IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received 14 November 2023, accepted 25 November 2023, date of publication 28 November 2023,
date of current version 4 December 2023.

Digital Object Identifier 10.1109/ACCESS.2023.3337375

RESEARCH ARTICLE

Community-Consideration Centrality, a Case Study of Lung Cancer Proteins

HERU CAHYA RUSTAMAJI[1,5], WISNU ANANTA KUSUMA[1,4], SRI NURDIATI[2], AND IRMANIDA BATUBARA[3,4]

[1]Department of Computer Science, Faculty of Mathematics and Natural Sciences, IPB University, Bogor 16680, Indonesia
[2]Department of Mathematics, Faculty of Mathematics and Natural Sciences, IPB University, Bogor 16680, Indonesia
[3]Department of Chemistry, Faculty of Mathematics and Natural Sciences, IPB University, Bogor 16680, Indonesia
[4]Tropical Biopharmaca Research Center, IPB University, Bogor 16680, Indonesia
[5]Department of Informatics, Faculty of Industrial Technology, UPN "Veteran" Yogyakarta, Yogyakarta 55283, Indonesia

Corresponding author: Wisnu Ananta Kusuma (ananta@apps.ipb.ac.id)

This work was supported by the Tropical Biopharmaca Research Center (TropBRC) and the Directorate of Scientific Publications and Strategic Information of IPB University.

ABSTRACT One of the most recent studies on the analysis of complex systems is to understand the role of community structure and centrality in analyzing the networks of complex systems such as protein and social networks. Traditional measures of centrality - degree centrality, closeness centrality, and betweenness centrality - cannot capture how community structures within these networks configure them. In this regard, we propose a new community-consideration centrality method to fill this gap. This method includes a weight of consideration, α, ranging from 0.0 to 1.0, to balance the focus between community and network-wide importance in the centrality calculations. Our analysis of two zachary karate and dolphin datasets shows that including community consideration in the degree, closeness, and betweenness centrality measures accurately captures the proportional significance of both communities and networks. In particular, for the lung adenocarcinoma cancer protein case study, our method not only identified more cancer hallmark genes than the traditional centrality measures without considering communities but also outperformed several other advanced centrality algorithms regarding the detection of crucial cancer-related genes. A balanced objective between network and community impacts was observed at an optimum performance α values of 0.1 and 0.2. It finds a strong significance of community structure in network analysis and features a more nuanced perspective on centrality in complex systems.

INDEX TERMS Degree centrality, closeness centrality, betweenness centrality, the hallmark of cancer.

I. INTRODUCTION

Centrality measures relevant to network analysis identify nodes that crucially influence information, communication, or interaction flows within a network [1]. Common centrality types – degree, closeness, and betweenness – offer insights into different aspects of network influence [2], [3]. For example, degree centrality represents the direct impact of a node on its first neighbors and has played a vital role in the discovery of essential biomarker proteins in diseases such as COVID-19 and type 2 diabetes mellitus [4], [5]. In disease multimorbidity networks, degree centrality is used to identify diseases that potentially have chronic influence [6]. Degree centrality in drug networks can also be used as a feature in machine learning to predict drug side effects [7]. Closeness centrality is a significant method in which essential proteins can be identified in biological networks owing to the swiftness of a node to reach others [8]. Closeness centrality was a significant predictor of susceptibility to the spread of Covid-19 [9]. Betweenness centrality, which indicates an intermediate node, is imperative for the identification of key tumor genes [10], [11].

However, the existence of communities in centrality analysis needs to be considered in network research.

The associate editor coordinating the review of this manuscript and approving it for publication was Vincenzo Conti.

Community detection is essential for understanding complex, modular network structures as they identify tightly-knit node groups within large networks [12]. In communities, groups of nodes are more strongly related, share similar traits, or play similar roles in a network [13]. This detection is particularly evident in biological networks and can lead to the discovery of protein clusters with shared functions that are fundamentally essential for deciphering intricate biological mechanisms [14], [15]. Community detection is critical for gene clustering and protein complex identification [16]. Community detection has been conducted using various algorithms ranging from traditional methods, such as disassembly greedy modularity [17], greedy modularity [18], Louvain [19], and Girvan and Newman [20], to those using deep learning techniques [21]. Healthcare application reviews in the recent past have highlighted the burgeoning use of community detection algorithms for extensive healthcare database analysis and the role that technology plays in knowledge discovery within health informatics [22].

Finally, our research innovates from existing methodologies by weaving community considerations into the traditional centrality measures. It is bound forward, and a dual-focused look at individual node impact and community roles enriches the current network structure understanding. From this viewpoint, the basic idea of our approach regarding lung adenocarcinoma is that it proactively searches for hallmark cancer proteins, focusing on capturing all the essential proteins affecting the disease trajectory. Therefore, from these viewpoints, our approach appears novel and more complete in terms of network structures by combining node-specific analysis and community structure. This approach offers profound and nuanced insights, particularly in complex biological systems, setting a new precedent for network analysis.

II. RELATED WORK

Over the years, various developments and fine-tuning of centrality measures have witnessed a continuous and dynamic evolution of network analysis. This development parallels diverse viewpoints and aims to assess the prominence of nodes within various network structures.

Eigenvector centrality [23] values the power of a node's connections, which counts the number and effectualness of the nodes connected to it. In other words, effective importance passes through a node owing to neighbors with influential friends. Katz Centrality [24] extended this idea to introduce a factor that attenuates the influence of distant nodes, thus tempering the instant effect instead of potentially extending neighborhood influences.

Information Centrality [25] and Load Centrality [26], [27] provide more intricate revelations on network structures. Information Centrality is a measure of the efficiency of a node concerning information the flow in the entire network, considering all direct and indirect interventions between nodes. Koad centrality focuses on the traffic volume in the network which is considered a node and accorded value to nodes sensitive to network link weights, such as the maintenance of network connectivity and the ability to host traffic.

Subgraph Centrality [28] and Communicability Betweenness Centrality [29] have introduced more nuanced approaches. Subgraph centrality totals the weights of all long and small closed walks that start from and end at a particular node, thus weighing a node's involvement in local motifs and a broader network structure. However, Communicability Betweenness Centrality generalizes the classical betweenness concept by integrating the communicability effect, which sums all paths binding any two nodes instead of being confined to the shortest routes. This provides a more comprehensive measure of how a node best facilitates the connectivity across a network.

Second-order centrality [30] and Harmonic centrality [31] are alternative centrality measures that offer new ways to influence measurement and accessibility. Second-order centrality provides a measure of the importance with which nodes return to a point in random walks, thus providing insights into node criticality and network structure. However, harmonic centralities are based on the harmonic mean of the distances from a node to all other nodes and effectively capture the reachability of a node within the network.

Viral Rank Centrality [32] and Distinctive Centrality [33] contribute to new perspectives on the centrality discourse. Viral Rank Centrality measures the potential of a node to initiate contagion through the spread of influence following the mechanisms of viral propagation. Distinctive Centrality emphasizes the uniqueness of a node's connections, focusing on the role of exclusive or rare links in determining a node's importance. These different centrality measures each have their own unique focus and methodology and are collected together to aggregate our understanding of the roles and importance of the nodes in complex networks.

Overall, they provide a more multi-dimensional view of the structure of entities encompassing direct influence and connectivity, as well as the subtleties of information flow and network resilience. While centrality metrics including Eigenvector, Katz, Load, and other centrality-based metrics provide deep insight into the importance of individual nodes in networks based on connections and information flow, they tend to waive away the large community structure within which these nodes operate, especially on a network that has a defined community. Consequently, particularly with networks characterized by well-outlined community structures, indicated by high modularity, these metrics could inadequately capture how contributions or interactions of nodes are performed within their community contexts, elements that could be crucial in establishing the dynamics and strengths of networks, whether social, biological, or technical.

III. MATERIAL AND METHODS

A. DATASET

To conduct our experiments, we began by selecting several datasets covering the required dimensions. The datasets used

were real-world networks referred to in various community detection studies, including the Zachary Karate Club and Dolphin Social Networks. The Zachary Karate dataset, based on a record of the pattern of interactions among 34 members of a karate club in the mid-1970s at a university in the U.S. [34], and the Dolphin dataset, which describes a community of 62 bottlenose dolphins in Doubtful Sound, New Zealand over seven years [35], [36], are publicly available and hosted open access at http://www-personal.umich.edu/ mejn/netdata/). These datasets are popular sources for network analysis and offer rich context for understanding the formation of social ties and communities. The genetic mutations input data of adenocarcinoma lung cancer were culled/extracted from four primary databases: TCGA Genomic Data Commons Data Portal [37] (https://portal.gdc.cancer.gov/), IntOGen cancer mutation database [38] (https://www.intogen.org), cBioPortal for cancer genomics [39] (https://www.cbioportal.org/), and Catalogue of Somatic Mutations in Cancer (COSMIC) [40] (https://cancer.sanger.ac.uk/cosmic). These databases provide a rich source of genetic data, which is crucial for our analysis in the context of lung cancer. We integrated data from four cancer databases. Duplicate records were carefully removed to validate the integrity of the datasets. Subsequently, Protein-Protein Interaction (PPI) networks were built using information recorded in the STRING database (https://string-db.org/). All datasets were publicly accessible.

B. METHOD

In our study, the ground truth embodied in the selected datasets played an essential role. The underlying community configurations of both the Zachary Karate Club and Dolphin social networks were recognized and incorporated. Incorporation was critical in addition to applying new measures of degree, closeness and betweenness centrality.

We performed community detection within the PPI network using the disassembly greedy modularity approach [17]. This method allows the identification of groups of proteins with high structural densities. Metascape (https://metascape.org) was used for enrichment analysis of each community to detect the possibility of carrying out the same function or participating in the same biological pathway. This allows the measurement of functional and structural relationships among protein communities.

The next stage involves research on community consideration centrality. This new approach features the adjustment of traditional centrality measures such as degree, closeness and betweenness, while considering the community structure inside the network. These adjustments were hypothesized to offer a more nuanced understanding of the involvement and significance of each protein in lung cancer.

With regard to tools and software, the experiments carried out in our case study mostly used network analysis based tools. Centrality calculations and community detection algorithms were implemented using Python with libraries, such as NetworkX, to implement graph-based network analysis.

We compared our analysis's proteins, which were characterized as central, and those involved in the hallmark of cancer [41], according to the COSMIC database. The jackknife curve was used to confirm the validity of the findings. This analysis involved a comparative study of other centrality measures within a broader scope of network analysis research.

C. COMMUNITY

In the network $G(V, E)$, where V is the set of vertices and E is the edge (connection) between vertices, partition $P = C_1, C_2, \ldots, C_k$ divides the vertices into k. Each community C_i refers to a group of vertices with a higher density in the network. The extent to which a network is divided into communities is measured using modularity, as shown in (1). The higher the modularity, the better is the partitioning Barabasi [12]. The toy network community is shown in Fig. 1, with two communities and a modularity of 0.3194.

$$M = \sum_{C=1}^{n_C} \left[\frac{L_C}{L} - \frac{k_C}{2L} \right] \tag{1}$$

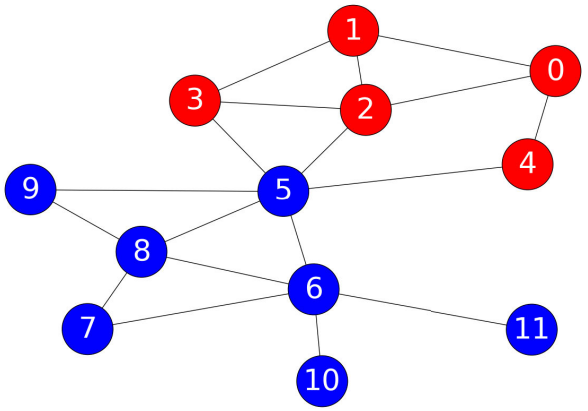

FIGURE 1. Communities in the Toy network consist of two communities, red and blue. Based on the overall network, node five has the highest degree, 6. However, looking at the red community, node five only has a degree of 3, lower than node six, with a degree of 5.

D. COMMUNITY-CONSIDERATION DEGREE CENTRALITY

We developed a community consideration degree centrality algorithm as shown in Algorithm 1, which algorithm uses a degree centrality calculation applied to the overall network and the community. The consideration proportion is expressed as $\alpha = [0, 1], \alpha \in R$. The smaller the value of α, the more the community is considered; the larger α, the more the network is considered. Each vertex is iterated to determine the community consideration degree centrality. The first forms a subgraph G from each community. Then, for each vertex contained in subgraph G, the community-consideration degree centrality is calculated according to and α and the degree value at vertex i. The time complexity for calculating the degree of centrality of each

node with a list adjacency is $O(n+m)$, where n is the number of nodes in the network and m is the number of edges. For all the nodes, it takes $O(n(n+m))$.

Algorithm 1 Algorithm for Calculating Community-Consideration Degree Centrality

Input: graph G, community, α
Output: Community-consideration closeness centrality
  for each community **do**
    Create subgraph C from the community
    for each $i \in C$ **do**
      $CCDC_\alpha(i) \leftarrow (d(i \in C) + \alpha(d(i \in G) - d(i \in C))$
      $/(n_C + \alpha(n_G - n_C) - 1))$
    end for
  end for

The degree centrality Freeman [2] of node i is the node degree $d(i)$ normalized by dividing it by the maximum node degree $(n-1)$, as expressed in (2).

$$DC(i) = \frac{d(i)}{n-1} = \frac{|(i,j) \in E|}{n-1} \tag{2}$$

If the network consists of several disjoint communities, take one community C, then $C \cap C' = \emptyset$ and $C \cup C' = V$. Equation (2) can be expressed as (3)

$$\begin{aligned} DC(i) &= \frac{|(i,j) \in E, j \in C| + |(i,j) \in E|, j \notin C}{n_C + n_{C'} - 1} \\ &= \frac{d_C(i) + d_{C'}(i)}{n_C + n_{C'} - 1} \end{aligned} \tag{3}$$

If we do not involve terms containing C', and isolate C alone, Equation (4) is formed.

$$DC(i) = \frac{d_C(i)}{n_C - 1} \tag{4}$$

We want to adjust the balance proportion of community and network involvement in the degree centrality. From (3) and (4), we can form a variable range $\alpha = [0, 1], \alpha \in \mathbb{R}$ and formulate the community consideration degree centrality as (5).

$$CCDC_\alpha(i) = \frac{d_C(i) + \alpha d_{C'}(i)}{n_C + \alpha n_{C'} - 1} \tag{5}$$

If $\alpha = 1$, then the formula will produce a degree of centrality at the network level without considering the formed community. Conversely, if $\alpha = 0$, only the degree of the community is considered. The value $0 < \alpha < 1$ represents the degree of centrality by proportionally considering the community and the network.

Because for each subgraph formed from community C, $V_{C'} = V_G - V_C$ is applied for each subgraph formed from community C, Equation (5) can be expressed as (6).

$$CCDC_\alpha(i) = \frac{d_C(i) + \alpha(d_G(i) - d_C(i))}{n_C + \alpha(n_G - n_C) - 1} \tag{6}$$

E. COMMUNITY-CONSIDERATION CLOSENESS CENTRALITY

We developed a community consideration closeness centrality algorithm. This algorithm uses closeness centrality calculations applied to the entire network and community. The proportion of consideration is expressed as $\alpha = [0, 1], \alpha \in \mathbb{R}$, which considers the community or network as a whole.

The steps to obtain the community consideration closeness centrality are expressed in Algorithm 2. First, the distance from node i to each node in the network was calculated. Next, the distance from each node i to each node in the community is calculated, and then the community-consideration closeness centrality is calculated. The time-complexity closeness centrality calculation for each node is $O(n(n+m))$. For all nodes, $O(n^2(n+m))$.

Algorithm 2 Algorithm for Calculating Community-Consideration Closeness Centrality

Input: graph G, community, α
Output: community-consideration closeness centrality
  for each $i \in G$ **do**
    $d(i) \leftarrow 0$
    for each $j \in G$ **do**
      $d(i) \leftarrow d(i) + d(i,j)$
    end for
  end for
  for each community **do**
    Create subgraph C from the community
    for each $i \in C$ **do**
      $d_C(i) \leftarrow 0$
      for each $j \in C$ **do**
        $d_C(i) \leftarrow d_C(i) + d(i,j)$
      end for
    end for
  end for
  for each $i \in G$ **do**
    $CCCC_\alpha(i) \leftarrow (n_C(i) + \alpha(n_G - n_C(i)) - 1)/$
    $(d_C(i) + \alpha(d(i) - d_C(i)))$
  end for

Closeness centrality Freeman [2] measures the shortest path from node i to all nodes and is expressed by dividing the number of destination nodes by the total topological distance of all nodes in the network, as expressed by (7).

$$CC(i) = \frac{n-1}{\sum_{j=1}^{n} d(i,j)} \tag{7}$$

If the network consists of several disjoint communities, take one community C, then $C \cap C' = \emptyset$ and $C \cup C' = V$. Equation (7) can be expressed as (8)

$$CC(i) = \frac{n_C + nC' - 1}{\sum_{j=1}^{n_C} d(i,j) + \sum_{j=1}^{n_{C'}} d(i,j)} \tag{8}$$

If we do not involve the term containing C' and isolate only C, we obtain (9).

$$CC(i) = \frac{n_C - 1}{\sum_{j=1}^{n_C} d(i,j)} \tag{9}$$

That is, if we want to set the proportion of the balance of community and network involvement in closeness centrality, we can form a variable range $\alpha = [0, 1]$, $\alpha \in \mathbb{R}$ and formulate the community-consideration closeness centrality as (10)

$$CCCC(i) = \frac{n_C + \alpha n C' - 1}{\sum_{j=1}^{n_C} d(i,j) + \alpha \sum_{j=1}^{n_{C'}} d(i,j)} \tag{10}$$

Because for each subgraph formed from community C, $V_{C'} = V_G - V_C$ applied, (10) can be expressed as (11)

$$CCCC(i) = \frac{n_C + \alpha(n_G - n_C) - 1}{\sum_{j=1}^{n_C} d(i,j) + \alpha(\sum_{j=1}^{n_G} d(i,j) - \sum_{j=1}^{n_C} d(i,j))} \tag{11}$$

F. COMMUNITY-CONSIDERATION BETWEENNESS CENTRALITY

The steps to obtain the community consideration betweenness centrality are presented in Algorithm 3. Iterations were performed for each community in graph G, form subgraph G, and the betweenness centrality in each subgraph was calculated. Next, we calculate the number of network and community path combinations. Furthermore, for each node in subgraph G, a community-consideration betweenness centrality calculation is performed according to the degree value at node i, the number of paths, and α. The time complexity betweenness centrality for each node was $O(n(n+m))$. For all nodes, $O(n^2(n+m))$.

Algorithm 3 Algorithm for Calculating Community-Consideration Betweenness Centrality

Input: graph, community, α
Output: community-consideration betweenness centrality
  for each community **do**
    Create subgraph C from the community
    $BCC(i) =$ betweenness centrality on C
    for each $i \in C$ **do**
      $PC(i) \leftarrow (n_C - 1)(n_C - 2)/2$
    end for
    $PG(i) \leftarrow (n-1)(n-2)/2$
    $BCG \leftarrow$ betweenness centrality on G
    $CCBC_\alpha \leftarrow (BCC + \alpha(BGC - BCG)/(PC + \alpha(PG - PC))$
  end for

The betweenness centrality Freeman [2] of node i is expressed as (12)

$$BC(i) = \sum_{j<k} \frac{p_{jk}(i)}{p_{jk}} \tag{12}$$

If the network consists of several disjoint communities, take one community C, then $C \cap C' = \emptyset$ and $C \cup C' = V$. With

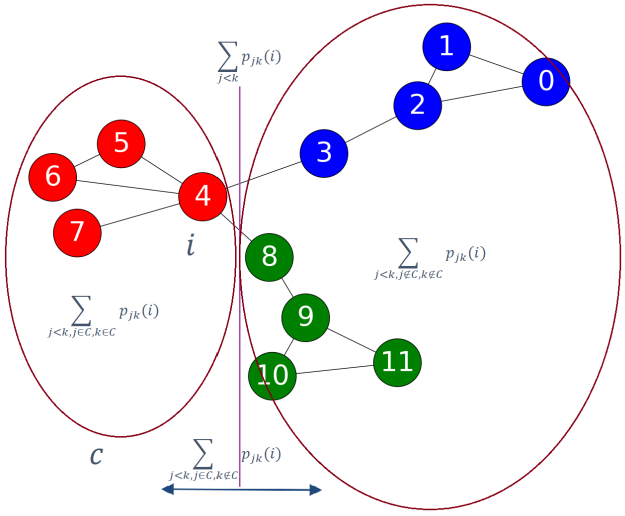

FIGURE 2. Illustration of betweenness centrality at node i, which is in the blue community. The left side of the vertical line is a node in the same community as *i*, and the right side is a node in a different community from *i*.

$i \in C$, here are three possibilities $p_{jk}(i)$ and p_{jk} according to the existence of j and k, illustrated in Fig. 2:

1) $j \in C, k \in C$; both nodes j and k are in the same community as i
2) $j \in C, k \notin C$; One node is in a community with i, and one node is not in a community with i
3) $j \notin C, k \notin C$; both vertices are not in the same community with i

Note that this graph is undirected; thus $\forall j, k \in V, ((j,k) \in E = (k,j) \in E)$, and is counted only once. From Equation (12), it can also be stated that (13)

$$BC(i) = \frac{\sum_{j<k, j\in C, k\in C} p_{jk}(i)}{\sum_{j<k, j\in C, k\in C} p_{jk}} + \frac{\sum_{j<k, j\in C, k\notin C} p_{jk}(i) + \sum_{j<k, j\notin C, k\notin C} p_{jk}(i)}{+\sum_{j<k, j\notin C, k\notin C} p_{jk} + \sum_{j<k, j\in C, k\notin C} p_{jk}} \tag{13}$$

If we do not involve terms containing C' and isolate C alone, then the formulation is formed as (14)

$$BC(i) = \frac{\sum_{j<k, j\in C, k\in C} p_{jk}(i)}{\sum_{j<k, j\in C, k\in C} p_{jk}} \tag{14}$$

That is, if we want to set the proportion of the balance of community and network involvement in closeness centrality, we can form a variable range $\alpha = [0, 1]$, $\alpha \in \mathbb{R}$ and formulate the community-consideration closeness centrality as (15).

$$BC(i) = \frac{\sum_{j<k, j\in C, k\in C} p_{jk}(i)}{\sum_{j<k, j\in C, k\in C} p_{jk}} + \frac{\alpha(\sum_{j<k, j\in C, k\notin C} p_{jk}(i) + \sum_{j<k, j\notin C, k\notin C} p_{jk}(i))}{+\alpha(\sum_{j<k, j\notin C, k\notin C} p_{jk} + \sum_{j<k, j\in C, k\notin C} p_{jk})} \tag{15}$$

From the explanation and Fig. 2 it can be stated that (16) and (17)

$$\sum_{j<k,j\in V,k\in V} p_{jk}(i) = \sum_{j<k,j\in C,k\in C} p_{jk}(i) + \sum_{j<k,j\in C,k\notin C} p_{jk}(i) + \sum_{j<k,j\notin C,k\notin C} p_{jk}(i) \tag{16}$$

$$\sum_{j<k,j\in V,k\in V} p_{jk} = \sum_{j<k,j\in C,k\in C} p_{jk} + \sum_{j<k,j\in C,k\notin C} p_{jk} + \sum_{j<k,j\notin C,k\notin C} p_{jk} \tag{17}$$

Then we substitute Equations (16) and (17) into Equation (15) which yields Equation (18)

$$BC(i) = \frac{\sum_{j<k,j\in C,k\in C} p_{jk}(i)}{\sum_{j<k,j\in C,k\in C} p_{jk}} \frac{+\alpha(\sum_{j<k,j\in V,k\in V} p_{jk}(i) - \sum_{j<k,j\in C,k\in C} p_{jk}(i))}{+\alpha(\sum_{j<k,j\in V,k\in V} p_{jk} - \sum_{j<k,j\in C,k\in C} p_{jk})} \tag{18}$$

Because $j, k \neq i$, based on the combination formula, two vertices are chosen from $n-1$ vertices. One of the excluded nodes is node i. The denominator can be described by (19) and (20).

$$\sum_{j<k,j\in C,k\in C} p_{jk} = \frac{(n_C-1)(n_C-2)}{2} \tag{19}$$

$$\sum_{j<k,j\in V,k\in V} p_{jk} = \frac{(n-1)(n-2)}{2} \tag{20}$$

Thus, from (18), (19) and (20) we obtain (21)

$$BC(i) = \frac{\sum_{j<k,j\in C,k\in C} p_{jk}(i)}{\frac{(n_C-1)(n_C-2)}{2}} \frac{+\alpha(\sum_{j<k,j\in V,k\in V} p_{jk}(i) - \sum_{j<k,j\in C,k\in C} p_{jk}(i))}{+\alpha(\frac{(n-1)(n-2)}{2} - \frac{(n_C-1)(n_C-2)}{2})} \tag{21}$$

G. TOY EXAMPLE

As an example, we use the toy example shown in Fig. 1. According to (6), the community-consideration degree centrality can be seen in Fig. 3, where community-consideration degree centrality node five without considering the community ($\alpha = 1$) has a value of 0.545 in the first place, followed by node 6 with a value of 0.455, and nodes 2 and 8 with a value of 0.364. If the community formed is considered with a value of $\alpha = 0.5$, then the first order ranking is node 6 with a value of 0.588, followed by nodes 5, 8, and 2, each worth 0.529, 0.471, and 0.467, respectively. Meanwhile, if we only consider the communities formed, without considering the network as a whole, we isolate each community separately

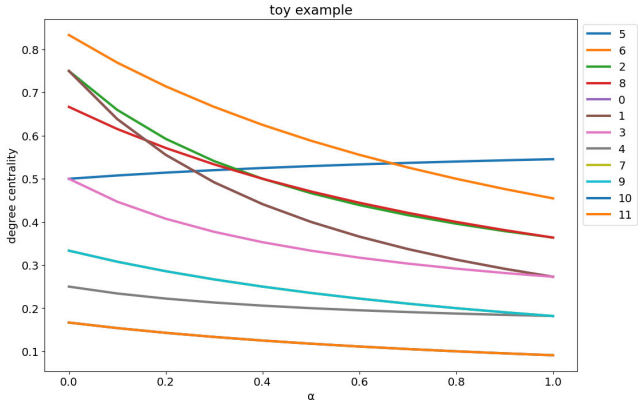

FIGURE 3. Community consideration degree centrality in the toy example. Vertex 5, first in network-level degree centrality, becomes 5th in the community level.

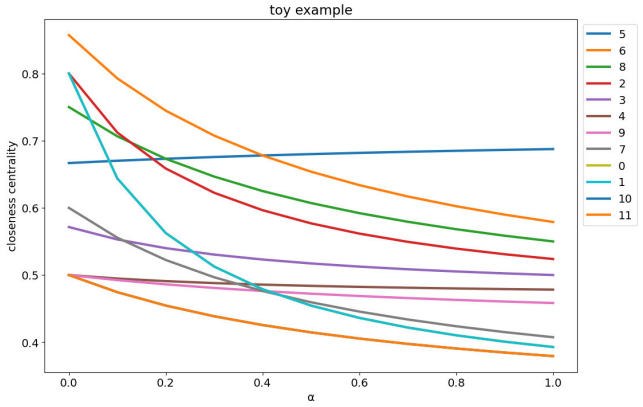

FIGURE 4. Community consideration closeness centrality in a toy example. There are sequence differences with α variations.

($\alpha = 0$). The first sequence was node 6, with a value of 0.833, followed by nodes 2, 0, and 1, which had the same value of 0.750.

The community consideration closeness centrality according to Equation (11) is shown in Fig. 4. The community consideration closeness centrality of node five without considering the community ($\alpha = 1$) was 0.688, followed by nodes 6, 8, and 2 with community-consideration closeness centralities of 0.579, 0.550 and 0.524, respectively. If the formed community is considered to have a value of $\alpha = 0.5$, then the ranking order remains the same: nodes 5, 6, 8, and 2. However, if the value of $\alpha = 0.3$, the order changes, namely node 6 first with a value of 0.708, followed by nodes 5, 8, and 2 with value of 0.676, 0.647, and 0.622, respectively. If the overall network element is not considered and only the community is considered ($\alpha = 0$), then the order is node 6, 2, 0, and 1 with values of 0.857, 0.800, 0.800, and 0.800, respectively.

The community consideration betweenness centrality according to (21) is shown in Fig. 5. For $\alpha = 1$, the order is nodes 5, 6, 2, and 8 with values of 0.600, 0.400, 0.136, and 0.100, respectively. The order did not change when the

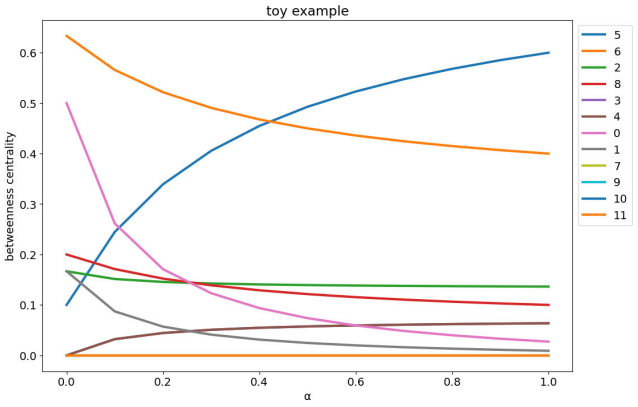

FIGURE 5. Community consideration betweenness centrality in the toy example. Vertices 6 and 5 are first and second for $\alpha > 0.2$, respectively.

value of $\alpha = 0.5$. When the value of $\alpha = 0.4$, the first order is node six, followed by nodes 5, 2, and 8, with values of 0.491, 0.406, 0.143, and 0.139, respectively. For $\alpha = 0$, the order is node 6, 0, 8, 2 with values of 0.633, 0.500, 0.200 and 0.167 respectively.

IV. RESULTS

A. ZACHARY KARATE

The Zachary Karate dataset reflects the structure of the interactions among 34 karate club members, with 78 connections representing relationships such as joint training and friendships [34]. Disputes among members of this club led to the formation of two different factions within the karate club. The first group followed the Instructor, Mr. Hi, represented by node one, while the second group followed club president, John A, represented by node 34 (Fig. 6).

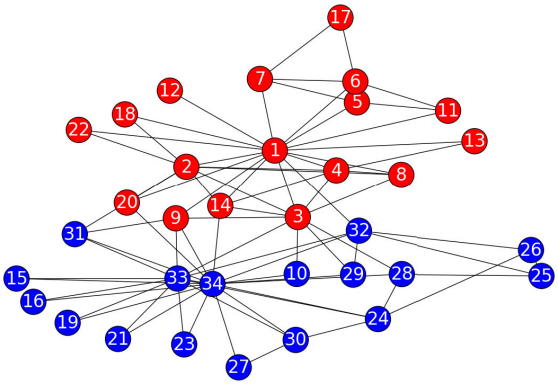

FIGURE 6. The community on the Zachary Karate dataset is divided into two groups: the red side with instructor Mr. Hi, node 1, and the red group that follows the club president, John A., node 34.

We sorted the results of the community-consideration degree, closeness, and betweenness centrality calculations with a range of α from 0 to 1, as presented in Table 1. Almost all the calculation results show that node one and node 34, which represent instructors and club presidents, respectively, occupy the top two ranks. The difference lies in their degree of centrality. Both nodes have the same value when $\alpha = 0.5$; when $\alpha > 0.5$, which means that it is more inclined to the network as a whole, node 34 is ranked first; conversely, when $\alpha < 0.5$, which is more community-oriented, node one is ranked first. Node 33 is in position 3, followed by nodes 2 and 3, which are ranked 4 and 5, respectively, depending on the value of α.

In community consideration closeness centrality, rank one is occupied by node 1 for each α value. Most rank two is occupied by node 34, except for $\alpha = 1.0$ node 34, and the club president is ranked three after node three, which is only a club member. This means that community considerations represented by the value of α can determine the order of centrality formed to provide better results. In ranks 3-5, nodes 33, 3, 32, and 2 are generally occupied, and their sequence is influenced by the value of α. Node 1 represents the instructor, and node 34 represents the club president, consistently ranking 1 and 2 on betweenness community-consideration centrality for all α values. Rank 3 is occupied by node 33, and ranks 4 and 5 are occupied by nodes 32, 3, 2, and 24, whose order depends on the value of α.

B. DOLPHINS

The dolphin social network, often used in community detection research, consists of 62 vertices labeled with names representing dolphins and 159 edges that associate the observed pair of dolphins occurring together Cheng et al. [36], Lusseau et al. [35]. The network can be partitioned into four groups: green, violet, yellow, and blue, each consisting of 23, 20, 12, and 7 dolphins as a ground-truth community structure, as shown in Fig. 7.

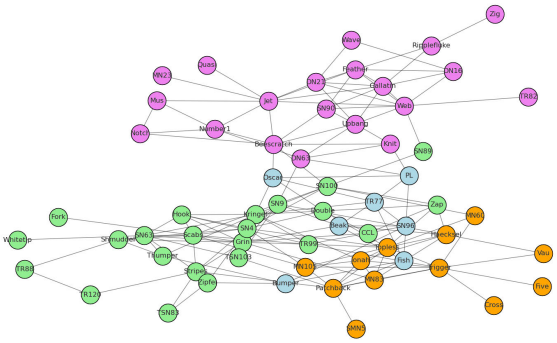

FIGURE 7. Dolphin Social Network Communities: illustrates the 62 dolphins network implementation divided into four communities (green, violet, yellow, blue) based on observed social interactions, showing the complex social structure of this network.

In community-consideration degree centrality, the top-five nodes for $\alpha \geq 0.5$ to 0.5 are dominated by the green and yellow communities, while for $0.1 \leq \alpha \leq 0.4$, it is more varied in the three communities. For $\alpha = 1$, it was dominated by small communities (blue and yellow). The order of community-consideration centrality in the Dolphin dataset is presented in Table 2.

In community-consideration closeness centrality, for $\alpha \leq 0.2$, the top five nodes are dominated by the green

TABLE 1. Community-consideration degree, closeness, and betweenness centrality on karate dataset.

Centrality	order	α										
		0.0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0
Degree	1	1 (0.938)	1 (0.853)	1 (0.784)	1 (0.725)	1 (0.675)	1,34 (0.633)	34 (0.603)	34 (0.577)	34 (0.554)	34 (0.534)	34 (0.515)
	2	34 (0.875)	34 (0.808)	34 (0.753)	34 (0.706)	34 (0.667)	1,34 (0.633)	1 (0.595)	1 (0.563)	1 (0.534)	1 (0.508)	1 (0.485)
	3	33 (0.625)	33 (0.576)	33 (0.536)	33 (0.502)	33 (0.474)	33 (0.449)	33 (0.427)	33 (0.409)	33 (0.392)	33 (0.377)	33 (0.364)
	4	2 (0.5)	2 (0.458)	2 (0.423)	2 (0.393)	2 (0.368)	2 (0.347)	2 (0.328)	3 (0.315)	3 (0.311)	3 (0.307)	3 (0.303)
	5	3 (0.375)	3 (0.362)	3 (0.351)	3 (0.341)	3 (0.333)	3 (0.327)	3 (0.321)	2 (0.312)	2 (0.297)	2 (0.284)	2 (0.273)
Closeness	1	1 (0.941)	1 (0.839)	1 (0.77)	1 (0.72)	1 (0.683)	1 (0.653)	1 (0.63)	1 (0.611)	1 (0.594)	1 (0.581)	1 (0.569)
	2	34 (0.889)	34 (0.797)	34 (0.735)	34 (0.69)	34 (0.655)	34 (0.628)	34 (0.606)	34 (0.589)	34 (0.574)	34 (0.561)	3 (0.559)
	3	33 (0.727)	33 (0.676)	33 (0.638)	33 (0.61)	33 (0.588)	33 (0.57)	3 (0.567)	3 (0.565)	3 (0.563)	3 (0.561)	34 (0.55)
	4	2 (0.64)	2 (0.604)	3 (0.581)	3 (0.577)	3 (0.573)	3 (0.57)	33 (0.555)	32 (0.549)	32 (0.546)	32 (0.543)	32 (0.541)
	5	3 (0.593)	3 (0.586)	2 (0.577)	32 (0.567)	32 (0.562)	32 (0.557)	32 (0.553)	33 (0.543)	33 (0.532)	33 (0.523)	33 (0.516)
Betweenness	1	1 (0.703)	1 (0.616)	1 (0.564)	1 (0.53)	1 (0.505)	1 (0.487)	1 (0.473)	1 (0.461)	1 (0.452)	1 (0.444)	1 (0.438)
	2	34 (0.57)	34 (0.483)	34 (0.431)	34 (0.396)	34 (0.372)	34 (0.353)	34 (0.339)	34 (0.328)	34 (0.318)	34 (0.311)	34 (0.304)
	3	33 (0.188)	33 (0.174)	33 (0.166)	33 (0.16)	33 (0.156)	33 (0.153)	33 (0.151)	33 (0.149)	33 (0.148)	33 (0.146)	33 (0.145)
	4	32 (0.119)	32 (0.125)	32 (0.129)	32 (0.131)	32 (0.133)	32 (0.135)	32 (0.136)	32 (0.137)	32 (0.137)	3 (0.141)	3 (0.144)
	5	24 (0.067)	2 (0.061)	3 (0.084)	3 (0.1)	3 (0.112)	3 (0.121)	3 (0.127)	3 (0.133)	3 (0.137)	32 (0.138)	32 (0.138)

TABLE 2. Community-consideration degree, closeness, and betweenness centrality on Dolphins dataset.

Centrality	order	α										
		0.0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0
Degree	1	SN96	Trigger	Trigger	Trigger	Trigger	Grin	Grin	Grin	Grin	Grin	Grin
	2	TR77	Topless	Topless	Topless	Grin	Trigger	Trigger	Topless	Topless	Topless	Topless
	3	Trigger	SN96	Grin	Grin	Topless	Topless	Topless	SN4	SN4	SN4	SN4
	4	Fish	TR77	Jet	SN4	SN4	SN4	SN4	Trigger	Trigger	Trigger	Trigger
	5	Topless	Grin	SN4	Jet	Jet	Scabs	Scabs	Scabs	Scabs	Scabs	Scabs
Closeness	1	SN96	Grin	SN4	SN4	SN4	SN4	SN100	SN100	SN100	SN100	SN100
	2	TR77	SN4	Grin	Grin	SN100	SN100	SN4	SN4	SN4	SN9	SN9
	3	Trigger	Trigger	Scabs	SN100	Grin	SN9	SN9	SN9	SN9	SN4	SN4
	4	Fish	Scabs	SN100	Scabs	SN9	Grin	Kringel	Kringel	Kringel	Kringel	Kringel
	5	Topless	Topless	SN9	SN9	Kringel	Kringel	Grin	Grin	Grin	Grin	Grin
Betweenness	1	Trigger	Jet	SN100	SN100	SN100	SN100	SN100	SN100	SN100	SN100	SN100
	2	Jet	SN100	Jet	Beescratch	Beescratch	Beescratch	Beescratch	Beescratch	Beescratch	Beescratch	Beescratch
	3	TR77	SN4	Beescratch	Jet	SN4	SN4	SN4	SN4	SN4	SN9	SN9
	4	SN96	Trigger	SN4	SN4	Jet	Jet	SN9	SN9	SN9	SN4	SN4
	5	SN4	Beescratch	Trigger	SN9	SN9	SN9	Jet	Jet	Jet	DN63	DN63

community, with variations in the nodes and their sequences. At $\alpha = 0.1$, blue community nodes appear in addition to green community nodes. For $\alpha = 0.0$, it is dominated by nodes in small communities, namely the blue and yellow communities.

The community consideration betweenness centrality is also in line, $\alpha \leq 0.2$; the top-five nodes are dominated by large communities, namely green and violet, whereas nodes in small communities appear at $\alpha \leq 0.1$.

C. HALLMARK OF CANCER

1) DATASET

We retrieved data on genetic mutations in lung adenocarcinoma from four databases, TCGA, IntOGen, Bioportal, and Cosmic, which are available in the supplementary files. Each dataset was showed using Interactivenn [42] http://www.interactivenn.net/ (Fig. 8). Almost all genes obtained from IntOGen and Cosmic were also found in the other databases (18%) and 434/742 (58%). All these data were combined, and the goal was to obtain as much protein as possible. Protein interactions were determined using the STRING database, with the highest confidence score of 0.9 (Fig. 9). The giant component of the network consists of 703 nodes with 4889 edges. The network forms a scale-free network based on the degree distribution of the nodes.

TABLE 3. Size of communities.

Community	Size	Community	Size
Community 1	230	Community 7	12
Community 2	209	Community 8	9
Community 3	82	Community 9	9
Community 4	69	Community 10	3
Community 5	43	Community 11	2
Community 6	35		

2) COMMUNITY DETECTION

We used the greedy modularity disassembly algorithm [17] to detect the communities. This is a community crisp algorithm, in which each node is only in one community. The results of this algorithm formed 11 communities of different sizes with two large communities, four medium communities, and five small communities (Fig. 10) as shown in Table 3.

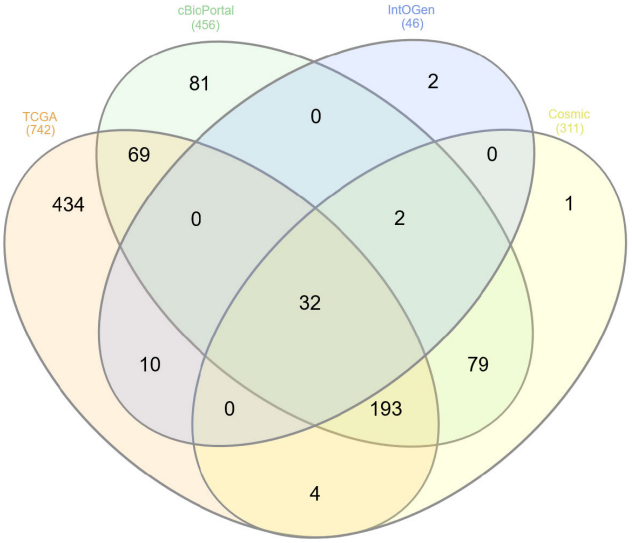

FIGURE 8. Venn Diagram of Lung Adenocarcinoma Genetic Mutations: Showing the overlapping and unique mutation data contribution of TCGA, IntOGen, Bioportal, and Cosmic database pointing to the combined dataset which is comprehensive.

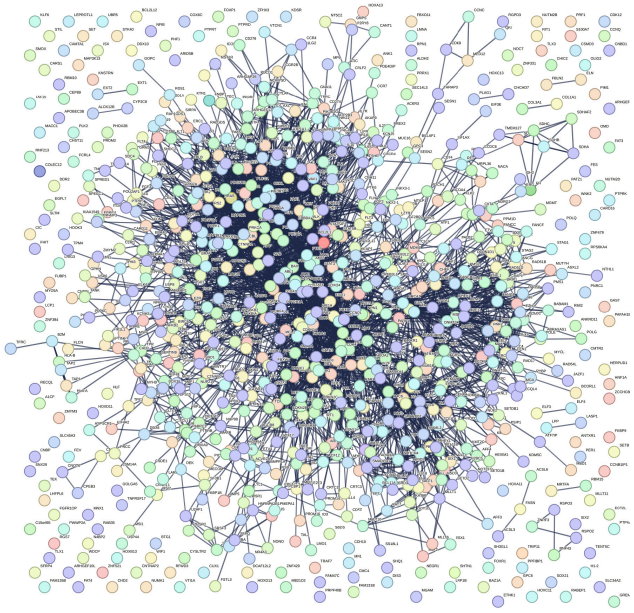

FIGURE 9. Network of Protein-Protein Interactions: This shows the complex interplay between proteins seen to result from the data derived from a genetic mutation in lung adenocarcinoma from the STRING database. It shows the relationships seen and possible interactions existing between these proteins that may imply their functional interaction.

Modularity, which shows the strength of the network division into communities, is 0.4684. This indicates that a community structure was formed in the network.

3) ENRICHMENT ANALYSIS

We used Metascape to analyze the results of structural groupings formed in the form of communities. The results

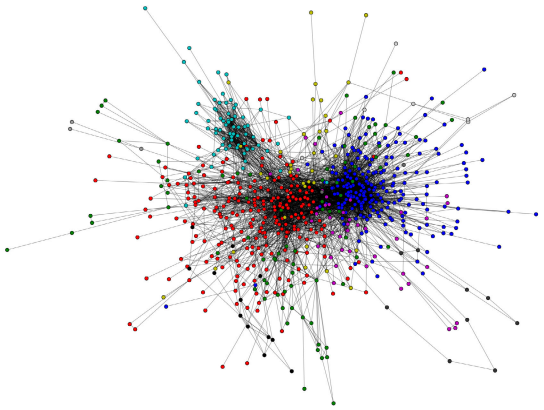

FIGURE 10. Lung Adenocarcinoma Protein Communities: shows protein communities, various size from the big to the small sub-groups, of the lung adenocarcinoma network and how they are related modularity-wise.

of the enrichment analysis of the majority of communities reflected functional grouping, in terms of biological processes, molecular functions, cellular components, and KEGG pathways, especially for large and medium-sized communities. In two major communities, 1 and 2, it was explicitly stated as a pathway to cancer in KEGG grouping. Some small communities did not produce functional groups (Table 4).

4) COMMUNITY-CONSIDERATION DEGREE CENTRALITY

We performed community-consideration degree centrality calculations and sorted the genes from the highest values for each variation in α values. At $\alpha = 1$ of the top 10 sequences (Table 5), it can be seen that degree centrality managed to obtain eight hallmark genes. With a degree of centrality that involves the community, that is, with $0.5 \leq \alpha \leq 0.8$, we can obtain more results, as many as nine genes.

We used the jackknife curve to test the performance of the ranking method [43], [44]. We compared the top-ranking protein results obtained using the number of cancer hallmarks from the Cosmic database. As the size of the shortlist increased, most genes became hallmarks of cancer (Fig. 11). Degree centrality, which considers only the community ($\alpha = 0$, red line), exhibits the worst performance. Degree centrality, which considers only the network as a whole ($\alpha = 1$,blue line), does not exhibit the best performance. However, when combined, it showed a better performance. For example, among the 50 gene lists, 35 (70%) were hallmarks of cancer ($\alpha = 0, 1$, green line). It can also be observed that the degree of centrality when considering the community is better than when it is not considering the community. The α values that give better results are 0.1 (green line) and 0.2 (gold line).

5) COMMUNITY-CONSIDERATION CLOSENESS CENTRALITY

We also tested the community-consideration degree centrality by ranking the genes with the highest values for each variation in α values. At $\alpha = 1$ of the top 10 sequences, it can be

TABLE 4. Enrichment analysis result.

Community	GO Biological Processes (%, log10(P))	GO Molecular Functions (%,log10(P))	GO Cellular Components (%,log10(P))	KEGG Pathway (%,log10(P))
1	chromatin organization (32.61%,-61.82)	transcription factor binding (36.52%,-82.58)	transcription regulator complex (31.30%,-69.75)	Transcriptional misregulation in cancer (22.61%,-65.58)
2	transmembrane receptor protein tyrosine kinase signaling pathway (37.80%,-91.97)	phosphotyrosine residue binding (8.61%,-26.63)	side of membrane (24.88%,-39.33)	Pathways in cancer (37.80%,-83.2)
3	regulation of innate immune response (19.51%,-14.13)	mRNA binding (18.29%,-10.12)	I-kappaB/NF-kappaB complex (3.66%,-7.12)	NF-kappa B signaling pathway (17.07%,-19.63)
4	DNA repair (84.06%,-91.7)	catalytic activity, acting on DNA (47.83%,-49.91)	nuclear chromosome (18.84%,-14.31)	Fanconi anemia pathway (27.54%,-37)
5	regulation of TOR signaling (25.58%,-16.99)	dendrite (18.60%,-5.65)	GTPase binding (11.63%,-4.11)	mTOR signaling pathway (34.88%,-23.49)
6	embryonic morphogenesis (37.14%,-13.36)	transcription factor binding (31.43%,-10.36)	Bcl-2 family protein complex (11.43%,-9.5)	Hedgehog signaling pathway (22.86%,-14.73)
7	histone lysine methylation (25.00%,-6.03)	histone binding (33.33%,-5.76)	transcription elongation factor complex (41.67%,-11.24)	-
8	ossification (44.44%,-6.06)	-	-	-
9	clathrin coat assembly (44.44%,-11.18)	septin ring (33.33%,-8.18)	cadherin binding (33.33%,-3.97)	Bacterial invasion of epithelial cells (44.44%,-8.31)
10	-	-	-	Fatty acid biosynthesis (100.00%,-9.75)
11	-	-	-	-

TABLE 5. Result of community-consideration degree centrality.

order	α										
	0.0	**0.1**	**0.2**	**0.3**	**0.4**	**0.5**	**0.6**	**0.7**	**0.8**	**0.9**	**1.0**
1	**ABL1**	SRC	SRC	SRC	SRC	SRC	SRC	SRC	**TP53**	**TP53**	**TP53**
2	ABL2	**PIK3R1**	**PIK3R1**	**PIK3R1**	**PIK3R1**	**PIK3R1**	**PIK3R1**	**TP53**	SRC	SRC	SRC
3	FASN	**PIK3CA**	**PIK3CA**	**PIK3CA**	**PIK3CA**	**PIK3CA**	**TP53**	**PIK3R1**	**PIK3R1**	**PIK3R1**	**PIK3R1**
4	MLLT1	**HRAS**	**HRAS**	**HRAS**	**HRAS**	**TP53**	**PIK3CA**	**PIK3CA**	**PIK3CA**	**PIK3CA**	**CTNNB1**
5	MLLT3	**BRCA1**	**EP300**	**EP300**	**EP300**	**HRAS**	**HRAS**	**EP300**	**EP300**	**CTNNB1**	**PIK3CA**
6	DOT1L	**EP300**	**KRAS**	**TP53**	**TP53**	**EP300**	**EP300**	**HRAS**	**CTNNB1**	**EP300**	**AKT1**
7	**BRCA1**	**ATM**	**PTPN11**	**KRAS**	**KRAS**	**CTNNB1**	**CTNNB1**	**CTNNB1**	**HRAS**	**AKT1**	**EP300**
8	**ATM**	**KRAS**	**TP53**	**PTPN11**	**CTNNB1**	**KRAS**	**AKT1**	**AKT1**	**AKT1**	**HRAS**	**HRAS**
9	**ACSL3**	**PTPN11**	FYN	FYN	**PTPN11**	**PTPN11**	**KRAS**	**KRAS**	**MAPK1**	**MAPK1**	**MAPK1**
10	ACSL6	FYN	SOS1	SOS1	FYN	**AKT1**	**PTPN11**	**MAPK1**	**KRAS**	MAPK3	MAPK3
# hallmark	4	8	7	7	8	9	9	9	9	8	8

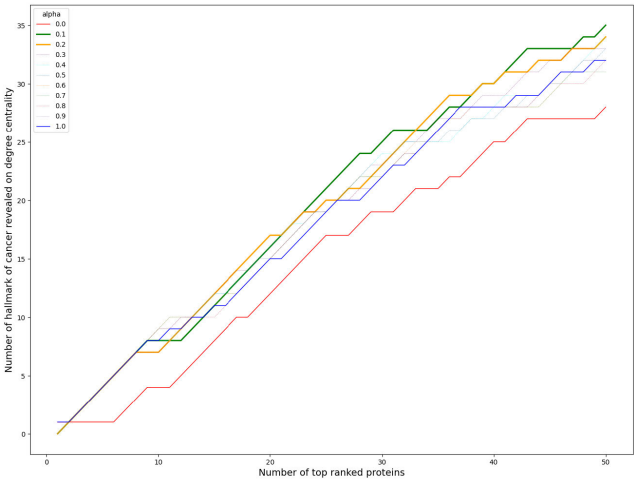

FIGURE 11. Jackknife curve analysis for the cancer hallmark genes identification through community consideration degree centrality.

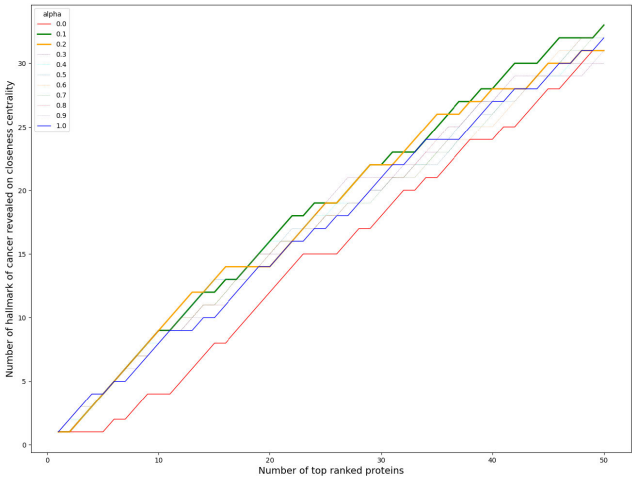

FIGURE 12. Jackknife curve analysis for the cancer hallmark genes identification through community consideration closeness centrality.

seen that closeness centrality managed to get eight hallmarks genes. This was similar to degree centrality, although there were differences in the proteins listed. By involving the community, community-consideration closeness centrality becomes higher results at $0.1 \leq \alpha \leq 0.3$; there were nine genes (Table 6). Furthermore, we extended the list

TABLE 6. Result of community-consideration closeness centrality.

order	α										
	0.0	**0.1**	**0.2**	**0.3**	**0.4**	**0.5**	**0.6**	**0.7**	**0.8**	**0.9**	**1.0**
1	**ABL1**	**PIK3CA**	**EP300**	**EP300**	**EP300**	**TP53**	**TP53**	**TP53**	**TP53**	**TP53**	**TP53**
2	ABL2	SRC	SRC	SRC	**TP53**	**EP300**	**EP300**	**EP300**	**AKT1**	**AKT1**	**AKT1**
3	FASN	**PIK3R1**	**PIK3CA**	**PIK3CA**	SRC	**AKT1**	**AKT1**	**AKT1**	**EP300**	**EP300**	**EP300**
4	MLLT3	**EP300**	**PIK3R1**	**TP53**	**PIK3CA**	SRC	SRC	**CTNNB1**	**CTNNB1**	**CTNNB1**	**CTNNB1**
5	MLLT1	**HRAS**	**TP53**	**PIK3R1**	**AKT1**	**PIK3CA**	**CTNNB1**	SRC	SRC	SRC	SRC
6	**ACSL3**	**PTPN11**	**HRAS**	**AKT1**	**CTNNB1**	**CTNNB1**	**PIK3CA**	**PIK3CA**	**PIK3CA**	**MAPK1**	**MAPK1**
7	ACSL6	**KRAS**	**AKT1**	**CTNNB1**	**PIK3R1**	**CREBBP**	**CREBBP**	**CREBBP**	**CREBBP**	**CREBBP**	MAPK3
8	**BRCA1**	**AKT1**	**CREBBP**	**CREBBP**	**CREBBP**	**MAPK1**	**MAPK1**	**MAPK1**	**MAPK1**	MAPK3	**CREBBP**
9	**ATM**	**CREBBP**	**CTNNB1**	**MAPK1**	**MAPK1**	MAPK3	MAPK3	MAPK3	MAPK3	**PIK3CA**	**MYC**
10	RAD51	**EGFR**	**KRAS**	**MYC**	MAPK3	**PIK3R1**	**MYC**	**MYC**	**MYC**	**MYC**	**ESR1**
# hallmark	4	9	9	9	8	8	8	8	8	8	8

to 50 genes with community-consideration closeness centrality with a variety of $\alpha = 0.1$ represented by the jackknife curve (Fig. 12). Community-consideration degree centrality showed better performance when we computed the closeness centrality considered for the community. For the top 50 shortlists at $\alpha = 0.1$, the community-consideration closeness centrality obtained 33 (66%) hallmarks of cancer. The best α that generates the most hallmarks of cancer is $\alpha = 0.1$ (green line) and $\alpha = 0.2$ (gold line)

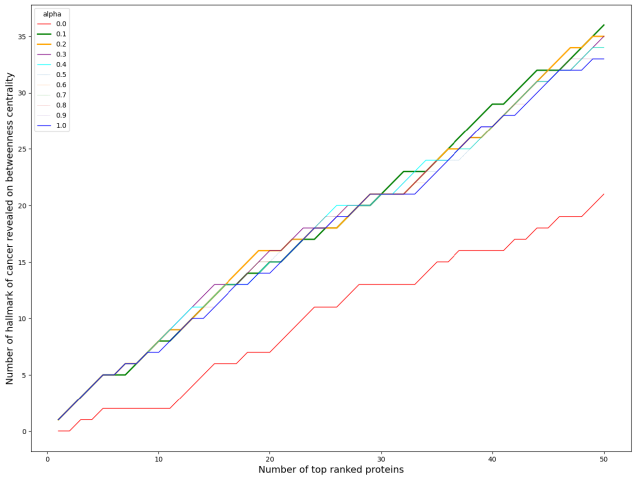

FIGURE 13. Jackknife curve analysis on the cancer hallmark genes identification through community-consideration betweenness centrality.

6) COMMUNITY-CONSIDERATION BETWEENNESS CENTRALITY

We did the same for community-consideration betweenness centrality, showing results for values ranging from $0.1 \leq \alpha \leq 0.7$ to 8 of the top 10 gene sequences (Table 7). Furthermore, we extended to 50 gene lists with community consideration betweenness centrality using the Jackknife curve (Fig. 13). It has become evident that most genes are hallmarks of cancer. Betweenness centrality, which considers only the community ($\alpha = 0$, red line) demonstrates the poorest performance. Betweenness centrality, which solely considers the network as a whole ($\alpha = 0$, blue line), also does not exhibit an optimal performance. However, when combined with community considerations, better performance was achieved. Optimal results are achieved at α values of 0.1 (green line) and 0.2 (gold line), followed by 0.3 and 0.4.

7) COMPARISON BETWEEN COMMUNITY-CONSIDERATION DEGREE, CLOSENESS AND BETWEENNESS CENTRALITY

We compared the best results of community-consideration degree, closeness and betweenness centrality at $\alpha = 0.1$ and $\alpha = 0.2$. As shown in Fig. 14, the top-ranked n protein, for $n = 5$ to $n = 16$, community-consideration closeness centrality gives the best results in obtaining the hallmark of cancer (thin line). For $n = 17$ to $n = 46$, community-consideration degree centrality (thick line) was the best, while for the rest, $n > 47$, community-consideration betweenness centrality (dotted line) is the best. It can be concluded that each community consideration can provide the most significant results for identifying hallmarks of cancer.

We also compared our community consideration centrality method with other popular centrality algorithms. The effectiveness of the method in identifying the critical nodes within these complex networks was ascertained in detail. We narrowed the focus on α values that yielded optimal results for community consideration centrality at $\alpha = 0.1$ and $\alpha = 0.2$. These α settings enabled us to fine-tune the weight between the local community and global network influence in our centrality calculations. Our findings showed that centrality with these α values outperformed traditional centrality, such as Information Centrality, Load Centrality, and Harmonic Centrality, among others, in detecting crucial nodes. Such a trend was all the more discernible concerning biological networks where our approach exhibited better performance than the other centrality measures towards detecting hallmark genes associated with lung adenocarcinoma cancer. The better performance of our method at $\alpha = 0.1$ and $\alpha = 0.2$ underlines the critical role of integrating community structures within network centrality to offer a more elaborate and accurate understanding of the influential nodes in complex systems.

Most benchmark algorithms, including Eigenvector Centrality, Katz Centrality and Information Centrality, have a computational complexity of $O(n^3)$. This is further explained by the fact that most of them depend on matrix operations

TABLE 7. Result of community-consideration betweenness centrality.

order	α										
	0.0	**0.1**	**0.2**	**0.3**	**0.4**	**0.5**	**0.6**	**0.7**	**0.8**	**0.9**	**1.0**
1	FASN	**TP53**	**TP53**	**TP53**	**TP53**	**TP53**	**TP53**	**TP53**	**TP53**	**TP53**	**TP53**
2	CREB1	**EP300**	**EP300**	**EP300**	**EP300**	**CTNNB1**	**CTNNB1**	**CTNNB1**	**CTNNB1**	**CTNNB1**	**CTNNB1**
3	**EPS15**	**CTNNB1**	**CTNNB1**	**CTNNB1**	**CTNNB1**	**EP300**	**EP300**	**EP300**	**EP300**	**EP300**	**EP300**
4	IGF2	**SMAD3**	**SMAD3**	**AKT1**	**AKT1**	**AKT1**	**AKT1**	**AKT1**	**AKT1**	**AKT1**	**AKT1**
5	**GPC3**	**AKT1**	**AKT1**	**SMAD3**	**SMAD3**	**SMAD3**	**SMAD3**	**SMAD3**	**SMAD3**	**SMAD3**	**SMAD3**
6	KLK2	SRC	SRC	SRC	SRC	SRC	SRC	SRC	SRC	SRC	SRC
7	PARK2	EIF4E	**SMARCA4**	**BRCA1**	**BRCA1**	**BRCA1**	**BRCA1**	**BRCA1**	**BRCA1**	**BRCA1**	**BRCA1**
8	EIF4E	**SMARCA4**	JUN	JUN	JUN	JUN	JUN	JUN	JUN	JUN	JUN
9	PKN1	**CASP8**	**BRCA1**	**SMARCA4**	**MYC**	**MYC**	**MYC**	**MYC**	**MYC**	**MYC**	**MYC**
10	SEPT5	**RHOA**	**MYC**	**MYC**	**SMARCA4**	**SMARCA4**	**SMARCA4**	**SMARCA4**	HSP90AA1	HSP90AA1	HSP90AA1
# hallmark	2	8	8	8	8	8	8	8	7	7	7

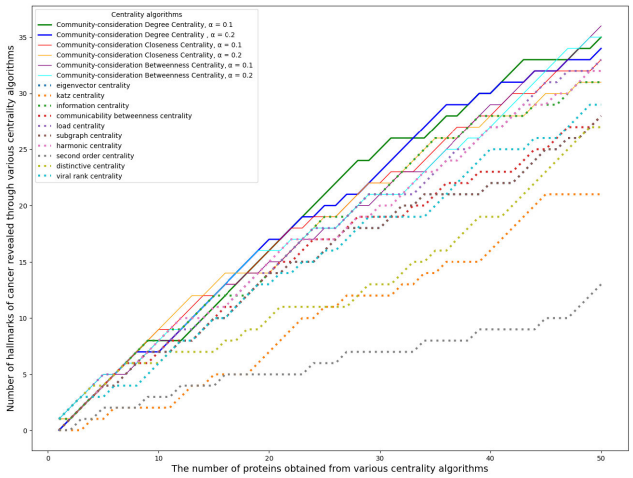

FIGURE 14. Jackknife curve analysis showing detection of cancer hallmarks with community-consideration centralities (Degree, Closeness and Betweenness) at $\alpha = 0.1$ and $\alpha = 0.2$ vis-a-vis as benchmarks.

such as eigenvalue computation and computation which involve matrix inversion, and have cubic characteristics in line with those of the network nodes. Harmonic Centrality works with an $O(nm)$ complexity where the m is the number of edges. This reduced complexity provides computational benefits, particularly for large-scale sparse networks. The proposed algorithm, community consideration centrality (including the degree, closeness and betweenness measures), complements all classical centrality measures from a traditional computational complexity perspective. Specifically, it shares the $O(n^3)$ complexity for the same reasons as its counterparts – intensive matrix operations standard for network analysis.

D. ANALYSIS OF HALLMARK GENES IN CANCER THROUGH COMMUNITY CONSIDERATION CENTRALITY

Using the community consideration centrality approach, which utilizes degree, closeness, and betweenness centrality, as outlined in Table 5, 6 and 7, we identified a set of genes that play a crucial role in the development and progression of cancer. Our analysis revealed that 18 genes, including AKT1, ATM, BRCA1, CASP8, CREBBP, CTNNB1, EGFR, EP300, HRAS, KRAS, MYC, PIK3CA, PIK3R1, PTPN11, RHOA, SMAD3, SMARCA4, and TP53, played a significant role in cancer hallmarks at $\alpha = 0.1$ and $\alpha = 0.2$. These genes have been mapped to the COSMIC database, and their specific roles in key cancer processes such as proliferative signaling, growth suppression, escaping immune response of cancer, replicative immortality, tumor-promoting inflammation, invasion and metastasis, angiogenesis, genome instability and mutations, escape programmed cell death, and changes in cellular energetics are shown in Table 8.

TABLE 8. Hallmark genes in cancer.

Gene	H1	H2	H3	H4	H5	H6	H7	H8	H9	H10
AKT1	P	P	-	-	-	P/S	P/S	S	P	P
ATM	-	-	-	P	P	S	-	S	P	-
BRCA1	-	-	S	-	-	-	-	-	-	-
CASP8	P	-	-	-	P/S	P/S	P	-	S	-
CREBBP	P	-	-	-	-	P	-	S	P/S	-
CTNNB1	P	P	P	P	-	P	P	S	P	P
EGFR	P	-	P	-	-	P	P	-	P	P
EP300	-	P	-	-	-	-	-	-	S	-
HRAS	P	-	-	-	P	P	P	P	P	-
KRAS	P	-	-	P	P	P	P	-	P	P
MYC	P	-	P	S	-	S	P	P	S	P
PIK3CA	P	-	-	-	-	P	P	-	P	-
PIK3R1	-	P	-	S	-	S	-	-	-	-
PTPN11	P	P	P	-	P/S	P	P/S	P/S	P/S	-
RHOA	-	P	-	-	S	P/S	-	S	P/S	-
SMAD3	P	P	-	-	-	P	-	-	P/S	-
SMARCA4	P	P	-	S	S	P/S	P	S	P	P
TP53	-	P	S	P/S	S	S	S	S	P/S	P

P = Promotes ; S = Suppresses
H1=proliferative signaling; H2=suppresion of growth; H3=escaping immune response of cancer; H4=cell replicative immortality; H5=tumour promoting inflammation; H6=invasion and metastasis; H7=angiogenesis; H8=genome instability and mutations; H9=escaping programmed cell death; H10=change of cellular energetics

Our analysis further investigated the categorization of these genes, identifying them as either oncogenes or tumor suppressor gene classes, based on their regulatory patterns in cancer. Oncogenes are normally overexpressed in cancerous tissues and promote tumorigenesis by stimulating cell proliferation, survival, and other oncogenic processes. However, tumor suppressor genes are significantly downregulated or inactivated in cancerous conditions, with frequent downregulation antagonizing the modulation of cell growth, DNA repair, and apoptosis mechanisms to reduce the incidence or promote regression of disease.

For example, KRAS is a classic prototype oncogene that is upregulated in various cancers, leading to continuous proliferative signaling. Conversely, TP53, which is considered a tumor suppressor gene, is mostly downregulated or mutated in cancer, leading to a loss of control over the immune response and cell death.

E. PARAMETER OPTIMIZATION IN COMMUNITY CONSIDERATION CENTRALITY

In the crucial discussion regarding consideration weight α, this parameter ranged from 0.0 (signifying full consideration of communities) to 1.0 (indicating sole reliance on network structure upon centrality calculation), corresponding sensitively to centrality measures. This is crucial for the studying networks with different community structures. Our empirical analysis revealed a strong influence of community structure on centrality within such networks, especially with α values close to zero ($\alpha = 0.1$ and $\alpha = 0.2$). This observation is essential within networks with community structures such as those obtained from our lung cancer datasets. In these networks, clustered and interconnected nodes within communities are critical for the functional integrity of the networks. It is illustrates the importance of low α values, revealing that more genes are hallmarks of cancer. However, the selection of α approximating 1 is more appropriate when no communities can be identified or when the modularity is low. Centrality measures are influenced more by the overall network structure than by community structure. Choosing $\alpha = 0.1$ and $\alpha = 0.2$ as near-optimal solutions demonstrates that community structures are essential in biological network analysis. This underscores the urgency of considering the individual properties of nodes and their roles within communities. It is more so for scientists who try to make sense amidst complex biological systems where understanding individual and collective behaviors becomes exceedingly significant. Thus, our choice of α value in this study not only advances the methodology of network analysis but also highlights community structures as essential components for identifying critical elements within biological networks. Therefore, our optimal α settings, indicate the strong influence of community structure and support the need for a nuanced approach that balances individual node significance with the community context and networks in which community structures reside.

F. EXPANDONG HORIZONS: COMMUNITY-CONSIDERATION CENTRALITY IN DIVERSE DOMAINS

Thus, this study offers significant development for network analysis literature because it innovatively proposes more consideration of community structures to be given as input when computing centrality measures. Given the potential value of our methodology beyond lung cancer, it has general implications and provides valuable insights across various domains. Traditional centrality measures often overlook the intricate sub-structures within a network. By incorporating these communities, this approach allows for a more detailed and accurate analysis of networks,thereby enhancing our understanding of the importance of individual nodes and the collective influence of communities.

Community-consideration centrality is not only limited to biological networks but also applies to other fields such as social network analysis, organizational studies, and communication networks. For instance, in social network analysis, community structure effects are essential for sufficiently identifying critical influencers or understanding group dynamics. Similarly, in organizational networks, this can be useful in identifying centralized units or departments that substantially impact the flow of information and, thus, organizational efficiency. Integrating community structures into epidemological network models can help delve deeper into the mechanisms of disease spread. This can help detect potential hotspots or super-spreaders in communities, thus helping to design more targeted and better intervention measures. This approach is also helpful for understanding the dynamics of disease spread in dense populations. Furthermore, our system biology method helps decode complex biological networks such as cellular processes, gene regulation, and metabolic pathways. Thus, understanding the role of these community structures in such networks may provide new biomarkers for diseases and potential drug targets, and a better understanding of the mechanisms associated with the plethora of biological functions being studied.

The concept of community consideration centrality opens up research in new areas that study the dynamics of communities among different types of networks. Further research may aim to develop more sophisticated models that dynamically adjust community weights according to network characteristics, apply them to overlapping communities, or translate this concept into large-scale networks to understand global patterns and behaviors. The broader implications of our research indicate the importance of community structure in contention with network analyses. This adds value to theoretical considerations in network science and provides practical tools and insights for managing real-life problems in all possible scientific and social areas.

V. CONCLUSION

This study proposes three centralities that consider the community: the community consideration degree, closeness, and betweenness centrality. Using this approach, community involvement with a certain weight can be considered,if a community is found in the network. We tested some real-world Zachary Karate and Dolphin datasets with proportional results, and we used community-consideration centrality to determine the hallmark of lung cancer and showed better results than using community-consideration centrality. Even though the performance is good, there is still potential for improvement and development: for example the relationship

between the community structures formed to determine α and determine the community-consideration centrality of overlapping communities.

REFERENCES

[1] S. Gómez, "Centrality in networks: Finding the most important nodes," in *Business and Consumer Analytics: New Ideas*. Switzerland: Springer, 2019, pp. 401–433.

[2] L. C. Freeman, "Centrality in social networks conceptual clarification," *Social Netw.*, vol. 1, no. 3, pp. 215–239, Jan. 1978.

[3] K. Das, S. Samanta, and M. Pal, "Study on centrality measures in social networks: A survey," *Social Netw. Anal. Mining*, vol. 8, no. 1, pp. 1–11, Dec. 2018.

[4] S. B. S. Omit, S. Akhter, H. K. Rana, A. R. M. Rana, N. K. Podder, M. I. Rakib, and A. Nobi, "Identification of comorbidities, genomic associations, and molecular mechanisms for COVID-19 using bioinformatics approaches," *BioMed Res. Int.*, vol. 2023, Jan. 2023, Art. no. 6996307.

[5] X.-Y. Zhang, T.-Y. He, C.-Y. Xu, K.-F. Cao, and X.-S. Zhang, "Theoretical investigation of the pathway-based network of type 2 diabetes mellitus-related genes," *Eur. Phys. J. B*, vol. 96, no. 6, p. 86, Jun. 2023.

[6] B. A. Monchka, C. K. Leung, N. C. Nickel, and L. M. Lix, "The effect of disease co-occurrence measurement on multimorbidity networks: A population-based study," *BMC Med. Res. Methodol.*, vol. 22, no. 1, p. 165, Jun. 2022.

[7] F. Zhou and S. Uddin, "Interpretable drug-to-drug network features for predicting adverse drug reactions," *Healthcare*, vol. 11, no. 4, p. 610, Feb. 2023.

[8] G. Li, M. Li, J. Wang, Y. Li, and Y. Pan, "United neighborhood closeness centrality and orthology for predicting essential proteins," *IEEE/ACM Trans. Comput. Biol. Bioinf.*, vol. 17, no. 4, pp. 1451–1458, Jul. 2020.

[9] S. Uddin, A. Khan, H. Lu, F. Zhou, and S. Karim, "Suburban road networks to explore COVID-19 vulnerability and severity," *Int. J. Environ. Res. Public Health*, vol. 19, no. 4, p. 2039, 2022.

[10] C. Durón, Y. Pan, D. H. Gutmann, J. Hardin, and A. Radunskaya, "Variability of betweenness centrality and its effect on identifying essential genes," *Bull. Math. Biol.*, vol. 81, no. 9, pp. 3655–3673, Sep. 2019.

[11] R. Jothi, "A betweenness centrality guided clustering algorithm and its applications to cancer diagnosis," in *Mining Intelligence and Knowledge Exploration*, A. Ghosh, R. Pal, and R. Prasath, Eds. Cham, Switzerland: Springer, 2017, pp. 35–42.

[12] A.-L. Barabasi, *Network Science*. Cambridge, U.K.: Cambridge Univ. Press, 2016.

[13] S. Fortunato, "Community detection in graphs," *Phys. Rep.*, vol. 486, nos. 3–5, pp. 75–174, Feb. 2010.

[14] F. Liu, S. Xue, J. Wu, C. Zhou, W. Hu, C. Paris, S. Nepal, J. Yang, and P. S. Yu, "Deep learning for community detection: Progress, challenges and opportunities," in *Proc. Int. Joint Conf. Artif. Intell.*, 2020, pp. 1–7.

[15] S. Rahiminejad, M. R. Maurya, and S. Subramaniam, "Topological and functional comparison of community detection algorithms in biological networks," *BMC Bioinf.*, vol. 20, no. 1, pp. 1–25, Dec. 2019.

[16] I. Manipur, M. Giordano, M. Piccirillo, S. Parashuraman, and L. Maddalena, "Community detection in protein–protein interaction networks and applications," *IEEE/ACM Trans. Comput. Biol. Bioinf.*, vol. 20, no. 1, pp. 217–237, Jan. 2023.

[17] H. C. Rustamaji, W. A. Kusuma, S. Nurdiati, and I. Batubara, "Community detection with greedy modularity disassembly strategy," Tech. Rep., 2023.

[18] M. E. J. Newman, "Fast algorithm for detecting community structure in networks," *Phys. Rev. E, Stat. Phys. Plasmas Fluids Relat. Interdiscip. Top.*, vol. 69, no. 6, Jun. 2004, Art. no. 066133.

[19] V. D. Blondel, J.-L. Guillaume, R. Lambiotte, and E. Lefebvre, "Fast unfolding of communities in large networks," *J. Stat. Mech., Theory Exp.*, vol. 2008, no. 10, Oct. 2008, Art. no. P10008.

[20] M. Girvan and M. E. J. Newman, "Community structure in social and biological networks," *Proc. Nat. Acad. Sci. USA*, vol. 99, no. 12, pp. 7821–7826, Jun. 2002.

[21] K. Berahmand, Y. Li, and Y. Xu, "DAC-HPP: Deep attributed clustering with high-order proximity preserve," *Neural Comput. Appl.*, vol. 35, no. 34, pp. 24493–24511, Dec. 2023.

[22] M. Rostami, M. Oussalah, K. Berahmand, and V. Farrahi, "Community detection algorithms in healthcare applications: A systematic review," *IEEE Access*, vol. 11, pp. 30247–30272, 2023.

[23] E. Landau, "Zur relativen wertbemessung der turnierresultate," *Deutsches Wochenschach*, vol. 11, pp. 366–369, 1895.

[24] L. Katz, "A new status index derived from sociometric analysis," *Psychometrika*, vol. 18, no. 1, pp. 39–43, Mar. 1953.

[25] K. Stephenson and M. Zelen, "Rethinking centrality: Methods and examples," *Social Netw.*, vol. 11, no. 1, pp. 1–37, Mar. 1989.

[26] K.-I. Goh, B. Kahng, and D. Kim, "Universal behavior of load distribution in scale-free networks," *Phys. Rev. Lett.*, vol. 87, no. 27, Dec. 2001, Art. no. 278701.

[27] M. E. J. Newman, "Scientific collaboration networks. II. Shortest paths, weighted networks, and centrality," *Phys. Rev. E, Stat. Phys. Plasmas Fluids Relat. Interdiscip. Top.*, vol. 64, no. 1, p. 16132, Jun. 2001.

[28] E. Estrada and J. A. Rodríguez-Velázquez, "Subgraph centrality in complex networks," *Phys. Rev. E, Stat. Phys. Plasmas Fluids Relat. Interdiscip. Top.*, vol. 71, no. 5, p. 56103, May 2005.

[29] E. Estrada, D. J. Higham, and N. Hatano, "Communicability betweenness in complex networks," *Phys. A, Stat. Mech. Appl.*, vol. 388, no. 5, pp. 764–774, Mar. 2009.

[30] A.-M. Kermarrec, E. Le Merrer, B. Sericola, and G. Trédan, "Second order centrality: Distributed assessment of nodes criticity in complex networks," *Comput. Commun.*, vol. 34, no. 5, pp. 619–628, Apr. 2011.

[31] P. Boldi and S. Vigna, "Axioms for centrality," 2013, *arXiv:1308.2140*.

[32] F. Iannelli, M. S. Mariani, and I. M. Sokolov, "Influencers identification in complex networks through reaction-diffusion dynamics," *Phys. Rev. E, Stat. Phys. Plasmas Fluids Relat. Interdiscip. Top.*, vol. 98, no. 6, p. 62302, Dec. 2018.

[33] A. Fronzetti Colladon and M. Naldi, "Distinctiveness centrality in social networks," *PLoS ONE*, vol. 15, no. 5, May 2020, Art. no. e0233276.

[34] W. W. Zachary, "An information flow model for conflict and fission in small groups," *J. Anthropological Res.*, vol. 33, no. 4, pp. 452–473, Dec. 1977.

[35] D. Lusseau, K. Schneider, O. J. Boisseau, P. Haase, E. Slooten, and S. M. Dawson, "The bottlenose dolphin community of doubtful sound features a large proportion of long-lasting associations," *Behav. Ecology Sociobiol.*, vol. 54, no. 4, pp. 396–405, Sep. 2003.

[36] J. Cheng, X. Yin, Q. Li, H. Yang, L. Li, M. Leng, and X. Chen, "Voting simulation based agglomerative hierarchical method for network community detection," *Sci. Rep.*, vol. 8, no. 1, pp. 1–11, May 2018.

[37] J. Liu, T. Lichtenberg, and K. A. Hoadley, "An integrated TCGA pan-cancer clinical data resource to drive high-quality survival outcome analytics," *Cell*, vol. 173, no. 2, pp. 400–416, 2018.

[38] F. Martínez-Jiménez, F. Muiños, I. Sentís, J. Deu-Pons, I. Reyes-Salazar, C. Arnedo-Pac, L. Mularoni, O. Pich, J. Bonet, H. Kranas, A. Gonzalez-Perez, and N. Lopez-Bigas, "A compendium of mutational cancer driver genes," *Nature Rev. Cancer*, vol. 20, no. 10, pp. 555–572, Aug. 2020.

[39] E. Cerami, J. Gao, U. Dogrusoz, B. E. Gross, S. O. Sumer, B. A. Aksoy, A. Jacobsen, C. J. Byrne, M. L. Heuer, E. Larsson, Y. Antipin, B. Reva, A. P. Goldberg, C. Sander, and N. Schultz, "The cBio cancer genomics portal: An open platform for exploring multidimensional cancer genomics data," *Cancer Discovery*, vol. 2, no. 5, pp. 401–404, May 2012.

[40] J. G. Tate, S. Bamford, H. C. Jubb, Z. Sondka, D. M. Beare, N. Bindal, H. Boutselakis, C. G. Cole, C. Creatore, E. Dawson, and P. Fish, "COSMIC: The catalogue of somatic mutations in cancer," *Nucleic Acids Res.*, vol. 47, pp. 941–947, Jan. 2019.

[41] D. Hanahan and R. A. Weinberg, "Hallmarks of cancer: The next generation," *Cell*, vol. 144, no. 5, pp. 646–674, Mar. 2011.

[42] H. Heberle, G. V. Meirelles, F. R. da Silva, G. P. Telles, and R. Minghim, "InteractiVenn: A web-based tool for the analysis of sets through venn diagrams," *BMC Bioinf.*, vol. 16, no. 1, pp. 1–7, May 2015.

[43] J. Zhong, C. Tang, W. Peng, M. Xie, Y. Sun, Q. Tang, Q. Xiao, and J. Yang, "A novel essential protein identification method based on PPI networks and gene expression data," *BMC Bioinf.*, vol. 22, no. 1, p. 248, Dec. 2021.

[44] S. Li, Z. Zhang, X. Li, Y. Tan, L. Wang, and Z. Chen, "An iteration model for identifying essential proteins by combining comprehensive PPI network with biological information," *BMC Bioinf.*, vol. 22, no. 1, p. 430, Sep. 2021.

HERU CAHYA RUSTAMAJI was born in Yogyakarta, Indonesia, in 1971. He received the bachelor's degree in computer science from Gadjah Mada University, and the master's degree from the Bandung Institute of Technology. He is currently pursuing the Ph.D. degree in computer science with IPB University. He is a Lecturer with UPN Veteran Yogyakarta Informatics Engineering. His research interests include network science, bioinformatics, and machine learning.

WISNU ANANTA KUSUMA received the bachelor's and master's degrees from the Bandung Institute of Technology, and the Ph.D. degree from the Tokyo Institute of Technology, in 2012. He is currently an Associate Professor with the Computer Science Department, IPB University. He is also the Executive Secretary of Institute for International Research on Advanced Technology, IPB University, coordinator of the Bioinformatics Working Group, Faculty of Mathematics and Natural Science, IPB University, and coordinator of the Bioinformatics and High-Performance Computing Research Group, Advanced Research Laboratory, IPB University. He has been the author of more than 60 articles and has reviewed international journals. His current research interests include machine learning, high-performance computing, and bioinformatic research. He is the Chairperson of the Indonesian Society of Bioinformatics and Biodiversity and an ExCo Member of the Asia Pacific Bioinformatics Network (APBioNet).

SRI NURDIATI received the bachelor's degree from the Department of Statistics, FMIPA, IPB University, the master's degree in computer science from the University of Western Ontario, Canada, and the doctoral degree in applied mathematics from Twente University, The Netherlands. She is currently a Professor of mathematical science with the Mathematics Department, Faculty of Mathematics and Natural Sciences, IPB University. She has conducted extensive research in applied mathematics and network science.

IRMANIDA BATUBARA received the bachelor's and master's degrees from the Chemistry Department, IPB University, Indonesia, and the Ph.D. degree from Gifu University, Japan, in 2009. She was the Director of the Tropical Biopharmaca Research Center, IPB University. She is currently a Professor with the Department of Chemistry, Faculty of Mathematics and Natural Sciences, IPB University.