}_j$ is recall.

Normalized mutual information (NMI)

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

$$NMI = \frac{-2\sum_{i=1}^{cm_x}\sum_{j=1}^{cm_y} N_{ij}\, log\left(\frac{N_{ij}\ N}{N_i N_j}\right)}{\sum_{i=1}^{cm_x} N_i \log\left(\frac{N_i}{N}\right) + \sum_{j=1}^{cm_y} \log\left(\frac{N_j}{N}\right)} \quad (23)$$

Here, cm_x is the number of original communities, cm_y is the number of communities identified, N is the number of nodes in the network, and N_{ij} is the number of nodes in the real community i which partitions x and the j^{th} found community which partitions y. N_i denotes the sum of the row matrix of N_{ij}. N_j denotes the sum of the column matrix of N_{ij}.

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:

$$Q\_M = \frac{1}{2n}\sum_{i,j\in Ver}\left(Adj_{i,j} - \frac{d(i)d(j)}{2n}\right) \times \delta\left(lbl_i, lbl_j\right) \quad (24)$$

Here, QM denotes the modularity, n is the number of edges in the network, and Adj is the adjacency matrix of the network. If the vertices ver_i and ver_j are connected directly then $Adj_{i,j} = 1$, else $Adj_{i,j} = 0$. Similarly, lbl_i, lbl_j are labels of the community of the vertices ver_i and ver_j. If $label_i = label_j$ then $\delta(lbl_i, lbl_j) = 1$, else $\delta(lbl_i, lbl_j) = 0$. Table 2 shows the F1-score evaluation criterion for the various algorithms with different data sets.

Table 2. *F1-score for different datasets.*

F1-score						
Data set	**InfoMap**	**SFSO**	**Louvain**	**LPA**	**GNCSA**	**I-SFSO (proposed)**
Dolphin	0.8451	0.9121	0.9012	0.9134	0.9552	0.9732
Facebook	0.8933	0.9211	0.9178	0.9342	0.9432	0.9652
Twitter	0.8718	0.9085	0.8982	0.9012	0.9376	0.9416
YouTube	0.8851	0.8956	0.8845	0.8934	0.9288	0.9478
Karate Club	0.8754	0.8954	0.8978	0.9086	0.9133	0.9256
Foot Ball	0.8987	0.9138	0.9215	0.9177	0.9487	0.9547

In Table 2, our proposed algorithm I-SFSO provides prominent results in the F1-score criterion compared to GNCSA of our previous work; it scored 0.9732 for the Dolphin data set, 0.9652 for Facebook, 0.9416 for Twitter, 0.9478 for the YouTube data set, 0.9256 for the karate club and 0.9547 for football. Similarly, the InfoMap algorithm yields 0.8451 for the Dolphin dataset, 0.8933 for Facebook, 0.8718 for Twitter, 0.8851 for the YouTube data set, 0.8754 for the karate club and 0.8987 for football.

Table 3. *NMI for different datasets.*

NMI						
Data set	**InfoMap**	**SFSO**	**Louvain**	**LPA**	**GNCSA**	**I-SFSO (proposed)**
Dolphin	0.9056	0.9362	0.9212	0.9398	0.9652	0.9767
Facebook	0.9145	0.9156	0.9055	0.9143	0.9531	0.9622
Twitter	0.9011	0.9177	0.9265	0.9056	0.9323	0.9435
YouTube	0.8956	0.9086	0.8986	0.9124	0.9321	0.9428
Karate Club	0.9045	0.9043	0.9149	0.8978	0.9033	0.9167
Foot Ball	0.9056	0.9077	0.9023	0.9124	0.9101	0.9234

In Table 3, our proposed algorithm provides prominent results in the NMI criterion compared to GNCSA of our previous work; it scored 0.9767 for the Dolphin data, 0.9622 for Facebook, 0.9435 for Twitter, 0.9428 for the YouTube data set, 0.9167 for the karate club and 0.9234 for the football data set. Similarly, the InfoMap algorithm yields 0.9056 for the Dolphin dataset, 0.9145 for Facebook, 0.9011 for Twitter, 0.8956 for the YouTube data set, 0.9045 for the karate club and 0.9056 for the football data set.
Figure 3 shows the computation time for detecting communities in the social networks for the various data sets (Dolphin, Facebook, Twitter, YouTube, karate club, football). It seems that our proposed I-SFSO algorithm needs less execution time for detection of communities in the social networks.

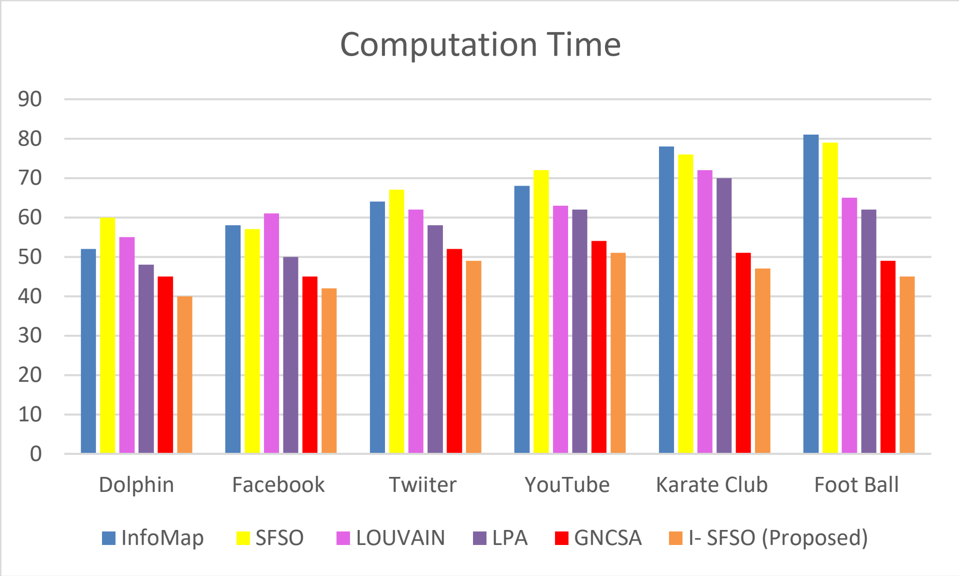

Figure 3. *Computation time.*

Figure 4 shows the accuracy rate of implementing various algorithms for different data sets.

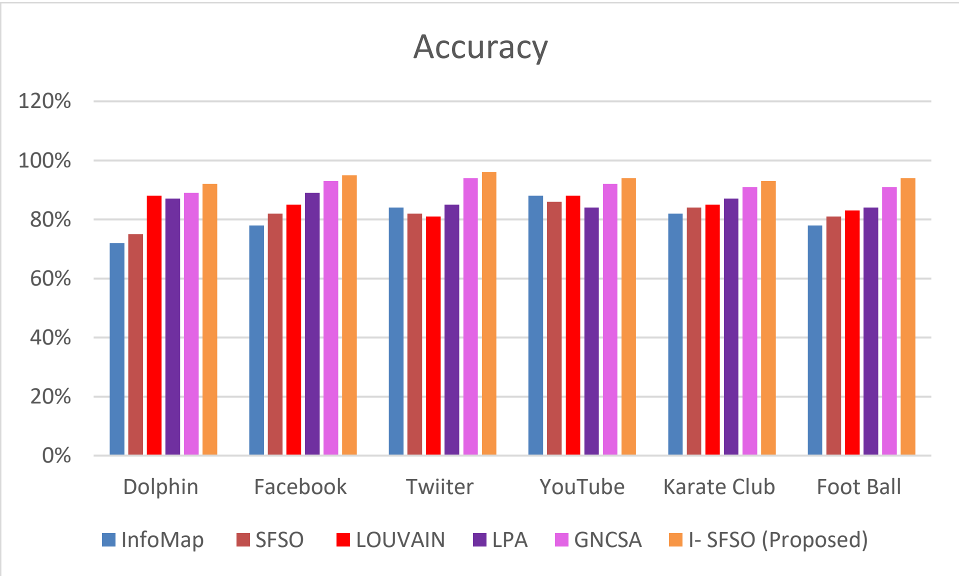

Figure 4*. Accuracy.*

Observing Figure 4, our proposed I-SFSO algorithm gives higher accuracy for the data sets of Dolphin (92%), Facebook (95%), Twitter (96%) YouTube (94%), the karate club (93%) and football (94%). Table 4 shows the detection of influence nodes or centrality in the various datasets.

Table 4. *Centrality measures.*

Centrality measure	Dolphin	Facebook	Twitter	YouTube	Karate club	Football
Degree centrality (DC)	0.7342	0.7844	0.8574	0.8798	0.8862	0.7448
Betweenness centrality (BC)	0.7225	0.8255	0.8282	0.8346	0.8834	0.8731
Closeness centrality (CC)	0.8566	0.8669	0.9631	0.9051	0.9274	0.7463
Katz centrality (BC)	0.8255	0.7225	0.8632	0.9173	0.8937	0.7826
Influence impact centrality	---	0.8467	0.9274	0.9342	---	---
Followers' centrality	--	0.7923	0.9045	0.8842	---	--

Table 4 shows the various centrality measures applied to our proposed I-SFSO algorithm. Figure 5 shows the Q-modularity of the various data sets.

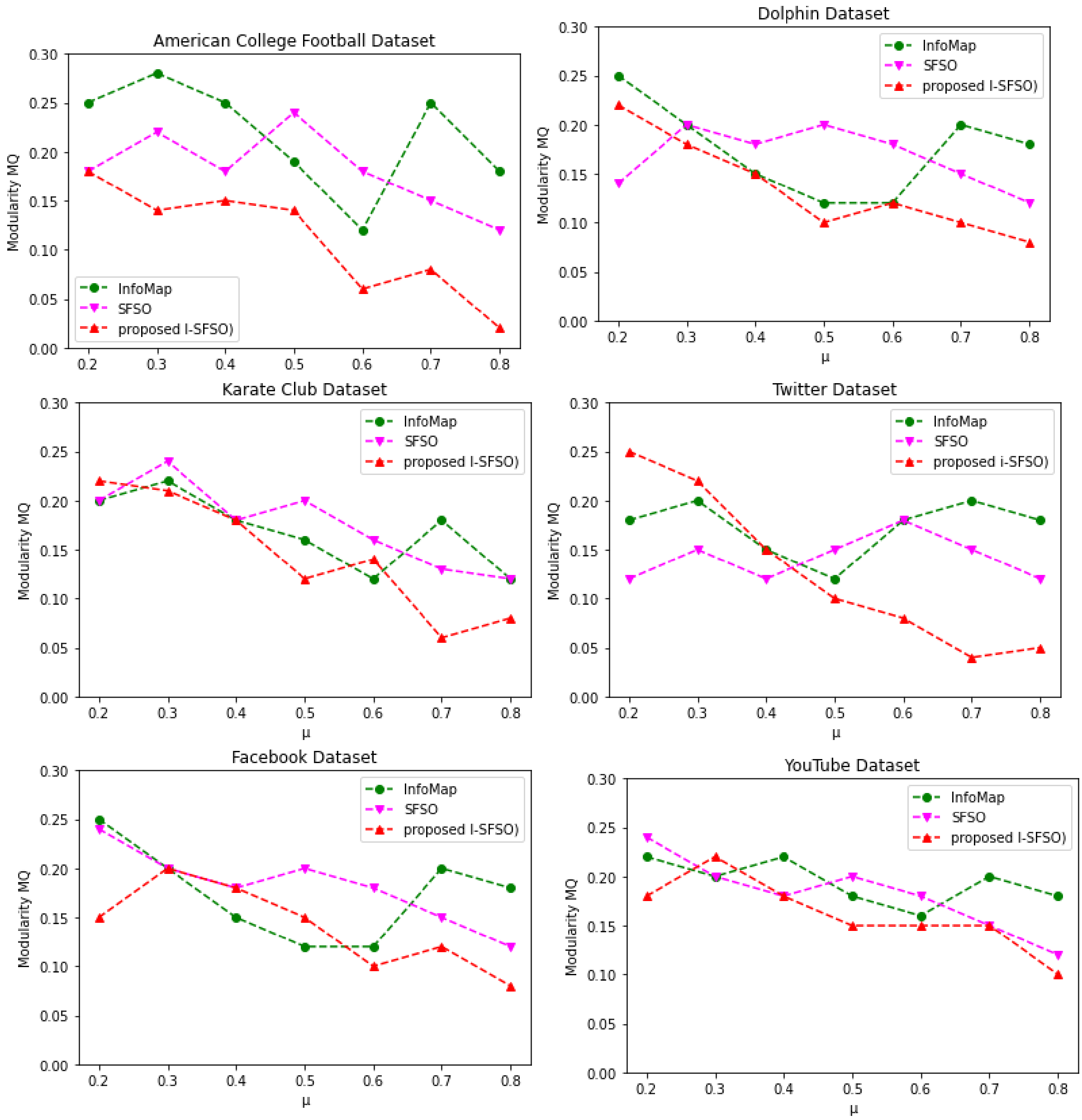

Figure 5. *Modularity in hybrid parameter μ.*

As can be seen from Figure 5, μ represents the hybrid parameter in the network for applying to overlapping modularity. In general, the range is between 0 and 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 I-SFSO algorithm provides the best results. It decreased significantly for the Dolphin data set ($\mu > 0.6$), the Facebook data set ($\mu > 0.6$), the Twitter data set ($\mu > 0.3$), the YouTube dataset ($\mu > 0.3$), the American college football data set ($\mu > 0.7$) and the karate club data set ($\mu > 0.7$). It seems that our proposed algorithm has good adaptability for the detection of communities in a complex network architecture. Figure 5 shows the sample community structure of the proposed algorithm for the various data sets. The use of the parameter μ as a hybrid parameter helps increase network complexity and improves community detection.

5 Conclusion

In this paper, we proposed an efficient community detection method using the InfoMap with the sigmoid fish swarm optimization algorithm (I-SFSO). For this work, data were collected from various data sets such as Dolphin, Twitter, Facebook, YouTube and American college football. The implementation part contains two phases: community detection by using the InfoMap algorithm, and the SFSO algorithm is used in order to get more accurate and optimized community detection and maximize the influence nodes. It merges similar communities together.

The performance analysis of this work is based on the aspects of accuracy, NMI, modularity and F1-score. Our proposed I-SFSO algorithm gives higher accuracy for the data sets of Dolphin (92%), Facebook (95%), Twitter (96%), YouTube (94%), the karate club (93%) and football (94%). The major advantage of influence node detection is that it helps in business development and trading. In future, this work will be extended to detection of communities in heterogenous complex network architectures.

Additional Information and Declarations

Conflict of Interests: The authors declare no conflict of interest.

Author Contributions: D.S.: Conceptualization, Methodology, Software, Data curation, Writing – Original draft preparation, Visualization, Investigation. R.M.: Supervision, Software, Validation, Writing – Reviewing and Editing.

Data Availability: The datasets used for evaluation are cited in Section 4.1.

References

Abduljabbar, D. A., Hashim, S. Z. M., & Sallehuddin, R. (2020). Nature-inspired optimization algorithms for community detection in complex networks: a review and future trends. *Telecommunication Systems*, 74(2), 225–252. https://doi.org/10.1007/s11235-019-00636-x

Adolfsson, A., Ackerman, M., & Brownstein, N. C. (2019). To cluster, or not to cluster: An analysis of clusterability methods. *Pattern Recognition*, 88, 13–26. https://doi.org/10.1016/j.patcog.2018.10.026

Ahmad, A., & Khan, S. S. (2019). Survey of State-of-the-Art Mixed Data Clustering Algorithms. IEEE Access, 7, 31883–31902. https://doi.org/10.1109/access.2019.2903568

Ahmad, Y., Ullah, M., Khan, R., Shafi, B., Khan, A., Zareei, M., Aldosary, A., & Mohamed, E. M. (2020). SiFSO: Fish Swarm Optimization-Based Technique for Efficient Community Detection in Complex Networks. *Complexity*, 2020, Article ID 6695032. https://doi.org/10.1155/2020/6695032

Al-Sahaf, H., Bi, Y., Chen, Q., Lensen, A., Mei, Y., Sun, Y., Tran, B., Xue, B., & Zhang, M. (2019). A survey on evolutionary machine learning. *Journal of the Royal Society of New Zealand*, 49(2), 205–228. https://doi.org/10.1080/03036758.2019.1609052

American college football network. (2012). American College Football Network Files. https://figshare.com/articles/dataset/American_College_Football_Network_Files/93179

Attea, B. A., Abbood, A. D., Hasan, A. A., Pizzuti, C., Al-Ani, M., Özdemir, S., & Al-Dabbagh, R. D. (2021). A review of heuristics and metaheuristics for community detection in complex networks: Current usage, emerging development and future directions. *Swarm and Evolutionary Computation*, 63, 100885. https://doi.org/10.1016/j.swevo.2021.100885

Babichev, S., Durnyak, B., Pikh, I., & Senkivskyy, V. (2019). An Evaluation of the Objective Clustering Inductive Technology Effectiveness Implemented Using Density-Based and Agglomerative Hierarchical Clustering Algorithms. In *Advances in Intelligent Systems and Computing*, (pp. 532–553). Springer https://doi.org/10.1007/978-3-030-26474-1_37

Blondel, V. D., Guillaume, J.-L., Lambiotte, R., & Lefebvre, E. (2008). Fast unfolding of communities in large networks. *Journal of Statistical Mechanics: Theory and Experiment*, 2008(10), P10008. https://doi.org/10.1088/1742-5468/2008/10/p10008

Chobe, S., & Zhan, J. (2019). Advancing community detection using Keyword Attribute Search. *Journal of Big Data*, 6(1). https://doi.org/10.1186/s40537-019-0243-y

Dolphin network. (2022). Dolphins (Social Networks). https://networkrepository.com/soc-dolphins.php

Doreian, P., Batagelj,V., & Ferligoj, A. (2020). *Advances in Network Clustering and Blockmodeling*. John Wiley & Sons.

Elbes, M., Alzubi, S., Kanan, T., Al-Fuqaha, A., & Hawashin, B. (2019). A survey on particle swarm optimization with emphasis on engineering and network applications. *Evolutionary Intelligence*, 12(2), 113–129. https://doi.org/10.1007/s12065-019-00210-z

Fortunato, S. (2010). Community detection in graphs. *Physics Reports*, 486(3-5), 75–174. https://doi.org/10.1016/j.physrep.2009.11.002

Ghalmane, Z., Cherifi, C., Cherifi, H., & Hassouni, M. E. (2019). Centrality in Complex Networks with Overlapping Community Structure. *Scientific Reports*, 9(1). https://doi.org/10.1038/s41598-019-46507-y

Karate club. (2022). Karate (Social Networks). https://networkrepository.com/soc-karate.php

Kou, F.-F., Du, J.-P., Yang, C.-X., Shi, Y.-S., Cui, W.-Q., Liang, M.-Y., & Geng, Y. (2018). Hashtag Recommendation Based on Multi-Features of Microblogs. *Journal of Computer Science and Technology*, 33(4), 711–726. https://doi.org/10.1007/s11390-018-1851-2

Kowald, D., Pujari, S. C., & Lex, E. (2017). Temporal Effects on Hashtag Reuse in Twitter. In *Proceedings of the 26th International Conference on World Wide Web* (pp. 1401–1410). ACM. https://doi.org/10.1145/3038912.3052605

Lee, S., Kim, J., Kang, H., Kang, D.-Y., & Park, J. (2021). Genetic Algorithm Based Deep Learning Neural Network Structure and Hyperparameter Optimization. *Applied Sciences*, 11(2), 744. https://doi.org/10.3390/app11020744

Liu, X., Du, Y., Jiang, M., & Zeng, X. (2020). Multiobjective Particle Swarm Optimization Based on Network Embedding for Complex Network Community Detection. *IEEE Transactions on Computational Social Systems*, 7(2), 437–449. https://doi.org/10.1109/tcss.2020.2964027

Li, J., Xu, H., He, X., Deng, J., & Sun, X. (2016). Tweet modeling with LSTM recurrent neural networks for hashtag recommendation. In *2016 International Joint Conference on Neural Networks* (IJCNN). IEEE. https://doi.org/10.1109/ijcnn.2016.7727385

Malhotra, D., & Chug, A. (2021). A modified label propagation algorithm for community detection in attributed networks. *International Journal of Information Management Data Insights*, 1(2), 100030. https://doi.org/10.1016/j.jjimei.2021.100030

Pan, Y., Yang, Y., & Li, W. (2021). A Deep Learning Trained by Genetic Algorithm to Improve the Efficiency of Path Planning for Data Collection With Multi-UAV. *IEEE Access*, 9, 7994–8005. https://doi.org/10.1109/access.2021.3049892

Raghavan, U. N., Albert, R., & Kumara, S. (2007). Near linear time algorithm to detect community structures in large-scale networks. *Physical Review E*, 76(3). https://doi.org/10.1103/physreve.76.036106

Sahan, A. M., Al-Itbi, A. S., & Hameed, J. S. (2021). COVID-19 detection based on deep learning and artificial bee colony. *Periodicals of Engineering and Natural Sciences*, 9(1), 29. https://doi.org/10.21533/pen.v9i1.1774

SNAP. (2022). Stanford Large Network Dataset Collection. https://snap.stanford.edu/data/index.html#onlinecoms

Traag, V. A., Waltman, L., & van Eck, N. J. (2019). From Louvain to Leiden: Guaranteeing well-connected communities. *Scientific Reports*, 9(1). https://doi.org/10.1038/s41598-019-41695-z

Tran-Ngoc, H., Khatir, S., Ho-Khac, H., De Roeck, G., Bui-Tien, T., & Abdel Wahab, M. (2021). Efficient Artificial neural networks based on a hybrid metaheuristic optimization algorithm for damage detection in laminated composite structures. *Composite Structures*, 262, 113339. https://doi.org/10.1016/j.compstruct.2020.113339

Wajahat, A., Nazir, A., Akhtar, F., Qureshi, S., ullah, F., Razaque, F., & Shakeel, A. (2020). Interactively Visualize and Analyze Social Network Gephi. In *2020 3rd International Conference on Computing, Mathematics and Engineering Technologies* (ICoMET). IEEE. https://doi.org/10.1109/icomet48670.2020.9073812

Wang, P., Huang, Y., Tang, F., Liu, H., & Lu, Y. (2021). Overlapping Community Detection Based on Node Importance and Adjacency Information. *Security and Communication Networks*, 2021, Article ID 8690662. https://doi.org/10.1155/2021/8690662

Yamada, Y., Masuyama, N., Amako, N., Nojima, Y., Loo, C. K., & Ishibuchi, H. (2020). Divisive Hierarchical Clustering Based on Adaptive Resonance Theory. In *2020 International Symposium on Community-Centric Systems* (CcS). IEEE. https://doi.org/10.1109/ccs49175.2020.9231474

YouTube. (2018). Trending YouTube Video Statistics. https://www.kaggle.com/datasets/datasnaek/youtube-new

Zeng, J., & Yu, H. (2018). A Distributed Infomap Algorithm for Scalable and High-Quality Community Detection. In *Proceedings of the 47th International Conference on Parallel Processing.* ACM. https://doi.org/10.1145/3225058.3225137

Editorial record: The article has been peer-reviewed. First submission received on 7 November 2022. Revision received on 29 November 2022. Accepted for publication on 7 December 2022. The editors coordinating the peer-review of this manuscript were Venkatachalam Kandasamy, Mohamed Abouhawwash, Nebojsa Bacanin. The editor in charge of approving this manuscript for publication was Zdenek Smutny.

Special Issue: Sustainable Solutions for Internet of Things Using Artificial Intelligence and Blockchain in Future Networks.

Acta Informatica Pragensia is published by Prague University of Economics and Business, Czech Republic.

ISSN: 1805-4951