IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 9 July 2024, accepted 26 July 2024, date of publication 1 August 2024, date of current version 15 August 2024.

Digital Object Identifier 10.1109/ACCESS.2024.3437215

RESEARCH ARTICLE

A New Framework for Pinpointing Crucial Proteins in Protein-Protein Interaction Networks

ABDUL MOIZ[1], UBAIDA FATIMA[2], AND M. ZEESHAN UL HAQUE[1], (Senior Member, IEEE)
[1]Department of Biomedical Engineering, Salim Habib University, Karachi 74900, Pakistan
[2]Department of Mathematics, NED University of Engineering & Technology, Karachi 75270, Pakistan
Corresponding author: Abdul Moiz (moizmazhar2015@gmail.com)

ABSTRACT Identifying crucial proteins in protein-protein interaction (PPI) networks is essential for understanding biological systems. However, the ambiguity of interaction strength hinders accurate identification of key proteins. To address this issue, this study proposes a new framework that introduces Bio-Link Strength, a fuzzy membership function that utilizes fuzzy set theory to quantifies interaction strength with continuous values between 0 and 1. Framework extends four commonly used traditional measures (degree, closeness, betweenness, and eigenvector) to fuzzy measures (fuzzy connectivity, fuzzy approachability, fuzzy bridge, and fuzzy influence centrality), enabling effective identification of crucial proteins. The efficacy of the proposed framework has been assessed on different commonly used real-world PPI network datasets (Saccharomyces cerevisiae, Escherichia coli, and Drosophila melanogaster), to prove the framework's scalability. Results show that proposed membership function effectively assesses protein interaction strength, with a strong positive Spearman's correlation between fuzzy and traditional measures. Furthermore, Gene Ontology analysis confirms the importance of top proteins identified by our fuzzy measures. Notably, our fuzzy connectivity and influence centrality measures outperform their traditional counterparts and other proposed fuzzy measures in identifying crucial proteins.

INDEX TERMS Crucial proteins, protein-protein interaction networks, edge strength, fuzzy membership function, fuzzy centrality measures, fuzzy biological networks.

I. INTRODUCTION

Proteins are the fundamental building blocks of living organisms, constituting biological cells and tissues, and maintaining life activities. As indispensable components of physiological functions, proteins are closely tied to the physiological states within living organisms [1]. The study of proteins is crucial, as predicting their functions, discovering their structures, and exploring protein-protein interactions are essential for understanding the intricacies of cellular processes [2]. Furthermore, protein interaction networks provide a comprehensive framework that integrates diverse information, including topological network properties, correlations between proteins, and functional relationships, offering a powerful tool for elucidating the complex interactions that underlie life processes [3].

In present time many studies and researches showed that there has been a marked surge in the inference of biological graph networks, particularly the investigation of protein- protein interaction networks (PPI) [4], [5], [6]. The considered networks are analyzed through depiction as graphs like $G_{PPI} = (\mathbb{N}_P, \mathscr{E}_I)$, where nodes or vertices ($\mathbb{N}_P$) depicts proteins, and edges ($\mathscr{E}_I$) indicate their associations or communications. Within graph theory, a significant area of research focuses on pinpointing crucial proteins. Proteins typically exhibit extensive interconnections with other components within the considered complex graphs, underscoring their significance for the network system proper operations [7]. The deletion of any of these crucial proteins is more prone to result in lethality as contrast to the

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

removal of non-crucial proteins, emphasizing their indispensable role in sustaining the protein graph integrity and functionality [8], [9].

A. NETWORK ANALYSIS TECHNIQUES

The concept of centrality in complex biological graph network analysis stands as a basic and widely employed concept as proved in proved by many researchers in literature in different domains. Diverse collections of centrality metrics been conceived and applied to pick the most vital proteins within the considered biological network graph datasets. The inference of these measures serves as invaluable tools for discerning the significance and influence of proteins to learn complex processes of biological graph interactions [10]. These methods mainly include Degree centrality (DC) [11], Closeness centrality (CC) [12], Betweenness centrality (BC) [13], Eigenvector centrality (EC) [14], Subgraph centrality (SC) [15], and Information centrality (IC) [16].

Popular and commonly used centrality methods in complex network analysis typically operate under the shared assumption that proteins exhibiting the highest centrality within the network are more arrowed to be indispensable compared to those with lower centrality values [17]. Researchers findings in Previous studies clearly indicate that methods grounded in network graph topology are proficient in the identification of vital nodes or proteins [18], [19], [20]. High-throughput techniques are employed to generate Protein-Protein Interaction (PPI) graph networks, leading to a substantial number of false-positive interactions (FPI). For catering this issue, many researches in the literature have suggested preprocessing phenomena that leverage topological properties to assess and eliminate these false positives nodes. As an example, Asur et al. [21] introduced an ensemble clustering framework that incorporates two topology-based similarity metrics. This approach aims to mitigate noise within Protein-Protein Interaction (PPI) graph networks and deduce biologically meaningful functional modules. Chen et al. [22] devised an iterative approach employing alternative paths and interaction generalities to enhance interactomes through refinement based on network graph topology. Additionally, these networks (PPI) do not consider the strength of protein interactions [23], which lowers their accuracy in predicting true crucial proteins. Some protein interactions are more crucial than others. However, the existing networks share the common attribute that the interaction strength is consistently crisp, with a value of either 0 or 1, shown in Fig. 1. However, this is not feasible with real-world data.

To overcome this problem, new measures have been proposed. For instance, Li et al. [24] proposed a method for assigning confidence scores to each interaction based on experimental and functional evidence, showing that using weighted centrality measures can improve the identification of crucial proteins in yeast. In their study Peng et al. [25]

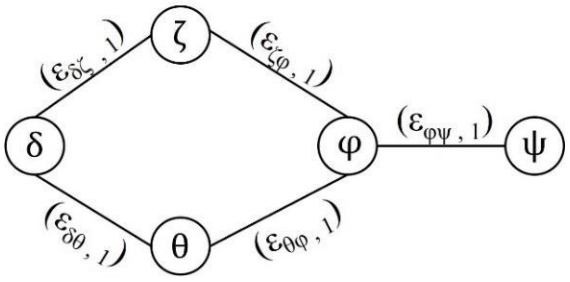


FIGURE 1. A simple protein-protein interaction network. Nodes $\delta, \zeta, \theta, \varphi$, and ψ represent proteins, and edges $\varepsilon_{\delta\zeta}, \varepsilon_{\delta\theta}, \varepsilon_{\zeta\varphi}, \varepsilon_{\theta\varphi}$, and $\varepsilon_{\varphi\Psi}$ signify interactions with their corresponding crisp weights representing interaction strengths.

introduced a new method, which integrates orthology and PPI networks to identify crucial proteins. Meng et al. [26] developed a method for crucial protein prediction based on a novel weighted protein domain interaction network that integrates gene expression data, protein domain associations, and topological features that minimize the noise and improve the precision of PPI networks. Huang et al. [27] have developed a series of methods for predicting protein-protein interactions using protein sequence data. They introduced a novel protein sequence representation, and used a weighted sparse representation-based classifier to achieve high prediction accuracies. Chen et al. [28] proposed a new predictive model that leverages GO terms and KEGG pathways to identify essential and non-essential genes. Li et al. [29] propose a new centrality measure by integrating protein-protein interaction and gene expression data. Unlike other centrality measures, it determines a protein's cruciality based on its connectivity and whether it is highly likely to be clustered and coexpressed with its neighbors. Berahmand et al. [30] proposed a new semilocal centrality measure that leverages the natural characteristics of complex networks to identify influential spreaders. Zhong et al. [31] proposed a new JDC metric based on PPI network and gene expression data. The proposed approach binarizes gene expression data using a dynamic threshold method and then combines the degree centrality and Jaccard similarity index to generate the JDC score for proteins in the PPI network. The JDC score shows the similarity of active and inactive gene expression states in a network cluster, effectively reducing the effects of false positives and negatives in PPI network. Similarly, [32] introduced a new GOS method that combines expression data, orthology, and sub-cellular localization information to identify crucial proteins. Moreover, new clustering methods [7], [33] have been proposed for detecting clusters in complex networks. Furthermore, Researchers [34] proposed a new method to purify protein interaction networks via gene expression and subcellular location information. Although these methods have improved prediction and reduced the effect of false interactions, they are often more complex and computationally

expensive to run. Additionally, their effectiveness depends on the availability and quality of the biological data.

This paper proposes a new framework that calculates the **membership degree of interacting protein strengths**, converts the network into a fuzzy network, and identifies crucial proteins. This is the core objective of our paper, which utilizes **fuzzy set theory** as a basis for constructing a fuzzy biological network model. Since each edge is linked to a membership degree, suitable definitions are proposed to measure centrality in fuzzy biological network graphs.

Four commonly used traditional centrality measures, degree [11], eigenvector [14], betweenness [13], and closeness centrality [12], are extended to determine the centrality of nodes. The extension considers that interaction strength is not a crisp value but a fuzzy one, where different values are associated with different membership degrees.

The efficacy of the proposed framework has been assessed on various real-world PPI networks, as different PPI networks can exhibit diverse topological features. The results show that the proposed measures perform well on all networks and demonstrate greater accuracy than traditional methods. The paper's key contribution is an approach that enables researchers to incorporate fuzzy properties **(membership degree)** into centrality calculations. The paper's specific contributions are summarized as follows:

- Development of membership function for calculating the interaction strength of each edge.
- Development of the extension of traditional centrality measures as fuzzy centrality measures.

The remaining sections are organized as follows: Section II describes the background knowledge, including the fundamental concepts of fuzzy logic and set theory. Section III introduces the proposed framework and algorithms used to calculate the strength of interacting edges and fuzzy centralities. The experimental datasets and findings are presented in Sections IV and V, respectively. Section VI discusses the limitations and deficiencies of the framework, while Section VII concludes the paper.

II. PRELIMINARIES

A. GRAPH AND ADJACENCY MATRIX

The mathematical notation for a graph is $G(\mathbb{N}, \mathscr{E})$, where $\mathbb{N}$ is a set of nodes $\{\mathbb{m}_1, \mathbb{m}_2, \mathbb{m}_3, \ldots, \mathbb{m}_n\}$ and $\mathscr{E}$ is a subset of $\mathbb{N} \times \mathbb{N}$ that denotes the edges of the graph. The presence of an edge $\varepsilon_{ij} \in \mathscr{E}$ denotes a connection between the nodes $\mathbb{m}_i$ and $\mathbb{m}_j$, commonly referred to as neighboring nodes. Fig. 1 depicts an example of a simple crisp graph G. The given graph comprises a vertex set $\mathbb{N}$ containing the elements $\{\delta, \zeta, \theta, \varphi, \text{and } \psi\}$, and an edge set $\mathscr{E}$ consisting of the edges $\{\varepsilon_{\delta\zeta}, \varepsilon_{\delta\theta}, \varepsilon_{\zeta\varphi}, \varepsilon_{\theta\varphi}, \text{and } \varepsilon_{\varphi\psi}\}$.

The adjacency matrix, A_d is a widely used graphical representation. It is a square matrix of size $\mathbb{N} \times \mathbb{N}$, where $\mathbb{N}$ is the number of nodes in the network. The (i, j) entry of the matrix A_d is 1 if there is an edge from node i to node j, and 0 otherwise. The term "adjacency matrix" is derived from the fact that it is constructed based on the adjacencies present in a graph [35]. The adjacency matrix of the graph present in Fig. 1 is:

$$\mathbf{A_d} = \begin{matrix} \delta \\ \zeta \\ \theta \\ \varphi \\ \varphi \end{matrix} \begin{bmatrix} \mathbf{0} & \mathbf{1} & \mathbf{1} & \mathbf{0} & \mathbf{0} \\ \mathbf{1} & \mathbf{0} & \mathbf{0} & \mathbf{1} & \mathbf{0} \\ \mathbf{1} & \mathbf{0} & \mathbf{0} & \mathbf{1} & \mathbf{0} \\ \mathbf{0} & \mathbf{1} & \mathbf{1} & \mathbf{0} & \mathbf{1} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{1} & \mathbf{0} \end{bmatrix} \tag{1}$$

B. CENTRALITY MEASURES

Centrality measures utilize a graph's topological properties to ascertain each node's significance [36]. Several metrics, including degree centrality [11], closeness centrality [12], eigenvector centrality [14], and betweenness centrality [13], are utilized to assess the vitality of nodes. Below are descriptions of these traditional and commonly employed centrality measures. Subsequently, in the manuscript, these metrics are expanded to assess the significance of nodes in a fuzzy biological network.

Definition 1: Degree centrality (DC)

The DC [11] quantifies the number of edges emanating from a network node. The formula for DC for node i is:

$$\mathbf{DC}(\boldsymbol{i}) = \sum_{j=1}^{n} \boldsymbol{a}(\boldsymbol{i}, \boldsymbol{j}) \tag{2}$$

where, $a(i, j)$ is the A_d element, representing the edge between node i and j.

Definition 2: Betweenness centrality (BC)

The BC [13] is a metric that quantifies the frequency with which a node serves as a bridge through the shortest path connecting two other nodes within a network. The formula for BC for node i is:

$$\mathbf{BC}(\boldsymbol{i}) = \sum_{m \neq n \neq i} \frac{\sigma_{\mathbf{mn}}(\boldsymbol{i})}{\sigma_{\mathbf{mn}}} \tag{3}$$

where, σ_{mn} represents the total count of shortest paths connecting nodes m and n, $\sigma_{\text{mn}}(i)$ denotes the quantity of shortest paths originating from node m and terminating at n while traveling via node i.

Definition 3: Closeness centrality (CC)

The CC [12] is a metric that quantifies the efficiency with which a node can establish connections with every other node within a network, taking into account the shortest path distances. The formula for CC for node i is:

$$\mathbf{CC}(\boldsymbol{i}) = \frac{1}{\sum_{j=1}^{n} \boldsymbol{l}(\boldsymbol{i}, \boldsymbol{j})} \tag{4}$$

where, $l(j, i)$ represents the length of the shortest distance from the beginning of node i to node j, and the sum is computed over every node j in the network.

Definition 4: Eigenvector centrality (EC)

EC [14] is a metric that determines a node's importance in a network based on the importance of its neighbours. The power iteration method is employed for its computation. The

formula for calculating eigenvector centrality is:

$$\mathbf{EC} = \frac{\boldsymbol{A_d} * \boldsymbol{v_k}}{\|\boldsymbol{A_d} * \boldsymbol{v_k}\|} \tag{5}$$

where, A_d is the network's adjacency matrix, v_k is the iteration vector. The algorithm begins with an initial vector v_0, then iteratively multiplies the vector by the adjacency matrix and normalizes it. This process continues until the normalized values converge and become the same.

C. FUNDAMENTALS OF FUZZY SET THEORY

Zadeh [37], [38] initially introduced the concept of fuzzy set theory as a means of addressing the challenges posed by vagueness and ambiguity in decision-making processes, from a mathematical perspective. Unlike crisp sets, fuzzy sets include partial membership qualities and respond better to realistic membership expressions used in real-world problems, such as modeling and analyzing human trafficking chains [39], computing protein similarities [40], and many others. The representation of a fuzzy set is denoted by the set $\tilde{Z}$, which is defined as $\tilde{Z} = \left\{\left(\mathrm{x}, \mu_{\tilde{Z}}(\mathrm{x})\right) \mid \mathrm{x} \in \mathrm{X}\right\}$. Here, X represents a set of elements, and $\mu_{\tilde{Z}}(\mathrm{x})$ is a membership function that assigns a membership degree, $\mu_{\tilde{Z}}(\mathrm{x}) \in [0, 1]$ to each element x in the set. The $\mu_{\tilde{Z}}(\mathrm{x})$ can be used to determine the degree of inclusion of an element x in a set. Specifically, when $\mu_{\tilde{Z}}(\mathrm{x}) = 0$, x is not a member of the set. In contrast, when $\mu_{\tilde{Z}}(\mathrm{x}) = 1$, x is entirely included in the set. For values of $0 < \mu_{\tilde{Z}}(\mathrm{x}) < 1$, x is partially included in the set. For any given fuzzy set $\tilde{Z}$, the cut $\tilde{Z}$ the cut Z is defined as the subset of elements whose degree of membership is equal to or greater than a particular threshold value α. The determination of the cut of $\tilde{Z}$, which is as follows:

$$\widetilde{\mathbf{Z}}^{(\boldsymbol{\alpha})}(\mathbf{x}) = \left\{\mathbf{x} \in \mathbf{X} \mid \mu_{\widetilde{\mathbf{Z}}}(\mathbf{x}) \geq \boldsymbol{\alpha}\right\} \tag{6}$$

The intersection of two distinct fuzzy sets, P and Q, on a given set X is denoted by the symbol P ∩ Q. The degree of membership for each element x in the intersection of sets P and Q is provided by the following expression:

$$\begin{aligned}\mu_{\boldsymbol{P}\cap\boldsymbol{Q}}(\mathbf{x}) &= \mu_{\boldsymbol{P}}(\mathbf{x}) \wedge \mu_{\boldsymbol{Q}}(\mathbf{x}) \\ &= \mathbf{min}\left\{\mu_{\boldsymbol{P}}(\mathbf{x}), \mu_{\boldsymbol{Q}}(\mathbf{x})\right\}\end{aligned} \tag{7}$$

Moreover, the union of sets P and Q is denoted as P ∪ Q. The degree of membership for each element x in the union of sets P and Q is provided by the following expression:

$$\begin{aligned}\mu_{\boldsymbol{P}\cup\boldsymbol{Q}}(\mathbf{x}) &= \mu_{\boldsymbol{P}}(\mathbf{x}) \vee \mu_{\boldsymbol{Q}}(\mathbf{x}) \\ &= \mathbf{max}\left\{\mu_{\boldsymbol{P}}(\mathbf{x}), \mu_{\boldsymbol{Q}}(\mathbf{x})\right\}\end{aligned} \tag{8}$$

D. MEMBERSHIP FUNCTION DEFINITIONS

There are various types of membership functions (MF); however, the most commonly used ones are triangular (TriMF), trapezoidal (TraMF), Gaussian (GauMF) functions [41].

A Triangular Membership Function (TriMF) is a function characterized via three parameters, namely $\{\mathrm{x}_1, \mathrm{x}_2, \mathrm{x}_3\}$.

$$\mathbf{triangle}\,(\mathbf{x}; \boldsymbol{x}_1, \boldsymbol{x}_2, \boldsymbol{x}_3) = \begin{cases} 0, & \boldsymbol{x} \leq \boldsymbol{x}_1 \\ \dfrac{\boldsymbol{x} - \boldsymbol{x}_1}{\boldsymbol{x}_2 - \boldsymbol{x}_1}, & \boldsymbol{x}_1 \leq \boldsymbol{x} \leq \boldsymbol{x}_2 \\ \dfrac{\boldsymbol{x}_3 - \boldsymbol{x}}{\boldsymbol{x}_3 - \boldsymbol{x}_2}, & \boldsymbol{x}_2 \leq \boldsymbol{x} \leq \boldsymbol{x}_3 \\ 0, & \boldsymbol{x}_3 \leq \boldsymbol{x} \end{cases} \tag{9}$$

where, $\mathrm{x}_1 < \mathrm{x}_2 < \mathrm{x}_3$. Fig. 2 (a) displays an example of a triangular-shaped MF defined by the triangle (x; 2, 5, 8).

A trapezoidal membership function (TraMF) is characterized by the following four variables, namely $\{\mathrm{x}_1, \mathrm{x}_2, \mathrm{x}_3, \mathrm{x}_4\}$.

$$\mathbf{trapezoidal}\,(\mathbf{x}; \boldsymbol{x}_1, \boldsymbol{x}_2, \boldsymbol{x}_3) = \begin{cases} 0, & \boldsymbol{x} \leq \boldsymbol{x}_1 \\ \dfrac{\boldsymbol{x} - \boldsymbol{x}_1}{\boldsymbol{x}_2 - \boldsymbol{x}_1}, & \boldsymbol{x}_1 \leq \boldsymbol{x} \leq \boldsymbol{x}_2 \\ 1, & \boldsymbol{x}_1 \leq \boldsymbol{x} \leq \boldsymbol{x}_3 \\ \dfrac{\boldsymbol{x}_4 - \boldsymbol{x}}{\boldsymbol{x}_3 - \boldsymbol{x}_2}, & \boldsymbol{x}_4 \leq \boldsymbol{x} \leq \boldsymbol{x}_3 \\ 0, & \boldsymbol{x}_4 \leq \boldsymbol{x} \end{cases} \tag{10}$$

The values of $\mathrm{x}_1, \mathrm{x}_2, \mathrm{x}_3$, and x_4 must follow the rule that x_1 is less than x_2 and x_3 is less than x_4. Fig. 2 (b) displays an example of a trapezoidal-shaped MF characterized by the trapezoid (x; 2, 4, 6, 8)

On the other hand, a Gaussian membership function (GauMF) is characterized by the variables: c and σ. c represents the mean, while σ represents the standard deviation. Fig. 2 (c) shows an illustration of a Gaussian-shaped MF defined by Gaussian (x; 2, 5).

$$\mathbf{gaussian}(\mathbf{x}; \mathbf{c}, \boldsymbol{\sigma}) = \boldsymbol{e}^{-\frac{1}{2}\left(\frac{\boldsymbol{x}-\boldsymbol{c}}{\boldsymbol{\sigma}}\right)^2} \tag{11}$$

E. FUZZY GRAPH

In a fuzzy graph represented as $G'_f\left(\mathbb{N}', \mathscr{E}'\right)$, the set of vertices is denoted as $\mathbb{N}' = \{\mathbb{m}_1, \mathbb{m}_2, \mathbb{m}_3, \ldots, \mathbb{m}_n\}$. The edges and their degree of membership to the set $\mathscr{E}'$ are denoted by $\mathscr{E}' = \left\{\left(\varepsilon_{\mathrm{ij}}, \mu_{\mathscr{E}}\left(\varepsilon_{\mathrm{ij}}\right)\right) \mid \mathbb{m}_{\mathrm{i}} \in \mathbb{N}', \mathbb{m}_{\mathrm{j}} \in \mathbb{N}'\right\}$ as a fuzzy set [27]. The degree of membership of an edge reflects the level of interaction or relationship between two vertices. Fig. 3 depicts a fuzzy graph that serves as a model for a biological network with five proteins, namely $\partial, \ell, и, \rho$, and ω. The edge set $\mathscr{E}$ comprising of $(\varepsilon_{\partial\ell}, 0.3)$, $(\varepsilon_{\partial и}, 0.4)$, $(\varepsilon_{\ell и}, 0.5)$, $\left(\varepsilon_{\ell\rho}, 0.7\right)$, $\left(\varepsilon_{и\rho}, 0.7\right)$, and $\left(\varepsilon_{\rho\omega}, 0.1\right)$, is utilized to model the interactions or relationships between the proteins. The membership degree of $\varepsilon_{\mathrm{ij}}$, denoted as $\mu_\delta\left(\varepsilon_{\mathrm{ij}}\right)$, represents the degree of interaction. The biological network graph presented in Fig. 3 illustrates that the interaction strength of all the edges in the network are not the same; some edges are more strongly connected than others. The membership degree of $\varepsilon_{\rho\omega}$ is 0.1, indicating that the interacting strength between proteins ρ and ω is 0.1, whereas the membership degree of the interaction between

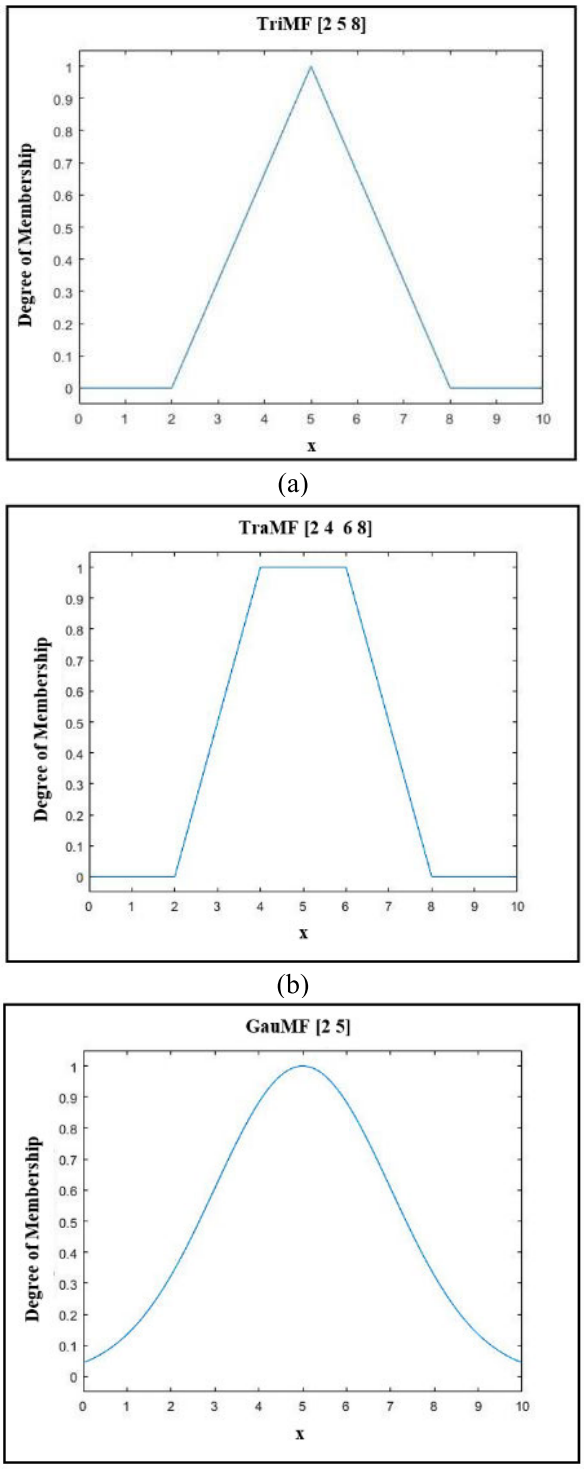


FIGURE 2. An example of membership functions (a) triangularshaped MF (b) trapezoidal-shaped MF (c) Gaussian-shaped MF.

proteins ℓ and $\imath$ is 0.5. Unlike traditional graphs, where all edges have the same interaction strength.

Theorem: The sum of the degrees of each vertex in a fuzzy graph equals two times the sum of the membership degree of all edges [42].

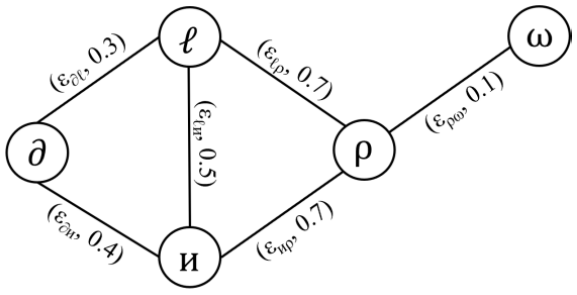


FIGURE 3. An illustration depicting example biological network graph with fuzzy connections.

Proof: Let $G'_f = \left(\mathbb{N}', \mathscr{E}'\right)$ be a fuzzy graph, where $\mathbb{N}' = \{\mathbb{m}_1, \mathbb{m}_2, \mathbb{m}_3, \ldots, \mathbb{m}_n\}$ represents the set of vertices and $\mathscr{E}' = \left\{\left(\varepsilon_{\mathrm{ij}},\ \mu_{\mathscr{E}}\left(\varepsilon_{\mathrm{ij}}\right)\right) \mid \mathbb{m}_{\mathrm{i}} \in \mathbb{N}', \mathbb{m}_{\mathrm{j}} \in \mathbb{N}'\right\}$ signifies the edges and their membership degree.

we know that;

$$\sum_{\mathbb{mI}\in\mathbb{N}} d(\mathbb{m}) = \sum_{k=1}^{n} d\left(\mathbb{m}_{\mathrm{K}}\right) = d\left(\mathbb{m}_1\right) + d\left(\mathbb{m}_2\right) + d\left(\mathbb{m}_3\right) + \ldots + d\left(\mathbb{m}_{\mathrm{N}}\right)$$

where, $d(\mathrm{n})$ = vertex's degree.

As per the definition of degree of a vertex,

$$\begin{aligned}
d(\mathbb{m}) &= \sum_{i\neq j}\left(\varepsilon_{ij}, \mu_{\varepsilon}\left(\varepsilon_{ij}\right)\right)\\
d\left(\mathbb{m}_1\right) &= \left(\varepsilon_{12}, \mu_{\varepsilon}\left(\varepsilon_{12}\right)\right) + \left(\varepsilon_{13}, \mu_{\varepsilon}\left(\varepsilon_{13}\right)\right)\\
&\quad + \left(\varepsilon_{14}, \mu_{\varepsilon}\left(\varepsilon_{14}\right)\right) + \ldots + \left(\varepsilon_{1n}, \mu_{\varepsilon}\left(\varepsilon_{1n}\right)\right)\\
d\left(\mathbb{m}_2\right) &= \left(\varepsilon_{21}, \mu_{\varepsilon}\left(\varepsilon_{21}\right)\right) + \left(\varepsilon_{23}, \mu_{\varepsilon}\left(\varepsilon_{23}\right)\right)\\
&\quad + \left(\varepsilon_{24}, \mu_{\varepsilon}\left(\varepsilon_{24}\right)\right) + \ldots + \left(\varepsilon_{2\mathrm{n}}, \mu_{\varepsilon}\left(\varepsilon_{2\mathrm{n}}\right)\right)\\
d\left(\mathbb{m}_3\right) &= \left(\varepsilon_{31}, \mu_{\varepsilon}\left(\varepsilon_{31}\right)\right) + \left(\varepsilon_{33}, \mu_{\varepsilon}\left(\varepsilon_{32}\right)\right)\\
&\quad + \left(\varepsilon_{34}, \mu_{\varepsilon}\left(\varepsilon_{34}\right)\right) + \ldots + \left(\varepsilon_{3\mathrm{n}}, \mu_{\varepsilon}\left(\varepsilon_{3\mathrm{n}}\right)\right)\\
&\ \ \vdots \qquad\qquad \vdots \qquad\qquad \vdots\\
d\left(\mathbb{m}_{\mathrm{n}}\right) &= \left(\varepsilon_{\mathrm{n}1}, \mu_{\varepsilon}\left(\varepsilon_{\mathrm{n}1}\right)\right) + \left(\varepsilon_{\mathrm{n}2}, \mu_{\varepsilon}\left(\varepsilon_{\mathrm{n}2}\right)\right)\\
&\quad + \left(\varepsilon_{\mathrm{n}3}\right)\right) + \ldots + \left(\varepsilon_{\mathrm{nn}-1}, \mu_{\varepsilon}\left(\varepsilon_{\mathrm{nn}-1}\right)\right)
\end{aligned}$$

Since each edge is connected to exactly two vertices, each edge gets counted twice, once at each end. Therefore,

$$\begin{aligned}
d\left(\mathbb{m}_1\right) + d\left(\mathbb{m}_2\right) + \ldots + d\left(\mathbb{m}_{\mathrm{n}}\right) &= 2\left[\left(\varepsilon_{12}, \mu_{\varepsilon}\left(\varepsilon_{12}\right)\right)\right.\\
&\quad \left. + \left(\varepsilon_{13}, \mu_{\varepsilon}\left(\varepsilon_{13}\right)\right) + \ldots + \left(\varepsilon_{\mathrm{n}-1\mathrm{n}}, \mu_{\varepsilon}\left(\varepsilon_{\mathrm{n}-1\mathrm{n}}\right)\right)\right]\\
\sum d(\mathbb{m}) &= \sum\nolimits_{i\neq j} 2\left(\varepsilon_{\mathrm{ij}}, \mu_{\varepsilon}\left(\varepsilon_{\mathrm{ij}}\right)\right).
\end{aligned}$$

Example: Consider a fuzzy graph, $G'_f = \left(\mathbb{N}', \mathscr{E}'\right)$ with a set of vertices, $\mathbb{N}' = \{\mathbb{m}_1, \mathbb{m}_2, \mathbb{m}_3\}$. The degree of each vertex is as follows: $d\left(\mathbb{m}_1\right) = 0.7$, $d\left(\mathbb{m}_2\right) = 0.8$ and $d\left(\mathbb{m}_3\right) = 0.9$. The membership values of the edges are $(\varepsilon_{12}, 0.3)$, $(\varepsilon_{23}, 0.4)$, and $(\varepsilon_{13}, 0.5)$. Here $\sum d(\mathbb{m}) = 2.4$ and $\sum\left(\varepsilon_{1j}, \mu_{\varepsilon}\left(\varepsilon_{1j}\right)\right) = 1.2$. Therefore, $\sum d(\mathbb{m}) = \sum_{i\neq j} 2\left(\varepsilon_{\mathrm{ij}}, \mu_{\varepsilon}\left(\varepsilon_{\mathrm{ij}}\right)\right)$.

F. FUZZY ADJACENCY MATRIX

If $G'_f\left(\mathbb{N}', \mathscr{E}'\right)$ is a fuzzy graph, with set of vertices $\mathbb{N}' = \{\mathbb{m}_1, \mathbb{m}_2, \mathbb{m}_3, \ldots, \mathbb{m}_n\}$ and $\mathscr{E}' = \left\{\left(\varepsilon_{\mathrm{ij}}, \mu_\varepsilon\left(\varepsilon_{\mathrm{ij}}\right)\right) \mid \mathbb{m}_{\mathrm{i}} \in \mathbb{N}', \mathbb{m}_j \in \mathbb{N}'\right\}$ represents the edges and their membership degree. The fuzzy adjacency matrix (F_{Ad}) of $G'_f\left(\mathbb{N}', \mathscr{E}'\right)$ is an $\mathbb{N}' \times \mathbb{N}'$ matrix defined as:

$$\boldsymbol{F_{Ad}} = \begin{cases} \mu_\varepsilon\left(\boldsymbol{\varepsilon_{ij}}\right), & \textit{if there is an edge between i and j} \\ \mathbf{0}, & \textit{otherwise} \end{cases} \tag{12}$$

When all the nonzero membership values, $\mu_\varepsilon\left(\varepsilon_{\mathrm{ij}}\right)$ in a fuzzy graph are equal to 1, the matrix of the fuzzy graph becomes the traditional adjacency matrix [43].

III. PROPOSED METHODOLOGY

In this section, to address the issues and limitations discussed in Section I, the new framework for crucial proteins identification is introduced and discussed in detail in ***A*** and demonstrated using a dummy researcher network in ***B***.

A. FRAMEWORK FOR IDENTIFYING CRUCIAL PROTEINS

One of the challenges in studying biological (PPI) networks is quantifying the strength of the interactions between the proteins. This is important because the strength of the interactions can affect the behavior of the network as a whole and make it hard to identify and comprehend a network's true crucial proteins. Traditionally, the strength of the interactions in biological networks are quantified using binary values, Fig. 4 (a). However, this approach is not always accurate, as the interactions in biological networks are often not binary. To address this issue, we propose a new framework, Fig. 4 (b), for identifying crucial proteins in PPI networks that takes into account the strength of protein interactions. The framework consists of three major steps:

Step 1: Assessing the strength of interactions between proteins using a proposed membership function, Bio Link Strength. This function assigns a degree of membership to the strength of edges in a network. Unlike existing membership functions such as triangular, trapezoidal, and Gaussian, which are not suitable for determining edge strength in PPI networks due to their inflexibility, Bio Link Strength considers the intricate dynamics of the network.

Step 2: Using the membership degrees obtained in Step 1, convert the crisp PPI network to a fuzzy PPI network. This means that each network edge now has a membership degree, which represents the intensity of the relationship. Algorithm 1 outlines the pseudocode for constructing a fuzzy biological (PPI) network.

Step 3: Identifying crucial proteins in the fuzzy PPI network using fuzzy centrality measures. These measures are extensions of traditional centrality measures and take into account the strength of protein interactions. The details are as follows.

1) BIO-LINK STRENGTH (μ_{BLS})

In this study, the proposed membership function is utilized, a method, to assess the membership degree of the strength of the edge between two interacting proteins. This method relies on the concept of shared neighbors [44], where a higher number of shared neighbors indicates a stronger interaction between two proteins. The computation of (μ_{BLS}) is determined using Eq. (13)

$$\boldsymbol{\mu_{\mathrm{BLS}}(i,j)} = \begin{cases} \dfrac{|\mathbb{P}(\boldsymbol{i}) \cap \mathbb{P}(\boldsymbol{j})|}{|\mathbb{P}(\boldsymbol{i}) \cup \mathbb{P}(\boldsymbol{j})|}, & \textit{if there is an edge between i and j} \\ \mathbf{0}, & \textit{otherwise} \end{cases} \tag{13}$$

where $|\mathbb{P}(i) \cap \mathbb{P}(j)|$ represents the common proteins that i and j share. On the other hand, $|\mathbb{P}(i) \cup \mathbb{P}(j)|$ refers to the entire set of distinct neighboring proteins connected to either protein i or j, or both. A higher membership degree score signifies a stronger interaction between two proteins, indicating a closer relationship and potential functional interdependence within the network.

Algorithm 1 Construction of a Fuzzy Biological (PPI) Network

Input: The crisp PPI network, $G = (\mathbb{N}, \mathscr{E})$
Output: The fuzzy PPI network, $G_{f'}\left(\mathbb{N}', \mathscr{E}'\right)$
where, $\mathscr{E}' = \left\{\left(\varepsilon_{ij}, \mu_{\mathscr{E}}\left(\varepsilon_{\mathrm{ij}}\right)\right)\right.$
1: **for** each edge (i, j) in $\mathscr{E}$ **do**
2: calculate the (μ_{BLS}) of each edge according to Eq. (13);
3: **if** there is an edge between i and j
4: **else** 0
5: **return** the fuzzy PPI network, $G'_f\left(\mathbb{N}', \mathscr{E}'\right)$

2) FUZZY CENTRALITY MEASURES

As per the proposed framework, there is a need to develop some metrics for identifying crucial proteins in fuzzy PPI networks; Fuzzy measures are proposed that include membership degree values for edge interaction strength by extending commonly used traditional centrality measures. The detail of each proposed centrality measure is defined as follows:

Definition 4: Fuzzy connectivity centrality (FCC)

It is a measure of the extent to which a node is connected to other nodes in the network, considering the strength of connections. It is a fuzzy version of degree centrality. The formula for FCC for node i is:

$$\mathbf{FCC}(\boldsymbol{i}) = \sum_{j=1}^{n} \boldsymbol{\mu(i,j)} \tag{14}$$

where, $\mu(i, j)$ is the element in the fuzzy adjacency matrix F_{Ad}, represents edge strength between node i and j.

Definition 5: Fuzzy bridge centrality (FBC)

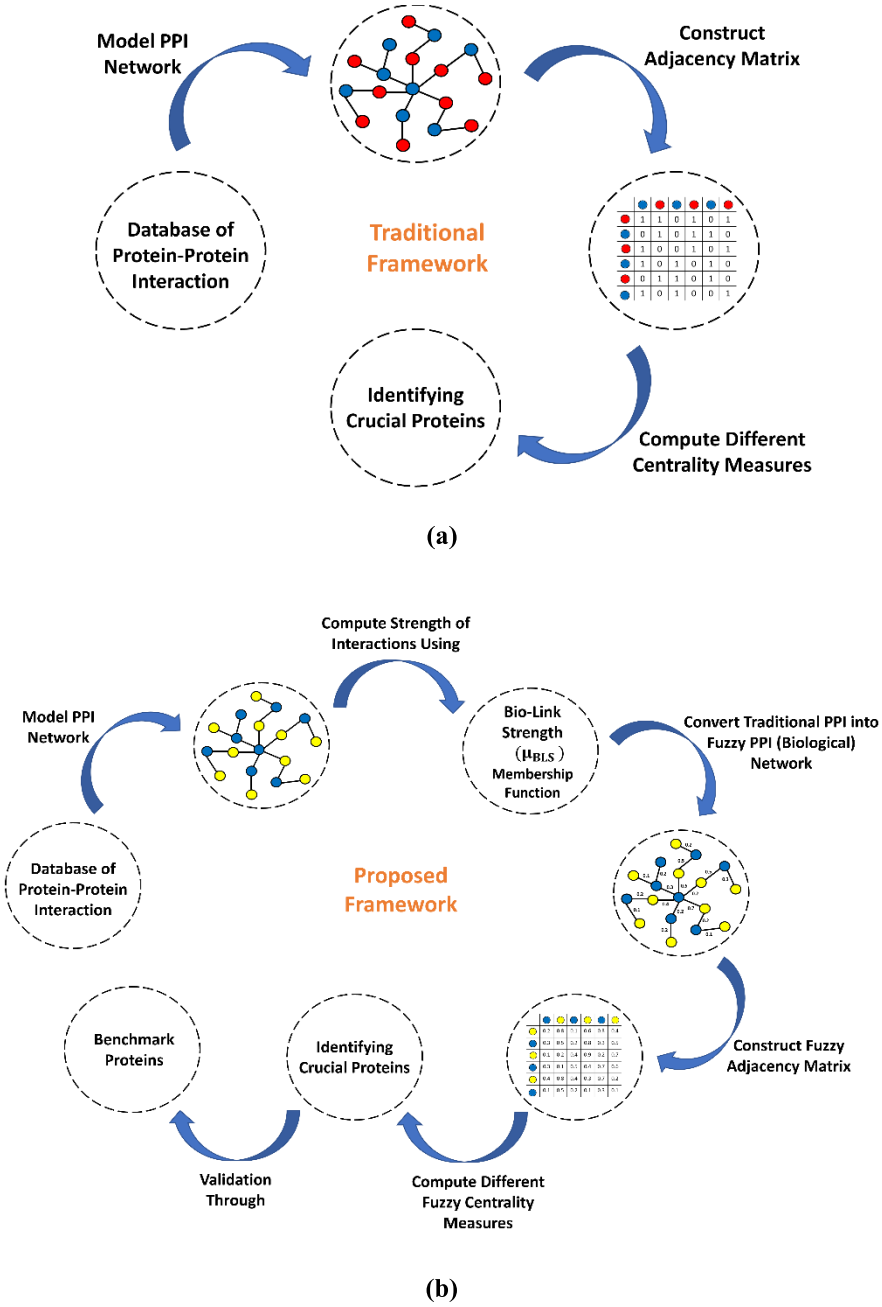


FIGURE 4. Illustration of the (a) traditional and (b) proposed framework.

FBC is a metric that measures how often a network node serves as a bridge across the shortest connection by considering the connecting strength between two other network nodes. It is a fuzzy version of betweenness centrality. The formula for calculating FBC for node i is:

$$\mathbf{FBC}(i) = \sum_{m \neq n \neq i} \frac{\delta_{\mathbf{mn}}(i)}{\delta_{\mathbf{mn}}} \tag{15}$$

where δ_{mn} represents the total count of shortest paths according to their interacting strength connecting nodes m and n, $\delta_{\mathrm{mn}}(i)$ denotes the quantity of shortest paths with their strengths originating from node m and terminating at n while traveling via node i.

Definition 6: Fuzzy approachability centrality (FAC)

The FAC is a metric that measures how fast a node can connect with all other nodes in a network, taking into account the shortest paths and their strength. It is a fuzzy version of closeness centrality. The formula for FAC for node i is:

$$\mathbf{FAC}(i) = \frac{1}{\sum_{j=1}^{n} s_I(i,j)} \tag{16}$$

where, s_l represents the strength of length of the shortest distance from the beginning of node i to node j.

Definition 7: Fuzzy influence centrality (FIC)

FIC is a metric that measures a node's impact on the network. based on the importance of its interacting nodes (neighbours). It is a fuzzy version of eigenvector centrality. The formula for calculating fuzzy influence centrality is:

$$\mathbf{FIC} = \frac{\boldsymbol{F_{Ad}} * \dot{\boldsymbol{v}}_{\boldsymbol{k}}}{\left\| \boldsymbol{F_{Ad}} * \dot{\boldsymbol{v}}_{\boldsymbol{k}} \right\|} \tag{17}$$

where F_{Ad} is the network's fuzzy adjacency matrix, v_k is the iteration vector. The algorithm starts with an initial vector $\dot{v}_0$, then iteratively multiplies the vector by the fuzzy adjacency matrix and normalizes it. This continues until the normalized values converge.

B. EXAMPLE EXPLANATION

To illustrate the proposed framework, consider a small dummy network of 8 researchers shown in Fig. 5 as an example.

Step 1 (Calculate Membership Degree for Edge Interaction Strength) According to Eq. (13) the membership degree of the strength of each edge is calculated:

$$\begin{aligned}
\varepsilon_{12} &= \text{edge between } \mathbb{m}_1 \text{ and } \mathbb{m}_2. \\
&= \frac{\{|\mathbb{m}_3|\}}{\{|\mathbb{m}_1, \mathbb{m}_2, \mathbb{m}_3, \mathbb{m}_4, \mathbb{m}_8|\}} = \frac{1}{5} = 0.2 \\
\varepsilon_{13} &= \text{edge between } \mathbb{m}_1 \text{ and } \mathbb{m}_3. \\
&= \frac{\{|\mathbb{m}_2|\}}{\{|\mathbb{m}_1, \mathbb{m}_2, \mathbb{m}_3, \mathbb{m}_4|\}} = \frac{1}{4} = 0.25
\end{aligned}$$

Similarly, the strength of remaining edges is calculated. The outcomes are shown in Table 1.

Step 2 (Construct Fuzzy Adjacency Matrix ($\boldsymbol{F_{Ad}}$)): Now construct a fuzzy adjacency matrix by using Eq. (12)

$$F_{Ad} = \begin{bmatrix}
0 & 0.2 & 0.25 & 0 & 0 & 0 & 0 & 0 \\
0.2 & 0 & 0.4 & 0.143 & 0 & 0 & 0 & 0 \\
0.25 & 0.4 & 0 & 0.167 & 0 & 0 & 0 & 0 \\
0 & 0.143 & 0.167 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0
\end{bmatrix}$$

Step 3 (Computed Proposed Fuzzy Centrality Metrics): Finally, values of each proposed fuzzy centrality measures are calculated by Eqs. (14) to (17), respectively. Values of each measure are shown in Table 2. For avoiding the calculation complexity in Fuzzy Bridge (FBC) and Accessibility (FAC) centrality we take edges of 0 strength as approximately 0.

Protein interaction networks are usually depicted as undirected graphs [19], where connections between proteins don't have a specific direction. Proposed framework is applicable to both undirected and directed networks, in the case of directed networks, the same steps can be followed.

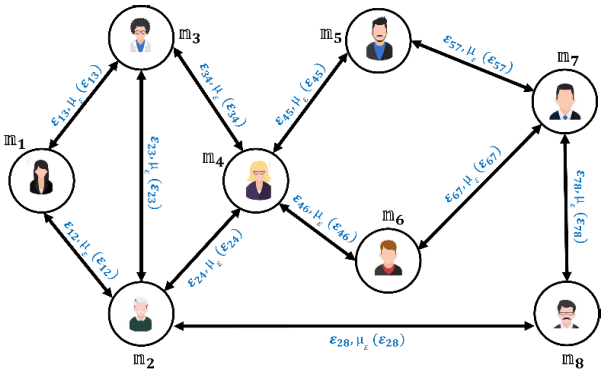


FIGURE 5. A small-scale network of 8 researchers (or 8 nodes).

TABLE 1. Membership degree of edges strength.

Edges	Connecting Nodes	Membership degree of the interacting strength
ε_{12}	Between $\mathbb{m}_1$and $\mathbb{m}_2$.	0.2
ε_{13}	Between $\mathbb{m}_1$and $\mathbb{m}_3$.	0.25
ε_{23}	Between $\mathbb{m}_2$and $\mathbb{m}_3$.	0.4
ε_{34}	Between $\mathbb{m}_3$and $\mathbb{m}_4$.	0.167
ε_{24}	Between $\mathbb{m}_2$and $\mathbb{m}_4$.	0.143
ε_{45}	Between $\mathbb{m}_4$and $\mathbb{m}_5$.	0.0
ε_{46}	Between $\mathbb{m}_4$and $\mathbb{m}_6$.	0.0
ε_{57}	Between $\mathbb{m}_5$and $\mathbb{m}_7$.	0.0
ε_{67}	Between $\mathbb{m}_6$and $\mathbb{m}_7$.	0.0
ε_{78}	Between $\mathbb{m}_7$and $\mathbb{m}_8$.	0.0
ε_{28}	Between $\mathbb{m}_2$and $\mathbb{m}_8$.	0.0

IV. EXPERIMENTAL ASSESMENT

A. TEST DATASETS

To evaluate the efficacy of the proposed framework, we applied it to Saccharomyces cerevisiae (Baker's yeast) and Escherichia coli (E. coli) to identify crucial proteins. These particular organisms were chosen as they have been extensively studied through experimental techniques and are commonly used as a benchmark in crucial protein evaluations. The protein-protein interaction data for this study was sourced from the Database of Interacting Proteins (DIP) [45]. The analyzed organisms, Saccharomyces cerevisiae, and Escherichia coli, initially contained 5,221 proteins and 24,918 interactions, and 2,994 proteins and 13,379 interactions, respectively, at the time of download (version 10.1.7.3). After filtering unnecessary (self and duplicate) interactions, a total of 24,743 interactions among 5,093 proteins in the Saccharomyces cerevisiae network and 11,803 interactions among 2,727 proteins in the E. coli network were included in our analysis. Some of the key traits of the organisms discussed in this study and the networks' protein-protein interaction (PPI) are shown in Fig. 6 (a and b).

B. BENCHMARK CRUCIAL PROTEIN LIST

A comprehensive list of crucial proteins in Saccharomyces cerevisiae and Escherichia coli was extracted from [31]. Based on the information obtained from the dataset, a total of

TABLE 2. Computed values of each fuzzy centrality metrics of researcher network.

Nodes	FCC	FBC	FAC	FIC
$\mathbb{m}_1$	0.450	0.000	0.689	0.428
$\mathbb{m}_2$	0.743	5.000	2.726	0.589
$\mathbb{m}_3$	0.817	0.000	0.799	0.617
$\mathbb{m}_4$	0.310	5.500	2.726	0.294
$\mathbb{m}_5$	0.000	3.500	2.726	0.000
$\mathbb{m}_6$	0.000	3.500	2.726	0.000
$\mathbb{m}_7$	0.000	11.500	2.726	0.000
$\mathbb{m}_8$	0.000	9.000	2.726	0.000

1,167 proteins out of the 5,093 proteins and 254 out of 2,727 proteins in the studied organisms were identified as crucial.

V. EXPERIMENTAL RESULTS AND DISCUSSION

A. VALIDATION OF PROPOSED MEMBERSHIP FUNCTION

In order to validate the effectiveness of the proposed membership function for evaluating edge strength in the PPI network, we utilized theorem, discussed in section II. The results presented in Table 3 demonstrated that the sum of the fuzzy connectivity (degree) of each protein in the saccharomyces cerevisiae and Escherichia coli network is twice the sum of the membership value of all edges. This suggests that the proposed function correctly evaluates the edge strength's membership degree.

B. COMPARING THE EFFECTIVENESS OF PROPOSED FUZZY TO TRADITIONAL CENTRALITIES

To evaluate the effectiveness of the proposed fuzzy centralities, namely Fuzzy Connectivity Centrality (FCC), Fuzzy Bridge Centrality (FBC), Fuzzy Approachability Centrality (FAC), and Fuzzy Influence Centrality (FIC), we compared their performance with that of traditional centrality measures: Degree (DC) [11], Betweenness (BC) [13], Closeness (CC) [12], and Eigenvector (EC).We ranked proteins based on their centrality values obtained from each measure and selected a particular number, $\dot{K}$, of top-ranking proteins as candidates for crucial proteins. The identified candidate proteins were then compared to benchmark proteins to determine the proportion of them were actually crucial proteins. Table 4 displays the comparison results obtained from the Saccharomyces cerevisiae and Escherichia coli PPI network.

The outcomes demonstrate that the proposed fuzzy centrality methods significantly outperform traditional centrality methods in predicting crucial proteins. For example, in the saccharomyces cerevisiae network, FCC and FIC identify nearly twice as many crucial proteins (66 and 55, respectively) at the top $\approx$ 2% as DC and EC (46 and 37, respectively). This pattern is consistent across all $\dot{K}$. These measures perform well in the E. coli network as well. Fig. 7 and Fig. 8 provide a more detailed representation of the performance of above discussed methods.

TABLE 3. Membership degree of edge strength in saccharomyces cerevisiae and escherichia coli network.

No. of Proteins	No. of Edges	Total fuzzy connectivity of each protein	Total membership value of all edges
Saccharomyces cerevisiae			
5093	24743	1,151.11	575.55
Escherichia coli			
2727	11803	720.23	360.11

C. ANALYSIS OF THE OVERLAPPING AND DIFFERENT TRUE CRUCIAL PROTEINS

To understand which proteins were identified by both the proposed and traditional methods and which were different, we looked at the top 100 ranked proteins. The results are shown in Table 5, where Tc represents one of these traditional methods (DC, CC, EC, and BC). $|* \cap \#|$ means the number of crucial proteins identified by both methods (intersection), while $|* - \#|$ means the number of crucial proteins found by method "*" instead of "#"(difference). Here, * and # are the arbitrary representation of the discussed measures. The results show that FBC has a higher number of overlapping crucial proteins with Tc compared to other methods in the Saccharomyces cerevisiae network. FCC and FIC, on the other hand, have a higher percentage of crucial proteins in their differences with Tc in both discussed organisms.

D. VALIDATION OF PROPOSED MEASURES BY SIX STATISTICAL METHODS

The validation and evaluation of the efficacy of any proposed method are commonly carried out using statistical methods, including negative predictive value (NPV), positive predictive value (PPV), accuracy, F-measure, specificity (S_P), and sensitivity (S_N) [46], [47]. In this paper, we validate our proposed fuzzy centrality measures using these measures.

The formulas for calculating these statistical measures are as follows:

Sensitivity: the ability to correctly identify true positives.

$$\textbf{\textit{Sensitivity}}\ (\boldsymbol{S_N}) = \frac{\boldsymbol{T_P}}{(\boldsymbol{T_P} + \boldsymbol{F_N})} \tag{18}$$

Specificity: the ability to correctly identify false negatives.

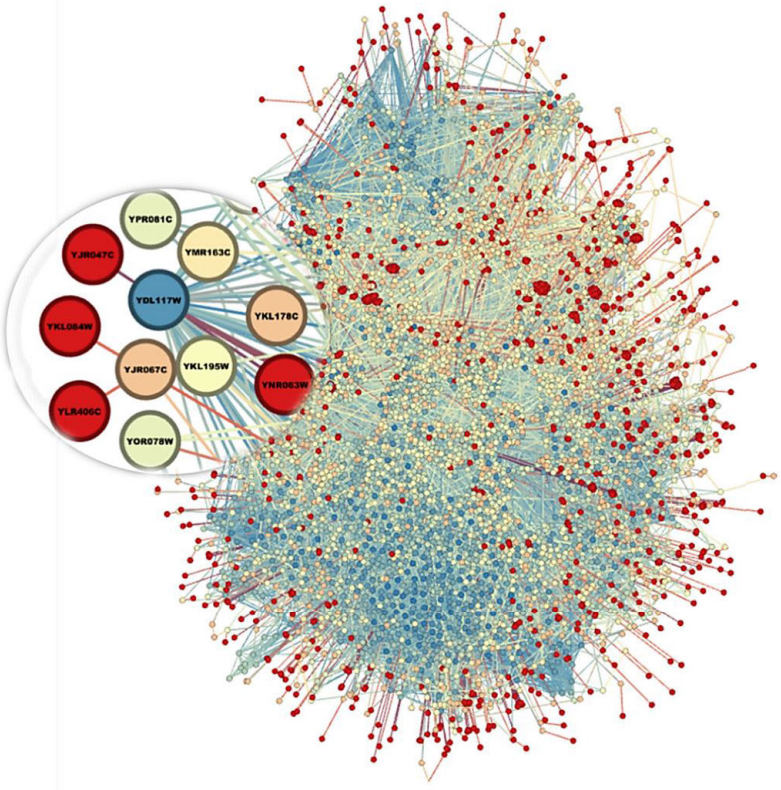
$$\textbf{\textit{Specificity}}\ (\boldsymbol{S_P}) = \frac{\boldsymbol{T_N}}{(\boldsymbol{T_N} + \boldsymbol{F_P})} \tag{19}$$

Positive Predictive Value: ratio of true positives out of all positive predictions.

$$\textbf{\textit{Positive Predicted Value}}\ (\boldsymbol{PPV}) = \frac{\boldsymbol{T_P}}{(\boldsymbol{T_P} + \boldsymbol{F_P})} \tag{20}$$

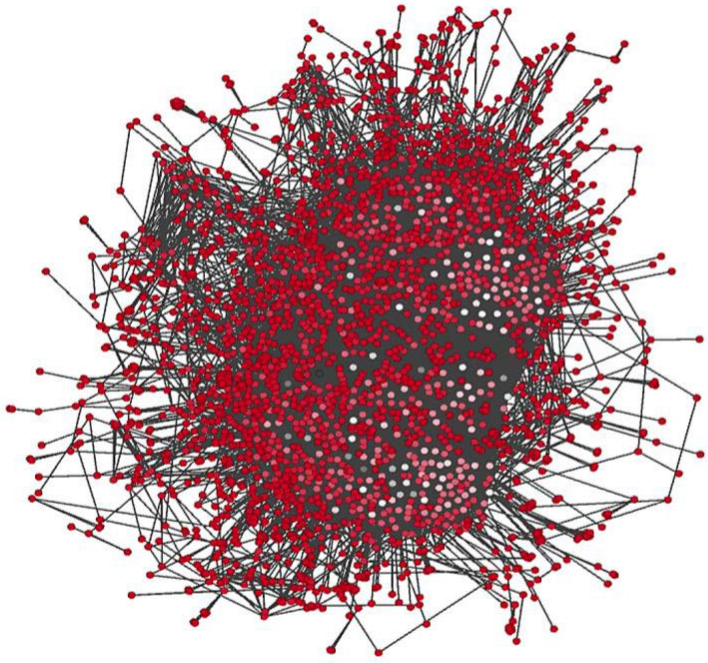
Negative predictive value: ratio of true negatives out of all negative predictions.

$$\textbf{\textit{Negative Predicted Value}}\ (\boldsymbol{NPV})\frac{\boldsymbol{T_N}}{(\boldsymbol{T_N} + \boldsymbol{F_N})} \tag{21}$$



Network Overview	
Nodes	5093
Edges	24743
Type	Undirected
Network Diameter	10
Average Path length	3.841029
Density	0.002
Modularity	0.497
Number of Communities	36

(a)



Network Overview	
Nodes	2727
Edges	11803
Type	Undirected
Network Diameter	12
Average Path length	3.985514
Density	0.003
Modularity	0.473
Number of Communities	89

(b)

FIGURE 6. A network of protein-protein interactions in the (a) Saccharomyces cerevisiae and (b) Escherichia coli, visualized using the Gephi software.

F-measure: harmonic mean of recall and precision.

$$F - Measure = \frac{2xS_N xPPV}{(S_N + PPV)} \tag{22}$$

Accuracy: ratio of accurate predictions to the overall number of predictions made.

$$Accuracy = \frac{T_P + T_N}{(T_P + T_N + F_P + F_N)} \tag{23}$$

where, the term T_P denotes proteins predicted to be crucial that are actually crucial, F_N represents proteins predicted to be non-crucial that are actually crucial, F_P represents proteins predicted to be crucial that are actually non-crucial and, T_N represents proteins predicted to be non-crucial that are actually non-crucial.

The results presented in Table 6 strongly support the precision of fuzzy connectivity (FCC) and fuzzy influence centrality (FIC). Each validation criterion for FCC and FIC in the Saccharomyces cerevisiae network has a significantly higher value than the other methods. FCC performs well on the Escherichia coli network as well. This suggests that FCC and FIC are more accurate than the other identification methods for crucial proteins.

E. VALIDATION BY GENE ONTOLOGY (GO)

To identify crucial proteins in biological networks, we selected the top 5 proteins from each centrality measure and analyzed them using the Gene Ontology (GO). We found that 4 of the top 5 proteins in fuzzy connectivity and fuzzy influence centrality was true crucial, indicating that these proposed approaches are highly significant for identifying crucial proteins in biological networks. Table 7 shows the results of our analysis. Using GO allowed us to systematically and accurately describe the concept of crucial proteins.

F. SPEARMAN CORRELATION WITH TRADITIONAL MEASURES

The Spearman rank correlation (ρ) is a statistical metric employed to evaluate the magnitude and orientation of the association between two variables or methodologies. It yields a numerical value within the range of -1 to 1 [48]. Our study employed this methodology to conduct a comparative analysis between proposed fuzzy and traditional centralities. A coefficient of 1 denotes a complete positive correlation, whereas a -1 signifies a complete negative correlation. A correlation coefficient of zero signifies the absence of any association between the centralities. The Spearman's rank correlation coefficients between fuzzy and traditional (crisp) methods on a network are computed by Eq. (24):

$$\rho = 1 - \frac{6\sum d_i^2}{n\left(n^2 - 1\right)} \tag{24}$$

The diagonal elements of the table 8 represent the correlations between the fuzzy and traditional versions of each centrality measure, which are all strongly positive and statistically significant. The off-diagonal elements represent the correlations between different centrality measures using the same method (either fuzzy or traditional). These correlations are also positive but generally weaker than the diagonal correlations. The results suggest that fuzzy centrality measures are highly correlated with their traditional counterparts and that different measures are positively correlated.

TABLE 4. Number of crucial proteins identified by proposed and traditional measures.

Methods	Top 100 (≈ 2%)	Top 200 (≈ 4%)	Top 300 (≈ 6%)	Top 400 (≈ 8%)	Top 500 (≈10%)	Top 600 (≈ 12%)
Saccharomyces cerevisiae						
FCC	**66**	**122**	**182**	**223**	**259**	**309**
DC	46	82	115	158	201	251
FIC	**55**	**99**	**143**	**187**	**231**	**276**
EC	37	77	119	158	192	221
FBC	41	71	101	139	178	212
BC	44	77	112	145	177	220
FAC	24	60	94	132	169	195
CC	29	64	107	139	175	217

Methods	Top 100	Top 1200	Top 2000
Escherichia coli			
FCC	**28**	**173**	**219**
DC	27	170	219
FIC	**25**	**169**	**223**
EC	22	169	220
FBC	**28**	**170**	**215**
BC	31	164	211
FAC	2	163	218
CC	2	168	219

G. COMPARING WITH OTHER RECENT TOPOLOGICAL AND BIOLOGICAL INFORMATION BASED METRICS

In above experiments, we compared the proposed measures with their traditional counterparts, validating the effectiveness of our approach. To further validate our method, we compared it with ten other previously proposed measures (Subgraph Centrality, Information Centrality, Bottle Neck, Density of Maximum Neighborhood Component, Range-Limited Centrality, L-Index, LeaderRank, Normalized α-Centrality, and Moduland-Centrality) mentioned in [29]. As shown in Table 9, our proposed fuzzy connectivity centrality (FCC) outperformed the aforementioned metrics in identifying crucial proteins within the yeast protein-protein interaction network, demonstrating a significant improvement over the results reported in [29].

In the top 100 rankings, FCC identified 66 crucial proteins, outperforming all other methods. SC identified 37, IC identified 44, BN identified 36, DMNC identified 55, RL identified 49, LI identified 41, LR identified 45, NC identified 37, and MC identified 39. In the top 200 rankings, FCC identified 122 crucial proteins, whereas SC, IC, BN, DMNC, RL, LI, LR, NC, and MC identified 77, 80, 76, 89, 80, 74, 82, 77, and 78 respectively. In the top 300 rankings, FCC identified 182 crucial proteins, whereas SC, IC, BN, DMNC, RL, LI, LR, NC, and MC identified 119, 118, 104, 136, 115, 123, 114, 120 and 124 respectively. In the top 400 rankings, FCC identified 223 crucial proteins, whereas SC, IC, BN, DMNC, RL, LI, LR, NC, and MC identified 158, 161, 145,

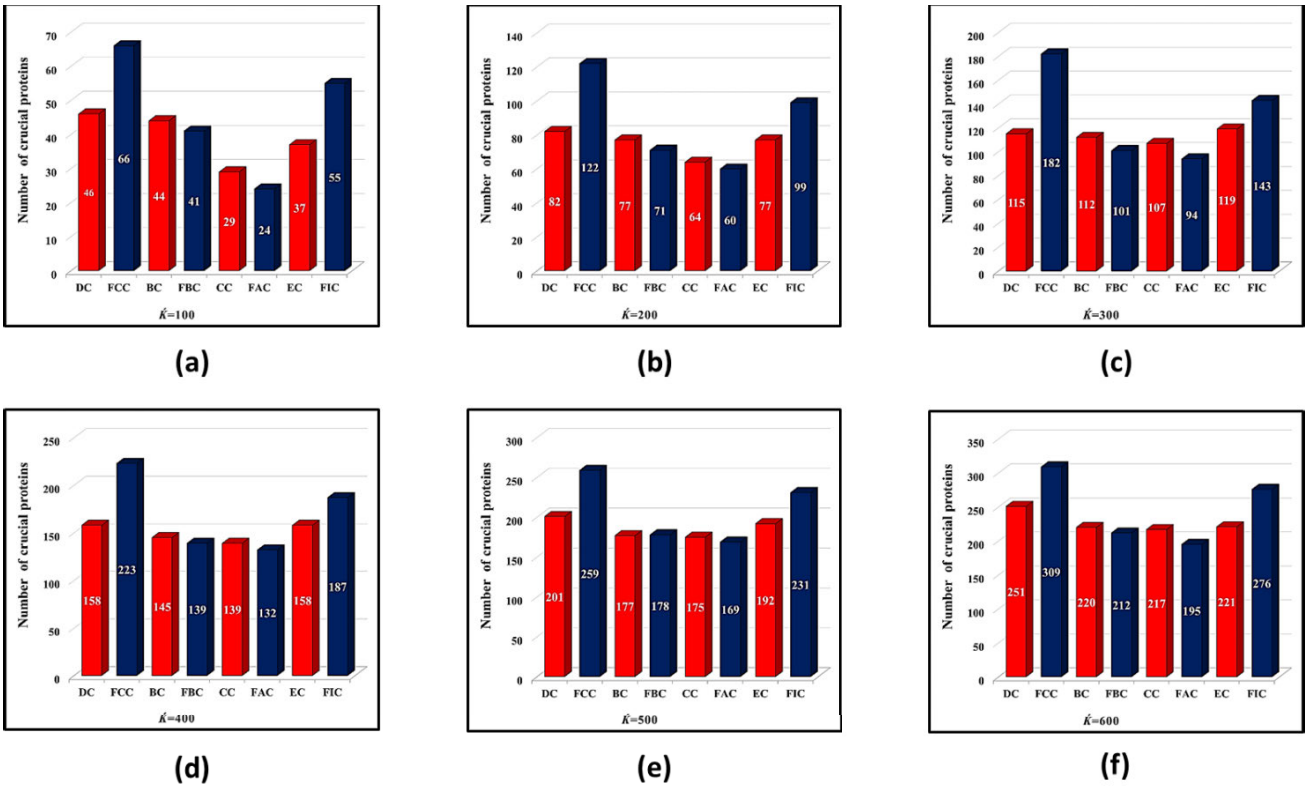


FIGURE 7. Comparison of the number of crucial proteins identified by proposed and traditional measures in Saccharomyces cerevisiae network (a) $\acute{K} = 100$ (b) $\acute{K} = 200$ (c) $\acute{K} = 300$ (d) $\acute{K} = 400$ (e) $\acute{K} = 500$ (f) $\acute{K} = 600$ top ranked proteins.

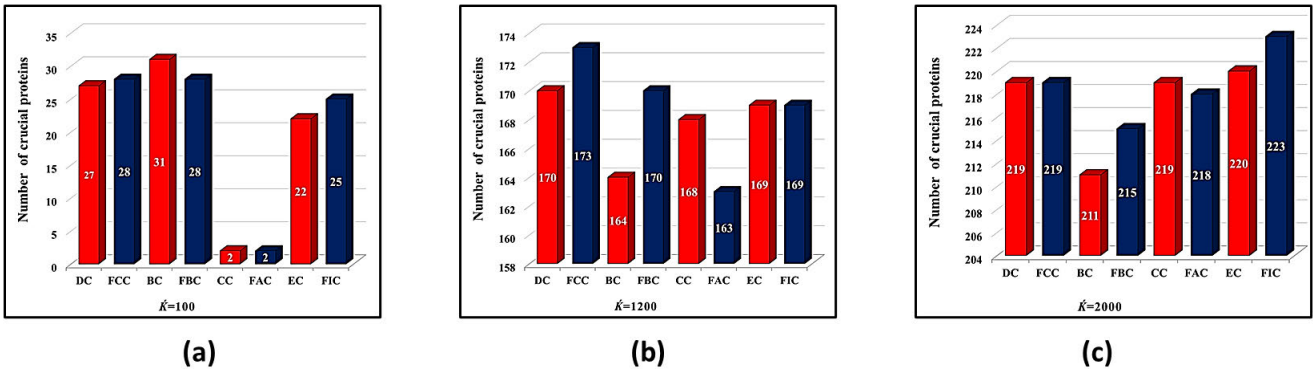


FIGURE 8. Comparison of the number of crucial proteins identified by proposed and traditional measures in Escherichia coli network (a) $\acute{K} = 100$ (b) $\acute{K} = 1200$ (c) $\acute{K} = 2000$ top ranked proteins.

182, 161, 158, 155, 157, and 166 respectively. In the top 500 rankings, FCC identified 259 crucial proteins, whereas SC, IC, BN, DMNC, RL, LI, LR, NC, and MC identified 192, 207, 175, 225, 190, 207, 204, 191, and 200 respectively. In the top 600 rankings, FCC identified 309 crucial proteins, whereas SC, IC, BN, DMNC, RL, LI, LR, NC, and MC identified 221, 251, 203, 265, 229, 256, 251, 221, and 247 respectively.

Moreover, our proposed FCC compared favorably with the recent PeC measure proposed by Li et al. [29], which integrates protein-protein interaction data and gene expression data. While PeC identified 328 crucial proteins in the top 600, our FCC identified 309, with a difference of less than 20 proteins. This highlights the significance of our proposed framework, which relies solely on topological properties and is easier to compute.

TABLE 5. Analysis of the overlapping and different true crucial proteins.

Centrality Measures (T_C)	FCC ∩T_C	Crucial proteins of T_C in $\|T_C - FCC\|$ (%)	Crucial proteins of FCC in $\|FCC - T_C\|$ (%)	FBC ∩T_C	Crucial proteins of T_C in $\|T_C - FBC\|$ (%)	Crucial proteins of FBC in $\|FBC - T_C\|$ (%)
Saccharomyces cerevisiae						
DC	7	84.78	89.39	32	30.43	21.95
BC	5	88.63	92.42	35	21.95	14.63
CC	6	79.31	90.90	18	37.93	56.09
EC	7	81.08	89.39	21	43.24	48.78
Escherichia coli						
DC	16	37.93	41.37	16	20.37	22.22
BC	15	28.07	22.80	21	30.3	21.21
CC	0	2.00	28.00	0	2.00	28.00
EC	13	34.69	57.69	11	16.66	25.75

Centrality Measures (T_C)	FAC ∩T_C	Crucial proteins of T_C in $\|T_C - FAC\|$ (%)	Crucial proteins of FAC in $\|FAC - T_C\|$ (%)	FIC ∩T_C	Crucial proteins of T_C in $\|T_C - FIC\|$ (%)	Crucial proteins of FIC in $\|FIC - T_C\|$ (%)
Saccharomyces cerevisiae						
DC	19	58.69	20.83	8	82.60	85.45
BC	18	50.09	25.00	7	84.09	87.27
CC	17	41.37	29.16	6	79.31	89.09
EC	18	51.35	25.00	11	70.27	80.00
Escherichia coli						
DC	0	27.00	2.00	17	34.48	27.58
BC	0	31.00	2.00	14	28.81	18.64
CC	2	0.00	0.00	0	2.00	25.00
EC	0	22.00	2.00	13	36.00	48.00

TABLE 6. Comparative analysis in terms of F-Measure, Accuracy, NPV, PPV, S_N, S_P between proposed and traditional measures.

Organism	Measures	S_N	S_P	PPV	NPV	F-Measure	Accuracy
Saccharomyces cerevisiae	FCC	**0.2648**	**0.9259**	**0.515**	**0.809**	**0.3497**	**0.7744**
	FIC	**0.2365**	**0.9175**	**0.46**	**0.8017**	**0.3124**	**0.7614**
	DC	0.2151	0.9111	0.4183	0.7961	0.2841	0.7516
	EC	0.1894	0.9035	0.3683	0.7895	0.2501	0.7398
	BC	0.1885	0.9032	0.3667	0.7892	0.249	0.7394
	CC	0.1859	0.9024	0.3617	0.7886	0.2456	0.7383
	FBC	0.1817	0.9012	0.3533	0.7874	0.24	0.7363
	FAC	0.1671	0.8968	0.325	0.7837	0.2207	0.7296
Organism	Measures	S_N	S_P	PPV	NPV	F-Measure	Accuracy
Escherichia coli	FCC	**0.6811**	**0.5847**	**0.1442**	**0.947**	**0.238**	**0.5937**
	DC	0.6693	0.5835	0.1417	0.945	0.2338	0.5915
	FBC	0.6693	0.5835	0.1417	0.945	0.2338	0.5915
	FIC	0.6654	0.5831	0.1408	0.9443	0.2325	0.5908
	EC	0.6654	0.5831	0.1408	0.9443	0.2325	0.5908
	CC	0.6614	0.5827	0.14	0.9437	0.2311	0.59
	BC	0.6457	0.5811	0.1367	0.9411	0.2256	0.5871
	FAC	0.6417	0.5807	0.1358	0.9404	0.2242	0.5864

AQ:4 **TABLE 7.** Gene ontology of top 5 ranked proteins of each measure.

Protein	Measure	Cruciality	Gene Ontology (GO)		
			Biological Process (BP)	Molecular Function (MF)	Cellular Component (CC)
YHR140W	FCC	×	No evidence	No evidence	Located in the ER (endoplasmic reticulum) [49]
YDR394W	FCC	✓	Promotes assembly of the RNA polymerase II transcription preinitiation complex [50]	Enables the activity of ATP hydrolysis [51]	A part of the proteasome regulatory particle [52]
YKR081C	FCC	✓	Involved in the assembly of large subunit precursor of pre-ribosome [53]	Facilitates the binding of 5S rRNA [54]	Located in nucleolus [55]
YDL097C	FCC, FIC	✓	Engaged in ubiquitin-dependent protein catabolic process [56]	Allows for structural molecule activity [56]	Found in proteasome storage granule [57]
YGR060W	FCC	✓	Implicated in the biosynthesis of ergosterol [58]	Facilitates C-4 methylsterol oxidase activity [58]	It is active in the endoplasmic reticulum membrane [59]
YJR091C	DC, BC, FBC	×	implicated in the localization of protein-containing complexes [60]	Facilitates mRNA binding [61]	Situated within the cytoplasm [62]
YBR160W	DC, BC	✓	Engaged in the processing of DNA double-strand breaks [63]	Facilitates the activity of protein serine/threonine kinases [64]	To be found in the endoplasmic reticulum [65]
YNL189W	DC, BC, FBC	✓	Engaged in the process of proteasome localization [66]	Facilitates the binding of protein-containing complexes [67]	Component of the NLS-dependent protein nuclear import complex [68]
YHR114W	DC, BC, FBC	×	Participating in actin nucleation [69]	Facilitates the binding of phospholipids [70]	Located in the plasma membrane [69]
YJR045C	DC, EC	✓	Engaged in the process of protein refolding [71]	Facilitates the activity of ATP hydrolysis [72]	Found in the mitochondrial nucleoid [73]
YMR047C	BC, FBC	✓	Engaged in the process of exporting mRNA from the nucleus [74]	Enables structural constituent of nuclear force [75]	Component of nuclear pore cytoplasmic filaments [76]
YLR295C	FBC	×	Engaged in the process of cristae formation [77]	Helps regulate the activity of ATP hydrolysis [78]	Component of mitochondrial proton-transporting ATP synthase complex [79]
YBL113C	CC, FAC	×	No evidence	No evidence	No evidence
YBR012C	CC, FAC	×	No evidence	No evidence	No evidence
YBR033W	CC, FAC	×	No evidence	Facilitates the binding of zinc ions [80]	No evidence
YBR056W	CC, FAC	×	No evidence	No evidence	Located in the intermembrane space of mitochondria [81]
YBR266C	CC, FAC	×	Participating in endocytosis [82]	No Evidence	No Evidence
YCL018W	EC	×	Participating in the glyoxylate cycle [83]	Facilitates the activity of 3-isopropylmalate dehydrogenase [84]	Within the cytosol [85]
YLR259C	EC	✓	Engaged in the maintenance of the mitochondrial genome [86]	Facilitates the binding of DNA replication origins [87]	Is functional in the mitochondrial nucleoid [86]
YBR127C	EC	×	Participating in pexophagy [88]	Facilitates the activity of the proton-transporting ATPase [89]	Within the vacuole-mitochondrion membrane contact site [90]
YML064C	EC	✓	Engaged in the facilitation of positive regulation of signalling [91]	Allows GTPase activity [92]	Seen in the spindle pole body [93]
YFR004W	FIC	✓	Engaged in the process of protein deubiquitination [94]	Facilitates the binding of zinc ions [80]	In mitochondrion [95]
YFR052W	FIC	✓	Engaged in the ubiquitin-dependent protein catabolic process [96]	No Evidence	Residing in the proteasome storage granule [57]
YGR232W	FIC	×	Engaged in the process of assembling the proteasome regulatory particle [97]	Facilitates the binding of the proteasome regulatory particle [98]	In nucleus [99]
YDL147W	FIC	✓	Engaged in the process of protein deneddylation [100]	Allows for protein binding [101]	Part of proteasome complex [100]

TABLE 8. Proposed and traditional centralities comparison using spearman's rank correlation coefficients.

ρ	FCC	FBC	FAC	FIC
Saccharomyces cerevisiae				
DC	**0.9300**	0.8667	0.7321	0.8098
BC	0.7854	**0.9682**	0.7001	0.6575
CC	0.7149	0.7354	**0.9477**	0.7872
EC	0.7399	0.7007	0.8498	**0.8752**
Escherichia coli				
DC	**0.9610**	0.9186	0.6137	0.7766
BC	0.8579	**0.9704**	0.5285	0.6573
CC	0.5739	0.5295	**0.9561**	0.6656
EC	0.7197	0.6777	0.6545	**0.9529**

TABLE 9. Number of crucial proteins identified by FCC and other recent metrics.

Methods	Top 100	Top 200	Top 300	Top 400	Top 500	Top 600
FCC	**66**	**122**	**182**	**223**	**259**	**309**
SC	36	77	119	158	192	221
IC	44	80	118	161	207	251
BN	36	76	104	145	175	203
DMNC	55	89	136	182	225	265
RL	49	80	115	161	190	229
LI	41	74	123	158	207	256
LR	45	82	114	155	204	251
NC	37	77	120	157	191	221
MC	39	88	124	166	200	247

H. DROSOPHILA MELANOGASTER (FRUIT FLY) PPI NETWORK (VALIDATION DATASET)

The proposed framework was also applied to the Drosophila melanogaster PPI network dataset mentioned in [102], which consists of 7,783 proteins (nodes) and 35,015 interactions (edges). Drosophila melanogaster, the fruit fly, is a widely used model organism in various biological disciplines, from fundamental genetics to tissue and organ development. Notably, the Drosophila genome shares 60% homology with the human genome, with fewer redundancies, and approximately 75% of human disease genes have homologs in flies [103].

To evaluate the top three proteins from each proposed centrality metric, we sourced biological essentiality data from UniProtKB. UniProt, the Universal Protein Resource, provides a comprehensive, freely accessible, and stable central resource on protein sequences and functional annotation. UniProt Knowledgebase (UniProtKB) is an expertly curated database that integrates protein information from multiple sources, offering a complete compendium of protein sequence data linked to functional information. UniProt integrates data from multiple resources, adding biological knowledge and metadata to protein records, and serves as a central hub linking to 180 other resources [104], [105].

The outcomes highlight the importance of the proposed metrics in identifying crucial proteins, demonstrating that their removal would significantly disrupt cellular function.

TABLE 10. Biological essentiality of top 3 ranked proteins.

Methods	Protein Name	Function	Ref.
FCC	Mec2	Enables protein binding (GO:0005515), involved in imaginal disc fusion, thorax closure (GO:0046529), involved in nephrocyte filtration (GO:0097206).	[106], [107], [108]
	Nucleoplasmin-like protein	Enables histone binding (GO:0042393), involved in chromatin remodeling (GO:0006338), involved in sperm DNA decondensation (GO:0035041).	[109]
	T-complex protein 1 subunit zeta	Involved in mitotic cell cycle (GO:0000278), involved in centriole replication (GO:0007099).	[110], [111]
FBC	LD27564p	No information	-
	p53 protein long form variant 1	Enables transcription cis-regulatory region binding (GO:0000976), enables RNA polymerase II general transcription initiation factor binding (GO:0001091), involved in regulation of DNA repair (GO:0006282).	[112], [113], [114]
	AT18705p	No information	-
	AT09807p	No information	-
FAC	Ephrin	Enables ephrin receptor binding (GO:0046875), involved in peripheral nervous system development (GO:0007422).	[115], [116]
	Lipid scramblase CLPTM1L	Located in endomembrane system (GO:0012505).	[117]
	Nucleoplasmin-like protein	Mentioned above	-
FIC	Calreticulin	Involved in olfactory behavior (GO:0042048), located in extracellular space (GO:0005615).	[118], [119]
	Histidine triad nucleotide binding protein 1, isoform B	No information	-

As shown in Table 10, the essentiality of proteins evaluated through mostly proposed centrality metrics aligns with their biological essentiality, supporting the value of this framework in understanding complex biological networks.

VI. LIMITATIONS AND DEFICIENCIES

A. EVALUATING FUZZY ACCESSIBILITY CENTRALITY PERFORMANCE USING GRAPHS

To investigate why fuzzy accessibility centrality did not perform better than the traditional methods or could not predict more crucial proteins, we randomly selected two noncrucial proteins with high values of fuzzy accessibility centrality. Network visualizations of these proteins are shown in Figs. 9 (a) and (b). These networks illustrate that although these proteins were more central in the network and provided

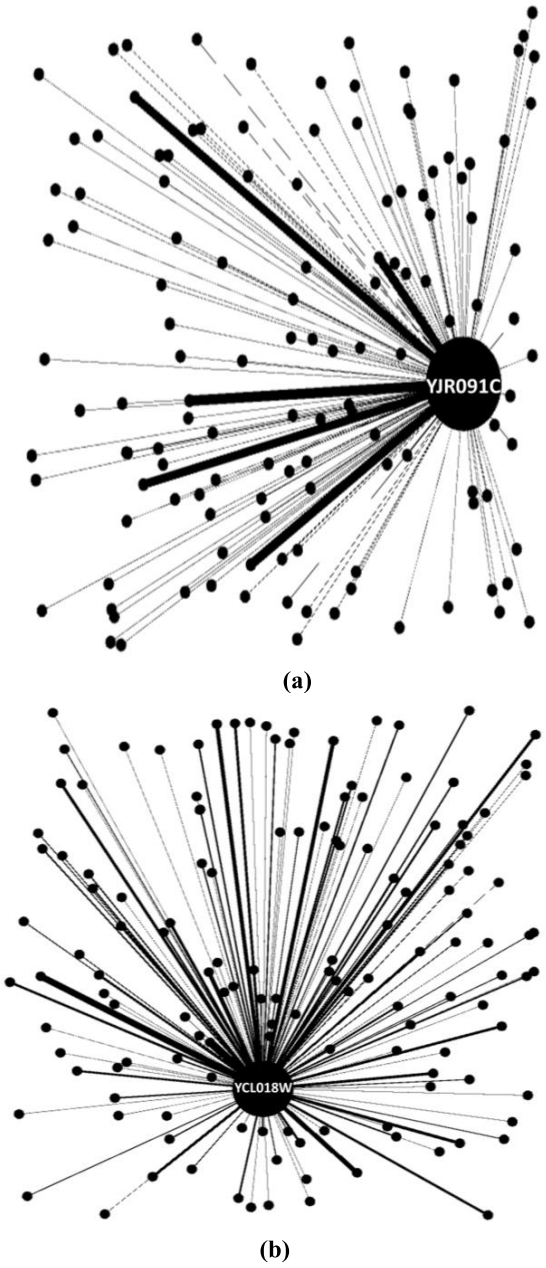


FIGURE 9. Protein-Protein Interaction (PPI) of (a) YJR091C and its interacting neighbours (b) YCL018W and its interacting neighbours.

shorter paths to other proteins, their neighboring proteins did not have higher strength degrees. This analysis suggests that the proposed fuzzy accessibility centrality method shares similarities with the traditional closeness centrality method.

B. EFFECT OF SCALE-FREE NETWORK PROPERTY

Biological networks are often characterized by a power-law (scale-free) node degree distribution, where numerous nodes have low degrees, while high-degree nodes (hubs) are relatively rare but still present. Scale-free networks exhibit distinct properties, including average degree and clustering

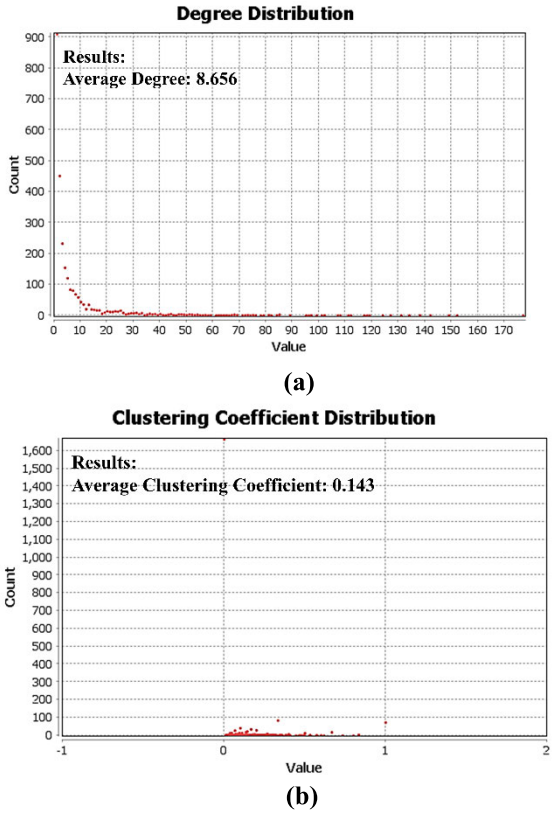


FIGURE 10. Network properties of the E-coli network (a) Average degree distribution (b) Clustering coefficient distribution.

coefficient [120]. In this study, we compute the membership degree of edge strength based on shared neighbors, which is influenced by the scale-free network property. To understand why our proposed fuzzy measures didn't outperform traditional measures in the E-coli network, we analyzed the network's properties using Gephi software. The results (Fig. 10 a and b) show a low average degree (8.656) and clustering coefficient (0.143), indicating potential scale-free characteristics.

VII. CONCLUSION AND FUTURE WORK

This paper discussed the pinpointing of crucial proteins in PPI networks. Most of the existing PPI network neglect the strength of the interacting proteins in a network, which make it difficult to find the network's true crucial proteins. A membership function (μ_{BLS}), for evaluating edge strength in a network was thus proposed in this paper. Also, extend four commonly used traditional measures into fuzzy ones to identify the true crucial proteins in the network. Considering the membership degrees of the interacting strength in their calculations. The proposed methods were evaluated on different real-world networks. By comparing proposed fuzzy measures with their traditional counterparts, the effectiveness of the proposed measures was validated. The importance of the top proteins identified by proposed fuzzy measures was also confirmed by gene ontology. Moreover, the correlation

between the proposed fuzzy and traditional measures was highly positive, indicating their accuracy. The best results were always provided by the proposed fuzzy connectivity and influence centrality methods among all proposed measures. Future work aims to address the limitations and deficiencies discussed and extend the application of the proposed measures to other biological networks.

DATA AND CODE AVAILABILITY

The processed dataset and source codes used in this study are available in https://github.com/abmoiz15/A-new-frameworkfor-pinpointing-crucial-proteins-in-protein-proteininteraction-networks-.git

REFERENCES

[1] M. Li, W. Li, F.-X. Wu, Y. Pan, and J. Wang, "Identifying essential proteins based on sub-network partition and prioritization by integrating subcellular localization information," *J. Theor. Biol.*, vol. 447, pp. 65–73, Jun. 2018, doi: 10.1016/j.jtbi.2018.03.029.

[2] M. Li, R. Zheng, H. Zhang, J. Wang, and Y. Pan, "Effective identification of essential proteins based on priori knowledge, network topology and gene expressions," *Methods*, vol. 67, no. 3, pp. 325–333, Jun. 2014, doi: 10.1016/j.ymeth.2014.02.016.

[3] J. Hou, Ed., "Chapter 7—Protein function prediction from functional connectivity," in *New Approaches of Protein Function Prediction from Protein Interaction Networks*. New York, NY, USA: Academic, 2017, pp. 97–105, doi: 10.1016/B978-0-12-809814-1.00007-8.

[4] S. Zhong, H. Zhang, and Y. Deng, "Identification of influential nodes in complex networks: A local degree dimension approach," *Inf. Sci.*, vol. 610, pp. 994–1009, Sep. 2022.

[5] P. G. Sun, Y. N. Quan, Q. G. Miao, and J. Chi, "Identifying influential genes in protein–protein interaction networks," *Inf. Sci.*, vols. 454–455, pp. 229–241, Jul. 2018, doi: 10.1016/j.ins.2018.04.078.

[6] N. Safari-Alighiarloo, M. Taghizadeh, M. Rezaei-Tavirani, B. Goliaei, and A. A. Peyvandi, "Protein–protein interaction networks (PPI) and complex diseases," *Gastroenterol. Hepatol. From Bed Bench*, vol. 7, no. 1, pp. 17–31, Winter 2014.

[7] K. Berahmand, E. Nasiri, R. Pir Mohammadiani, and Y. Li, "Spectral clustering on protein–protein interaction networks via constructing affinity matrix using attributed graph embedding," *Comput. Biol. Med.*, vol. 138, Nov. 2021, Art. no. 104933.

[8] J. E. Dickerson, A. Zhu, D. L. Robertson, and K. E. Hentges, "Defining the role of essential genes in human disease," *PLoS One*, vol. 6, no. 11, Nov. 2011, Art. no. e27368, doi: 10.1371/journal.pone.0027368.

[9] H. Jeong, S. P. Mason, A.-L. Barabási, and Z. N. Oltvai, "Lethality and centrality in protein networks," *Nature*, vol. 411, no. 6833, pp. 41–42, May 2001, doi: 10.1038/35075138.

[10] A. Özgur, T. Vu, G. Erkan, and D. R. Radev, "Identifying gene-disease associations using centrality on a literature mined gene-interaction network," *Bioinformatics*, vol. 24, no. 13, pp. i277–i285, Jul. 2008, doi: 10.1093/bioinformatics/btn182.

[11] S. Wuchty and P. F. Stadler, "Centers of complex networks," *J. Theor. Biol.*, vol. 223, no. 1, pp. 45–53, Jul. 2003.

[12] L. C. Freeman, *Lausanne-Printed*. Amsterdam, The Netherlands: Elsevier, 1978.

[13] M. E. J. Newman, "A measure of betweenness centrality based on random walks," *Social Netw.*, vol. 27, no. 1, pp. 39–54, Jan. 2005, doi: 10.1016/j.socnet.2004.11.009.

[14] P. Bonacich, "Some unique properties of eigenvector centrality," *Social Netw.*, vol. 29, no. 4, pp. 555–564, Oct. 2007, doi: 10.1016/j.socnet.2007.04.002.

[15] 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, pp. 1–12, May 2005, doi: 10.1103/physreve.71.056103.

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

[17] M. Ashtiani, A. Salehzadeh-Yazdi, Z. Razaghi-Moghadam, H. Hennig, O. Wolkenhauer, M. Mirzaie, and M. Jafari, "A systematic survey of centrality measures for protein–protein interaction networks," *BMC Syst. Biol.*, vol. 12, no. 1, pp. 1–26, Jul. 2018, doi: 10.1186/s12918-018-0598-2.

[18] M. Li, Y. Lu, J. Wang, F.-X. Wu, and Y. Pan, "A topology potential-based method for identifying essential proteins from PPI networks," *IEEE/ACM Trans. Comput. Biol. Bioinf.*, vol. 12, no. 2, pp. 372–383, Mar. 2015, doi: 10.1109/TCBB.2014.2361350.

[19] F. Zaidi, U. Fatima, B. Usmani, and A. Jafri, "Comprehending nodes essentiality through centrality measures in biological networks," *IJCSNS*, vol. 19, no. 9, pp. 1–23, 2019.

[20] U. Fatima, S. Hina, and M. Wasif, "A novel global clustering coefficient-dependent degree centrality (GCCDC) metric for large network analysis using real-world datasets," *J. Comput. Sci.*, vol. 70, Jun. 2023, Art. no. 102008, doi: 10.1016/j.jocs.2023.102008.

[21] S. Asur, D. Ucar, and S. Parthasarathy, *An Ensemble Framework for Clustering Protein-protein Interaction Networks*. London, U.K.: Oxford Univ. Press, Jul. 2007, doi: 10.1093/bioinformatics/btm212.

[22] J. Chen, W. Hsu, M. L. Lee, and S.-K. Ng, "Increasing confidence of protein interactomes using network topological metrics," *Bioinformatics*, vol. 22, no. 16, pp. 1998–2004, Aug. 2006, doi: 10.1093/bioinformatics/btl335.

[23] N. M. Ahmed, L. Chen, B. Li, W. Liu, and C. Dai, "A random walk-based method for detecting essential proteins by integrating the topological and biological features of PPI network," *Soft Comput.*, vol. 25, no. 14, pp. 8883–8903, Jul. 2021, doi: 10.1007/s00500-021-05780-8.

[24] M. Li, J.-X. Wang, H. Wang, and Y. Pan, "Identification of essential proteins from weighted protein–protein interaction networks," *J. Bioinf. Comput. Biol.*, vol. 11, no. 3, Jun. 2013, Art. no. 1341002, doi: 10.1142/s0219720013410023.

[25] W. Peng, J. Wang, W. Wang, Q. Liu, F.-X. Wu, and Y. Pan, "Iteration method for predicting essential proteins based on orthology and protein–protein interaction networks," *BMC Syst. Biol.*, vol. 6, no. 1, pp. 1–32, Dec. 2012.

[26] Z. Meng, L. Kuang, Z. Chen, Z. Zhang, Y. Tan, X. Li, and L. Wang, "Method for essential protein prediction based on a novel weighted protein-domain interaction network," *Frontiers Genet.*, vol. 12, pp. 1–28, Mar. 2021, doi: 10.3389/fgene.2021.645932.

[27] Y.-A. Huang, Z.-H. You, X. Li, X. Chen, P. Hu, S. Li, and X. Luo, "Construction of reliable protein–protein interaction networks using weighted sparse representation based classifier with pseudo substitution matrix representation features," *Neurocomputing*, vol. 218, pp. 131–138, Dec. 2016, doi: 10.1016/j.neucom.2016.08.063.

[28] L. Chen, Y.-H. Zhang, S. Wang, Y. Zhang, T. Huang, and Y.-D. Cai, "Prediction and analysis of essential genes using the enrichments of gene ontology and KEGG pathways," *PLoS ONE*, vol. 12, no. 9, Sep. 2017, Art. no. e0184129, doi: 10.1371/journal.pone.0184129.

[29] M. Li, H. Zhang, J.-X. Wang, and Y. Pan, "A new essential protein discovery method based on the integration of protein–protein interaction and gene expression data," *BMC Syst. Biol.*, vol. 6, no. 1, p. 15, Mar. 2012, doi: 10.1186/1752-0509-6-15.

[30] K. Berahmand, A. Bouyer, and N. Samadi, "A new centrality measure based on the negative and positive effects of clustering coefficient for identifying influential spreaders in complex networks," *Chaos, Solitons Fractals*, vol. 110, pp. 41–54, May 2018, doi: 10.1016/j.chaos.2018.03.014.

[31] 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, pp. 1–15, Dec. 2021, doi: 10.1186/s12859-021-04175-8.

[32] M. Li, Z. Niu, X. Chen, P. Zhong, F. Wu, and Y. Pan, "A reliable neighbor-based method for identifying essential proteins by integrating gene expressions, orthology, and subcellular localization information," *Tsinghua Sci. Technol.*, vol. 21, no. 6, pp. 668–677, Dec. 2016.

[33] K. Berahmand, S. Bahadori, M. N. Abadeh, Y. Li, and Y. Xu, "SDAC-DA: Semi-supervised deep attributed clustering using dual autoencoder," *IEEE Trans. Knowl. Data Eng.*, vol. 2, no. 1, pp. 1–10, Apr. 2024, doi: 10.1109/tkde.2024.3389049.

[34] M. Li, X. Chen, P. Ni, J. Wang, and Y. Pan, "Identifying essential proteins by purifying protein interaction networks," in *Iartificial Intelligence*. Cham, Switzerland: Springer, 2016, pp. 106–116.

[35] L. Metcalf and W. Casey, "Graph theory," in *Cybersecurity and Applied Mathematics*. Amsterdam, The Netherlands: Elsevier, 2016, pp. 67–94, doi: 10.1016/B978-0-12-804452-0.00005-1.
[36] A. Zareie and R. Sakellariou, "Centrality measures in fuzzy social networks," *Inf. Syst.*, vol. 114, Mar. 2023, Art. no. 102179, doi: 10.1016/j.is.2023.102179.
[37] L. A. Zadeh, "Fuzzy sets and systems," *Int. J. Gen. Syst.*, vol. 17, nos. 2–3, pp. 129–138, Jun. 1990, doi: 10.1080/03081079008935104.
[38] L. A. Zadeh, "Fuzzy sets," *Inf. Control*, vol. 8, no. 3, pp. 338–353, 1965, doi: 10.1016/S0019-9958(65)90241-X.
[39] A. Sebastian, S. Mathew, and J. N. Mordeson, "A new fuzzy graph parameter for the comparison of human trafficking chains," *Fuzzy Sets Syst.*, vol. 450, pp. 27–46, Dec. 2022, doi: 10.1016/j.fss.2022.04.016.
[40] D. Pelta, N. Krasnogor, C. Bousono-Calzon, J. L. Verdegay, J. Hirst, and E. Burke, "A fuzzy sets based generalization of contact maps for the overlap of protein structures," *Fuzzy Sets Syst.*, vol. 152, no. 1, pp. 103–123, May 2005, doi: 10.1016/j.fss.2004.10.017.
[41] O. Adil, A. Ali, M. Ali, A. Y. Ali, and B. S. Sumait, "Comparison between the effects of different types of membership functions on fuzzy logic controller performance the applications of E-systems in developing countries view project comparison between the effects of different types of membership functions on fuzzy logic controller performance," *Int. J. Emerg. Eng. Res. Technol.*, vol. 3, p. 76, Jun. 2015.
[42] A. N. Gani and M. B. Ahamed, "Order and size in fuzzy graphs," *Bull. Pure Appl. Sci.*, vol. 22, no. 1, pp. 145–148, 2003.
[43] S. Mathew. (2020). *Energy of a Fuzzy Graph*. [Online]. Available: http://www.afmi.or.kr
[44] X. Peng, J. Wang, W. Peng, F.-X. Wu, and Y. Pan, "Protein–protein interactions: Detection, reliability assessment and applications," *Briefings Bioinform.*, vol. 18, no. 5, pp. 798–819, Sep. 2017, doi: 10.1093/bib/bbw066.
[45] *DIP: STAT: Database*. Accessed: Jul. 2, 2023. [Online]. Available: https://dip.doe-mbi.ucla.edu/dip/Stat.cgi?SM=0
[46] A. Elahi and S. M. Babamir, "Identification of essential proteins based on a new combination of topological and biological features in weighted protein–protein interaction networks," *IET Syst. Biol.*, vol. 12, no. 6, pp. 247–257, Dec. 2018, doi: 10.1049/iet-syb.2018.5024.
[47] J. Wang, M. Li, H. Wang, and Y. Pan, "Identification of essential proteins based on edge clustering coefficient," *IEEE/ACM Trans. Comput. Biol. Bioinf.*, vol. 9, no. 4, pp. 1070–1080, Jul. 2012, doi: 10.1109/TCBB.2011.147.
[48] P. Sedgwick, "Spearman's rank correlation coefficient," *BMJ*, p. g7327, Nov. 28, 2014, doi: 10.1136/bmj.g7327.
[49] H. S. Garapati, G. Male, and K. Mishra, "Predicting subcellular localization of proteins using protein–protein interaction data," *Genomics*, vol. 112, no. 3, pp. 2361–2368, May 2020, doi: 10.1016/j.ygeno.2020.01.007.
[50] S. Malik, A. Shukla, P. Sen, and S. R. Bhaumik, "The 19 s proteasome subcomplex establishes a specific protein interaction network at the promoter for stimulated transcriptional initiation in vivo," *J. Biol. Chem.*, vol. 284, no. 51, pp. 35714–35724, Dec. 2009, doi: 10.1074/jbc.m109.035709.
[51] M. H. Glickman, D. M. Rubin, V. A. Fried, and D. Finley, "The regulatory particle of the *Saccharomyces cerevisiae* proteasome," *Mol. Cellular Biol.*, vol. 18, no. 6, pp. 3149–3162, Jun. 1998, doi: 10.1128/MCB.18.6.3149.
[52] H. Fu, "Subunit interaction maps for the regulatory particle of the 26S proteasome and the COP9 signalosome," *EMBO J.*, vol. 20, no. 24, pp. 7096–7107, Dec. 2001, doi: 10.1093/emboj/20.24.7096.
[53] J. Talkish, J. Zhang, J. Jakovljevic, E. W. Horsey, and J. L. Woolford, "Hierarchical recruitment into nascent ribosomes of assembly factors required for 27SB pre-rRNA processing in Saccharomyces cerevisiae," *Nucleic Acids Res.*, vol. 40, no. 17, pp. 8646–8661, Sep. 2012, doi: 10.1093/nar/gks609.
[54] N. Asano, K. Kato, A. Nakamura, K. Komoda, I. Tanaka, and M. Yao, "Structural and functional analysis of the Rpf2-Rrs1 complex in ribosome biogenesis," *Nucleic Acids Res.*, vol. 43, no. 9, pp. 4746–4757, May 2015, doi: 10.1093/nar/gkv305.
[55] K. A. Wehner and S. J. Baserga, "The sigma(70)-like motif: A eukaryotic RNA binding domain unique to a superfamily of proteins required for ribosome biogenesis," *Mol. Cell*, vol. 9, no. 2, pp. 329–339, Feb. 2002, doi: 10.1016/s1097-2765(02)00438-0.
[56] P. G. Santamaría, D. Finley, J. P. G. Ballesta, and M. Remacha, "Rpn6p, a proteasome subunit from Saccharomyces cerevisiae, is essential for the assembly and activity of the 26 S proteasome," *J. Biol. Chem.*, vol. 278, no. 9, pp. 6687–6695, Feb. 2003, doi: 10.1074/jbc.m209420200.
[57] D. Laporte, B. Salin, B. Daignan-Fornier, and I. Sagot, "Reversible cytoplasmic localization of the proteasome in quiescent yeast cells," *J. Cell Biol.*, vol. 181, no. 5, pp. 737–745, Jun. 2008, doi: 10.1083/jcb.200711154.
[58] S. Darnet, "Enzymological properties of sterol-C4-methyl-oxidase of yeast sterol biosynthesis," *Biochimica Et Biophysica Acta Mol. Cell Biol. Lipids*, vol. 1633, no. 2, pp. 106–117, Jul. 2003.
[59] C. Mo, M. Valachovic, S. K. Randall, J. T. Nickels, and M. Bard. (1996). *Protein–protein Interactions Among C-4 Demethylation Enzymes Involved in Yeast Sterol Biosynthesis*. [Online]. Available: https://www.pnas.orgcgidoi10.1073pnas.112202799
[60] K. L. Fehrenbacher, I. R. Boldogh, and L. A. Pon, "A role for Jsn1p in recruiting the Arp2/3 complex to mitochondria in budding yeast," *Mol Biol Cell*, vol. 16, pp. 5094–5102, Jun. 2005.
[61] Y. Yosefzon, Y. Y. Koh, J. J. Chritton, A. Lande, L. Leibovich, L. Barziv, C. Petzold, Z. Yakhini, Y. Mandel-Gutfreund, M. Wickens, and Y. Arava, "Divergent RNA binding specificity of yeast Puf2p," *RNA*, vol. 17, no. 8, pp. 1479–1488, Aug. 2011, doi: 10.1261/rna.2700311.
[62] S. F. Mitchell, S. Jain, M. She, and R. Parker, "Global analysis of yeast mRNPs," *Nature Structural Mol. Biol.*, vol. 20, no. 1, pp. 127–133, Jan. 2013, doi: 10.1038/nsmb.2468.
[63] K. Matsuzaki, M. Terasawa, D. Iwasaki, M. Higashide, and M. Shinohara, "Cyclin-dependent kinase-dependent phosphorylation of Lif1 and Sae2 controls imprecise nonhomologous end joining accompanied by double-strand break resection," *Genes Cells*, vol. 17, no. 6, pp. 473–493, Jun. 2012, doi: 10.1111/j.1365-2443.2012.01602.x.
[64] T. Zimniak, V. Fitz, H. Zhou, F. Lampert, S. Opravil, K. Mechtler, P. Stolt-Bergner, and S. Westermann, "Spatiotemporal regulation of Ipl1/Aurora activity by direct Cdk1 phosphorylation," *Current Biol.*, vol. 22, no. 9, pp. 787–793, May 2012, doi: 10.1016/j.cub.2012.03.007.
[65] E. Vergés, N. Colomina, E. Garí, C. Gallego, and M. Aldea, "Cyclin Cln3 is retained at the ER and released by the J chaperone Ydj1 in late G1 to trigger cell cycle entry," *Mol. Cell*, vol. 26, no. 5, pp. 649–662, Jun. 2007, doi: 10.1016/j.molcel.2007.04.023.
[66] L. Chen and K. Madura, "Yeast Importin-α (Srp1) performs distinct roles in the import of nuclear proteins and in targeting proteasomes to the nucleus," *J. Biol. Chem.*, vol. 289, no. 46, pp. 32339–32352, Nov. 2014, doi: 10.1074/jbc.m114.582023.
[67] S. A. Synowsky, M. van Wijk, R. Raijmakers, and A. J. R. Heck, "Comparative multiplexed mass spectrometric analyses of endogenously expressed yeast nuclear and cytoplasmic exosomes," *J. Mol. Biol.*, vol. 385, no. 4, pp. 1300–1313, Jan. 2009, doi: 10.1016/j.jmb.2008.11.011.
[68] M. A. MacKinnon, A. J. Curwin, G. J. Gaspard, A. B. Suraci, J. P. Fernández-Murray, and C. R. Mcmaster, "The Kap60-Kap95 karyopherin complex directly regulates phosphatidylcholine synthesis," *J. Biol. Chem.*, vol. 284, no. 11, pp. 7376–7384, Mar. 2009, doi: 10.1074/jbc.m809117200.
[69] Y. Sun, A. C. Martin, and D. G. Drubin, "Endocytic internalization in budding yeast requires coordinated actin nucleation and myosin motor activity," *Develop. Cell*, vol. 11, no. 1, pp. 33–46, Jul. 2006, doi: 10.1016/j.devcel.2006.05.008.
[70] K. Moravcevic, D. Alvarado, K. R. Schmitz, J. A. Kenniston, J. M. Mendrola, K. M. Ferguson, and M. A. Lemmon, "Comparison of Saccharomyces cerevisiae F-BAR domain structures reveals a conserved inositol phosphate binding site," *Structure*, vol. 23, no. 2, pp. 352–363, Feb. 2015, doi: 10.1016/j.str.2014.12.009.
[71] B. von Janowsky, T. Major, K. Knapp, and W. Voos, "The disaggregation activity of the mitochondrial ClpB homolog Hsp78 maintains Hsp70 function during heat stress," *J. Mol. Biol.*, vol. 357, no. 3, pp. 793–807, Mar. 2006, doi: 10.1016/j.jmb.2006.01.008.
[72] Q. Liu, J. Krzewska, K. Liberek, and E. A. Craig, "Mitochondrial Hsp70 ssc1: Role in protein folding," *J. Biol. Chem.*, vol. 276, no. 9, pp. 6112–6118, Mar. 2001, doi: 10.1074/jbc.m009519200.
[73] P. D. Ward, J. Botha, R. Buick, M. O. De Kock, D. H. Erwin, G. H. Garrison, J. L. Kirschvink, and R. Smith, "Abrupt and gradual extinction among late Permian land vertebrates in the karoo basin, south Africa," *Science*, vol. 307, no. 5710, pp. 709–714, Feb. 2005, doi: 10.1126/science.1107068.

[74] L. A. Strawn, T. Shen, and S. R. Wente, "The GLFG regions of Nup116p and Nup100p serve as binding sites for both Kap95p and Mex67p at the nuclear pore complex," *J. Biol. Chem.*, vol. 276, no. 9, pp. 6445–6452, Mar. 2001, doi: 10.1074/jbc.m008311200.

[75] N. P. C. Allen, L. Huang, A. Burlingame, and M. Rexach, "Proteomic analysis of nucleoporin interacting proteins," *J. Biol. Chem.*, vol. 276, no. 31, pp. 29268–29274, Aug. 2001, doi: 10.1074/jbc.m102629200.

[76] E. Kiseleva, T. D. Allen, S. Rutherford, M. Bucci, S. R. Wente, and M. W. Goldberg, "Yeast nuclear pore complexes have a cytoplasmic ring and internal filaments," *J. Structural Biol.*, vol. 145, no. 3, pp. 272–288, Mar. 2004, doi: 10.1016/j.jsb.2003.11.010.

[77] V. Goyon, R. Fronzes, B. Salin, J.-P. di-Rago, J. Velours, and D. Brèthes, "Yeast cells depleted in Atp14p fail to assemble Atp6p within the ATP synthase and exhibit altered mitochondrial cristae morphology," *J. Biol. Chem.*, vol. 283, no. 15, pp. 9749–9758, Apr. 2008, doi: 10.1074/jbc.m800204200.

[78] W. C. Y. Lau, L. A. Baker, and J. L. Rubinstein, "Cryo-EM structure of the yeast ATP synthase," *J. Mol. Biol.*, vol. 382, no. 5, pp. 1256–1264, Oct. 2008, doi: 10.1016/j.jmb.2008.08.014.

[79] G. V. Arselin, J. Vaillier, P.-V. Graves, and J. Velours, "ATP synthase of yeast mitochondria: Isolation Of The Subunit H and disruption of the ATP14 gene," *J. Biol. Chem.*, vol. 271, no. 34, pp. 20284–20290, 1996, doi: 10.1074/jbc.271.34.20284.

[80] J. Wang, S. U. Andersen, and P. Ratet, "Editorial: Molecular and cellular mechanisms of the Legume-Rhizobia symbiosis," *Frontiers Plant Sci.*, vol. 9, pp. 1–27, Dec. 2018, doi: 10.3389/fpls.2018.01839.

[81] F.-N. Vögtle, J. M. Burkhart, S. Rao, C. Gerbeth, J. Hinrichs, J.-C. Martinou, A. Chacinska, A. Sickmann, R. P. Zahedi, and C. Meisinger, "Intermembrane space proteome of yeast mitochondria," *Mol. Cellular Proteomics*, vol. 11, no. 12, pp. 1840–1852, Dec. 2012, doi: 10.1074/mcp.m112.021105.

[82] A. Wiederkehr, K. D. Meier, and H. Riezman, "Identification and characterization of Saccharomyces cerevisiae mutants defective in fluid-phase endocytosis," *Yeast*, vol. 18, no. 8, pp. 759–773, Jun. 2001, doi: 10.1002/yea.726.

[83] P. Branduardi, V. Longo, N. M. Berterame, G. Rossi, and D. Porro. (2013). *A Novel Pathway to Produce Butanol and Isobutanol in Saccharomyces Cerevisiae*. [Online]. Available: http://www.biotechnologyforbiofuels.com/content/6/1/68

[84] A. Andreadis, Y.-P. Hsu, G. B. Kohlhaw, and P. Schimmel, "Nucleotide sequence of yeast LEU2 shows 5'-noncoding region has sequences cognate to leucine," *Cell*, vol. 31, no. 2, pp. 319–325, Dec. 1982.

[85] E. D. Ryan, J. W. Tracy, and G. B. Kohlhaw, "Subcellular localization of the leucine biosynthetic enzymes in yeast," *J. Bacteriology*, vol. 116, no. 1, pp. 222–225, Oct. 1973.

[86] B. A. Kaufman, J. E. Kolesar, P. S. Perlman, and R. A. Butow, "A function for the mitochondrial chaperonin Hsp60 in the structure and transmission of mitochondrial DNA nucleoids in Saccharomyces cerevisiae," *J. Cell Biol.*, vol. 163, no. 3, pp. 457–461, Nov. 2003, doi: 10.1083/jcb.200306132.

[87] B. A. Kaufman, S. M. Newman, R. L. Hallberg, C. A. Slaughter, P. S. Perlman, and R. A. Butow. (2023). *In Organello Formaldehyde Crosslinking of Proteins to MtDNA: Identification of Bifunctional Proteins*. [Online]. Available: https://www.pnas.orgcgidoi10.1073pnas.140063197

[88] F. Boutouja, C. M. Stiehm, T. Mastalski, R. Brinkmeier, C. Reidick, F. El Magraoui, and H. W. Platta, "Vps10-mediated targeting of Pep4 determines the activity of the vacuole in a substrate-dependent manner," *Sci. Tech. Rep.*, vol. 9, no. 1, pp. 1–28, Jul. 2019, doi: 10.1038/s41598-019-47184-7.

[89] C. T. Yamashiro, P. M. Kane, D. F. Wolczyk, R. A. Preston, and T. H. Stevens, "Role of vacuolar acidification in protein sorting and zymogen activation: A genetic analysis of the yeast vacuolar proton-translocating ATPase," *Mol. Cellular Biol.*, vol. 10, no. 7, pp. 3737–3749, Jul. 1990.

[90] Y. Elbaz-Alon, E. Rosenfeld-Gur, V. Shinder, A. H. Futerman, T. Geiger, and M. Schuldiner, "A dynamic interface between vacuoles and mitochondria in yeast," *Develop. Cell*, vol. 30, no. 1, pp. 95–102, Jul. 2014, doi: 10.1016/j.devcel.2014.06.007.

[91] I. W. Campbell, X. Zhou, and A. Amon, "Spindle pole bodies function as signal amplifiers in the Mitotic Exit Network," *Mol. Biol. Cell*, vol. 31, no. 9, pp. 906–916, Apr. 2020, doi: 10.1091/mbc.E19-10-0584.

[92] I. Scarfone, M. Venturetti, M. Hotz, J. Lengefeld, Y. Barral, and S. Piatti, "Asymmetry of the budding yeast Tem1 GTPase at spindle poles is required for spindle positioning but not for mitotic exit," *PLOS Genet.*, vol. 11, no. 2, Feb. 2015, Art. no. e1004938, doi: 10.1371/journal.pgen.1004938.

[93] F. Monje-Casas and A. Amon, "Cell polarity determinants establish asymmetry in MEN signaling," *Develop. Cell*, vol. 16, no. 1, pp. 132–145, Jan. 2009, doi: 10.1016/j.devcel.2008.11.002.

[94] W. Mansour, M. A. Nakasone, M. von Delbruck, Z. Yu, D. Krutauz, N. Reis, O. Kleifeld, T. Sommer, D. Fushman, and M. H. Glickman, "Disassembly of Lys11 and mixed linkage polyubiquitin conjugates provides insights into function of proteasomal deubiquitinases Rpn11 and Ubp6," *J. Biol. Chem.*, vol. 290, no. 8, pp. 4688–4704, Feb. 2015, doi: 10.1074/jbc.m114.568295.

[95] L. Hofmann, R. Saunier, R. Cossard, M. Esposito, T. Rinaldi, and A. Delahodde, "A nonproteolytic proteasome activity controls organelle fission in yeast," *J. Cell Sci.*, vol. 122, no. 20, pp. 3673–3683, Oct. 2009, doi: 10.1242/jcs.050229.

[96] E. Bailly and S. I. Reed, "Functional characterization of Rpn3 uncovers a distinct 19S proteasomal subunit requirement for ubiquitin-dependent proteolysis of cell cycle regulatory proteins in budding yeast," *Mol. Cellular Biol.*, vol. 19, no. 10, pp. 6872–6890, Oct. 1999.

[97] J. Roelofs, S. Park, W. Haas, G. Tian, F. E. McAllister, Y. Huo, B.-H. Lee, F. Zhang, Y. Shi, S. P. Gygi, and D. Finley, "Chaperone-mediated pathway of proteasome regulatory particle assembly," *Nature*, vol. 459, no. 7248, pp. 861–865, Jun. 2009, doi: 10.1038/nature08063.

[98] R. Verma, S. Chen, R. Feldman, D. Schieltz, J. Yates, J. Dohmen, and R. J. Deshaies, "Proteasomal proteomics: Identification of nucleotide-sensitive proteasome-interacting proteins by mass spectrometric analysis of affinity-purified proteasomes," *Mol. Biol. Cell*, vol. 11, no. 10, pp. 3425–3439, Oct. 2000.

[99] Y. Saeki, A. Toh-e, T. Kudo, H. Kawamura, and K. Tanaka, "Multiple proteasome-interacting proteins assist the assembly of the yeast 19S regulatory particle," *Cell*, vol. 137, no. 5, pp. 900–913, May 2009, doi: 10.1016/j.cell.2009.05.005.

[100] Z. Yu, O. Kleifeld, A. Lande-Atir, M. Bsoul, M. Kleiman, D. Krutauz, A. Book, R. D. Vierstra, K. Hofmann, N. Reis, M. H. Glickman, and E. Pick, "Dual function of Rpn5 in two PCI complexes, the 26S proteasome and COP9 signalosome," *Mol. Biol. Cell*, vol. 22, no. 7, pp. 911–920, Apr. 2011, doi: 10.1091/mbc.e10-08-0655.

[101] N. Smith, D. J. Adle, M. Zhao, X. Qin, H. Kim, and J. Lee, "Endoplasmic reticulum-associated degradation of pca1p, a polytopic protein, via interaction with the proteasome at the membrane," *J. Biol. Chem.*, vol. 291, no. 29, pp. 15082–15092, Jul. 2016, doi: 10.1074/jbc.m116.726265.

[102] H. Zhao, G. Liu, and X. Cao, "A seed expansion-based method to identify essential proteins by integrating protein–protein interaction sub-networks and multiple biological characteristics," *BMC Bioinf.*, vol. 24, no. 1, pp. 1–29, Dec. 2023.

[103] Z. Mirzoyan, M. Sollazzo, M. Allocca, A. M. Valenza, D. Grifoni, and P. Bellosta, "Drosophila melanogaster: A model organism to study cancer," *Frontiers Genet.*, vol. 10, Mar. 2019, Art. no. 425859.

[104] R. Apweiler, "The universal protein resource (uniprot)," *Nucleic Acids Res*, vol. 36, pp. 1–12, Jan. 2008, doi: 10.1093/NAR/GKM895.

[105] A. Bateman et al., "UniProt: The universal protein knowledgebase in 2021," *Nucleic Acids Res.*, vol. 49, no. D1, pp. D480–D489, Jan. 2021, doi: 10.1093/nar/gkaa1100.

[106] H. Weavers, S. Prieto-Sánchez, F. Grawe, A. Garcia-López, R. Artero, M. Wilsch-Bräuninger, M. Ruiz-Gómez, H. Skaer, and B. Denholm, "The insect nephrocyte is a podocyte-like cell with a filtration slit diaphragm," *Nature*, vol. 457, no. 7227, pp. 322–326, Oct. 2008, doi: 10.1038/nature07526.

[107] J. Lu, Y. Wang, X. Wang, D. Wang, G. O. Pflugfelder, and J. Shen, "The Tbx6 transcription factor dorsocross mediates dpp signaling to regulate Drosophila thorax closure," *Int. J. Mol. Sci.*, vol. 23, no. 9, p. 4543, Apr. 2022, doi: 10.3390/ijms23094543.

[108] F. Zhang, Y. Zhao, and Z. Han, "An in vivo functional analysis system for renal gene discovery in Drosophila pericardial nephrocytes," *J. Amer. Soc. Nephrology*, vol. 24, no. 2, pp. 191–197, Feb. 2013, doi: 10.1681/asn.2012080769.

[109] A. V. Emelyanov, J. Rabbani, M. Mehta, E. Vershilova, M. C. Keogh, and D. V. Fyodorov, "Drosophila TAP/P32 is a core histone chaperone that cooperates with NAP-1, NLP, and nucleophosmin in sperm chromatin remodeling during fertilization," *Genes Develop.*, vol. 28, no. 18, pp. 2027–2040, Sep. 2014, doi: 10.1101/gad.248583.114.

[110] D. Ducat, S.-I. Kawaguchi, H. Liu, J. R. Yates, and Y. Zheng, "Regulation of microtubule assembly and organization in mitosis by the AAA+ ATPase pontin," *Mol. Biol. Cell*, vol. 19, no. 7, pp. 3097–3110, Jul. 2008, doi: 10.1091/mbc.e07-11-1202.

[111] J. Dobbelaere, F. Josué, S. Suijkerbuijk, B. Baum, N. Tapon, and J. Raff, "A genome-wide RNAi screen to dissect centriole duplication and centrosome maturation in drosophila," *PLoS Biol.*, vol. 6, no. 9, p. e224, Sep. 2008, doi: 10.1371/journal.pbio.0060224.

[112] P. Merlo, B. Frost, S. Peng, Y. J. Yang, P. J. Park, and M. Feany, "P53 prevents neurodegeneration by regulating synaptic genes," *Proc. Nat. Acad. Sci. USA*, vol. 111, no. 50, pp. 18055–18060, Dec. 2014, doi: 10.1073/pnas.1419083111.

[113] O. Bereczki, Z. Ujfaludi, N. Pardi, Z. Nagy, L. Tora, I. M. Boros, and E. Balint, "TATA binding protein associated factor 3 (TAF3) interacts with p53 and inhibits its function," *BMC Mol. Biol.*, vol. 9, no. 1, p. 57, Jun. 2008, doi: 10.1186/1471-2199-9-57.

[114] M. H. Brodsky, B. T. Weinert, G. Tsang, Y. S. Rong, N. M. McGinnis, K. G. Golic, D. C. Rio, and G. M. Rubin, "Drosophila melanogaster MNK/Chk2 and P53 regulate multiple DNA repair and apoptotic pathways following DNA damage," *Mol. Cellular Biol.*, vol. 24, no. 3, pp. 1219–1231, Feb. 2004, doi: 10.1128/mcb.24.3.1219-1231.2004.

[115] T. Bossing and A. H. Brand, "Dephrin, a transmembrane ephrin with a unique structure, prevents interneuronal axons from exiting the drosophilaembryonic CNS," *Development*, vol. 129, no. 18, pp. 4205–4218, Sep. 2002, doi: 10.1242/dev.129.18.4205.

[116] R. E. Dearborn, Y. Dai, B. Reed, T. Karian, J. Gray, and S. Kunes, "Reph, a regulator of eph receptor expression in the Drosophila melanogaster optic lobe," *PLoS One*, vol. 7, no. 5, May 2012, Art. no. e37303, doi: 10.1371/journal.pone.0037303.

[117] D. J. L. Tan, H. Dvinge, A. Christoforou, P. Bertone, A. Martinez Arias, and K. S. Lilley, "Mapping organelle proteins and protein complexes in Drosophila melanogaster," *J. Proteome Res.*, vol. 8, no. 6, pp. 2667–2678, Jun. 2009, doi: 10.1021/pr800866n.

[118] J. R. Stoltzfus, W. J. Horton, and M. S. Grotewiel, "Odor-guided behavior in Drosophila requires calreticulin," *J. Comparative Physiol. A: Sensory, Neural, Behav. Physiol.*, vol. 189, no. 6, pp. 471–483, Jun. 2003, doi: 10.1007/s00359-003-0425-z.

[119] M. J. Walker, C. M. Rylett, J. N. Keen, N. Audsley, M. Sajid, A. D. Shirras, and R. E. Isaac, "Proteomic identification of Drosophila melanogaster male accessory gland proteins, including a pro-cathepsin and a soluble γ-glutamyl transpeptidase," *Proteome Sci.*, vol. 4, no. 1, p. 9, May 2006, doi: 10.1186/1477-5956-4-9.

[120] R. Albert, "Scale-free networks in cell biology," *J. Cell Sci.*, vol. 118, no. 21, pp. 4947–4957, Nov. 2005.

ABDUL MOIZ received the B.E. degree in biomedical engineering and the M.S. degree in applied mathematics from the NED University of Engineering & Technology, Karachi, Pakistan, in 2021 and 2023, respectively. He is currently a Lecturer with the Department of Biomedical Engineering, Salim Habib University, Karachi. His current research interests include bioinformatics, biological networks, fuzzy logic, and biomaterials.

UBAIDA FATIMA received the B.E. degree in biomedical engineering and the M.S. and Ph.D. degrees in applied mathematics from the NED University of Engineering & Technology, Karachi, Pakistan, in 2011, 2014, and 2020, respectively. She is currently an Assistant Professor with the Department of Mathematics, NED University of Engineering & Technology. Her research interests include big data analysis, social network analysis, artificial intelligence, data mining, fuzzy logic, neural networks, data science, health care, fitness, and applied statistics.

M. ZEESHAN UL HAQUE (Senior Member, IEEE) received the B.S. degree in biomedical engineering from the Sir Syed University of Engineering and Technology, Karachi, Pakistan, in 2003, the M.Sc. degree in medical electronics and physics from Queen Mary University, London, U.K., in 2005, and the Ph.D. degree in bioengineering from the University of Auckland, New Zealand. He is currently a HEC approved Ph.D. Supervisor, having more than 17 years of experience in teaching, research, and administration. He has a good understanding of productive and operative teaching methods that promote exciting learning environment. He has an excellent knowledge of accreditation process of engineering program. He is also a Pakistan Engineering Council's (PEC) Program Evaluator (PEV) for Biomedical Engineering Education at national level. He is a member of various statutory bodies and project evaluation committees and journals reviewer. He also served as a Co-Convener for Biomedical Engineering of National Curriculum Review Committee, in 2017, constituted by Higher Education Commission, Pakistan.