Journal of Intelligent Systems and Internet of Things
Vol. 09, No. 02, PP. 178-193, 2023
DOI: https://doi.org/10.54216/JISIoT.090213

Diagnosis of Overlapping Communities and Coherent Groups Using Structural Centrality based Methodology

Tamarah Alaa Diame[1], Sajad Ali Zearah[2], Sahar R. Abdul Kadeem[3], Hiba Abdulameer Hasan[4,*], Munqith Saleem[5], Narjes Benameur[6], M. A. Burhanuddin[7]

[1]Technical Computer Engineering Department, Al-Kunooze University College, Basrah, Iraq
[2]Scientific Research Center, Al-ayen University, Thi-Qar, Iraq
[3]Department of Medical Devices Engineering Technologies, National University of Science and Technology, Dhi Qar, Nasiriyah, Iraq
[4]Computer Technologies Engineering, Al-Turath University College, Baghdad, Iraq
[5]Medical instruments engineering techniques, Al-farahidi University, Baghdad, Iraq
[6]Laboratory of Biophysics and Medical Technology, Higher Institute of Medical Technologies of Tunis, University of Tunis El Manar, Tunis 1006, Tunisia
[7]Faculty of Information & Communication Technology, Universiti Teknikal Malaysia Melaka, Durian Tunggal, Melaka, Malaysia
Emails: Tamarah.Alaa@Kunoozu.Edu. Iq; sajad@alayen.edu.iq; sahar@nust.edu.iq; hiba.abdalameer@turath.edu.iq; Munqith Saleem@uoalfarahidi.edu.iq; narjes.benameur@istmt.utm.tn; burhanuddin@utem.edu.my
[*]Corresponding Author: hiba.abdalameer@turath.edu.iq

Abstract

Community detection in complex networks has become an important step in understanding the structure and behaviour of networks in many fields. However, both standard node clustering and the relatively new link clustering methods have problems that make it hard to find combined clusters. Networks have been used to depict many types of real-world systems, such as those involving the transmission of information, the movement of funds, and biological processes. Communities are key structures for comprehending the structure of actual networks. The purpose of community detection is to identify meaningful subsets of these networks. Mesoscopically, a community consists of highly interconnected nodes within each subcommunity yet less strong connections across subcommunities. Communities can share a node or numerous nodes with overlapping. Evaluating the performance of a community detection method is crucial. Grouping the network's nodes into a family of subsets called clusters such that each cluster comprises similar nodes concerning the overall network structure is the problem of detecting overlapping communities in a network. Meanwhile, it has been shown that many methods for finding cluster centers have inherent flaws. Most methods are vulnerable to initial seeding and built-in variables, while others fail to highlight substantial overlaps. This article proposes the Structural Centrality Approach for Local Overlapping Community Detection (SCA-LOCD). It provides a novel approach to regional development that emphasizes the role of systems in identifying cluster centers. The fundamental concept behind this strategy is to identify structural centers in societies with coherent structures and then increase these centers using weighted methods and search engine techniques. Experimental results on synthetic and network systems show that the suggested technique is efficient and fascinating for detecting overlapped communities. It shows the success of regional extension strategies in identifying coherent groups and producing reliable classification results.

Received: March 02, 2023 Revised: May 28, 2023 Accepted: September 06, 2023

Keywords: density peaks; overlapping community detection; node density; structural centrality

1. Introduction

The broad adoption and use of social media sites (SMSs) have revolutionized interaction channels by enabling people from distinct places to engage and interest, thanks to the advent of Web 3.0 applications [1]. Many scholars and the mainstream have paid attention to platforms like Instagram, Facebook, Orkut, LinkedIn, Friendster, and Multiply, which garnered billions of members. While these services can help users, they also present difficult big data challenges and create more demands on activity recognition [2].

Community, also called clusters or units, are vertex groups in a network with similar traits or responsibilities [3]. The capacity to unravel these sub-structures in a networking site can reveal how the entire network and activities interact and higher stability for interested forecasting, friend referral, and incident development [4]. The technique of detecting all groups in a network is called clustering algorithms [5]. This single operation is accomplished using a wide range of ways.

There is no universally acknowledged definition of plant communities; it is widely understood that a formation has strong inner and sparse exterior linkages. Because it is frequently associated with management systems and functionality of network layers, as reported in some studies is regarded as a major structural component in complex patterns [6]. Yet, several studies have found that the role of the communal in complex patterns includes widespread overlap. It is good knowledge that users of online communities belong to many social communities [7].

It concentrates on traditional or complicated systems as a basic paradigm for clustering algorithms. A client will likely be connected to various social groups, such as family, colleagues, and hobbies. Other complicated networks, like bipartite networks, include cluster centers. As a result, the overlapped clustering algorithm, which aids in discovering group traits and structural aspects, is receiving more and more attention.

Various approaches for identifying communities in complicated systems have been presented in the last ten years, including rough sets, clustering, parameter variations, and methods involving optimization techniques [8]. These strategies aim to investigate community organizations from a variety of angles. Graph cut methods split a network into predetermined groups to reduce the overall amount of inter-community connections.

Cluster analysis is a popular method for revealing a graph's multilevel classification structures. Spectral techniques take advantage of the graph's spectrum features [9]. Iterations of the universal max of modularity are sought using an approach based on modular optimization. On the other hand, all the techniques listed above allocate every node to a particular community and the ability to identify multiple communities in the network [10]. As a result, there is a renewed emphasis on discovering the fundamental networks and groups that overlap. Structural centrality is a measure of the influence an actor has on a social network. It shows how different or better an actor's place or benefit is in the structure of the social network. Discovering the community that includes the provided node based on local information is called local community overlapping detection. This is especially beneficial when global knowledge about the network is inaccessible or too costly. Local overlapping community is determined by the nodes in numerous communities to which the given node fits. Then selecting, illustrative nodes from the nodes obtained tend to be in different communities. Finally, determine the communities to which these representative nodes belong. Finding local communities that meet is important for analyzing network structures, looking into how groups connect, learning how networks work, and understanding how networks change over time.

The major contribution of the article is

- Designing the Structural Centrality approach for local overlapping community detection (SCA-LOCD) is used to detect overlapped communities.
- To examine natural cluster centers, it concentrates on locating community architectural centers with a hierarchical coherence. The networks' localized concentration and the distance among nodes are considered when calculating structural relevance. It grows a structure and function from the architectural center to the boundary with a localized search technique and a weighted approach.
- Unlike earlier work, the suggested algorithm increases algorithm robustness and accelerates converging to optimum solutions using a unique seed selection method. Additionally, it avoids artificial parameter choice.

The upcoming section of the article is as follows: section 2 discusses the background of the local overlapping community detection models. The proposed Structural Centrality approach for the local overlapping community detection (SCA-LOCD) approach is considered and established in section 3. Section 4 illustrates the software analysis and comparison. The conclusion and limitations of the system are portrayed in section 5.

2. Related Works

Recently, several approaches for recognizing overlapping groups have been created. Yazdanparast et al. created the Clique Infiltration Methodology (CIM), which was predicated on the idea that society consists of overlapped core sets [11]. CIM finds communities by looking for nearby cliques and identifying all social circles of size k. However, the communal building's rigid rule was inappropriate for systems with densely coupled components. Chen et al. proposed a Link Clusters (LC) technique for community discovery founded on the assumption of link grouping, which involved hierarchical grouping of nearby connections using a similarity metric [12].

LC naturally described the overlapping among societies by changing link groups into nodal groups. Yet, there was no assurance that LC outperformed other techniques [13]. Because it promoted internal community relations while ignoring quality backlinks, it resulted in many tiny communities. Furthermore, by enabling a node to have several names, the proposed clustering method expanded to discover overlap communities. Proposed clustering approaches attained linear speed for densely connected discovery, but labeled grouping is still non-deterministic [14].

These approaches discovered communities are separated from one another and cannot have the same connections. Different groups frequently intersect in a network system. Ramesh et al. devised the clique infiltration approach to find an overlapped group [15]. However, recent investigations revealed its limited accuracy, as evidenced by normalized measure information (NMI). When LC does k-clique templates rolling, it only analyzes a small grouping range, which might decrease overall connectedness across all vertices in the collective [16].

The Lancichinetti–Fortunato methodology (LFM) extended the basic k-clique pattern rollover to the pessimistic fitness function optimization considering local and global areas. Generous clique extension (GCE) enhanced significantly over LFM by exploiting maximum cliques. It was vulnerable to the condition of a seed, which serves as the beginning point for growth. Yu et al. suggested a seed choice approach to improve the discovery of overlapped communities [17].

Moreover, expanding regional approaches were frequently employed in densely connected identification due to their great efficiency. One of the great benefits of the regional expansion approach over others was whether it detects communities using local area network data instead of the global system. As a result, it benefited computing difficulty and adaptability in large-scale and fuzzy systems. On the other hand, local expansion techniques were frequently sensitive to randomized seeding and built-in variables [18].

To find overlapped and multilayered communities, the LFM method introduced by Sheng et al. maintains choosing seeds randomly and extending groups around certain seeds by maximizing a fitness value regionally [19]. Because LFM heavily relied on chosen samples and a quality factor, it generated unreliable results. According to many studies, establishing groups depended on core members. The seedlings' integrity directly impacted the methods' effectiveness for seed-extension methods. Several seed-qualified candidates have been presented in the recent decade. Cheng et al. presented a simple technique based on seed production randomly [20].

Unpredictability promoted productivity but also rendered it difficult to find high-quality groups. Mahabadi et al. suggested an approach that used the program's k-cliques as starting communities to find highly clustered centers [21]. The algorithm, on the other hand, overlooked solitary sub-networks with insufficient size. According to the research, seeds were localized level center nodes with levels roughly equivalent to their neighbors' degrees. On the surface, the fundamental residents appeared to be regional degree key nodes. The technique identified all of a community's key members.

It might choose non-core individuals. Some techniques estimated the conductivity of every base station and chose the nodes with the lowest local conductivity as seeds. This approach identified greater seedlings. However, it was inefficient in terms of time. To manage communal size, Fatima et al. employed the local town neighborhood ratio tool [22]. Zhang et al.'s technique greedily developed communities depending on maximal cliques [23]. The effectiveness of discovering generalizations communities depends on built-in factors like the subgroup size k.

Zhang et al. also presented maximum sub-graphs and a related degree to determine overlaps [24]. For absorption neighbors, Zhang et al. recommended expanding core-vertex to an interpersonal degree [25]. An undirected system towards its center maximum tree structure was used. Yet, like all other local procedures, it was fed with appropriate parameters. As a result, local community identification still confronted certain difficulties.

Based on the survey, there are numerous drawbacks to the existing system in attaining high robustness and reliable classification results. Hence, this article proposes the structural Centrality approach for the local overlapping community detection (SCA-LOCD). It provides a novel approach to regional development that emphasizes the role of systems in identifying cluster centers.

3. Proposed Structural Centrality approach for local overlapping community detection (SCA-LOCD) approach

The ability to manage the regional grouping of vertices and the division of the graphs into communities is enabled by local town identification [26, 27]. As a result, two major issues in local town identification must be solved. The first is how to choose the origin networks to grow, which impacts the grouping outcomes' durability, and the second is architectural optimizing depending on local data. This paper presents an efficient method for selecting seed nodes that minimize susceptibility to origin seeds. Hence, this article proposes the structural Centrality approach for the local overlapping community detection (SCA-LOCD). It provides a novel approach to regional development that emphasizes the role of systems in identifying cluster centers.

The two steps of the suggested SCA-LOCD technique are architectural center discovery and community-based growth. It studies structural centers as beginning seeds and extends every seed regionally to identify cluster centers built on the principle that neighbors border them with a less-density distribution. Still, they're far away from any networks with a greater density distribution by a relatively huge distance. In contrast to prior research, this method seeks to find cluster centers in networks rather than disjoint groups. It uses fundamental centralization to locate structural centers in a system during the locating phase.

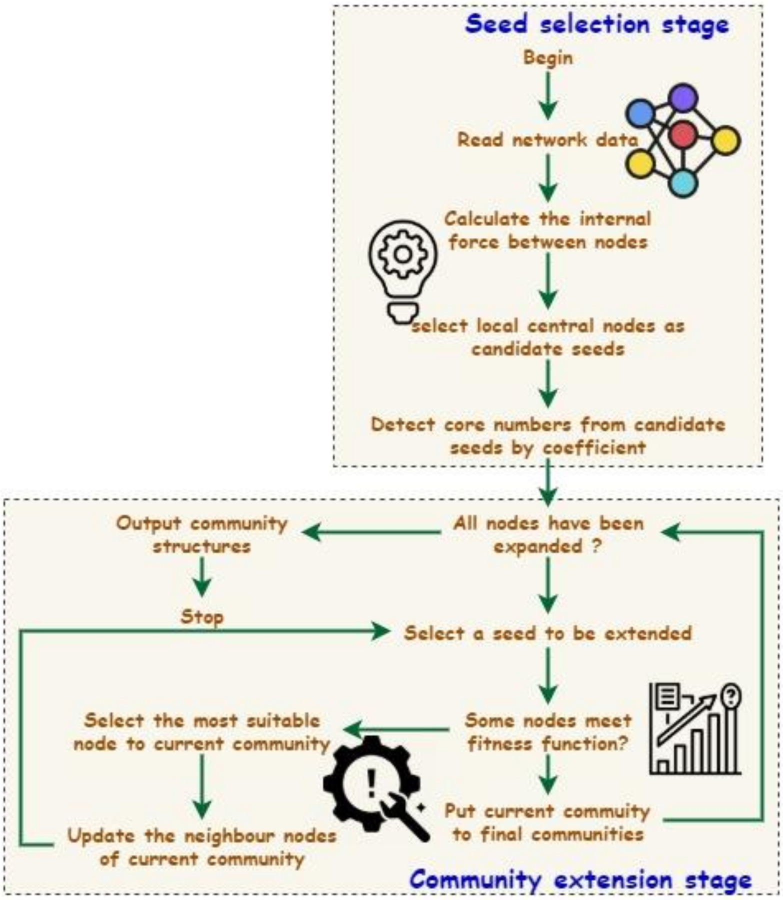

Figure 1: The system design of the proposed SCA-LOCD approach

The system design of the recommended SCA-LOCD approach is described in Figure 1. It has two stages such as the seed selection stage and the community extension stage. It uses seeds to find the local overlapping community detection model. In contrast to current centrality metrics like level, adjacency, eigenvalues relevance, infiltration importance, and localized Fiedler vectors centrality, structured centrality assesses a datatype local grouping central importance from a two-dimensional viewpoint. Based on a frequency graph, it analyses the algebraic connectedness of a graph following node or edge deletions. It was used to find deep networks by deleting nodes or connections with the highest centrality score.

The suggested strategy makes objects centrality to find seeds and then extends those seedlings locally to find cluster centers. This method may be considered a systematic enhancement of the LFM method. As previously stated, LFM picks seeds randomly, resulting in unforeseeable consequences. This instability is addressed here by establishing structural centers in communal structures.

3.1 User modeling

In this investigation, it chose Chinese microblogging as the process of gathering information. Yet, the suggested technique extended to every microblogging service and social networking platform written in other dialects with a few tweaks.

3.1.1. Ontology base

Because the user behavior data on a microblogging site are not reliant on a particular field but encompass all elements of real life, a broad ontology is used rather than a specialized one. Wikipedia is the most extensive and important Chinese ontology, created to construct a Chinese data-collecting platform covering all learning domains. There are 12 term channel categories in the first level, 89 in the second level, and 360,697 phrases in all four levels of the Encyclopedia Ontological Foundation as of June 2018. This research considers Wikipedia as the ontology basic for user modeling for these qualities.

3.1.2. User interest quantization

It measures user preferences by attention levels on certain terms in the ontologies foundation depending on their microblogging experiences. Frequent-inverse document repetition (TI) is used to create user preference profiles. The symbol for TI is expressed in Equation (1)

$$TI(D_x, s_y, D) = \frac{tf(s_y, D_x)}{\sum_{q=0}^{D_x} tf(s_y, D_x)} i(s_y, D) \quad (1)$$

When $TI(D_x, s_y, D)$ indicates the TI score of the search term s_y to article D_x in management structure D; $tf(s_y, D_x)$indicates the number of times that phrase s_y occurs in documents D_x; and $i(s_y, D)$ is denoted in Equation (2)

$$i(s_y, D) = \log\left[\frac{|D|}{(D_x \in D \rightarrow s_y \in D_x)}\right] \quad (2)$$

The seed is denoted s_y, and the management data is denoted D. The four main processes comprise measuring the levels of user preferences.

Step 1: Initialization

Users' most current microblogs are gathered and saved in a formal document for keyword extraction. Then, just the terms retrieved from the built ontologies basis are retrieved, and those that aren't eliminated.

Step 2: Quantification of degrees

The TI score of every phrase in every article is computed. All TI results connected with the same customer are transmitted to a four-layer ontological basis of the accompanying user preference modeling. It dubbed the customer's user profile interest prototype.

Step 3: Propagation of value

The straight application of the original user demand paradigm doesn't represent semantic relationships between terms, but it does need a lot of processing. As a result, only the two highest ontological foundations are retained in the user attention research model. It suggests that the two lowest ontological bases are destroyed, but the two lower phrases' attention levels are transmitted to the highest search terms.

The One-way Upwards Drift Theory spreads interest levels from lower levels to the surface layer. That, within the fourth stage, the attention level on a term s_x is increased by a factor of x and given to an interesting level on its parental phrase, etc. As a result, the propagation might occur in the following and second levels, where x is the inverse of the number of sister phrases of keywords. s_x. The diffusion system is presented in Equation (3) and modified in Equation (4)

$$J_y^* = J_y + \sum_{x=0}^{n_c^y} \propto_x J_x^y \tag{3}$$

$$J_y^* = J_y + \sum_{x=0}^{n_c^y} \frac{J_x^y}{n_c^y} \tag{4}$$

WhereJ_y involves the minimum interest level on keyword s_x; J_y^* is the ultimate function after transmission, n_c^y is the rate of child phrases, and J_x^y is the interest level on the x-th child phrase.

Step 4: Reduction of the model

The two highest phrases and the stable interest levels are kept to depict user preferences after transmission. Every level's attention level is normalized without losing generalization; the total is close to unity after homogeneous scale magnification.

3.1.3. User social attribute tagging

The nearer the sending and discussing postings are to the present point, the more personal the relationships among appropriate individuals become. Because these often recurring exchanges are founded on subscriber social relationships, the exchanges between individuals are asymmetrical. For clarity, the social features of the customer framework for this study solely include friends and admirers. See the associated literature for specific uses, like cooperative filtration systems, that need the inclusion of social affiliation levels.

3.1.4. User model representation

Following the quantification of user interests and the labeling of social relations, the formalized customer model is divided into three sections:

- The customer's basic characteristics, like creating an account, username, nationality, registration location, labels, etc.
- The Encyclopedia has the two highest ontology foundations with 100 terms with varying interest levels.
- Social qualities include the customer's admirers and supporters.

3.2 Algorithm description

The suggested approach is divided into seed picking and network expansion. It blends internal pressure among vertices with fitness values in the outreach stage to expand the community. Local grade designated points and the Jaccard value are used to discover the community's core individuals during the seed decision stage. The seedling choice and communal expansion stages are detailed in-depth in the subsequent subsections.

3.2.1 Seed selection

The grade of seedlings directly influences the efficacy of groups recognized by improvement in accuracy on grain. The chosen seeds must generally have a significant "impact" on the spatial relationship and be an integral part of societies. The seed selection algorithm is denoted in the below algorithm.

Seed selection model
Input – Graph $\boldsymbol{G(N,E),\delta}$
Output – seed selection
$\boldsymbol{S = 0}$
For every $\boldsymbol{v \in V}$ ***do***
If v is the regional centre node
$\boldsymbol{S = S \cup \{v\}}$
End if
End for
For every $\boldsymbol{v \in S}$ ***do***
$\boldsymbol{F(v) = \sum_{i=0}^{n} F(p,v) \mid p \in M(v)}$

End for
$\boldsymbol{F(max) = max(F(v)\|v \in S)}$
For every $\boldsymbol{v \in S}$ ***do***
$\boldsymbol{F'(v) = \frac{F(v)}{F(max)}}$
If $\boldsymbol{F'(v) < \delta}$
$\boldsymbol{S = S - \{v\}}$
End if
End for

The platform's regional degree core nodes are chosen as prospective seeds. These networks have a significant "impact" on their neighbors. Secondly, it involves the calculation of every option seed's Jaccard score with its nearby nodes to quantify the "impact" of options seeded on their neighboring nodes. The higher the value, the stronger the applicant seed's "impact" on neighboring nodes. Lastly, it normalizes every option seed's Jaccard correlation and determines a selection criterion.

When a nominee seed's normalized value exceeds, it is classified as a permanent part and utilized as the seeding in the research and extension phase. In the seeding selection phase, the parameters were set to 0.5 by the standard. It demonstrates the effectiveness of the seeding selection strategy using the well-studied Karate system. The system is divided into a class teacher and a membership manager.

3.2.2 Community extension

It explains the procedure of group expansion. It distributes seeds in the system to discover groups after getting them. The steps for extending a society can be stated as follows:

(1) The community welfare step is completed after all network elements are expanded. The seeds from the seedling selection phase are classified in a non-ascending sequence by degree. If the seeding set isn't blank, it chooses the seeding with the greatest degree to be expanded; otherwise, it chooses a node that hasn't been stretched arbitrarily as the seeding.

(2) The worth of every surrounding data fitness value to the present society is evaluated. The top k sites that maximize the present value of society's fitness values are chosen.

The cluster that can optimize the present society's function value is chosen. (3) The optimization algorithm with applied stress among nodes reliably calculates the k nearest vertices with the greatest result in step (2). It modifies the present society's adjacent nodes, and the procedure returns to step (2); the present society becomes an ultimately divided society, and the procedure returns to step (1).

(4) If the based on limited boosts the quality of the existing group's fitness values, it is merged.

It uses the value of k of 3 because of the program's tiny size. This method is repeated at the community welfare step until all network elements are expanded. It expands the seedlings acquired in the seedling selection step of the Karate networks. The method correctly separates the system into two groups, but every node is assigned to the appropriate community.

3.3 Locating Community Structural Centres

Interconnections have a mesoscale property called public architecture. Most techniques of installations presume that a data share is primarily a local system made up of things about subgroups and their extended neighbors. It makes sense in big networks when every node relies on its neighbors rather than the majority of its neighbors. For example, in social media, the role of the community is localized groupings with no connection to the rest of the system. Local extension techniques have a significant benefit; nonetheless, they are vulnerable to beginning seeding and creating variables. The workflow of the recommended SCA-LOCD approach is depicted in Figure 2.

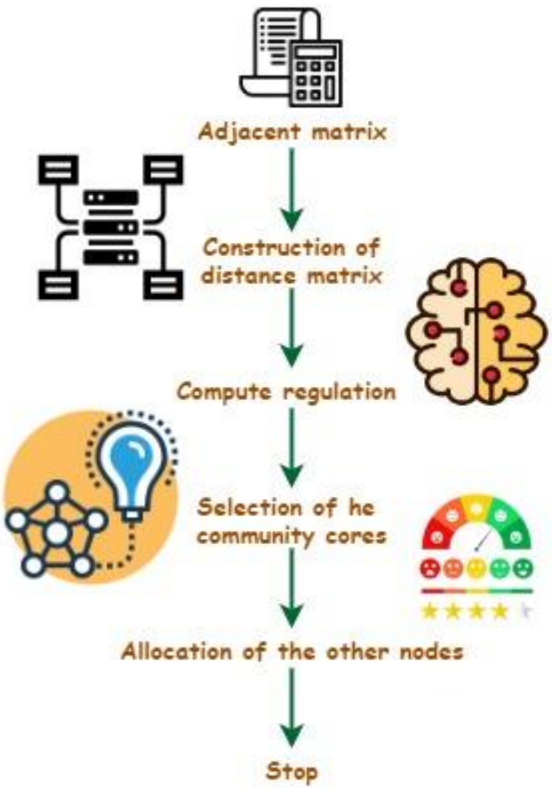

Figure 2: The workflow of the suggested SCA-LOCD approach

Adjacent Matrix

Finite graphs may be represented by the adjacency matrix. Whether or not a set of nodes are neighbors in a network may be inferred from the matrix's values. The adjacency matrix will be symmetric if the graph has no direction. It uses an adjacent matrix and a distance construction matrix to find the distance between nodes. On either side, they are separated from networks with greater density by a significant distance.

Construction of Distance Matrix

A distance matrix is a matrix that represents the weighted adjacency between nodes in a network. The distance between any two nodes in a network (a directed graph with weights given to the arcs) is the minimum of the weights on the shortest routes between those nodes. This paper looks into communal structural centers, a key device in the network. When contrasted to certain other nodes, architectural centers have two distinct characteristics. On the one side, structural centers have a larger density than their nearby nodes.

Compute Regulation

The regulation is computed, the community is selected, and the overlapping model is detected. Furthermore, they fail to estimate the number of clusters ahead of time.

Selection of the Community Cores

The proposed algorithm is based on integrated community core and community expansions. The community cores are the sub-set of the community, and the node in the core has greater contact frequencies, and the community structure alters slowly. If a node only links with one community core, this study can add these nodes to the corresponding communities.

Allocation of the Other Nodes

These architectural centers are evenly distributed among communities, meaning nodes are calculated intuitively. A community is an organizational system where units are coupled through the network under this premise.

It presents an architectural centrality that integrates the described context and the spatial variations to discover structural centers in a system. Both numbers are dependent on the function of distance. $G(N,E)$ is a network with vertex set N and angular distance $E \in N \times N$ that is undirected and sample mean. Assume A is G's neighboring matrix. The local concentration $\delta_x$of node x, provided nodes in system $G = (N, E)$, is described in Equation (5)

$$\delta_x = \sum_{y=0}^{n} \mu(l_{xy} - l_c) \quad (5)$$

$\mu(i) = 1$ if $i \leq 0$ or $\mu(i) = 0$; l_{xy} denotes the number of hops x and y, and l_c is the cutting length. Inside the threshold length l_c. A node x's local frequency covers its surrounding nodes. To compute the difference among nodes, it uses the shortest duration. With various distance measures, the amount of l_c varies. The procedure has been proven to be responsive solely to the relative importance of δ_x. It implies that the outcomes are unaffected by the selection of l_c.

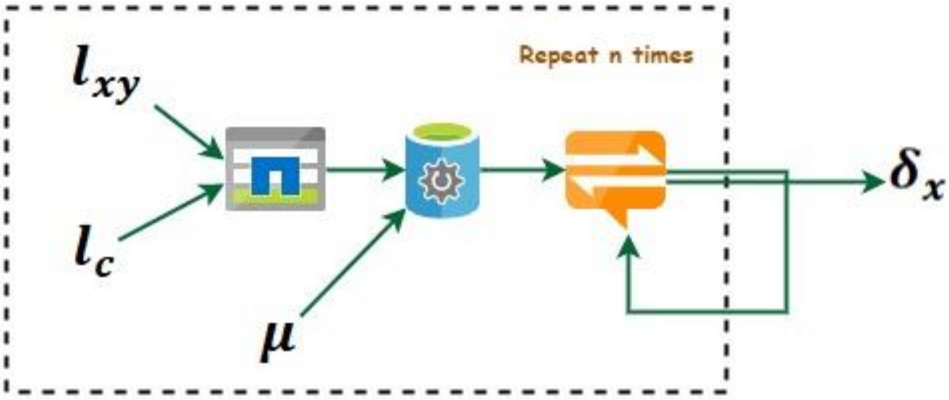

Figure 3: The pictorial depiction of the local concentration of node δ_x

The pictorial depiction of δ_x is denoted in Figure 3. It uses calculated distance, distance center point, and learning rate. The comparative length δ_x of node x is described as the smallest route among node x or any cluster with a greater density and expressed in Equation (6)

$$\delta_x = minimum\ (l_{xy}) \quad (6)$$

It usually uses $\delta_x = minimum_y\ (l_{xy})$ for the unit with the highest density distribution. It deduces that the comparative distance between a node with functional density maxima and its neighboring pixels is significantly bigger. As a result, structural centers in a system are recognized as vertices with abnormally large density distributions and distance values.

The architectural centers have a greater density than their neighbors and are separated from networks with greater density by a wide distance. The component x 's architectural significance s_{ex} is described in Equation (7)

$$sc_x = \frac{\beta_x}{\delta_x} \quad (7)$$

The density is signified δ_x, and the learning rate is represented β_x. Architectural centrality measures its local importance and the influence of greater-density hubs. As a result, this measure successfully prevents nearby vertices with functional density maxima from being classified as structural centers.

3.4 Local Community Expansion

After the architectural centers of a system were established, an extending method around certain structural centers was used to identify overlap communities. The goal of the regional expansion phase is to incrementally add neighbors to a structural center to optimize a communal showed well. As a result, a subgraph found by the optimizer is classified as a plant community. Although community development is comparable to other local strategies, many key variances exist.

Above everything, it grows communities following established structural centers, representing the number of groupings, but other techniques, such as those stated above, investigate the community using random seeding or clans. Secondly, to increase the accuracy and durability of the results, it uses a local search algorithm in the extension operation. Finally, it incorporates a weighted method into functionality to improve community closeness based on system similarities.

The architectural resemblance SS_{xy} among x and y is described in Equation (8).

$$SS_{xy} = \frac{|V(x) \cap V(y)|}{\sqrt{V(x) \times V(y)}} \tag{8}$$

Let $V(x)$ be the neighborhood of node x. Let $V(y)$ be the neighborhood of node y. The correlation is a local metric that calculates network entities' sets of common nodes. The higher the structural resemblance, the more probable it is that two nodes belong to the same group. As a result, the expansion improves the compactness of communities. This metric is used to evaluate every connection in the growing community. Given a community composition C with v_c vertices, the group's subgraph densities SD_c is given in Equation (9)

$$SD_c = \frac{\sum_{x=0}^{v_c} \sum_{y=0}^{v_c} k_{xy} \times s_{xy}}{\sum_{x=0}^{v_c} \beta_x} \tag{9}$$

β_x signifies the neighboring relation among node x and node y in K, s_{xy} represents the regional concentration of node x. The goal of the regional product increases is to establish a community composition beginning from an architectural center, such that adding a new neighbor or removing one node from the plant communities reduces the density SD_c. Furthermore, the balanced technique improves communities discovered via installations.

This section's suggested SCA-LOCD approach is designed with the structural centrality approach. The distance from the nodes is calculated from the cluster center to find the overlapping of the community as early as possible.

4. Software analysis and comparisons

This paper presents the structural Centrality approach for the local overlapping community detection (SCA-LOCD). It provides a novel approach to regional development that emphasizes the role of systems in identifying cluster centers. Studies are run on a machine with dual four 2.34 GHz CPUs, 32 GB of RAM, and Windows Server 2009 R2 as the operating system. Java 1.9 (32-bit) software platform and MySQL 5.7 database system are used in the design (32-bit). The experimental databases are well and frequently utilized in the community detection range. Real communities are collected from various areas with varying sizes and degree ranges, allowing authors to test communities' recognition methods and put them to the test in terms of resilience and adaptability.

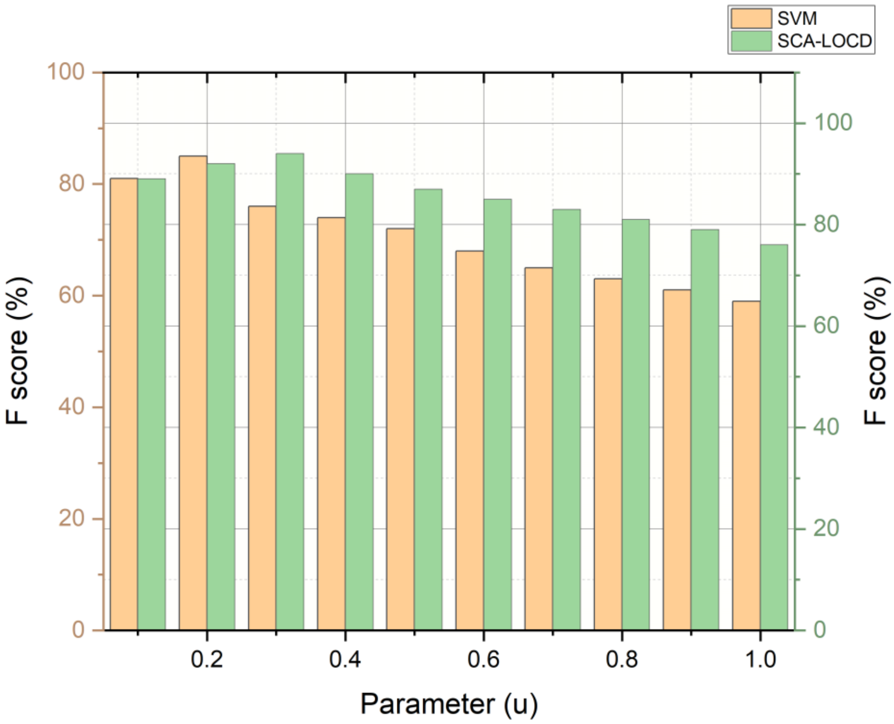

Figure 4(a): F score analysis of the suggested SCA-LOCD method

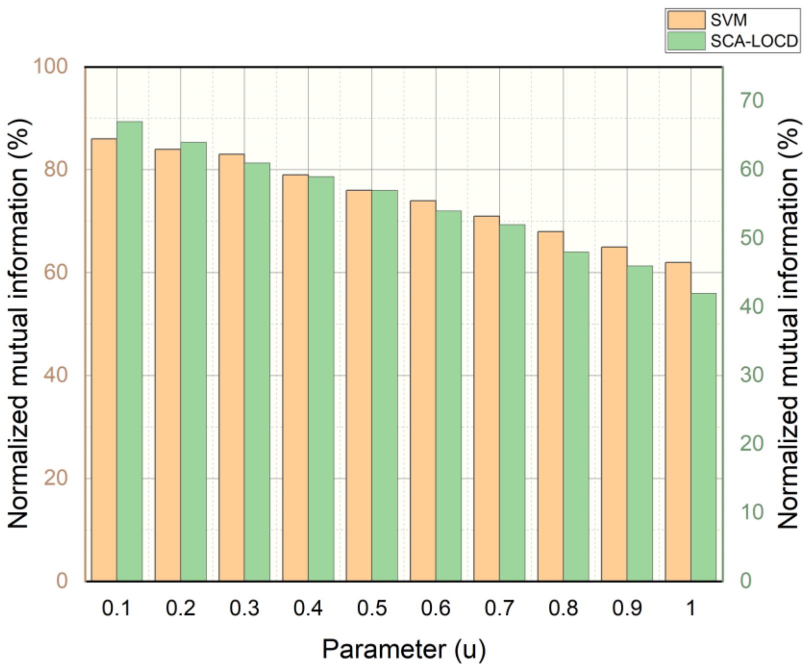

Figure 4(b): Normalized mutual information analysis of the suggested SCA-LOCD method

Figures 4(a) and 4(b) demonstrate the F score analysis and normalized mutual information (NMI) analysis of the suggested SCA-LOCD approach, respectively. The simulation analysis of the recommended SCA-LOCD approach is done under the given simulation environment. The simulation results are compared with the existing SVM model, like the F score and NMI of the suggested SCA-LOCD approach. The simulation mutation parameter varies from least to high values, and the relevant results are monitored and plotted.

Table 1: F score analysis of the suggested SCA-LOCD approach

Parameter (u)	SVM (%)	SCA-LOCD (%)
0.1	81	89
0.2	85	92
0.3	76	94
0.4	74	90
0.5	72	87
0.6	68	85
0.7	65	83
0.8	63	81
0.9	61	79
1	59	76

Table 1 displays the F score analysis of the recommended SCA-LOCD method. The simulation analysis of the recommended SCA-LOCD approach regarding the F score is analyzed, and the outcomes are compared with the existing SVM model. The simulation mutation parameter is varied from a least of 0.1 to a high of 1 with step sizes of 0.1. The suggested SCA-LOCD approach with structural centrality and machine learning model produces better simulation results. As the parameter value increases, the system weightage increases, which results in better simulation outcomes.

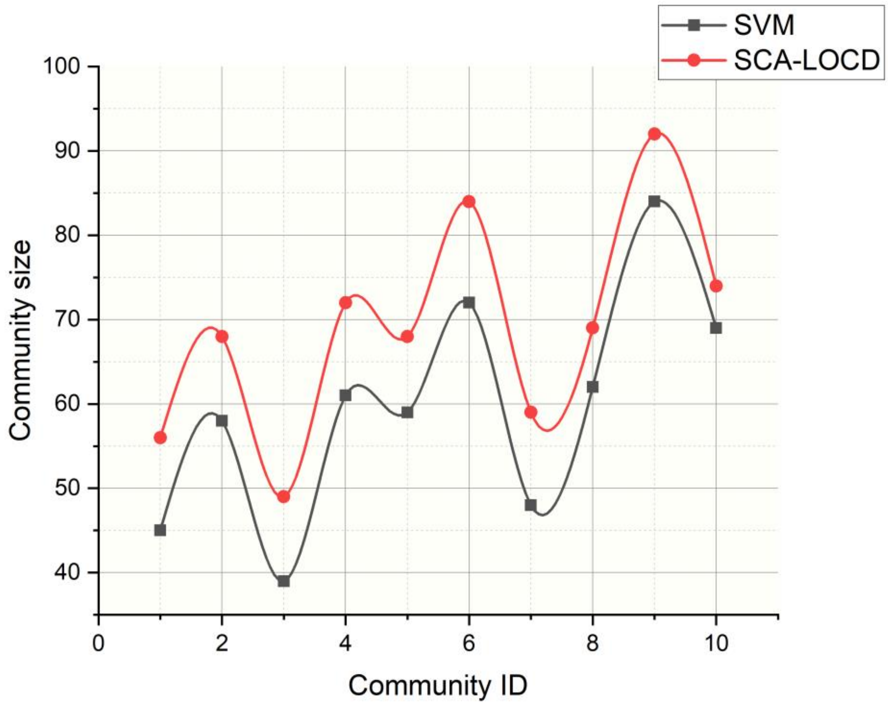

Figure 5(a): Community size analysis of the recommended SCA-LOCD method

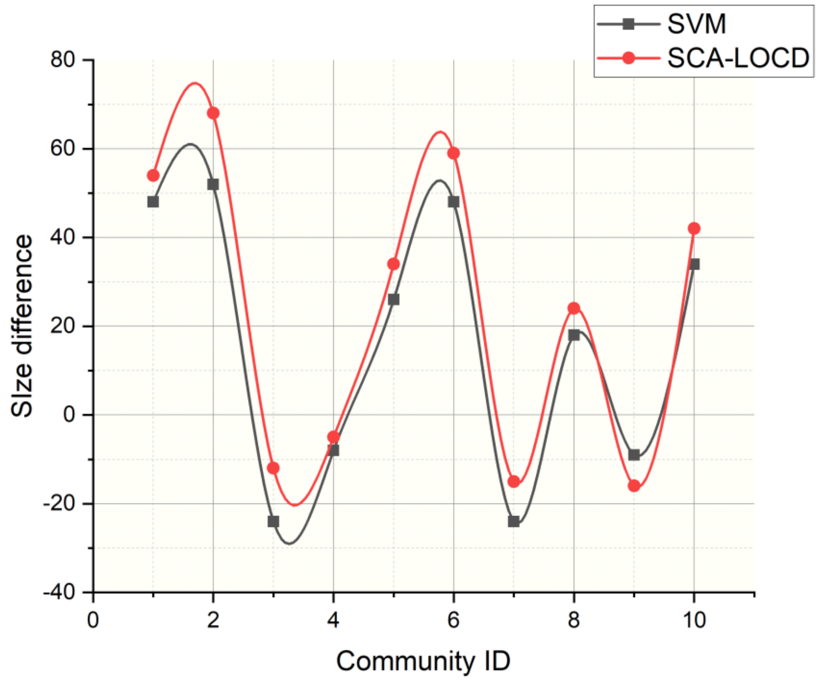

Figure 5(b): Size difference analysis of the recommended SCA-LOCD method

The community size and size difference analysis of the recommended SCA-LOCD approach is depicted in Figures 5(a) and 5(b). The recommended SCA-LOCD approach is examined by varying the community ID. There are ten communities considered for the simulation analysis. The different communities' size and size differences are analyzed under the suggested SCA-LOCD approach and the existing SVM model. The simulation findings indicate that the suggested SCA-LOCD approach has higher simulation results than the structural centrality model.

Table 2: Community size analysis of the suggested SCA-LOCD approach

Community ID	SVM	SCA-LOCD
1	45	56
2	58	68
3	39	49
4	61	72
5	59	68
6	72	84
7	48	59
8	62	69
9	84	92
10	69	74

The community size analysis of the recommended SCA-LOCD approach is represented in Table 2. The simulation analysis of the suggested SCA-LOCD method is examined under different communities. The respective simulation results regarding community size are analyzed using existing and suggested SCA-LOCD approaches. The suggested SCA-LOCD approach with structural centrality and position analysis model exhibits higher simulation results than the existing model. The larger community size leads to higher prediction results in the suggested SCA-LOCD approach.

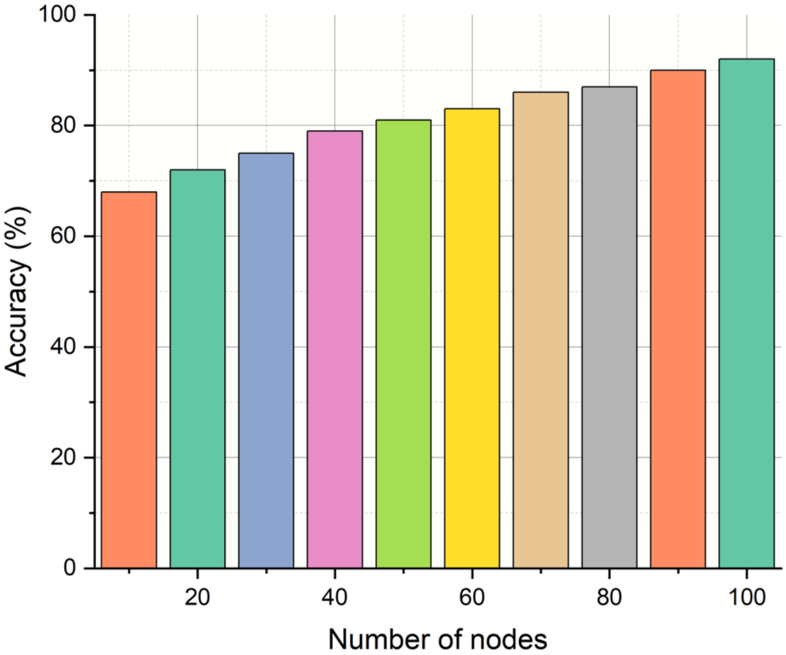

Figure 6(a): Accuracy analysis of the suggested SCA-LOCD method

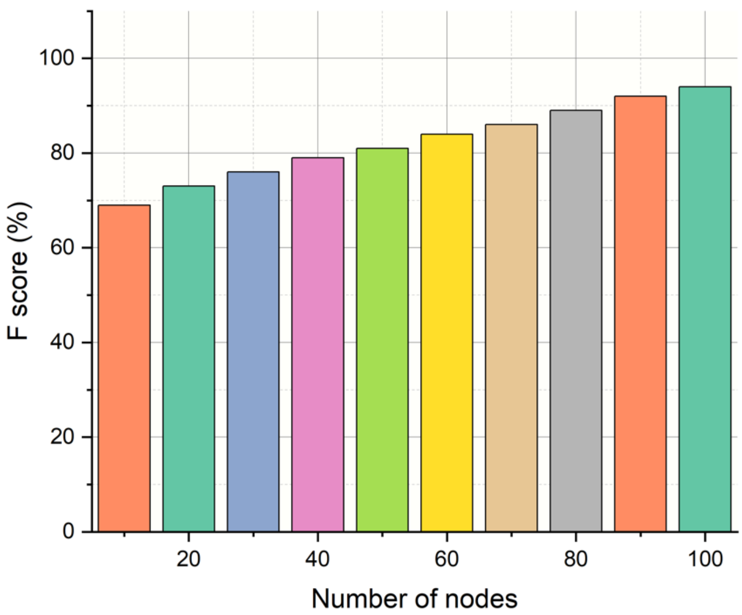

Figure 6(b): F score analysis of the suggested SCA-LOCD method

Figures 6(a) and 6(b) display the accuracy and F score analysis of the recommended SCA-LOCD method. The simulation analysis of the recommended SCA-LOCD approach is done by changing the number of nodes in a community from a least 10 nodes to an extreme of 100 nodes with step sizes of 10 nodes. The higher number of nodes helps to better connectivity and enhances the overall detection accuracy. As the number of nodes rises, the relevant simulation outcomes of the recommended SCA-LOCD approach improve.

The recommended SCA-LOCD approach is designed, examined, assessed, and compared with the existing models. The simulation findings demonstrate that the suggested SCA-LOCD approach with the structural centrality model exhibits higher simulation outcomes.

5. Conclusion and limitations

This paper presents the structural Centrality approach for the local overlapping community detection (SCA-LOCD). It provides a novel approach to regional development that emphasizes the role of systems in identifying cluster centers. It proposed a local expansion process based on system centers to identify overlapped communities efficiently. Structural centers are networks with a greater concentration than their neighbors and are separated from vertices with greater density by a relatively wide distance. Architectural centers are often evenly distributed throughout communities, meaning societies are calculated intuitively. As a result, it established the concept of architectural centrality and a finding approach for locating architectural centers in systems. On this foundation, the suggested SCA-LOCD approach is described as a locally enabling environment for discovering naturally overlapped communities in complicated networks. In experiments, the suggested SCA-LOCD approach demonstrates benefits divided into two categories. First, establishing structural centers establishes the number of communities for a system. This methodology is used to help other approaches, particularly grouping methods that require the clustering results to be provided. Secondly, unlike previous local approaches, the growth model around architectural centers avoids choosing seeds at whim and physically tweaking variables, resulting in a faster resolution to the optimum solution and a more consistent method. Furthermore, community detection techniques handle various issues, including topic recognition, picture grouping, and viral transmission. The suggested SCA-LOCD approach is limited to recognizing artificial community fluctuations, which are affected by community detection algorithms. Therefore population detection's implementation potential is considered in future studies.

References

[1] Bello-Orgaz, G., Salcedo-Sanz, S., & Camacho, D. (2018). A multiobjective genetic algorithm for overlapping community detection based on edge encoding. *Information Sciences*, *462*, 290-314.

[2] Shakeel, P. M., Baskar, S., Sampath, R., & Jaber, M. M. (2019). Echocardiography image segmentation using feed-forward artificial neural network (FFANN) with fuzzy multi-scale edge detection (FMED). International Journal of Signal and Imaging Systems Engineering, 11(5), 270-278.

[3] Baskar, S., Dhulipala, V. S., Shakeel, P. M., Sridhar, K. P., & Kumar, R. (2019). Hybrid fuzzy based spearman rank correlation for cranial nerve palsy detection in MIoT environment. Health and Technology, 1-12.

[4] Al-Yousif, S., Nabeel, A., Ibrahim, W.K., Jaber, M.M., Ali, M.H., Jaber, M., Hameed, A.S., Al-Khayyat, A.H., Omer, A.F., Mustafa, N., Jabbar, K.A., and Abbood, A.A.A., 2023. Intelligent Multilevel Fusion System for Wireless Sensor Network Virtualization Using Deep Reinforcement Learning in Education. *Fusion: Practice and Applications*, 10(1), pp.116–127.

[5] Jaber, M.M., Hassan, M., Sura, A., Abd, K., Mohammed, M., Alkhayyat, A., Jassim, M., Rashid, A., and Lahib, A., 2022. Q-learning based task scheduling and energy-saving MAC protocol for wireless sensor networkss. *Wireless Networks*, 6.

[6] Billah, M. F. R. M., Saoda, N., Gao, J., & Campbell, B. (2021, May). BLE Can See: A Reinforcement Learning Approach for RF-based Indoor Occupancy Detection. In Proceedings of the 20th International Conference on Information Processing in Sensor Networks (co-located with CPS-IoT Week 2021) (pp. 132-147).

[7] Gao, J., Wang, H., & Shen, H. (2020, August). Machine learning based workload prediction in cloud computing. In 2020 29th international conference on computer communications and networks (ICCCN) (pp. 1-9). IEEE.

[8] Ngo, T. D., Bui, T. T., Pham, T. M., Thai, H. T., Nguyen, G. L., & Nguyen, T. N. (2021). Image deconvolution for optical small satellite with deep learning and real-time GPU acceleration. Journal of Real-Time Image Processing, 1-14..

[9] Ngo, T. D., Bui, T. T., Pham, T. M., Thai, H. T., Nguyen, G. L., & Nguyen, T. N. (2021). Image deconvolution for optical small satellite with deep learning and real-time GPU acceleration. Journal of Real-Time Image Processing, 1-14.

[10] Ali, M.H., Jaber, M.M., Abd, S.K., Alkhayyat, A., and Jameel, H.A., 2022. Model for wireless image correlation assisted by sensors based on 3D display technology. Optik, 268.

[11] Yazdanparast, S., Havens, T. C., & Jamalabdollahi, M. (2020). Soft overlapping community detection in large-scale networks via fast fuzzy modularity maximization. *IEEE Transactions on Fuzzy Systems*, *29*(6), 1533-1543.

[12] Chen, J., Gong, Z., Mo, J., Wang, W., Wang, C., Dong, X., ... & Wu, K. (2021). Self-Training Enhanced: Network Embedding and Overlapping Community Detection With Adversarial Learning. *IEEE Transactions on Neural Networks and Learning Systems*.
[13] Wang, Z., Sun, C., Rui, X., Philip, S. Y., & Sun, L. (2021). Localization of multiple diffusion sources based on overlapping community detection. *Knowledge-Based Systems*, *226*, 106613.
[14] He, C., Liu, H., Tang, Y., Liu, S., Fei, X., Cheng, Q., & Li, H. (2021). Similarity preserving overlapping community detection in signed networks. *Future Generation Computer Systems*, *116*, 275-290.
[15] Ramesh, A. C., & Srivatsun, G. (2021). Evolutionary algorithm for overlapping community detection using a merged maximal cliques representation scheme. *Applied Soft Computing*, *112*, 107746.
[16] Jiang, W., Pan, S., Lu, C., Zhao, Z., Lin, S., Xiong, M., & He, Z. (2021). Label entropy-based cooperative particle swarm optimization algorithm for dynamic overlapping community detection in complex networks. *International Journal of Intelligent Systems*.
[17] Yu, Q., Yu, Z., Wang, Z., Wang, X., & Wang, Y. (2020). Estimating posterior inference quality of the relational infinite latent feature model for overlapping community detection. *Frontiers of Computer Science*, *14*(6), 1-15.
[18] Chai, Z., & Liang, S. (2020). A node-priority-based large-scale overlapping community detection using multiobjective evolutionary optimization. *Evolutionary Intelligence*, *13*(1), 59-68.
[19] Sheng, J., Wang, K., Sun, Z., Wang, B., Khawaja, F., Lu, B., & Zhang, J. (2019). Overlapping community detection via preferential learning model. *Physica A: Statistical Mechanics and its Applications*, *527*, 121265.
[20] Cheng, J., Chen, M., Zhou, M., Gao, S., Liu, C., & Liu, C. (2018). Overlapping community change-point detection in an evolving network. *IEEE Transactions on Big Data*, *6*(1), 189-200.
[21] Mahabadi, A., & Hosseini, M. (2021). SLPA-based parallel overlapping community detection approach in large complex social networks. *Multimedia Tools and Applications*, *80*(5), 6567-6598.
[22] Fatima, S., & Badugu, S. (2020). A study on overlapping community detection for the multimedia social network. *Advances in Decision Sciences, Image Processing, Security and Computer Vision*, 572-578.
[23] Zhang, H., Niu, X., King, I., & Lyu, M. R. (2018). Overlapping community detection with preference and locality information: a non-negative matrix factorization approach. *Social Network Analysis and Mining*, *8*(1), 1-14.
[24] Zhang, Y., Zhang, Y., Chen, Q., Ai, Z., & Gong, Z. (2018). True-link clustering through signaling process and subcommunity merge in overlapping community detection. *Neural Computing and Applications*, *30*(12), 3613-3621.
[25] Meenakshi, N., Jaber, M.M., Pradhan, R., Kamruzzaman, M.M., Maragatham, T., Ramamoorthi, J.S., and Murugesan, M., 2023. Design systematic wireless inventory trackers with prolonged lifetime and low energy consumption in future 6G network. *Wireless Networks*.
[26] Obeid, N. (2023). On The Product and Ratio of Pareto and Erlang Random Variables. International Journal of Mathematics, Statistics, and Computer Science, 1, 33–47. https://doi.org/10.59543/ijmscs.v1i.7737
[27] Saeed Kolahi-Randji, S., Nejad Attari, M.Y. & Ala. A. (2023). Enhancement the Performance of Multi-Level and Multi-Commodity in Supply Chain: A Simulation Approach. Journal of Soft Computing and Decision Analytics, 1(1), 18-38. https://doi.org/10.31181/jscda1120232