DOI 10.15622/ia.23.5.4

A.H. IBRAHIM, M.A. BOUDREF, L. BADIS

CLUSTERING OF NETWORKS USING THE FISH SCHOOL SEARCH ALGORITHM

Ibrahim A.H., Boudref M.A., Badis L. **Clustering of Networks Using the Fish School Search Algorithm.**

Abstract. A network is an aggregation of nodes joined by edges, representing entities and their relationships. In social network clustering, nodes are organized into clusters according to their connectivity patterns, with the goal of community detection. The detection of community structures in networks is essential. However, existing techniques for community detection have not yet utilized the potential of the Fish School Search (FSS) algorithm and modularity principles. We have proposed a novel method, clustering with the Fish School Search algorithm and modularity function (FSC), that enhances modularity in network clustering by iteratively partitioning the network and optimizing the modularity function using the Fish School Search Algorithm. This approach facilitates the discovery of highly modular community structures, improving the resolution and effectiveness of network clustering. We tested FSC on well-known and unknown network structures. Also, we tested it on a network generated using the LFR model to test its performance on networks with different community structures. Our methodology demonstrates strong performance in identifying community structures, indicating its effectiveness in capturing cohesive communities and accurately identifying actual community structures.

Keywords: clustering, fish school search algorithm, modularity function, network structures.

1. Introduction. In network theory, a graph or network refers to a group of nodes linked through edges or lines, representing different entities and their relationships or interactions [1]. For example, in metabolic networks, nodes could represent biochemical compounds or enzymes, and edges could represent metabolic reactions or pathways [2]. In the case of the Internet, nodes could represent web pages or computers, and edges could represent hyperlinks or network connections [3]. The representation of a network is a graph G (V, E), where V refers to the set of nodes or vertices and E refers to the set of edges or links [4]. The number of vertices and edges determines the size of a graph. In the case of a social network, the size of the graph depends on the number of people in the network and their relationships or interactions [5]. In social network clustering, nodes are grouped into clusters based on their connectivity patterns, which are determined by the links or edges that connect them. This clustering process relies on community recognition, where a community refers to a group of nodes that are densely connected with only a few connections to nodes outside the community [6]. The objective of community detection is to identify clusters or communities of nodes within a network, where the nodes within a community exhibit denser connections among themselves compared to nodes in other communities [7].

Clustering a social network can help us identify groups of individuals that are more closely connected, revealing social substructures such as cliques, subgroups, or communities. Understanding these substructures can provide insights into the dynamics of social interactions and help us identify influential individuals or groups within the network [8]. For example, recognizing communities and clusters can help identify groups of proteins that work together to perform specific functions in the cell [9]. In the case of a transportation network, clustering can help identify clusters of nodes that are more interconnected in traffic planning and routing [10]. Several methods can structure networks [11], including modularity-based methods [12], spectral clustering [13], hierarchical clustering [14], and modularity optimization [15, 16].

Maximizing modularity is a challenging task due to its computational intractability, as finding an optimal solution has been proven to be NP-complete [17]. Heuristic algorithms are often used, but they may not always produce the best partitions. Modularity maximization is a promising approach for community detection, aiming to maximize within-community interconnectedness while minimizing between-community interconnectedness.

Our main contribution is the enhancement of modularity in network clustering through the innovative application of the Fish School Search Algorithm [18, 19]. This algorithm, inspired by the collective behavior and intelligence of fish schools, serves as a powerful metaheuristic for navigating the complex landscape of community detection. By iteratively partitioning the network and judiciously pruning edges, our approach focuses on maximizing the modularity function [20], which quantifies the strength of the division of a network into modules or communities.

The Fish School Search Algorithm's ability to simulate the dynamic and collaborative [21] search strategies of fish enables the discovery of highly modular community structures, thereby improving the resolution and effectiveness of network clustering. This method not only aligns with the principles of modularity optimization but also introduces a novel perspective to the field, leveraging natural processes to address computational challenges inherent in the task of community detection.

In the clustering process, the modularity function serves as the objective function, assessing the quality of a clustering solution [22]. The proposed method works by initializing a population of fish, each representing a potential clustering solution. Until they reach an optimal solution, the fish iteratively evaluate and update their solution based on the modularity function.

The method was tested on well-known and unknown network structures and compared to other commonly used clustering methods [23]. The findings demonstrate the method's superior performance in both modularity and computational efficiency. The paper is organized in a logical and easy-to-follow manner. The second section, titled Background and Related Works, is further divided into four subsections. The first subsection provides background on community detection and its importance. The second subsection discusses modularity and modularity maximization. The third subsection introduces the Fish School Search (FSS) algorithm and its use in the proposed network clustering method, while the fourth subsection reviews related works in the field. The third section presents the proposed approach, while the fourth section of the paper presents a performance assessment of our method. The fifth section discusses the results and their implications, and the sixth section concludes the study and suggests future work.

2. Background and related works

2.1. Community detection. Community detection in complex networks is the process of identifying clusters of nodes that share common properties, depicted as nodes and edges in a graph. These algorithms group nodes based on denser connections within groups compared to connections between groups. Communities can represent groups of individuals with shared interests and interactions within human society, and they can be identified in networks based on connection patterns [24].

Hierarchical clustering is a popular approach to community detection that builds a hierarchy of partitions either by merging smaller communities into bigger ones or by dividing larger communities into smaller ones based on similarity measures. Initially, each node is considered a separate community, and similarity measures between pairs of communities are calculated using factors such as shared neighbors or connection strength. The algorithm then merges the most similar communities iteratively, creating a hierarchical structure [25].

Hierarchical clustering allows flexible exploration of community structure at different granularity levels, ranging from broad to specific communities. This approach provides a comprehensive view of the network's community structure, enabling the analysis of nested communities and different levels of detail. The FSC proposed was developed to solve community detection problems by maximizing modularity, and is a widely recognized and commonly used approach for discovery communities in structure networks. It involves finding the split of the network into clusters that maximize the modularity value, which is a measure of the quality of the division.

2.2. Modularity. Modularity, a concept pioneered by [20], is a metric for assessing the efficacy of partitioning a graph into distinct communities. While its original formulation was designed for undirected graphs, subsequent research has broadened its applicability to encompass directed and weighted graphs [26 – 28]. The notion of modularity within a partition is expressed through a singular numerical measure (with a range reaching 1) used to evaluate the level of interconnectedness among components residing within communities, as opposed to the connections linking different communities together. A heightened positive modularity value indicates an enhanced organization of communities.

Modularity maximization is a widely used approach in network analysis for resolving community detection problems. Modularity quantifies the quality of the community structure in a network by comparing the number of edges each community has to the number expected by chance. The goal of modularity maximization is to find a partition of nodes into communities that maximize the modularity score. This involves iteratively assigning nodes to communities and evaluating the change in modularity until no further enhancement can be performed.

Modularity maximization has several advantages for resolving community detection problems. It is a flexible and scalable method that can be used in networks of various sizes and types. It can find the optimal division of the network into communities without prior knowledge of their number or size, allowing for an unbiased exploration of the community structure.

Maximizing modularity to find the best partition of a network into communities is not feasible for large and complex networks due to the NP-complete [29] nature of the problem. Several heuristic algorithms have been proposed to approximate the optimal solution. Despite its drawbacks, such as the resolution limit and high computational requirements, modularity maximization remains a popular and effective approach for community detection. The FSC performance will be evaluated by comparing the communities obtained using different algorithms with the ground truth and assessing the quality and reliability of community detection results through modularity scores.

2.3. Artificial Fish School Search Algorithm. The study of collective behavior in decentralized, self-organized systems, encompassing both natural and artificial systems, falls within the realm of swarm intelligence [30]. The inspiration behind swarm intelligence stems from observing the collective behavior exhibited in natural societies, including the coordinated movements of birds, fish, ants, bees, termites, and other species. This behavior emerges from the interactions of numerous individuals acting in unison. It is a form of distributed problem-solving in which a group of individuals cooperates to achieve a common goal [31]. Algorithm 1, known as Artificial Fish School Search (FSS) [32], is a prime example of swarm intelligence, drawing inspiration from the collective behavior observed in fish schools. The foundational algorithmic structure of the FSS optimization algorithm, as outlined in [32], is as follows:

Algorithm 1. FSS optimization algorithm for community detection

```
Input:
    Graph (G = (V, E)): The input graph with vertices (V) and edges (E).
    Maximum Cycle Number: The maximum number of iterations for the algorithm.
    Maximum CPU Time: The maximum allowed CPU time for the optimization process.
Output:
    π: The resulting cluster assignment that maximizes modularity Q
Do
    Attraction fish phase:
        Move each fish towards the center of mass in the school.
    Aggregation fish phase:
        Increase the step size of each fish
    Movement fish phase:
        Move each fish based on its current position and step size.
    Evaluate the modularity Q of each fish's cluster assignment
    Save the best solution (cluster assignment π) found so far that maximizes Q.
While Cycle < Maximum Cycle Number and CPU Time < Maximum CPU Time
Return the best cluster assignment π that maximizes modularity Q.
```

The FSS algorithm is based on the idea that fish in a school move together, following some simple rules of behavior such as attraction, repulsion, and alignment. Within the domain of swarm intelligence, the Artificial Fish School Search (FSS) algorithm stands out as a notable exemplar. Taking cues from the coordinated behavior displayed by fish in their schools, the FSS algorithm derives its principles from this collective phenomenon. Like their aquatic counterparts, the algorithm envisions a scenario where virtual fish move in unison, guided by simple yet impactful behavioral rules encompassing attraction, repulsion, and alignment [33]. By harnessing these fundamental principles, the FSS algorithm aims to address complex problems through the power of collective intelligence. The process of fish locating their food typically unfolds in three stages: attraction, aggregation, and coordinated movement [34]. Initially, in the attraction phase, fish employ a myriad of sensory cues, ranging from chemical to visual and auditory signals, to sense the availability of food. This phase often sees fish swimming in the direction of a positive chemical gradient, which is a reliable indicator of food's existence in the water. A plethora of fish species possess specialized sensory apparatus, like olfactory organs, that are exceptionally responsive to waterborne chemical cues. This sensitivity enables fish to discern even minor chemical traces released by potential food sources and subsequently follow this scent trail to discover food [35]. In the aggregation phase, fish may form schools or shoals to more efficiently locate and access the food source. In these groups, individuals may benefit from the presence of others in terms of increased feeding opportunities and protection from predators. In the last stage, referred to as the coordinated movement phase, fish within the collective meticulously adjust their movements to ensure the maintenance of group unity and synchronously approach the food source [36]. This often entails each fish attentively responding to their neighbors' actions, facilitating a seamless and efficient group movement. The creation of collective wisdom in fish schools and other kinds of animal swarms heavily hinges on effective information exchange. Fish utilize diverse sensory signals to communicate and disseminate information among each other. This information exchange often involves the use of auditory, visual, and tactile cues [37]. Based on the concept of information exchange, the Artificial Fish School Search (FSS) algorithm can be divided into four stages to find maximum modularity. These stages include an initialization phase, where the FSS algorithm starts by randomly searching for a food source, which is considered a potential solution. Following the initialization phase, the process transitions into the individual swimming phase, which is analogous to conducting a local search in the current location of each fish or solution. Each instance of individual swimming generates a novel candidate solution, guided randomly and exhibiting distinct values. Upon the conclusion of the individual phase, an evaluation or update of the fitness function is performed. If there's no enhancement in the fish's position, it is assumed that this specific fish or solution remains static. Only those fish or solutions showing improvements in their fitness functions will transition to a new position. During the instinctive-collective phase, the overall positioning of the school of fish or set of solutions is adjusted, taking into consideration the alterations in the fitness function of each fish or solution from the prior iteration [38]. The process culminates with the collective-volitional phase, where the fish or solutions are moved if there has been an improvement. In algorithmic terms, this signifies a refinement of the solution for the optimization problem. If there hasn't been an enhancement in the position of the entire school or set of solutions compared to its previous position, it indicates the need for another round of food search or, in algorithmic terms, a return to the initialization phase.

2.4. Related Works. Community detection is a rapidly evolving field, with researchers continuously proposing new and innovative approaches to identify and analyze community structures in complex networks. These approaches range from leveraging the concept of modularity to using artificial intelligence algorithms for clustering. In this review, we will discuss some notable works in the literature that focus on harnessing the power of modularity and artificial intelligence algorithms for community detection.

The authors in [39] proposed a method for module partitioning in complex products using stable overlapping community detection and component allocation.

It effectively handles intricate component correlations, demonstrating superiority over existing methods on a CNC grinding machine. The study model (CDFSE) [40] presents a dynamic model simulating fish school behavior to showcase the formation of larger, stable groups based on shared attributes. Additionally, the study [41] introduces a new network clustering method using the Bee Colony Algorithm and Modularity Function. This method involves iterative edge removal for network partitioning and uses the Bee Colony Algorithm to optimize the modularity function, revealing optimized clusters. The study [42] highlights encord's role in enhancing the efficiency and accuracy of medical data annotation in cancer subtype classification for identifying patient clusters, advancing targeted therapies, and biomarker discovery. Automated and collaborative features in encord improve annotation speed and consistency, benefiting AI-driven diagnostics and treatment planning. Lastly, the study [43] introduces an algorithm for disjoint communities in complex networks. It’s designed for undirected, unweighted graphs and uses cosine similarity for weight determination. It starts with individual nodes as communities and merges them based on modularity values.

Our proposed approach, FSC-Fish School Clustering, addresses gaps in previous work by offering a comprehensive solution for community detection in unipartite, unweighted, and undirected networks. While existing methods often have limited scope, our approach widens their applicability by specifically targeting these types of networks. Moreover, we introduce a dynamic modeling aspect inspired by fish school behavior, enabling the formation of larger, stable groups based on shared attributes. To overcome optimization challenges, we integrate the Fish School Search Algorithm with the Modularity function, enhancing accuracy and efficiency in community detection. Additionally, scalability concerns are addressed by leveraging the efficiency of the Fish School Search Algorithm, making our approach suitable for handling large-scale networks. By bridging these gaps, our proposal aims to advance the field of network analysis and provide more accurate and efficient tools for understanding complex network structures.

The FSC provides a unique approach to optimization problems, drawing inspiration from natural phenomena. Its strengths lie in its adaptability and collective problem-solving power. It mimics fish school behavior to navigate complex search spaces effectively. However, it may lag behind modern network assembly methods in computational efficiency and scalability. While FSC is ideal for intricate problems requiring exploration, modern methods suit tasks demanding precision and large-scale network management. The choice between the two depends on the specific needs of the optimization challenge, balancing the strengths of FSC against the requirements for efficiency and control in network assembly.

3. The proposed approach (FSC-Fish School Clustering). Hierarchical methods are clustering algorithms that can be either divisive or agglomerative. Agglomerative methods recursively merge clusters into larger clusters, while divisive methods recursively split clusters into smaller clusters. Divisive methods in community detection remove edges based on various criteria. For example, Edge Density [44] involves removing edges based on the density of connections to identify densely connected subgraphs as communities. Topological [45] measures utilize measures like centrality or clustering coefficient to determine edge importance; high centrality edges connecting communities are preserved, while low centrality edges are removed. Additionally, structural properties [46] involve removing edges that bridge distinct clusters to promote clearer separation between communities and identify internally connected communities. Some graph-splitting methods create separate partitions by deleting edges that connect vertices with very high or very low weights [47]. In this study, we propose a novel network clustering approach, referred to as FSC, which uses the Fish School Search Algorithm and the modularity function introduced by [20].

In this study, we propose a novel network clustering method that focuses on identifying community structures in unipartite, unweighted, and undirected networks. We utilize the Fish School Search Algorithm to iteratively split the network and remove edges while maximizing the modularity function. During this splitting phase, the algorithm discovers a set of clusters that represent the community structure. This process is repeated iteratively until each cluster consists of a single vertex. FSC leverages the modularity function as an objective metric to measure the power of the community structure. It guides the determination of the optimal number of network communities (clusters). By maximizing modularity, FSC aims to identify the community structure that best fits the given network. This is done through the use of the Fish School Search Algorithm and the Modularity function. For a graph $G(V, E)$, our method identifies the most optimal community structure $\pi = \{c_1, \dots, c_{nc}\}$, where:

– $\bigcup_{i=1}^{nc} c_i = V$,
– $c_i \neq \emptyset$,
– $c_i \cap c_j = \emptyset \; if \; i \neq$ j,
– nc: number of clusters.

Modularity assesses the cohesion within different segments of a network by assigning a numerical value to each segment, known as a community structure. Modularity Q ranges between 0 and 1, where higher scores typically indicate better-defined partitions, while lower scores may imply less cohesive groupings. However, it's important to note that the quality of community structures can vary depending on the specific context. The modularity function is formulated around the comparison between the observed fraction, denoted as $e(c_i)$of edges within communities, and the expected fraction, denoted as $a(c_i)$, of edges within the same communities. The modularity, represented as Q, is calculated using the expression:

$$Q = \sum_{c_i} e(c_i) - a(c_i)^2. \tag{1}$$

Consider an undirected, unweighted graph G, with n vertices and m edges, and a partition denoted π = $(c_1, \dots, c_{nc})$.

In this context, the modularity [48], can be expressed as follows:

$$Q = \frac{1}{(2m)} \sum_{vw} \left[A[v, w] - \frac{K_v K_w}{(2m)} \right] \delta(v, w), \tag{2}$$

where

$$\delta(v, w) = \begin{cases} 1 & if\ v\ and\ w\ belong\ to\ the\ same\ community \\ 0 & otherwise \end{cases},$$

where:

– v, w – are vertices of G,
– $K_{v,} K_w$ – the degree of v and w which is the number of edges linked to each vertex,
– m – total number of edges in the graph G,
– A – the adjacency matrix of the graph G.

The overarching framework of our algorithm designed to uncover community structures within networks is outlined as follows:

Algorithm 2. FSC Algorithm

```
Input: G = (V, E) – The input graph with vertices V and edges E.
Result cluster: π, maximizes Q, the resulting cluster assignment cluster.
Cluster ← G
π ← Cluster
repeat
 Set_Cluster ← π
    for i ∈ | Set_Cluster | do
        Cluster = Set_Cluster[i]
        π' ← Fish School Search (Cluster)
        Divide Cluster based on π'
        Update π'
    end for
   until | π | ≥ |V|
 return cluster that maximizes Q
```

The FSC algorithm, utilizing the Artificial Fish School Search (FSS) algorithm, iteratively partitions a network to maximize modularity. Starting with the entire graph as a single cluster, the algorithm refines cluster assignments through FSS, aiming to enhance modularity within each cluster. This iterative process continues until each cluster comprises a single node, culminating in the selection of the assignment with the highest modularity. The resulting clusters represent distinct communities, offering flexibility to tailor the number of communities. Additionally, we construct a tree diagram to visually depict the hierarchical organization of communities, elucidating the structural composition of the network.

4. Performance Assessment. We tested FSC on known and unknown network topologies. These networks are well-studied and have been used to benchmark other community detection algorithms. This means that we can compare our results to those of other algorithms and see how FSC performs. Networks tested include Karate Club, College Football, US Politics Books, and Dolphins. We additionally created a network using the Lancichinetti et al. random graph model, specifically designed to generate networks with desired community structures. This enables us to assess the performance of FSC on diverse network structures with varying communities. We conducted a comparative analysis of our approach against several established methods, such as:

– **The Infomap method** [49] identifies community structure by minimizing the description length of a random walker's navigation, using iterative node agglomeration to achieve improved partitions.

– **The label propagation method** [50] efficiently detects communities in complex networks by assigning unique labels to nodes and adopting the most prevalent label among neighbors, but its accuracy may suffer due to its random nature and potential tie situations.

– **The Louvain method** [51] optimizes modularity by iteratively reassigning nodes between communities to enhance network division strength.

– **The Fast Greedy method** [52] utilizes community analysis by greedily enhancing modularity scores through the continuous merging of community pairs until maximal improvements are achieved. The Fish School Search (FSS) algorithm chosen for network clustering stands out due to its unique nature-inspired approach that simulates the behavior of a school of fish, enabling it to explore the search space more effectively and efficiently. Unlike other algorithms such as Infomap, Label Propagation, Louvain, and Fast Greedy, FSS does not require a predefined number of clusters, making it highly adaptable to various datasets. It further optimizes the modularity function, a unique feature that quantifies the strength of the division of a network into modules or communities, enhancing the resolution and effectiveness of network clustering. Demonstrating robust performance, FSS excels in identifying community structures in both well-known and unknown network structures, making it a compelling, versatile, and reliable choice for network clustering. This dynamic modeling aspect, inspired by fish school behavior, forms larger, stable groups based on shared attributes, addressing gaps in previous work. It also integrates the Fish School Search Algorithm with the Modularity function, enhancing accuracy and efficiency in community detection and making it suitable for handling large-scale networks. In this part, we assess the performance and effectiveness of our method. This required two measures to compare and evaluate it against existing methods: normalized mutual information and computer-generated networks.

4.1. Normalized Mutual Information (NMI). We can leverage the NMI scale to compare the outcomes of our approach with those of various community detection methods. It is a widely used measure of community quality and is a good predictor of human-annotated ground-truth communities. In [53] proposed the utilization of the Normalized Mutual Information (NMI) metric for the comparative assessment of community-based screening techniques. NMI requires a confusion matrix M, where the rows in M indicate the real communities and the columns in M indicate the detected communities. Each element M , M_{ij} in the matrix represents the count of nodes belonging to the actual community i that are present in the discovered community j. This confusion matrix allows us to quantify the overlap and agreement between the two sets of communities. The NMI measure employs information theory principles to evaluate the similarity between the partitions. It calculates the mutual information between the real and found communities, taking into account the distribution of nodes across the communities. The NMI measure captures the shared information and dependence between the partitions by considering the probabilities of nodes belonging to specific communities. The formula derived from data analysis principles [53] for measuring the likeness between two community structures A and B is as follows:

$$I(A,B) = \frac{-2\sum_{i=1}^{c_A}\sum_{j=1}^{c_B} M_{ij}\log\left(M_{ij}M/M_{i.}M_{.j}\right)}{\sum_{i=1}^{c_A} M_i \log(M_i/M) + \sum_{j=1}^{c_B} M_{.j}\log\left(M_{.j}/M\right)}. \quad (3)$$

Here, c_A reflects the number of actual communities, while c_B represents the number of discovered communities. The total of the elements in the row i vector of the matrix M_{ij} is denoted by M_i, and the summation over column j is denoted by $M_{.j}$. When our method's community structure perfectly aligns with the actual community structure, $I(A,B)$ achieves its maximum value of 1. This means that the two sets of clusters are a perfect fit, and the NMI value will be 1, indicating a complete match between the clusters. Conversely, if the detected communities have no resemblance to the actual communities in the network, the NMI value will be 0, signifying a complete lack of similarity. This means that the two sets of clusters have no connection or resemblance. When the discovered community structure is similar but not identical to the actual community structure, it means that the clusters share some commonalities while also having some differences; the NMI value will fall between 0 and 1, capturing the partial agreement between the two sets of clusters.

4.2. Computer-Generated Networks. To validate the effectiveness of our method, we can employ computer-generated grids as a means of testing. These computer-generated networks possess a predetermined community structure, which renders them highly suitable for evaluating the accuracy and robustness of community detection algorithms. By leveraging these networks, we can thoroughly assess the performance and reliability of FSC in accurately identifying and uncovering community structures. To create networks with a community structure, we utilized the LFR model proposed by [54]. The LFR model is a widely used framework for generating networks with diverse properties, including power-law degree distributions, community structure, and even overlapping communities. In our study, we employed the LFR model to generate a network consisting of 128 nodes. The degree exponent distribution was set to 2, determining the number of links per node. The community size distribution exponent was 3, governing the sizes of the communities. The average degree was 16, showing the ratio of the actual connections to the potential connections in a network. The network consisted of 3 communities, and the mixing parameter varied from 0.1 to 0.9, influencing the interconnectivity between communities. In our method, the fitness function is defined as the modularity function. We set 500 (the number of fish) as food sources and specified the maximum number of iterations as 1000 for the optimization FSC.

5. Results and Discussion. In this section, we present an elaborate examination of the authentic social networks utilized to assess the effectiveness of our method. We have tested our method on the following networks: The Zachary Club Network [55], American College Football [56], The Dolphin Social Network [57], The Book about US Politics Network [58] Facebook [59], Amazon [60], Les Miserable Network [61], The Jazz Collaboration Network [62], The HIV Network [63], and The Contiguous USA [64] Network. Table 1 summarizes the fundamental characteristics of real benchmark networks, including the number of nodes (|V|), the number of edges (|E|), the average degree (〈k〉), and whether the community structure (CS) is known or unknown.

Table 1. Summarizes characteristics of Benchmark Social Networks

Networks	CS	|V|	|E|	〈k〉	Description
Zachary	Known	34	78	4.58	Derived from a university karate club.
Football	Known	115	613	10.66	Fall 2000 Division IA college games.
Dolphin	Known	62	159	5.12	62 New Zealand dolphins' interactions.
Book	Known	105	441	8.4	2004 election US politics book list.
Facebook	Unknown	4039	88234	18.02	Facebook social network dataset.
Amazon	Unknown	334863	925872	5.52	Amazon product co-purchasing network.
Miserable	Unknown	77	245	6.36	Victor Hugo's character network.
Jazz	Unknown	198	2742	27.69	Jazz collaborations from 1912-1940.
HIV	Unknown	40	41	2.05	Early HIV spread in USA contacts.
USA	Unknown	48	107	4.45	US shared border network with DC.

Table 2 presents a comparative analysis of the modularity metric for our proposed algorithm against four established algorithms. The modularity metric is calculated using the formula (2) (Section 3).

In Table 3, we present two metrics:

1. The number of clusters (|C|) is the number of clusters obtained after the iterative splitting process.

2. The normalization mutual information (NMI) is a measure of similarity between the clusters obtained using formula (3) (See section 4.1).

Table 2. Modularity for real networks with community structures

Methods	Q					
	Karate	Football	Books	Dolphins	Facebook	Amazon
Infomap	0.37	0.**60**	**0**.52	0.52	0.05	0.82
Label Propagation	0.37	0.57	0.47	0.51	0.65	0.78
Louvain	0.41	0.60	0.52	0.52	0.68	0.92
Fast Greedy	0.38	0.54	0.50	0.49	0.64	0.87
FSC	0.40	0.64	0.54	0.52	0.72	0.94

Table 3. NMI for real networks with community structures

Methods	Karate		Football		Books		Dolphins		Facebook		Amazon	
	|C|	NMI	|C|	NMI	|C|	NMI	|C|	NMI	|C|	NMI	|C|	NMI
Infomap	2	0.59	**2**	0.92	6	0.49	5	0.53	239	0.09	17296	0.1
Label Propagation	2	0.1	10	0.83	3	0.48	4	0.47	15	0.18	22496	0.01
Louvain	4	0.50	9	0.85	5	0.50	4	0.49	11	0.18	240	0.02
Fast Greedy	3	0.69	5	0.65	3	0.53	3	0.41	25	0.1	1532	0.1
FSC	2	0.87	7	0.1	3	0.60	4	0.81	150	0.74	260	0.1

To evaluate the performance of our proposed algorithm, we conducted experiments in a controlled environment using the Python programming language. Our algorithm's code was developed from scratch, and after multiple executions, we recorded the best modularity value achieved. In contrast, the baseline algorithms were sourced from the igraph library, which is well-known for its comprehensive collection of network analysis tools.

According to Table 2 and Table 3, FSC performed well, achieving high NMI and Q values. For example, on the Karate network, it detected 2 clustering with an NMI of 0.87 and a Q value of 0.40. On the football network, it detected 7 clusters with an NMI of 0.1 and a Q value of 0.64. FSC achieved the highest or second-highest NMI and Q scores on all four networks, suggesting it is more effective than other methods considered.

FSC identified meaningful communities in the karate network (Figure 1) and detected three clusters in the network of books about US politics (Figure 2), with high modularity and NMI values indicating good performance. This provides insights into social dynamics.

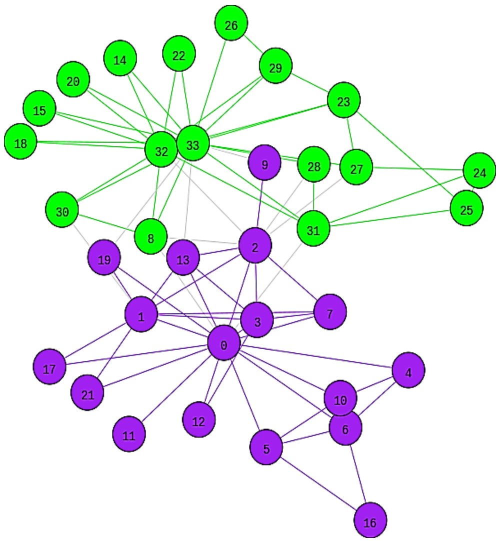


Fig. 1. FSC identifies 2 clusters in the Zachary club network structure

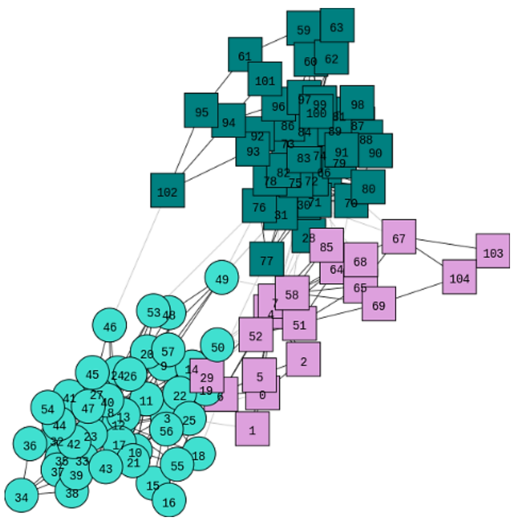
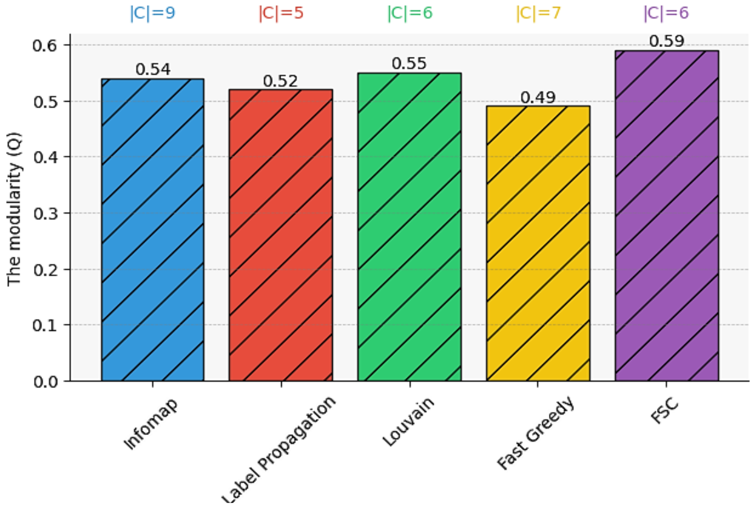
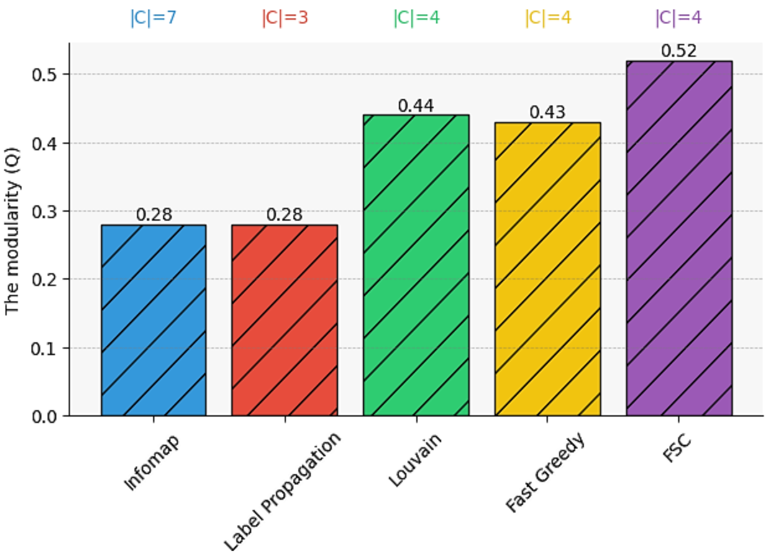


Fig. 2. FSC identifies 3 clusters in the structure of the Books about US politics

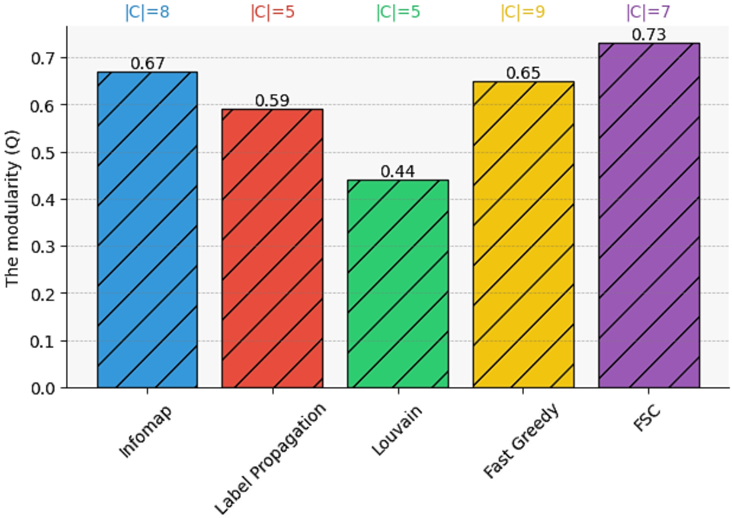
We tested FSC on various unknown networks (Les Miserable, Jazz, HIV, USA, Facebook, and Amazon), as shown in Figure 3.



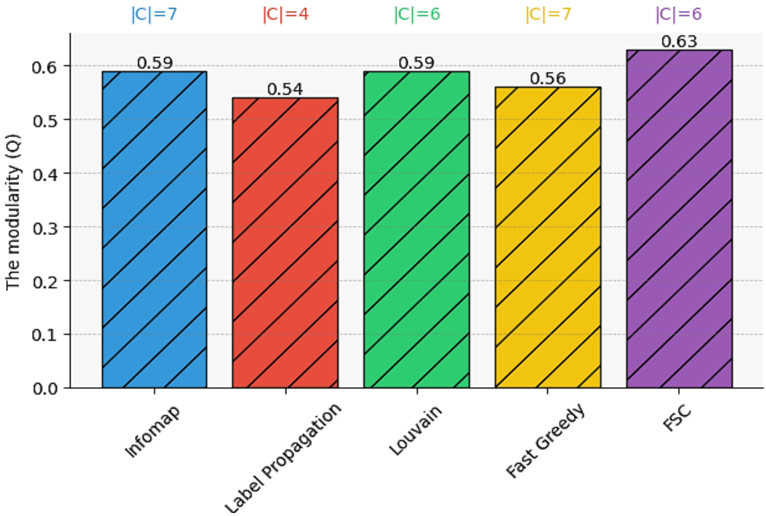
a) Les Miserables network



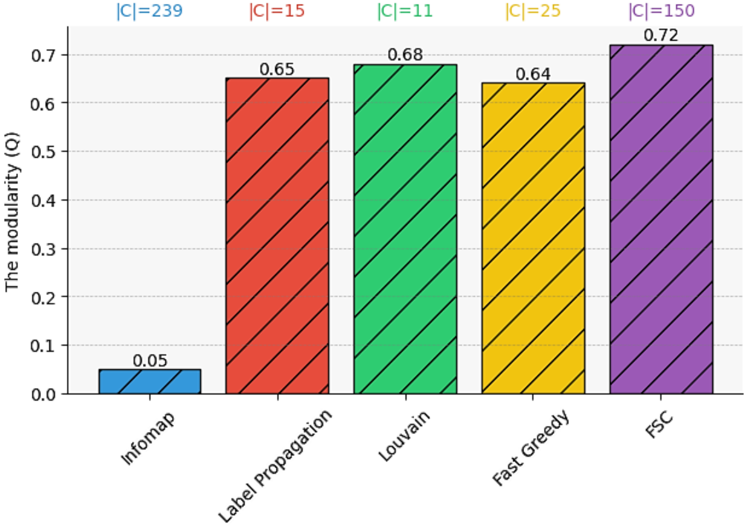
b) Jazz network



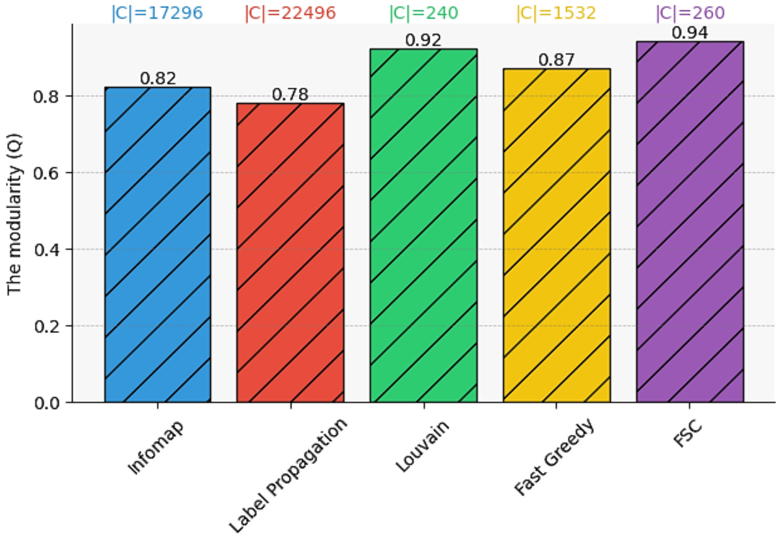
c) The HIV network



d) The Contiguous USA network



e) Facebook network



f) Amazon network

Fig. 3. Compares FSC with different unknown networks: a) Les Miserable Network; b) Jazz Network; c) HIV Network; d) USA Network; e) Facebook Network; f) Amazon Network

FSC outperformed alternative approaches in several cases, although the specific networks where it excelled varied across different contexts. The FSC approach has demonstrated exceptional performance when applied to the intricate networks of Facebook and Amazon, exhibiting commendable quality in community detection. Its success in the Amazon network particularly stands out, showcasing unparalleled excellence with Q index values surpassing 0.90. Moreover, within the Facebook network, the FSC method has achieved a remarkable Q index value of 0.72, significantly eclipsing the performance of alternative algorithms, as illustrated in the accompanying Figure 4. Figure 4 shows our novel method for partitioning the contiguous USA into six clusters, providing a new way to analyze regional patterns. Using a mixing parameter denoted as μ, FSC accurately partitions the graph and uncovers clear communities (Figure 5). This value of $\mu = 0.1$ is commonly used in community detection within networks because it indicates a clear community structure where only 10% of the edges are between different communities, while 90% of the edges are within the same community. The choice of $\mu = 0.1$ provides stability and robustness for accurate detection and has been empirically successful in various network studies [65]. By using this mixing parameter value, the FSC algorithm can effectively identify distinct communities within the graph, as shown in Figure 5.

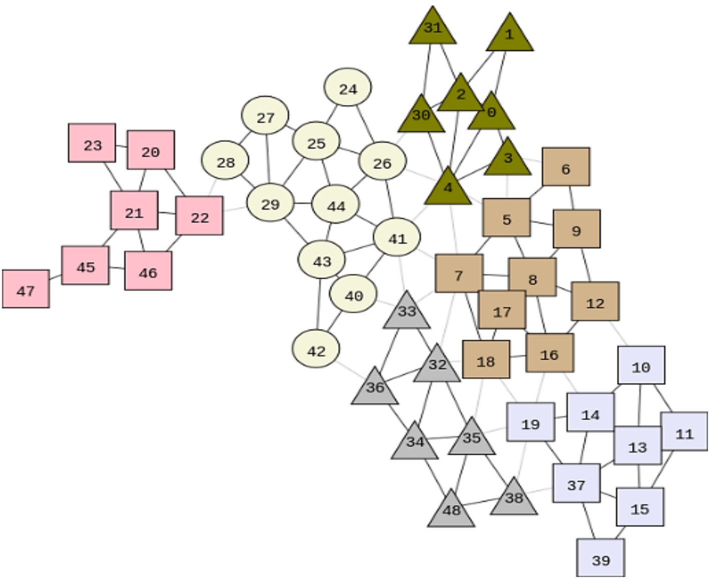


Fig. 4. FSC identifies 6 clusters in the Contiguous USA network

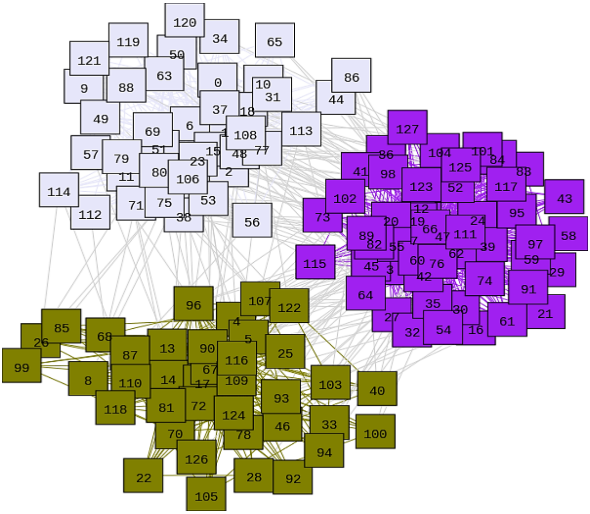


Fig. 5. FSC detects computer-generated networks at μ = 0.1

Figure 6 shows the FSC achievement of a maximum NMI value of 1 when the mixing parameter ranges from 0 to 0.4, indicating the successful identification of robust and well-defined community structures. The connections within communities are denser compared to other methods, highlighting our approach's effectiveness. However, other methods like Fast Greedy and Louvain have NMI values less than 1, indicating accuracy in identifying true communities, particularly for mixing parameters between 0 and 0.3. As the parameter exceeds 0.3, both Infomap and Label Propagation encounter challenges in defining distinct communities. Overall, as the mixing parameter increases, all methods face difficulties in accurately uncovering the true community structure. This decline in performance suggests that as more edges are added between different communities, making the network more interconnected, the task of identifying clear divisions and separating the network into distinct communities becomes increasingly challenging for all methods. According to [52], a network's community structure is significant when its modularity exceeds 0.3. FSC observed a high modularity value, indicating a strong community structure and the ability to uncover densely connected and cohesive communities (Figure 7). The higher modularity values imply strong internal connections and fewer interconnections with other communities, indicating that FSC is effective in capturing meaningful community divisions and that the community structure is robust.

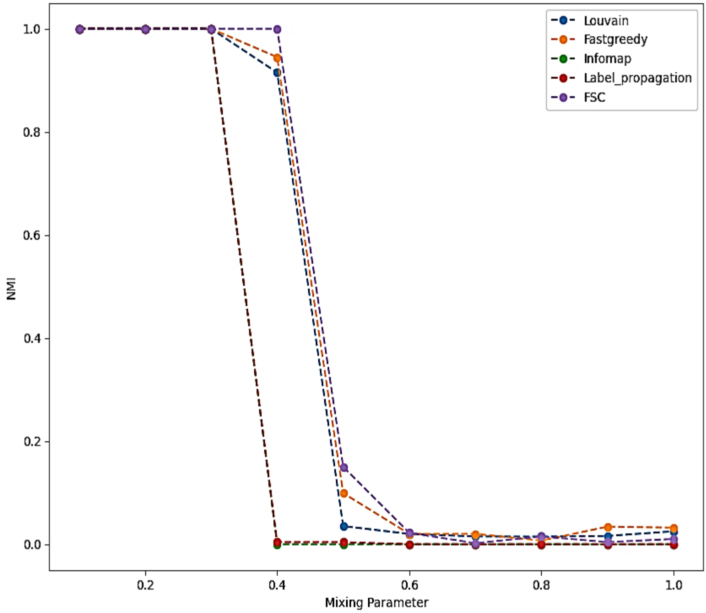


Fig. 6. How NMI varies with the level of community mixing

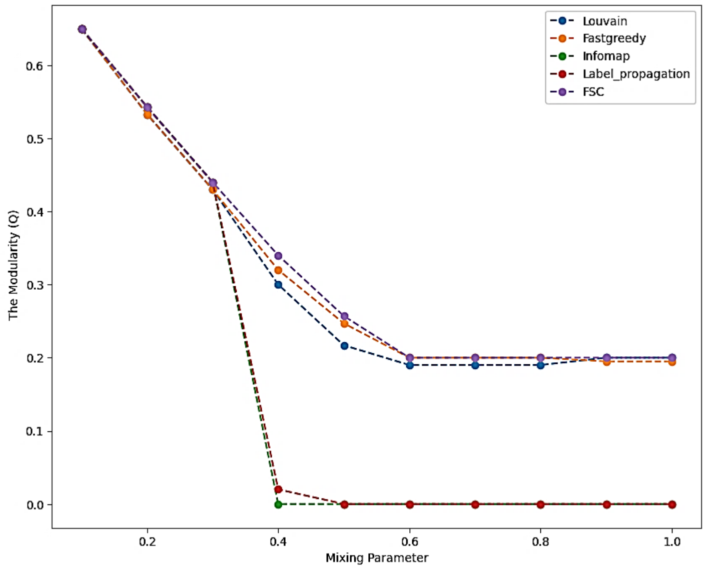


Fig. 7. How modularity changes with the degree of inter-community mixing

6. Conclusion and future prospects. We introduce a new method for clustering networks that uses the Fish School Search Algorithm to enhance the modularity function. The approach iteratively removes edges from the network to maximize the modularity function. Through rigorous testing on both established and novel network structures, including those generated by the LFR model, our FSC methodology has demonstrated exceptional prowess in discerning cohesive communities and pinpointing authentic community structures. This breakthrough signifies a leap forward in network analysis, offering a potent tool for researchers and practitioners alike to navigate the ever-evolving landscape of networks. As we stand on the brink of this new horizon, it is clear that the FSC method is more than just an algorithm; it is a beacon that guides us toward a deeper comprehension of the networks that encompass and connect us all. The implications are vast, from enhancing social network analysis to optimizing transportation systems and even unraveling the collaborative networks within our very cells. The future of network clustering is bright, and it is our conviction that the FSC method will be at the forefront of this transformative journey. In future work, FSC can be extended to handle diverse network types, including directed, bipartite, or weighted networks. Additionally, research can explore alternative optimization algorithms or modularity functions to enhance FSC's performance in identifying community structures. Incorporating side information, such as node attributes or interactions with other nodes, is also a promising avenue for improvement.

References

1. Pavlopoulos G. Using graph theory to analyze biological networks. Pavlopoulos et al. BioData Mining. 2011. vol. 4. pp. 1–27.
2. Sorokina M., Medigue C., Vallenet D. A new network representation of the metabolism to detect chemical transformation modules. BMC Bioinformatics. 2015. vol. 16. DOI: 10.1186/s12859-015-0809-4.
3. Clauset A., et al. The Structure and Function of Complex Networks. Trends in Ecology and Evolution. 2011. vol. 45. no. 2. p. 78.
4. Mathew J., Rejikumar K. Communication Structures, its graph representation and decomposition possibilities. Journal of Physics: Conference Series. 2020. DOI: 10.1088/1742-6596/1706/1/012050.
5. Sarker I. Data Science and Analytics: An Overview from Data-Driven Smart Computing, Decision-Making and Applications Perspective. SN Computer Science. 2021. vol. 2. no. 5. DOI: 10.1007/s42979-021-00765-8.
6. Lee M., Buckley C., Zhang X., Louhivuori L., Uhlen P., Wilson C., McCarron J. Small-world connectivity dictates collective endothelial cell signaling. Proceedings of the National Academy of Sciences. 2022. vol. 119. no. 18. DOI: 10.1073/PNAS.2118927119.
7. Hoffman M., Steinley D., Gates K., Prinstein M., Brusco M. Detecting Clusters/Communities in Social Networks. Multivariate behavioral research. 2018. vol. 53. no. 1. p. 57–73. DOI: 10.1080/00273171.2017.1391682.

8. Felmlee D., Kreager D. The invisible contours of online dating communities: A social network perspective. Journal of Social Structure. 2017. vol. 18. no. 1. pp. 1–28. DOI: 10.21307/JOSS-2018-004.
9. Inuwa-Dutse I., Liptrott M., Korkontzelos I. A multilevel clustering technique for community detection. Neurocomputing. 2021. vol. 441. pp. 64–78. DOI: 10.1016/J.NEUCOM.2021.01.059.
10. Bramson A., Hori M., Zha B., Inamoto H. Scoring and classifying regions via multimodal transportation networks. Applied Network Science. 2019. vol. 4. no. 1. DOI: 10.1007/S41109-019-0191-7.
11. Copic J., Jackson M., Kirman A. Identifying Community Structures from Network Data via Maximum Likelihood Methods. The BE Journal of Theoretical Economics. 2009. vol. 9. no. 1. DOI: 10.2202/1935-1704.1523.
12. Yang L., Cao X., He D., Wang C., Wang X., Zhang W. Modularity based community detection with deep learning. International Joint Conference on Artificial Intelligence. 2016. vol. 2016. pp. 2252–2258.
13. Roy S. Spectral clustering of graphs. 2018. pp. 1–65.
14. Cohen-Addad V., Kanade V., Mallmann-trenn F., Mathieu C., Hierarchical Clustering: Objective Functions and Algorithms. Journal of the ACM. 2019. vol. 66. no. 4. DOI: 10.1145/3321386.
15. Ghosh S., Halappanavar M., Tumeo A., Kalyanaraman A. Scaling and Quality of Modularity Optimization Methods for Graph Clustering. 2019. pp. 1–6. DOI: 10.1109/HPEC.2019.8916299.
16. Zhang X. et al. Modularity optimization in community detection of complex networks. Europhysics Letters. 2009. vol. 87. no. 3. DOI: 10.1209/0295-5075/87/38002.
17. Brandes U., Delling D., Gaertler M., Gorke R., Hoefer M., Nikoloski Z., Wagner D. Maximizing Modularity is hard. arXiv preprint physics/0608255. 2006.
18. Bastos Filho C., de Lima Neto F., Lins A., Nascimento A., Lima M. Fish school search. Nature-inspired algorithms for optimisation. 2009. vol. 193. pp. 261–277. DOI: 10.1007/978-3-642-00267-0_9.
19. Bastos-Filho C., Guimarães A. Multi-Objective Fish School Search. International Journal of Swarm Intelligence Research. 2015. vol. 6. no. 1. pp. 23–40. DOI: 10.4018/IJSIR.2015010102.
20. Newman M., Girvan M. Finding and evaluating community structure in networks. Physical Review E. 2004. vol. 69. no. 2. DOI: 10.1103/PhysRevE.69.026113.
21. Lukac Z. Metaheuristic optimization. Proceedings of the 11th International Symposium on Operational Research in Slovenia, SOR 2011. 2011. pp. 17–22. DOI: 10.4249/SCHOLARPEDIA.11472.
22. Kim J., Luo S., Cong G., Yu W. DMCS : Density Modularity based Community Search. Proceedings of the ACM SIGMOD International Conference on Management of Data. 2022. pp. 889–903. DOI: 10.1145/3514221.3526137.
23. Tantardini M., Ieva F., Tajoli L., Piccardi C. Comparing methods for comparing networks. Scientific Reports 2019vol. 9. no. 1. DOI: 10.1038/s41598-019-53708-y.
24. Rustamaji H., Kusuma W., Nurdiati S., Batubara I. Community detection with greedy modularity disassembly strategy. Scientific Reports. 2024. vol. 14. no. 1. DOI: 10.1038/s41598-024-55190-7.
25. Drazdilová P., Prokop P., Platoš J., Snášel V. A hierarchical overlapping community detection method based on closed trail distance and maximal cliques. Information Sciences. 2024. vol. 662. DOI: 10.1016/J.INS.2024.120271.
26. Newman M. Fast algorithm for detecting community structure in networks. Physical Review E – Statistical Physics, Plasmas, Fluids, and Related Interdisciplinary Topics. 2004. vol. 69. no. 6. DOI: 10.1103/PhysRevE.69.066133.

27. Dorfler F., Bullo F. Kron reduction of graphs with applications to electrical networks. IEEE Transactions on Circuits and Systems I: Regular Papers. 2012. vol. 60. no. 1. pp. 150–163. DOI: 10.1109/TCSI.2012.2215780.
28. Bickel P., Chen A. A nonparametric view of network models and Newman-Girvan and other modularities. Proceedings of the National Academy of Sciences. 2009. vol. 106. no. 50. pp. 21068–21073. DOI: 10.1073/pnas.0907096106.
29. Brandes U., Delling D., Gaertler M., Hoefer R., M Nikoloski Z., Wagner D. On Modularity Clustering. IEEE Transactions on knowledge and data engineering. 2008. vol. 20. no. 2. pp. 172–188. DOI: 10.1109/TKDE.2007.190689.
30. Hu Y. Swarm intelligence Ant colony optimization algorithms. 2012. pp. 1–11.
31. Dehuri S., Ghosh S., Coello C. An introduction to swarm intelligence for multi-objective problems. Swarm Intelligence for Multi-objective Problems in Data Mining. Studies in Computational Intelligence. 2009. vol. 242. DOI: 10.1007/978-3-642-03625-5_1.
32. Herbert-Read J., Perna A., Mann R., Schaerf T., Sumpter D., Ward A. Inferring the rules of interaction of shoaling fish. Proceedings of the National Academy of Sciences. 2011. vol. 108. no. 46. pp. 18726–18731. DOI: 10.1073/PNAS.1109355108.
33. A novel search algorithm based on fish school behavior. 2008 IEEE International Conference on Systems, Man and Cybernetics. Available at: https://sci-hub.se/10.1109/ICSMC.2008.4811695 (accessed: 14/05/2024).
34. Molina D., Poyatos J., Del Ser J., García S., Hussain A., Herrera F. Comprehensive Taxonomies of Nature- and Bio-inspired Optimization: Inspiration Versus Algorithmic Behavior, Critical Analysis Recommendations. Cognitive Computation. 2020. vol. 12. pp. 897–939. DOI: 10.1007/S12559-020-09730-8.
35. Abangan A., Kopp D., Faillettaz R. Artificial intelligence for fish behavior recognition may unlock fishing gear selectivity. Front Mar Sci. 2023. vol. 10. DOI: 10.3389/FMARS.2023.1010761.
36. Krongauz D., Lazebnik T. Collective evolution learning model for visionbased collective motion with collision avoidance. PLoS One. 2023. vol. 18. no. 5. DOI: 10.1371/JOURNAL.PONE.0270318.
37. Malone T. Collective intelligence. 2007 International Symposium on Collaborative Technologies and Systems. 2008. DOI: 10.1109/CTS.2007.4621716.
38. Huang Z., Application of the Artificial Fish School Algorithm and Particle Filter Algorithm in the Industrial Process Control Particle Filtering Algorithm for Industrial Process Control. Math Probl Eng. 2020. vol. 2020. DOI: 10.1155/2020/3070539.
39. Liu Z., Zhong P., Liu H., Jia W., Sa G., Tan J. Module partition for complex products based on stable overlapping community detection and overlapping component allocation. Research in Engineering Design. 2024. vol. 35. pp. 269–288. DOI: 10.1007/S00163-024-00432-Y.
40. Sun Y., Sun Z., Chang X., Pan Z., Luo L. Community Detection Based on Fish School Effect. IEEE Access. 2022. vol. 10. pp. 48523–48538. DOI: 10.1109/ACCESS.2022.3172298.
41. Saoud B. Networks clustering with bee colony. Artificial Intelligence Review. 2019. vol. 52. no. 2. pp. 1297–1309. DOI: 10.1007/s10462-018-9657-8.
42. Encord E. Available at: https://encord.com/try-it-free/?&utm_campaign=cta-blog-medical-dark (accessed: 10/03/2024).
43. Cetin P., Amrahov Ş. A new network-based community detection algorithm for disjoint communities. Turkish Journal of Electrical Engineering and Computer Sciences. 2022. vol. 30. no. 6. pp. 2190–2205. DOI: 10.55730/1300-0632.3933.
44. Bonchi F., García-Soriano D., Miyauchi A., Tsourakakis C. Finding densest k-connected subgraphs. Discrete Applied Mathematics. 2021. vol. 305. pp. 34–47. DOI: 10.1016/J.DAM.2021.08.032.

45. Oldham S., Fulcher B., Parkes L., Arnatkeviciute A., Suo C., Fornito A. Consistency and differences between centrality measures across distinct classes of networks. PLoS One. 2019. vol. 14. no. 7. DOI: 10.1371/JOURNAL.PONE.0220061.
46. Leskovec J., Lang K., Dasgupta A., Mahoney M. Community Structure in Large Networks: Natural Cluster Sizes and the Absence of Large Well-Defined Clusters. Internet Mathematics. 2008. vol. 6. no. 1. pp. 29–123.
47. Adoni W., Nahhal T., Krichen M., El byed A., Assayad I. DHPV: a distributed algorithm for large-scale graph partitioning. Journal of Big Data. 2020. vol. 7. DOI: 10.1186/S40537-020-00357-Y.
48. Newman M. Modularity and community structure in networks. Available at: https://sci-hub.se/10.1073/pnas.0601602103 (accessed: 14/05/2024).
49. Rosvall M., Bergstrom C. Maps of random walks on complex networks reveal community structure. Proc Natl Acad Sci U S A. 2008. vol. 105. no. 4. pp. 1118–1123. DOI: 10.1073/PNAS.0706851105.
50. Cordasco G., Gargano L. Label propagation algorithm: a semi-synchronous approach. International Journal of Social Network Mining. 2012. vol. 1. no. 1. DOI: 10.1504/IJSNM.2012.045103.
51. Blondel V., Guillaume J.-L., Lambiotte R., Lefebvre E. Fast unfolding of communities in large networks. Journal of statistical mechanics: theory and experiment. 2008. vol. 2008. no. 10. DOI: 10.1088/1742-5468/2008/10/P10008.
52. Clauset A., Newman M., Moore C. Finding community structure in very large networks. Physical Review E – Statistical, Nonlinear, and Soft Matter Physics. 2004. vol. 70. no. 6. DOI: 10.1103/PhysRevE.70.066111.
53. Danon L., DIaz-Guilera A., Duch J., Arenas A. Comparing community structure identification. Journal of statistical mechanics: Theory and experiment. 2005. vol. 2005. no. 09. DOI: 10.1088/1742-5468/2005/09/P09008.
54. Lancichinetti A., Fortunato S., Radicchi F. Benchmark graphs for testing community detection algorithms. Physical Review E – Statistical, Nonlinear, and Soft Matter Physics. 2008. vol. 78. no. 4. DOI: 10.1103/PHYSREVE.78.046110.
55. Zachary W. An Information Flow Model for Conflict and Fission in Small Groups. Journal of anthropological research. 1977. vol. 33. no. 4. pp. 452–473. DOI: 10.1086/JAR.33.4.3629752.
56. Girvan M., Newman M. Community structure in social and biological networks. Proceedings of the national academy of sciences. 2002. vol. 99. no. 12. pp. 7821–7826. DOI: 10.1073/pnas.122653799.
57. Lusseau D., Schneider K., Boisseau O., Haase P., Slooten E., Dawson S. The bottlenose dolphin community of doubtful sound features a large proportion of long-lasting associations: Can geographic isolation explain this unique trait? Behavioral Ecology and Sociobiology. 2003. vol. 54. pp. 396–405. DOI: 10.1007/S00265-003-0651-Y.
58. Newman M. Network data. Network data collection. 2013. Available at: http://www-personal.umich.edu/~mej. (accessed 26.10.2023).
59. Leskovec J., Mcauley J. Learning to Discover Social Circles in Ego Networks. Advances in neural information processing systems. 2012. vol. 25.
60. Leskovec J., Lada A., Huberman B. The Dynamics of Viral Marketing. ACM Transactions on the Web (TWEB). 2007. vol. 1. no. 1. DOI: 10.1145/1232722.1232727.
61. Knuth D. The Stanford GraphBase: a platform for combinatorial computing. New York: ACM Press, Addison-Wesley Publishing Company. 2009. 577 p.
62. Gleiser P., Danon L. Community structure in jazz. Advances in complex systems. 2003. vol. 6. no. 04. pp. 565–573.

63. Kunegis J. Konect: The Koblenz Network Collection. Proceedings of the 22nd international conference on world wide web. 2013. pp. 1343–1350. DOI: 10.1145/2487788.248817.
64. Knuth D. The Art of Computer Programming: Volume 4 Fascicle 0: Introduction to Combinatorial Algorithms and Boolean Functions. Addison-Wesley Professional, 2008. 228 p.
65. Yang Z., Algesheimer R., Tessone C. A comparative analysis of community detection algorithms on artificial networks. Scientific reports. 2016. vol. 6. no. 1 DOI: 10.1038/srep30750.

Ibrahim Abuzer — Researcher, Department of computer science, Lim Laboratory, Akli Mohand Oulhadj University of Bouira. Research interests: computer networking, network communication, networking, network architecture, routing, information security. The number of publications — 1. h.abouzer@univ-bouira.dz; Rue Drissi Yahia Bouira, 10000, Bouira, Algeria; office phone: +213(558)606-511.

Boudref Mohamed — Researcher, Department of mathematics, Lim Laboratory, Akli Mohand Oulhadj University of Bouira. Research interests: computer networking, network communication, networking, network architecture, routing, information security. The number of publications — 3. m.boudref@univ-bouira.dz; Rue Drissi Yahia Bouira, 10000, Bouira, Algeria; office phone: +213(550)251-333.

Badis Lyes — Researcher, Department of computer science, Lim Laboratory, Akli Mohand Oulhadj University of Bouira. Research interests: computer networking, network communication, networking, network architecture, routing, information security. The number of publications — 7. l.badis@univ-bouira.dz; Rue Drissi Yahia Bouira, 10000, Bouira, Algeria; office phone: +213(659)564-064.

DOI 10.15622/ia.23.5.4

А.Х. Ибрагим, М.А. Будреф, Л. Бадис

КЛАСТЕРИЗАЦИЯ СЕТЕЙ С ИСПОЛЬЗОВАНИЕМ АЛГОРИТМА ПОИСКА КОСЯКОВ РЫБ

Ибрагим А.Х., Будреф М.А., Бадис Л. **Кластеризация сетей с использованием алгоритма поиска косяков рыб.**

Аннотация. Сеть представляет собой совокупность узлов, соединенных ребрами, которые представляют сущности и их взаимосвязи. В кластеризации социальных сетей узлы организованы в кластеры в соответствии с их шаблонами соединений с целью обнаружения сообществ. Выявление структур сообществ в сетях является важным. Однако существующие методы обнаружения сообществ еще не использовали потенциал алгоритма поиска косяков рыб (FSS) и принципов модулярности. Мы предложили новый метод, основанный на кластеризации с использованием алгоритма поиска рыбной школы и функции модулярности (FSC), который улучшает модулярность в кластеризации сети путем итерационного разбиения сети и оптимизации функции модулярности. Этот подход облегчает обнаружение высокомодулярных структур сообществ, улучшая разрешение и эффективность кластеризации сети. Мы протестировали FSC на известных и неизвестных структурах сетей. Также мы протестировали его на сети, сгенерированной с использованием модели LFR, чтобы проверить его производительность на сетях с различными структурами сообществ. Наша методология демонстрирует высокую эффективность в выявлении структур сообществ, что указывает на ее способность эффективно захватывать сплоченные сообщества и точно определять фактические структуры сообществ.

Ключевые слова: кластеризация, алгоритм поиска косяков рыб, функция модульности, сетевые структуры.

Литература

1. Pavlopoulos G. Using graph theory to analyze biological networks. Pavlopoulos et al. BioData Mining. 2011. vol. 4. pp. 1–27.
2. Sorokina M., Medigue C., Vallenet D. A new network representation of the metabolism to detect chemical transformation modules. BMC Bioinformatics. 2015. vol. 16. DOI: 10.1186/s12859-015-0809-4.
3. Clauset A., et al. The Structure and Function of Complex Networks. Trends in Ecology and Evolution. 2011. vol. 45. no. 2. p. 78.
4. Mathew J., Rejikumar K. Communication Structures, its graph representation and decomposition possibilities. Journal of Physics: Conference Series. 2020. DOI: 10.1088/1742-6596/1706/1/012050.
5. Sarker I. Data Science and Analytics: An Overview from Data-Driven Smart Computing, Decision-Making and Applications Perspective. SN Computer Science. 2021. vol. 2. no. 5. DOI: 10.1007/s42979-021-00765-8.
6. Lee M., Buckley C., Zhang X., Louhivuori L., Uhlen P., Wilson C., McCarron J. Small-world connectivity dictates collective endothelial cell signaling. Proceedings of the National Academy of Sciences. 2022. vol. 119. no. 18. DOI: 10.1073/PNAS.2118927119.
7. Hoffman M., Steinley D., Gates K., Prinstein M., Brusco M. Detecting Clusters/Communities in Social Networks. Multivariate behavioral research. 2018. vol. 53. no. 1. p. 57–73. DOI: 10.1080/00273171.2017.1391682.

8. Felmlee D., Kreager D. The invisible contours of online dating communities: A social network perspective. Journal of Social Structure. 2017. vol. 18. no. 1. pp. 1–28. DOI: 10.21307/JOSS-2018-004.
9. Inuwa-Dutse I., Liptrott M., Korkontzelos I. A multilevel clustering technique for community detection. Neurocomputing. 2021. vol. 441. pp. 64–78. DOI: 10.1016/J.NEUCOM.2021.01.059.
10. Bramson A., Hori M., Zha B., Inamoto H. Scoring and classifying regions via multimodal transportation networks. Applied Network Science. 2019. vol. 4. no. 1. DOI: 10.1007/S41109-019-0191-7.
11. Copic J., Jackson M., Kirman A. Identifying Community Structures from Network Data via Maximum Likelihood Methods. The BE Journal of Theoretical Economics. 2009. vol. 9. no. 1. DOI: 10.2202/1935-1704.1523.
12. Yang L., Cao X., He D., Wang C., Wang X., Zhang W. Modularity based community detection with deep learning. International Joint Conference on Artificial Intelligence. 2016. vol. 2016. pp. 2252–2258.
13. Roy S. Spectral clustering of graphs. 2018. pp. 1–65.
14. Cohen-Addad V., Kanade V., Mallmann-trenn F., Mathieu C., Hierarchical Clustering: Objective Functions and Algorithms. Journal of the ACM. 2019. vol. 66. no. 4. DOI: 10.1145/3321386.
15. Ghosh S., Halappanavar M., Tumeo A., Kalyanaraman A. Scaling and Quality of Modularity Optimization Methods for Graph Clustering. 2019. pp. 1–6. DOI: 10.1109/HPEC.2019.8916299.
16. Zhang X. et al. Modularity optimization in community detection of complex networks. Europhysics Letters. 2009. vol. 87. no. 3. DOI: 10.1209/0295-5075/87/38002.
17. Brandes U., Delling D., Gaertler M., Gorke R., Hoefer M., Nikoloski Z., Wagner D. Maximizing Modularity is hard. arXiv preprint physics/0608255. 2006.
18. Bastos Filho C., de Lima Neto F., Lins A., Nascimento A., Lima M. Fish school search. Nature-inspired algorithms for optimisation. 2009. vol. 193. pp. 261–277. DOI: 10.1007/978-3-642-00267-0_9.
19. Bastos-Filho C., Guimarães A. Multi-Objective Fish School Search. International Journal of Swarm Intelligence Research. 2015. vol. 6. no. 1. pp. 23–40. DOI: 10.4018/IJSIR.2015010102.
20. Newman M., Girvan M. Finding and evaluating community structure in networks. Physical Review E. 2004. vol. 69. no. 2. DOI: 10.1103/PhysRevE.69.026113.
21. Lukac Z. Metaheuristic optimization. Proceedings of the 11th International Symposium on Operational Research in Slovenia, SOR 2011. 2011. pp. 17–22. DOI: 10.4249/SCHOLARPEDIA.11472.
22. Kim J., Luo S., Cong G., Yu W. DMCS : Density Modularity based Community Search. Proceedings of the ACM SIGMOD International Conference on Management of Data. 2022. pp. 889–903. DOI: 10.1145/3514221.3526137.
23. Tantardini M., Ieva F., Tajoli L., Piccardi C. Comparing methods for comparing networks. Scientific Reports 2019vol. 9. no. 1. DOI: 10.1038/s41598-019-53708-y.
24. Rustamaji H., Kusuma W., Nurdiati S., Batubara I. Community detection with greedy modularity disassembly strategy. Scientific Reports. 2024. vol. 14. no. 1. DOI: 10.1038/s41598-024-55190-7.
25. Drazdilová P., Prokop P., Platoš J., Snášel V. A hierarchical overlapping community detection method based on closed trail distance and maximal cliques. Information Sciences. 2024. vol. 662. DOI: 10.1016/J.INS.2024.120271.
26. Newman M. Fast algorithm for detecting community structure in networks. Physical Review E – Statistical Physics, Plasmas, Fluids, and Related Interdisciplinary Topics. 2004. vol. 69. no. 6. DOI: 10.1103/PhysRevE.69.066133.

27. Dorfler F., Bullo F. Kron reduction of graphs with applications to electrical networks. IEEE Transactions on Circuits and Systems I: Regular Papers. 2012. vol. 60. no. 1. pp. 150–163. DOI: 10.1109/TCSI.2012.2215780.
28. Bickel P., Chen A. A nonparametric view of network models and Newman-Girvan and other modularities. Proceedings of the National Academy of Sciences. 2009. vol. 106. no. 50. pp. 21068–21073. DOI: 10.1073/pnas.0907096106.
29. Brandes U., Delling D., Gaertler M., Hoefer R., M Nikoloski Z., Wagner D. On Modularity Clustering. IEEE Transactions on knowledge and data engineering. 2008. vol. 20. no. 2. pp. 172–188. DOI: 10.1109/TKDE.2007.190689.
30. Hu Y. Swarm intelligence Ant colony optimization algorithms. 2012. pp. 1–11.
31. Dehuri S., Ghosh S., Coello C. An introduction to swarm intelligence for multi-objective problems. Swarm Intelligence for Multi-objective Problems in Data Mining. Studies in Computational Intelligence. 2009. vol. 242. DOI: 10.1007/978-3-642-03625-5_1.
32. Herbert-Read J., Perna A., Mann R., Schaerf T., Sumpter D., Ward A. Inferring the rules of interaction of shoaling fish. Proceedings of the National Academy of Sciences. 2011. vol. 108. no. 46. pp. 18726–18731. DOI: 10.1073/PNAS.1109355108.
33. A novel search algorithm based on fish school behavior. 2008 IEEE International Conference on Systems, Man and Cybernetics. Available at: https://sci-hub.se/10.1109/ICSMC.2008.4811695 (accessed: 14/05/2024).
34. Molina D., Poyatos J., Del Ser J., García S., Hussain A., Herrera F. Comprehensive Taxonomies of Nature- and Bio-inspired Optimization: Inspiration Versus Algorithmic Behavior, Critical Analysis Recommendations. Cognitive Computation. 2020. vol. 12. pp. 897–939. DOI: 10.1007/S12559-020-09730-8.
35. Abangan A., Kopp D., Faillettaz R. Artificial intelligence for fish behavior recognition may unlock fishing gear selectivity. Front Mar Sci. 2023. vol. 10. DOI: 10.3389/FMARS.2023.1010761.
36. Krongauz D., Lazebnik T. Collective evolution learning model for visionbased collective motion with collision avoidance. PLoS One. 2023. vol. 18. no. 5. DOI: 10.1371/JOURNAL.PONE.0270318.
37. Malone T. Collective intelligence. 2007 International Symposium on Collaborative Technologies and Systems. 2008. DOI: 10.1109/CTS.2007.4621716.
38. Huang Z., Application of the Artificial Fish School Algorithm and Particle Filter Algorithm in the Industrial Process Control Particle Filtering Algorithm for Industrial Process Control. Math Probl Eng. 2020. vol. 2020. DOI: 10.1155/2020/3070539.
39. Liu Z., Zhong P., Liu H., Jia W., Sa G., Tan J. Module partition for complex products based on stable overlapping community detection and overlapping component allocation. Research in Engineering Design. 2024. vol. 35. pp. 269–288. DOI: 10.1007/S00163-024-00432-Y.
40. Sun Y., Sun Z., Chang X., Pan Z., Luo L. Community Detection Based on Fish School Effect. IEEE Access. 2022. vol. 10. pp. 48523–48538. DOI: 10.1109/ACCESS.2022.3172298.
41. Saoud B. Networks clustering with bee colony. Artificial Intelligence Review. 2019. vol. 52. no. 2. pp. 1297–1309. DOI: 10.1007/s10462-018-9657-8.
42. Encord E. Available at: https://encord.com/try-it-free/?&utm_campaign=cta-blog-medical-dark (accessed: 10/03/2024).
43. Cetin P., Amrahov Ş. A new network-based community detection algorithm for disjoint communities. Turkish Journal of Electrical Engineering and Computer Sciences. 2022. vol. 30. no. 6. pp. 2190–2205. DOI: 10.55730/1300-0632.3933.
44. Bonchi F., García-Soriano D., Miyauchi A., Tsourakakis C. Finding densest k-connected subgraphs. Discrete Applied Mathematics. 2021. vol. 305. pp. 34–47. DOI: 10.1016/J.DAM.2021.08.032.

45. Oldham S., Fulcher B., Parkes L., Arnatkeviciute A., Suo C., Fornito A. Consistency and differences between centrality measures across distinct classes of networks. PLoS One. 2019. vol. 14. no. 7. DOI: 10.1371/JOURNAL.PONE.0220061.
46. Leskovec J., Lang K., Dasgupta A., Mahoney M. Community Structure in Large Networks: Natural Cluster Sizes and the Absence of Large Well-Defined Clusters. Internet Mathematics. 2008. vol. 6. no. 1. pp. 29–123.
47. Adoni W., Nahhal T., Krichen M., El byed A., Assayad I. DHPV: a distributed algorithm for large-scale graph partitioning. Journal of Big Data. 2020. vol. 7. DOI: 10.1186/S40537-020-00357-Y.
48. Newman M. Modularity and community structure in networks. Available at: https://sci-hub.se/10.1073/pnas.0601602103 (accessed: 14/05/2024).
49. Rosvall M., Bergstrom C. Maps of random walks on complex networks reveal community structure. Proc Natl Acad Sci U S A. 2008. vol. 105. no. 4. pp. 1118–1123. DOI: 10.1073/PNAS.0706851105.
50. Cordasco G., Gargano L. Label propagation algorithm: a semi-synchronous approach. International Journal of Social Network Mining. 2012. vol. 1. no. 1. DOI: 10.1504/IJSNM.2012.045103.
51. Blondel V., Guillaume J.-L., Lambiotte R., Lefebvre E. Fast unfolding of communities in large networks. Journal of statistical mechanics: theory and experiment. 2008. vol. 2008. no. 10. DOI: 10.1088/1742-5468/2008/10/P10008.
52. Clauset A., Newman M., Moore C. Finding community structure in very large networks. Physical Review E – Statistical, Nonlinear, and Soft Matter Physics. 2004. vol. 70. no. 6. DOI: 10.1103/PhysRevE.70.066111.
53. Danon L., DIaz-Guilera A., Duch J., Arenas A. Comparing community structure identification. Journal of statistical mechanics: Theory and experiment. 2005. vol. 2005. no. 09. DOI: 10.1088/1742-5468/2005/09/P09008.
54. Lancichinetti A., Fortunato S., Radicchi F. Benchmark graphs for testing community detection algorithms. Physical Review E – Statistical, Nonlinear, and Soft Matter Physics. 2008. vol. 78. no. 4. DOI: 10.1103/PHYSREVE.78.046110.
55. Zachary W. An Information Flow Model for Conflict and Fission in Small Groups. Journal of anthropological research. 1977. vol. 33. no. 4. pp. 452–473. DOI: 10.1086/JAR.33.4.3629752.
56. Girvan M., Newman M. Community structure in social and biological networks. Proceedings of the national academy of sciences. 2002. vol. 99. no. 12. pp. 7821–7826. DOI: 10.1073/pnas.122653799.
57. Lusseau D., Schneider K., Boisseau O., Haase P., Slooten E., Dawson S. The bottlenose dolphin community of doubtful sound features a large proportion of long-lasting associations: Can geographic isolation explain this unique trait? Behavioral Ecology and Sociobiology. 2003. vol. 54. pp. 396–405. DOI: 10.1007/S00265-003-0651-Y.
58. Newman M. Network data. Network data collection. 2013. Available at: http://www-personal.umich.edu/~mej. (accessed 26.10.2023).
59. Leskovec J., Mcauley J. Learning to Discover Social Circles in Ego Networks. Advances in neural information processing systems. 2012. vol. 25.
60. Leskovec J., Lada A., Huberman B. The Dynamics of Viral Marketing. ACM Transactions on the Web (TWEB). 2007. vol. 1. no. 1. DOI: 10.1145/1232722.1232727.
61. Knuth D. The Stanford GraphBase: a platform for combinatorial computing. New York: ACM Press, Addison-Wesley Publishing Company. 2009. 577 p.
62. Gleiser P., Danon L. Community structure in jazz. Advances in complex systems. 2003. vol. 6. no. 04. pp. 565–573.

63. Kunegis J. Konect: The Koblenz Network Collection. Proceedings of the 22nd international conference on world wide web. 2013. pp. 1343–1350. DOI: 10.1145/2487788.248817.
64. Knuth D. The Art of Computer Programming: Volume 4 Fascicle 0: Introduction to Combinatorial Algorithms and Boolean Functions. Addison-Wesley Professional, 2008. 228 p.
65. Yang Z., Algesheimer R., Tessone C. A comparative analysis of community detection algorithms on artificial networks. Scientific reports. 2016. vol. 6. no. 1 DOI: 10.1038/srep30750.

Ибрагим Абузер Хусейн — научный сотрудник, факультет компьютерных наук, Университет Буира – Акли Моханд Улхадж. Область научных интересов: компьютерные сети, сетевая связь, сети, сетевая архитектура, маршрутизация, информационная безопасность. Число научных публикаций — 1. h.abouzer@univ-bouira.dz; Рю Дрисси Яхья Буира, 10000, Буира, Алжир; р.т.: +213(558)606-511.

Будреф Мохамед Ахмед — научный сотрудник, факультет математики, Университет Буира – Акли Моханд Улхадж. Область научных интересов: компьютерные сети, сетевая связь, сети, сетевая архитектура, маршрутизация, информационная безопасность. Число научных публикаций — 3. m.boudref@univ-bouira.dz; Рю Дрисси Яхья Буира, 10000, Bouira, Алжир; р.т.: +213(550)251-333.

Бадис Лайес — научный сотрудник, факультет компьютерных наук, Университет Буира – Акли Моханд Улхадж. Область научных интересов: компьютерные сети, сетевая связь, сети, сетевая архитектура, маршрутизация, информационная безопасность. Число научных публикаций — 7. l.badis@univ-bouira.dz; Рю Дрисси Яхья Буира, 10000, Буира, Алжир; р.т.: +213(659)564-064.