

Visualizing the Detected Communities Using Traditional Algorithms on Keyword Co-occurrence Networks.

Kiruthika. R^{1*}, Dr. Krishnaveni Sakkarapani²

¹*Research Scholar, Department of Computer Science, PSGR Krishnammal College for Women, Coimbatore, India
Email: kirthikamole@gmail.com - ORCID: 0009-0001-1636-3151*

²*Assistant Professor, Department of Data Analytics (PG), PSGR Krishnammal College for Women, Coimbatore, India
Email: krishnavenis@psgrkcw.ac.in - ORCID: 0000-0002-0735-1263*

ABSTRACT

Keyword co-occurrence analysis is essential for understanding emerging trends in research and discovering specific studies. The process of detecting communities by the group nodes in a network based on their interconnection as a structure is called community detection. The community detection algorithm helps analyze and detect the real connection as clusters as a structure within the network. Visualization is one of the significant ways to understand complex networks like community structures. The main aim of this work focuses on visualizing the detected communities based on the co-occurrence of keywords using traditional community detection algorithms. The methodology involves a process of gathering deep learning-based articles from Scopus Bibliographic Dataset (SBD) information based on three major time frames as network datasets, namely SBD_1 as 2006-2013, SBD_2 as 2014-2016 and SBD_3 as 2017. This data is mainly worked with Indexed keyword fields as nodes and their weighted co-occurrences as edges into networks. This work proposed a framework for converting the bibliographic data into graphs for visualizing the detected communities. This work helps scholars to understand the connections among keywords and patterns for their effective research works like extracting academic research articles through exact keyword matching

Keywords: Community Detection Visualization Scopus Bibliometric Data co-occurrence Networks Complex Network.

How to cite this article: R K, Sakkarapani K, Visualizing the Detected Communities Using Traditional Algorithms on Keyword Co-occurrence Networks.. Int J Drug Deliv Technol. 2026;16(2s): 1009-1019: DOI: 10.25258/ijddt.16. 1009-1019

Source of support: Nil.

Conflict of interest: None

INTRODUCTION

Detecting communities from the network has become a requisite topic in social network analysis and applications research domain [1–4]. Finding collections of nodes that are closely connected to each other but distinct from the other parts of the network is generally known as community detection [5–8]. These interrelated groupings of nodes are frequently labeled as communities. Community detection in various applications based on keyword co-occurrence networks which include extraction of exact relevant research articles from the academic database, modeling topics in large documents, clustering research papers by their keywords, social network analysis to identify groups with shared interests and search engine optimization through keyword relationship analysis [9–13].

Co-occurrence networks provide a way to approach analyzing similar patterns among keywords, abstracts, authors, or citations [14–16]. However, this work focuses only on Indexed keywords of Scopus Bibliographic Dataset (SBD). These networks are created via interconnecting pairs of keywords by detecting groupings of nodes based on their interconnections [17–19]. Networks consist of nodes and links that show connections between pairs of nodes [20–22]. Co-occurrence network analysis is used to observe structures and research interests within a certain domain [23,24].

Visualizing the bibliometric data into networks is an efficient approach for discovering the more complicated interconnections among scientific articles [25–27]. This process includes the implementation of visualization methods and processes to analyze the interconnection among various scientific articles based on this bibliometric data, such as citations, co-authorship, and co-occurrence [28,29]. Visualizations would benefit scholars to find significant authors, articles and emerging trends for enhancing academic research and attract the researcher's attention for a long time. This provides a significant understanding of the dynamics of scholarly communication and cooperation. Visualizing techniques are also extended to envision larger-scale complex graphs by using detected communities to compress the original graph. Social media along with other online platforms enable the formation of enormous digital community networks during the recent period of rapid expansion. Social scientists require comprehensive knowledge of community structures together with their behavioral patterns to achieve accurate results in both rumor detection processes and political sentiment analysis operations. Multiple detection strategies exist to analyze these social interaction clusters together with their meanings specifically when detecting misinformation and analyzing social polarization. The research by Xu et al. (2023) applied graph neural networks to recognize social media rumors because network-based

*Author for Correspondence: kirthikamole@gmail.com

methods excel at detecting patterns of community behavior [30]. Traditional community detection algorithms keep their value because of their scalability and clarity and minimal computational requirements even when the latest solutions from deep learning and GNNs exist. Networked data becomes easily understandable through these methods because they show community structures while offering visual insight. The assessment of complex community patterns requires visual tools as essential elements which support their understanding. Safuan et al. (2024) proved that Tableau delivers effective visualizations of financial information which supports wider applications for visual analytics during pattern recognition processes. According to Riskhan et al. (2024) and Aziz et al. (2023) the identification of relevant clusters holds essential value when analyzing both religious mindsets and phishing activity because it enables the identification of tightly bonded groups in extensive networks. Research adopts widely-used algorithms to find and represent communities in network data with the purpose of merging analytical information with visual interpretation capabilities [31-34]. The methodology for analyzing and visualizing the keywords from bibliographic data becomes efficient by using keyword co-occurrence networks with the help of NLP techniques [35–41] for preprocessing the data. This implementation is carried out based on Python-based approaches for community detection [42–48]. This helps to reveal hidden structures within the networks using community detection methods. This supports researchers to track scientific article domain evolution as clusters from academic publications. These visual representations of detected communities built from keyword co-occurrence networks provide an understanding of interconnection among the deep learning research publication structures by applying traditional community detection methods using Louvain, SpinGlass, GMO, Infomap, SLPA, Leiden, CNM, Walktrap and K-Clique [49–58].

2. Methodology

This article explores a significant approach to visualize detected communities in word co-occurrence networks of the Scopus Bibliographic Dataset through various traditional community detection algorithms. The methodology starts with collecting the bibliographic information followed by preprocessing steps and graph construction and network analysis measures before visualizing communities along with different community discovery methods comparison. The research objective focuses on understanding keyword network structures by identifying major communities. This work contributes to the researchers to discover modern academic trends and identify new research areas. The process flow for our methodology is illustrated in Fig.1.

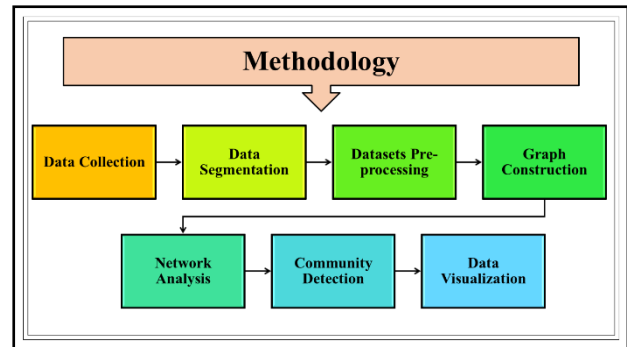


Figure 1. Methodology for visualizing the Detected Communities

The first phase in our methodology is the data extraction process which is essential for real-time analysis. The data extraction process retrieves essential bibliographic records from the SBD but focuses specifically on deep learning-related articles. The Data Segmentation process divides the extracted data into three yearly segments which include 2006-2013, 2014-2016 and 2017 because Scopus only allows 20,000 article retrievals. Data preprocessing removes noise and extraneous information, by using different techniques like cleaning, normalizing keywords, and eliminating ambiguity to ensure network consistency. The preprocessing technique enables the effective network construction of keyword co-occurrence by creating an adjacency matrix. Network analysis is essential for visualizing detected communities to analysis the network structure and relationships. Network analysis measures include nodes, edges, weighted edge, type of graph, average clustering coefficient, and edge density which are used for analyzing the community structures with the network. This network consists of nodes as keywords and weighted edges as keyword co-occurrence frequencies. Each edge weight represents the extent of the semantic connection between two keywords.

Traditional community discovery techniques, such as Louvain, SpinGlass, GMO, Infomap, SLPA, Leiden, CNM, Walktrap, and K-Clique, are used in this study for detecting and visualizing complicated networks. These techniques detect strongly interconnected regions in graphs and provide keyword clusters. The approach discovers communities within the term co-occurrence network by identifying nodes with greater connectivity. These communities are then evaluated to reveal hidden patterns and study fields, giving information about knowledge structures and subject evolution. The primary objective is to find communities by applying these standard approaches to keyword co-occurrence networks built from scientific collections of data via Python. This technique improves the recognition of scientific areas by enabling the detection of new trends and significant research.

2.1 Datasets Description and preprocessing

The rapid growth in the number of research publications in recent times has prompted the creation of scholarly databases that track publications and their citation records [59,60]. The metadata fields present in these databases

serve multiple functions including information search and retrieval as well as research assessment [61]. Emerging bibliographic databases transform the way of scholarly article retrieval and analysis of bibliographic information. Scopus serves as the main source for scholarly databases. Scopus by Elsevier is one of the bibliographic indexing repositories that contains articles from a predetermined set of publications.

The data collecting technique is defined as an organized framework for collecting and transmitting data frequently by ensuring the correctness and maintaining consistency of the data. The Scopus online interface export tools were used to extract bibliographic source information. Each frequently used deep learning-based article was identified by using the option that Searches within “Article title, Abstract, Keywords” from Scopus. The list of deep learning-based articles was exported as a CSV file as bibliographic data for a different set of years. The main reason for doing the data Segmentation process is that Scopus allows up to 20,000 records per query to be exported. In the data segmentation process, segments of these retrieved datasets are primarily separated into three datasets namely SBD_1 as 2006-2013, SBD_2 as 2014-2016 and SBD_3 as 2017.

Data preprocessing is essential for identifying and correcting any inaccuracies present in the dataset. The method for finding and visualizing the communities in Scopus bibliographic keyword co-occurrence networks via data processing is described in this article. The initial dataset contains complete bibliographic information yet this research focuses on extracting the indexed keyword field. The cleaned keywords eliminate duplicate entries complete records and redundant items from the dataset. The standardized keywords result from converting text to lowercase and removing punctuation while also removing stop words. The process creates a co-occurrence adjacency matrix that displays keyword relationships according to their document appearances. Self-loops get removed from the network structure because this step maintains meaningful connections between keywords while eliminating the formation of edges that link a keyword to itself. There are numerous studies related to this has been conducted [62-67].

2.2 Keyword Co-Occurrence Network

Keyword co-occurrence networks are used to identify the most frequent keyword to investigate the interrelationship connections of keywords among the scientific articles. These networks are used for identifying emerging trends and topics by tracking and analyzing the frequency of certain keywords as co-occurrence patterns of keywords over time. Keyword co-occurrence networks are denoted by nodes for each keyword and links for each time a pair of words occurs together. The number of times a pair of terms appears in various articles determines the weight of the connection connecting them. The nodes within the network represent keywords from journal articles and the links demonstrate co-occurring word relationships while link thickness reflects weight values. A sample example of how

the Keyword Co-Occurrence Network works is shown in Fig. 2.

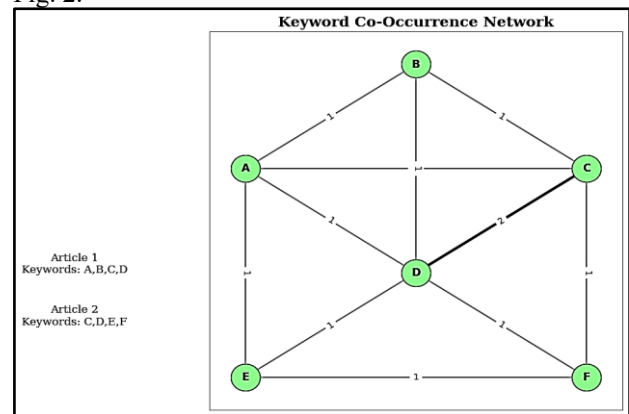


Figure 2. Sample Example of Keyword Co-Occurrence Network.

2.3 Graph Construction

Keyword co-occurrence graphs are constructed by expressing keywords as nodes, their co-occurrences in documents as edges and the frequencies of their co-occurrences as weighted edges. Initially, keywords are extracted from the dataset to calculate the co-occurrence among them to construct the graph. In Figure 3, each character term in the table is considered as a keyword which shows how often a pair of keywords appears together. This number becomes the weight of the connection, which shows how closely the keywords co-occur with each other. This table is turned into another table by using a keyword pair with frequency if keywords appear together. The connection between them is two-way, which shows how the points are connected to the graph. Self-loops are avoided to maintain meaningful connections. The constructed graph is typically undirected, although directed graphs can be used if the order of keyword occurrences is considered. The example working procedure of graph construction for community detection is exemplified in Fig. 3.

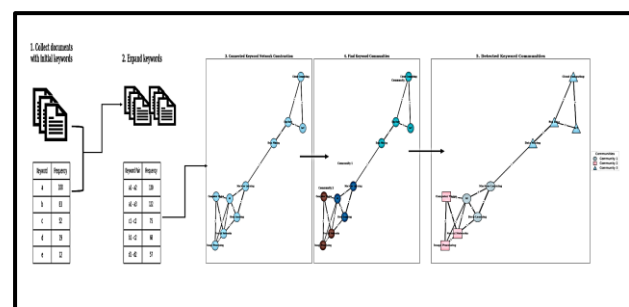


Figure 3. Sample Example of Visualizing the Community Detection for Keyword Co-Occurrence Network

The graph is constructed using this approach by defining keywords as nodes and representing their co-occurrences as weighted edges. After converting into graphs, these constructed graphs are separated into three networks namely SBD_1 as 2006-2013, SBD_2 as 2014-2016 and SBD_3 as 2017. These three graphs are constructed to encompass a range of network sizes, including small,

medium and large graphs to evaluate the effectiveness of community detection using numerous traditional algorithms across different network scales. Network analysis metrics are necessary for graph building which provides perspectives on the structural characteristics and interconnections within keyword co-occurrence networks. The separated visualization of constructed graphs for network datasets SBD_1, SBD_2, and SBD_3 individually are visualized as graphs as shown in Fig. 4 which contains a) SBD_1 as 2006-2013, b) SBD_2 as 2014-2016 and c) SBD_3 as 2017.

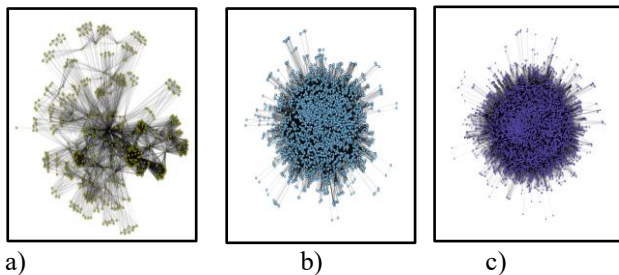


Figure 4. Graphical Representation of SBD Networks.

The main aim of this graph construction process involved examining these three networks to detect community structures. Network SBD_1 contained 599 nodes and 8085 edges with network analysis measures like average clustering coefficient as 0.9119 and edge density as 0.0485 to form a less complex system. The Network SBD_2 was constructed with medium complex network patterns through its 3934 nodes which joined 65587 edges with an average clustering coefficient of 0.8764 and edge Density of 0.0085. The third Network SBD_3 demonstrated the highest network scales of network interconnection through its 8202 nodes which connected to 155975 edges with an average clustering coefficient of 0.8614 and edge Density of 0.0046. The various network sizes and connection densities enable us to detect communities and analyze the connections that exist in the dataset.

2.4 Community Detection

Community detection algorithms are essential tools for understanding the structure of complex networks. The communities are detected within the network as clusters where nodes establish more connections within their group than externally. Some common methods for finding communities are SpinGlass, GMO, Walktrap, CNM, Louvain, K-Clique, Infomap, SLPA, Leiden, etc. The SpinGlass method uses physics ideas to find the best community setup that uses the least energy in the network. The GMO method tries to find the best community setup that makes the network's modularity as high as possible. The Walktrap method uses random paths to find communities by noting which nodes are visited often. The Clauset, Newman, and Moore (CNM) algorithm is a community discovery approach that uses modularity to find communities in broad networks. This implements a greedy approach of merging vertices based on modularity obtained which makes it suitable for sparse networks. The findings of the Greedy Modularity Optimization method focus on

the changes and structural patterns of the detected communities over three different time intervals. The Louvain method is a dynamic method that finds the community based on modularity of the network as high as possible. The K-Clique method finds communities by looking for groups of nodes that are all connected. The Infomap method makes the path of a random walker in a network as short as possible. SLPA method uses label sharing to locate communities by repeatedly changing node labels based on their neighbor labels. Leiden methods perform better than Louvain since it is an advanced method. This method finds communities that are strongly linked inside and separate from each other.

3. Experimental Results and Discussions

Visualizing the identified communities using traditional methods on term co-occurrence networks constructed from the Scopus bibliographic dataset is the main objective of the experimental results. Indexed terms from deep learning publications are used to construct the networks. Community detection is performed by using SpinGlass, GMO, Walktrap, CNM, Louvain, K-Clique, Infomap, SLPA, and Leiden methods. The performance results are compared and evaluated based on the number of detected communities which identifies the meaningful communities by highlighting differences in community structures across the numerous applied algorithms.

3.1 Evaluating the Performance of Detected Communities

Community detection consists of grouping nodes based on their strong connectivity into different communities within the network. The performance of detected communities should be evaluated based on the number of communities within the network. Network structural complexity and detection methods provide insights for analyzing detected communities. The number of identified communities serves as a vital indicator to understand the network structure. A low number of detected communities has the effect of combining distinct sub-groups which results in information loss. An excessive number of detected communities indicates over-sensitivity because it finds irrelevant distinctions between small variations in groups which makes interpretation difficult. A large number of communities demonstrates complex structures with multiple distinct subgroups but fewer communities suggest tightly linked clusters with unified characteristics. The appropriate optimal and compact number of detected communities represents natural divisions within the network by identifying significant clusters without creating excessive subdivisions. The network characteristics together with the analysis goals determine how these factors should be balanced. The evaluation of community detection performance across SBD_1, SBD_2, and SBD_3 revealed significant variations in the number of communities identified by different methods as shown in Table 1.

Table 1. Number of Detected Communities for SBD_1, SBD_2, SBD_3

Sno	Method	SBD 1	SBD 2	SBD 3
1	SpinGlass	14	25	24
2	GMO	12	43	75
3	Walktrap	17	146	263
4	CNM	10	51	74
5	Louvain	12	29	25
6	K-Clique	14	-	-
7	Infomap	29	211	365
8	SLPA	22	49	21
9	Leiden	11	19	15

Traditional algorithms help establish a reference for evaluating the performance of enhanced methods through network community detection in keyword co-occurrence networks because they show the differences in thematic cluster identification. Our model incorporated nine traditional community detection methods compared with these networks. The selection of community detection methods leads to significant differences in the detected community results and node distribution patterns. The number of communities discovered by various algorithms during different time intervals of SBD_1, SBD_2 and SBD_3 data are visualized in Fig. 5.

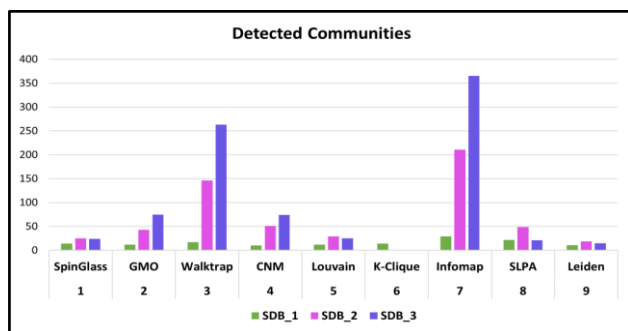


Figure 5. Comparison of Detected Communities for SBD_1, SBD_2, SBD_3.

3.2 Visualizing the Detected Communities

Visual representations of detected communities in keyword co-occurrence networks help researchers understand the outcomes of community detection algorithms. It enables researchers to observe both the network structure along its community structure. Visual presentation of the communities reveals their distribution as well as keyword relationships throughout the thematic structure. The visualization contributes the researchers to discovering patterns where keywords frequently appear together to indicate common research interests or emerging trends. The interpretation of complicated keyword networks becomes simpler through visualization because it helps researchers detect dominant themes and new topics along with important keywords throughout different communities. The visual presentation of research findings becomes more effective through keyword network organization which enables researchers to present keyword clusters and their

research domain importance. The visual depiction helps researchers complete multiple tasks including research domain identification thematic module comprehension and emerging keyword discovery in the keyword network. The visualization of detected communities for each method contains a legend that represents the number of detected communities separated with different color representations. If the legend in the figures is not mentioned or cropped means the detected communities are not compact in the network.

The SpinGlass algorithm identified 14 communities in SBD_1, 25 in SBD_2 and 24 in SBD_3. The performance analysis indicates that it is stable and efficient for medium to large data sets which ensures consistent detection as the network size changes. The picturing of the identified communities is signified in Fig. 6 which contains a) SBD_1 as 2006-2013, b) SBD_2 as 2014-2016 and c) SBD_3 as 2017.

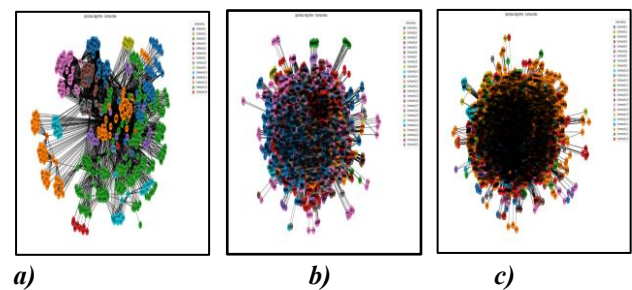


Figure 6. Visualization of Communities Detected by SpinGlass Algorithm.

The GMO algorithm found 12 communities for SBD_1, 43 for SBD_2 and 75 for SBD_3. The method effectively handles both moderate and large networks which demonstrates adaptability. The identified communities are represented graphically in Fig. 7 which contains a) SBD_1 as 2006-2013, b) SBD_2 as 2014-2016 and c) SBD_3 as 2017.

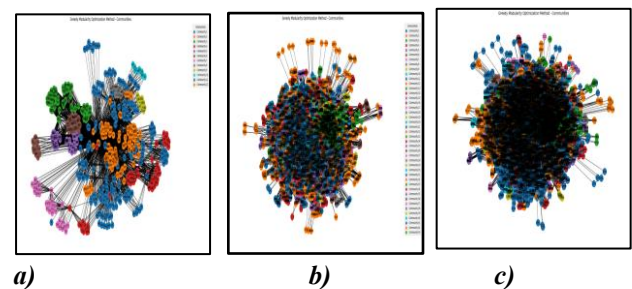


Figure 7. Visualization of Communities Detected by GMO Algorithm.

Walktrap identified 17 communities in SBD_1; this increased to 146 in SBD_2 and 263 in SBD_3. Performance evaluation also shows that it has a good degree of flexibility and is suitable for use with large data sets. The increase in the number of detected communities further proves that Walktrap is not effectively capable of identifying new structures in large networks that were observed during the

expansion of the 2017 dataset. The community structures that were detected are depicted in the graphic shown in Fig. 8 which contains a) SBD_1 as 2006-2013, b) SBD_2 as 2014-2016 and c) SBD_3 as 2017.

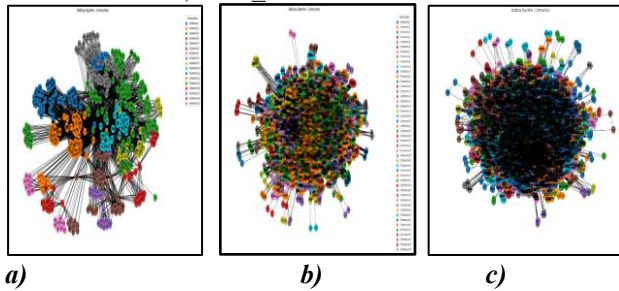


Figure 8. Visualization of Communities Detected by Walktrap Algorithm

CNM algorithm discovered ten communities for SBD_1, 51 for SBD_2 and 74 in SBD_3. Performance evaluation indicated consistent detection in medium-sized networks, although this could be challenging for large-scale networks. This approach works best for small to medium-sized data sets since it is like a hierarchical grouping. This gradual community growth shows that this method effectively identifies clusters of data and could have difficulties in significantly strongly linked networks, as observed in 2017. A graphical representation of the identified communities is illustrated in Fig. 9 which contains a) SBD_1 as 2006-2013, b) SBD_2 as 2014-2016 and c) SBD_3 as 2017.

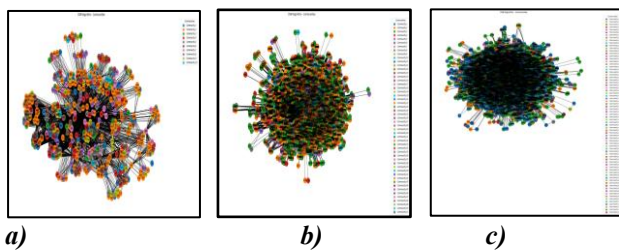


Figure 9. Visualization of Communities Detected by CNM Algorithm.

Louvain algorithm discovered 12 communities for SBD_1, 29 SBD_2 and 25 SBD_3. Performance analysis demonstrates that the technique is very adaptive to moderate and big datasets. The algorithm is most appropriate for networks of moderate size with dynamical communities. The identified communities from the networks using the Louvain algorithm are shown in Fig. 10 which contains a) SBD_1 as 2006-2013, b) SBD_2 as 2014-2016 and c) SBD_3 as 2017.

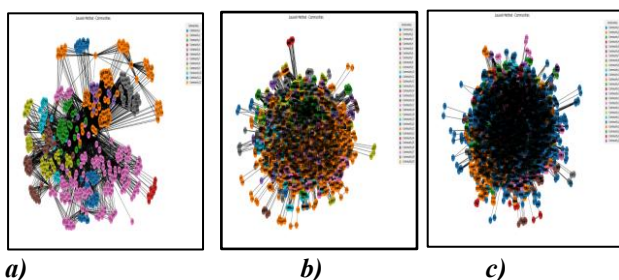


Figure 10. Visualization of Communities Detected by Louvain Algorithm.

K-Clique algorithm identified 14 communities for SBD_1. Community detection using K-Clique for SBD_2 and SBD_3 is not computationally possible due to insufficient run time, so results are not mentioned in the table. Performance evaluation of K-Cliques shows its usefulness in identifying compact group-sized data, performing well with less complex network topologies, but efficiency may decrease with increasing complexity. The graphical layout of community detection by K-Clique for SBD_1 alone is illustrated in Fig. 11.

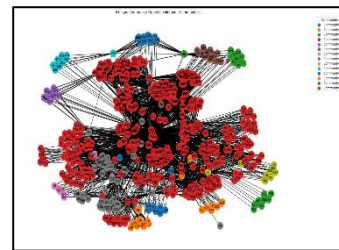


Figure 11. Visualization of Communities Detected by K-Clique for SBD_1.

Infomap detected 29 communities for SBD_1, increased to 211 in SBD_2 and 365 by SBD_3. Infomap struggles with large datasets that make it less suitable for collecting important community patterns in complex networks. This method became insufficient and not an optimized method for the massive identification of communities. The detected communities are illustrated graphically in Fig. 12 which contains a) SBD_1 as 2006-2013, b) SBD_2 as 2014-2016 and c) SBD_3 as 2017.

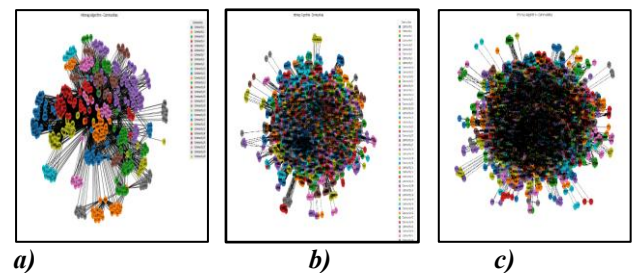


Figure 12. Visualization of Communities Detected by Infomap Method.

SLPA algorithm identified 22 communities for SBD_1, 49 for SBD_2 and 21 for SBD_3. Performance evaluation shows how it is scalable to different networks and might have difficulty in capturing all the communities in large data sets. The ability of this algorithm to detect communities in various network sizes demonstrates its utility in structures of varying densities. The graphical presentation of identified communities is shown in Fig. 13 which contains a) SBD_1 as 2006-2013, b) SBD_2 as 2014-2016 and c) SBD_3 as 2017.

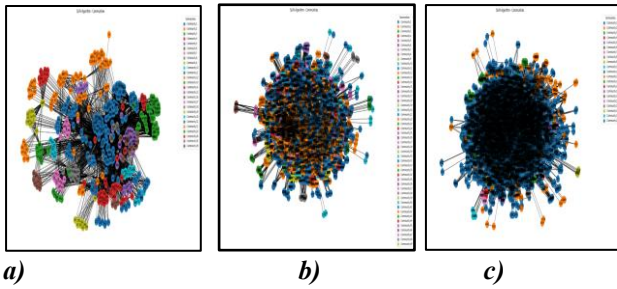


Figure 13. Visualization of Communities Detected by SLPA Algorithm.

Leiden algorithm detected 11 communities in SBD_1, 19 in SBD_2 and 15 in SBD_3. Evaluations prove its efficiency in preserving the consistency of the community structure across different sizes of the dataset which makes it optimal for any-sized networks. This is a consistent community detection method suitable for numerous dataset sizes. The outcome of community detection by the Leiden algorithm is illustrated in Fig. 14 which contains a) SBD_1 as 2006-2013, b) SBD_2 as 2014-2016 and c) SBD_3 as 2017 in the form of a graphical layout.

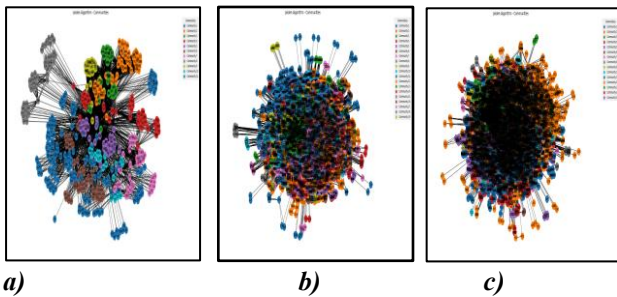


Figure 14. Visualization of Communities Detected by Leiden Algorithm.

3.3 Discussion

The detection of fewer compact communities leads algorithms to identify generalized network structures effectively. Leiden and Louvain produce the most effective results. Leiden detects 11 communities within the small dataset which form bigger and denser subdivisions. Louvain divides the same dataset into 12 communities which preserves both structural clarity and generalized results. Leiden groups the medium dataset into 19 large communities yet Louvain detects 29 communities that present a more refined community structure. Within the complex dataset, Leiden provides efficient detection of 15 communities while Louvain finds 25 communities that balance between general and detailed structural representation.

The Leiden algorithm detects densely connected and well-separated communities throughout all datasets. The SBD_1 scenario demonstrates that Leiden identifies strong cohesive communities. The Leiden algorithm detects balanced community structures in SBD_2. Leiden stands out as the optimal algorithm for large networks because it detects well-defined substantial communities in SBD_3. From the observation, the Leiden algorithm detects distinct

community clusters that maintain the complete network structure. Louvain achieves successful detection of moderately cohesive communities throughout its analysis of all datasets. Louvain detects specific meaningful community structures within SBD_1 SBD_2 SBD_3 which establishes a proper balance between generalization and structural clarity. SpinGlass works efficiently to detect evenly distributed community structures which makes it appropriate for networks with uniform connection patterns. The detection of distinct communities by Leiden makes it an appropriate method for high-level analysis of larger community structures. Louvain maintains a balance between broad community detection and detailed network structure analysis which makes it efficient in complex network systems. SpinGlass provides precise community detection by identifying balanced structures between communities thus demonstrating its strength in detecting community structures. The Leiden method proves to be the optimal solution for identifying cohesive communities across different data ranges. The community detection capabilities of Louvain and SpinGlass prove to be highly efficient which makes them important tools for studying complex networks.

4. Conclusion

Traditional community detection algorithms prove their effectiveness for visualizing networks of keyword co-occurrence in this work. Analysis of keyword interconnections in the SBD during three separate periods enables researchers to identify important research topic structural patterns. The Leiden algorithm excels at identifying cohesive well-defined communities which makes it ideal for high-level analysis through its ability to detect and be suitable for diverse network sizes. Louvain operates as a method that successfully combines the ability to generalize patterns while preserving detailed network structure qualities and is suitable for various network scales to study complex systems effectively. SpinGlass demonstrates its ability to identify balanced community structures.

This work achieves better identification of optimal communities by using network general structures. The comparative evaluation based on detected communities reveals that Leiden and Louvain represent the best community detection methods which are suitable for different network sizes. SpinGlass is best at analyzing structured networks with uniform connection distribution. This work presents an organized method to visualize keyword communities which enhance the ability to track research developments and topic changes. This work adds value to keyword co-occurrence research through its effective visual framework for community detection which helps researchers find connections between keywords and spot new research fields. The proposed framework facilitates a deeper understanding of academic research patterns, guiding scholars in keyword selection for literature exploration and research development.

4. Future Work and Scope

The scope of future work is to evaluate the community detection methods by using measures like modularity, conductance, internal density, and partition density which will improve the performance of the model. Exploring new measures for evaluating the community detection methods will enhance the community cohesiveness and network partition which offer an expanded perspective of community structures. Comparative investigations using numerous datasets and methods would provide further insight into the network structures. Furthermore, developing an adaptive framework using various network analysis measures and properties takes into consideration for evaluating community detection models.

Author Statements

Ethical approval: The conducted research is not related to either human or animal use.

Conflict of interest: The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper

Acknowledgement: We thank PSGR Krishnammal College for Women for the Motivation and Encouragement to make this work a successful one by providing GRG Trust Fellowship.

Author contributions: The authors declare that they have equal right on this paper.

Funding information: The authors declare that there is a GRG Trust Fellowship by PSGR Krishnammal College for Women to be acknowledged.

Data availability statement: The data that support the findings of this study are available on request from the corresponding author. The data are not publicly available due to privacy or ethical restrictions.

REFERENCE

1. Malode, Y., Aylani, A., Bhardwaj, A., & Hajoary, D. (2025). Comparative Analysis of Community Detection Algorithms on the SNAP Social Circles Dataset. <http://arxiv.org/abs/2502.04341>
2. Ehlers, H., Brich, N., Krone, M., Nöllenburg, M., Yu, J., Natsukawa, H., Yuan, X., & Wu, H. Y. (2025). An introduction to and survey of biological network visualization. *Computers and Graphics*, 126. <https://doi.org/10.1016/j.cag.2024.104115>
3. Fortunato, S., & Hric, D. (2016). Community detection in networks: A user guide. *Physics Reports*, 659, 1–44. <https://doi.org/10.1016/J.PHYSREP.2016.09.002>
4. Rani, S., & Mehrotra, M. (2019). Community Detection in Social Networks: Literature Review. *Journal of Information and Knowledge Management*, 18(2). <https://doi.org/10.1142/S0219649219500199>
5. Yao, J., & Liu, B. (2024). Citation: Community-Detection Method of Complex Network Based on Node Community-Detection Method of Complex Network Based on Node Influence Analysis. <https://doi.org/10.3390/sym16060754>
6. Brahim, L., Mouad, M., Chihab-Eddine, C., & Ali, I. (2024). A SURVEY ON COMMUNITY DETECTION: APPLICATIONS, ALGORITHMS, AND CHALLENGES. *Journal of Theoretical and Applied Information Technology*, 30(12). www.jatit.org
7. Chunaev, P. (2020). Community detection in node-attributed social networks: A survey. *Computer Science Review*, 37, 100286. <https://doi.org/10.1016/J.COSREV.2020.100286>
8. El-Moussaoui, M., Agouti, T., Tikniouine, A., & El Adnani, M. (2019). A comprehensive literature review on community detection: Approaches and applications. *Procedia Computer Science*, 151, 295–302. <https://doi.org/10.1016/J.PROCS.2019.04.042>
9. Lin, W., Wu, X., Wang, Z., Wan, X., & Li, H. (2022). Topic Network Analysis Based on Co-Occurrence Time Series Clustering. *Mathematics 2022*, Vol. 10, Page 2846, 10(16), 2846. <https://doi.org/10.3390/MATH10162846>
10. Wartena, C., & Brussee, R. (2008). Topic detection by clustering keywords. *Proceedings - International Workshop on Database and Expert Systems Applications, DEXA*, 54–58. <https://doi.org/10.1109/DEXA.2008.120>
11. Weng, M. H., Wu, S., & Dyer, M. (2022). Identification and Visualization of Key Topics in Scientific Publications with Transformer-Based Language Models and Document Clustering Methods. *Applied Sciences 2022*, Vol. 12, Page 11220, 12(21), 11220. <https://doi.org/10.3390/APP122111220>
12. Zhang, H., Xu, H., Long, F., Wang, X., & Gao, K. (2024). Unsupervised Multimodal Clustering for Semantics Discovery in Multimodal Utterances. <http://arxiv.org/abs/2405.12775>
13. Priyanka, S., Rama Krishna, S., & Tech, M. (2021). COMMUNITY DETECTION IN SOCIAL NETWORK USING GRAPH CLUSTERING METHODS. *Turkish Journal of Computer and Mathematics Education (TURCOMAT)*, 12(13), 5687–5697. <https://doi.org/10.17762/TURCOMAT.V12I13.9812>
14. Sood, S. K., Rawat, K. S., & Kumar, D. (2022). Analytical mapping of information and communication technology in emerging infectious diseases using CiteSpace. *Telematics and Informatics*, 69, 101796. <https://doi.org/10.1016/J.TELE.2022.101796>
15. Narong, D. K., & Hallinger, P. (2023). A Keyword Co-Occurrence Analysis of Research on Service Learning: Conceptual Foci and Emerging Research Trends. *Education Sciences*, 13(4), 339. <https://doi.org/10.3390/EDUCSCI13040339/S1>
16. Li, H., An, H., Wang, Y., Huang, J., & Gao, X.

- (2016). Evolutionary features of academic articles co-keyword network and keywords co-occurrence network: Based on two-mode affiliation network. *Physica A: Statistical Mechanics and Its Applications*, 450, 657–669. <https://doi.org/10.1016/J.PHYSA.2016.01.017>
17. He, D., Jin, D., Chen, Z., & Zhang, W. (n.d.). Identification of hybrid node and link communities in complex networks. <https://doi.org/10.1038/srep08638>
18. Zhao, N., Yang, S., Wang, H., Zhou, X., Luo, T., & Wang, J. (2024). A Novel Method to Identify Key Nodes in Complex Networks Based on Degree and Neighborhood Information. *Applied Sciences* 2024, Vol. 14, Page 521, 14(2), 521. <https://doi.org/10.3390/APP14020521>
19. Chigarev, B. (2024). Keyword Co-Occurrence Analysis Using the FPGrowth Algorithm. An Example of Energies Journal Bibliometric Data for 2023-2024. <https://doi.org/10.20944/PREPRINTS202406.1380.V1>
20. Muppidi, S., & Reddy, K. T. (2020). Co-occurrence analysis of scientific documents in citation networks. *International Journal of Knowledge-Based and Intelligent Engineering Systems*, 24(1), 19–25. https://doi.org/10.3233/KES-200025/ASSET/IMAGES/LARGE/10.3233_KES-200025-FIG8.JPEG
21. Yuan, C., Li, G., Kamarthi, S., Jin, X., & Moghaddam, M. (2022). Trends in intelligent manufacturing research: a keyword co-occurrence network based review. *Journal of Intelligent Manufacturing*, 33(2), 425–439. <https://doi.org/10.1007/S10845-021-01885-X/FIGURES/8>
22. Fudolig, M. I., Alshaabi, T., Arnold, M. V., Danforth, C. M., & Dodds, P. S. (2022). Sentiment and structure in word co-occurrence networks on Twitter. *Applied Network Science*, 7(1), 1–27. <https://doi.org/10.1007/S41109-022-00446-2/FIGURES/12>
23. Kumar, D., Librarian, K., & College, B. S. (n.d.). Network Visualization of ChatGPT Research: a study based on term and keyword co-occurrence network analysis. Retrieved February 10, 2025, from <https://www.lens.org/>
24. Chen, X., Tian, W., & Fang, H. (2025). Bibliometric analysis of natural language processing using CiteSpace and VOSviewer. *Natural Language Processing Journal*, 10, 100123. <https://doi.org/10.1016/J.NLP.2024.100123>
25. Zhao, W., Luo, J., Fan, T., Ren, Y., & Xia, Y. (2021). Analyzing and visualizing scientific research collaboration network with core node evaluation and community detection based on network embedding. *Pattern Recognition Letters*, 144, 54–60. <https://doi.org/10.1016/J.PATREC.2021.01.007>
26. Kiruthika, R., & Radha, N. (2025). Author-Centric Pattern Detection in Scopus Citation Network via Community Structures. 1–15. https://doi.org/10.1007/978-981-97-7839-3_1
27. Huang, Z., Wu, J., Zhu, W., Wang, Z., Mehrotra, S., & Zhao, Y. (2021). Visualizing complex networks by leveraging community structures. *Physica A: Statistical Mechanics and Its Applications*, 565, 125506. <https://doi.org/10.1016/J.PHYSA.2020.125506>
28. Kiruthika, R., & Radha, N. (2025). Analyzing the Citation Networks Using Community Detection Approaches: A Review. 398, 181–194. https://doi.org/10.1007/978-981-97-5200-3_13
29. Duvvuru, A., Kamarthi, S., & Sultornsanee, S. (2012). Undercovering research trends: Network analysis of keywords in scholarly articles. *JCSSE 2012 - 9th International Joint Conference on Computer Science and Software Engineering*, 265–270. <https://doi.org/10.1109/JCSSE.2012.6261963>
30. S. Xu et al., “Rumor detection on social media using hierarchically aggregated feature via graph neural networks,” *Applied Intelligence*, vol. 53, no. 3, pp. 3136–3149, 2023.
31. H. A. J. Safuan, C. K. K. Nyein, B. Riskhan, and F. Ashfaq, “Visualizing NASDAQ Composite Index Historical Data Using Tableau,” in *International Conference on Computational Intelligence in Pattern Recognition*, Springer, 2024, pp. 263–274.
32. B. Riskhan, M. S. Hossain, and F. Ashfaq, “Religious Sentiment Analysis and Detection on Social Media,” in *International Conference on Computational Intelligence in Pattern Recognition*, Springer, 2024, pp. 245–262.
33. M. A. A. Aziz, B. Riskhan, N. H. Zakaria, and M. N. Jambli, “An Exploratory Study of Automated Anti-phishing System,” in *International Conference on Computing and Informatics*, Springer, 2023, pp. 58–69.
34. M. S. Hossain, M. D. R. Islam, B. R. Riskhan, M. D. M. H. HASAN, and R. I. ISLAM, “Political sentiment analysis using natural language processing on social media,” *International Journal of Applied Methods in Electronics and Computers*, vol. 12, no. 4, pp. 81–89, 2024.
35. Wang, X., & Zhao, D. (2025). A comparative experimental study of citation sentiment identification based on the Athar-Corpus. *Data Science and Informetrics*. <https://doi.org/10.1016/J.DSIM.2025.01.008>
36. Gao, P., Yang, C., Sun, N., & Zitikis, R. (2025). Predicting classification errors using NLP-based machine learning algorithms and expert opinions. *Machine Learning with Applications*, 19, 100630. <https://doi.org/10.1016/J.MLWA.2025.100630>
37. Xu, Q. (2025). Application of an intelligent English text classification model with improved KNN

- algorithm in the context of big data in libraries. *Systems and Soft Computing*, 7, 200186. <https://doi.org/10.1016/J.SASC.2025.200186>
38. Chen, K., Fei, C., Bi, Z., Liu, J., Peng, B., Zhang, S., Pan, X., Xu, J., Wang, J., Yin, C. H., Zhang, Y., Feng, P., Wen, Y., Wang, T., Li, M., Ren, J., Niu, Q., Chen, S., Hsieh, W., ... Liu, M. (2024). Deep Learning and Machine Learning -- Natural Language Processing: From Theory to Application. <http://arxiv.org/abs/2411.05026>
39. Probierz, B., Kozak, J., & Hrabia, A. (2022). Clustering of scientific articles using natural language processing. *Procedia Computer Science*, 207, 3449–3458. <https://doi.org/10.1016/J.PROCS.2022.09.403>
40. Young, T., Hazarika, D., Poria, S., & Cambria, E. (2018). Recent trends in deep learning based natural language processing [Review Article]. *IEEE Computational Intelligence Magazine*, 13(3), 55–75. <https://doi.org/10.1109/MCI.2018.2840738>
41. Otter, D. W., Medina, J. R., & Kalita, J. K. (2021). A Survey of the Usages of Deep Learning for Natural Language Processing. *IEEE Transactions on Neural Networks and Learning Systems*, 32(2), 604–624. <https://doi.org/10.1109/TNNLS.2020.2979670>
42. Yang, Z., Algesheimer, R., & Tessone, C. J. (2016). A Comparative Analysis of Community Detection Algorithms on Artificial Networks. *Scientific Reports* 2016 6:1, 6(1), 1–18. <https://doi.org/10.1038/srep30750>
43. Cherifi, H., Palla, G., Szymanski, B. K., & Lu, X. (2019). On community structure in complex networks: challenges and opportunities. *Applied Network Science* 2019 4:1, 4(1), 1–35. <https://doi.org/10.1007/S41109-019-0238-9>
44. Toaza, B., & Esztergár-Kiss, D. (2024). Automated bibliometric data generation in Python from a bibliographic database. *Software Impacts*, 19, 100602. <https://doi.org/10.1016/J.SIMPA.2023.100602>
45. Pereira, V., Basilio, M. P., & Santos, C. H. T. (2023). pyBibX -- A Python Library for Bibliometric and Scientometric Analysis Powered with Artificial Intelligence Tools. <https://arxiv.org/abs/2304.14516>
46. Rossetti, G., Milli, L., & Cazabet, R. (2019). CDLIB: a python library to extract, compare and evaluate communities from complex networks. *Applied Network Science*, 4(1), 52. <https://doi.org/10.1007/s41109-019-0165-9>
47. Bhanot, N., Singh, H., Sharma, D., Jain, H., Jain, S., & Management, O. (n.d.). Python vs. R: A Text Mining Approach for analyzing the Research Trends in Scopus Database.
48. Hagberg, A. A., Schult, D. A., & Swart, P. J. (2008). Exploring network structure, dynamics, and function using NetworkX. 7th Python in Science Conference (SciPy 2008), SciPy, 11–15.
49. Blondel, V. D., Guillaume, J. L., Lambiotte, R., & Lefebvre, E. (2008). Fast unfolding of communities in large networks. *Journal of Statistical Mechanics: Theory and Experiment*, 2008(10). <https://doi.org/10.1088/1742-5468/2008/10/p10008>
50. Newman, M. E. J. (2004). Analysis of weighted networks. *Physical Review. E, Statistical, Nonlinear, and Soft Matter Physics*, 70(5), 9. <https://doi.org/10.1103/PHYSREVE.70.056131>
51. F. Al-Mukhtar, A., & S. Al-Shamery, E. (2018). Greedy Modularity Graph Clustering for Community Detection of Large Co-Authorship Network. *International Journal of Engineering and Technology*, 7(4.19), 857–863. <https://doi.org/10.14419/IJET.V7I4.19.28058>
52. Traag, V., Waltman, L., & van Eck, N. J. (2018). From Louvain to Leiden: guaranteeing well-connected communities. *Scientific Reports*, 9(1). <https://doi.org/10.1038/s41598-019-41695-z>
53. Rosvall, M., & Bergstrom, C. T. (2008). Maps of random walks on complex networks reveal community structure. *Proceedings of the National Academy of Sciences of the United States of America*, 105(4), 1118–1123. https://doi.org/10.1073/PNAS.0706851105/SUPPL_FILE/06851SUPPAPPENDIX.PDF
54. Xie, J., Szymanski, B. K., & Liu, X. (2011). SLPA: Uncovering overlapping communities in social networks via a speaker-listener interaction dynamic process. *Proceedings - IEEE International Conference on Data Mining, ICDM*, 344–349. <https://doi.org/10.1109/ICDMW.2011.154>
55. Clauset, A., Newman, M. E. J., & Moore, C. (2004). Finding community structure in very large networks. *Physical Review. E, Statistical, Nonlinear, and Soft Matter Physics*, 70(6), 6. <https://doi.org/10.1103/PHYSREVE.70.066111>
56. Pons, P., & Latapy, M. (2005). Computing communities in large networks using random walks. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 3733 LNCS, 284–293. https://doi.org/10.1007/11569596_31
57. Reichardt, J., & Bornholdt, S. (2006). Statistical mechanics of community detection. *Physical Review E - Statistical, Nonlinear, and Soft Matter Physics*, 74(1). <https://doi.org/10.1103/PHYSREVE.74.016110/FIGURES/13/THUMBNAI>
58. Palla, G., Derényi, I., Farkas, I., & Vicsek, T. (2005). Uncovering the overlapping community structure of complex networks in nature and society. *Nature* 2005 435:7043, 435(7043), 814–818. <https://doi.org/10.1038/nature03607>

59. Martín-Martín, A., Thelwall, M., Orduna-Malea, E., & Delgado López-Cózar, E. (2021). Google Scholar, Microsoft Academic, Scopus, Dimensions, Web of Science, and OpenCitations' COCI: a multidisciplinary comparison of coverage via citations. *Scientometrics*, 126(1), 871–906. <https://doi.org/10.1007/S11192-020-03690-4/FIGURES/1060>.
60. Cascajares, M., Alcayde, A., Salmerón-Manzano, E., & Manzano-Agugliaro, F. (2021). The Bibliometric Literature on Scopus and WoS: The Medicine and Environmental Sciences Categories as Case of Study. *International Journal of Environmental Research and Public Health*, 18(11), 5851. <https://doi.org/10.3390/IJERPH18115851>
61. Singh, P., Singh, V. K., & Kanaujia, A. (2024). Exploring the Publication Metadata Fields in Web of Science, Scopus and Dimensions: Possibilities and Ease of doing Scientometric Analysis. *Journal of Scientometric Research*, 13(3), 715–731. <https://doi.org/10.5530/JSCIRES.20041144>
62. Attaullah, M., Ali, M., Almufareh, M. F., Ahmad, M., Hussain, L., Jhanjhi, N., & Humayun, M. (2022). Initial Stage COVID-19 Detection System Based on Patients' Symptoms and Chest X-Ray Images. *Applied Artificial Intelligence*, 36(1). <https://doi.org/10.1080/08839514.2022.2055398>
63. Aherwadi N, Mittal U, Singla J, Jhanjhi NZ, Yassine A, Hossain MS. Prediction of Fruit Maturity, Quality, and Its Life Using Deep Learning Algorithms. *Electronics*. 2022; 11(24):4100. <https://doi.org/10.3390/electronics11244100>
64. Dogra, V., Singh, A., Verma, S., Kavita, Jhanjhi, N.Z., Talib, M.N. (2021). Analyzing DistilBERT for Sentiment Classification of Banking Financial News. In: Peng, S.L., Hsieh, S.Y., Gopalakrishnan, S., Duraisamy, B. (eds) *Intelligent Computing and Innovation on Data Science*. Lecture Notes in Networks and Systems, vol 248. Springer, Singapore. https://doi.org/10.1007/978-981-16-3153-5_53
65. C. Diwaker et al., "A New Model for Predicting Component-Based Software Reliability Using Soft Computing," in *IEEE Access*, vol. 7, pp. 147191-147203, 2019, doi: 10.1109/ACCESS.2019.2946862.
66. Ahmed, Q.W.; Garg, S.; Rai, A.; Ramachandran, M.; Jhanjhi, N.Z.; Masud, M.; Baz, M. AI-Based Resource Allocation Techniques in Wireless Sensor Internet of Things Networks in Energy Efficiency with Data Optimization. *Electronics* 2022, 11, 2071. <https://doi.org/10.3390/electronics11132071>
67. Aldughayfiq B, Ashfaq F, Jhanjhi NZ, Humayun M. Explainable AI for Retinoblastoma Diagnosis: Interpreting Deep Learning Models with LIME and SHAP. *Diagnostics*. 2023; 13(11):1932. <https://doi.org/10.3390/diagnostics13111932>.