Machine Learning for Human Intelligence
2024, Vol. 2, Issue 2

Research Article

Clustering the Digital Footprint: Revealing User Interaction Communities

Numan Asif[1,⋆], Iqra Rashid[2], and Adnan Ali[3]

[1,2,3]Department of Computer Science, University of Management and Technology, Lahore, Pakistan.; Email: numan.asif@post.umt.edu.pk, iqra.rashid@yahoo.com, aadnanali@gmail.com
⋆Corresponding author: Numan Asif (numan.asif@post.umt.edu.pk)

Article History

Academic Editor:
Dr. Muhammad Sajid
Submitted: March 21, 2024
Revised: May 1, 2024
Accepted: September 1, 2024

Keywords:
Clustering, communities, social networks, closeness centrality, eigenvector centrality, strong entities, and weak entities

Abstract

The proliferation of social networking sites (SNS) and the expansion of the internet have enabled seamless communication among individuals on a unified platform. A graph with nodes and edges connecting the nodes might represent a social network. The nodes symbolize individuals or entities, while the edges illustrate their interactions. Individuals who associate inside social networks and share analogous decisions, tastes, and preferences create virtual clusters or communities. Identifying these communities can be advantageous for several objectives, such as discovering a common study domain in collaborative networks, identifying a target audience for marketing and recommendations, and mapping protein interaction networks inside biological systems. Community identification serves as an effective instrument for discerning intricate networks. Diverse methodologies have been suggested for community detection, each addressing the issue from a distinct viewpoint. Consequently, extensive graph-processing community recognition techniques have become essential due to the emergence of vast and intricate networks across several areas. This research presents an innovative method for community detection that integrates node space similarity and utilizes local knowledge. We utilize eigenvector centrality and proximity metrics to improve community detection in social networks. Comprehensive studies on both synthetic and real-world networks demonstrate the effectiveness of the suggested hybrid paradigm. The results demonstrate that the hybrid technique is more effective in large-scale graphs compared to other established algorithms, exhibiting significant robustness and efficiency.

1 Introduction

As more and more of our daily activities are conducted online, there is an increasing need for social data. People can interact and voice their thoughts on goods and policies via social media platforms [1]. Therefore, everyone from heads of state to small business owners uses them as a source of information. Social media platforms make everything available to a global audience without regard to demographic limitations. People now congregate in communities and organizations to communicate and exchange information in a virtualized social environment made possible by the widespread use of social media [2]. A social network is a type of networking referred to by this designation. Currently, some of the most

 Â© 2024 Copyright by the Authors.

prominent platforms are Instagram, LinkedIn, Facebook, and Twitter, among others. The research on these networks expands the boundaries of transdisciplinary fields. Researchers and scientists must perform extensive data calculations due to the vast volume of data generated by the regular use of this communication method [3]. Social network analysis (SNA) facilitates the examination of social phenomena within a specific social context. The study predominantly utilizes data from a limited community or social networking group [4, 5].

A cohort of readers with a shared interest in publications of the same subject matter and age demographic plans to enroll in an introductory college course [6]. Graph theory is a popular method for modeling the relationships and interactions of elements or entities in real systems. In mathematics, a graph consists of a set of vertices interconnected by a set of edges [7]. Graph theory attributes are utilized to comprehend user behavior, consumer preferences, and interactions [8]. Furthermore, learner interactions in social learning environments are defined by graph methodologies. Understanding network science and its applications is essential for representing and assessing data derived from social networks [11, 12]. These nodes are designated as leaders who possess exceptional skills in community building [13, 14]. Currently, the predominant focus in Social Network Analysis (SNA) is the identification of communities and significant nodes due to their applications in recommender systems [16], e-learning [15], and healthcare [17]. Community discovery (CD) is the process of identifying groups of users inside a network who share analogous traits. Community detection is employed to ascertain the network's structure and functionality by identifying the distinct connections among the nodes [18]. To do this, three methodologies may be employed: utilizing topological features, incorporating supplementary node and edge data, or integrating both [19]. The temporal complexity and size limitations of social networks render the selection of an appropriate community structure a formidable challenge. Although certain methods from the aforementioned categories can rapidly analyze large-scale graphs, they may also uncover subpar community structures [20]. Elevated modularity signifies that the community detection algorithm effectively clustered the nodes into high-density, functionally distinct communities [21]. Communities inside a network are discovered utilizing the proposed methodology. The procedure commences with an input network depicted by an adjacency matrix. Consequently, prominent nodes are identified by their extensive connections and numerous interactions [22]. It has demonstrated considerable efficacy in transforming high-dimensional graphs into continuous, dense, low-dimensional vector spaces [16].

The centrality metrics indicate the significance of nodes inside the network framework. Influential nodes characterized by high eigenvector centrality and strategically situated nodes exhibiting high proximity centrality constitute primary communities [23]. Graph embeddings, such as DeepWalk and node2vec, learn representations that uphold the proximity and structural responsibilities of nodes. Nodes with analogous vector representations are likely to possess comparable network neighbors, signifying community affiliation. The current modularity-based approaches face several challenges [24]. For example, they can't find communities that are smaller than a certain size because of a resolution problem. This means that they miss complex structures within large networks. Another technique tends to create large, poorly defined communities by grouping different nodes, resulting in excessively broad and less significant community boundaries. Along with centrality metrics and graph embeddings [25], this work created a hybrid model that aims to improve the limitations of modularity-based approaches. We utilized the concept of centrality measurements, specifically eigenvector centrality, to discover highly influential nodes that constitute the center of tiny networks. These nodes can serve as the primary mechanism for identifying smaller groups that modularity may overlook. Researchers have shown that graph embeddings hold complex structural details, which makes it easier to find subtle patterns that separate small communities [26] from the larger network. Proximity centrality, according to research, effectively identifies nodes centrally situated within a community, leading to a more precise delineation of community boundaries. Such an effect can inhibit the establishment of excessively expansive communities. The methods we look at are modularity-based methods that focus on global network structure, centrality, and graph embeddings that give us information at the local and node levels. Integrating these methodologies provides a more holistic perspective of the network [5].

Graph embedding is an exceptionally efficient method for tackling challenges in network analysis. Moreover, the objective of graph embedding is to convert a network into a lower-dimensional vector space while maintaining the structural aspects of the network [27]. Moreover, in a low-dimensional space, it is feasible to represent the nodes proximal to the network with a uniform vector. This streamlines responsibilities related to recognizing and classifying communities. The proposed paradigm consists of three separate phases. We first create an embedding space in which the nodes are shown as vectors. We identify nodes with exceptional capacity for community formation and significant influence over others through degree centrality measurements. Subsequently, we establish a preliminary community structure by clustering nodes that exhibit the highest similarity to the prominent nodes within the same community, as determined by the Jaccard coefficient in the embedding space. In the concluding stage, the robust communities are amalgamated with the frail communities that were excluded from the initial community structure developed in the second phase.

1.1 Motivation

This proposed method seeks to detect possible personalities that may impact social network communities. Influencer targets may encompass users with elevated eigenvector centrality to establish brand connections among networks. A further source of motivation was uncovering the interconnections among other cultures. Individuals with elevated closeness centrality can serve as intermediaries between distinct communities, facilitating communication and the exchange of knowledge. Comprehending these connections may facilitate the formulation of strategies to promote collaboration among groups or to disseminate information more efficiently across the network. Integrating graph clustering with eigenvector and proximity centrality may enhance our understanding of information flow within these communities, as well as the interactions among users and their influence on one another.

1.2 Research Problem

Despite the plethora of user interaction data offered by online platforms, comprehending how individuals engage and establish communities continues to pose a challenge. Traditional approaches to community identification may inadequately reflect the complexities of user behavior and information dissemination. This research proposal seeks to establish a more thorough methodology for detecting concealed communities inside user interaction networks.

- What is the level of connectivity among users inside designated user societies?
- Who are the principal contributors in every the neighborhood?
- Can a hybrid methodology integrating eigenvector and closeness centrality improve the precision and comprehension of social network recognition?
- Which findings may graph segmentation offer on user actions and interaction routines?

1.3 Contribution

The primary innovations and contributions of our work are summarised as follows:

- Community cohesiveness is assessed by measures like edge weight, adaptability, and the mean grouping ratio.
- We analyze the impact of these individuals in molding user experiences, disseminating knowledge, and connecting different communities.

2 RELATED WORK

2.1 An Index of Community Detection Methodologies

Community detection has been the subject of extensive research [20], and various algorithms [29] exist for this purpose. The methods can be classified into the following categories: modularity-based methods, spectral analysis-based methods, hierarchical structures, clustering methods, random walk methods, label-propagation methods, graph-based methods, and information-theoretic measure methods [30].

2.2 Modularity-Driven Group Recognition

According to Equation 1, the Girvan-Newman algorithm [31] employed modularity, a recognized standard metric, to detect communities within the network. Subsequently, modularity served as the foundation for the development of further algorithms. These algorithms produce robust comparative results and are widely utilized across various fields, including product recommendations and research group identification [32]. The modularity metric is modified and linked to the spanning tree for community detection [33].

$$Qm = \frac{1}{2n} \sum_{in,jn} \left[A_{injn} - \frac{Km_{in} * km_{jn}}{2n} \right] \delta\left(C_{in}, C_{jn}\right) \tag{1}$$

A_{injn} represents the adjacency matrix between vertices in and jn. n indicates how many edges there are in the graph. C_{in} indicates the class that is associated with node i. As stated in Equation 2, the Kronecker delta is (C_{in}, C_{jn}). It equals 1 in the case when c1 equals c2 and 0 otherwise.

$$\delta\left(C_{in}, C_{jn}\right) = \begin{cases} 1 & \text{if } in \text{ and } jn \text{ in similar community} \\ 0 & \text{else} \end{cases} \tag{2}$$

A density-based technique is an alternative method for community discovery [34]; however, this method requires the algorithm to accept the resolution parameter as input. By recognizing and addressing its deficiencies, the society attains more cohesion [35]. Nonetheless, the application of this technique results in inferior modularity and NMI performance measures for certain networks in comparison to alternative algorithms.

2.3 Evaluating Contemporary Community Detection Methods

Various methodologies have been employed historically to tackle the community detection issue. Communities can be identified by several methodologies, including networks, modularity, mathematical models, and evolutionary computing. Instances of these models encompass fuzzy logic [38], matrix factorization [37], and statistical methods [36]. Clans [40], local communities [39], and network embedding [41] exemplify applications of the network approach in research. The modularity technique optimizes community quality, as indicated by Louvain [42], Leiden [43, 44], Girvan Newman [31], and Greedy modularity [45]. Evolutionary computing techniques employ abstract principles from biological evolutionary theory to create optimization algorithms or methodologies. This method combines the tenets of biological evolution with computational technologies, including particle swarm optimization [47] and genetic algorithms [46]. Nonetheless, several strategies utilized to achieve this maximum modularity yield suboptimal results. Moreover, several methods generate groups of either substantial or negligible size that may lack practical significance. Some algorithms demonstrate reduced adaptability to alterations in the network, especially those involving the addition or removal of edges or nodes. The results differ when diverse methods are utilized to examine a network for community identification. Each technique produces unique modularity and community results [48].

2.4 Challenges in Community Detection: Surpassing Local Optima

Multiple factors may contribute to the emergence of local optima in community detection. Due to a resolution restriction, modularity-based community recognition algorithms may overlook small communities [49]. The generalized modularity density technique may discern communities of diverse sizes and configurations by assessing node density within the network [50]. Modularity utilizing Z-scores, which standardizes the modularity score, is a further method that may detect communities of differing sizes [51]. A significant concern is the inadequate community infrastructure [52]. Researchers have proposed many techniques, like concealed community identification and poor supervision, to address this issue [53]. Hidden communities denote clandestine or obscure groupings that are challenging to locate through traditional community detection methods. An alternative method for recognizing community structure is weak supervision, employing the node2vec technique [54] to detect communities of diverse sizes and shapes. Communities with a low degree of embedding present challenges in identification, as noted by [55] in their research on node2vec.

2.5 Hybrid Approach for Community Detection Utilizing Enhanced Modularity

Most techniques for community detection utilize modularity and similarity metrics independently. The proposed hybrid method utilizes the modularity of network measurements, such as proximity and eigenvector centrality, to improve the ultimate community structure. Moreover, our approach surpasses prior methods regarding collaborative results, NMI metrics, and node categorization. It also exhibits exceptional modularity.

3 Prelude and Definition

This section provides a concise overview of the proposed hybrid model and illustrates an example of the utilized graph measurements. The proposed paradigm has five stages. Initially, we establish an embedding space in which vector representations of the nodes are located. Through degree centrality assessments, we uncover nodes with a notable ability to establish communities and exert considerable influence over others. Furthermore, we employ the Jaccard coefficient similarity within the embedding space to cluster nodes that closely resemble prominent nodes within the same community, so establishing an initial community structure. Moreover, the communities are categorized into classifications that are either weak or strong. Moreover, the less resilient groups that were initially excluded from the community structure generated during the s phase are integrated with the more robust communities. Finally, the ultimate communities are identified and prioritized based on modularity and NMI criteria.

3.1 Problem Denotation

This study depicts a community of people as an unweighted and undirected graph $G_r = (N_o, E_d)$, which is composed of a set of m_e edges $E_d \subset N_o x N_o$, where $E_d = u_i, u_j \epsilon \frac{N_o}{2}$. The nodes in the social network are their users, and the edges represent their ties or interactions with each other. In this case, our objective is to divide graph G into a collection of separate communities, ensuring that each user u_i in the neighborhood N_o is distinct inside a community. The primary goal is to identify a community arrangement in which users exhibit strong connections with other users within the same community $Ci\epsilon D$ while having weak connections with users in different communities $Cj \neq i\epsilon D$. Furthermore, our objective is to identify significant actors or leaders inside each community $Oi\epsilon D$ to improve our comprehension of the internal organization of each community.

3.2 Significance of Nodes

3.2.1 Adjacency Matrix

The matrix representing the adjacency A of graph G $G = (N, D)$ is a nxn matrix, where $N = u_1, u1_2, ..., u_n$ and $E = E_{ui}|(u_i, u_j)\epsilon N$. $AG = [a_i, j]1 \leq i, j \leq n$ as shown in Equation 3

$$a_{i,j} = \begin{cases} 1 & \text{if } e_{u_i,u_j} \in E \\ 0 & \text{otherwise} \end{cases} \tag{3}$$

3.2.2 Degree of Node

The degree of a graph $G = (N, E)$ is the number of edges that connect a node. Equation 4 provides the algorithm for computing the degree of a collection of neighbors of a node, represented by (u):

$$\mathrm{Deg}_G(u_i) = |\rho_G(u_i)| = |\{u_j \in N \mid a_{i,j} = 1\}| \tag{4}$$

where $a_{i,j} = 1$ denotes the presence of an edge between u_i and u_j, and $|PG(u_i)|$ is the cardinality of the collection of neighbors. Formally speaking, the degree of node $u_i \epsilon N$ with $AG = [a_{i,} xn$ is given in Equation 5:

$$\mathrm{Deg}_G(u_i) = \sum_{j=1}^{n} a_{ij} \tag{5}$$

Users engaging in a higher volume of interactions than their counterparts may wield greater influence and enjoy enhanced access to information. Individuals possessing the highest educational attainment within the network are seen as active nodes, or hubs, proficient in disseminating knowledge within a certain region of the graph. In community detection, it is crucial to concentrate on these nodes, as they are generally the most significant and possess a high likelihood of creating communities.

Table 1: Time complexity of different centrality measures

Approach	**Centrality Measure**	**Time Complexity**
Freeman et al. [63]	Degree Neighbors based Centrality	$0(n)$
Freeman et al. [64]	Closeness diameter based Centrality	$0(n.log(n) + n.m$
Borgatti et al. [65]	Eigenvector values based Centrality	$0(n^3)$
Borgatti et al. [65]	Betweenness flow based Centrality	$0(n^3)$

3.2.3 Node Degree Centrality

The relative significance of a vertex inside the network is quantified by a metric referred to as degree centrality. To enable comparison, it is often beneficial to normalize the degree value specified in the equation. 6. The degree centrality of node $u_i \epsilon N$ is represented as $DC_G(u_i)$ whenever there is an adjacency matrix $AG = [a_{i,} xn$.

$$DC_G(u_i) = \frac{\mathrm{Deg}_G(u_i)}{n-1} = \frac{1}{n-1} \sum_{j=1}^{n} a_{ij} \tag{6}$$

Eigenvector centrality, betweenness centrality, and proximity centrality are among them. Each statistic signifies a unique focal point of attention. The centrality measures are thoroughly elucidated in Table 1.

3.2.4 Closeness Centrality

According to [66], the closeness centrality CC_e of a node in a network is calculated by taking the reciprocal of the total length of the shortest paths that connect the node to all other nodes. This calculation may be seen in Figure 3. Equation 7 provides the estimated normalized CC_e of node j.

$$CC_e[j] = \frac{N_o - 1}{\sum_{k=1}^{N_o} d_G(j,k)} \tag{7}$$

The value of $|V|$ is equal to N. The closeness centrality values of the nodes are often denoted as the CC_e vector when the $CC_e[j]$ values are organized into a vector of length N. Importantly, the normalized CC_e of (1) adheres to the fundamental principle of centrality, wherein a greater CC_e value signifies greater significance. However, for the sake of making things easier, we take into account the combined length of all the shortest routes, as given by Equation 8, from each given node to every other node:

$$dl[k] = \sum_{j=1}^{N_o} dl_G(k,j) \tag{8}$$

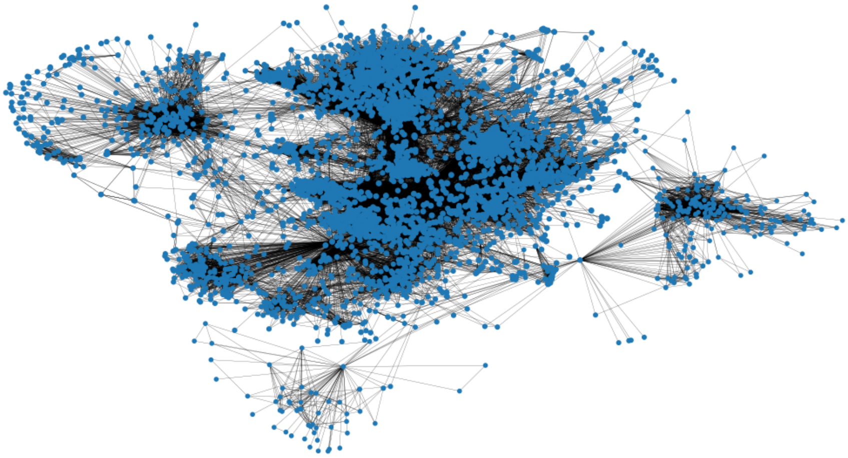

Figure 1: Community with Closeness Centrality

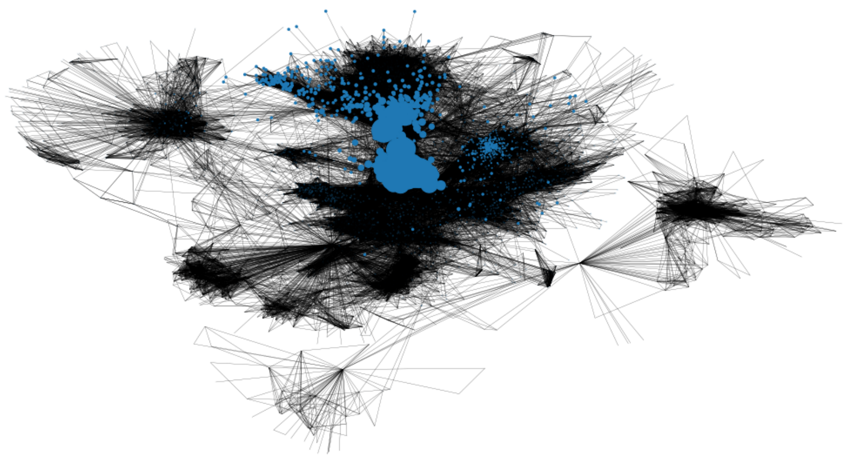

Figure 2: Community with Eigenvector Centrality

3.2.5 Eigenvector Centrality

Eigenvector centrality [67], illustrated in Figure 2, accounts for the significance of neighboring nodes in conjunction with the total count of neighboring nodes. In contrast, degree centrality is calculated solely by tallying all directly connected nodes, as delineated in Equation 9. In eigenvector centrality, the significance of links varies. The influence of an individual is typically more significant in connections with prominent individuals than with those of lesser influence. In addition to its connections, the score of the connected node (eigenvector centrality) is significant in eigenvector centrality. Eigenvector centrality is determined by evaluating an individual's connectedness to the network's most significantly related components. Individuals with elevated eigenvector analysis scores exhibit extensive connectivity, with several connections extending to the network's conclusion. Eigenvector dominance of the adjacency matrix is referred to as eigenvector centrality. Google's PageRank is a form of eigenvector centrality, developed by [68]. SCAN++ is based on the finding that real-world graphs, such as web graphs, exhibit elevated clustering coefficient scores [69]. The density of a node is ascertained by its clustering coefficient [70]. It is anticipated that a node and its two-hop-away node, especially in real-world graphs [71], will have a substantial overlap in their neighbors. This feature underpins SCAN++'s pruning of the density assessment for shared nodes between a node and its two-hop distant node.

$$Av = \lambda v \tag{9}$$

An eigenvector of a square matrix A is a non-zero vector v such that the product of A and v yields a scalar multiple of v. The constant multiple is generally denoted by the symbol λ. The eigenvalue λ is associated with the vector v in matrix A. Individuals with elevated eigenvector centrality occupy leadership roles within the network. They are frequently prominent individuals possessing an extensive network of associations with other distinguished figures. Consequently, they often function as prominent thinking leaders. Conversely, individuals with high Eigenvector centrality may not consistently fulfill responsibilities characterized by high betweenness and high closeness. The time complexity of integrating proximity centrality and eigenvector centrality for community detection can be reduced by mitigating the computational bottlenecks associated with each metric. We diminished the temporal complexity of proximity centrality through the application of the random sampling technique. Consider use a random sampling strategy rather than calculating closeness centrality for each user. This necessitates fewer computations and yields a satisfactory estimate of average proximity centrality inside the network. To diminish the temporal complexity of eigenvector centrality measurements, we utilized iterative approaches. It asserts that iterative techniques, such as the Power Method, can be employed to compute eigenvector centrality. Although these methods may require additional iterations to achieve convergence, they frequently outperform the explicit computation of the eigenvectors of the adjacency matrix in terms of speed.

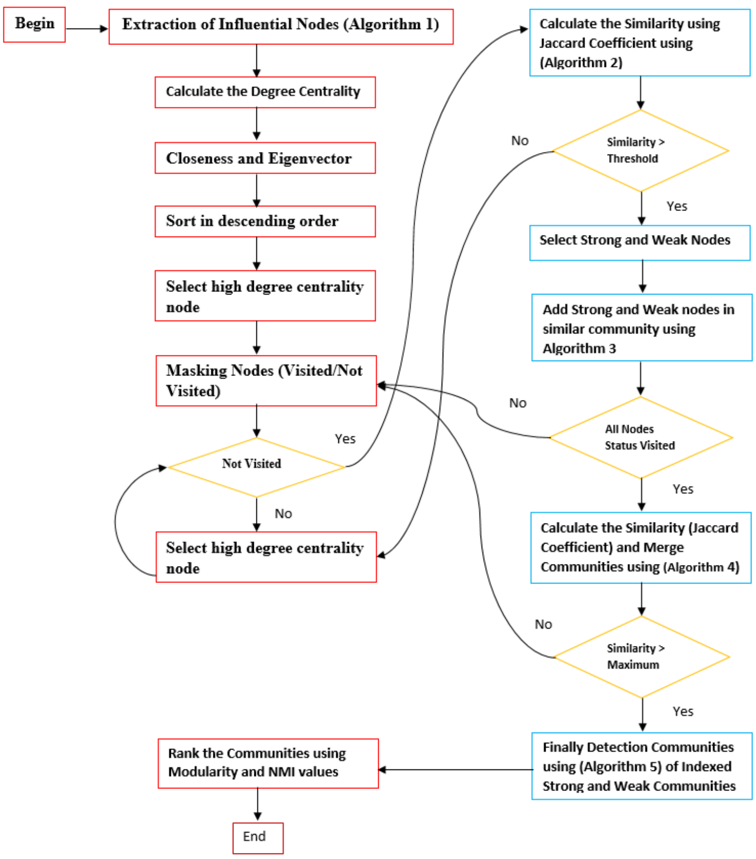

Figure 3: Flow diagram for the selection of community

4 Proposed Ideology

We delineate the essential terminology of our proposed hybrid model and, thereafter, provide an in-depth explanation of each step of the model. Figure 3 illustrates the flow diagram for the Proposed Hybrid paradigm. The proposed methodology consists of the following primary steps:

1. Performing the extraction of influential nodes and the generation of an embedding space.
2. Determining the initial configuration of the community.
3. Choosing strong and weak communities.
4. Community final merging.
5. Community detection and ranking based on modularity and NMI values.

The degree of a node i, represented as di, is the count of edges linked to node i, computed as $di = \sum Aij$. The degree matrix of the graph G is defined as D = diag(d1, d2, ..., dn), where D is a diagonal matrix containing the degrees of each node on its diagonal. The Laplacian matrix L of the graph G is defined as $L = D - A$. The diagonal elements of L correspond to the degrees of the nodes, $Lii = di$, whereas the off-diagonal elements are represented as $Lij = -Aij$. Given that A is symmetric for an undirected graph, L is likewise symmetric, and the sum of each row in L is zero. The Laplacian matrix L is recognized as positive semi-definite, indicating that all its eigenvalues, collectively termed the Laplacian spectrum, are nonnegative. We represent the eigenvalues in descending order as in Equation 10:

$$\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n \tag{10}$$

4.1 Nodes with considerable influence

It facilitates the identification of individuals who are significantly pertinent to various professions due to their capacity to swiftly spread knowledge and information within the network. This research uses Laplacian eigenmaps for dimensionality reduction. This technique utilizes vertices and edges to project the vertices into a lower-dimensional space (d). The parameter "d" regulates the number of features utilized to represent each node in the embedded space. By selecting a smaller value of d in the network, we condense the information regarding the node into a more succinct representation to efficiently identify the influential nodes. A larger "d" will preserve more information but will be less compressed. Consequently, we select a lesser value of d. To mitigate computational complexity, we diminish the dimensionality of space for various tasks. This work employs the Laplacian Eigenmap approach, which aims to preserve the links between nodes in a lower-dimensional space, where the parameter (d) affects the fidelity of these interactions. The centers of the communities serve as potent nodes. Finally, we use the LE method to build the embedding space after putting each node in the network in decreasing order based on its degree centrality value. The depiction of nodes as vectors in the embedding space facilitates the analysis of the network's structure and connections. Upon the establishment of the embedding space, each node within the network is designated the label Not visited, signifying that they have not yet been allocated to a community. This document outlines the essential phases in community detection.

1. After determining the level of centrality of each node, place the nodes in decreasing order.
2. Create the embedding space using the Laplacian Eigenvectors technique.
3. Give each node a "No visited" flag.

4.2 Preliminary Community Identification

The similarity is computed within the embedding space generated by the Laplacian Eigenvectors technique, employing the Jaccard Coefficient metric. Nodes exhibiting significant similarity to the chosen core should be categorized together. To do this, we allocate the core community to all connected nodes whose similarity value surpasses a certain threshold, S. Upon establishing the initial community and designating its members as "Visited," we advance to the next unvisited node with the highest centrality rating. In summary, comprises a sequence of meticulously organized steps:

1. Select the node with the maximum centrality and a state of Not-visited.
2. Calculate the Jaccard Coefficient of Similarity between the most influential node and the remaining nodes that have the status of not-visited.
3. Combine the most important node in the community with additional nodes that are comparable to it, and then designate them as "Visited."
4. Continue doing the identical procedures for the subsequent significant vertex that has not been visited until all the vertices in the graph have been marked as visited.

To establish the preliminary community structure, we initially select the node exhibiting the highest degree centrality as the core and compute its similarity with the remaining nodes in the graph. The similarity is calculated in the embedding space generated by the LE algorithm utilizing the jaccard coefficient measure using Equations 11 - 13. The objective is to cluster nodes that exhibit a significant degree of similarity with the chosen core. To accomplish this, we allocate all analogous nodes to the core that possess a similarity value over a specified threshold k to the core's community. Upon the completion of the initial community, we designate the status of its members as "Visited" and proceed to the next unvisited node exhibiting the highest degree centrality. We implement the identical procedure on the second designated core and persist until every node in the graph has been designated as "visited." The preliminary community structure is now established.

4.3 Identification of Robust and Fragile Communities

Weak communities are smaller than external ones; they may consist of singleton communities or exhibit reduced interactions among members. Therefore, certain acquired communities must be merged to achieve the ideal community structure for the graph.

4.4 Consolidation of Communities

The predominant techniques for optimizing community structure have been articulated in the literature and focus on maximizing or decreasing a specific target function. The subsequent phase, detailed, involves identifying communities that were deficient in the initial community structure and amalgamating them with robust communities. To minimize the temporal complexity as much as feasible, we calculated the Jaccard Coefficient similarity [72] between the cores of strong communities and the members of weak communities. The initial stage in merging weak and strong communities is to find the Jaccard Coefficient similarity between each weak community's nodes and each strong community's core that can be calculated using Equation 11.

$$\mathrm{sim}(u,v)^{\mathrm{Jaccard}} = \frac{|N_u \cap N_v|}{|N_u \cup N_v|} \tag{11}$$

N_u and N_v represent the collection of things that users u and v, respectively, have rated. The addition rule theorem is applied in this case to form $|N_u \cap N_v| = |N_u| + |N_v| - |N_u \cap N_v|$, since N_u and N_v are not mutually exclusive. On the other hand, according to Equation 12, $|Nu|$ and $|Nv|$ represent the cardinality of the sets N_u and N_v, respectively.

$$\mathrm{sim}(u,v)^{\mathrm{Jaccard}} = \frac{|N_u \cap N_v|}{|N_u| + |N_v| - |N_u \cap N_v|} \tag{12}$$

Suppose $|\bar{N}_u|$, $|\bar{N}_v|$ are the cardinality of the set of items un-co-rated by users u and v respectively. Hence, $|\bar{N}_u| = |N_u|$ - $|N_u \cap N_v|$ and $|\bar{N}_v| = |N_v|$ - $|N_u \cap N_v|$. As a result, the Jaccard similarity can be written as in Equation 13,

$$\text{sim}(u,v)^{\text{Jaccard}} = \frac{|N_u \cap N_v|}{(|\overline{N_u}| + |N_u \cap N_v|) + (|\overline{N_v}| + |N_u \cap N_v|) - |N_u \cap N_v|} = \frac{|N_u \cap N_v|}{|\overline{N_u}| + |\overline{N_v}| + |N_u \cap N_v|} \tag{13}$$

Modularity is the primary criterion to consider in discussions of community detection. The modularity of a network quantifies its capacity for division into distinct groups [73]. Optimization frameworks employ modularity to identify community networks. This concerns the discrepancy between the actual and expected amounts of edges. The notation used in Equation 14 represents modularity Q:

$$Q = \sum_i \left(e_{ii} - a_i^2\right) \tag{14}$$

Two communities should be merged if there are a greater number of connections between them compared to other groupings as can be extracted using algorithm . The variable l_{ij} is defined as the count of inter-community linkages between C_i and C_j, as stated in Equation 15.

$$l_{ij} =| (v_i, v_j) : v_i \in C_i and v_j \in C_j | \tag{15}$$

We use Equation 16 to determine if, among all the communities in the community setting, the community C_j and C_i should merge.

$$S_{ij} = \frac{l_{ij}}{dc_i dc_j} \tag{16}$$

Let Qm_j represent the community set's modularity before merging. If $Qm_j > Qm$, merge Com_i and Com_j into a single community to update the community structure. This process should be continued until there is no more room for improvement in modularity; at that time, the resulting community structure will have the highest feasible modularity. Inter- and intra-community edges are used to visually portray the identified communities to improve understanding of the relationships between nodes and communities.

5 Performance Evaluation Metrics

Modularity can be utilized to evaluate the results of multiple algorithms and identify the optimal method through community discovery.

5.1 Modularity

The preliminary metric is well-acknowledged in the literature. This method contrasts the real connections within a community against the probability of encountering those links in a randomly generated network. The efficacy of a network is maximized when there is a high concentration of connections within communities and a low concentration of connections between communities. The division with the highest modularity score is considered the most optimal in this context. Equation 17 delineates the modularity of division D given a graph G as follows:

$$Q(D) = \sum_{i=1}^{|D|} \left(e_{ii} - a_i^2\right) \tag{17}$$

The likelihood of an intra-community link in the community C_i is denoted by e_{ij}, while the likelihood of a relationship with at least one extremity is indicated by a_i. The information that is normalized mutually (NMI) Normalized mutual information (NMI), normalized to a number between 0 and 1, is used to determine the amount of information about two variables. The NMI is calculated using Equation 18, which involves taking the logarithm of the ratio between the joint probability of communities U and V and the product of the probabilities of each community, denoted as $logPUV(i,j)\ PU(i)PV(j)$. Values approaching 1 imply a robust connection between two variables, while values approaching 0 signify a feeble one.

$$\mathrm{NMI}(U,V) = \frac{2\sum_{i=1}^{R}\sum_{j=1}^{C} P_{UV}(i,j)\log\frac{P_{UV}(i,j)}{P_U(i)P_V(j)}}{-\sum_{i=1}^{R} P_U(i)\log P_U(i) - \sum_{i=1}^{R} P_V(i)\log P_V(i)} \tag{18}$$

5.2 Clustering Coefficient

We now give the primary observation utilized in both network average and global grouping coefficient estimators.

$$\mathbb{E}\left[\phi_k f(x_k)\right] = \sum_{i=1}^{n} p_i \mathbb{E}\left[\phi_k f(x_k) \mid x_k = i\right] \tag{19}$$

$$= \sum_{i=1}^{n} \frac{d_i}{D}\frac{2l_i}{d_i^2} f(v_i) \tag{20}$$

$$= \sum_{i=1}^{n} \frac{1}{D}\frac{2l_i}{d_i} f(v_i). \tag{21}$$

Equations 19 - 21, shows the initial equality is a consequence of the law of total expectation. The second equality is valid because there are 2 equal probability combinations $xk-1, vi, xk+1$, of which only 2 form a triangle vj, vi, vk or a reverse triangle vk, vi, vj. Observe that in a triangle or a reverse triangle, vj is connected to $vk(Aj, k = 1)$. The third equality is established using algebraic manipulation.

6 Experiments and Results

This section delineates the datasets and evaluates the performance of the most advanced community detection algorithms. The primary objective of this axis is to conduct an experimental investigation to determine the feasibility of our idea. We achieve this by evaluating the model's efficacy on both simulated and actual networks. We employ industry-standard metrics, such as modularity and normalized mutual information (NMI), as performance indicators.

Table 2: Experimental Settings

Configuration	Parameters
CPU	Intel(R) Xeon(R) Gold 6154
GPU	NVidia GeForce RTX 3090
Operating System	Windows 10
Environment	CUDA 11.1
Development Platform	Jupiter Notebook
Python Library	PyTorch
Video Memory	32GB
System Drive	1TB

6.1 Experimental setup

Table 2 shows the hardware requirements for the proposed model. The code is written in Python, and the other techniques were executed using the Python igraph package [74]. The network module in Python [75] is utilized to enhance the visualization of the detected communities.

6.2 Real World Datasets

1. The network of Zachary's Karate Club is examined in the study by [76]. Zachary established a concrete network by leveraging the social relationships among the 34 members of a karate group. The network has been divided into two portions due to a political dispute between the club's leadership and the instructor. For this investigation, we employ the fundamental iteration of this network.

2. Between 1994 and 2001, a social network known as The Dolphin's Network [77] documented that 62 bottlenose dolphins seen in New Zealand consistently established associations with each other. Two groups exist inside the network.

6.3 Experimental Results

Several specific modifications are required for the proposed algorithm to operate more effectively. Employing the Eigenmaps methodology, we initially identify the most significant nodes. The acquired vectors in the embedding space are utilized in the second and third phases of the proposed methodology to develop and enhance the original community structure [35, 37]. Consequently, the value of d directly influences the identification of communities. Consequently, it significantly influences the efficacy of the proposed model. The real network architecture will adjust this number to ascertain the suitable dimension d. The subsequent phase of the suggested methodology involves clustering the node-representation vectors according to their similarity. The objective is to create a preliminary community structure by clustering related nodes, which can subsequently be refined in the third phase. A node is considered part of the community of an influential node if there is a significant degree of similarity between them. In the trial phase, a node is classified as a member of the core community if the similarity between the two nodes exceeds 0.8. It is essential to note that this value was employed by every analyzed network [40]. We applied Algorithm to six real-world networks possessing established

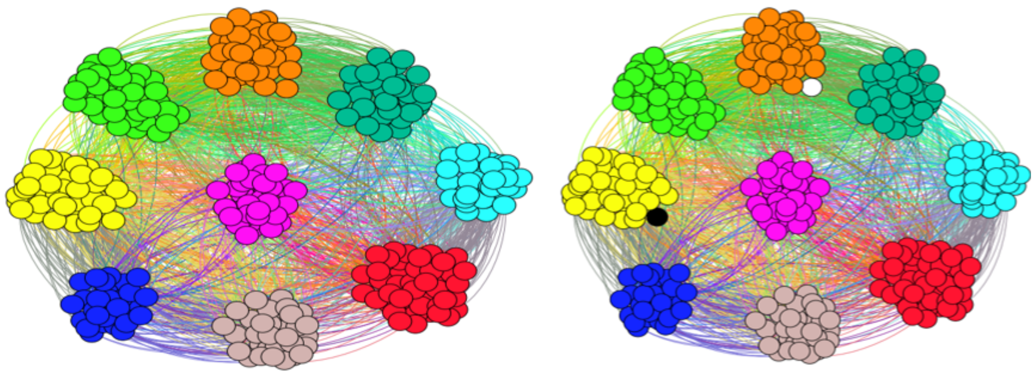

Figure 4: Karate Club Community structure (Left) Ground Truth (Right) Community Structure detected by Hybrid Model

community structures. The identified community structures were evaluated for modularity and Normalized Mutual Information (NMI) utilizing Algorithm and advanced methodologies. The results of Algorithm for each network on the list were examined and reported separately. The graphics illustrate the ground truth and the community visualizations produced by Algorithm for each data collection. The designated communities are marked with various colors. This section delineates the intermediary phases for implementing algorithm to obtain the final community structure from the initial community structure of the karate club network. In each step, two communities are chosen and subsequently merged, based on the quantity of edges both internally and externally among the communities. The communities will reconvene if the modularity continues to improve. The karate club network depicted in Figure 4 was analyzed using algorithm , yielding many communities in the outcome, in contrast to the two present in the ground truth. Nonetheless, the modularity and NMI surpass those of the alternative approaches [29, 31]. Figure 5 illustrates the resultant dolphin social network, comprising six communities as opposed to the four communities of the ground truth. The NMI of the identified communities is superior compared to the various algorithms. Moreover, modularity is superior to the Louvain [42] and Infomap [82] algorithms. The communities exhibiting the highest modularity and normalized mutual information have been identified, among other methodologies. Our approach utilizing methods yields superior NMI and modularity in comparison to the ground truth, despite variations in the number of identified communities. We employ the suggested algorithm in conjunction with associated approaches [20, 27] to extract communities from the six real-world networks delineated in Table following the collection of the experimental setup and data. The proposed approach effectively

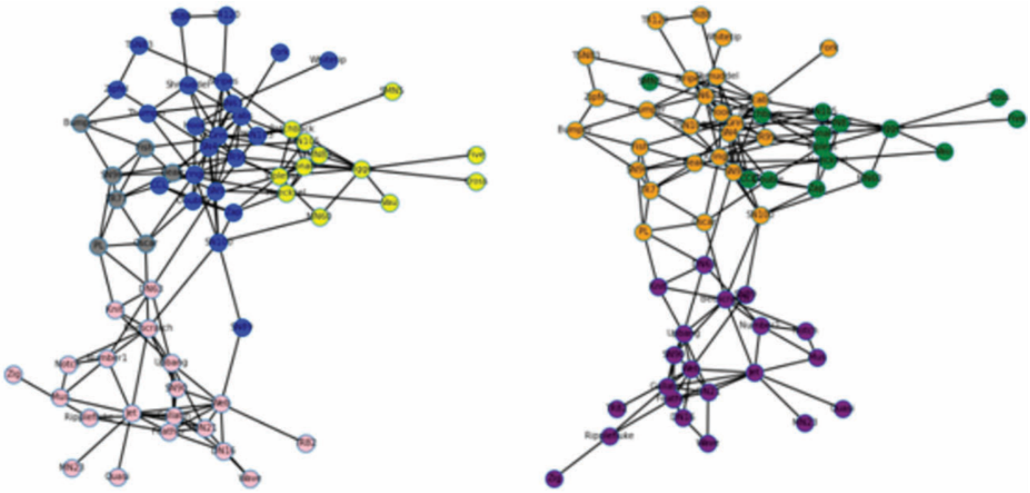

Figure 5: Dolphin community structure (left) versus ground truth (right) community structure identified by the hybrid model.

detects the communities inside the Karate network, characterized by distinct membership attributes, as seen in Figure 4. This is further substantiated by the NMI metrics, illustrating that the proposed algorithm outperforms alternative methods in achieving superior values. This figure 4 illustrates that the proposed technique performs more effectively for the Karate network, however its inferior value compared to the Walktrap [81], FLPA [22], Louvain [42], and Spinglass [80] algorithms. Irrespective

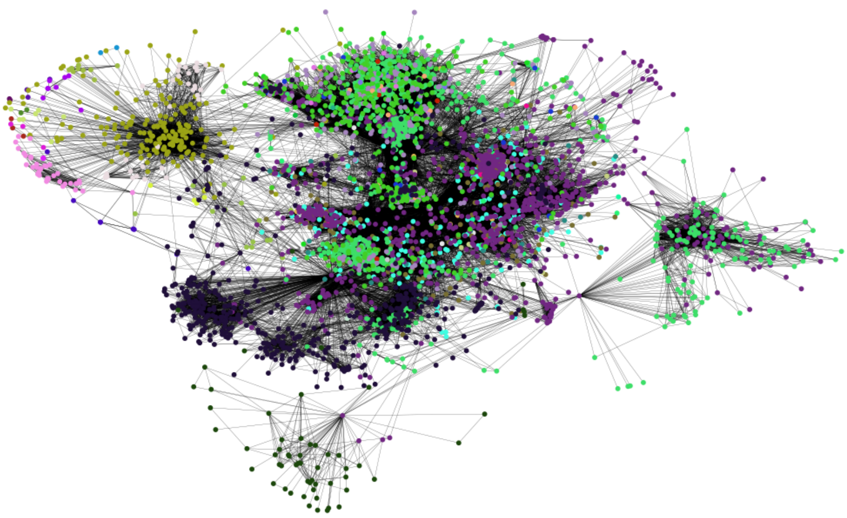

Figure 6: Clustering of different Communities

of dataset size, our strategy yields a community structure that aligns more closely with the ground truth than competing methodologies. Our proposed hybrid model algorithm exhibits greater stability than other algorithms, such as Louvain [42], FLPA [22], and Spinglass [80]. This is accurate as our concept does not depend on a random process. Moreover, our proposed hybrid model performs effectively in dense graphs, yielding substantial NMI values, which is recognized as a crucial advantage for community detection in social networks. Figure 6 illustrates the community structure indicated by the hybrid model. In summary, as articulated in [19], the rapid identification of a community structure that closely resembles the actual structure is emphasized. Consequently, our methodology excels at identifying prominent communities within social networks.

The experiment selects baseline approaches, encompassing many classic community detection methods and established clustering techniques to illustrate the superiority of the recommended method. To evaluate the efficacy of the proposed strategy, it will be juxtaposed with established embedding-based baselines and existing graph embedding methodologies [1, 2, 3] for community discovery, into which networks will be integrated. Subsequently, we derive community divisions from the low-dimensional vector space obtained by. We employed the modularity value (Q) and the Normalized Mutual Information (NMI) metric in the evaluation process as outlined in [6, 8, 9]. The technique attains all six highest NMI values and all five highest Q values in the real-world community datasets. Despite a minor mismatch between the Q values derived from the Ego-Facebook dataset and those from the Dolphins dataset, the maximum NMI value of 1 is achieved, indicating that it accurately represents the correct community for the actual categorization. In the DBLP dataset, the proposed method outperforms alternative techniques [4, 5] and produces more precise results, achieving the highest NMI value of 0.91. The Amazon dataset has elevated Q and NMI values, with the NMI approaching 1, indicating greater alignment with data from authentic communities. DBLP exhibited the most favorable outcomes, demonstrating superior logic and effectiveness relative to other algorithms, which produced the least modularity.

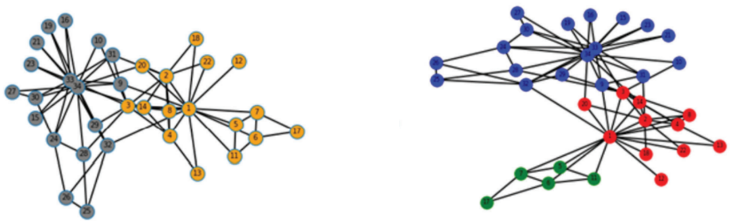

Figure 7: (a) Ground Truth Karate Club Community, (b) Karate Club Community detected by the proposed method

6.4 Community Visualization and Evaluation Results

In Figure 7, the karate club network was analyzed using the suggested model, yielding three communities, although the ground truth had only two communities. However, the modularity and NMI surpass those of the other algorithms. Figure 8 illustrates the football social network, revealing nine

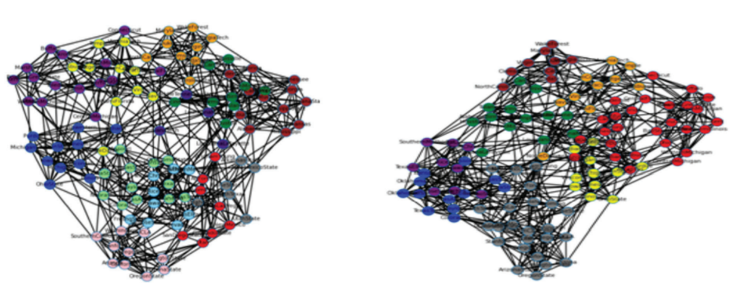

Figure 8: (a) Ground Truth Football network Community, (b) Football network Community detected by the proposed method

identified communities, but twelve communities existed in actuality. The modularity and NMI for the communities identified by the proposed model are superior to those of existing algorithms. Although the quantity of identified communities varies from the ground truth, our model demonstrates superior modularity and normalized mutual information (NMI).

6.5 Discussion

The results show that picking the right centrality-based clustering method is very important in social networks, recommendation systems, and fraud. Degree and betweenness centrality are valuable metrics that are particularly effective for identifying key interaction hubs or influencers. The experiment demonstrates that the clustering coefficient is not a robust predictor of significant clusters. This means that eigenvector centrality and proximity centrality are no longer important. Degree and betweenness have a strong correlation value of 0.85 to 0.86, which means that nodes with direct connections act as bridges in the network. These measures do a good job of showing how users interact within clusters, though not as well as when eigenvector centrality is added. The clustering coefficient exhibits a negative correlation with degree (-0.47) and betweenness (-0.42), and an even stronger negative correlation with itself (-0.98). The result suggests that nodes in densely clustered regions are less likely to operate as pivotal connectors or to affect significant interaction patterns. In the absence of eigenvector centrality, we forfeit insights on influential users linked to other influential users, which may be vital in hierarchical or authority-driven networks. Eigenvector centrality is a way to figure out how powerful a node is on a global scale by looking at both the number of connections and how those connections are linked to other important nodes. Eliminating it may impair the capacity to identify central figures in intricate networks, such as influencers in social media and pivotal decision-makers in communication networks.

6.6 Comparison of Performance Evaluation of Proposed Model with Existing Approaches

Table 3: Comparison of modularity and NMI for different approaches

Ref	Modularity (%)	NMI (%)	Degree (%)	Betweenness (%)	Clustering Coefficient (%)
Palacio [83]	91	90	81	82	80
Yuan [84]	92	89	82	74	85
Farrokh [85]	86	89	81	83	84
Werner [86]	85	88	83	89	85
Cao [87]	89	87	84	78	74
Proposed	95.05	92.50	91.62	90.65	86

Table 3 indicates that the study [83] attained a modularity of 91%, a normalized mutual information (NMI) of 90%, a degree of 81%, and a betweenness of 82%. The evaluation metric values of alternative methods exhibit modest variations. Table 3 illustrates that study [86] attained a modularity of 85%, an NMI of 88%, a Degree of 83%, and a Betweenness of 89%, whereas the proposed model reached a modularity of 95.05%, an NMI of 92.50%, a Degree of 91.62%, and a Betweenness of 90.65%. The experimental results indicate that the clustering coefficient has minimal impact on community discovery. Nonetheless, in certain instances, the clustering coefficient can be significant.

7 Conclusion

The emergence of social networking sites and the proliferation of the Internet have enabled effortless contact among individuals on a unified platform. The nodes signify individuals or entities, while the edges illustrate their interactions. Individuals who engage in social networks and have analogous decisions, tastes, and preferences establish virtual clusters or communities. Community identification serves as an effective mechanism for discerning intricate networks. Scholars have proposed a variety of methodologies for community detection, each approaching the problem from a unique perspective. Large-scale graph-based community detection techniques have had to be created because of the growth of large and complex networks in many areas. This paper presents an innovative method for community finding that integrates node space similarity with localized knowledge. We utilize eigenvector centrality and proximity metrics to improve community detection in social networks. This research has introduced a five-phase hybrid methodology for identifying communities within social networks.

References

[1] Bolorunduro, J. O., & Zou, Z. (2024). Community detection on multi-layer graph using intra-layer and inter-layer linkage graphs (cdmiilg). *Expert Systems with Applications*, *238*, 121713.

[2] Yang, C., Liu, J., Yan, Y., & Shi, C. (2024). FairSIN: Achieving fairness in graph neural networks through sensitive information neutralization. In *Proceedings of the AAAI Conference on Artificial Intelligence* (Vol. 38, No. 8, pp. 9241-9249).

[3] Gadár, L., & Abonyi, J. (2024). Finding multifaceted communities in multiplex networks. *Scientific Reports*, *14*(1), 14521.

[4] Cavoretto, R., De Rossi, A., Lancellotti, S., & Romaniello, F. (2024). Node-bound communities for partition of unity interpolation on graphs. *Applied Mathematics and Computation*, *467*, 128502.

[5] Christensen, A. P., Garrido, L. E., Guerra-Peña, K., & Golino, H. (2024). Comparing community detection algorithms in psychometric networks: A Monte Carlo simulation. *Behavior Research Methods*, *56*(3), 1485-1505.

[6] Xia, Y., Xu, Q., Fang, J., Tang, R., & Du, P. (2024). Bipartite graph-based community-to-community matching in local energy market considering socially networked prosumers. *Applied Energy*, *353*, 122245.

[7] Baruah, B., Dutta, M. P., Banerjee, S., & Bhattacharyya, D. K. (2024). A novel density based community detection algorithm and its application in detecting potential biomarkers of ESCC. *Journal of Computational Science*, *81*, 102344.

[8] Zheng, J., Wang, S., Li, D., & Zhang, B. (2019). Personalized recommendation based on hierarchical interest overlapping community. *Information Sciences*, *479*, 55-75.

[9] Fortunato, S. (2010). Community detection in graphs. *Physics Reports*, *486*(3-5), 75-174.

[10] Adraoui, M., Retbi, A., Idrissi, M. K., & Bennani, S. (2022). Maximal cliques based method for detecting and evaluating learning communities in social networks. *Future Generation Computer Systems*, *126*, 1-14.

[11] Yilmaz, L. S., & Walhout, A. J. M. (2017). Metabolic network modeling with model organisms. *Current Opinion in Chemical Biology*, *36*, 32-39.

[12] Newman, M. E. J., & Girvan, M. (2004). Finding and evaluating community structure in networks. *Physical Review E*, *69*(2), 026113.

[13] Ahajjam, S., El Haddad, M., & Badir, H. (2018). A new scalable leader-community detection approach for community detection in social networks. *Social Networks*, *54*, 41-49.

[14] Azaouzi, M., Rhouma, D., & Ben Romdhane, L. (2019). Community detection in large-scale social networks: state-of-the-art and future directions. *Social Network Analysis and Mining*, *9*(1), 23.

[15] Rostami, M., Oussalah, M., Berahmand, K., & Farrahi, V. (2023). Community detection algorithms in healthcare applications: a systematic review. *IEEE Access*, *11*, 30247-30272.

[16] Rostami, M., Farrahi, V., Ahmadian, S., Jalali, S. M. J., & Oussalah, M. (2023). A novel healthy and time-aware food recommender system using attributed community detection. *Expert Systems with Applications*, *221*, 119719.

[17] Rostami, M., Muhammad, U., Forouzandeh, S., Berahmand, K., Farrahi, V., & Oussalah, M. (2022). An effective explainable food recommendation using deep image clustering and community detection. *Intelligent Systems with Applications*, *16*, 200157.

[18] Alotaibi, N., & Rhouma, D. (2022). A review on community structures detection in time evolving social networks. *Journal of King Saud University-Computer and Information Sciences*, *34*(8), 5646-5662.

[19] Xu, M. (2021). Understanding graph embedding methods and their applications. *SIAM Review*, *63*(4), 825-853.

[20] Kumar, P., Jain, R., Chaudhary, S., & Kumar, S. (2021). Solving community detection in social networks: A comprehensive study. In *2021 5th International Conference on Computing Methodologies and Communication (ICCMC)* (pp. 239-345).

[21] Bhattacharya, R., Nagwani, N. K., & Tripathi, S. (2023). A community detection model using node embedding approach and graph convolutional network with clustering technique. *Decision Analytics Journal*, *9*, 100362.

[22] Traag, V. A., & Šubelj, L. (2023). Large network community detection by fast label propagation. *Scientific Reports*, *13*(1), 2701.

[23] Rabiei, M., Fartash, M., & Nazari, S. (2025). A hybrid optimization approach for graph embedding: leveraging Node2Vec and grey wolf optimization. *The Journal of Supercomputing*, *81*(4), 1-32.

[24] Wu, X., Teng, D., Zhang, H., Hu, J., Quan, Y., Miao, Q., & Sun, P. G. (2025). Graph reconstruction and attraction method for community detection. *Applied Intelligence*, *55*(5), 1-17.

[25] Ullah, A., & Meng, Y. (2025). Finding influential nodes via graph embedding and hybrid centrality in complex networks. *Chaos, Solitons & Fractals*, *194*, 116151.

[26] Abed, I. H., & Bahadori, S. (2025). Using density criterion and increasing modularity to detect communities in complex networks. *InfoTech Spectrum: Iraqi Journal of Data Science*, *2*(1), 1-15.

[27] Chen, G., & Zhou, S. (2024). A novel overlapping community detection strategy based on Core-Bridge seeds. *International Journal of Machine Learning and Cybernetics*, *15*(6), 2131-2147.

[28] Zhang, W., Shang, R., & Jiao, L. (2023). Large-scale community detection based on core node and layer-by-layer label propagation. *Information Sciences*, *632*, 1-18.

[29] Mohamed, E., Agouti, T., Tikniouine, A., & El Adnani, M. (2019). A comprehensive literature review on community detection: Approaches and applications. *Procedia Computer Science*, *151*, 295-302.

[30] Mothe, J., Mkhitaryan, K., & Haroutunian, M. (2017). Community detection: Comparison of state of the art algorithms. In *2017 Computer Science and Information Technologies (CSIT)* (pp. 125-129).

[31] Newman, M. E. J. (2004). Fast algorithm for detecting community structure in networks. *Physical Review E*, *69*(6), 066133.

[32] Chintalapudi, S. R., & Prasad, M. H. M. K. (2016). Finding research groups using modularity based community detection algorithm. In *2016 3rd International Conference on Computing for Sustainable Global Development (INDIACom)* (pp. 65-69).

[33] Behera, R. K., Rath, S. K., & Jena, M. (2016). Spanning tree based community detection using min-max modularity. *Procedia Computer Science*, *93*, 1070-1076.

[34] Krishnan, S. G., Karthika, S., & Bose, S. (2016). Detection of communities in dynamic social networks. In *2016 International Conference on Recent Trends in Information Technology (ICRTIT)* (pp. 1-6).

[35] Jiang, T., Zhang, C., Zhu, H., & Deng, G. (2018). Energy-efficient scheduling for a job shop using grey wolf optimization algorithm with double-searching mode. *Mathematical Problems in Engineering*, *2018*(1), 8574892.

[36] Zhu, J., Liu, Y., Wu, H., Chen, Z., Zhang, Y., Yang, S., Yang, C., Yang, W., & Wu, X. (2020). A no self-edge stochastic block model and a heuristic algorithm for balanced anti-community detection in networks. *Information Sciences*, *518*, 95-112.

[37] Huang, M., Jiang, Q., Qu, Q., & Rasool, A. (2021). An overlapping community detection approach in ego-splitting networks using symmetric nonnegative matrix factorization. *Symmetry*, *13*(5), 869.

[38] Gutiérrez, I., Gómez, D., Castro, J., & Espínola, R. (2022). A new community detection problem based on bipolar fuzzy measures. In *Computational Intelligence and Mathematics for Tackling Complex Problems 2* (pp. 91-99). Springer.

[39] Tabarzad, M. A., & Hamzeh, A. (2017). A heuristic local community detection method (HLCD). *Applied Intelligence*, *46*, 62-78.

[40] Yin, Y., Zhao, Y., Li, H., & Dong, X. (2021). Multi-objective evolutionary clustering for large-scale dynamic community detection. *Information Sciences*, *549*, 269-287.

[41] Zhang, X., Wang, C., Su, Y., Pan, L., & Zhang, H. F. (2017). A fast overlapping community detection algorithm based on weak cliques for large-scale networks. *IEEE Transactions on Computational Social Systems*, *4*(4), 218-230.

[42] Zhou, X., Yang, K., Xie, Y., Yang, C., & Huang, T. (2019). A novel modularity-based discrete state transition algorithm for community detection in networks. *Neurocomputing*, *334*, 89-99.

[43] Blondel, V. D., Guillaume, J. L., Lambiotte, R., & Lefebvre, E. (2008). Fast unfolding of communities in large networks. *Journal of Statistical Mechanics: Theory and Experiment*, *2008*(10), P10008.

[44] Traag, V. A., Waltman, L., & Van Eck, N. J. (2019). From Louvain to Leiden: guaranteeing well-connected communities. *Scientific Reports*, *9*(1), 1-12.

[45] Girvan, M., & Newman, M. E. J. (2002). Community structure in social and biological networks. *Proceedings of the National Academy of Sciences*, *99*(12), 7821-7826.

[46] Alhijawi, B., & Awajan, A. (2024). Genetic algorithms: Theory, genetic operators, solutions, and applications. *Evolutionary Intelligence*, *17*(3), 1245-1256.

[47] Ghoshal, A. K., Das, N., Bhattacharjee, S., & Chakraborty, G. (2019). A fast parallel genetic algorithm based approach for community detection in large networks. In *2019 11th International Conference on Communication Systems & Networks (COMSNETS)* (pp. 95-101).

[48] Zeng, X., Wang, W., Chen, C., & Yen, G. G. (2019). A consensus community-based particle swarm optimization for dynamic community detection. *IEEE Transactions on Cybernetics*, *50*(6), 2502-2513.

[49] Rustamaji, H. C., Suharini, Y. S., Permana, A. A., Kusuma, W. A., Nurdiati, S., Batubara, I., & Djatna, T. (2022). A network analysis to identify lung cancer comorbid diseases. *Applied Network Science*, *7*(1), 30.

[50] Fortunato, S., & Barthélemy, M. (2007). Resolution limit in community detection. *Proceedings of the National Academy of Sciences*, *104*(1), 36-41.

[51] Guo, J., Singh, P., & Bassler, K. E. (2023). Resolution limit revisited: community detection using generalized modularity density. *Journal of Physics: Complexity*, *4*(2), 025001.

[52] Miyauchi, A., & Kawase, Y. (2016). Z-score-based modularity for community detection in networks. *PLoS One*, *11*(1), e0147805.

[53] Fortunato, S., & Hric, D. (2016). Community detection in networks: A user guide. *Physics Reports*, *659*, 1-44.

[54] He, K., Li, Y., Soundarajan, S., & Hopcroft, J. E. (2018). Hidden community detection in social networks. *Information Sciences*, *425*, 92-106.

[55] Chattopadhyay, S., & Ganguly, D. (2021). Node2vec with weak supervision on community structures. *Pattern Recognition Letters*, *150*, 147-154.

[56] Hafez, A. I., Hassanien, A. E., Fahmy, A. A., & Tolba, M. F. (2013). Community detection in social networks by using Bayesian network and expectation maximization technique. In *13th International Conference on Hybrid Intelligent Systems (HIS 2013)* (pp. 209-214).

[57] Srinivas, S., & Rajendran, C. (2019). Community detection and influential node identification in complex networks using mathematical programming. *Expert Systems with Applications*, *135*, 296-312.

[58] Cheng, J., Su, X., Yang, H., Li, L., Zhang, J., Zhao, S., & Chen, X. (2019). Neighbor similarity based agglomerative method for community detection in networks. *Complexity*, *2019*(1), 8292485.

[59] You, X., Ma, Y., & Liu, Z. (2020). A three-stage algorithm on community detection in social networks. *Knowledge-Based Systems*, *187*, 104822.

[60] Kasoro, N., Kasereka, S., Mayogha, E., Vinh, H. T., & Kinganga, J. (2019). PercoMCV: A hybrid approach of community detection in social networks. *Procedia Computer Science*, *151*, 45-52.

[61] Tahir, N., Hassan, A., Asif, M., & Ahmad, S. (2019). MCD: mutually connected community detection using clustering coefficient approach in social networks. In *2019 2nd International Conference on Communication, Computing and Digital Systems (C-CODE)* (pp. 160-165).

[62] Bai, L., Liang, J., Du, H., & Guo, Y. (2018). A novel community detection algorithm based on simplification of complex networks. *Knowledge-Based Systems*, *143*, 58-64.

[63] Freeman, L. C. (1977). A set of measures of centrality based on betweenness. *Sociometry*, 35-41.

[64] Freeman, L. C. (2002). Centrality in social networks: Conceptual clarification. *Social Network: Critical Concepts in Sociology*, *1*, 238-263.

[65] Borgatti, S. P., & Everett, M. G. (2006). A graph-theoretic perspective on centrality. *Social Networks*, *28*(4), 466-484.

[66] Bavelas, A. (1950). Communication patterns in task-oriented groups. *The Journal of the Acoustical Society of America*, *22*(6), 725-730.

[67] Spizzirri, L. (2011). Justification and application of eigenvector centrality. *Algebra in Geography: Eigenvectors of Network*.

[68] Ding, D., & He, X. (2010). Notice of retraction: Application of eigenvector centrality in metabolic networks. In *2010 2nd International Conference on Computer Engineering and Technology* (Vol. 1, pp. V1-89).

[69] Shiokawa, H., Fujiwara, Y., & Onizuka, M. (2015). Scan++ efficient algorithm for finding clusters, hubs and outliers on large-scale graphs. *Proceedings of the VLDB Endowment*, *8*(11), 1178-1189.

[70] Gmati, H., Mouakher, A., Gonzalez-Pardo, A., & Camacho, D. (2024). A new algorithm for communities detection in social networks with node attributes. *Journal of Ambient Intelligence and Humanized Computing*, 1-13.

[71] Guerreiro, L., Silva, F. N., & Amancio, D. R. (2024). Identifying the perceived local properties of networks reconstructed from biased random walks. *PLoS One*, *19*(1), e0296088.

[72] Hamedani, M. R., & Kim, S. W. (2017). JacSim: An accurate and efficient link-based similarity measure in graphs. *Information Sciences*, *414*, 203-224.

[73] Newman, M. E. J. (2006). Modularity and community structure in networks. *Proceedings of the National Academy of Sciences*, *103*(23), 8577-8582.

[74] Csárdi, G., & Nepusz, T. (2006). The igraph software. *Complex Systems*, *1695*, 1-9.

[75] Hagberg, A., Swart, P. J., & Schult, D. A. (2008). *Exploring network structure, dynamics, and function using NetworkX* (Technical Report). Los Alamos National Laboratory.

[76] Avrachenkov, K. E., Kondratev, A. Y., & Mazalov, V. V. (2017). Cooperative game theory approaches for network partitioning. In *Computing and Combinatorics: 23rd International Conference, COCOON 2017, Hong Kong, China, August 3-5, 2017, Proceedings 23* (pp. 591-602). Springer.

[77] Lusseau, D., Schneider, K., Boisseau, O. J., Haase, P., Slooten, E., & Dawson, S. M. (2003). The bottlenose dolphin community of doubtful sound features a large proportion of long-lasting associations: can geographic isolation explain this unique trait? *Behavioral Ecology and Sociobiology*, *54*, 396-405.

[78] Yang, J., & Leskovec, J. (2012). Defining and evaluating network communities based on ground-truth. In *Proceedings of the ACM SIGKDD Workshop on Mining Data Semantics* (pp. 1-8).

[79] Leskovec, J. (2014). SNAP Datasets: Stanford large network dataset collection. Retrieved from http://snap.stanford.edu/data

[80] Eaton, E., & Mansbach, R. (2012). A spin-glass model for semi-supervised community detection. In *Proceedings of the AAAI Conference on Artificial Intelligence* (Vol. 26, No. 1, pp. 900-906).

[81] Smith, N. R., Zivich, P. N., Frerichs, L. M., Moody, J., & Aiello, A. E. (2020). A guide for choosing community detection algorithms in social network studies: The question alignment approach. *American Journal of Preventive Medicine*, *59*(4), 597-605.

[82] Zeng, J., & Yu, H. (2018). A distributed infomap algorithm for scalable and high-quality community detection. In *Proceedings of the 47th International Conference on Parallel Processing* (pp. 1-11).

[83] Palacio-Niño, J. O., & Berzal, F. (2024). Enhancing community detection in networks: A comparative analysis of local metrics and hierarchical algorithms. *arXiv preprint arXiv:2408.09072*.

[84] Yuan, L., Zhang, X., Ke, Y., Lu, Z., Li, X., & Liu, C. (2025). Temporal community detection and analysis with network embeddings. *Mathematics*, *13*(5), 698.

[85] Farrokhpour Dizaj, Y., Lotfi, S., & Hajipour, J. (2025). A multi-objective social-based algorithm for community detection. *Multimedia Tools and Applications*, 1-30.

[86] Werner, M. A., de Ron, J., Fried, E. I., & Robinaugh, D. J. (2025). Iterated community detection in psychological networks. *Psychological Methods*.

[87] Cao, J., Xu, W., Jin, D., Zhang, X., Liu, L., Miller, A., Shi, Z., & Ding, W. (2025). A unified framework of semi-supervised community detection integrating network topology and node content. *Information Sciences*, *686*, 121349.