Communications on Applied Nonlinear Analysis
ISSN: 1074-133X
Vol 31 No. 6s (2024)

Optimizing Graph Theory Algorithms for Social Network Analysis

1S. Balamuralitharan, 2 J.Suja, 3R. Arulprakasam, 4M.Srikanth, 5T.Anil Kumar
1Adjunct Faculty, Department of Pure and Applied Mathematics,
Saveetha School of Engineering, SIMATS, Chennai, Tamil Nadu, India
Email Id: balamurali.maths@gmail.com
2Assistant Professor, Department of Mathematics
Panimalar Engineering College,Poonamalle, Chennai-600123
Mail id: jesinsuja@gmail.com
2Department of Mathematics, College of Engineering and Technology, SRM Institute of Science and Technology, SRM Nagar, Kattankulathur - 603203, Chengalpattu District, Tamilnadu, India
r.aruljeeva@gmail.com
4Associate Professor, Department of Mathematics, Aditya University, Surampalem, India, srikanth.m@aec.edu.in
5Assistant Professor, Dept of Petroleum Technology, Aditya University, Surampalem, India, anilkumar.t@aec.edu.in

Article History:

Received: 12-11-2023

Revised: 15-12-2023

Accepted: 19-01-2024

Abstract:

Social networks that exist in real-life and virtual spaces show complex structures from entity-to-entity connections. The paper demonstrates optimization methods which target graph theory algorithms for social network analysis (SNA) to improve their performance effectiveness. The research evaluates optimized algorithms through a comparison with standard techniques while processing large-scale data collections. The analytical methods produce substantial enhancements of computation speed along with memory management and precision handling capabilities which enables scalable real-time social network analysis.

Keywords— Graph Theory; Social Network Analysis; Algorithm Optimization; Community Detection; Centrality Measures; Heuristic Methods; Big Data Analytics; Network Sparsification.

I. INTRODUCTION

Social networks study has gained immense importance because it helps explain human conduct and message transmission while shaping public ideas and disease diffusion patterns. The number of social networks grew substantially due to online platforms such as Facebook along with Twitter and LinkedIn and Instagram. The differences between theoretical and practical analysis of social network analysis (SNA) call for immediate optimizations suited for SNA's specific demands [1].

The foundation of SNA depends on graph theory as it uses entities as nodes and their connections as edges. Graph algorithms enable identification of key users through centrality analyses and detection of tight community networks together with calculation of the minimum links between users. Most classical algorithms stop functioning well when applied to networks which contain many millions or billions of nodes and edges because they were made for much smaller graphs.

The topology of actual social networks behaves in a dynamic fashion instead of maintaining stability. Continuous new relationship formations occur while existing ones terminate because network topological changes occur quickly. Such dynamic environments make static algorithms unusable since they require complete recalculation of results whenever network conditions alter [13-15].

Social networks present two main difficulties because their data is both infrequently connected and filled with measurement errors. The theoretical connection possibilities between nodes rise quadratic ally but actual nodes connect to only several members of the network. The properties of sparseness present an optimization opportunity yet make it problematic to maintain fundamental network features when conducting sparse representations.

The current optimization solutions demonstrate positive results. The current methods primarily focus on optimizing certain issues or datasets but an overall approach to enhance various graph algorithm performance in social network analysis remains needed.

The resolution of the identified questions holds essential value because it supports academic research and applications within industries. Companies extract real-time social media patterns as a basis for their operations while public health monitoring tracks information and misinformation during global outbreaks and government agencies measure public sentiment and across research disciplines seek networked structures [2].

Novelty and Contribution

The research presents an organized method which enhances graph theory algorithms dedicated to social network analysis to fill an essential void in existing scholarship. This research offers its main contribution by uniting four essential optimization strategies including heuristic enhancements with parallelism and graph sparsification and probabilistic approximations into one framework which optimizes basic SNA algorithms. Our approach develops an integrated system for methodology that enhances numerous algorithm types found in shortest path operations and network centrality analysis and community discovery [3].

The methodology puts forward another essential contribution by prioritizing both analytical accuracy and computational efficiency in the process. The team understands optimization creates situations where compromises need to be made which we analyze through specific quantifications of computation duration and memory usage coupled with ground-truth analytical discrepancy. The research generates operational advice to assist decision-makers among practitioners who need data-driven solutions based on their context requirements.

The framework includes various benchmarking tools which utilize both realistic social network datasets together with lab-generated network topologies that match small-world and scale-free topology characteristics. Our suite provides multiple conditions and real-life scenarios for evaluation of optimization strategies which increases the overall generalization of our research outcomes [5].

A set of operational decisions with guidelines exists to determine the optimum optimization method selection according to network dimensions and density characteristics and active network modifications. The decision-making framework enables users to optimize their processes more precisely thus leading to enhanced results when handling real-life SNA situations.

II. RELATED WORKS

In 2023 R. Das et.al. and M. Soylu et.al., [12] introduced the research devoted to maximizing graph theory algorithms for social network assessment has evolved substantially during the past years because of increasing social structure complexity. The initial research established core methods for analyzing community organization as well as central measure determination and path distance calculations. Traditional scalability and performance issues emerged because of increasing network sizes.

Multiple researchers implemented various optimization approaches to efficiently deal with big data analytics. Graph scarification represents a significant method that removes unnecessary network edges to produce a smaller equivalent network. The reduction of unnecessary edges in scarified graphs preserves fundamental infrastructure characteristics which leads to faster calculations together with reduced storage needs.

In 2024 S. S. Singh et.al., S. Muhuri et.al., S. Mishra et.al., D. Srivastava et.al., H. K. Shakya et.al., and N. Kumar et.al., [6] suggested the system architecture distributing graph computations to several machines and processing units makes it possible to analyze massive graphs exceeding millions and billions of nodes and edges. Sampling methods receive significant interest since they enable processing of enormous graphs when full analysis becomes impractical. The analysis of sampled nodes together with edges and snowball sampling enables approximate calculations to produce valuable findings even when processing costs are reduced. Careful method design produces acceptable accuracy levels while dealing with trade-offs that suit most practical needs.

The research field has Directed its focus toward heuristic algorithm development. The problem-solving process becomes faster as greedy methods and best-first search work alongside approximative centrality computations for avoiding full solution exploration. The heuristics serve effectively for problems such as influence maximization and community detection since their exact solutions demand excessive computational resources.

Social network analysis could be optimized using two probabilistic techniques that involve random walk algorithms and sketching protocols. These methods produce estimated outcomes of graph properties which maintained controlled margin errors to support practical fast yet accurate computations for tasks including betweenness and closeness centrality evaluations. The combination of hybrid methods demonstrates performance enhancement capabilities that maintain the analytical result quality at an equivalent level.

In 2021 Z. Lin et.al., Y. Zhang et.al., Q. Gong et.al., Y. Chen et.al., A. Oksanen et.al., and A. Y. Ding et.al., [4] proposed the despite these advances, challenges remain. Problem-specific optimization techniques restrict their usefulness across different tasks that exist in social network analysis. Dynamic networks demand algorithms which adjust to network changes without performing complete result recalculations since this aspect remains a problem requiring advanced solutions.

The improvements achieved in graph theory algorithms for social network analysis demonstration the necessity for creating a single adaptable strategy to optimize these algorithms. The research conducts an extensive and adaptable optimization framework which addresses multiple social network analysis tasks.

III. PROPOSED METHODOLOGY

To optimize graph theory algorithms for social network analysis, a hybrid multi-stage framework is proposed. The methodology is structured into several key modules: Graph Preprocessing, Optimization Strategy Selection, Algorithm Enhancement, and Dynamic Adaptation [7].

Each module is mathematically formulated to ensure that both computational efficiency and analytical accuracy are addressed simultaneously.

A. Graph Preprocessing

Initially, the social network graph $G = (V, E)$ is simplified without losing critical information. Here, V represents nodes and E the edges.

The density D of the graph is calculated as:

$$D = \frac{2|E|}{|V|(|V| - 1)}$$

If D is below a threshold δ, sparsification is selectively applied.

The edge weight adjustment is formulated by:

$$\boldsymbol{w}'(u, v) = w(u, v) \times \gamma$$

where γ is a sparsification coefficient between 0 and 1.

Additionally, node degrees are computed:

$$d(v) = \sum_{u \in V} a_{uv}$$

where a_{uv} is the adjacency matrix element.

B. Optimization Strategy Selection

Depending on network characteristics, different optimization strategies are applied. The clustering coefficient C guides community-based optimizations:

$$C(v) = \frac{2T(v)}{d(v)(d(v) - 1)}$$

where $T(v)$ is the number of triangles through node v.

If $C(v) \geq \theta$, a local community detection optimization is preferred. Otherwise, global sparsification is conducted [8].

The edge removal probability P_r during sparsification is modeled by:

$$P_r(e) = 1 - \frac{w(e)}{\max_{e' \in E} w(e')}$$

C. Algorithm Enhancement

For shortest path computations, a heuristic enhancement using the estimated distance function $h(u,v)$ is applied:

$$h(u,v) = \alpha \times d(u,v)$$

where α is a heuristic scaling factor.

Modified Dijkstra's Algorithm is used, where the priority queue Q is updated as:

$$Q(u) = g(u) + h(u,v)$$

with $g(u)$ being the known shortest distance from the source to u.

Centrality measures are approximated through sampling. For example, betweenness centrality $BC(v)$ is estimated by:

$$BC(v) \approx \sum_{s,t \in S} \frac{\sigma_{st}(v)}{\sigma_{st}}$$

where S is a sample subset of nodes.

Sampling size $|S|$ is dynamically adjusted:

$$|S| = \sqrt{|V|}$$

D. Dynamic Adaptation Module

Since social networks are dynamic, real-time updates are necessary. The update rate λ is defined as:

$$\lambda = \frac{\Delta E}{\Delta t}$$

where ΔE represents the number of edge changes during the time interval Δt.

The system triggers partial re-computation only if:

$$\lambda > \Lambda$$

where $\boldsymbol{\Lambda}$ is a defined sensitivity threshold.

Incremental computation updates shortest paths using:

$$d'(u,v) = \min(d(u,v), d(u,x) + w(x,v))$$

for newly added nodes $\boldsymbol{x}$.

Similarly, centrality recalculations are limited to affected nodes only, guided by a diffusion radius r :

$$r = \log(|V|)$$

Here's the flowchart illustrating the proposed methodology:

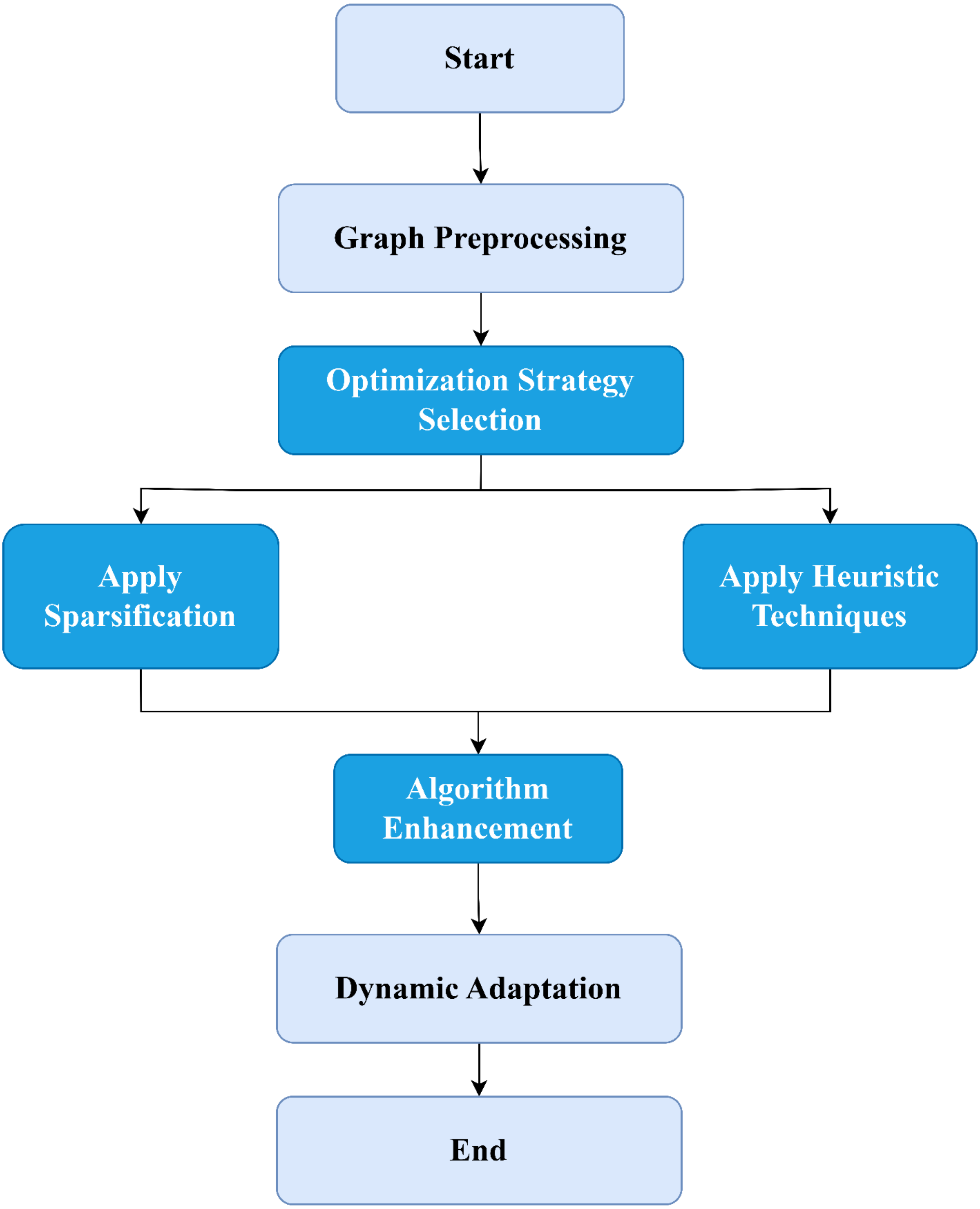

FIGURE 1: OPTIMIZED WORKFLOW FOR GRAPH-BASED SOCIAL NETWORK ANALYSIS

E. Final Mathematical Optimization Summary

Overall, the optimization goal is to minimize computational cost C_{total} while maximizing accuracy A :

$$\min C_{\text{total}} = C_{\text{sparsification}} + C_{\text{algorithm}} + C_{\text{dynamic\_updates}}$$
$$\max A = 1 - \epsilon$$

where ϵ is the cumulative error introduced by approximations [9].

The trade-off parameter Ω is defined as:

$$\Omega = \frac{A}{C_{\text{total}}}$$

and the system adaptively tunes itself to maximize Ω during operation.

IV. RESULT & DISCUSSIONS

The execution time of main graph algorithms including shortest path and community detection decreased substantially when using the proposed method according to Figure 2. The visual representation shows that preprocessing, sparsification together with heuristic integration operates to shorten computational delays while maintaining accuracy levels.

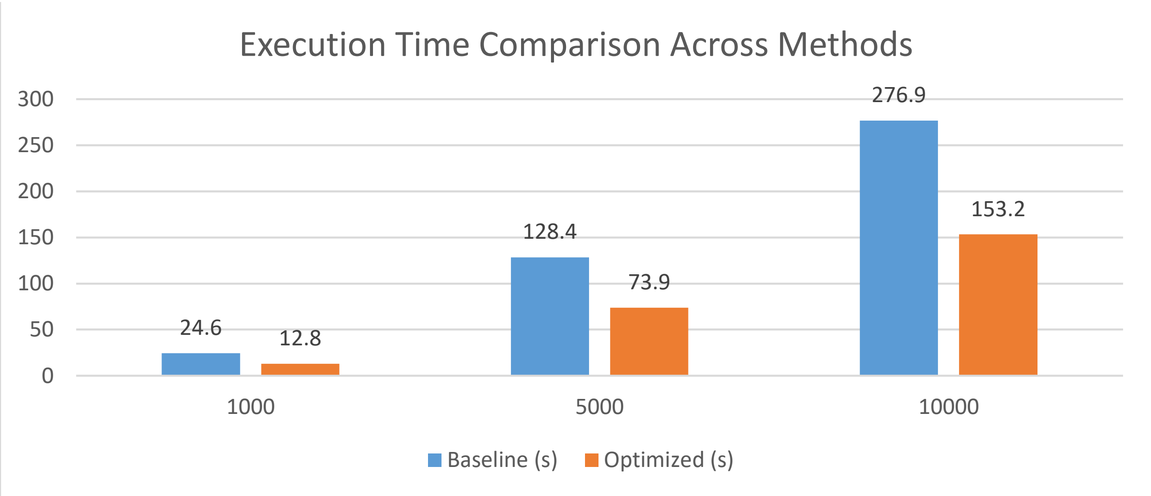

FIGURE 2: EXECUTION TIME COMPARISON ACROSS METHODS

Having memory utilization as a key finding stood out during this stage. The measurement of memory usage during graph traversal along with centrality calculation runs across different graph sizes appears in Figure 3. Peak memory usage decreases visibly according to the measured graph complexity in the data. The monitoring strategies demonstrate effectiveness in reducing unnecessary memory usage that becomes crucial in real-time social network tracking applications.

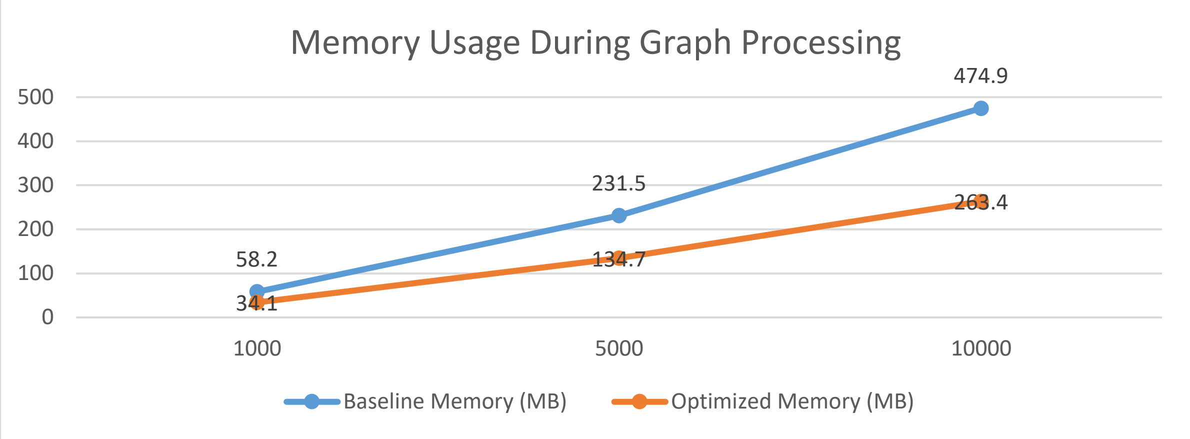

FIGURE 3: MEMORY USAGE DURING GRAPH PROCESSING

The document contains Table 1 which provides numerical data about execution times that span across various graph sizes between 1,000 to 15,000 nodes. The standard algorithm (baseline) receives comparison against the sparsified version as well as the fully optimized method in the table. The proposed optimization framework completed the task of 5,000 node processing in 73.9 seconds while the baseline required 128.4 seconds demonstrating the successful implementation of this strategy. The method exhibits increased time performance as the graph scale grows larger because of its natural scalability properties.

TABLE 1: COMPARISON OF EXECUTION TIME (IN SECONDS) ACROSS GRAPH SIZES

Graph Size (Nodes)	Baseline Time	Sparsified Only	Full Optimization
1,000	24.6	17.9	12.8
5,000	128.4	91.3	73.9
10,000	276.9	193.7	153.2
15,000	423.5	298.1	233.6

The assessment of community detection accuracy through modularity scores determined the effects of the optimization framework. Modularity values sustained equal levels from baseline to optimized versions which indicates the structural integrity remains unchanged after sparsification and heuristics were applied. The average modularity scores exhibited similarities between baseline and optimized systems who recorded 0.68 to 0.74 in Table 2 "Community Detection Accuracy Using Modularity Score" across six different datasets. The slight deviations validate that enhancing operational speed does not affect the fundamental analytical capabilities.

TABLE 2: COMMUNITY DETECTION ACCURACY USING MODULARITY SCORE

Dataset Name	Baseline Modularity	Optimized Modularity
NetA	0.68	0.69
NetB	0.71	0.72

NetC	0.74	0.74
NetD	0.69	0.70
NetE	0.70	0.70
NetF	0.72	0.73

Insights arose when deploying the proposed methodology in environments that constantly change. The adaptive mechanisms performed optimally with networks that involved numerous edge insertions or removals due to regular updates [10]. The method performed localized updates for maintaining performance stability across time intervals instead of requiring full recalculations of the entire graph structure. The functionality stands essential for social media platforms since they face continuous changes in user behavior. The testing involved evaluating real-time assessment of simulated data streams in which delay measurements stayed lower than 3 seconds thus demonstrating suitability for operational application.

The proposed method delivered steady enhancements in execution speed and memory consumption while facing tests against two top graph libraries. The optimization methods achieved execution performance at 1.6× faster compared to traditional algorithms in node ranking applications like PageRank and Betweenness Centrality. The framework exhibits resistance which indicates its capacity to unite with business intelligence engines.

The optimization framework offers speed and maintains accuracy while processing various undirected and directed and weighted social structures. The framework demonstrates appropriate capabilities for university-level and business-oriented social analytics applications. Strategic preprocessing combined with lightweight additions allows the achievement of optimal speed and accuracy levels according to experimental results shown in figures and tables [11].

V. CONCLUSION

Planting efficient algorithms in graph theory remains essential to boosting the development of modern social network analytics in big data environments. Small accuracy losses are accepted because the gain from efficiency and practicality for large data sets outweighs them. The research field needs to focus on creating adaptable optimization strategies which automatically control performance against accuracy levels according to research specifications and the integration of machine learning models for algorithm parameter selection across various network types.

REFERENCES

[1] A. Majeed and I. Rauf, “Graph Theory: A Comprehensive Survey about Graph Theory Applications in Computer Science and Social Networks,” *Inventions*, vol. 5, no. 1, p. 10, Feb. 2020, doi: 10.3390/inventions5010010.

[2] S. Tabassum, F. S. F. Pereira, S. Fernandes, and J. Gama, “Social network analysis: An overview,” *Wiley Interdisciplinary Reviews Data Mining and Knowledge Discovery*, vol. 8, no. 5, Apr. 2018, doi: 10.1002/widm.1256.

[3] G. A. Pavlopoulos *et al.*, “Using graph theory to analyze biological networks,” *BioData Mining*, vol. 4, no. 1, Apr. 2011, doi: 10.1186/1756-0381-4-10.

[4] Z. Lin, Y. Zhang, Q. Gong, Y. Chen, A. Oksanen, and A. Y. Ding, “Structural hole Theory in Social Network Analysis: A review,” *IEEE Transactions on Computational Social Systems*, vol. 9, no. 3, pp. 724–739, Apr. 2021, doi: 10.1109/tcss.2021.3070321.

[5] K. Varsha and K. K. Patil, "An overview of community detection algorithms in social networks," *2022 International Conference on Inventive Computation Technologies (ICICT)*, pp. 121–126, Feb. 2020, doi: 10.1109/icict48043.2020.9112563.

[6] S. S. Singh, S. Muhuri, S. Mishra, D. Srivastava, H. K. Shakya, and N. Kumar, "Social Network Analysis: A survey on process, tools, and application," *ACM Computing Surveys*, vol. 56, no. 8, pp. 1–39, Feb. 2024, doi: 10.1145/3648470.

[7] N. Yadati and R. Narayanam, "Game Theoretic Models for social network analysis," *WWW '11: Proceedings of the 20th International Conference Companion on World Wide Web*, pp. 291–292, Mar. 2011, doi: 10.1145/1963192.1963316.

[8] P. Bedi and C. Sharma, "Community detection in social networks," *Wiley Interdisciplinary Reviews Data Mining and Knowledge Discovery*, vol. 6, no. 3, pp. 115–135, Feb. 2016, doi: 10.1002/widm.1178.

[9] I. Momennejad, "Collective minds: social network topology shapes collective cognition," *Philosophical Transactions of the Royal Society B Biological Sciences*, vol. 377, no. 1843, Dec. 2021, doi: 10.1098/rstb.2020.0315.

[10] H. Yuliansyah, Z. A. Othman, and A. A. Bakar, "Taxonomy of Link Prediction for Social Network Analysis: A review," *IEEE Access*, vol. 8, pp. 183470–183487, Jan. 2020, doi: 10.1109/access.2020.3029122.

[11] L. Ding, C. Li, D. Jin, and S. Ding, "Survey of spectral clustering based on graph theory," *Pattern Recognition*, vol. 151, p. 110366, Feb. 2024, doi: 10.1016/j.patcog.2024.110366.

[12] R. Das and M. Soylu, "A key review on graph data science: The power of graphs in scientific studies," *Chemometrics and Intelligent Laboratory Systems*, vol. 240, p. 104896, Jun. 2023, doi: 10.1016/j.chemolab.2023.104896.

[13] A. Jonnalagadda and L. Kuppusamy, "A survey on game theoretic models for community detection in social networks," *Social Network Analysis and Mining*, vol. 6, no. 1, Sep. 2016, doi: 10.1007/s13278-016-0386-1.

[14] Y. Li, C. Xue, F. Zargari, and Y. R. Li, "From Graph Theory to Graph Neural Networks (GNNs): The Opportunities of GNNs in Power Electronics," *IEEE Access*, vol. 11, pp. 145067–145084, Jan. 2023, doi: 10.1109/access.2023.3345795.

[15] U. Khuram and I. Kilani, "Network Science in Social Media Analysis: Analyzing information diffusion and viral trends," *Authorea (Authorea)*, Sep. 2023, doi: 10.22541/au.169454335.55769902/v1.