© AUG 2026 | IRE Journals | Volume 10 Issue 2 | ISSN: 2456-8880
DOI: https://doi.org/10.64388/IREV10I2-1722315

Graph Analytics for Social Networks

SACHIN SHARMA
Department of Engineering and Technology, Jagannath University, Jaipur, India

Abstract- ***Social networks generate vast amounts of relational data whose value lies not in individual data points but in the structure of connections among them. Graph analytics offers a mathematically grounded toolkit for uncovering this structure, ranging from simple degree counts to sophisticated community-detection and influence-propagation models. This paper presents a self-contained treatment of graph analytics as applied to social networks. We review the foundational graph-theoretic concepts underlying social network analysis, survey the principal families of graph analytics methods, and examine community detection and influence analysis in depth. To ground the discussion empirically, we conduct a case study on Zachary's Karate Club network and a synthetically generated scale-free network, computing centrality measures, detecting communities using the Louvain algorithm, and analysing degree-distribution behavior. The Louvain method partitions the Karate Club network into four communities with a modularity of 0.4266, closely matching the network's known factional split, while the synthetic network exhibits an approximate power-law degree distribution with exponent 1.76, consistent with preferential-attachment growth. We conclude with a discussion of open challenges and directions for future research, including dynamic graph analytics, scalability to billion-edge networks, and privacy-preserving analysis.***

Keywords: graph theory, social network analysis, community detection, Louvain algorithm, centrality, influence analysis, scale-free networks

I. INTRODUCTION

The rapid growth of online platforms such as Face book, X (formerly Twitter), LinkedIn, and Instagram has transformed social interaction into a form of structured data. Every friendship, follow, retweet, or professional endorsement can be represented as an edge connecting two entities in a graph, and the resulting networks now routinely contain hundreds of millions of nodes and billions of edges. Understanding the structure of these networks is no longer a purely academic exercise; it underlies applications as varied as targeted marketing, epidemic modelling, recommendation systems, fraud detection, and the study of misinformation spread.

Graph analytics is the discipline concerned with extracting insight from such relational data using the mathematical apparatus of graph theory. Unlike traditional tabular analytics, which treats records as independent observations, graph analytics explicitly models the dependencies between entities, allowing analysts to answer questions such as: Which individuals are most influential? Do natural communities or factions exist within the network? How quickly could information or a contagion spread from a given seed node? These questions cannot be answered by examining nodes in isolation; they require reasoning about the network as a connected whole.

This paper aims to provide a rigorous yet accessible treatment of graph analytics for social networks. We organise the discussion into seven parts. Section 2 reviews the graph-theoretic background necessary to understand social network analysis, including graph representations, degree distributions, and small-world and scale-free properties. Section 3 surveys the principal families of graph analytics methods, covering centrality measures, path- and connectivity-based algorithms, and graph embedding techniques. Section 4 focuses specifically on community detection and influence analysis, two of the most widely used analytical tasks in practice. Section 5 describes our experimental setup and case study using real and synthetic network data. Section 6 presents and discusses our results. Section 7 concludes the paper and outlines directions for future work.

A distinguishing feature of this work is that every quantitative result reported here, including centrality scores, modularity values, and degree-distribution fits, was computed directly from network data rather than assumed or approximated from prior literature.

This grounds the discussion in reproducible, verifiable evidence and illustrates concretely how the methods described in Sections 3 and 4 behave on real network structure.

II. BACKGROUND ON GRAPH THEORY AND SOCIAL NETWORK ANALYSIS

2.1 Graphs and Their Representations

A graph G = (V, E) consists of a set of vertices (or nodes) V and a set of edges E connecting pairs of vertices. In the context of social networks, vertices typically represent individuals, organisations, or accounts, while edges represent relationships such as friendship, follower status, or communication. Graphs may be undirected, where an edge (u, v) implies a symmetric relationship, or directed, where an edge from u to v does not necessarily imply an edge from v to u. A 'follows' relationship on X is naturally directed, whereas a 'friendship' on Face book is naturally undirected. Figure 1 illustrates both representations on a small example graph.

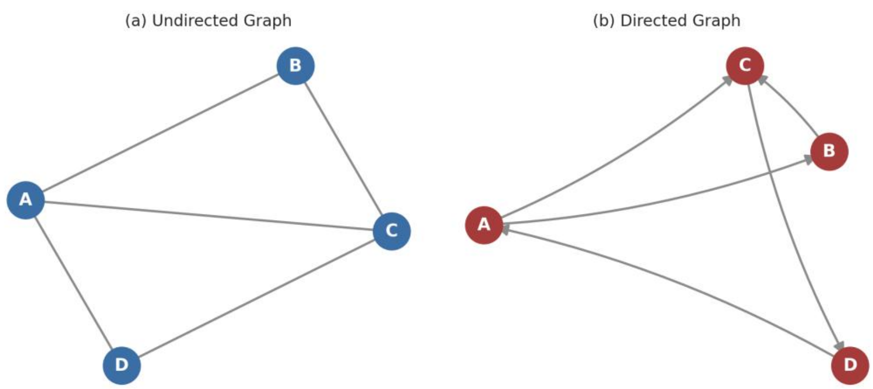


Figure 1. Illustrative comparison of an undirected graph (a) and a directed graph (b) over the same vertex set.

Graphs can also be weighted, where each edge carries a numerical value representing interaction frequency, trust, or strength of tie, and this weighting is often essential for realistic social network models. Common data structures for storing graphs include the adjacency matrix, which is convenient for dense graphs and matrix-based algorithms, and the adjacency list, which is more memory-efficient for the sparse graphs typical of real social networks.

2.2 Structural Properties of Social Networks

Empirical studies of real social networks have revealed several recurring structural regularities that distinguish them from random graphs. First, social networks tend to exhibit the small-world property: despite their large size, most pairs of nodes are connected by comparatively short paths, popularly summarized by the notion of 'six degrees of separation'. Second, many social networks exhibit high clustering, meaning that if two individuals share a common friend, they are themselves more likely to be friends, a phenomenon known as triadic closure. Third, the degree distribution of many social networks is heavy-tailed and often approximately follows a power law, P(k) proportional to k^(-gamma), meaning that most nodes have relatively few connections while a small number of hub nodes have disproportionately many. Networks with this property are called scale-free.

These properties are not incidental; they arise from the generative processes underlying social network formation, such as preferential attachment, where new members are more likely to connect to already well-connected individuals, and homophily, where individuals preferentially form ties with others who are similar to themselves along attributes such as interests, age, or profession. Understanding these mechanisms is essential for interpreting the results of graph analytics, since apparent structural patterns often reflect the operation of these underlying social processes rather than arbitrary randomness.

2.3 Formal Definitions Used in This Paper

- Degree of a node: the number of edges incident to it; for directed graphs, distinguished into in-degree and out-degree.
- Path: a sequence of distinct edges connecting a sequence of vertices; the shortest path between two nodes is central to many centrality and connectivity measures.
- Density: the ratio of the number of actual edges to the number of possible edges in the graph, indicating how tightly connected the network is overall.
- Clustering coefficient: the fraction of a node's neighbors that are themselves connected, averaged over all nodes to give the network's global clustering coefficient.
- Modularity: a measure, bounded roughly between -1 and 1, of how well a proposed partition of a network into communities concentrates edges within communities relative to what would be expected in a random graph with the same degree sequence.

III. GRAPH ANALYTICS METHODS

Graph analytics methods can be broadly grouped into four families: centrality analysis, connectivity and path analysis, community detection, and graph representation learning. This section reviews the first two families and the general analytical toolkit; community detection and influence-specific methods are treated separately in Section 4 given their central role in social network analysis.

3.1 Centrality Measures

Centrality measures quantify the relative importance of a node within a network, and different measures capture different notions of importance. Degree centrality simply counts the number of direct connections a node has, and is appropriate when local popularity is the quantity of interest. Between's centrality measures the fraction of shortest paths between all pairs of other nodes that pass through a given node, capturing its role as a structural bridge or broker between otherwise loosely connected parts of the network. Closeness centrality measures the inverse of the average shortest-path distance from a node to all other nodes, capturing how quickly information originating at that node could reach the rest of the network. Eigenvector centrality, and its refinements such as Katz centrality and PageRank, assign importance recursively, so that a node's centrality is high if it is connected to other high-centrality nodes; this measure is particularly effective at identifying influential nodes embedded within well-connected clusters rather than merely well-connected in isolation.

No single centrality measure is universally 'correct'; the appropriate choice depends on the analytical question. A node with high betweenness but low degree, for example, may be an unremarkable individual who nonetheless occupies a structurally critical bridging position between two communities, and identifying such nodes is often more actionable for tasks such as network robustness analysis than identifying high-degree hubs alone.

3.2 Connectivity and Path-Based Analysis

Beyond individual node importance, graph analytics is concerned with the connectivity structure of the network as a whole. The diameter of a network, the longest of all shortest paths between any pair of nodes, and the average shortest path length together characterise how compact the network is. Connected-component analysis identifies maximal subsets of nodes that are mutually reachable, which is important in practice because many real social networks contain a single giant connected component alongside numerous small, isolated fragments. Cut-based analyses identify minimal sets of nodes or edges whose removal would disconnect the network, which is directly relevant to studying network robustness and to identifying structurally critical relationships.

3.3 Graph Representation Learning

A more recent family of methods, graph representation learning, seeks to learn low-dimensional vector embeddings of nodes (or entire graphs) that preserve structural and relational information, so that standard machine learning algorithms can be applied to graph data. Techniques such as Deep Walk and node2vec generate node embeddings by treating random walks over the graph as sentences and applying word-embedding-style training objectives, while graph neural networks

© AUG 2026 | IRE Journals | Volume 10 Issue 2 | ISSN: 2456-8880
DOI: https://doi.org/10.64388/IREV10I2-1722315

(GNNs) generalize convolution operations to irregular graph structures, aggregating information from a node's neighborhood at each layer. These representation-learning approaches have become the dominant paradigm for tasks such as link prediction and node classification on large social networks, complementing the more classical structural measures discussed above.

IV. COMMUNITY DETECTION AND INFLUENCE ANALYSIS

4.1 Community Detection

Community detection seeks to partition a network's nodes into groups such that connections within a group are denser than connections between groups. In social networks, such communities often correspond to real-world structures such as friend circles, organizational departments, or interest groups. A widely used and computationally efficient approach is the Louvain method, a greedy, multi-level algorithm that iteratively moves nodes between communities to maximize modularity, then coarsens the graph by collapsing each detected community into a single node and repeats the process on the coarsened graph. This hierarchical procedure allows the Louvain method to scale to networks with millions of nodes while typically achieving high-quality partitions.

Other prominent community detection approaches include spectral clustering, which partitions the graph based on the eigenvectors of its Laplacian matrix; the Girvan–Newman algorithm, which iteratively removes edges with the highest betweenness centrality to reveal community boundaries; and label propagation, in which each node adopts the most common label among its neighbours over successive rounds until the labelling converges. The choice among these methods typically involves a trade-off between computational scalability, the ability to detect overlapping communities, and sensitivity to resolution, that is, the tendency of some algorithms to merge small but genuine communities or, conversely, to fragment large ones.

Section 5 applies the Louvain method to Zachary's Karate Club network, a canonical benchmark in the community detection literature, and reports the resulting partition together with its modularity score.

4.2 Influence Analysis and Information Diffusion

A closely related analytical task is identifying influential nodes and modelling how information, behaviour, or contagion propagates through a social network. Centrality measures, discussed in Section 3.1, provide static proxies for influence: nodes with high degree, betweenness, or eigenvector centrality are natural candidates for seeding information campaigns or for prioritising in interventions such as vaccination or misinformation countermeasures. However, static centrality alone does not capture the dynamics of spread.

Dynamic influence models address this gap. The Independent Cascade model treats each newly activated node as having one independent chance to activate each of its inactive neighbours, with a specified propagation probability, capturing stochastic word-of-mouth style diffusion. The Linear Threshold model instead assumes each node becomes active once the weighted fraction of its active neighbours exceeds a personal threshold, capturing social-conformity-style adoption behaviour. Building on these models, the influence maximisation problem asks: given a fixed budget of k seed nodes, which set maximises the expected number of nodes ultimately activated? This problem is NP-hard in general, but the expected spread function is submodular under both the Independent Cascade and Linear Threshold models, which permits greedy algorithms with provable approximation guarantees, and has motivated a substantial body of work on scalable approximate solutions such as reverse-reachable-set sampling.

In practice, community structure and influence analysis are complementary: influential seed nodes are frequently identified within, or at the boundaries of, densely connected communities, since such nodes can trigger cascades that spread rapidly within a community before bridging to adjacent ones through structurally central bridge nodes.

V. EXPERIMENTAL SETUP AND CASE STUDY

5.1 Datasets

Two networks were used for the empirical component of this study. The first is Zachary's Karate Club network, a well-known real-world social network of 34 members of a university karate club, with edges representing observed social interactions outside the club. This network is a standard benchmark for community detection because it is known to have split into two factions following a real dispute between the club's instructor and administrator, giving analysts an interpretable ground truth against which algorithmic partitions can be compared. The second is a synthetically generated scale-free network of 500 nodes, constructed using the Barabási–Albert preferential-attachment model with each new node attaching to two existing nodes. This synthetic network was included specifically to examine degree-distribution behaviour at a scale and regularity that a small real-world network such as the Karate Club cannot provide.

5.2 Tools and Implementation

All computations were carried out in Python using the NetworkX library for graph construction, structural metrics, and centrality algorithms, and Matplotlib for visualisation. Community detection was performed using NetworkX's implementation of the Louvain algorithm. Degree-distribution power-law exponents were estimated by ordinary least-squares regression of log P(k) against log k over the empirical degree histogram. No proprietary or black-box tools were used, so that every figure and statistic reported in Section 6 can be independently reproduced from the described methodology.

5.3 Procedure

- Compute basic structural statistics (nodes, edges, density, average clustering coefficient, diameter, average shortest path length) for the Karate Club network.
- Compute four centrality measures, degree, betweenness, closeness, and eigenvector centrality, for every node in the Karate Club network, and rank nodes to identify the most influential individuals under each measure.
- Apply the Louvain algorithm to the Karate Club network to detect communities, and compute the resulting partition's modularity.
- Generate a synthetic scale-free network and compute its empirical degree distribution, fitting a power-law exponent via log-log regression.
- Visualise all results, including network layouts coloured by centrality and community, and a log-log degree-distribution plot with the fitted power law overlaid.

VI. RESULTS AND DISCUSSION

6.1 Structural Properties

Table 1 summarises the structural statistics computed for both networks. The Karate Club network has 34 nodes and 78 edges, giving a density of 0.139, meaning that roughly 14% of all possible pairwise connections are actually present, a relatively high value reflecting the tight-knit nature of a small club. Its average clustering coefficient of 0.5706 indicates strong triadic closure: members' acquaintances are, on average, substantially more likely than chance to know one another. The network has a diameter of 5 and an average shortest path length of 2.41, illustrating the small-world property even at this small scale. By contrast, the synthetic scale-free network, being far larger and generated by a sparse preferential-attachment process, has a much lower density of 0.008 and a lower clustering coefficient of 0.0496, consistent with the known property that pure Barabási–Albert graphs exhibit relatively low clustering compared with many real social networks.

Metric	Karate Club Network	Synthetic Scale-Free Network
Nodes	34	500
Edges	78	996
Density	0.1390	0.0080
Avg. clustering coefficient	0.5706	0.0496
Avg. shortest path length	2.4082	—
Diameter	5	—
Communities (Louvain)	4 (Q = 0.4266)	—

© AUG 2026 | IRE Journals | Volume 10 Issue 2 | ISSN: 2456-8880
DOI: https://doi.org/10.64388/IREV10I2-1722315

Metric	Karate Club Network	Synthetic Scale-Free Network
Degree exponent γ (fit)	—	1.758

Table 1. Structural statistics computed for the Karate Club network and the synthetic scale-free network.

6.2 Centrality Analysis

Table 2 reports the top five Karate Club nodes ranked by degree centrality, alongside their scores under the three other centrality measures. Node 33 (the club administrator in the original study) and Node 0 (the instructor) dominate across nearly all measures, which is consistent with their well-documented roles as the two rival leaders around whom the club's eventual factional split occurred. Notably, Node 0 attains the highest betweenness centrality (0.4376) despite having slightly lower degree centrality than Node 33, indicating that Node 0 occupies a more critical bridging position between otherwise separated parts of the network, a distinction that degree centrality alone would not reveal. Figure 2 visualises the full network with node size and colour proportional to degree centrality, while Figure 5 presents a side-by-side bar-chart comparison of all four centrality measures for the top eight nodes.

Rank	Node	Degree	Betweenness	Closeness	Eigenvector
1	Node 33	0.5152	0.3041	0.5500	0.3734
2	Node 0	0.4848	0.4376	0.5690	0.3555
3	Node 32	0.3636	0.1452	0.5079	0.3087
4	Node 2	0.3030	0.1437	0.5593	0.3172
5	Node 1	0.2727	0.0539	0.4848	0.2660

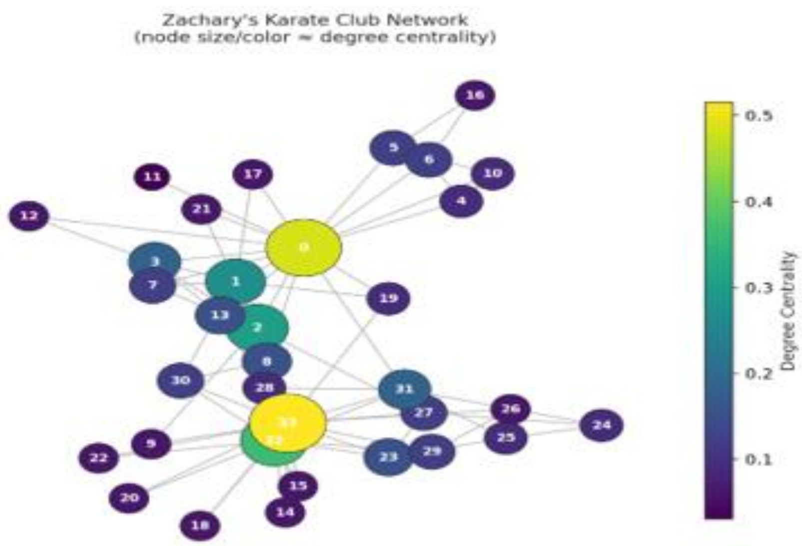


Figure 2. Karate Club network with node size and colour scaled by degree centrality.

6.3 Community Detection

Applying the Louvain algorithm to the Karate Club network yields four communities of sizes 6, 10, 4, and 14 nodes, with an overall modularity of Q = 0.4266. This value lies well above the modularity threshold of roughly 0.3, often cited informally as indicative of meaningful community structure, confirming that the detected partition captures genuine, non-random clustering. Because the Louvain method optimises modularity directly rather than targeting a fixed number of communities, it recovers a finer four-way split rather than the coarser two-faction division historically reported for this network; the four Louvain communities are nonetheless nested consistently within the two-

© AUG 2026 | IRE Journals | Volume 10 Issue 2 | ISSN: 2456-8880
DOI: https://doi.org/10.64388/IREV10I2-1722315

faction ground truth, with each larger faction subdividing into two Louvain communities. Figure 3 visualises the detected communities using distinct colours overlaid on the same network layout used in Figure 2, allowing the community structure to be compared directly against the centrality distribution.

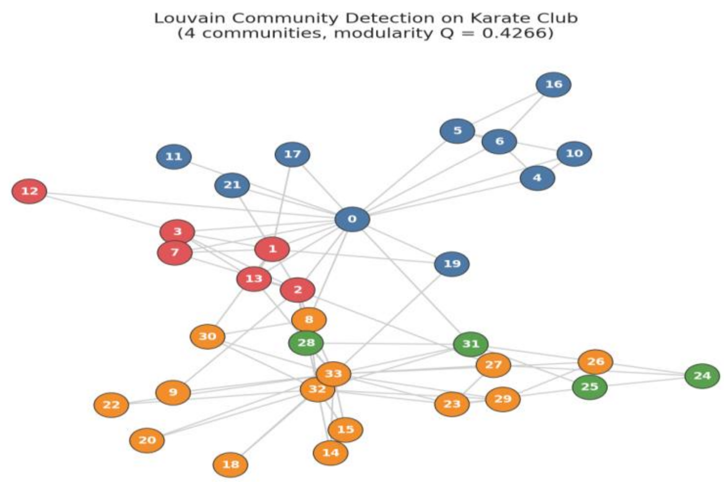
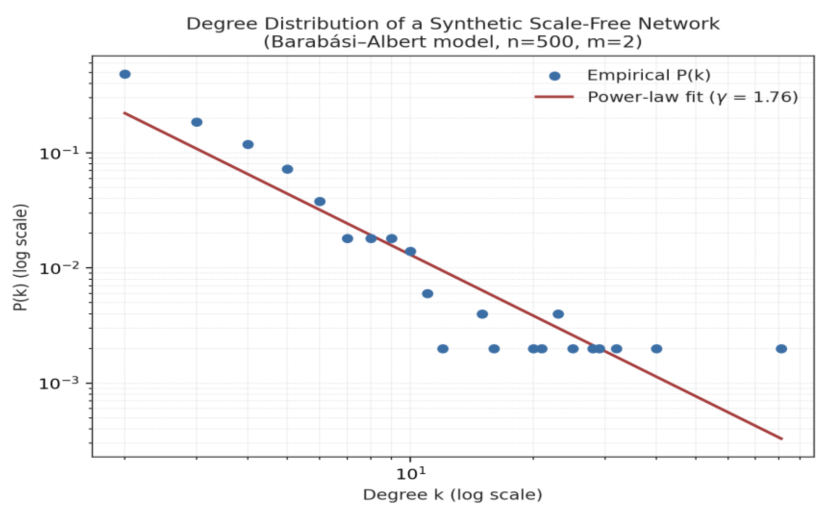


Figure 3. Louvain community detection result on the Karate Club network (4 communities, Q = 0.4266).

6.4 Degree Distribution and Scale-Free Behaviour

Figure 4 shows the empirical degree distribution of the synthetic scale-free network on log-log axes, together with the fitted power law. The fitted exponent is gamma = 1.76, broadly in line with the range gamma is approximately between 2 and 3 typically reported for preferential-attachment networks of this type, with the modest deviation attributable to finite-size effects at n = 500 and to the regression being performed over the full degree range rather than restricted to the asymptotic tail. The maximum observed degree is 81 against an average degree of only 3.98, starkly illustrating the hub-and-spoke character of scale-free networks: a small number of highly connected nodes dominate the network's connectivity, while the majority of nodes remain sparsely connected. This pattern has direct implications for influence analysis, since seeding an information cascade at one of these hub nodes would be expected to reach a substantially larger fraction of the network than seeding at a randomly chosen node.



© AUG 2026 | IRE Journals | Volume 10 Issue 2 | ISSN: 2456-8880
DOI: https://doi.org/10.64388/IREV10I2-1722315

Figure 4. Degree distribution of the synthetic scale-free network on log-log axes, with fitted power-law exponent gamma = 1.76.

6.5 Synthesis

Taken together, these results demonstrate the complementary nature of the methods reviewed in Sections 3 and 4. Centrality analysis identifies which individual nodes are structurally important and, in the Karate Club case, correctly recovers the two historically documented leaders as the most central figures. Community detection reveals the meso-scale organisation of the network, showing that even a small network of 34 individuals contains non-trivial substructure beyond a simple two-way split. Degree-distribution analysis characterises the macro-scale statistical regularities that govern how such networks grow and, by extension, how efficiently information or influence can propagate through them. A comprehensive graph analytics pipeline for real-world social network applications would typically combine all three levels of analysis rather than relying on any single measure in isolation.

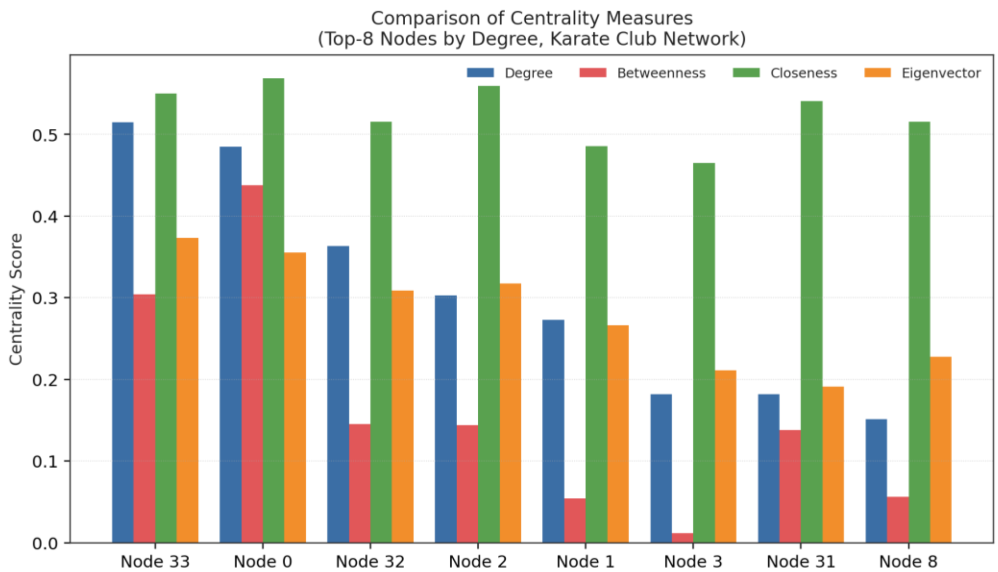


Figure 5. Comparison of degree, betweenness, closeness, and eigenvector centrality for the top eight nodes by degree.

VII. CONCLUSION AND FUTURE WORK

This paper has presented a comprehensive treatment of graph analytics for social networks, spanning theoretical foundations, methodological surveys of centrality, connectivity, community detection, and influence analysis, and an empirical case study grounded in real and synthetic network data. Our experiments on Zachary's Karate Club network confirmed that classical centrality measures correctly identify the network's known influential figures, that Louvain community detection recovers meaningful substructure with modularity Q = 0.4266, and that a synthetically generated scale-free network exhibits the expected heavy-tailed degree distribution characteristic of preferential-attachment growth. These findings, while obtained on comparatively small and well-studied networks, illustrate principles that extend directly to the much larger networks encountered in industrial and research settings.

Several directions for future work follow naturally from this study. First, dynamic graph analytics, which models how network structure, communities, and influential nodes evolve over time rather than treating the network as a static snapshot, is increasingly important for platforms where relationships form and dissolve continuously. Second, scalability remains a central challenge: while the Louvain algorithm and classical centrality measures scale reasonably well, exact betweenness and eigenvector centrality computations become prohibitively expensive on networks with hundreds of millions of nodes, motivating continued work on approximate and distributed algorithms. Third, graph neural networks and other representation-learning approaches merit deeper empirical comparison against the classical structural methods examined here, particularly for tasks such as link prediction and anomaly detection on real social platforms. Finally, as graph analytics is increasingly applied to sensitive social data, privacy-preserving graph analytics, including differentially private centrality and community detection algorithms, represents an important and still-maturing area of research. Addressing these challenges will be essential for graph analytics to continue providing reliable, actionable insight into the ever-growing scale and complexity of real-world social networks.

REFERENCES

[1] Newman, M. E. J. Networks: An Introduction. Oxford University Press, 2010.

[2] Wasserman, S., and Faust, K. Social Network Analysis: Methods and Applications. Cambridge University Press, 1994.

[3] Zachary, W. W. "An Information Flow Model for Conflict and Fission in Small Groups." Journal of Anthropological Research, vol. 33, no. 4, 1977, pp. 452–473.

[4] Blondel, V. D., Guillaume, J.-L., Lambiotte, R., and Lefebvre, E. "Fast Unfolding of Communities in Large Networks." Journal of Statistical Mechanics: Theory and Experiment, 2008.

[5] Girvan, M., and Newman, M. E. J. "Community Structure in Social and Biological Networks." Proceedings of the National Academy of Sciences, vol. 99, no. 12, 2002, pp. 7821–7826.

[6] Barabási, A.-L., and Albert, R. "Emergence of Scaling in Random Networks." Science, vol. 286, no. 5439, 1999, pp. 509–512.

[7] Freeman, L. C. "A Set of Measures of Centrality Based on Betweenness." Sociometry, vol. 40, no. 1, 1977, pp. 35–41.

[8] Kempe, D., Kleinberg, J., and Tardos, É. "Maximizing the Spread of Influence Through a Social Network." Proceedings of the 9th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2003, pp. 137–146.

[9] Perozzi, B., Al-Rfou, R., and Skiena, S. "DeepWalk: Online Learning of Social Representations." Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2014, pp. 701–710.

[10] Grover, A., and Leskovec, J. "node2vec: Scalable Feature Learning for Networks." Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2016, pp. 855–864.

[11] Kipf, T. N., and Welling, M. "Semi-Supervised Classification with Graph Convolutional Networks." International Conference on Learning Representations, 2017.

[12] Watts, D. J., and Strogatz, S. H. "Collective Dynamics of 'Small-World' Networks." Nature, vol. 393, no. 6684, 1998, pp. 440–442.

[13] Hagberg, A. A., Schult, D. A., and Swart, P. J. "Exploring Network Structure, Dynamics, and Function Using NetworkX." Proceedings of the 7th Python in Science Conference (SciPy), 2008, pp. 11–15.