

Topology-Enhanced and Label Correlation-Aware Model for Protein-Protein Interaction Prediction

Bin Deng, Huifang Ma^{*}, Ruijia Zhang, Meihuizi Jia, Rui Bing

College of Computer Science and Engineering, Northwest Normal University
dengbin151@gmail.com, mahuifang@yeah.net, zrj20327@gmail.com, {jiameihuizi, bingrui}@nwnu.edu.cn

Abstract

Protein-Protein Interactions (PPIs) prediction is crucial for understanding cellular functions and disease mechanisms. Existing deep learning-based methods primarily rely on direct interaction within the PPI network to update protein representations. However, (1) such networks overlook the potential associations between functionally similar proteins, limiting the smoothing capability of Graph Neural Networks (GNNs) in learning representations for similar nodes. (2) Additionally, most approaches fail to adequately model the latent dependencies among interaction types (edge labels), which hinders their performance in PPI prediction tasks. To address these limitations, we propose **TELC-PPI**, a topology-enhanced and label correlation-aware model for protein-protein interactions prediction. Specifically, TELC-PPI first identifies similar proteins by leveraging both the topological information of the PPI network and the label distributions of nodes, constructing similarity edges. Then, it incorporates label co-occurrence statistics into the learning of label embeddings. Experimental results on multiple datasets and under various data split settings demonstrate that TELC-PPI significantly outperforms existing methods, validating the effectiveness of our model design.

Code — <https://github.com/dengbin151/TELC-PPI>

Introduction

Protein-Protein Interactions (PPIs) play a central role in essential biological processes, including cell signaling, metabolic regulation, and many others (Acuner Ozbabacan et al. 2011). Accurately identifying and predicting the specific interaction types between proteins is crucial for elucidating molecular mechanisms, advancing disease research and accelerating drug development. While traditional wet-lab methods (Clegg 1995; Kodama and Hu 2012; Zhou, Li, and Wang 2016) offer direct observation of these interactions, they remain resource-intensive, time-consuming, and challenging to scale. Consequently, computational methods for PPI prediction have emerged as a prominent research focus due to their efficiency and scalability.

As an effective class of computational approaches, Graph Neural Networks (GNNs) have become a dominant

^{*}Corresponding author. Email: mahuifang@yeah.net
Copyright © 2026, Association for the Advancement of Artificial Intelligence (www.aaai.org). All rights reserved.

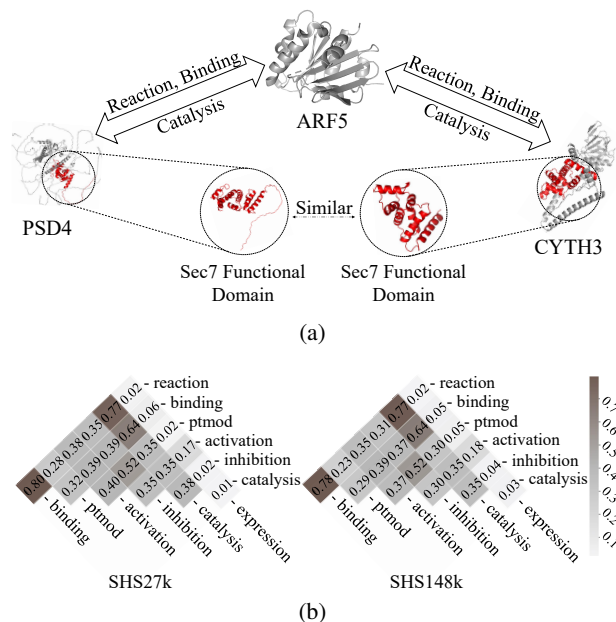


Figure 1: (a) Illustration of the relationships among ARF5, CYTH3, and PSD4. Red: Sec7 domains; gray: other regions. (b) Label co-occurrence on SHS27k and SHS148k. Values are average conditional probabilities: $(p(i|j) + p(j|i))/2$, where i, j are interaction types.

paradigm for PPI prediction, owing to their powerful ability to model graph-structured data (Soleymani et al. 2022; Tang et al. 2023; Durham et al. 2023). These methods represent proteins as nodes and interactions as edges, learning protein representations via neighborhood aggregation. However, existing GNN-based approaches suffer from two key limitations. (1) **Heterophily in PPI graphs**. PPI networks derived from widely used datasets (e.g., SHS27k, SHS148k, and STRING; see Experiments) contain only experimentally validated interactions, which predominantly connect functionally complementary protein pairs, rather than functionally similar ones. As Figure 1(a) illustrates, while CYTH3 and PSD4 both interact with ARF5 due to their activation capability, the graph fails to capture their shared Sec7 functional domain essential for this activation. The omission violates

the **homophily assumption** central to GNNs (Wang et al. 2022; Ma et al. 2021), which states that connected nodes exhibit feature/label similarity, thereby limiting representation learning efficacy. **(2) Ignored dependencies among interaction types.** Existing methods typically treat interaction types (i.e., edge labels) as independent (Lv et al. 2021; Gao et al. 2023; Wu et al. 2024), disregarding their strong co-occurrence relationships. Figure 1(b) shows consistent co-occurrence patterns across datasets (e.g., frequent *binding–reaction/catalysis* pairs). Ignoring these interaction dependencies contradicts the **collaborative assumption** of multi-label learning (Feng, An, and He 2019; Du et al. 2024), which indicates that labels exhibit semantic correlations, consequently impairing performance in multi-type PPI prediction.

To address the aforementioned limitations, we propose **TELC-PPI (Topology-Enhanced and Label Correlation-Aware Model for Protein–Protein Interaction Prediction)**. TELC-PPI jointly captures the functional similarity between protein pairs and the latent dependencies among different interaction types, aiming to improve the accuracy of multi-label PPI prediction. Specifically, we introduce a topology enhancement mechanism to better model functionally similar protein pairs. We observe that proteins sharing similar interaction types with a common partner (e.g., CYTH3 and PSD4 both interacting with ARF5) are likely to be functionally related. Based on this observation, we identify and explicitly connect such hidden similar protein pairs using graph topology and edge label distributions. This approach enriches the semantic expressiveness of protein representations. To address the insufficient modeling of label dependencies, we design a label correlation-aware module. This module learns representations for each interaction type by incorporating maximum likelihood loss and label co-occurrence information, thereby enabling the model to effectively capture dependencies among the labels.

In summary, our key contributions are as follows:

- **Graph Topology Enhancement.** We introduce a topology-enhanced mechanism that explicitly connects functionally similar protein pairs based on semantic associations. This approach significantly improves node representation quality by mitigating the heterophily inherent in PPI graphs.
- **Label Dependency Modeling.** We propose a novel correlation-aware module that integrates label co-occurrence constraints with maximum likelihood loss. This enables the model to explicitly capture dependencies among interaction types, thereby overcoming the limitations of modeling labels independently.
- **State-of-the-Art Performance.** Extensive experiments demonstrate that TELC-PPI achieves state-of-the-art performance on most datasets (ranging from SHS27k to STRING) across various data splits, establishing a new benchmark for multi-label PPI prediction.

Related Work

Protein-Protein Interaction Prediction

Predicting protein–protein interactions (PPIs) is fundamental for understanding cellular mechanisms and disease processes. Existing machine learning–based PPI prediction methods can be broadly categorized into three groups. (1) Traditional machine learning methods rely on handcrafted sequence features and conventional classifiers. For example, DNN-PPI (Li et al. 2018) employs amino acid composition features with an MLP. However, such methods are constrained by limited feature expressiveness. (2) Deep learning methods extract representations directly from raw protein sequences in an end-to-end manner. Representative approaches include DPPI (Hashemifar et al. 2018), which applies CNNs to capture local patterns, and PIPR (Chen et al. 2019), which combines CNNs with BiGRU to model long-range dependencies. Despite improved performance, these methods struggle to exploit structural information from PPI networks. (3) GNN-based methods leverage graph neural networks to model topological structures in PPI networks. For instance, GNN-PPI (Lv et al. 2021) enhances predictions for unseen protein pairs, SemiGNN-PPI (Zhao et al. 2023) improves semi-supervised learning via multi-graph fusion, and HIGH-PPI (Gao et al. 2023) adopts hierarchical graph modeling. More recently, MAPE-PPI (Wu et al. 2024) integrates sequence and structural cues through a microenvironment-aware mechanism. Although effective, these approaches often overlook protein-pair similarity relationships or label correlations, leading to suboptimal multi-type PPI prediction.

Multi-label Classification

Multi-label classification (Yu and Zhang 2021; Lin et al. 2024) involves objects associated with multiple labels simultaneously, making it more realistic and attracting increasing attention compared to traditional single-label classification. Existing methods can be broadly categorized into two groups: (1) Methods that ignore label correlations. These approaches predict labels directly based on node representations (Ridnik et al. 2021; Li et al. 2023; Saidabad et al. 2024) without explicitly modeling semantic relationships among labels. While simple, this often limits prediction performance by neglecting structured information within labels. (2) Methods that model label relationships (Bei et al. 2025; Gao, Zhang, and Zhou 2019; Shi, Tang, and Zhu 2019). For example, CAMEL (Feng, An, and He 2019) integrates learned label correlations into training, allowing each label’s prediction to depend not only on its own features but also on predictions from other labels, enabling collaborative decision-making. CorGCN (Bei et al. 2025) introduces a correlation-aware graph decomposition module that constructs label embeddings and guides node representations to be decomposed on their corresponding label graphs, explicitly modeling dependencies between nodes and labels as well as among labels themselves, thereby improving classification accuracy. However, these methods primarily address multi-label node classification and do not consider multi-label edge classification tasks important in PPI prediction.

Preliminaries

Given a set of proteins and their multiple types of interactions, a multi-relational PPI graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ can be constructed with an adjacency matrix $\mathbf{A} \in \mathbb{R}^{N \times N}$, where $\mathcal{V} = \{v_1, v_2, \dots, v_N\}$ denotes the set of protein nodes and $N = |\mathcal{V}|$ is the number of nodes. $\mathcal{E} \subseteq \mathcal{V} \times \mathcal{V}$ represents the set of interaction edges between protein pairs. Each protein node $v_i \in \mathcal{V}$ is associated with a feature vector $\mathbf{x}_i \in \mathbb{R}^d$, forming a node feature matrix $\mathbf{X} \in \mathbb{R}^{N \times d}$. During training, the edge set \mathcal{E} is divided into labeled edges \mathcal{E}_L and unlabeled edges $\mathcal{E}_U = \mathcal{E} \setminus \mathcal{E}_L$. Each labeled edge $e_{u,v} \in \mathcal{E}_L$ is annotated with a multi-hot vector $\mathbf{y}_{u,v} \in \{0, 1\}^C$, where the c -th entry is 1 if the edge involves the c -th interaction type.

Multi-type PPI prediction aims to learn a multi-label edge classifier f_θ that, given graph \mathcal{G} , node feature matrix \mathbf{X} , and the labeled edges \mathcal{E}_L , predicts a probability vector $\hat{\mathbf{y}}_{i,j} \in [0, 1]^C$ representing interaction types for each unlabeled edges $e_{i,j} \in \mathcal{E}_U$:

$$\hat{\mathbf{y}}_{i,j} = f_\theta(i, j, \mathcal{G}, \mathbf{X}), \quad \forall e_{i,j} \in \mathcal{E}_U. \quad (1)$$

Methodology

As shown in Figure 2, TELC-PPI consists of five key components: **(a)** The input module, which includes the original PPI graph, edge labels, label prototypes, and their corresponding label indices; **(b)** The structure enhancement module, which improves semantic representations by connecting functionally similar nodes based on label distribution and graph topology; **(c)** The edge representation learning module, which updates node features on the enhanced graph and generates informative edge representations; **(d)** The label modeling module, which learns label embeddings by combining label co-occurrence information—used to capture inter-label correlations—with a maximum likelihood loss; and **(e)** The PPI prediction module, which predicts protein-protein interaction types by computing the dot product between edge and label embeddings.

Topology Enhancement Module

H2 Principle. As previously observed, the original PPI graph exhibits significant heterophily—functionally similar proteins are often not directly connected. This weakens the expressive power of GNN models that rely on the homophily assumption. To address this, inspired by the studies of Chua, Sung, and Wong (2006); Kovács et al. (2019), we propose identifying functionally similar protein pairs based on the following insight: **two proteins that interact with sufficient neighbors via similar interaction types are likely to share functional relevance.** However, directly analyzing all common neighbors is computationally expensive and sensitive to noise. Therefore, we introduce an efficient and robust strategy—termed the H2 Principle (2-Hop Similarity Principle)—to identify potential associations by jointly evaluating the similarity of interaction type distributions and the number of shared neighbors. The definition is as follows:

Definition (H2 Principle): For a node pair in the original PPI graph, a potential functional association exists according to the H2 Principle provided that:

- **Sufficient Shared Neighbors:** The node pair shares a sufficient number of common neighbors;
- **Similar Interaction Patterns:** The edge interaction type distributions from each node to the common neighbors exhibit high similarity.

H2 Edges Construction. To identify potential functionally related protein pairs, we compute the square of the adjacency matrix \mathbf{A} , denoted \mathbf{A}^2 , to obtain all two-hop node pairs connected via a common intermediate node. Subtracting the original adjacency \mathbf{A} , then removing negative values and self-loops, yields the H2 candidate edge adjacency matrix $\mathbf{A}^{\text{cand}} \in \mathbb{R}^{N \times N}$. Non-zero elements (i.e., edge weights) are normalized using min-max normalization:

$$\mathbf{A}^{\text{cand}} = \text{MinMaxNorm}(\max(\mathbf{A}^2 - \mathbf{A}, 0) \odot (\mathbf{1} - \mathbf{I})), \quad (2)$$

where $\max(\cdot, 0)$ zeroes negatives, $\mathbf{1}$ and \mathbf{I} are all-ones and identity matrices of the same shape as \mathbf{A} , respectively, \odot denotes element-wise multiplication, and $\text{MinMaxNorm}(\cdot)$ represents min-max normalization.

Proteins connected by two-hop paths may differ functionally, so indiscriminately adding edges can introduce noise. Thus, assessing functional similarity for two-hop pairs is necessary. Since protein functions are reflected in their direct interaction patterns—for example, proteins involved in *binding* and *catalysis* share specific sites and activities—we apply a semantic constraint based on local interactions. Specifically, for each node v_i , we aggregate and normalize incident edge labels to derive a probability distribution $\mathbf{p}_i \in \mathbb{R}^C$ representing its local functional profile:

$$\mathbf{p}_i = \frac{\sum_{u \in \mathcal{N}_i} \mathbf{y}_{i,u}}{\sum_{c \in C} \sum_{u \in \mathcal{N}_i} \mathbf{y}_{i,u}^c}, \quad (3)$$

where \mathcal{N}_i denotes the neighbors of v_i , and $\mathbf{y}_{i,u}^c$ is the value of label l_c on edge $e_{i,u}$.

To jointly incorporate topological and semantic information, we design a similarity scoring function to evaluate the importance of each H2 candidate edge $e_{i,j}^{\text{cand}}$. The overall similarity score is defined as:

$$\text{sim}(i, j) = \alpha \cdot \mathbf{A}_{i,j}^{\text{cand}} + (1 - \alpha) \cdot \frac{\mathbf{p}_i^T \mathbf{p}_j}{\|\mathbf{p}_i\|_2 \cdot \|\mathbf{p}_j\|_2}, \quad (4)$$

where $\|\cdot\|_2$ is the L2 norm, and $\alpha \in [0, 1]$ controls the contribution between topological strength and semantic similarity. Note that $\text{sim}(i, j)$ is computed only for pairs with $\mathbf{A}_{i,j}^{\text{cand}} > 0$, ensuring that only candidate edges whose endpoints share at least one common neighbor are considered in the subsequent procedure. Subsequently, we normalize the similarity scores over the candidate edge set to obtain the sampling probability of each candidate edge:

$$\text{Pr}(i, j) = \frac{\text{sim}(i, j)}{\sum_{(u,v) \in \mathcal{E}_{\text{cand}}} \text{sim}(u, v)}, \quad (5)$$

where $\mathcal{E}_{\text{cand}}$ denotes the set of nonzero-weight edges in \mathbf{A}^{cand} . We then use the normalized probability distribution

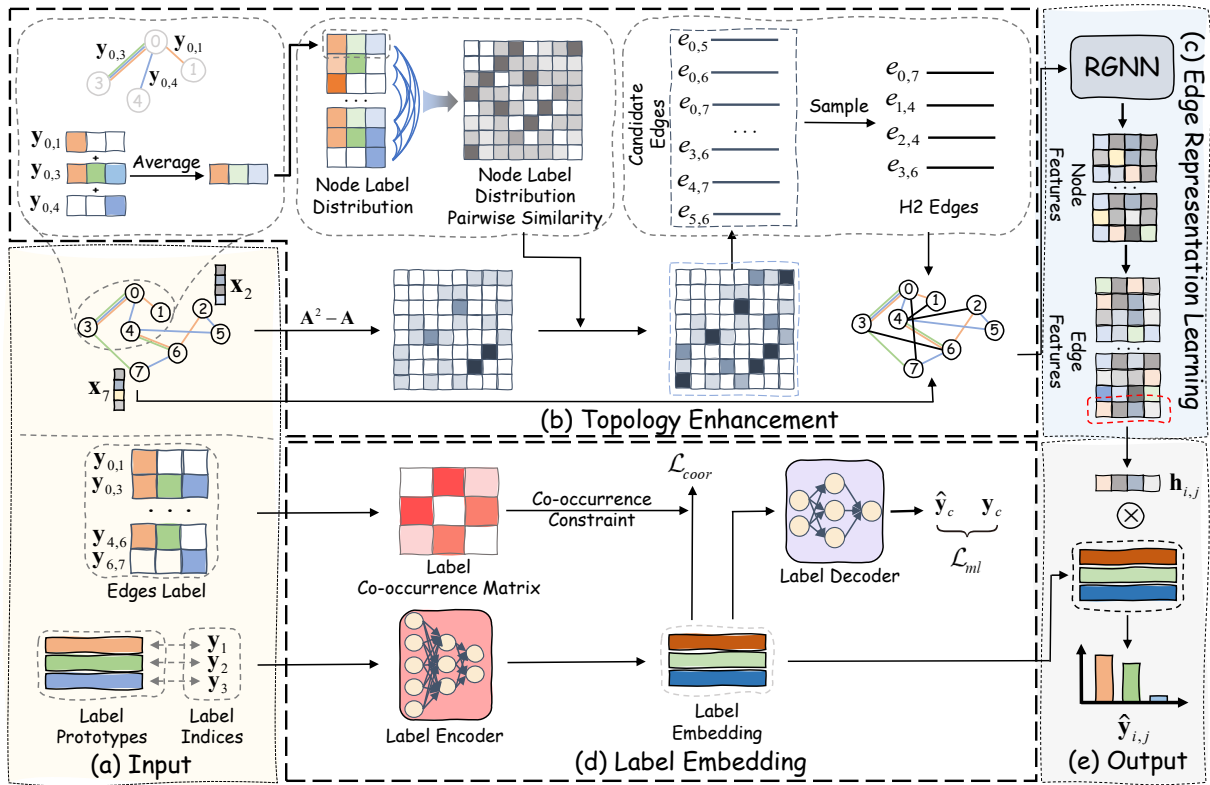


Figure 2: The model flow of TELC-PPI framework. TELC-PPI includes five key components: (a) Model inputs. (b) Topology enhancement, adding H2 edges to the original PPI graph. (c) Edge representation learning, updating node embeddings and deriving edge features. (d) Label embedding, optimizing label representations with co-occurrence constraints and maximum likelihood. (e) Model outputs.

$\text{Pr}(i, j)$ to stochastically sample a proportion of edges from the candidate set:

$$\mathcal{E}_{\text{H2}} = \text{Sample}(\mathcal{E}_{\text{cand}}, \text{Pr}, \rho), \quad (6)$$

where $\rho \in (0, 1]$ is the sampling ratio. This probabilistic strategy favors protein pairs with strong structural and functional similarity, while still allowing rare but meaningful interactions (e.g., *inhibition*), thereby balancing reliability and diversity. Once sampled, H2 edges remain fixed during training to ensure stability and convergence.

The augmented graph is formed by adding the H2 edges to the original PPI graph:

$$\mathcal{G}' = (\mathcal{V}, \mathcal{E} \cup \mathcal{E}_{\text{H2}}). \quad (7)$$

Edge Representation Learning Module

We apply relational graph convolution on the enhanced heterogeneous graph \mathcal{G}' to update node representations, where the initial representation \mathbf{h}_i^0 is set as the node feature \mathbf{x}_i :

$$\mathbf{h}_i^{(k+1)} = \sigma \left(\text{BN} \left(\sum_{r \in \mathcal{R}} \mathbf{W}_r \sum_{j \in \mathcal{N}_r(i)} \mathbf{h}_j^{(k)} \right) \right) + \mathbf{h}_i^{(k)}, \quad (8)$$

where $\mathbf{h}_i^{(k)}$ denotes the representation at layer k , $\mathcal{N}_r(i)$ is the set of r -type neighbors of node v_i , \mathbf{W}_r is the relation-specific weight matrix, \mathcal{R} contains two relation types: the original PPI edges and the added H2 edges in \mathcal{G}' , $\text{BN}(\cdot)$ denotes batch normalization, and $\sigma(\cdot)$ is a non-linear activation function.

After obtaining the updated node representations, the corresponding edge representations are computed as follows:

$$\mathbf{h}_{i,j}^{(k+1)} = \mathbf{h}_i^{(k+1)} \odot \mathbf{h}_j^{(k+1)}. \quad (9)$$

Label Embedding Module

Label Prototype Initialization. To incorporate label information, we introduce a trainable prototype matrix $\mathbf{E}^p \in \mathbb{R}^{C \times d_p}$, where each row \mathbf{E}_c^p represents the prototype of label l_c . A multilayer perceptron is employed to project these prototypes into a latent space and obtain expressive label embeddings:

$$\mathbf{E} = \sigma(\mathbf{E}^p \mathbf{W}_1 + \mathbf{b}_1), \quad (10)$$

where $\mathbf{W}_1 \in \mathbb{R}^{d_p \times d_l}$ is a trainable weight matrix, $\mathbf{b}_1 \in \mathbb{R}^{d_l}$ is a bias term.

Label Co-occurrence Constraint Modeling. In multilabel learning, label co-occurrence is a key prior for modeling dependencies (Kurata, Xiang, and Zhou 2016; Hang

and Zhang 2021; Wang and Zhang 2023). This is evident in multi-type PPI prediction, where interactions like *reaction* and *binding* frequently co-occur, reflecting their functional synergy. Thus, leveraging co-occurrence statistics can enhance label embedding expressiveness.

To model label co-occurrence, we construct a label co-occurrence matrix $\mathbf{M}^{\text{coor}} \in \mathbb{R}^{C \times C}$ based on the empirical co-occurrence frequencies in the training set:

$$\mathbf{M}_{i,j}^{\text{coor}} = \frac{1}{2} [p(l_i|l_j) + p(l_j|l_i)], \quad (11)$$

where $p(l_j|l_i)$ denotes the conditional probability of label l_j given label l_i . To eliminate self-dependency, diagonal entries of \mathbf{M}^{coor} are set to zero.

To encourage semantically similar labels (i.e., those with high co-occurrence) to have closer embeddings, we introduce a label embedding consistency loss:

$$\mathcal{L}_{\text{coor}} = \sum_{i=1}^C \sum_{j=1}^C \mathbf{M}_{i,j}^{\text{coor}} \cdot \|\mathbf{E}_i - \mathbf{E}_j\|_2^2. \quad (12)$$

By minimizing the embedding distance between frequently co-occurring labels, the model effectively captures label correlations and improves generalization in multi-label classification.

Label Discrimination via Maximum Likelihood. While modeling label co-occurrence brings correlated labels closer in embedding space, it may blur semantic distinctions. To address this, we introduce a maximum-likelihood loss that trains a decoder ψ to reconstruct each label’s class index from its embedding, thereby enhancing the discriminative power of label embeddings:

$$\mathcal{L}_{\text{ml}} = \sum_{c=1}^C \text{CE}(\hat{\mathbf{y}}_c^l, \mathbf{y}_c^l), \quad \hat{\mathbf{y}}_c^l = \psi(\mathbf{E}_c), \quad (13)$$

where $\hat{\mathbf{y}}_c^l$ denotes the predicted probability distribution, and \mathbf{y}_c^l is the one-hot vector for class c , $\psi(\cdot)$ denotes the label decoder, $\text{CE}(\cdot, \cdot)$ denotes the cross-entropy loss function.

Prediction and Training Strategy

Prediction Module. After obtaining edge and label embeddings, the model projects them into a shared space, where prediction scores are computed via inner product:

$$\hat{\mathbf{y}}_{i,j} = (\mathbf{h}_{i,j} \mathbf{W}_2)(\mathbf{E} \mathbf{W}_3)^T, \quad (14)$$

where $\hat{\mathbf{y}}_{i,j}$ denotes the predicted scores for C interaction types on edge $e_{i,j}$, and $\mathbf{W}_2 \in \mathbb{R}^{d_e \times d'}$, $\mathbf{W}_3 \in \mathbb{R}^{d_l \times d'}$ are learnable projection matrices.

To train the model, we adopt the multi-label binary cross-entropy loss:

$$\mathcal{L}_{\text{sup}} = \frac{1}{|\mathcal{E}_L|} \sum_{e_{i,j} \in \mathcal{E}_L} \text{BCE}(\hat{\mathbf{y}}_{i,j}, \mathbf{y}_{i,j}), \quad (15)$$

where $\text{BCE}(\cdot, \cdot)$ represents the element-wise binary cross-entropy loss.

Joint Training Objective. During training, the model optimizes a joint loss that combines the supervised loss \mathcal{L}_{sup} , the label co-occurrence regularization loss $\mathcal{L}_{\text{coor}}$, and the maximum likelihood loss \mathcal{L}_{ml} :

$$\mathcal{L}_{\text{total}} = \lambda_1 \cdot \mathcal{L}_{\text{sup}} + \lambda_2 \cdot \mathcal{L}_{\text{coor}} + (1 - \lambda_1 - \lambda_2) \cdot \mathcal{L}_{\text{ml}}, \quad (16)$$

where $\lambda_1, \lambda_2 \in [0, 1]$ control the relative contributions of each loss term, with $\lambda_1 + \lambda_2 \leq 1$. Joint optimization improves both edge prediction and label embedding quality, yielding more robust multi-label PPI prediction.

Experiments

Dataset

We conducted extensive experiments on three datasets: STRING, SHS27k, and SHS148k. The STRING dataset (Szklarczyk et al. 2019) comprises 15,355 proteins and 593,397 multi-label human PPI records. SHS27k and SHS148k (Lv et al. 2021) are curated subsets of STRING, consisting of proteins with more than 50 amino acids and less than 40% sequence similarity. Specifically, SHS27k comprises 1,690 proteins and 7,624 PPIs, while SHS148k comprises 5,189 proteins and 44,488 PPIs. Each PPI is annotated with at least one of seven interaction types: *activation*, *binding*, *catalysis*, *expression*, *inhibition*, *ptmod*, and *reaction*. The 3D structures of all proteins are predicted by AlphaFold2 (Jumper et al. 2021).

Experimental Details

Experimental Settings. To evaluate the effectiveness of the proposed method, we divide each dataset into training, validation, and test sets with a ratio of 6:2:2. Additionally, we adopt three partitioning strategies proposed by GNN-PPI (Lv et al. 2021), including Random, Breadth-First Search (BFS), and Depth-First Search (DFS). Considering the significant class imbalance among different PPI types, we employ the micro-F1 score as the evaluation metric. For each partitioning strategy on every dataset, we perform five runs with different random seeds, select the model with the best validation performance, and report the average micro-F1 score on the test set. In our experiments, protein representations pre-trained by MAPE-PPI are used as the initial protein embeddings.

Baseline Methods. We compare our model with two categories of baseline methods. The first category includes deep learning-based approaches: DPPI (Hashemifar et al. 2018), DNN-PPI (Li et al. 2018), and PIPR (Chen et al. 2019). The second category consists of GNN-based methods: GNN-PPI (Lv et al. 2021), SemiGNN-PPI (Zhao et al. 2023), HIGH-PPI (Gao et al. 2023), and MAPE-PPI (Wu et al. 2024).

Results and Analysis

Benchmark Comparison. We compare our proposed method with several baseline approaches across different datasets and partition settings. Table 1 summarizes the results, based on which we make the following observations:

Method	SHS27k			SHS148k			STRING		
	Random	DFS	BFS	Random	DFS	BFS	Random	DFS	BFS
DPPI	70.45	43.69	43.87	76.10	51.43	50.80	92.49	63.41	54.41
DNN-PPI	75.18	48.90	51.59	85.44	56.70	54.56	81.91	61.34	51.53
PIPR	79.59	52.19	47.13	88.81	61.38	58.57	93.68	64.97	53.80
GNN-PPI	83.65	66.52	63.08	90.87	75.34	69.53	94.53	84.28	75.69
SemiGNN-PPI	85.57	69.25	67.94	91.40	77.62	71.06	94.80	84.85	77.10
HIGH-PPI	86.23	70.24	68.40	91.26	78.18	72.87	OOM	OOM	OOM
MAPE-PPI	<u>88.91</u>	<u>71.98</u>	<u>70.38</u>	<u>92.38</u>	<u>79.45</u>	74.76	<u>96.12</u>	<u>86.50</u>	<u>78.26</u>
TELC-PPI	89.37	73.45	71.59	92.63	80.77	<u>73.61</u>	96.94	87.98	78.73

Table 1: Comparison of the overall performance on STRING and its subsets, measured by F1-score, where **bold** entries and underlined entries indicate the best and second-best results, respectively, and OOM denotes out-of-memory errors.

Method	SHS27k	SHS148k	STRING	Average
w/o SE	89.09	92.29	96.24	92.54 (-0.44)
w/o TI	89.15	<u>92.50</u>	<u>96.76</u>	92.80 (-0.18)
w/o NS	88.77	92.43	96.58	92.59 (-0.39)
w/o LE	<u>89.21</u>	92.49	96.72	<u>92.81</u> (-0.17)
TELC-PPI	89.37	92.63	96.94	92.98

Table 2: Ablation study of modules on three datasets.

(1) Graph-based methods generally outperform traditional deep learning approaches, which may be attributed to their ability to better exploit the topological structure of the PPI network. (2) Our model outperforms baselines on many datasets and under various split settings, with over 1% F1-score gains in the majority of DFS and BFS settings, highlighting its effectiveness and robustness to distribution shifts. (3) An exception occurs in the BFS split of the SHS148k dataset, where our model ranks second-best. This may be due to the introduction of noise during the construction of structural enhancement edges (H2 edges), which slightly impairs performance.

Ablation Study. To evaluate the contributions of the structure enhancement and label embedding modules, we conducted ablation studies under the random split on all datasets. We compared the full **TELC-PPI** model with four variants: (1) **w/o TE** – removes the topology enhancement module, using the original PPI graph; (2) **w/o TI** – excludes topological information in edge sampling by setting $\alpha = 0$; (3) **w/o NS** – ignores node similarity in sampling by setting $\alpha = 1$; and (4) **w/o LE** – removes the label embedding module, using edge representations directly for prediction. Results in Table 2, measured by F1-score, yield the following observations:

(1) Removing any module leads to a performance drop, demonstrating the effectiveness of each component. (2) The topological enhancement module contributes more to overall performance than the label embedding module. (3) Node similarity proves more critical than topological information in scoring candidate H2 edges.

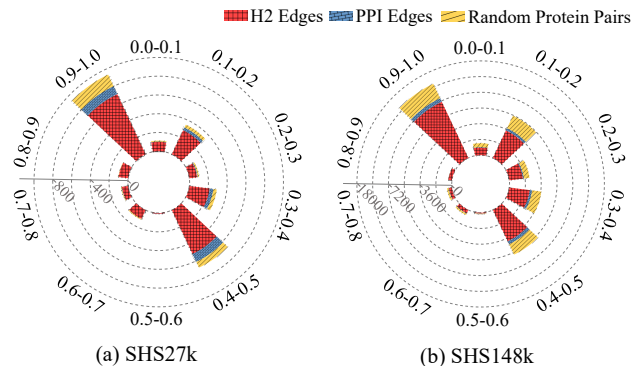


Figure 3: Distribution of Jaccard similarities of functional domains among different groups of protein pairs.

Edge Connection Quality Evaluation. To validate the biological relevance of H2 edges, we retrieved functional domain annotations for proteins in SHS27k and SHS148k from UniProt (Consortium 2023), yielding 646 and 1,914 proteins with valid annotations. Then, we formed three protein-pair groups: (1) pairs corresponding to original PPI edges, (2) H2-connected pairs, and (3) random pairs matched in size to the H2-connected pairs. Jaccard similarity of functional domains was computed for each group and discretized into 10 bins over $[0.0, 1.0]$. Figure 3 presents radial stacked bar charts for comparison, revealing the following observations:

(1) Across all Jaccard similarity intervals, H2-connected pairs significantly outnumber original PPI edges and random pairs, supporting their biological relevance. (2) H2 edges capture functional associations across a range of similarities, with peaks at $[0.1, 0.2)$, $[0.4, 0.5)$, and notably $[0.9, 1.0]$, indicating their ability to identify both weak and strong functional correlations.

Label Embeddings Similarity Analysis. To evaluate whether the learned label embeddings capture interaction-type co-occurrence relationships, we visualize pairwise similarities among the seven PPI labels using chord diagrams on SHS27k and SHS148k. Specifically, we compute co-

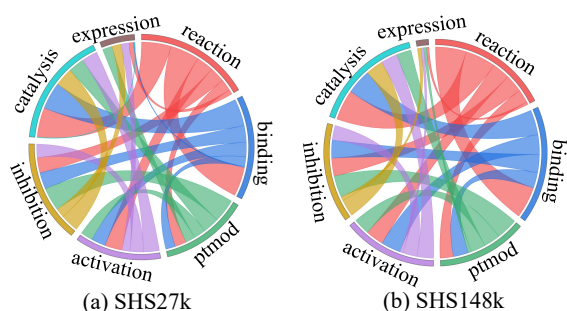


Figure 4: Label embedding similarities on (a) SHS27k and (b) SHS148k datasets.

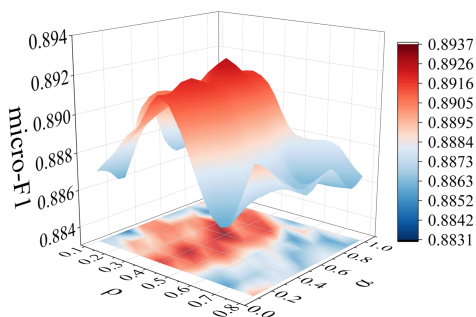


Figure 5: Hyperparameter sensitivity analysis of model performance with respect to the hyperparameters α and ρ .

sine similarities between label embeddings and map them to chord diagrams, where edge thickness indicates similarity strength. Figure 4 presents the visualizations, from which we derive the following observations:

(1) The learned embeddings capture core biological relationships. Across SHS27k and SHS148k, *reaction*, *binding*, and *catalysis* exhibit strong embedding similarity, consistent with the co-occurrence patterns in Figure 1(b), confirming the model’s ability to encode fundamental biological dependencies. (2) The representations exhibit good generalizability. Relative similarity strengths among most interaction types remain stable across datasets, indicating that the embeddings capture intrinsic label correlations rather than dataset-specific artifacts. (3) Rare labels show sample-size sensitivity. The *expression* label displays notable divergence between datasets, likely due to its limited instances in SHS27k (622 samples), suggesting that reliable embeddings require sufficient observations.

Hyperparameter Sensitivity Analysis. We conducted a grid search over the topological-semantic trade-off parameter α and sampling ratio ρ on SHS27k, evaluating micro F1-scores. Figure 5 presents the results and key trends from this sensitivity analysis.

(1) When the sampling ratio $\rho \in \{0.3, 0.4, 0.5\}$, the model achieves consistently good and stable performance across different values of α , indicating robustness to α under certain conditions. (2) When $\alpha \in \{0.3, 0.4, 0.5\}$ and $\rho \in \{0.3, 0.4\}$, the model reaches a performance peak with an F1-score around 0.8937, suggesting that a moder-

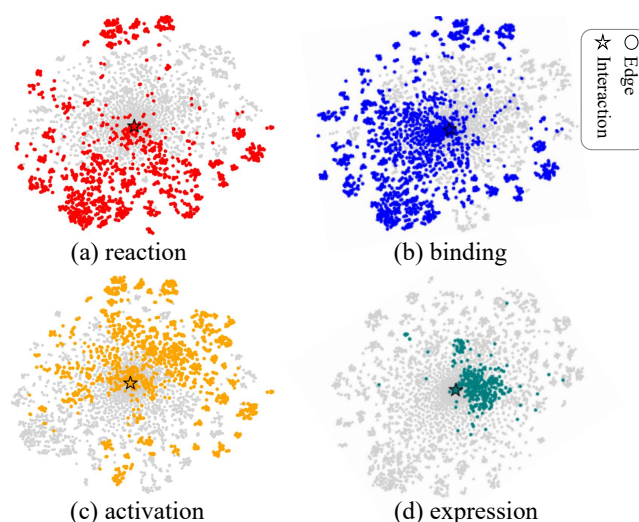


Figure 6: Visualization of label embeddings and their corresponding edge representations for four interaction types: (a) *reaction*, (b) *binding*, (c) *activation*, and (d) *expression*.

ate trade-off and sampling rate lead to the most informative edge augmentation. (3) When ρ is too small (e.g., 0.1) or too large (e.g., 0.8), performance slightly decreases, likely due to insufficient augmentation or the introduction of noisy edges, respectively.

Visualization of Edge Representations and Label Embeddings.

To analyze spatial relationships between edge representations and label embeddings, we visualize their t-SNE projections on SHS27k. Since each edge may represent multiple interaction types, we generate subfigures corresponding to each label embeddings. Each subfigure in Figure 6 highlights a specific interaction type, where edge representations are shown as colored circles and the corresponding label embedding as a colored star (e.g., *reaction*: red circles/star; *binding*: blue circles/star). Unrelated edges are shown in gray. We make the following key observations:

(1) The learned label embeddings cluster near the centroid of the edge representation space, indicating a unified embedding space across interaction types. (2) Distinct spatial separation among labels, e.g., *activation* in upper regions and *expression* in center-right areas, highlights functional specialization preserved in the embedding space.

Conclusion

In this paper, we propose TELC-PPI, a framework for protein–protein interaction prediction that integrates topology enhancement and label correlation modeling. It employs a topology enhancement module based on node similarity and topology information, alongside a label embedding module capturing inter-label dependencies. Our model achieved superior performance across multiple datasets and various data split settings, demonstrating its effectiveness. However, the current H2 edge construction may introduce noise in complex cases, and refining this process to better capture protein similarity is an interesting future direction.

Acknowledgments

This study is supported by Gansu Province Key Foundation Project (No. 24JRRA123) and National Natural Science Foundation of China (No.62567007 and 62441701).

References

- Acuner Ozbabacan, S. E.; Engin, H. B.; Gursoy, A.; and Keskin, O. 2011. Transient Protein-Protein Interactions. *Protein Engineering, Design & Selection*, 24(9): 635–648.
- Bei, Y.; Chen, W.; Chen, H.; Zhou, S.; Yang, C.; Fan, J.; Huang, L.; and Bu, J. 2025. Correlation-Aware Graph Convolutional Networks for Multi-Label Node Classification. In *Proceedings of the ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, 37–48.
- Chen, M.; Ju, C. J.-T.; Zhou, G.; Chen, X.; Zhang, T.; Chang, K.-W.; Zaniolo, C.; and Wang, W. 2019. Multifaceted Protein-Protein Interaction Prediction Based on Siamese Residual RCNN. *Bioinformatics*, 35(14): i305–i314.
- Chua, H. N.; Sung, W.-K.; and Wong, L. 2006. Exploiting indirect neighbours and topological weight to predict protein function from protein-protein interactions. *Bioinformatics*, 22(13): 1623–1630.
- Clegg, R. M. 1995. Fluorescence Resonance Energy Transfer. *Current Opinion in Biotechnology*, 6(1): 103–110.
- Consortium, T. U. 2023. UniProt: the Universal Protein Knowledgebase in 2023. *Nucleic Acids Research*, 51(D1): D523–D531.
- Du, G.; Zhang, J.; Zhang, N.; Wu, H.; Wu, P.; and Li, S. 2024. Semi-supervised imbalanced multi-label classification with label propagation. *Pattern Recognition*, 150: 110358.
- Durham, J.; Zhang, J.; Humphreys, I. R.; Pei, J.; and Cong, Q. 2023. Recent Advances in Predicting and Modeling Protein-Protein Interactions. *Trends in Biochemical Sciences*, 48(6): 527–538.
- Feng, L.; An, B.; and He, S. 2019. Collaboration Based Multi-Label Learning. In *Proceedings of the AAAI Conference on Artificial Intelligence*, 3550–3557.
- Gao, K.; Zhang, J.; and Zhou, C. 2019. Semi-supervised Graph Embedding for Multi-label Graph Node Classification. In *Proceedings of the International Conference on Web Information Systems Engineering*, 555–567.
- Gao, Z.; Jiang, C.; Zhang, J.; Jiang, X.; Li, L.; Zhao, P.; Yang, H.; Huang, Y.; and Li, J. 2023. Hierarchical Graph Learning for Protein-Protein Interaction. *Nature Communications*, 14(1): 1093.
- Hang, J.-Y.; and Zhang, M.-L. 2021. Collaborative Learning of Label Semantics and Deep Label-Specific Features for Multi-Label Classification. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 44(12): 9860–9871.
- Hashemifar, S.; Neyshabur, B.; Khan, A. A.; and Xu, J. 2018. Predicting Protein-Protein Interactions Through Sequence-Based Deep Learning. *Bioinformatics*, 34(17): i802–i810.
- Jumper, J.; Evans, R.; Pritzel, A.; Green, T.; Figurnov, M.; Ronneberger, O.; Tunyasuvunakool, K.; Bates, R.; Židek, A.; Potapenko, A.; et al. 2021. Highly Accurate Protein Structure Prediction with AlphaFold. *Nature*, 596(7873): 583–589.
- Kodama, Y.; and Hu, C.-D. 2012. Bimolecular Fluorescence Complementation (BiFC): A 5-Year Update and Future Perspectives. *Biotechniques*, 53(5): 285–298.
- Kovács, I. A.; Luck, K.; Spirohn, K.; Wang, Y.; Pollis, C.; Schlabach, S.; Bian, W.; Kim, D.-K.; Kishore, N.; Hao, T.; et al. 2019. Network-Based Prediction of Protein Interactions. *Nature Communications*, 10(1): 1240.
- Kurata, G.; Xiang, B.; and Zhou, B. 2016. Improved Neural Network-Based Multi-Label Classification with Better Initialization Leveraging Label Co-Occurrence. In *Proceedings of the Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, 521–526.
- Li, H.; Chen, Z.; Li, Z.; Zheng, Q.; Zhang, P.; and Zhou, S. 2023. GIPA: A General Information Propagation Algorithm for Graph Learning. In *Proceedings of the International Conference on Database Systems for Advanced Applications*, volume 13973 of *Lecture Notes in Computer Science*, 465–476.
- Li, H.; Gong, X.-J.; Yu, H.; and Zhou, C. 2018. Deep Neural Network Based Predictions of Protein Interactions Using Primary Sequences. *Molecules*, 23(8): 1923.
- Lin, Y.; Chen, M.; Zhang, K.; Li, H.; Li, M.; Yang, Z.; Lv, D.; Lin, B.; Liu, H.; and Cai, D. 2024. TagCLIP: A Local-to-Global Framework to Enhance Open-Vocabulary Multi-Label Classification of CLIP Without Training. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 38, 3513–3521.
- Lv, G.; Hu, Z.; Bi, Y.; and Zhang, S. 2021. Learning Unknown from Correlations: Graph Neural Network for Inter-novel-protein Interaction Prediction. In *Proceedings of the International Joint Conference on Artificial Intelligence*, 3677–3683.
- Ma, Y.; Liu, X.; Zhao, T.; Liu, Y.; Tang, J.; and Shah, N. 2021. A Unified View on Graph Neural Networks as Graph Signal Denoising. In *Proceedings of the ACM International Conference on Information and Knowledge Management*, 1202–1211.
- Ridnik, T.; Ben-Baruch, E.; Zamir, N.; Noy, A.; Friedman, I.; Protter, M.; and Zelnik-Manor, L. 2021. Asymmetric Loss for Multi-Label Classification. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 82–91.
- Saidabad, M. Y.; Hassanzadeh, H.; Ebrahimi, S. H. S.; Khezri, E.; Rahimi, M. R.; and Trik, M. 2024. An Efficient Approach for Multi-Label Classification Based on Advanced Kernel-Based Learning System. *Intelligent Systems with Applications*, 21: 200332.
- Shi, M.; Tang, Y.; and Zhu, X. 2019. MLNE: Multi-Label Network Embedding. *IEEE Transactions on Neural Networks and Learning Systems*, 31(9): 3682–3695.
- Soleymani, F.; Paquet, E.; Viktor, H.; Michalowski, W.; and Spinello, D. 2022. Protein-Protein Interaction Prediction

with Deep Learning: A Comprehensive Review. *Computational and Structural Biotechnology Journal*, 20: 5316–5341.

Szklarczyk, D.; Gable, A. L.; Lyon, D.; Junge, A.; Wyder, S.; Huerta-Cepas, J.; Simonovic, M.; Doncheva, N. T.; Morris, J. H.; Bork, P.; et al. 2019. STRING v11: Protein–Protein Association Networks with Increased Coverage, Supporting Functional Discovery in Genome-Wide Experimental Datasets. *Nucleic Acids Research*, 47(D1): D607–D613.

Tang, T.; Zhang, X.; Liu, Y.; Peng, H.; Zheng, B.; Yin, Y.; and Zeng, X. 2023. Machine Learning on Protein–Protein Interaction Prediction: Models, Challenges and Trends. *Briefings in Bioinformatics*, 24(2): bbad076.

Wang, T.; Jin, D.; Wang, R.; He, D.; and Huang, Y. 2022. Powerful Graph Convolutional Networks with Adaptive Propagation Mechanism for Homophily and Heterophily. In *Proceedings of the AAAI Conference on Artificial Intelligence*, 4210–4218.

Wang, W.; and Zhang, M.-L. 2023. Label Specific Multi-Semantics Metric Learning for Multi-Label Classification: Global Consideration Helps. In *Proceedings of the International Joint Conference on Artificial Intelligence*, 3193–3199.

Wu, L.; Tian, Y.; Huang, Y.; Li, S.; Lin, H.; Chawla, N. V.; and Li, S. Z. 2024. MAPE-PPI: Towards Effective and Efficient Protein-Protein Interaction Prediction via Microenvironment-Aware Protein Embedding. In *Proceedings of the International Conference on Learning Representations*, 1–13.

Yu, Z.-B.; and Zhang, M.-L. 2021. Multi-Label Classification with Label-Specific Feature Generation: A Wrapped Approach. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 44(9): 5199–5210.

Zhao, Z.; Qian, P.; Yang, X.; Zeng, Z.; Guan, C.; Tam, W. L.; and Li, X. 2023. SemiGNN-PPI: Self-Ensembling Multi-Graph Neural Network for Efficient and Generalizable Protein-Protein Interaction Prediction. In *Proceedings of the International Joint Conference on Artificial Intelligence*, 4984–4992.

Zhou, M.; Li, Q.; and Wang, R. 2016. Current Experimental Methods for Characterizing Protein–Protein Interactions. *ChemMedChem*, 11(8): 738–756.