IAES International Journal of Artificial Intelligence (IJ-AI)
Vol. 15, No. 1, February 2026, pp. 919~928
ISSN: 2252-8938, DOI: 10.11591/ijai.v15.i1.pp919-928

Evaluating the detected communities using traditional algorithms on keyword co-occurrence networks

Kiruthika R.[1], Krishnaveni Sakkarapani[2]
[1]Department of Computer Science, PSGR Krishnammal College for Women, Coimbatore, India
[2]Department of Data Analytics (PG), PSGR Krishnammal College for Women, Coimbatore, India

Article Info

Article history:

Received Feb 2, 2025
Revised Nov 6, 2025
Accepted Dec 15, 2025

Keywords:

Community detection
Complex network
Keyword co-occurrence
Network analysis
Scopus bibliometric data

ABSTRACT

Community detection is one of the most significant research areas in network analysis, which helps to understand the internal structure of large networks. This work utilizes the traditional community detection methods on a keyword co-occurrence graph derived from the Scopus bibliographic database. This research article primarily focused on the index keywords of deep learning-driven publications obtained from three major network Scopus bibliometric datasets (SBD), namely SBD_1 as 2006-2013, SBD_2 as 2014-2016, and SBD_3 as 2017. For this proposed model framework, the existing traditional algorithms, including Louvain, greedy modularity optimization (GMO), Leiden, Infomap, speaker-listener label propagation algorithm (SLPA), Walktrap, SpinGlass, K-Clique, and Clauset, Newman and Moore (CNM) methods are applied to detect communities from the network and carried out through Python. Comparisons among these algorithms, Leiden, SpinGlass, and Louvain are considered as better algorithms for our work based on the detected communities, modularity score and other metrics to evaluate the performance of detected communities from the network. This research proposes an ideology for the selection process of algorithms that depends on different factors like network characteristics, network structure, dataset size, and computational efficiency. This analysis suggests a unique perspective on the effectiveness of each method in the Scopus bibliometric network and its potential to enhance research topic exploration.

Corresponding Author:

Kiruthika R.
Department of Computer Science, PSGR Krishnammal College for Women
Coimbatore, India
Email: kirthikamole@gmail.com

1. INTRODUCTION

The process of detecting the group of nodes in the network is based on their interconnection, referred to as community detection [1]. Scopus is a significant bibliographic database that distributes metadata of an article, which is essential for consistent research activities [2]. The growth in academic publications demands an effective technique for capturing and categorizing the information efficiently. The majority of the literature on community detection methods has concentrated on author collaboration or citation networks and the use of keyword-based co-occurrence networks has been relatively under-researched. This is due to that most of the benchmark datasets like Cora, Citeseer, and PubMed are based on citation networks. This creates scope for exploring keyword-based co-occurrence networks to uncover thematic relationships beyond citation perspectives. There is a lack of comparative analysis among various traditional algorithms to analyze bibliometric keywords networks, particularly when considering Scopus datasets on deep learning research. These research gaps establish the necessity to consider the efficiency and appropriateness of various traditional community detection methods for detecting meaningful communities of keywords that can reflect emerging research topics. From this, the main purpose of this work is to examine keyword co-occurrence networks through various conventional community detection algorithms and evaluate their performance in terms of modularity and other structural measures. Keyword analysis [3] is very important for identifying connections and patterns across scientific papers. Co-occurrence networks are the frequency of their co-occurrence facilitates researchers to discover and understand the complex connections between terms across several domains [4], [5].

This work evaluates the Scopus bibliometric datasets (SBD) [6] for keyword collection and traditional community identification methods to find keyword-based communities. The first phase is to build a keyword co-occurrence network [7], in which nodes are the indexed keywords and the edges are the connections between articles sharing the keywords. The network analysis phase provides community structures [8], which are the foundation for identifying research communities. This phase helps to gain more insights into related concepts using the best detection algorithms. The next phase is to examine better community discovery algorithms [9] based on the characteristics of graphs analysed from the network. For this proposed model, the existing traditional community detection methods include the Louvain method, greedy modularity optimization method (GMO), Leiden algorithm, Infomap, speaker-listener label propagation algorithm (SLPA), Clauset, Newman, and Moore (CNM), Walktrap algorithm, SpinGlass, and K-Clique methods are applied to detect communities [10]–[19]. The quality of the detected communities is measured through the modularity, conductance and partition density. This framework enables identifying the relationships between different keywords as well as determining different patterns and clusters in the given dataset. The main aim of this work is to evaluate the community discovery algorithms [20] using keyword co-occurrence networks generated from academic datasets and implemented using Python programming language-based methodology [21].

From the methodology and outcome of this article, understanding of recognizing keyword communities from scholarly collections using basic community detection techniques. This will provide the effectiveness for identifying significant communities and analysing their interactions among keywords. This research work suggests the necessity of effectively choosing algorithm [22] to get an efficient outcome of the detected community may vary based on the community structure and distribution of nodes within communities.

2. RELATED WORKS

Several studies have focused on community detection and network analysis, which is an extremely important topic in bibliometric and scientometric research. Keyword co-occurrence networks [23] are essential for discovering term connections, where traditional community identification methods find associated keyword groups, presenting the structure of data and patterns. This includes determining clusters of connected terms [24] inside a network using numerous community detection approaches. Efficient preprocessing, feature extraction and embeddings all improve accuracy for this research area. Understanding keyword communities supports mapping discipline evolution, discovering interdisciplinary linkages and identifying prominent themes in datasets.

A lot of work has been done to compare community detection algorithms [25] in terms of their performance in finding communities in the networks. Some related work discussed the co-occurrence relations that enhanced modularity and computational complexity. These studies evaluated several approaches, such as label propagation [26], [27], WalkTrap and infomap [8] on a large number of artificial and real-world datasets.

Lozano *et al.* [7] analyzed the frequency with which 39 keywords appeared in data envelopment analysis (DEA) literature between 2008 and 2017 and established that sustainability-related themes had begun to appear in DEA literature. The network showed a small-world topology with a power law exponent and the disassortativity increased from 0.102 to 0.157 and the average path length was constant at 6.50. Considering challenges with keyword standards and temporal scope, this study suggests that additional research on keyword maturity trends and link prediction is possible.

The related works highlight the strengths and limitations of different algorithms. Community detection on a keyword co-occurrence is an emerging field that contains a building collection of topics, models and approaches. Keyword research in this sector can help to understand the network structure, interpret co-occurrences and evaluate complicated networks.

3. METHOD

The methodology is proposed for analyzing and detecting communities in keyword co-occurrence networks from the SBD. The overall phases for this work are data extraction, data preprocessing, graph construction, network analysis measures, and traditional community detection algorithms. The reason for choosing traditional algorithms is based on their ability to efficiently detect well-defined communities. First, the process begins with data acquisition for deep learning-based keywords as identification and then removing any irrelevant data to create an acceptance co-occurrence network, with keywords forming the nodes of edges being co-occurrence frequencies. Detailed explanation of data preprocessing, graph construction and network analysis measures is explained in the next section. For this proposed model framework, the existing traditional community detection methods, including Louvain, GMO, Leiden, Infomap, SLPA, CNM, Walktrap, SpinGlass, and K-Clique methods, are applied to detect communities. The performance evaluation metrics used in work are modularity, conductance, internal and partition density. The primary goal of this work is to detect communities using traditional discovery methods on the keyword co-occurrence networks derived from academic datasets through Python.

3.1. Datasets description and data collection

Scopus is one of the bibliographic scientific data sources that retrieves the different fields as bibliographic data [28]. These fields include authors, DOI, year, source title, volume, abstract, author keywords, index keywords, references, and numerous other instances of metadata fields [29]. But this work mainly focuses on index keywords of the deep learning-based articles retrieved as three primary datasets: SBD_1 as 2006-2013, SBD_2 as 2014-2016, and SBD_3 as 2017.

3.2. Data pre-processing

Data preprocessing is a significant phase in network analysis that involves the process of discovering and fixing incorrect entries in a dataset. The data processing [30] was utilized for detecting communities in Scopus bibliographic keyword co-occurrence networks. First, the raw dataset contains all bibliographic data, but for this work, the indexed keyword field was retrieved. These keywords are cleaned to remove replications, insufficient records and redundant items. The indexed keywords get standardized by transforming them to lowercase, eliminating punctuation and stop words are removed. A co-occurrence adjacency matrix is then created to capture the associations between keywords based on their existence in documents. Self-loops are removed to retain meaningful network structure by avoiding the keyword connecting to itself to form an edge.

3.3. Graph construction

Graph construction is one of the significant components for developing the community detection model. A Scopus bibliographic keyword co-occurrence adjacency matrix is converted into a graph by computing the weight of each edge connecting keywords, with the matrix value representing the frequency of co-occurrence. The basic structural metrics of the graph consist of nodes and edges. The representation of keywords as nodes and the frequency of the co-occurrence of keywords between the articles are considered as weighted edges to form an effective keyword co-occurrence network. The final keywords are converted into node numbers to construct the simplified and optimistic form of the graph. Figure 1 presents the graphical representation of the three distinct Scopus bibliographic keyword co-occurrence networks constructed after data preprocessing. Figure 1 are three different and separately constructed networks, which are the final results of preprocessing and which are analyzed separately in this article. Figure 1(a) shows the SBD_1 dataset is a set of deep learning-based indexed keywords between 2006 and 2013 in the form keyword co-occurrence network. Figure 1(b) shows the SBD_2 data, which includes indexed keywords between 2014 and 2016, and these are a medium-sized keyword co-occurrence network. Figure 1(c) shows the SBD_3 dataset, which comprises indexed keywords of 2017 and these are constructed for analyzing the larger data as a massive dense simple complex keyword co-occurrence network.

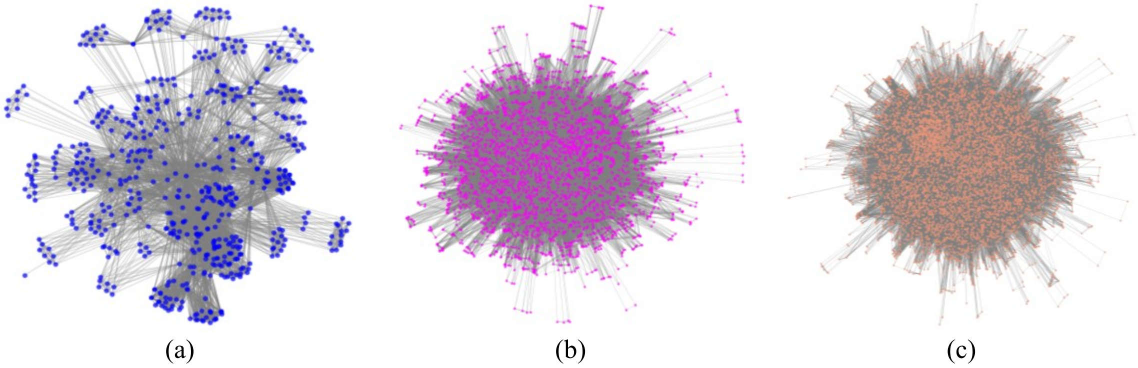

Figure 1. Graphical representation of SBD networks of (a) SBD_1 as 2006-2013, (b) SBD_2 as 2014-2016, and (c) SBD_3 as2017

3.4. Network analysis

Network analysis is essential for community detection as it uncovers key patterns and structures within data. For detecting communities, network analysis needs to ensure the properties and structure of the network. So that it will not reflect and affect the effectiveness of the algorithm. The importance of network analysis depends on its ability to optimize community detection by selecting the most appropriate algorithm based on considering these network characteristics. The characteristics of the network for SBD_1, SBD_2, and SBD_3 are presented in Table 1.

This analysis identifies groups of nodes that are more strongly connected than to those outside the group. These visualizations enhance the exactness of community detection and provide valuable understanding in areas such as social networks, collaboration and information exchange. Density gives the measure of how connected the graph is in terms of percentage. Average degree gives the measure of how many nodes, on average, are connected in the graph. If the graph is weighted, the average link weight is the measure of the average strength of the edges. The average clustering coefficient gives the probability of the nodes in the network clustering and the average path length is the average distance between any two nodes in the connected graphs. The diameter of a graph is the greatest distance between two nodes.

Table 1. Network characteristics of SBD

Network characteristics	Node	Edge	Density (%)	Average degree	Average link weight	Average clustering coefficient	Edge density	Average path length
SBD_1	599	8085	0.451	26.99	1.242	0.9119	0.0485	1.95
SBD_2	3934	65587	0.85	33.34	1.37	0.88	0.0085	2.01
SBD_3	8202	155975	0.46	38.03	1.42	0.8614	0.0046	2.0136

3.5. Community detection

Community detection is used to identify the groups within a network through various algorithms. These community detection algorithms are important in network analysis due to their contribution to uncovering relationships among structural key interconnections. This gives a better understanding of how information flows throughout a network. The Scopus bibliometric networks were improved by using various algorithms which include Louvain, GMO, Leiden, Infomap, SLPA, Walktrap, SpinGlass, K-Clique, and CNM algorithms. The Louvain method functions in a way that merges small communities over and over again through modularity optimization. This can discover the hierarchical community structures in the large networks and also dynamic in nature. GMO works by initializing every node as belonging to exactly one community and finding new partnerships between the communities to optimize modularity. The Leiden algorithm is an advancement of the Louvain algorithm. But both methods are almost similar ways that enhance community splitting for better modularity and for coping with large networks. Infomap identifies the most efficient division of a given network based on the description length of random walks. The SLPA identifies communities by allowing nodes to share and adopt labels representing their membership in specific network clusters based on interactions with neighbouring nodes. The SpinGlass algorithm identifies communities by modelling the network and optimizing modularity to group nodes into clusters. The K-Clique method identifies communities by locating cliques or groups of nodes well-connected in the graph. The CNM algorithm is a greedy approach that searches for communities by trying to maximize modularity and it is sensitive to hierarchical structures of the network.

4. EXPERIMENTAL RESULTS AND DISCUSSION

The results are obtained and analyzed with the help of traditional community detection algorithms on the keyword co-occurrence networks built from the Scopus bibliographic network dataset. This research aims to find and analyses the network communities using indexed terms extracted from different deep learning publications. The experiments compare the results between the applied traditional community detection approaches like Louvain, SLPA, CNM, Infomap, Walktrap, and the Spin-glass model. This analysis shows the values of modularity, community size distribution to determine the presence of meaningful communities.

4.1. Evaluating the performance of detected communities

Community detection is a process used to discover patterns in a network based on the nodes that share similar density are grouped into communities. Nine existing community detection algorithms are chosen and applied to our proposed model based on their high importance in explicit implementations. This research strongly indicates that the importance of selecting the method for community detection depends on the outcomes, such as the number of detected communities and node distribution within each community. The comparison of the number of detected communities for 2006-2013, 2014-2016, and 2017 as SBD_1, SBD_2, and SBD_3 among different algorithms is tabulated in Table 2. The comparison among numerous algorithms is visualized in Figure 2.

The findings of research highlight that Leiden, SpinGlass, and Louvain as the best methods for considering the performance of discovered communities on datasets of different sizes. The Leiden algorithm detects communities with strong connectivity that show excellent results in small datasets. Its ability to handle large networks proves to be its greatest strength. SpinGlass achieves balanced community structure in small networks, along with medium input requirements for precise partitioning and it efficiently detects highly organized communities in large networks. Louvain divides small datasets into coherent segments while supporting consistent results for medium-sized datasets and delivering reliable results for large datasets.

Table 2. Number of detected communities for SBD_1, SBD_2, and SBD_3

Method	CNM	GMO	Infomap	K-Clique	Leiden	Louvain	SLPA	SpinGlass	Walktrap
SBD_1	10	12	29	14	11	12	22	14	17
SBD_2	51	43	211	-	19	29	49	25	146
SBD_3	74	75	365	-	15	25	21	24	263

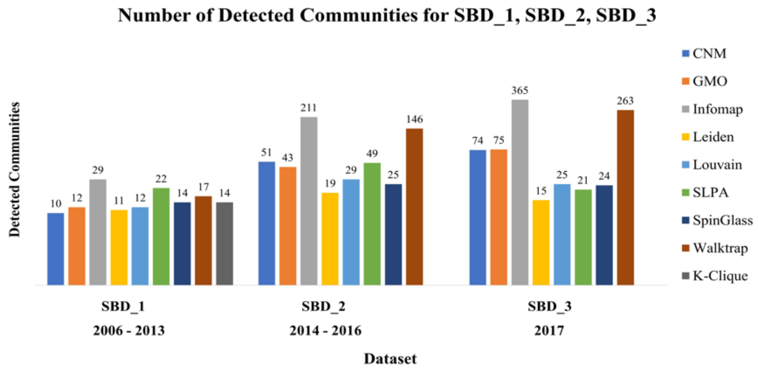

Figure 2. Comparison of detected communities on existing algorithms

4.2. Evaluation metrics

Evaluation metrics for these measures evaluate the overall performance and characteristics of discovered communities, working on aspects of community size, modularity, conductance and density. These metrics will evaluate the effectiveness of the community detection method and the network partitioning structure. Detected communities are the groups of nodes that are produced through the community detection methods, in which nodes of a given community are more closely connected than to nodes of other communities. Modularity, conductance, internal and partition density are the main performance evaluation metrics used in the work.

4.3. Modularity and conductance-based performance evaluation

Modularity (M) provides the measure of how well the divisions are set within communities. Conductance (C) measures partition quality by comparing inter and intra-community edges and if low conductance indicates densely connected nodes. For SBD_1, Leiden has the highest modularity of 0.6402 and the lowest conductance of 0.2246, which makes it efficient in identifying well-connected and coherent communities. SpinGlass got a modularity of 0.6372 and conductance of 0.2916, thus making it suitable for balanced community structures. Walktrap delivers good results in terms of M of 0.6135 and C of 0.2282. Infomap obtains an M of 0.5928 and C of 0.3393. GMO presents a fairly balanced result with M of 0.5864 and C of 0.2459. Louvain has a modularity of 0.6181 with a conductance of 0.2219. The CNM method gives an M of 0.5953, but the C of 0.8884 shows that the algorithm is not efficient in separating clear communities. SLPA offers a M of 0.5765 and C of 0.3245. K-Clique gives the lowest modularity at 0.1648 and high conductance at 0.7396, and is not scalable.

For SBD_2, strong modularity of 0.4203 and conductance of 0.4557 are attained in Leiden, which maintains the efficiency of the method in balanced community detection. SpinGlass has the M of 0.431 and C of 0.5132, which offers balanced performance. For Walktrap, it performs well with M being 0.3666 and C being 0.482. The Infomap has an M of 0.3302 and a C of 0.563. GMO proves to be consistent with M as 0.3425 and C 0.4503. Louvain presents modularity at 0.3938 and conductance at 0.3774. CNM has a modularity of 0.3794 and a conductance of 0.9736. SLPA achieves moderate performance with M of 0.2836 and C of 0.4526.

For SBD_3, Leiden got a modularity of 0.3991 and conductance 0.4944, which makes it suitable for large networks. SpinGlass attains modularity at 0.3981 and conductance at 0.5451. Walktrap has an M of 0.3142 and a C of 0.5356. Infomap has M of 0.2669 and C of 0.6191, which is efficient. GMO got M with 0.3215 and C 0.5155. Louvain gives a reliable M of 0.352 and C of 0.4283. CNM with M at 0.3573 and C at 0.9786 does not seem to perform very well in terms of identifying well-defined communities. SLPA yields a low performance in terms of M, which is 0.0283 and C, which is 0.3528. K-Clique does not perform effectively and does not get any results for the SBD_2 and SBD_3 datasets due to the computational complexity. The comparison among the evaluated results is tabulated in Table 3 and presented in Figure 3. The parameters of evaluation are indicated on the X-axis as metrics and the Y-axis as score to show the parameters and their value in Figure 3, respectively.

Therefore, based on the analysis above, it can be concluded that the Leiden algorithm outperforms all the other algorithms in large datasets with the highest modularity and low conductance, and for all the years. SpinGlass has good performance with appropriate modularity and conductance. Louvain performs stably across datasets with a reasonable level of modularity and community cohesiveness. Walktrap is appropriate for a large number of edges, but its performance decreases gradually. GMO and CNM are effective when a set of objects is compact, but their effectiveness significantly decreases when the set is large. Infomap and Walktrap perform best for the smaller-scale hierarchical network, but are not suitable for more complex data. CNM shows limitations due to high conductance, while K-Clique algorithms struggle with the effectiveness, scalability and performance across the complex datasets.

Table 3. M and C-based performance evaluation

Method	SBD_1		SBD_2		SBD_3	
	M	C	M	C	M	C
CNM	0.5953	0.8884	0.3794	0.9736	0.3573	0.9786
GMO	0.5864	0.2459	0.3425	0.4503	0.3215	0.5155
Infomap	0.5928	0.3393	0.3302	0.5630	0.2669	0.6191
K-Clique	0.1648	0.7396	-	-	-	-
Leiden	**0.6402**	0.2246	**0.4203**	0.4557	**0.3991**	0.4944
Louvain	0.6181	**0.2219**	0.3938	**0.3774**	0.3520	**0.4283**
SLPA	0.5765	0.3245	0.2836	0.4526	0.0283	**0.3528**
SpinGlass	0.6372	0.2916	**0.4310**	0.5132	0.3981	0.5451
Walktrap	0.6135	0.2282	0.3666	0.4820	0.3142	0.5356

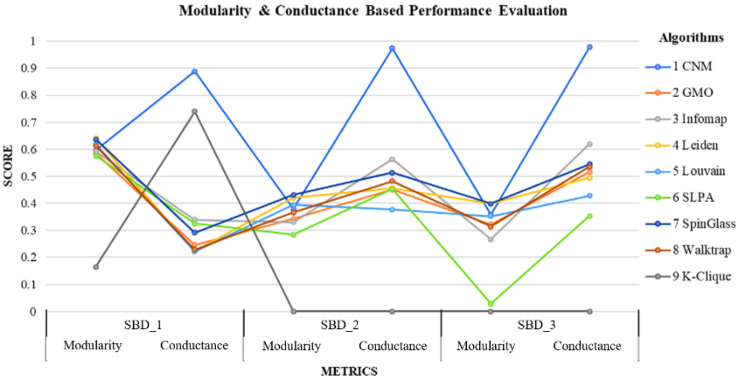

Figure 3. Performance evaluation comparison of modularity and conductance visualization

4.4. Internal and partition density-based performance evaluation

Internal density measures the degree of connectedness within communities, while partition density quantifies the separation of communities. If the smaller community size naturally indicates a higher internal density value. If the connecting edges are divided into communities means then the partition density values may decrease. For SBD_1, K-Clique has the best internal density of 0.8956 and partition density of 0.747, demonstrating that this method can form highly dense and well-separated communities. SLPA has a high internal density of 0.8312 and partition density of 0.6851. Infomap follows with an internal density of 0.7947 and a partition density of 0.6446. GMO got a reasonable performance with an internal density of 0.6375 and partition density of 0.5214. Walktrap also works well, as it has an internal density of 0.7078, and the partition density is 0.5870. Leiden has a moderate internal density of 0.4864 and partition density of 0.434, which are ideal for good community detection. Louvain gives a similar performance with an internal density of 0.3210 and partition density of 0.5847 and it is used for identifying bigger, relatively fewer compact groups. The internal density of SpinGlass is relatively low at 0.5768, and the partition density represents 0.5002, meaning that it is well suited for small to large networks with lesser inter-cluster compactness. CNM is not efficient for community detection in this case, as it has a low internal density of 0.0562 and a partition density of -0.0033. The negative value representation in partition density indicates less separation between the communities, which means that the algorithm faces a great challenge in identifying and partitioning the network into different, distinct communities.

For SBD_2, Walktrap outperforms the other algorithms with the highest internal of 0.8904 and partition density of 0.6230. SLPA with an internal density of 0.7495 and partition of 0.5827. Infomap has an internal of 0.6598 and a partition density of 0.4759. GMO has moderate results with an internal density of 0.7655 and a partition of 0.5440. Leiden and Louvain present moderate results, which are reasonable in the general network partitions. Leiden got an internal density of 0.2368 and a partition of 0.1917. Louvain got an internal density of 0.5955 and a partition density of 0.4414. CNM is low again with internal density 0.8764 and partition density -0.1798. The outcome of an internal density and partition density using K-Clique is not mentioned in the table. The reason is that K-Clique does not perform well and generate any result for the SBD_2 and SBD_3 datasets due to the high computation time. SpinGlass got an internal and partition density of 0.3478, 0.2668, respectively.

For SBD_3, the results indicate that Walktrap still outperforms the other methods by achieving the highest internal density of 0.8974 and partition density of 0.6197. SLPA gets an internal density of 0.8506 and a partition density of 0.6686. Infomap demonstrates an internal density of 0.5573 and a partition density of 0.3949 for complex networks. GMO yields reliable results with an internal density of 0.7720 and partition density of 0.5349. SpinGlass offers lower results with the internal density of 0.2499 and partition density of 0.1432, which can be interpreted as the lower efficiency of the algorithm in distinguishing communities. Leiden and Louvain give moderate results for internal and partition densities. Leiden gives internal density as 0.1206 and partition density as 0.0911, whereas Louvain gives internal density as 0.5145 and partition density as 0.3361. The comparison of the evaluated outcomes for internal and partition density is presented in Table 4 and illustrated in Figure 4.

Table 4. Internal and partition density-based performance evaluation

Method	SBD_1		SBD_2		SBD_3	
	Internal	Partition	Internal	Partition	Internal	Partition
CNM	0.0562	-0.0033	0.8764	-0.1798	0.0479	-0.2151
GMO	0.6375	0.5214	0.7655	0.544	0.772	0.5349
Infomap	0.7947	0.6446	0.6598	0.4759	0.5573	0.3949
K-Clique	0.8956	0.747	-	-	-	-
Leiden	0.4864	0.434	0.2368	0.1917	0.1206	0.0911
Louvain	0.321	0.5847	0.5955	0.4414	0.5145	0.3361
SLPA	0.8312	0.6851	0.7495	0.5827	0.8506	0.6686
SpinGlass	0.5768	0.5002	0.3478	0.2668	0.2499	0.1432
Walktrap	0.7078	0.587	0.8904	0.623	0.8974	0.6197

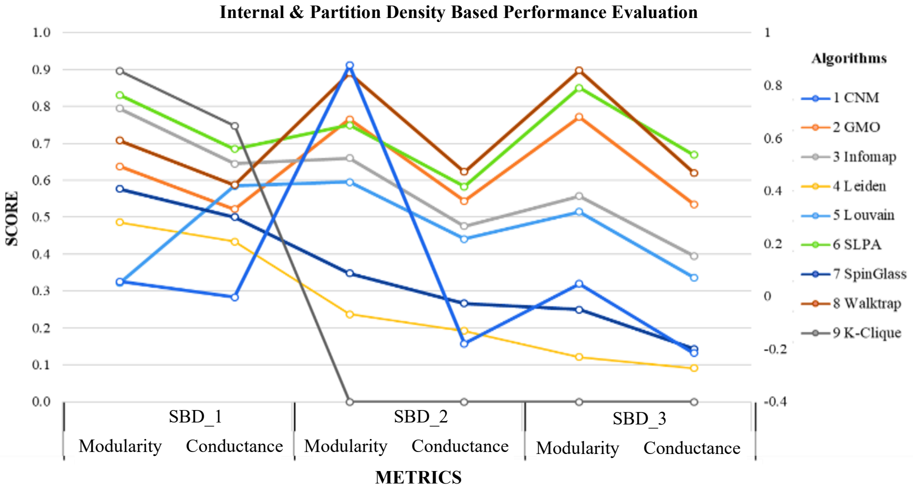

Figure 4. Performance evaluation comparison of internal and partition density visualization

Consolidating the Walktrap algorithm shows the highest internal and partition densities in all the datasets. This shows that this type of algorithm works well for density-based methods, and SpinGlass works well for community structures on moderate to large networks. The findings show that algorithms identifying a lower number of communities can indeed identify larger structures in the network. Leiden identified 11 communities in SBD_1 with an internal density of 0.4864 and a partition density of 0.434, which makes a large compact network. SBD_2 identified 19 communities with high cohesiveness, having an internal density of 0.2368 and a partition density of 0.1917. In SBD_3, the algorithm of Leiden identified 15 communities, and as it was already established, it defines well recognizable wide groups. Louvain found 12 communities in SBD_1 with an internal density of 0.321 and a partition density of 0.5847, which were able to provide a balance between generalization and structural differentiation. Louvain used in SBD_2 identified 29 communities with a higher internal as 0.5955 and partition density as 0.4414, while SBD_3 identified 25 communities with internal as 0.5145 and partition density as 0.3361, providing a clear vision of network structures.

4.5. Discussion

The Leiden algorithm produces the highest results in all datasets for both the modularity and conductance score than the other algorithms. In SBD_1 as the first scenario, the modularity is 0.6402 and the conductance of 0.2246, which shows that Leiden is capable of finding the communities that are densely connected and well separated. For SBD_2, it also performs well with a modularity of 0.4203 and conductance of 0.4557, which shows a good performance in the balanced community detection. In the last SBD_3, Leiden also comes out as the most suitable for large networks with a modularity of 0.3991 and conductance of 0.4944. The high modularity and low conductance in all four datasets imply that Leiden offers the best solution to the community detection problem by providing large and clear communities while maintaining structural properties. Louvain also provides a good result with moderate modularity and conductance values of 0.6181, 0.3938, 0.352, and 0.2219, 0.3774, 0.4283 for SBD_1, SBD_2, SBD_3, respectively. SpinGlass performs well, having modularity of 0.6372, 0.431, and 0.3981; conductance of 0.2916, 0.5132, and 0.5451 for the three datasets, indicating that it is appropriate for balanced community structures.

The findings reveal that the compact and smaller number of detected communities by algorithms are actually better placed to identify the more generalized structures within the network. Leiden and Louvain were found to be the most appropriate techniques. Leiden was able to identify 11 communities in the small dataset with an internal density of 0.4864 and partition density of 0.434, indicating that it can create larger and denser clusters. Louvain found the same number of 12 communities in the same dataset with an internal density of 0.321 and slightly higher partition density 0.5847, which is a good balance between generalization and structural clarity. In the medium dataset, Leiden targeted larger clusters with 19 identified communities and a lower density. Louvain identified 29 communities with improved density values in this dataset. For the complex dataset Leiden algorithm efficiently identifies 15 communities while the Louvain algorithm identifies 25. This shows that Leiden is suitable for higher-level analysis using fewer and larger communities, whereas Louvain has a balance between generality and structure, hence making the methods appropriate for community detection in complex networks.

The greater values in modularity score and other related metrics achieved better results on the Leiden and Louvain algorithms using real-time large network data have established the effectiveness of the algorithms in a high ability to discover well-defined communities. The dataset utilized in this work is not comparable to citation-based benchmark datasets like Cora or PubMed. The SBD network is built on the real-time Scopus information, so it cannot be directly compared with the other benchmark networks. But the obtained community structures and modular performance for these real-world networks are consistent, as Leiden, Louvain, and SpinGlass have demonstrated good modular performance. When compared with previous results reported in the related work section, this novel work is devoted to observing the algorithmic performance by evaluating the detected communities using traditional algorithms within the dataset itself, instead of comparison with external sources. Hence, this work achieves by detecting the better and compact communities on SBD keyword co-occurrence network of various data sizes using Leiden, Louvain, and Spinglass methods are the next best efficient methods for community detection. This will assist in choosing appropriate traditional algorithms for improving community detection in real-world data, particularly when handling complex or large-scale datasets.

5. CONCLUSION

This work shows that standard methods for finding communities in large networks can indeed be used for bibliometric analysis on the co-occurrence of keywords. As a result, this research finds that Leiden is a more successful, efficient method for finding generic, compact communities detecting from the networks. These results highlight the strengths of each method and their suitability for distinct network types and dataset characteristics. Louvain and Spinglass are also the next efficient approaches for community detection compared with the other methods, including GMO, Infomap, SLPA, Walktrap, K-Clique and CNM methods. The outcome shows that the choice of the algorithm should be directed by the network properties, analysis and the objective of the study. In conclusion, the evaluation based on modularity, conductance, internal and partition density reveals that Leiden, SpinGlass and Louvain are better algorithms for our different dataset sizes of SBD_1, SBD_2, and SBD_3, respectively. Future research work will focus on incorporating node features by integrating additional nodes, such as article title, year, publication type, and location as features to enhance the identification of communities within complex keyword networks as a community detection model.

ACKNOWLEDGEMENTS

We thank PSGR Krishnammal College for Women for the motivation and encouragement to make this work a success.

FUNDING INFORMATION

Authors state no funding involved.

AUTHOR CONTRIBUTIONS STATEMENT

This journal uses the Contributor Roles Taxonomy (CRediT) to recognize individual author contributions, reduce authorship disputes, and facilitate collaboration.

Name of Author	C	M	So	Va	Fo	I	R	D	O	E	Vi	Su	P	Fu
Kiruthika R.	✓	✓	✓	✓	✓		✓	✓	✓	✓	✓		✓	
Krishnaveni Sakkarapani	✓			✓	✓	✓	✓			✓		✓	✓	

C : **C**onceptualization
M : **M**ethodology
So : **So**ftware
Va : **Va**lidation
Fo : **Fo**rmal analysis

I : **I**nvestigation
R : **R**esources
D : **D**ata Curation
O : Writing - **O**riginal Draft
E : Writing - Review & **E**diting

Vi : **Vi**sualization
Su : **Su**pervision
P : **P**roject administration
Fu : **Fu**nding acquisition

CONFLICT OF INTEREST STATEMENT

Authors state no conflict of interest.

DATA AVAILABILITY

The data that support the findings of this study are available from the corresponding author, [KR], upon reasonable request.

REFERENCES

[1] P. Bedi and C. Sharma, “Community detection in social networks,” *WIREs Data Mining and Knowledge Discovery*, vol. 6, no. 3, pp. 115–135, Feb. 2016, doi: 10.1002/widm.1178.
[2] R. Pranckutė, “Web of Science (WoS) and Scopus: the titans of bibliographic information in today’s academic world,” *Publications*, vol. 9, no. 1, Mar. 2021, doi: 10.3390/publications9010012.
[3] R.-Z. Wei, X.-Y. Liu, and P.-H. Lyu, “Bibliometrics of public administration research hotspots: topic keywords, author keywords, keywords plus analysis,” *Heliyon*, vol. 10, no. 21, Nov. 2024, doi: 10.1016/j.heliyon.2024.e39352.
[4] X. Yang, Z. Liu, J. Li, and Q. Xie, “Communities of co-occurrence network of financial firms in news,” *Procedia Computer Science*, vol. 221, pp. 821–825, 2023, doi: 10.1016/j.procs.2023.08.056.
[5] T. You, J. Yoon, O.-H. Kwon, and W.-S. Jung, “Tracing the evolution of physics with a keyword co-occurrence network,” *Journal of the Korean Physical Society*, vol. 78, no. 3, pp. 236–243, Jan. 2021, doi: 10.1007/s40042-020-00051-5.
[6] Y. Zhang *et al.*, “Deep learning meets bibliometrics: a survey of citation function classification,” *Journal of Informetrics*, vol. 19, no. 1, Feb. 2025, doi: 10.1016/j.joi.2024.101608.
[7] S. Lozano, L. C.-Infante, B. A.-Díaz, and S. García, “Complex network analysis of keywords co-occurrence in the recent efficiency analysis literature,” *Scientometrics*, vol. 120, no. 2, pp. 609–629, Jun. 2019, doi: 10.1007/s11192-019-03132-w.
[8] G. K. Orman, V. Labatut, and H. Cherifi, “On accuracy of community structure discovery algorithms,” *Journal of Convergence Information Technology*, vol. 6, no. 11, pp. 283–292, Nov. 2011, doi: 10.4156/jcit.vol6.issue11.32.
[9] K. Varsha and K. K. Patil, “An overview of community detection algorithms in social networks,” in *2020 International Conference on Inventive Computation Technologies (ICICT)*, IEEE, Feb. 2020, pp. 121–126, doi: 10.1109/ICICT48043.2020.9112563.
[10] V. D. Blondel, J.-L. Guillaume, R. Lambiotte, and E. Lefebvre, “Fast unfolding of communities in large networks,” *Journal of Statistical Mechanics: Theory and Experiment*, vol. 2008, no. 10, p. P10008, Oct. 2008, doi: 10.1088/1742-5468/2008/10/P10008.
[11] M. E. J. Newman, “Analysis of weighted networks,” *Physical Review E*, vol. 70, no. 5, Nov. 2004, doi: 10.1103/PhysRevE.70.056131.

[12] A. F. Al-Mukhtar and E. S. Al-Shamery, “Greedy modularity graph clustering for community detection of large co-authorship network,” *International Journal of Engineering & Technology*, vol. 7, no. 4.19, pp. 857–863, Nov. 2018, doi: 10.14419/ijet.v7i4.19.28058.

[13] V. A. Traag, L. Waltman, and N. J. van Eck, “From Louvain to Leiden: guaranteeing well-connected communities,” *Scientific Reports*, vol. 9, no. 1, Mar. 2019, doi: 10.1038/s41598-019-41695-z.

[14] M. Rosvall and C. T. Bergstrom, “Maps of random walks on complex networks reveal community structure,” *Proceedings of the National Academy of Sciences*, vol. 105, no. 4, pp. 1118–1123, Jan. 2008, doi: 10.1073/pnas.0706851105.

[15] J. Xie, B. K. Szymanski, and X. Liu, “SLPA: uncovering overlapping communities in social networks via a speaker-listener interaction dynamic process,” in *2011 IEEE 11th International Conference on Data Mining Workshops*, IEEE, Dec. 2011, pp. 344–349, doi: 10.1109/ICDMW.2011.154.

[16] A. Clauset, M. E. J. Newman, and C. Moore, “Finding community structure in very large networks,” *Physical Review E*, vol. 70, no. 6, Dec. 2004, doi: 10.1103/physreve.70.066111.

[17] P. Pons and M. Latapy, “Computing communities in large networks using random walks,” in *Computer and Information Sciences - ISCIS 2005*, Springer Berlin Heidelberg, 2005, pp. 284–293, doi: 10.1007/11569596_31.

[18] J. Reichardt and S. Bornholdt, “Statistical mechanics of community detection,” *Physical Review E*, vol. 74, no. 1, Jul. 2006, doi: 10.1103/PhysRevE.74.016110.

[19] G. Palla, I. Derényi, I. Farkas, and T. Vicsek, “Uncovering the overlapping community structure of complex networks in nature and society,” *Nature*, vol. 435, pp. 814–818, Jun. 2005, doi: 10.1038/nature03607.

[20] M. Vasudevan, H. Balakrishnan, and N. Deo, “Community discovery algorithms: an overview,” *Congressus Numerantium*, vol. 196, pp. 127–142, 2009.

[21] B. Toaza and D. E.-Kiss, “Automated bibliometric data generation in Python from a bibliographic database,” *Software Impacts*, vol. 19, Mar. 2024, doi: 10.1016/j.simpa.2023.100602.

[22] N. R. Smith, P. N. Zivich, L. M. Frerichs, J. Moody, and A. E. Aiello, “A guide for choosing community detection algorithms in social network studies: the question alignment approach,” *American Journal of Preventive Medicine*, vol. 59, no. 4, pp. 597–605, Oct. 2020, doi: 10.1016/j.amepre.2020.04.015.

[23] S. Radhakrishnan, S. Erbis, J. A. Isaacs, and S. Kamarthi, “Novel keyword co-occurrence network-based methods to foster systematic reviews of scientific literature,” *PLOS ONE*, vol. 12, no. 3, Mar. 2017, doi: 10.1371/journal.pone.0172778.

[24] A. Hamm and S. Odrowski, “Term-community-based topic detection with variable resolution,” *Information*, vol. 12, no. 6, May 2021, doi: 10.3390/info12060221.

[25] Z. Yang, R. Algesheimer, and C. J. Tessone, “A comparative analysis of community detection algorithms on artificial networks,” *Scientific Reports*, vol. 6, no. 1, Aug. 2016, doi: 10.1038/srep30750.

[26] A. Kakisim and I. Sogukpinar, “Weighting links based on co-occurrence relationship for community detection enhancement,” in *Proceedings of the 2017 International Conference on Data Mining, Communications and Information Technology*, in DMCIT ’17. New York, United States: ACM, May 2017, pp. 1–5, doi: 10.1145/3089871.3089895.

[27] R. K. Darst, Z. Nussinov, and S. Fortunato, “Improving the performance of algorithms to find communities in networks,” *Physical Review E*, vol. 89, no. 3, Mar. 2014, doi: 10.1103/PhysRevE.89.032809.

[28] R. Harder, “Using Scopus and OpenAlex APIs to retrieve bibliographic data for evidence synthesis. A procedure based on Bash and SQL,” *MethodsX*, vol. 12, Jun. 2024, doi: 10.1016/j.mex.2024.102601.

[29] A. N. A. Nuar and S. C. Sen, “Examining the trend of research on big data architecture: bibliometric analysis using Scopus database,” *Procedia Computer Science*, vol. 234, pp. 172–179, 2024, doi: 10.1016/j.procs.2024.04.010.

[30] S. V Mahadevkar, S. Patil, K. Kotecha, L. W. Soong, and T. Choudhury, “Exploring AI-driven approaches for unstructured document analysis and future horizons,” *Journal of Big Data*, vol. 11, no. 1, Jul. 2024, doi: 10.1186/s40537-024-00948-z.

BIOGRAPHIES OF AUTHORS

Kiruthika R. holds a Bachelor of Science (B.Sc.) in Computer Science, Master of Science (M.Sc.) in Computer Science and Master of Philosophy (M.Phil.) in Computer Science, besides several professional certificates and skills. She is currently pursuing Doctor of Philosophy (Ph.D.) in Computer Science at PSGR Krishnammal College for Women, Peelamedu, Coimbatore. Her research areas of interest include data mining, web mining, social network analysis, community detection, and bibliometric analysis. She can be contacted at email: kirthikamole@gmail.com.

Krishnaveni Sakkarapani completed MCA., M.Phil., Ph.D., in Computer Science and is currently working as an assistant professor in Department of Data Analytics (PG), PSGR Krishnammal College for Women, Coimbatore, Tamil Nadu, India. Fourteen years of experience in teaching and published 80+ papers in international journals and chapters. Also presented 30+ papers in various national and international conferences. Interested research areas are data mining and warehousing, software engineering, bioinformatics, computer networks, and neural networks. She is a reviewer in national and international journals. She can be contacted at email: krishnavenis@psgrkcw.ac.in.