e-ISSN 2549-7472
Edumatic: Jurnal Pendidikan Informatika
Vol. 9 No. 2, Agustus, 2025, **Hal. 580-588**
DOI: 10.29408/edumatic.v9i2.30441

Community Detection in Scientific Collaboration Networks using the Louvain Method

Ghina Briliana Fatin Octariana [1, *], Heri Wijayanto [1], I Gede Darmawan Adi Pratama Putra [1]

[1] Department of Informatics Engineering, Universitas Mataram, Indonesia
[*] Correspondence: ghinabrilianafatinocktariana@gmail.com

Received: 23 May 2025 | Revised: 10 June 2025 | Accepted: 11 Agustus 2025 | Published: 17 August 2025

Abstract
Collaboration among authors in scientific publications reflects the dynamics and productivity of an institution's research. Understanding these collaboration patterns is important for supporting decision-making in research management and scientific policy development. This study aims to identify author collaboration patterns in academic publications using the Louvain algorithm. This study is quantitative descriptive research that uses metadata from 500 scientific publications collected from an academic institution in Mataram between 2021 and 2024. The data consist of author information, which were modeled as a collaboration graph. The analysis was conducted using the Louvain algorithm to detect communities, and the modularity score obtained was 0.7689, indicating a strong community structure, which reflects a strong community structure. Additionally, centrality metric calculations reveal the strategic role of authors as collaboration hubs, group connectors, or information disseminators. Three centrality metrics were used: degree centrality, betweenness centrality, and closeness centrality. These findings also demonstrate the effectiveness of Louvain in mapping scientific collaboration and can be utilized to support collaboration planning, expertise mapping, and decision-making based on scientific publication data.

Keywords: centrality; collaboration; clustering; detection community; louvain

INTRODUCTION

Scientific publications play an important role in the development of science today. Published research may introduce new findings or corroborate previous research results(Jafar et al., 2022). In addition, scientific publications are one of the main indicators in assessing the performance and academic career of a researcher (Masud et al., 2020). According to data from Scimago Journal & Country Rank, the number of scientific publications from Indonesia has increased significantly, reaching 10,594 publications since 2020. However, this data reflects publications at the national and international levels that are indexed in specific databases, not the entirety of publications at the institutional level. Therefore, in this study, the focus of the data used is limited to the metadata of scientific publications originating from the faculty environment of one university in the city of Mataram. This is done to ensure that the analysis remains relevant and does not generalize the overall condition of national scientific publications.

As the trend of scientific publication increases, collaboration between researchers has also become increasingly complex, resulting in dense and dynamic collaboration networks (Masud et al., 2020; Yu et al., 2024). Clustering is needed, especially in understanding who tends to work together, the frequency of collaboration, and the structure of the collaboration network itself (Xiong & Fan, 2021). Clustering techniques usually stand alone as a very useful technique in machine learning due to their ability to detect patterns in a set of data that are not visible (Elgazzar et al., 2021)Sarker, 2021; Zhou et al., 2024). So, it is necessary to find how the collaboration pattern between authors of scientific publications in a community. Not only that, community detection is an important research field in network analysis because it allows to find out the basic structure of complex networks.

This study stems from the limited understanding of collaboration patterns among authors in scientific publications, which has an impact on the suboptimal research management strategies and publication policies. To address this issue, this study proposes the application of the Louvain algorithm, a community detection method based on modularity optimization that is capable of identifying collaborative groups in a network without the need to determine the number of communities in advance (Grass-Boada et al., 2025; Wang et al., 2019). The visualization that will be generated using the Louvain method is a graph. Graphs help make it easier to analyze the resulting collaboration patterns (’Mao & ’Wu, 2024; Yi et al., 2022). From this method, it is expected to provide optimal comparison results to analyze and visualize the network of authors of scientific publications.

Theoretically and methodologically, the main contribution of this research is the application of the Louvain algorithm to detect author communities in scientific collaboration networks using a modularity optimization approach. Practically, centrality metrics such as degree centrality, betweenness, and closeness are used as quantitative inputs in a decision support system (DSS). DSS is a computer-based system that assists decision-makers in processing large-scale data, providing recommendations, and presenting information to support more accurate and strategic decision-making (Riehle et al., 2024; Wang et al., 2019). DSS assists decision makers by processing large amounts of data, providing recommendations and information that help in making the right choices (Moslem & Pilla, 2023).

There have been several previous studies analyzing community detection. Bhattacharya et al. (2023) Developing a community detection model based on node embedding and graph theory, capable of efficiently handling large-scale networks. The results show that the model can identify communities with significant accuracy and handle large-scale networks more effectively.

Meanwhile, Gupta & Singh (2022) developed the Clique-Based Louvain Algorithm (CBLA) to detect overlapping communities in synthetic graphs. The results show that the CBLA algorithm is capable of classifying vertices that cannot be classified in the Clique Percolation Method (CPM). Another relevant research was conducted by Paul et al. (2025) introduced a model called Commerce Graph to analyze the interactions between buyers and products in the digital commerce ecosystem. They used the Louvain algorithm to detect the community structure in the transaction network. Their findings suggest that nodes with high centrality values play a strategic role and can be targeted in marketing strategies. The integration of community detection and centrality-based segmentation in the model provides a deeper understanding of consumer and product relationships and enables the formulation of more targeted business strategies.

Most prior studies rely on bibliometric indicators or co-authorship counts without analyzing the deeper structure of collaboration networks. The use of network metrics and community detection algorithms like Louvain remains limited, despite their potential to reveal meaningful and practical collaboration patterns. This research aims to map collaboration structures, detect author communities using Louvain, and evaluate its real-world applicability. The findings can support strategic decisions in academic institutions, such as identifying faculty expertise and forming effective research teams.

METHOD

This study began with a literature review to understand the theories and concepts underlying the topic, through searches of journals, publications, and other relevant sources. Next, data in the form of an Excel file with 416 rows and 13 columns was obtained from an academic institution in Mataram, Indonesia. The data contains information about the authors of scientific publications, ranging from the chairperson's data to the data of members 1 to 12, which will be processed through the data preprocessing stage. The data needs to be processed first because a well-processed dataset can produce better results. Once the data is ready, the method is implemented in the computation stage and then continues to the analysis stage.

Three centrality metrics are used to determine collaboration patterns among authors of scientific publications. The three-centrality metrics are Degree Centrality, Betweenness Centrality, and Closeness Centrality. The method used in the computational stage is the Louvain method. The Louvain method is a community detection method based on the calculation of modularity values. To identify the optimal division of communities in the network, the Louvain method will seek to maximize modularity. The effectiveness of the Louvain algorithm in detecting the modular structure of a collaboration network lies in grouping authors into communities based on the high level of connections between community members. In the louvain algorithm, the modularity between each node will be calculated, modularity that is close to 1 will be combined and so on. The equation for louvain can be seen in equation (1).

$$Q = \frac{1}{2m}\sum_{i,j}\left(A_{ij} - \frac{K_i K_j}{2m}\right)\delta(c_i, c_j) \quad (1)$$

In equation (1) modularity, denoted by Q, is a measure used to evaluate the density of links within a cluster. The value of this modularity varies between -1 to 1, where higher values indicate that there are more links within clusters compared to the links connecting between clusters. In this case, Aij represents the weight of the edge connecting vertices i (i) and j (j). Meanwhile, Ki denotes the total weight of all the edges connected to vertex i. In addition, Ci refers to the cluster to which vertex i is assigned. The function δ (u, v) (δ (*u, v*)) has a value of 1 if u is equal to v, and 0 otherwise.

Then m is used to calculate the total number of weights in the graph, the previously described process will proceed iteratively until there is no more increase in the modularity value. m is defined as listed in equation (2) (Tang et al., 2021). The value of m is very important in calculating modularity in the Louvain algorithm, because it is used as a normalization factor to measure how dense the connections within a community are compared to random connections across the entire network.

$$m = \frac{1}{2}\sum_{ij} A_{ij} \quad (2)$$

RESULT AND DISCUSSION

Result

The preprocessing of the data resulted in clean data, in which each author's name was converted into a unique 4-digit code to facilitate further computation. This data was then processed to generate network visualizations and metric calculations aimed at revealing patterns of collaboration between authors in scientific publications while evaluating the performance of the Louvain method in detecting community structures.

Figure 2 shows the visualization results using the Louvain method, which successfully identified communities in a structured manner. There is a clear separation between large groups, while small communities maintain connections with other communities. This demonstrates the effectiveness of Louvain in detecting the modular structure of collaboration networks. Larger nodes indicate that authors play a more central role in the network, as they have more connections or collaborations than other authors, as can be seen in Table 1. Different colors in each node group represent communities or collaborative groups detected using the Louvain algorithm.

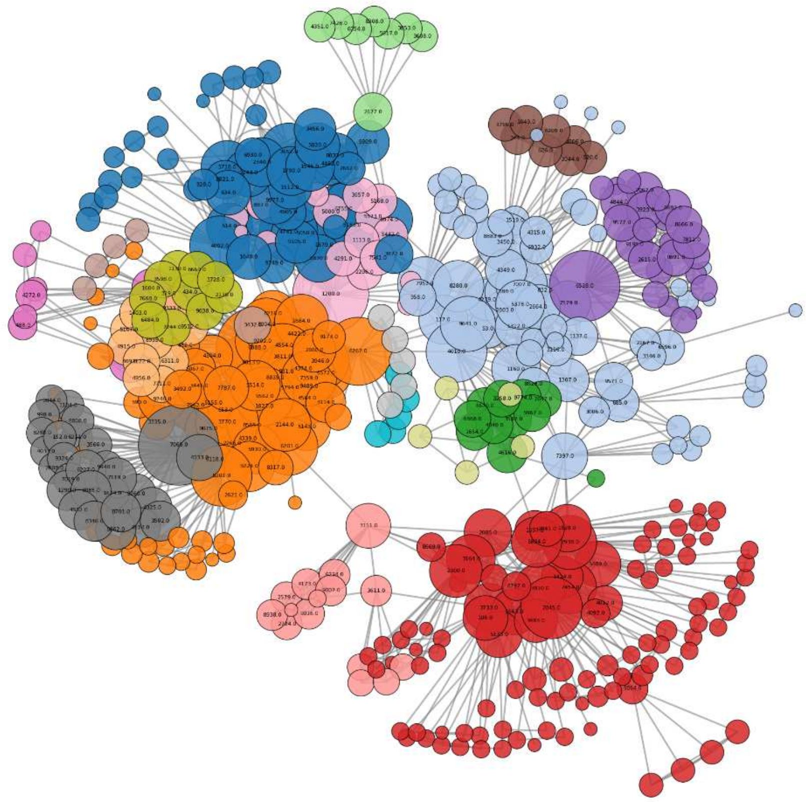

Figure 2. Louvain visualization results

To determine collaboration among authors of scientific publications, three centrality metrics were used: degree centrality, betweenness centrality, and closeness centrality. These metrics not only describe the strategic position of individuals in the network, but also show the extent to which an author can influence the flow of knowledge and connectivity between communities. Identifying central authors helps in understanding key actors in the dissemination of ideas, building cross-institutional collaboration, and detecting centers of excellence in specific fields. Thus, these results contribute to the strategic mapping of scientific networks and the development of data-driven research policies. The results of the three-centrality metrics degree, betweenness, and closeness are presented in Table 1 to highlight the most influential authors within the collaboration network.

Table 1. Result metric centrality

ID Author	***Degree***	***ID Author***	***Betweenness***	***ID Authors***	***Closeness***
P107	0,11	P177	0,23	P174	0,31
P119	0,11	P423	0,22	P432	0,31
P202	0,09	P093	0,16	P093	0,31
P109	0,08	P075	0,14	P107	0,31
P117	0,08	P202	0,14	P182	0,31

Table 1 shows the results of measuring three centrality metrics, namely degree, betweenness, and closeness, to identify key authors in scientific collaboration networks. Based on the results of the centrality metrics, it can be identified that several authors show consistent high values on more than one centrality metric. For example, the author with ID P107 ranks highest in the degree centrality metric (0.11) and fourth in closeness centrality (0.31). This indicates that P107 is not only active in collaboration (having many direct connections) but also has a high average closeness to other authors, thereby playing a crucial role in reaching various parts of the network.

Other authors like P202 also hold significant positions in degree (0.09) and betweenness (0.14), indicating that they are actively involved in collaboration while also serving as a bridge between author communities. Meanwhile, P093 has high values for betweenness (0.16) and closeness (0.31), indicating its ability to bridge communities while maintaining efficient access to the entire network. On the other hand, there are authors like P119 who have a high degree (0.11) but do not appear in other metrics, indicating high collaborative activity but limited to one community.

In contrast, P423 has a high betweenness value (0.22) without appearing in the degree or closeness metrics, indicating its role as a connector between communities even though it does not have many direct connections. These differences show that each metric provides a different perspective on the role of authors in scientific collaboration networks.

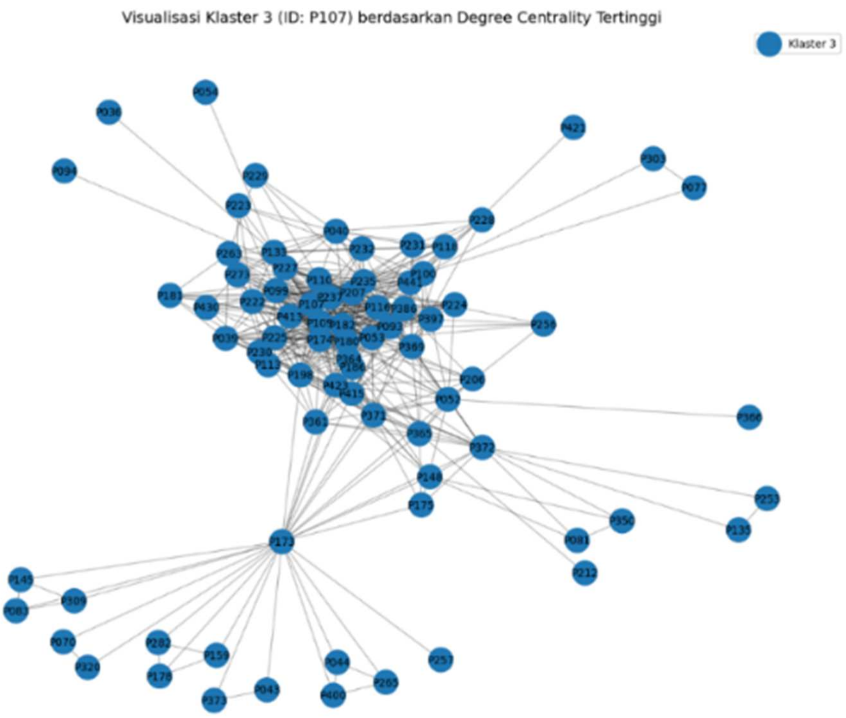

Figure 3. Degree centrality visualization

The degree centrality metric shows how many direct connections an author has in the network. From the results, the author with ID P107 in cluster three has the highest degree centrality value of 0.11, which indicates that he is the author most involved in direct collaboration with other authors. In this case, the author with ID P107 in cluster three has the highest degree centrality value of 0.11, which means that he has direct connections with approximately 10.7% of all authors in the network. Followed by P119 (0.11), P202 (0.09), P109 (0.08) and P117 (0.08) also show a high level of involvement in collaboration, as illustrated in the degree centrality visualization in figure 3.

Then, betweenness centrality describes how often an author acts as a liaison or intermediary between other author groups. The author with ID P177 has the highest betweenness centrality, which is 0.23, meaning that he is an intermediary point in approximately 23.1% of all shortest paths between pairs of authors. This indicates that P177 plays a strategic role as a bridge between clusters in the collaboration network, as visualized in the Betweenness Centrality graph in Figure 4.

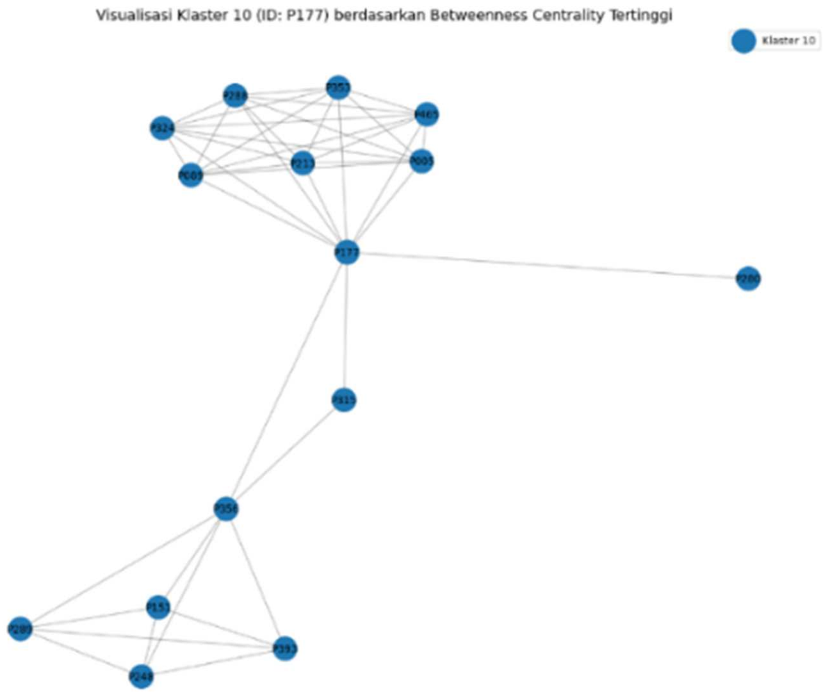

Figure 4. Betweenness centrality visualization

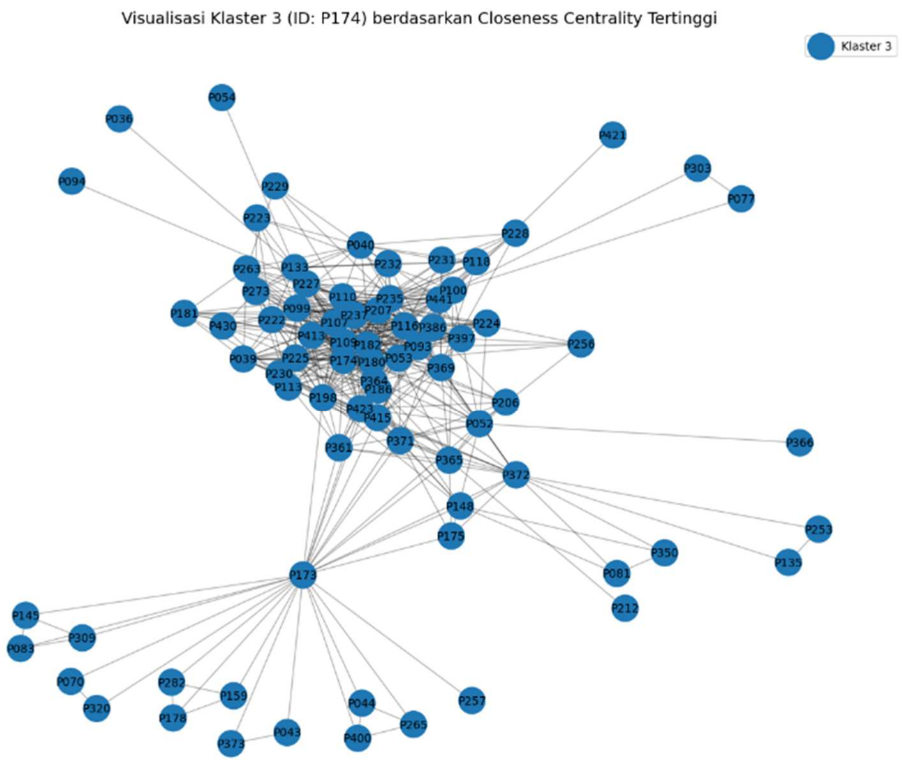

Figure 5. Closeness centrality visualization

The last one is closeness centrality which will measure how close or fast an author can reach all other authors in the network. Author with ID P174 recorded the highest closeness value of 0.31, this value can be interpreted as P174 being approximately 31% of the way toward perfect accessibility, assuming a fully connected network would yield a value close to 1 which means he has a strategic position in the network and can spread information efficiently. This is followed by P432 (0.31), P093 (0.31), P107 (0.31), and P182 (0.31), who also demonstrate fast access to other authors in the network. Overall, the table highlights key authors in the collaboration network, with the Closeness Centrality visualization presented in Figure 5.

Each node represents a writer with a unique ID, the size of the node indicates the degree centrality value, the edges represent direct collaborative relationships between writers, and the colors indicate the communities detected by Louvain. To measure the quality of the communities formed, the Modularity Score value is used, because the Louvain algorithm is designed based on modularity optimization. The resulting Modularity Score is 0.7689, indicating a fairly good ability to identify community structures in the network because the closer the Modularity Score value is to 1, the better the community structure formed. Therefore, the value of 0.7705 signifies that the modular structure in this network is very strong, even exceeding many similar studies.

Overall, the visualization results show that this method is able to group nodes optimally, with a fairly dense community and clear boundaries between each other. Based on visual observation and node distribution, communities tend to consist of authors who frequently appear in the same publications, originating from similar affiliations or research interests. This indicates that the Louvain algorithm not only forms groups structurally, but also reflects the reality of scientific collaboration that is thematic or institutional in nature. Thus, the high modularity values achieved are not only numerically valid but also conceptually meaningful, as they capture collaboration patterns that are indeed relevant in the context of the academic world and research organizations.

Discussion

This study applies the Louvain algorithm to detect communities in a collaboration network of scientific publication authors based on real data from the faculty. Louvain was chosen for its superiority in optimizing modularity without the need to determine the number of clusters at the outset, as well as its efficiency for medium-scale data. The modularity value obtained was 0.7689, indicating a strong community structure, and higher than several similar studies such as Tunali (2021) which reports modularity of 0.63–0.70. This approach differs from embedding methods such as node2vec or GNN used in large-scale studies, and is more stable than alternative methods such as Label Propagation or Girvan-Newman, which tend to form small and fragmented communities.

The findings of this study reveal different roles among authors. For example, P093 has high betweenness and closeness values as a connector between communities, while P107 has a high degree but a low connector role, reflecting a hub-and-spoke pattern. Additionally, P177 appears to be important as a bridge between clusters, and P174 is at the center of the network with quick access to other nodes. These findings reinforce the scientific contribution in local bibliometric studies, particularly in mapping key actors in scientific collaboration.

The practical implications of these findings include mapping expertise, forming potential research groups, and making data-driven collaboration policies. However, limitations remain, such as potential bias from using publication metadata without information on individual contributions or temporal dynamics. For future development, integration with topic, citation, or temporal data is recommended to improve the validity and relevance of the results.

Compared to the study by Bhattacharya et al. (2023), which uses a node embedding approach to model large-scale networks, this research emphasizes the actual representation of local collaboration networks through static graphs. Although node embedding excels in complex networks, the Louvain-based approach and centrality metrics are considered more stable and suitable for the limited bibliometric context. Meanwhile, Gupta & Singh (2022) developed the CBLA to detect overlapping communities using synthetic graphs, whereas this study chose not to use modifiable graphs as they were deemed less reflective of the actual complexity of scientific collaboration. Instead, static graphs were used to represent collaboration based on actual relationships between authors. This study aims to identify collaboration structures, detect communities, and evaluate the effectiveness of the Louvain algorithm in a real-world data context, contributing to expertise mapping, research team formation, and data-driven collaboration policy development. Moving forward, integration with topic, citation, or temporal dimension data is recommended to enhance the validity and relevance of findings

CONCLUSION

This study successfully identified patterns of collaboration among authors of scientific publications within the Faculty of Engineering through the application of an optimized Louvain algorithm based on modularity. A modularity value of 0.7689 indicates a strong community structure and confirms the effectiveness of this method in mapping collaborative groups based on author metadata. Centrality analysis further enriches our understanding of the strategic roles of authors within the network, ranging from community connectors, collaboration hubs, to actors with rapid access to various clusters. The primary scientific contribution of this study lies in the application of the Louvain method to local data, which has been rarely discussed in bibliometric studies, as well as the integration of modularity and centrality analysis within a single graph-based visual framework. This approach offers a replicable analytical model for mapping collaboration in other educational institutions. However, limitations remain, particularly due to the use of static publication metadata that does not fully represent individual contributions or the temporal dynamics of collaboration. Therefore, further research is recommended to integrate topic data, citation networks, and temporal dimensions to produce a more comprehensive and contextual mapping of collaboration.

REFERENCES

Bhattacharya, R., Nagwani, N. K., & Tripathi, S. (2023). A community detection model using node embedding approach and graph convolutional network with clustering technique. *Decision Analytics Journal*, *9*, 100362. https://doi.org/10.1016/j.dajour.2023.100362

Bilal, M., Ali, G., Iqbal, M. W., Anwar, M., Malik, M. S. A., & Kadir, R. A. (2022). Auto-Prep: Efficient and Automated Data Preprocessing Pipeline. *IEEE Access*, *10*, 107764–107784. https://doi.org/10.1109/ACCESS.2022.3198662

Elgazzar, H., Spurlock, K., & Bogart, T. (2021). Evolutionary clustering and community detection algorithms for social media health surveillance. *Machine Learning with Applications*, *6*, 100084. https://doi.org/10.1016/j.mlwa.2021.100084

Grass-Boada, D. H., González-Montesino, L., & Armañanzas, R. (2025). Community detection in networks: A rough sets and consensus clustering approach. *Applied Soft Computing*, *178*, 113219. https://doi.org/10.1016/j.asoc.2025.113219

Gupta, S. K., & Singh, D. D. P. (2022). CBLA: A Clique Based Louvain Algorithm for Detecting Overlapping Community. *Procedia Computer Science*, *218*, 2201–2209. https://doi.org/10.1016/j.procs.2023.01.196

Lu, M., Perez, E. J., & Mason, K. (2025). Complex network analysis of China's integrated air-high-speed rail network: Topological characteristics, centrality measures, and cluster analysis. *Journal of Transport Geography*, *127*, 104292. https://doi.org/10.1016/j.jtrangeo.2025.104292

Ma, J., & Fan, J. (2020). Local Optimization for Clique-Based Overlapping Community Detection in Complex Networks. *IEEE Access*, *8*, 5091–5103. https://doi.org/10.1109/ACCESS.2019.2962751

Mao, M., Wu, Z., Xu, D., Xu, J., & Hu, Q. (2021). Community-detection-based Approaches for Distribution Network Partition. *CSEE Journal of Power and Energy Systems*, 10(5), 1965-1976.

Masud, N., Masuadi, E., Moukaddem, A., Omair, A., Mohamud, M., Al Dubayee, M., Althubaiti, A., Alnamshan, M. K., Bawazeer, M., & AlJasser, M. I. (2020). Development and Validation of Authorship Order Score (AOS) for Scientific Publication. *Health Professions Education*, *6*(3), 434–443. https://doi.org/10.1016/j.hpe.2020.04.006

Moslem, S., & Pilla, F. (2023). Addressing last-mile delivery challenges by using euclidean distance-based aggregation within spherical Fuzzy group decision-making. *Transportation Engineering*, *14*. https://doi.org/10.1016/j.treng.2023.100212

Paul, S., Chakraborty, A., Dutta, S., Ghosh, K. Das, & Mitra, A. (2025). Leveraging the Commerce Graph for Optimizing Digital Commerce. *Procedia Computer Science*, *259*, 1072–1081. https://doi.org/10.1016/j.procs.2025.04.061

Riehle, D. M., Wolters, A., & Muller, K. (2024). Adopting Artificial Intelligence in a Decision Support System - Learnings from Comment Moderation Systems. *Procedia Computer Science*, *239*, 1847–1855. https://doi.org/10.1016/j.procs.2024.06.366

Sarker, I. H. (2021). Machine learning: Algorithms, real-world applications and research directions. *SN computer science*, *2*(3), 160. https://doi.org/10.1007/s42979-021-00592-x

Tang, Y., Wang, B., & Wang, P. (2021). RefineDCN: An Improved Community Detection Algorithm Based on Center Finding. *IEEE Access*, *9*, 14630–14639. https://doi.org/10.1109/ACCESS.2020.3048187

Tunali, V. (2021). Large-Scale Network Community Detection Using Similarity-Guided Merge and Refinement. *IEEE Access*, *9*, 78538–78552. https://doi.org/10.1109/ACCESS.2021.3083971

Wang, Y., Yuan, Y., Ma, Y., & Wang, G. (2019). Time-Dependent Graphs: Definitions, Applications, and Algorithms. In *Data Science and Engineering* (Vol. 4, Issue 4, pp. 352–366). Springer Science and Business Media Deutschland GmbH. https://doi.org/10.1007/s41019-019-00105-0

Xiong, H., & Fan, Y. (2021). How to better identify venture capital network communities: Exploration of a semi-supervised community detection method. *Journal of Social Computing*, *2*(1), 27–42. https://doi.org/10.23919/JSC.2020.0008

Yi, H. C., You, Z. H., Huang, D. S., & Kwoh, C. K. (2022). Graph representation learning in bioinformatics: trends, methods and applications. *Briefings in Bioinformatics*, *23*(1), bbab340. https://doi.org/10.1093/bib/bbab340

Yu, G., Jiao, Y., Dehmer, M., & Emmert-Streib, F. (2024). Community detection in directed networks based on network embeddings. *Chaos, Solitons & Fractals*, *189*, 115630. https://doi.org/10.1016/j.chaos.2024.115630

Zhou, S., Xu, H., Zheng, Z., Chen, J., Li, Z., Bu, J., ... & Ester, M. (2024). A comprehensive survey on deep clustering: Taxonomy, challenges, and future directions. *ACM Computing Surveys*, *57*(3), 1-38. https://doi.org/10.1145/3689036