Enhancing Real-World Network Understanding through Centrality Measures and Improved Clustering Coefficient Methods

Touseef Ali
Department of Mathematics
NED University of Engineering and Technology
Karachi, Pakistan

Ubaida Fatima, PhD
Department of Mathematics
NED University of Engineering and Technology
Karachi, Pakistan
https://orcid.org/0000-0003-0372-0858

ABSTRACT
Network analysis has become an essential tool for understanding the complex structures and dynamics of large datasets across various disciplines. The quick growth of data in size and complexity presents significant challenges in accuracy and explanation of existing methods. This research proposes the development and application of advanced community detection algorithms related to large scale networks. Particular emphasis of this will be placed on addressing challenges such as overlapping communities, dynamic network structures, and the balance between computational cost and detection quality. This study seeks to advance the understanding of community detection in large networks and its implications for real world data driven problems.

Keywords
Network Analysis (NA), Community detection Methods, Centrality Measures, and Clustering Coefficients.

1. INTRODUCTION
The rapid growth of large-scale networks in domains such as social media, biology, finance, and cybersecurity has created a strong demand for advanced analytical methods [1]. Traditional statistical tools often fail to capture the dynamic and high-dimensional properties of these systems, making specialized approaches such as network analysis essential. Network analysis provides frameworks including centrality [2], influence [3], flow [4] and dynamics [5], which together enable understanding of complex interactions. A key method in this area is community detection, which identifies densely connected groups and reveals hidden structures in social, biological, and communication networks [6]. Traditional approaches such as partitioning [7], hierarchical clustering [8], spectral clustering [9] and density-based clustering [10] are interpretable but struggle with scalability. Modern deep learning approaches including Graph Neural Networks [11]. Graph Convolutional Networks [12] and Graph Attention Networks [13] offer scalability but face issues of interpretability and computational cost. To overcome these limitations, hybrid strategies that integrate graph-based methods with deep learning are emerging as promising solutions. This study contributes to this direction by combining community detection with centrality measures to improve accuracy, scalability, and applicability in real-world networks such as social, biological, and financial systems.

1.1 Motivation
The rapid growth of large networks in areas like social media, biology, finance, and cybersecurity has created a need for effective tools to study their complex structures. Community detection helps by finding groups of closely connected nodes, revealing hidden patterns, important modules, and influential clusters. Traditional graph-based methods are easy to understand but often cannot handle large or changing networks. Deep learning methods, like Graph Neural Networks and Graph Convolutional Networks, are accurate and can handle big networks, but they are harder to interpret and require a lot of computing power. Many existing studies also overlook the importance of centrality measures, which identify influential nodes that affect how communities behave and stay connected. This research aims to combine traditional and modern approaches by integrating community detection with centrality analysis. The goal is to develop frameworks that are robust, scalable, and easy to interpret, showing both the key nodes driving network behavior and the overall community structure.

1.2 Aims and Objectives
- To improve the study of large and complex networks by combining community detection with centrality measures.
- To study and compare traditional graph methods (like graph partitioning, spectral clustering, and hierarchical clustering) with deep learning methods (like GNNs, GCNs, GATs, and Autoencoders).
- To study how well centrality measures (degree, closeness, betweenness, eigenvector, PageRank, Katz, gravity) can find important nodes in communities.
- To create a combined framework that uses both community detection and centrality analysis to make results more accurate, easier to understand, and able to handle large networks.
- To test the proposed framework using real-world datasets from areas like social media, biology, finance, and cybersecurity.
- To show how combining these methods can help with decision-making, managing risks, detecting fraud, and making networks more resilient.

2. LITERATURE REVIEW
- **Network Analysis (NA):** Network Analysis provides methods to study the structure and behavior of complex systems. Foundational work by Wasserman and Faust (1994) [14] defined structural measures for relational data. Barabási and Albert (1999) [15] introduced the scale-free model, showing many real-world networks have highly connected hubs following a

power-law distribution. Newman (2003) [16] highlighted key network properties, including clustering, assortativity, and path lengths, which remain central to modern network research.

- **Community detection methods:** Community detection finds groups of closely connected nodes using approaches like modularity optimization (Newman, 2006) [17], spectral clustering (von Luxburg, 2007) [18], hierarchical clustering (Müllner, 2011) [19], and probabilistic models such as the Stochastic Block Model (Holland et al., 1983) [20]. These methods capture both structural and statistical patterns in static and dynamic networks. These methods capture both structural and statistical patterns in static and dynamic networks.
- **Centrality measures:** Centrality measures quantify the importance of nodes in a network. Classical metrics like degree, closeness, and betweenness (Freeman, 1977) [21] assess local connectivity, path efficiency, and control over flows. Eigenvector centrality (Bonacich, 1987)[22], PageRank (Page et al., 1999) [23], and Katz centrality capture global influence, while gravity centrality (Zhang, Shuai & Lü, 2022) [24] combines node degree with distance to identify influential nodes in large networks.

3. METHODOLOGY

To study large-scale networks, the research framework combines graph-theoretic and deep learning-based methods. To find influential nodes, structural connectivity, and cohesive communities, the method evaluates centrality measures, clustering coefficients, and community detection methods. Both traditional and modern approaches are applied to ensure robustness and scalability.

3.1 Data collection and processing

3.1.1 Main Datasets

i. Small undirected network having 8 nodes are described:

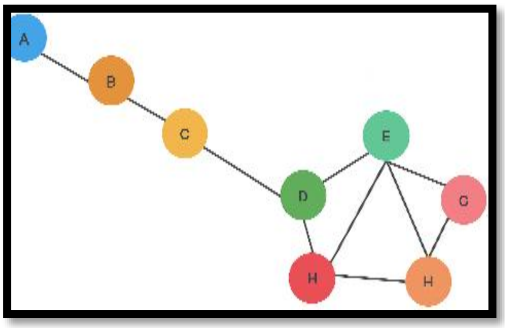

Figure 1

ii. Dolphin ***Social Network Dataset:***
The Dolphin Social Network dataset captures social interactions among 62 bottlenose dolphins in Doubtful Sound, New Zealand. Nodes represent individual dolphins, and edges indicate frequent associations. The network is undirected and unweighted, with a total of 159 edges.

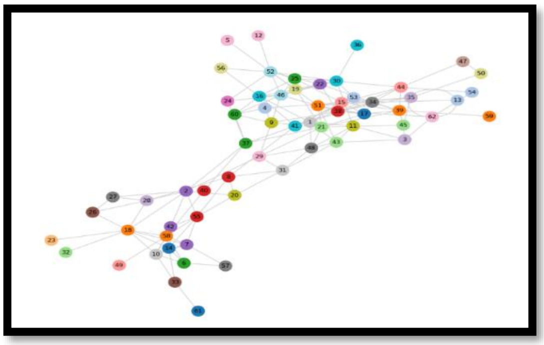

Figure 2

i. **Facebook social network dataset**
The Facebook Social Network dataset, provided by the Stanford Network Analysis Project (SNAP), includes 4,039 users (nodes) and 88,234 friendship connections (edges). Nodes represent users, and undirected edges represent friendships. It is widely used in social network analysis, community detection, and centrality studies due to its large scale and real-world relevance.

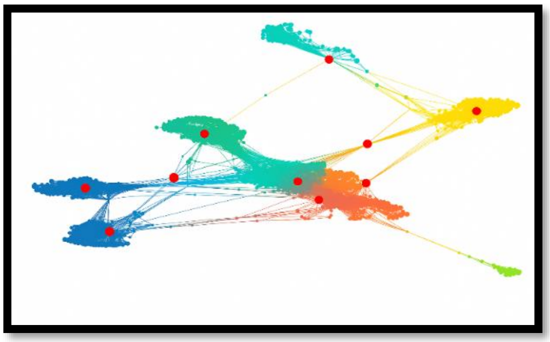

Figure 3

3.2 Techniques

3.2.1 Centrality Measures:

Centrality measures quantify the importance or influence of nodes in a network based on their connectivity and position [21]. Multiple centrality measures are computed to capture node influence and structural importance:

- **Degree Centrality (DC):** Degree centrality measures the number of direct connections a node has in a network, indicating its immediate influence [21].
- **Closeness Centrality (CC):** Closeness centrality measures how quickly a node can reach all other nodes in a network [2].

$$CC(v) = \frac{N-1}{\sum_{u \neq v} d(u,v)} \quad (1)$$

Equation (1) is mathematical formula to find CC.

- **Betweenness Centrality (BC):** Betweenness centrality measures a node's role as a bridge by counting how often it lies on shortest paths between other nodes [21].

$$BC(v) = \sum_{u,v \in N} \frac{\sigma uw(v)}{\sigma uw} \quad (2)$$

Equation (2) is mathematical formula to find BC.

- **Eigenvector Centrality (EVC):** Eigenvector centrality measures a node's influence based on the importance of its connected neighbors [22].

$$xi = \frac{1}{\lambda}\sum_{j\in N(i)} Aij \times xj \qquad (3)$$

Equation (3) is mathematical formula to find EVC.

- **PageRank Centrality (PRC):** PageRank measures the importance of a node based on the quantity and quality of links pointing to it [23].

$$\boldsymbol{PRC(i) = \frac{1-d}{N} + d\sum_{j\in N(i)} \frac{PRC(j)}{\deg(j)}} \qquad (4)$$

Equation (4) is mathematical formula to find PRC.

- **Katz Centrality (KC):** Katz centrality measures a node's influence by considering the total number of walks connecting it to other nodes, giving more weight to closer nodes [24].

$$KC = (I - \alpha A)^{-1}.\beta.1 \qquad (5)$$

Equation (5) is mathematical formula to find KC.

- **Gravity Centrality (GC):** Gravity centrality identifies influential nodes by combining their degree with the shortest-path distances to other nodes [25].

$$GC(i) = \sum_{j\neq i} \frac{Mi \times Mj}{d(i,j)^{\alpha}} \qquad (6)$$

Equation (6) is mathematical formula to find GC.

All measures are computed using **MATLAB**

3.2.2 Clustering Coefficients Analysis

- **Local Clustering Coefficient (LCC):** The local clustering coefficient measures the likelihood that a node's neighbors are also connected, indicating the tendency to form tightly-knit groups [26]

$$LCC = \frac{Actual\ number\ of\ links\ between\ neighbors\ of\ a\ node}{Maximum\ Possible\ links\ between\ neighbors\ of\ a\ node} \qquad (7)$$

Equation (7) is mathematical formula to find LCC.

- **Global Clustering Coefficient (GCC):** The global clustering coefficient measures the overall tendency of a network to form tightly connected triangles among nodes [26].

$$GCC = \frac{3\times Number\ of\ triangles\ in\ the\ Graph}{Number\ of\ connected\ triplets\ of\ nodes} \qquad (8)$$

Equation (8) is mathematical formula to find GCC.

4. RESULTS AND INTERPRETATIONS:

Centrality Measures:

i- The results of the above Figure 1 containing 8 nodes are given in the Table 1:

Table 1

Nodes	DC	CC	BC	EVC	PRC	KC	GC
A	1.00	0.29	0.00	0.02	0.05	1.12	2.36
B	2.00	0.39	6.00	0.06	0.11	1.24	3.52
C	2.00	0.50	10.0	0.15	0.11	1.25	5.52
D	3.00	0.58	12.0	0.39	0.14	1.30	19.7
E	4.00	0.54	6.50	0.54	0.19	1.53	41.7
F	2.00	0.39	0.00	0.33	0.10	1.36	25.4
G	3.00	0.41	0.50	0.45	0.14	2.32	29.1
H	3.00	0.50	2.50	0.46	0.15	0.25	34.2

ii- The results of top 10 nodes of the above Figure 2 (Dolphin Social Network Dataset)containing 62 nodes are given in the table 2:

Table 2: Dolphin Social Network Dataset

Node	**DC**	**BC**	**CC**	**EC**	**PRC**	**KC**	**GC**
37	7	454.3	0.007	0.0235	0.0206	8.355	700.6
2	8	390.9	0.006	0.0075	0.0247	5.817	688.1
41	8	261.9	0.007	0.0369	0.0219	11.54	876.6
38	11	253.6	0.007	0.0534	0.0298	15.82	1327.6
8	5	216.4	0.006	0.00762	0.0157	4.505	383.3
18	9	209.8	0.005	0.0031	0.0318	5.526	676.9
21	9	187.8	0.006	0.0328	0.0246	10.69	931.99
55	7	181.4	0.005	0.0041	0.0217	5.276	540.83
52	10	154.9	0.005	0.0374	0.0313	11.57	895.24
58	9	154.1	0.005	0.0031	0.0302	5.77	667.34

iii- Centrality measure of top 10 nodes of the above Figure 3 (Facebook Social Network Dataset) containing 4039 nodes are given in the Table 3:

Table 3: Facebook Social Network Dataset

Node	DC	CC	BC	EVC	PRC	KC	GC
108	1045	0.00011	3.9×10^{6}	1.1×10^{-5}	0.007	9.56	8.3×10^{7}
1685	792	9.8×10^{-5}	2.8×10^{6}	3.9×10^{-7}	0.006	6.17	4.2×10^{7}
3438	547	7.8×10^{-5}	1.9×10^{6}	4.9×10^{-9}	0.008	4.09	1.4×10^{7}
1913	755	8.7×10^{-5}	1.9×10^{6}	0.0062	0.004	11.1	5.6×10^{7}
1086	66	8.9×10^{-5}	1.2×10^{6}	1.1×10^{-7}	0.001	1.47	1.9×10^{6}
1	347	8.8×10^{-5}	1.2×10^{6}	2.1×10^{-6}	0.006	2.96	1.9×10^{7}
699	68	6.7×10^{-5}	9.4×10^{5}	5×10^{-11}	0.001	1.39	1.1×10^{6}

568	63	8.1×10^{-5}	7.9×10^{5}	6.2×10^{-7}	0.001	1.48	1.6×10^{6}
59	12	9.8×10^{-5}	6.9×10^{5}	3.8×10^{-5}	0.002	1.19	5.1×10^{5}
429	115	9.8×10^{-5}	5.2×10^{5}	3.8×10^{-5}	0.001	1.91	4.9×10^{6}

Clustering Coefficient Analysis:

i- The results of the above Figure 1 containing 8 nodes are given in the table 4:

Table 4

Node	**Degree**	Local Clustering **Coefficients**
A	1	0.000
B	2	0.000
C	2	0.000
D	3	0.333
E	4	1.500
F	2	1.000
G	3	0.667
H	3	0.333

Global Clustering Coefficient (GCC) ≈ 0.354

ii- Clustering Coefficients of top 10 nodes of the above Figure 2 (Dolphin Social Network Dataset) containing 62 nodes are given in the table 5:

Table 5: Dolphin Social Network Dataset

Node	Degree	Local Clustering Coefficients
26	3	0.66667
27	3	0.66667
17	6	0.60000
42	5	0.60000
7	6	0.53333
22	6	0.53333
25	6	0.53333
10	7	0.52381
19	7	0.52381
6	4	0.50000

Global Clustering Coefficient (GCC) = 0.2590

iii- Clustering Coefficients of top 10 nodes of the above Figure 3 (Facebook Social Network Dataset) containing 4039 nodes are given in the table 6:

Table 6: Facebook Social Network Dataset

Node	Degree	Local Clustering Coefficients
136	10	1
310	9	1
79	9	1
196	9	1
219	9	1
274	9	1
307	9	1
329	9	1
889	9	1
1017	9	1

Global clustering coefficient: 0.605

Clustering Coefficients of lowest 10 nodes of the above Figure 3 (Facebook Social Network Dataset) containing 4039 nodes are given in the table 7:

Table 7: Facebook Social Network Dataset

Node	Degree	Local Clustering Coefficients
0	347	0.041962
1	17	0.419118
2	10	0.888889
3	17	0.632353
4	10	0.866667
5	13	0.333333
6	6	0.933333
7	20	0.431579
8	98	0.678571
9	57	0.397243

4.1 Main dataset Results

Centrality Measures:

- The results of the centrality analysis for the 8-node network presented in Table 1 shown in **Figure 1** are:

  Node E emerges as the most influential node overall, leading in degree, eigenvector, PageRank, and gravity centrality. Node D is crucial as a bridge, dominating betweenness and closeness centrality. Node G shows importance through indirect connections (Katz centrality), while Node A remains peripheral across all measures. Collectively, these metrics highlight hubs (E), key connectors (D), indirect influencers (G), and low-impact nodes (A) within the network.

- The results of the centrality analysis for the Dolphin Social Network, which contains 62 nodes, are presented in **Table 2** for the top 10 ranked dolphins (**Figure 2):**

  Node 38 is the most influential dolphin, leading in degree, eigenvector, Katz, and gravity centrality. Node 37 is the key bridge with the highest betweenness, while Nodes 18, 21, and 52 also show strong influence in global measures like PageRank, Katz, and gravity centrality. Less central dolphins, such as Nodes 8 and 55, play smaller but noticeable roles in the network.

- The Facebook Social Network dataset contains **4039 nodes,** and the centrality results for the **top 10 nodes** are summarized in **Table 3 (Fig. 3):**

  Node 108 is the dominant hub, leading in degree, betweenness, closeness, and gravity centrality. Node 1913 is highly influential via eigenvector and Katz centrality, while Node 3438 stands out in PageRank. The analysis shows that influence in large social networks is distributed across major hubs (Node 108), globally connected influencers (Node 1913), and structurally embedded actors (Node 3438).

Clustering Coefficient Analysis:

- The results of the clustering coefficients for the 8-node presented in **Table 4** shown in **Figure 1**, are:

  Node E shows the highest local clustering, followed by Node F, while Nodes A, B, and C have no clustering. Nodes D, G, and H have moderate clustering. The overall global clustering coefficient (~0.354) indicates a moderately clustered network, lower than smaller networks due to increased size and connections.

- The results of the Clustering Coefficients for the Dolphin Social Network consisting of 62 nodes, are presented in **Table 5** for the top 10 ranked nodes **(Figure 2):**

  Nodes 26 and 27 have the highest local clustering (0.667), while moderately connected nodes like 17 and 42 also show strong clustering. Highly connected dolphins (e.g., 10 and 19) have lower local clustering, indicating links across groups. The network's global clustering coefficient (0.259) reflects moderate clustering, balancing local cohesion and broader connectivity.

- The results of the Clustering Coefficients for the Facebook Social Network dataset consisting of 4039 nodes, are presented in **Table 6** for the top 10 ranked nodes **(Figure 3):**

  Top-ranked nodes in the Facebook network exhibit perfect local clustering (LCC = 1), forming fully interconnected friendship cliques. The high global clustering coefficient (GCC = 0.6055) indicates that the network strongly favors community-based structures, where dense local groups are embedded within the broader social graph. These results highlight the presence of tightly-knit social circles and the overall tendency of users to form cohesive friendship communities.

  The results of the Clustering Coefficients for the Facebook Social Network dataset consisting of 4039 nodes, are presented in **Table 7** for the lowest 10 ranked nodes (Figure 3):

5. CONCLUSION

This research enhances the understanding of real-world networks by integrating centrality measures with improved clustering coefficient methods to achieve a more comprehensive view of network structure and node influence. Through detailed experimentation on synthetic, Dolphin, and Facebook social network datasets, the study demonstrates that combining traditional graph-based techniques with modern analytical frameworks yields more accurate, scalable, and interpretable results.

The analysis revealed that centrality measures such as degree, betweenness, eigenvector, PageRank, Katz, and gravity effectively identify influential nodes, while clustering coefficients capture the cohesiveness and structural balance of communities. Nodes with high centrality values often serve as bridges or hubs that sustain network connectivity, whereas clustering coefficients highlight the formation of tightly knit local groups within larger networks.

The integration of these methods provides a dual perspective—quantifying both global influence and local cohesion—thereby improving the interpretability of complex network dynamics. This combined framework offers practical significance in fields such as social media analytics, biological systems modeling, financial risk assessment, and cybersecurity, where understanding influence and community structure is critical.

Future work may extend this framework to dynamic and weighted networks, leveraging deep learning approaches such as Graph Neural Networks and Graph Convolutional Networks to further enhance performance and adaptability. Overall, the study contributes a robust, interpretable, and scalable methodology for advancing real-world network analysis and community detection research.

6. REFRENCES

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

[2] Freeman, L. C. (1979). *Centrality in social networks: Conceptual clarification.* Social Networks.

[3] Kempe, D., Kleinberg, J., & Tardos, É. (2003). Maximizing the spread of influence through a social network.

[4] Ford, L. R., & Fulkerson, D. R. (1962). *Flows in Networks*. Princeton University Press.

[5] Holme, P., & Saramäki, J. (2012). Temporal networks.

[6] Fortunato, S. (2010). Community detection in graphs. *Physics Reports*

[7] Kernighan, B. W., & Lin, S. (1970). An efficient heuristic procedure for partitioning graphs. *Bell System Technical Journal.*

[8] Johnson, D. S. (1967). *The Bell System Technical Journal.*

[9] Ng, A. Y., Jordan, M. I., & Weiss, Y. (2002). On spectral clustering: Analysis and an algorithm.

[10] Ester, M., Kriegel, H.-P., Sander, J., & Xu, X. (1996). A density-based algorithm for discovering clusters in large spatial databases with noise.

[11] Scarselli, F., Gori, M., Tsoi, A. C., Hagenbuchner, M., & Monfardini, G. (2009). The graph neural network model.

[12] Kipf, T. N., & Welling, M. (2017). Semi-supervised classification with graph convolutional networks.

[13] Veličković, P., Cucurull, G., Casanova, A., Romero, A., Liò, P., & Bengio, Y. (2018). Graph attention networks.

[14] Wasserman, S., & Faust, K. (1994). *Social network analysis: Methods and applications.*

[15] Barabási, A.-L., & Albert, R. (1999). Emergence of scaling in random networks.

[16] Newman, M. E. J. (2003). The structure and function of complex networks.

[17] Newman, M. E. J. (2006). Modularity and community structure in networks.

[18] Von Luxburg, U. (2007). A tutorial on spectral clustering. *Statistics and Computing.*

[19] Müllner, D. (2011). Modern hierarchical, agglomerative clustering algorithms.

[20] Holland, P. W., Laskey, K. B., & Leinhardt, S. (1983). Stochastic blockmodels: First steps.

[21] Freeman, L. C. (1977). A set of measures of centrality based on betweenness.

[22] Bonacich, P. (1987). Power and centrality: A family of measures. *American Journal of Sociology.*

[23] Page, L., Brin, S., Motwani, R., & Winograd, T. (1999). The PageRank citation ranking: Bringing order to the web.

[24] Zhang, S., & Lü, W. (2022). *De-Bias for Generative Extraction in Unified NER Task.*

[25] Zhang, S., Shuai, B., & Lü, M. (2014). A novel method to identify influential nodes in complex networks based on gravity centrality.

[26] Watts, D. J., & Strogatz, S. H. (1998). Collective dynamics of 'small-world' networks.

.