

## Clustering-Based: Assessing the Impact of Overlapping Nodes and Centrality Measures on Influencer Detection in Social Networks



Mustafa K. Alasadi<sup>1\*</sup>, Waffaa M. Ali<sup>2</sup>, Abdulkadhem A. Abdulkadhem<sup>3</sup>

<sup>1</sup> Faculty of Computer Science and Information Technology, University of Sumer, Al-Rifai 64005, Thi Qar, Iraq

<sup>2</sup> Veterinary Medicine College, University of Shutrah, Shutrah 64007, Thi Qar, Iraq

<sup>3</sup> Al-Mustaqbal Center for AI Applications, Al-Mustaqbal University, Hillah 51001, Babil, Iraq

Corresponding Author Email: [mustafa.kamil@uos.edu.iq](mailto:mustafa.kamil@uos.edu.iq)

Copyright: ©2024 The authors. This article is published by IETA and is licensed under the CC BY 4.0 license (<http://creativecommons.org/licenses/by/4.0/>).

<https://doi.org/10.18280/ria.380501>

### ABSTRACT

**Received:** 26 May 2024

**Revised:** 21 August 2024

**Accepted:** 30 August 2024

**Available online:** 23 October 2024

#### Keywords:

*social network, influential nodes, overlapping nodes, network communities, centrality measure*

All social networks possess a community organization of actors, and these actors are unequal in importance and influence. In the last decade, identifying influential nodes in social networks has drawn significant interest due to their numerous applications, as they play a crucial role in influence maximization, rumor control, and other practical applications. In social network analysis, most works rely on centrality measures for defining the influential nodes in a network. Finding the centrality score of the nodes in real-world large-scale networks involve extensive computational complexity, making it necessary to address this research gap and find a solution. The main contribution of this paper is its innovative methodology, which improves the efficiency of identifying the influencers by reducing the time and resources required for network analysis. The research objective is to categorize nodes and identify overlapping regions indicating individuals' memberships in multiple clusters or communities. It also incorporates centrality measures like degree, closeness, and betweenness to assess the influence levels of nodes within these overlapping clusters. The paper investigates the relationship between the node centrality score and the location of the nodes in the network communities and, at the same time, proposes a graph-based method to identify the influential nodes. The focus is on the overlapping nodes among the communities to understand how this factor influences the effectiveness of influencers. Through empirical analysis and simulations, the findings show a strong correlation between the node centrality score and the characteristic of overlapping nodes, which plays a crucial role in determining influencer efficacy. The research concludes that it is effectively possible to define the overlapping nodes as influencers since a vast majority of overlapping nodes have a high score of centrality compared to other nodes in network communities.

## 1. INTRODUCTION

Social networks play a central role in our daily lives, especially in light of the ongoing growth of web-based services and mobile devices. The content diffused by actors through social relations has direct influence on our beliefs, political opinions, and economic decisions are mainly affected [1]. The social network analysis use graph theory to present the network as a graph with nodes and edges, in which the nodes represent actors, and the edges define the relationship among the actors [2]. Real-world networks possess a modular organization of nodes, the so-called cluster or community structure [3, 4]. Despite the massive work that has gone into defining this property, there is still no formal agreement on the term that best describes a community [5]. It is intuitively understood as a tightly connected set of nodes where actors communicate more intensely with one another than with other network members [6, 7].

Communities in real networks usually overlap and even nest within one another, meaning that different communities might have some members in common, as in the case of social

networks, where people can belong to several groups, such as family, friends, and colleagues [8]. These members are unequal in importance and influence. Influencers are the people with more spreading ability and can significantly affect audience behavior and choices in social networks [9, 10].

In the last decade, identifying influential nodes in a graph has become a key research area due to its numerous applications [11], and it is worthwhile to provide an efficient way to identify these nodes. These vital nodes play a central role in controlling the spread of epidemics [12], ensuring efficient information diffusion [13], and essential for managing the spread of rumors [14].

Much previous research demonstrated how the community structure has a significant role in addressing some problems, such as modeling information spread and marketing applications [15]. Community detection helps to simplify large-scale social network analysis because grouping nodes by clustering methods is the basis of their connectivity or other attributes [16]. Despite these studies, there was no effective utilization of this feature for measuring the influential nodes in social networks, and the focus remained on traditional

centrality measures that are characterized by their high complexity.

In the context of related works and the systematic literature review as clarified in the study [17], five methods are commonly used to identify influencers: “Data mining techniques, machine-learning-based approaches, metaheuristic algorithms, graph-based methods, and hybrid approaches”. Given the characteristics of a social network structure based on member nodes and links between them, it is easy to model these structures in graph form. Thus, the graph-based is the most popular approach for solving social network problems, including influencer detection.

Numerous studies were conducted following a graph-based approach, like the work [18] that introduces the enhanced K-shell algorithm to uncover the most influential nodes and identify the communities to which they reside. On the other hand, the research [19] suggests a method to identify opinion leaders using node input and output degree parameters, in addition to a study [20] that finds influencers using a parallel algorithm based on user behavior. Despite the promising results of these studies, most of their techniques rely on finding the centrality score of nodes, where the flow of information among communities can be controlled by high centrality nodes, making them strategic targets for maximizing the spread of information or marketing campaigns. However, there is a challenge with finding the centrality score of the graph nodes that requires lots of calculations to be processed. Research evidencing the overlapping nodes and centrality measures in social networks shows a remarkable relationship between location of nodes within network communities and node centrality. The Overlapping nodes are important for influencer detection and demonstrate higher centrality scores as they which belong to multiple communities. These empirical studies prove that these nodes can be effectively recognized as influencers due to their significant positions that perform an extensive information distribution. An example, Sabharwal and Kaur [21] used several centrality measures to highlight key nodes within community structures and to identify the necessity of combining multiple measures for better accuracy. Other recent study [22], trying to combine the characteristics of local and global networks to validate its effectiveness in pinpointing influential nodes. Same scenario [23] leveraged skill-based user profiles and a unique PageRank algorithm. Their mission is to identify influential nodes in online social networks. Also, research [24] demonstrated the community detection role and bridge detection in enhancing the efficiency of node centrality computations, and they are crucial for identifying influencers in large networks. A high-quality subgraph extension in local core regions to develop a novel overlapping community detection algorithm was developed by researchers [25], showing the role of these nodes in supporting the stability and cohesion of network. From this, we can draw that the overlapping nodes and advanced centrality measures will effectively facilitating information flow, innovation, and community cohesion within complex social networks.

Based on this, there was an incentive to find an efficient way to identify influential nodes without the need to calculate the centrality score of the nodes. The idea was to consider the overlapping nodes as influential nodes and validate this assumption by measuring the centrality of overlapping nodes using the four basic centrality measures to check if there is a correlation between the fact that the node is overlapping and the degree of its centrality.

This paper utilizes a graph-based method that focuses on

overlapping nodes, where the unique positions of these overlapping nodes enable them to bridge various communities and facilitate or block the flow of information across the network. Three real networks were used for evaluation purposes in the experimental study, and four centrality measures (degree, eigenvector, betweenness, and closeness centrality) were used with generated graphs from datasets. The results show that the vast majority of overlapping nodes are characterized by high centrality, as they are located in the first quadrant, arranging nodes according to the degree of centrality. The paper concludes that overlapping nodes play a vital role in information propagation through social relations, and it is possible to adopt them as influencers, avoiding the complexity of finding the centrality score. The following section details the four basic centrality measures used in the evaluation process:

**Degree Centrality** [26]: The degree centrality of a node refers to the number of ties incident upon a node. For a given graph with vertices and edges, the degree of centrality for node  $i$  is defined as in Eq. (1) [27].

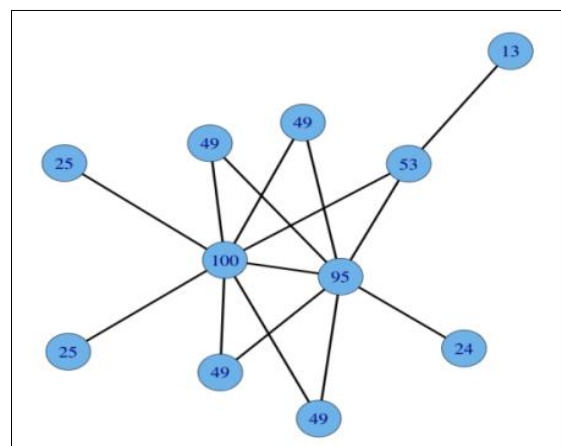
$$\alpha_d(i) = \sum_{j=1}^N a_{ij} \tag{1}$$

where,  $a_{ij}$  obtained from the adjacency matrix (one-step neighborhood) representing the graph  $G$  connectivity, where  $a_{ij}=1$ , if node  $i$  and node  $j$  are connected and  $a_{ij}=0$ , otherwise.

**Eigenvector centrality**: Is a natural extension of degree centrality [28], where an actor has more influence if it is in relation to influencer actors. To put it another way, the node's centrality depends not only on the number of its neighbors but also on the value of their centrality. It is calculated by finding the largest absolute eigenvalue and its associated eigenvector (leading eigenvector) of the adjacency matrix, Eq. (2).

$$X_i = \lambda^{-1} \sum_j A_{ij} X_j \tag{2}$$

where,  $X_i$  is the score at the node  $i$ ,  $A_{ij}$  is the correspondent value on the adjacency matrix, and  $\lambda$  is the eigenvalue. Figure 1 illustrates the calculation of this metric, where the graph nodes are labeled with their eigenvector centrality [29].



**Figure 1.** Illustration of the eigenvector centrality measure

**Closeness Centrality**: In graph theory, this metric is defined as a sophisticated centrality measure of a node [30].

Nodes have higher closeness if they tend to have short path lengths to other nodes in the graph. The work [31] demonstrates that “closeness centrality is the inverted sum of topological distances to every other node from a given node. It is calculated as illustrated in the Eq. (3).

$$\alpha_c(i) = \frac{N - 1}{\sum_{j=1}^{N-1} d(i, j)} \quad (3)$$

where, the length of the shortest path between nodes  $i$  and  $j$  is denoted by  $d(i, j)$ .

**Betweenness Centrality** [32, 33]: The betweenness measure is usually utilized to detect the amount of influence a node has over the diffusion of information in a network. Each node in the graph is given a score based on the number of shortest paths between every pair of nodes that travel through that node. In other words, a graph node is considered well connected if it lies on many of the shortest paths between other nodes. It is clarified as in Eq. (4).

$$\alpha_b(i) = \sum_{s, t \neq i} \frac{\sigma_i(s, t)}{\sigma(s, t)} \quad (4)$$

where,  $\sigma(s, t)$  represents the number of shortest paths between nodes  $s$  and  $t$  whereas  $\sigma_i(s, t)$  is the number of those paths that pass through node  $i$ .

**The Dataset:** Three datasets were used to validate the experiments, as detailed in Table 1.

**Table 1.** Summary statistics of used datasets

The Datasets	No. Nodes	No. Edges
Karate club	34	78
Dolphins online social network	62	159
Wikipedia who-votes-on-whom network	889	2914

Focusing on nodes that bridge multiple network clusters, and then maximizing their influence spread potential could be led by defining influencers based on their centrality score within overlapping communities. This new approach underscores the importance of nodes that serve as inter-community connectors in comparing to the state-of-the-art methods like global and local centrality measures, which evaluate nodes based on their individual connections or overall network paths. Such work [34] proposed a novel centrality measure combining degree centrality and network path characteristics. This method is to validate by Spearman and Pearson correlations on benchmark dataset. Such work showed superior performance in identifying influencers. In addition, this approach has been tested with the SIR model for virus spread and confirmed its sufficiency in predicting influential nodes capable of maximizing information diffusion. Another study [35] achieved high accuracy in diverse network scenarios by the efficacy of blending local and global centrality measures. Therefore, the traditional methods of identifying influential brokers are limited to consistently identify influencers across different network structures so they need for integrative improvements. In conclusion, introducing the overlapping nodes and enhanced centrality measures provide an excellent improvement over traditional benchmarks by effectively capturing nodes that utilizing extensive inter-community influence.

## 2. LITERATURE REVIEW

In this section, we follow the common approach for a literature review process, which includes stages like defining the research question, searching for relevant literature, screening and selecting sources, analyzing and synthesizing the findings, and finally comparing such existing related work and identifying the research gap of our proposed work. In this review, the literature scanned covers the past five years. Furthermore, the majority of these publications are from reputable journals, especially those published within the last three years, starting from 2022. Based on the chosen paper’s topic, some aspects are considered as objectives of our research question. The used dataset, methodology applied, results evaluation, and most importantly, the measures utilized in each work. In addition, the review methodology we follow is a synthesis manner, so that we criticize the related work and compare results in order to conduct our research question.

The work [36] present a study whose primary goal is to identify the most influential social media influencers in the food and beverage industry on Twitter by using centrality measures. The dataset used for the study consists of Twitter data with the hashtag #pizzahut, which was collected using the Twitter API. The work utilized four centrality measurements for the comparative analysis. Degree centrality to measure the number of connections of a node. Betweenness to quantify the node’s position as a bridge between other nodes. Closeness to find the closest nodes for others. There are some shortcomings in their work, it focuses only on Pizza Hut, which may limit how the findings can be applied more broadly. Also a lack of detail in the methodology, which could affect the study’s transparency and the ability for others to replicate the results.

Additional work [37] utilized three real popular networks: Les Miserables network (LesM), the Adjectives and Nouns network (Word), and Zachary network (Zachary). It is important to mention that the metrics utilized in this study are degree centrality, PageRank, HITS, and the global structure model. The SIRIR model as a ranking mechanism is employed to supply a susceptible infected recovered model and to measure the spreading influence of the top-ranking nodes. Each metric captures different aspects of node influence, and integrating them offers a more complete view of node centrality. In fact, individual centrality metrics may have biases or limitations in a particular network. Another issue is the computational complexity, making it challenging to apply on large networks, and the assumption of homogeneity, by which centrality indices may assume homogeneity in the network structure, causing the overlooking of the diversity of node roles and functions. Another study [38] applied their approach to the social network data from Huawei Company. Their results showed that different centrality measures do not produce similar influencers across different social media platforms. The betweenness, eigenvector, closeness, and degree centrality measures were primarily utilized in this work. The noticeable shortcoming of their work is the dataset size. It only included relationships among 1000 people which may limit the generalizability of the findings. However, their contribution was excellent with respect to their conclusion, which states that regardless of the number of individuals chosen as influencers, the Jaccard indices consistently remain below 20%, indicating a lack of similarity in identified influencers across platforms.

The author in corresponding, such as: Yang et al. [39] introduced a new model called the centrality measure, DCC. It

involves clustering, degree, coefficient, and neighbor information to detect influential nodes within large complex networks. This novel work offers highly enhanced results in identifying influential nodes in comparison with traditional centrality measures. The study utilized four real networks: USAir97, Karate club for social relationships, Email network, and Jazz musicians jazz musicians. The study utilized a variety of measures: degree centrality, closeness centrality, betweenness centrality, eigencentrality, K-shell centrality, Local centrality, and the centrality proposed in a previous study (NP) were applied for comparison with the DCC centrality measure. In terms of computational resources, the study [39] requires significant computational resources to analyze the datasets. A demonstrated work [40] utilized a dataset involving information on team attributes, relationships, and expertise within product development project organizations. There were a couple of indices used: modularity index, silhouette index, expertise overlap, and centrality indices to assess the clustering results and organizational network. This work was implemented by employing a social network analysis (SNA) approach to analyze the structural and attribute similarities among product development teams. Meanwhile, this work is not free of limitations. It may include assumptions made in modeling team attribute similarity, potential biases in data collection, and the complexity of interpreting social network analysis results. Also, it requires sufficient computational resources for analyzing large datasets and conducting social network analysis on complex organizational structures.

Based on social sites, an experimental study [41] used four real network datasets. These have a variety number of nodes and edges to evaluate the ranking efficiency and time consumption of the tested method. The study employed a clustering method in conjunction with centrality measures to rank nodes in large networks. Such a shortcoming suggests that future research directions to include exploring additional accuracy metrics, memory usage evaluation, and the complexity of centrality measures when combined with clustering algorithms. Another drawback is in large graphs, the computational complexity of these calculations can increase significantly, leading to longer execution times, and storing the necessary information for centrality calculations in large graphs can require substantial memory resources. Large number of datasets were utilized by the work [42]. The study applied 33 real-world networks from various domains; biological, social, infrastructural, collaboration, and ecological networks. The community structure of these networks was uncovered using the Louvain and Infomap detection algorithms. Also, this work applied diverse measures like; centrality measures were employed, including degree, closeness, maximum neighborhood component betweenness, and to quantify the importance of nodes within the networks. However, this work needs further research to address performance issues in networks with few and large communities and to adapt the ranking scheme for networks with overlapping community structures. Such work [43] used five data networks for evaluation, including Facebook, NS, Jazz, PB, Email, and USAir. The results were best compared to other algorithms, especially for datasets that have numerous links and closely related neighbor nodes using a new model called DKGM\_CLC algorithm. Although the study aimed to improve the gravity model by employing the local clustering methods to improve influential node detection in social networks and node position information, there are potential

biases or assumptions made during the model development and evaluation. So, future research should aim to test the model on a more extensive range of datasets to assess its generalizability across different network structures and characteristics. This work [44] utilized 13 real-world networks in their experiments for assessing the proposed LGI-VIKOR algorithm performance in complex networks for influential node detection. The metrics used are degree centrality, Shannon entropy, clustering coefficient, and the VIKOR method to evaluate the local and global influences of nodes within complex networks. The paper lacks detailed information on the implementation and computational aspects of the LGI-VIKOR algorithm. The significant contribution proposed by LGI-VIKOR is the enhancement of the traditional measures such as degree centrality and the k-shell by integrating local and global influences to assess node importance in complex networks comprehensively.

Another recent research [45] utilized four datasets: Cora, Citeseer, PubMed, and Amazon Computers. Each with varying numbers of nodes, edges, classes, and features. The study provided a significant advancement in influence maximization by the bet-clus method over traditional graph metrics for seed node selection. There are various metrics used: Betweenness centrality, clustering coefficient, closeness centrality, and degree centrality were compared with bet-clus for seed node selection based on network coverage. However, the study has some limitations in scalability for large networks and the need for heuristic solutions due to the complexity of the influence maximization problem. Three datasets used in the study [46] consist of online social network data, potentially from platforms like Facebook, Twitter, or Orkut. Meanwhile, centrality measures, such as Principal Component Centrality (PCC) and Eigen Vector Centrality (EVC), are utilized for node influence evaluation within the network. The Principal Component Analysis (PCA) was applied by this model to reduce the dataset dimensionality while preserving essential information. However, one of the most obvious limitations of the model is the focus on influential neighborhoods rather than individual nodes, which may overlook specific node-level influences. Researchers proposed a novel model for the GISR centrality measure, which is a parameterizable metric designed within influence spread models identifying influential actors in social networks [47]. The research was carried out on the Crime-Moreno dataset from the Network Repository. The difference in relevance of specific parameters actually testifies to the difference in network characteristics. Parameterizable centrality from GISR is highly appreciated because of this feature; researchers can adjust centrality measures to network characteristics and further to research requirements. GISR is, therefore, a method that may bring forth new insights into the location of influential actors for different social network contexts, through the adjustment of parameters such as depth levels and probabilities.

### 3. PROPOSED METHODOLOGY

The main research objectives include three steps. The first step is to create graphs (for evaluation purposes, the centrality of all graph nodes is measured using the four measures of centrality degree, eigenvector, betweenness, and closeness centrality).

The second step consists of detecting network communities and identifying overlapping nodes. Finally, in the third step,

the key influencers are selected from the overlapping nodes. The main question we are trying to answer is whether the overlapping (inter-community) nodes are often highly centralized compared to the centrality score of the remaining nodes. Figure 2 summarizes the proposed method and evaluation process.

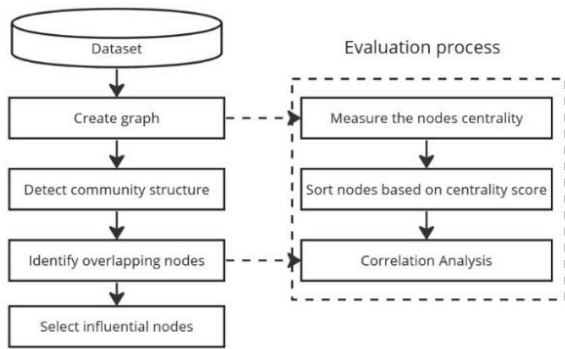


Figure 2. Summary of research methodology

### 3.1 Graph creation

To model social relations, the information imputed in each dataset is represented as undirected graph  $G(V,E)$ , where  $V$  is the set of nodes and  $E$  is the set of edges [48]. At the end of this step, three graphs are generated, one from each dataset. Figure 3 illustrates this on the three datasets.

### 3.2 Community detection

In this step, the communities are uncovered for every generated graph, and the associated overlapped nodes are identified. We adopt the "agglomerative hierarchical clustering based on maximal clique" (EAGLE) algorithm [7] to detect the community structure and, in turn, identify the overlapped nodes of the graph.

Table 2. The outcome of community detection

The Datasets	No. Communities	No. Overlapping Nodes
Karate club	3	3
Dolphins online social network	4	7
Wikipedia who-votes-on-whom network	21	42

The algorithm was utilized for its simplicity due to its minimal required parameters, and its quality function of modularity ensured accurate detection results for both overlapping and hierarchical communities. "Hierarchical means that communities may be further divided into subcommunities" [7]. Instead of dealing with a set of single vertices, the EAGLE algorithm uses an agglomerative framework to handle a list of maximal cliques. A clique that is not a part of any other clique is said to be the maximum. The EAGLE initially identifies all the maximal cliques in the network using the well-known Bron-Kerbosch parallel algorithm for its simplicity. To guarantee the algorithm's efficiency, it is necessary to discard certain cliques whose vertices are part of larger maximal cliques, and these are referred to as Subordinate maximal cliques. A dendrogram and a relevant cut are selected to divide the dendrogram into

communities [49]. Due to the fact that the densely linked community typically comprises a sizable clique that could be considered the community's nucleus and could be considered the core, the EAGLE is presented as an agglomerative hierarchical clustering algorithm. The computational complexity of this algorithm may not be the most efficient, but it is suitable for the networks size used in the experiment and yields satisfactory results. Table 2 illustrates the outcome of this step.

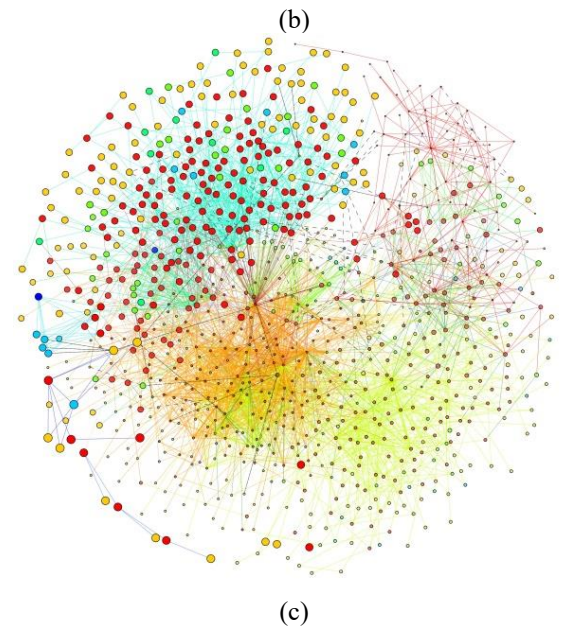
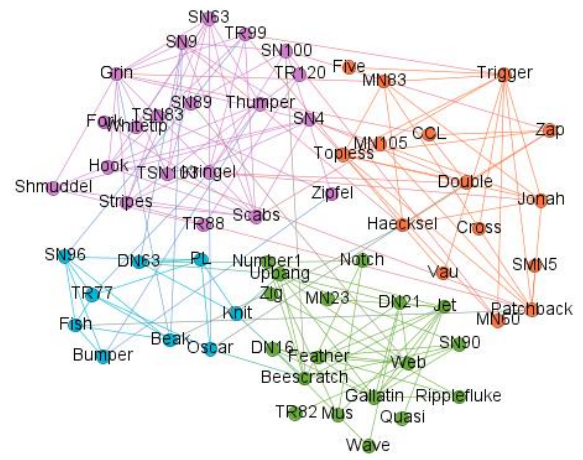
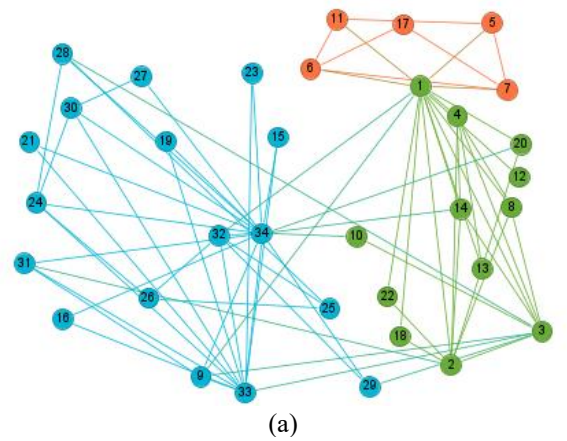


Figure 3. Visualizations of the community structure (a) Karate club (b) Dolphins online social network (c) Wikipedia who-votes-on-whom network

### 3.3 Centrality measure

For evaluation purposes, the four popular fundamental centrality measures have been used in the study: degree centrality, eigenvector centrality, closeness centrality, and betweenness centrality. In this step, each centrality measure that was previously explained is applied to the configured graphs, and the node centrality is calculated for each measure separately. As a result of this step, four lists of nodes are created for each graph (one for each centrality measure), in which the nodes are sorted in descending order based on the centrality score. The outcome of this step is used for comparison and validation tasks.

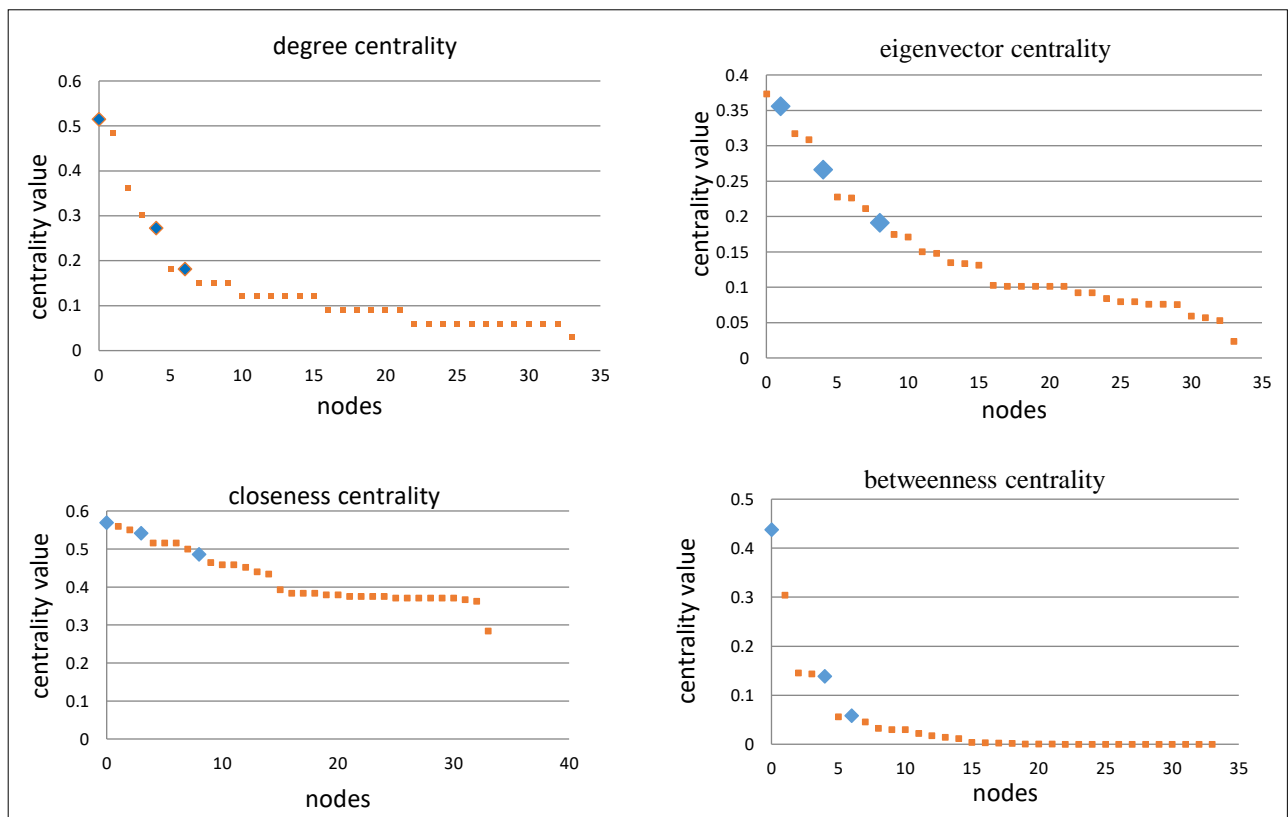
### 3.4 Correlation result and influential nodes selection

Based on the outcome of the previous steps, a comparison was conducted to investigate the relationship between node centrality and the characteristic of being one of the overlapping nodes.

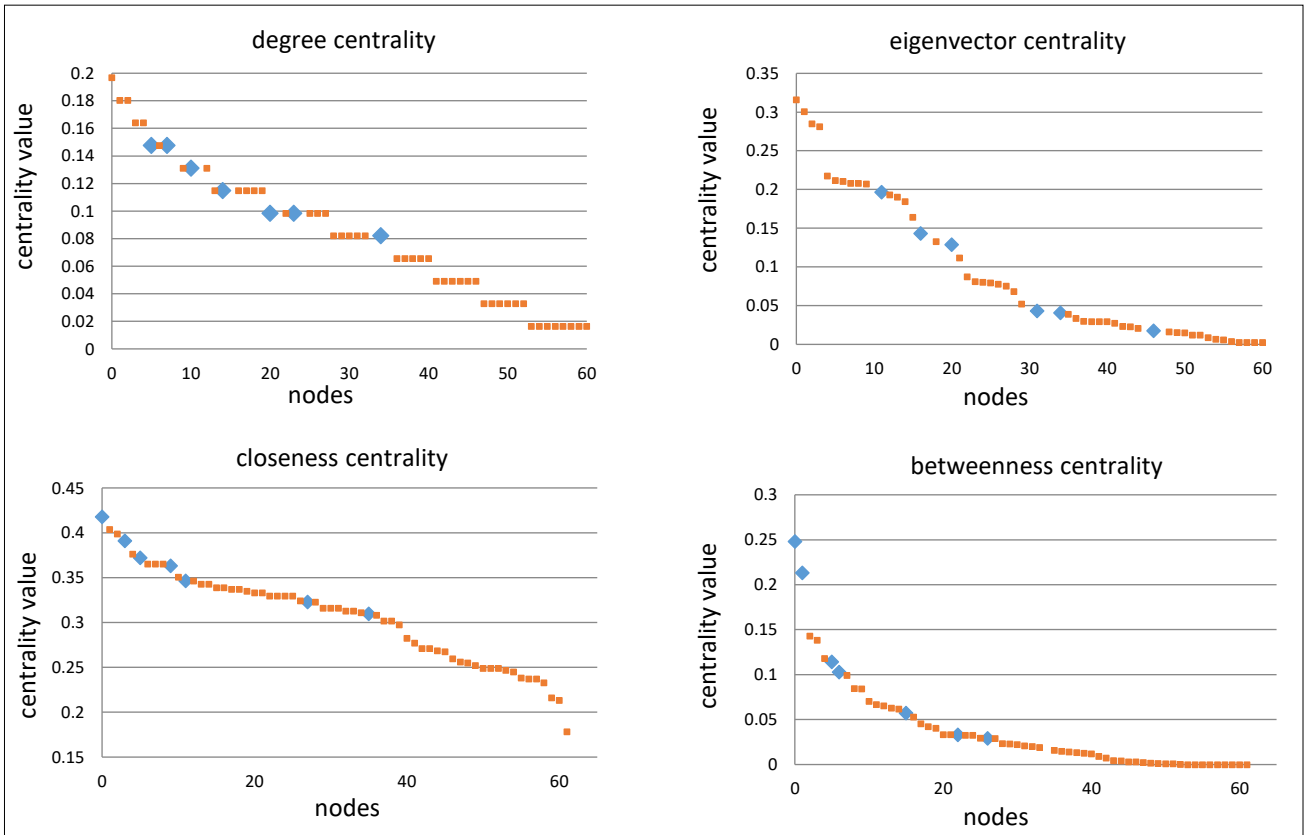
We claim that the overlapping nodes are characterized by high centrality scores compared with other nodes and they can be selected as influencer nodes directly without the need to find the centrality measure of graph nodes. It is important to confirm this claim and reveal the extent of the correlation, in addition to understanding the behavior of the overlapping nodes concerning each centrality measure. As explained earlier, the previous steps produced each dataset with four lists of nodes arranged based on their score on the basic four centrality measures, and the overlapping nodes are identified by their appearance in the sorted lists. We divided each of these lists into four quarters (Q1, Q2, Q3, and Q4) and counted the number of overlapping nodes in each quarter to examine the relationship between the degree of centrality and the merit that the nodes were overlapping (see Table 3). For instance, in Karate club graph, almost all the overlapping nodes are within the first quarter of the sorted nodes for all centrality measures. Figure 4(a-c) shows the distribution of nodes based on their centrality score in each centrality measure, where the overlapping nodes appear in blue color, and non-overlap nodes with red color.

**Table 3.** The number of overlapping nodes in each quadrant (Q) of sorted node list

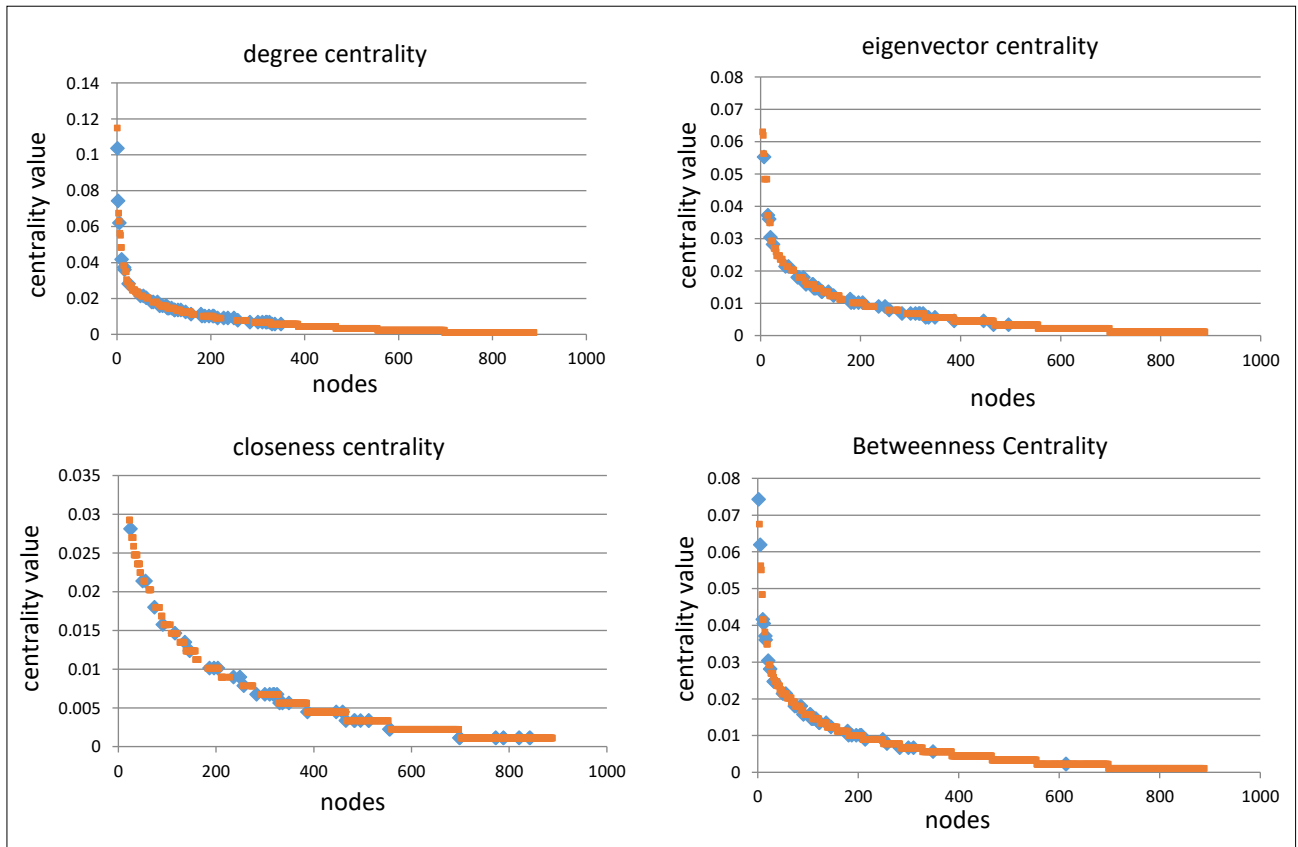
Mesures Dataset	Degree				Eigenvector				Closeness				Betweenness			
	Q1	Q2	Q3	Q4	Q1	Q2	Q3	Q4	Q1	Q2	Q3	Q4	Q1	Q2	Q3	Q4
Karate club.	3	0	0	0	2	1	0	0	2	1	0	0	3	0	0	0
Dolphin.	5	2	0	0	3	2	1	0	5	1	1	0	5	2	0	0
Wikipedia.	33	9	0	0	28	7	5	3	15	12	7	6	36	5	1	0



**Figure 4a.** The distribution of nodes on the basis of their centrality, Karate club



**Figure 4b.** The distribution of nodes on the basis of their centrality, Dolphins network



**Figure 4c.** The distribution of nodes on the basis of their centrality, Wikipedia dataset

## 4. DISCUSSION

In the Karate club dataset, the patterns across all four measures suggest that overlapping nodes are likely to be more central and influential in the network. In closeness centrality, the overlapping nodes tend to have higher closeness centrality values, which indicates they are generally closer to all other nodes in the network. It could mean they are more efficient in disseminating information within the network. With betweenness centrality, the overlapping nodes start with a higher betweenness centrality but then align with non-overlapping nodes as the node number increases. It suggests that the majority but not all of overlapping nodes are crucial in connecting different parts of the network. However, by Eigenvector Centrality, overlapping nodes show higher eigenvector centrality values for the initial nodes, but like betweenness centrality, this influence does not extend across all overlapping nodes. Finally, with Degree centrality, the overlapping nodes demonstrate a higher degree of centrality at the beginning, which sharply declines. It might mean that most overlapping nodes have many connections, but some do not maintain this level of direct connectivity. Thus, the overlapping nodes have the potential to be key influencers due to their higher centrality measures.

With the Dolphin dataset, Degree centrality reveals most nodes have low connectivity; overlapping nodes have marginally higher connections, hinting at their central role. And, Eigenvector centrality selects few nodes as influential, with overlapping status not strongly affecting influence. With closeness centrality, some overlapping nodes are central, potentially reaching others more quickly. However, by Betweenness centrality, an overlapping node stands out as a critical connector, significantly affecting the network's flow paths. The network has a few pivotal nodes that ensure connectivity, with overlapping nodes being slightly more central. Overall, the network is not overly dependent on any single node for its functionality. Across these centrality measures, the network appears to have a few nodes that are crucial in terms of connectivity and flow. The overlapping nodes sometimes have higher centrality scores, suggesting they might have a slightly more central role in network cohesion and flow. However, the network generally lacks a large number of highly central nodes, indicating that it's not heavily reliant on a small number of nodes to maintain its structure or function. This could suggest robustness against the removal of any single node but also a potential lack of efficiency in information or resource distribution.

With Wikipedia, in Degree Centrality, a small number of nodes (both overlapping and non-overlapping) have very high degrees, suggesting such nodes are connected with many other nodes. However, most of the nodes have significantly lower degrees, implying a network structure that might be skewed towards a few highly connected nodes. However, with the Eigenvector centrality, while some nodes have relatively high centrality values, indicating they are not only connected to many nodes but also to nodes that are themselves highly connected, the centrality scores decrease rapidly as we move down the list of nodes. In Closeness, the network might have a dense core of nodes that can broadcast information and resources across the network efficiently because of the shorter path lengths to other nodes. Meanwhile, the betweenness centrality a few nodes have significantly higher betweenness centrality scores than the rest. This implies that these nodes act as critical bridges within the network that control the

information and resources flow within the network. The steep decline in betweenness centrality values suggests that after these key nodes, the role of other nodes as intermediaries drops off quickly. Thus, we conclude that the presence of nodes with high degree and eigenvector centrality values indicates a network with influential hubs that are not connected to other well-connected nodes, enhancing their importance. The closeness centrality results highlight the efficiency of the network's core in disseminating information, while the betweenness centrality underscores the critical role certain nodes play in connecting disparate parts of the network.

In real-world applications like marketing and information diffusion, overlapping nodes as influential in social networks has high impact on such fields. Because the nodes in overlapping communities have high connections across multiple communities, they could possess high centrality scores and prone to hold more significant positions within the network. Such a position provides them ability to act as bridges, facilitating the information diffusion across diverse communities more efficiently than nodes confined to a single group. We would have campaigns more effective and widespread in the context of information dissemination due to targeting overlapping nodes can enhance the speed. In example, in public health campaigns, leveraging these nodes can achieve mission of crucial information to reach a broader audience more efficiency. Overlapping nodes can be considered an effective marketing strategies as they can enhance brand visibility and reach because these nodes naturally propagate information to a wider and more varied audience. This ability allows businesses to identify and target the best impactful nodes in their advertising campaigns. Our research has been motivated by recent work [50-54] within last four years. So, such work underscores the value of incorporating centrality measures and community structure into the process of influencers identification, thereby maximizing the impact of promotional efforts in social networks and optimizing resource allocation.

## 5. LIMITATIONS

Beside these significant findings, we would like to mention some limitations that should be outlined for future work. One, being the reliance on centrality measures could overlook other critical factors which influencing node influence, like the nature of connections or node activity levels. Also, the simulation models have inherent assumptions that may not fully capture real-world complexities. Future research needs to explore additional metrics beyond centrality and address these limitations by incorporating various types of networks. Furthermore, the improvements in the methodology may include the impact of dynamic network evolution on influencer efficacy and longitudinal studies to observe changes in influencer roles over time. This enhancement will offer comprehensive understanding of practical applications in social network analysis and influencer detection.

## 6. CONCLUSIONS

Through the experiments conducted on real networks and according to the proposed methodology, we conclude that overlapping nodes tend to exhibit higher centrality values across various measures compared to non-overlapping nodes.



However, It is worth mentioning that the overlapping nodes have a very high centrality of betweenness compared with the rest of the centrality measures. It suggests overlapping nodes play more central and influential roles in the network, acting as key connectors, information disseminators, and potentially critical influencers in network dynamics. Understanding and targeting these overlapping nodes can significantly impact network dynamics, connectivity, and overall network performance. Through the foregoing, it is possible to define the overlapping nodes as influencer nodes and avoid the high power of calculation needed to uncover the influencer by finding the centrality score of graph nodes. It is promising to take advantage of the feature that the nodes overlap among the communities in many applications, especially in social networking applications that aim to maximize the spread of information and other applications such as rumor control.

## REFERENCES

- [1] Al-Sabaawi, M. (2014). The comparative analysis of social network in international and local corporate business. *International Journal of Advances in Engineering & Technology*, 7(3): 723-732. [https://www.ijaet.org/media/9I21-IJAET0721299\\_v7\\_iss3\\_723-732.pdf](https://www.ijaet.org/media/9I21-IJAET0721299_v7_iss3_723-732.pdf).
- [2] Tabassum, S., Pereira, F.S., Fernandes, S., Gama, J. (2018). Social network analysis: An overview. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 8(5): e1256. <https://doi.org/10.1002/widm.1256>
- [3] Girvan, M., Newman, M.E. (2002). Community structure in social and biological networks. *Proceedings of the National Academy of Sciences*, 99(12): 7821-7826. <https://doi.org/10.1073/pnas.122653799>
- [4] Roy, A. (2011). Online communities and social networking. *IGI Global*, 2011: 74-83. <http://doi.org/10.4018/978-1-60566-014-1.ch145>
- [5] Porter, M.A., Onnela, J.P., Mucha, P.J. (2009). Communities in networks. *Notices of the American Mathematical Society*, 56(9): 1082-1097. <https://ssrn.com/abstract=1357925>.
- [6] Adalı, S., Lu, X., Magdon-Ismail, M. (2014). Local, community and global centrality methods for analyzing networks. *Social Network Analysis and Mining*, 4: 1-18. <https://doi.org/10.1007/s13278-014-0210-8>
- [7] Shen, H., Cheng, X., Cai, K., Hu, M.B. (2009). Detect overlapping and hierarchical community structure in networks. *Physica A: Statistical Mechanics and its Applications*, 388(8): 1706-1712. <https://doi.org/10.1016/j.physa.2008.12.021>
- [8] Devi, J.C., Poovammal, E. (2016). An analysis of overlapping community detection algorithms in social networks. *Procedia Computer Science*, 89: 349-358. <https://doi.org/10.1016/j.procs.2016.06.082>
- [9] Tsapatsoulis, N., Anastasopoulou, V., Ntalianis, K. (2019). The central community of Twitter ego-networks as a means for fake influencer detection. In 2019 IEEE Intl Conf on Dependable, Autonomic and Secure Computing, Intl Conf on Pervasive Intelligence and Computing, Intl Conf on Cloud and Big Data Computing, Intl Conf on Cyber Science and Technology Congress (DASC/PiCom/CBDCCom/CyberSciTech), Fukuoka, Japan, pp. 177-184. <https://doi.org/10.1109/DASC/PiCom/CBDCCom/CyberSciTech.2019.00042>
- [10] Rodríguez-Vidal, J., Gonzalo Arroyo, J., Plaza Morales, L. (2020). Detecting Influencers in social media using information from their followers. *Procesamiento del Lenguaje Natural*, 64: 21-28. <https://doi.org/10.26342/2020-64-2>
- [11] Shang, Q., Deng, Y., Cheong, K.H. (2021). Identifying influential nodes in complex networks: Effective distance gravity model. *Information Sciences*, 577: 162-179. <https://doi.org/10.1016/j.ins.2021.01.053>
- [12] Britton, T. (2020). Epidemic models on social networks—With inference. *Statistica Neerlandica*, 74(3): 222-241. <https://doi.org/10.1111/stan.12203>
- [13] Alasadi, M.K., Arb, G.I. (2021). Community-based framework for influence maximization problem in social networks. *Indonesian Journal of Electrical Engineering and Computer Science*, 24(3): 1604-1609. <http://doi.org/10.11591/ijeecs.v24.i3.pp1604-1609>
- [14] Askarizade, M., Najafi, S., Tork Ladani, B. (2022). Modeling and analysis of rumor control strategies in social networks. *Computer and Knowledge Engineering*, 5(1): 59-68. <https://doi.org/10.22067/cke.2022.74885.1049>
- [15] Yang, L., Li, Z., Giua, A. (2020). Containment of rumor spread in complex social networks. *Information Sciences*, 506: 113-130. <https://doi.org/10.1016/j.ins.2019.07.055>
- [16] Rajeh, S., Savonnet, M., Leclercq, E., Cherifi, H. (2021). Comparing community-aware centrality measures in online social networks. In *International Conference on Computational Data and Social Networks*, pp. 279-290. [https://doi.org/10.1007/978-3-030-91434-9\\_25](https://doi.org/10.1007/978-3-030-91434-9_25)
- [17] Seyfosadat, S.F., Ravanmehr, R. (2023). Systematic literature review on identifying influencers in social networks. *Artificial Intelligence Review*, 56(Suppl 1): 567-660. <https://doi.org/10.1007/s10462-023-10515-2>
- [18] Ma, T., Liu, Q., Cao, J., Tian, Y., Al-Dhelaan, A., Al-Rodhaan, M. (2020). LGIEM: Global and local node influence based community detection. *Future Generation Computer Systems*, 105: 533-546. <https://doi.org/10.1016/j.future.2019.12.022>
- [19] Rehman, A.U., Jiang, A., Rehman, A., Paul, A., Din, S., Sadiq, M.T. (2023). Identification and role of opinion leaders in information diffusion for online discussion network. *Journal of Ambient Intelligence and Humanized Computing*, 14(11): 1-13. <https://doi.org/10.1007/s12652-019-01623-5>
- [20] Mnasri, W., Azaouzi, M., Romdhane, L.B. (2021). Parallel social behavior-based algorithm for identification of influential users in social network. *Applied Intelligence*, 51(10): 7365-7383. <https://doi.org/10.1007/s10489-021-02203-x>
- [21] Sabharwal, S., Kaur, P. (2024). Identification of efficient central nodes using centrality measures within community structures. In *2024 2nd International Conference on Disruptive Technologies (ICDT)*, Greater Noida, India, pp. 1062-1067. <https://doi.org/10.1109/ICDT61202.2024.10489597>
- [22] Bendahman, N., Lotfi, D. (2023). A novel centrality based measure for influential nodes detection in social networks. In *2023 10th International Conference on Wireless Networks and Mobile Communications (WINCOM)*, Istanbul, Turkiye, pp. 1-7.

- <https://doi.org/10.1109/WINCOM59760.2023.10322968>
- [23] Rashmi, C., Kodabagi, M.M. (2023). Influential node detection in online social networks using skill features of user profiles. In 2023 International Conference on Network, Multimedia and Information Technology (NMITCON), Bengaluru, India, pp. 1-6. <https://doi.org/10.1109/NMITCON58196.2023.10276247>
- [24] Chung, L., Park, K.Y. (2022). Implementation of computing node centrality with bridge and community detection. *Journal of Student Research*, 11(3): 2703. <https://doi.org/10.47611/jsrhs.v11i3.2703>
- [25] Zhao, Y., Deng, K., Liu, X., Yao, J. (2023). Overlapping community detection algorithm based on high-quality subgraph extension in local core regions of network. *Wireless Communications and Mobile Computing*, 2023(1): 4988601. <https://doi.org/10.1155/2023/4988601>
- [26] Bendahman, N., Lotfi, D. (2024). Unveiling influence in networks: A novel centrality metric and comparative analysis through graph-based models. *Entropy*, 26(6): 486. <https://doi.org/10.3390/e26060486>
- [27] Ayoub, J., Lotfi, D., Hammouch, A. (2022). Mean received resources meet machine learning algorithms to improve link prediction methods. *Information*, 13(1): 35. <https://doi.org/10.3390/info13010035>
- [28] Freeman, L.C. (2002). Centrality in social networks: Conceptual clarification. *Social network: Critical concepts in sociology*. Londres: Routledge, 1: 238-263. [https://doi.org/10.1016/0378-8733\(78\)90021-7](https://doi.org/10.1016/0378-8733(78)90021-7)
- [29] Rajeh, S., Savonnet, M., Leclercq, E., Cherifi, H. (2021). Investigating centrality measures in social networks with community structure. In *Complex Networks & Their Applications IX: Volume 1, Proceedings of the Ninth International Conference on Complex Networks and Their Applications COMPLEX NETWORKS 2020*, pp. 211-222. [https://doi.org/10.1007/978-3-030-65347-7\\_18](https://doi.org/10.1007/978-3-030-65347-7_18)
- [30] Ruhbau, B. (2000). Eigenvector-centrality – A node-centrality?. *Social Networks*, 22(4): 357-365. [https://doi.org/10.1016/S0378-8733\(00\)00031-9](https://doi.org/10.1016/S0378-8733(00)00031-9)
- [31] Landherr, A., Friedl, B., Heidemann, J. (2010). A critical review of centrality measures in social networks. *Business & Information Systems Engineering*, 2: 371-385. <https://doi.org/10.1007/s12599-010-0127-3>
- [32] Sharma, S., Purohit, G.N. (2012). A new centrality measure for tracking online community in social networks. *International Journal of Information Technology and Computer Science*, 4(4): 47-53. <http://dx.doi.org/10.5815/ijitcs.2012.04.07>
- [33] Zhang, J., Luo, Y. (2017). Degree centrality, betweenness centrality, and closeness centrality in social network. In 2017 2nd International Conference on Modelling, Simulation and Applied Mathematics (MSAM2017), pp. 300-303. <https://doi.org/10.2991/msam-17.2017.68>
- [34] Klein, D.J. (2010). Centrality measure in graphs. *Journal of Mathematical Chemistry*, 47(4): 1209-1223. <https://doi.org/10.1007/s10910-009-9635-0>
- [35] Goh, K.I., Oh, E., Kahng, B., Kim, D. (2003). Betweenness centrality correlation in social networks. *Physical Review E*, 67(1): 017101. <https://doi.org/10.1103/PhysRevE.67.017101>
- [36] Wang, Y., Li, H., Zhang, L., Zhao, L., Li, W. (2022). Identifying influential nodes in social networks: Centripetal centrality and seed exclusion approach. *Chaos, Solitons & Fractals*, 162: 112513. <https://doi.org/10.1016/j.chaos.2022.112513>
- [37] Mukhtar, M., Abas, Z., Rasib, A. (2023). Global structure model modification to improve influential node detection. *Journal of Engineering and Applied Sciences*, 18(3): 220-225.
- [38] Dong, Y., Xu, G., Meng, L., Yang, P. (2022). A novel algorithm for finding influential nodes in complex networks based on communication probability and relative entropy. *Physica A: Statistical Mechanics and its Applications*, 603: 127797. <https://doi.org/10.1016/j.physa.2022.127797>
- [39] Yang, Y., Wang, X., Chen, Y., Hu, M., Ruan, C. (2020). A novel centrality of influential nodes identification in complex networks. *IEEE Access*, 8: 58742-58751. <https://doi.org/10.1109/ACCESS.2020.2983053>
- [40] Ali, F., Waffaa, M.A., Mustafa, A.H. (2024). Applying hybrid clustering with evaluation by AUC classification metrics. *International Journal of Computing and Digital Systems*, 15(1): 1091-1102. <http://doi.org/10.12785/ijcds/150177>
- [41] Marjai, P., Szabari, B., Kiss, A. (2021). An experimental study on centrality measures using clustering. *Computers*, 10(9): 115. <https://doi.org/10.3390/computers10090115>
- [42] Rajeh, S., Cherifi, H. (2022). Ranking influential nodes in complex networks with community structure. *Plos One*, 17(8): e0273610. <https://doi.org/10.1371/journal.pone.0273610>
- [43] Van Duong, P., Dinh, X.T., Son, L.H., Van Hai, P. (2022). Enhancement of gravity centrality measure based on local clustering method by identifying influential nodes in social networks. In *International Conference on Multimedia Technology and Enhanced Learning*, pp. 614-627. [https://doi.org/10.1007/978-3-031-18123-8\\_48](https://doi.org/10.1007/978-3-031-18123-8_48)
- [44] Qiu, L., Liu, Y., Zhang, J. (2024). A new method for identifying influential spreaders in complex networks. *The Computer Journal*, 67(1): 362-375. <https://doi.org/10.1093/comjnl/bxac180>
- [45] Saxena, R., Jadeja, M., Vyas, P. (2023). An efficient influence maximization technique based on betweenness centrality measure and clustering coefficient (Bet-Clus). In 2023 15th International Conference on Computer and Automation Engineering (ICCAE), Sydney, Australia, pp. 565-569. <https://doi.org/10.1109/ICCAE56788.2023.10111177>
- [46] Kumar, R., Sharma, U. (2021). Influential nodes in online social network: A principal component centrality based approach. In 2021 9th International Conference on Reliability, Infocom Technologies and Optimization (Trends and Future Directions)(ICRITO), Noida, India, pp. 1-9. <https://doi.org/10.1109/ICRITO51393.2021.9596353>
- [47] Riquelme, F., Vera, J.A. (2022). A parameterizable influence spread-based centrality measure for influential users detection in social networks. *Knowledge-Based Systems*, 257: 109922. <https://doi.org/10.1016/j.knosys.2022.109922>
- [48] Alamsyah, A., Rahardjo, B. (2021). Social network analysis taxonomy based on graph representation. *arXiv preprint arXiv:2102.08888*. <https://doi.org/10.48550/arXiv.2102.08888>
- [49] Bron, C., Kerbosch, J. (1973). Algorithm 457: Finding

- all cliques of an undirected graph. *Communications of the ACM*, 16(9): 575-577. <https://doi.org/10.1145/362342.362367>
- [50] Zhou, L., Jin, F., Wu, B., Chen, Z., Wang, C.L. (2023). Do fake followers mitigate influencers' perceived influencing power on social media platforms? The mere number effect and boundary conditions. *Journal of Business Research*, 158: 113589. <https://doi.org/10.1016/j.jbusres.2022.113589>
- [51] Azaouzi, M. (2023). Influential nodes detection in multiplex social networks: A comprehensive study. <http://dx.doi.org/10.2139/ssrn.4658122>
- [52] Kim, J., Kim, M. (2022). Rise of social media influencers as a new marketing channel: Focusing on the roles of psychological well-being and perceived social responsibility among consumers. *International Journal of Environmental Research and Public Health*, 19(4): 2362. <https://doi.org/10.3390/ijerph19042362>
- [53] Hajarathaiyah, K., Enduri, M.K., Anamalamudi, S., Subba Reddy, T., Tokala, S. (2022). Computing influential nodes using the nearest neighborhood trust value and pagerank in complex networks. *Entropy*, 24(5): 704. <https://doi.org/10.3390/e24050704>
- [54] Zhuang, Y.B., Li, Z.H., Zhuang, Y.J. (2021). Identification of influencers in online social networks: measuring influence considering multidimensional factors exploration. *Heliyon*, 7(4): e06472. <https://doi.org/10.1016/j.heliyon.2021.e06472>