Article

Predicting Protein–Protein Interactions via Gated Graph Attention Signed Network

Zhijie Xiang [1], Weijia Gong [1], Zehui Li [1], Xue Yang [1], Jihua Wang [1] and Hong Wang [1,2,*]

[1] School of Information Science and Engineering, Shandong Normal University, Jinan 250014, China; 201811010415@stu.sdnu.edu.cn (Z.X.); 201811990134@stu.sdnu.edu.cn (W.G.); 201811010322@stu.sdnu.edu.cn (Z.L.); 2020020935@stu.sdnu.edu.cn (X.Y.); Wangjihua@sdnu.edu.cn (J.W.)

[2] Shandong Provincial Key Laboratory for Distributed Computer Software Novel Technology, Shandong Normal University, Jinan 250014, China

[*] Correspondence: 111052@sdnu.edu.cn

Abstract: Protein–protein interactions (PPIs) play a key role in signal transduction and pharmacogenomics, and hence, accurate PPI prediction is crucial. Graph structures have received increasing attention owing to their outstanding performance in machine learning. In practice, PPIs can be expressed as a signed network (i.e., graph structure), wherein the nodes in the network represent proteins, and edges represent the interactions (positive or negative effects) of protein nodes. PPI predictions can be realized by predicting the links of the signed network; therefore, the use of gated graph attention for signed networks (SN-GGAT) is proposed herein. First, the concept of graph attention network (GAT) is applied to signed networks, in which “attention” represents the weight of neighbor nodes, and GAT updates the node features through the weighted aggregation of neighbor nodes. Then, the gating mechanism is defined and combined with the balance theory to obtain the high-order relations of protein nodes to improve the attention effect, making the attention mechanism follow the principle of “low-order high attention, high-order low attention, different signs opposite”. PPIs are subsequently predicted on the Saccharomyces cerevisiae core dataset and the Human dataset. The test results demonstrate that the proposed method exhibits strong competitiveness.

Keywords: protein–protein interactions (PPIs); PPI signed network; link sign prediction; attention mechanism; gating mechanism

Citation: Xiang, Z.; Gong, W.; Li, Z.; Yang, X.; Wang, J.; Wang, H. Predicting Protein–Protein Interactions via Gated Graph Attention Signed Network. *Biomolecules* **2021**, *11*, 799. https://doi.org/10.3390/biom11060799

Academic Editor: Brigita Urbanc

Received: 1 May 2021
Accepted: 26 May 2021
Published: 28 May 2021

1. Introduction

Proteins inside cells do not function alone, and they must interact with other proteins to perform their functions. Therefore, studying protein–protein interactions (PPIs) is necessary for understanding various biological processes in cells, such as gene transcription, that involve multiple protein interactions. The accurate prediction of unknown PPIs reveals the function of proteins at the molecular level and is critical for revealing life activity rules, e.g., growth, development, differentiation, and apoptosis. In addition, accurate PPI prediction provides an important theoretical basis for discussing the mechanisms of major diseases, disease treatment, disease prevention, and new drug development.

PPIs control nearly all cellular processes and play an important role in the execution of various physiological functions. Therefore, PPI prediction has been extensively studied, and as such, many methods have been proposed, including biological experimental methods and calculation methods. Currently, the experimental methods for PPI identification mainly include affinity purification mass spectrometry (AP-MS) [1] and yeast two-hybrid system (Y2H) [2]. In recent decades, proteomics based on mass spectrometry (MS) has become an important technique for identifying PPIs. One method of AP-MS is to label the cells of the experimental group and the control group, respectively, by the method of stable isotope labeling with amino acids in cell cultures (SILAC) [3], and then carry out a co-immunoprecipitation (Co-IP) [4] experiment. The immune complex is separated by the specific reaction between the antigen and antibody, and then the protein in the immune complex is detected by liquid chromatography tandem mass spectrometry (LC-MS/MS). When the content of a protein in the experimental group and the control group reaches statistical difference, it can be observed that the protein interacts with the studied protein. This method can greatly reduce the possibility of false positive results of protein–protein interaction. Y2H was initially established by Fields et al. [2] when they studied the properties of yeast transcription factor Gal4. After continuous improvement, Y2H has developed into a mature protein–protein interaction research tool. Y2H is a system in which the two proteins to be studied are cloned into a DNA-binding domain (DNA-BD) and activation domain (AD) of a transcription activator (Gal4, etc.) of yeast expression plasmid, respectively, to construct a fusion expression vector, and then the interaction between the two proteins is analyzed from the expression products. Y2H can also sensitively detect weak and transient interactions between proteins through the expression products of reporter genes. This method is a highly sensitive technology to study the relationship between proteins. In addition, there are many other experimental methods for PPI prediction, such as phage display technology [5] protein chip technology [6], and surface plasmon resonance technology [7]. However, detecting PPIs in biological experiments is inefficient, time-consuming, and laborious; thus, such methods are unsuitable for large-scale PPI detection.

In recent years, structure- and sequence-based PPI prediction calculation methods have been proposed. Structure-based prediction methods are constrained by the experimentally determined protein structure, and sequence-based prediction methods include statistical and machine learning methods. Statistics-based prediction methods primarily include the mirror tree [8] and the co-evolutionary differences (CD) [9]. The mirror tree method is based on observing the correspondence between the phylogenetic trees of related proteins in systems, such as ligands and receptors, and employs a new method to discover possible protein interactions by comparing the evolutionary distances between related protein family sequences. The co-evolutionary differences method does not use multivariate comparison. Thus, it requires less time than other comparison methods. Prediction methods based on machine learning primarily include autocovariance (AC) and support vector machine (SVM) [10], similarity comparison [11], the amino acid composition (AAC) method [12], universal in silico predictor of PPIs (UNISPPI) [13] and the ETB-Viterbi [14]. The advantage of the AC and SVM method is that autocovariance contains the remote interaction information of amino acid residues, which is crucial in PPI identification. The similarity comparison method predicts PPIs according to the pairwise similarity of the primary protein structure, and the amino acid composition method is suitable for any protein sequence, particularly when domain information is lacking. The universal in silico predictor of the PPI method uses a small number of features to train decision tree classifiers. The advantages of this method are low calculation cost and simple implementation. The disadvantage is that decision tree classifiers typically suffer overfitting problems. The ETB-Viterbi method can capture long-distance correlation to improve prediction accuracy and is unaffected by the sequence direction.The calculation methods for predicting PPIs can not only compute large-scale protein interaction data but also have advantages of high accuracy and low cost.

In the real world, many application scenarios can be represented as graphs or networks, wherein nodes represent entities, and edges represent the relationships among entities. Compared with the traditional network that exclusively comprises positive edges, the signed network can express more abundant semantic information and more accurate expression of the actual scene; thus, signed networks are used widely, e.g., in Epinions consumer review networks, Slashdot news review networks, and organizations or groups interaction networks [15]. The edge signs signify the user's emotional tendency (e.g., like or do not like) regarding a comment. Inspired by this, we express PPIs as a signed network, wherein nodes in the network represent proteins, and connecting edges represent positive or negative interactions between proteins, where positive means the presence and negative means the absence of interaction. The PPI prediction task can be transformed into a link prediction task for the PPI signed network created herein. Research has revealed that the negative edge of a signed network has an important impact on the network structure and node representation [16]. The positive edge renders nodes increasingly similar, whereas the negative edge renders nodes increasingly different. Therefore, in the analysis and study of PPI networks, positive and negative edges must be comprehensively analyzed. However, the structure of a signed network is more complex than that of traditional unsigned networks. In the low-dimensional node representation space of signed networks, the distance between two nodes connected by a positive edge should be less than that of two nodes connected by a negative edge. The traditional representation learning method of unsigned networks cannot be applied directly to signed networks; consequently, researching PPIs based on signed networks is extremely challenging.

Currently, graph neural networks (GNNs) have attracted significant research interest in the field of deep learning, particularly in machine learning tasks, such as link prediction [17,18]. GNNs introduce a neural network into graph data by defining convolution [19] and attention [20]. After the graph convolution network (GCN) was proposed, Velickovic et al. further proposed the use of GAT [20], which utilizes attention coefficients (i.e., weight coefficients) to aggregate the features of neighbor nodes to the central node and uses local stationarity on the graph to learn new node feature expressions. By learning the weight of neighbors, GAT can realize the weighted aggregation of neighbors, rendering GAT more robust to noisy neighbors and giving the model certain interpretability via the attention mechanism [20]. To some extent, GAT is more powerful than GCN because, in the GAT method, the correlation between node features is better integrated into the model. In addition, GAT can not only simultaneously handle multiple nodes of different degrees that are highly parallel, but it also does not rely on graph structure information, thus making it exhibit strong generalizability. Furthermore, the GAT model can be applied to graph-based inductive and transductive learning problems effectively. However, traditional GAT cannot compute a network with negative edges and can only aggregate and update nodes on an unsigned network. A network comprising PPIs is a signed network; therefore, the above limitation of GAT inspired our research on its application to signed networks. In addition, GAT has scope for improvement: first, GAT only considers the influence of low-order neighbors and ignores the influence of high-order neighbors. Second, the balance theory is critical in signed network research; however, it has not been integrated into GAT-based methods. Overall, current GAT-based methods do not have the expected features, i.e., focusing on the sign of the edges, considering high-order neighbor information, and the interpretability of the balance theory.

To overcome the limitation of the inapplicability of GAT to signed networks, the attention mechanism must be improved to effectively predict the unknown links on the signed networks to realize PPI prediction on PPI signed networks. This is the primary focus of the present study.

To solve these problems, the method of gated graph attention for signed networks (SN-GGAT) is proposed herein and subsequently used to predict the links in a PPI network. The primary contributions of this study are summarized as follows:

1. PPIs are transformed into a signed network with rich semantic information, and the proposed SN-GGAT method is applied to the network, thus enabling the accurate prediction of unknown sign information (i.e., interactions) between proteins.
2. We propose the application of GAT to signed networks while retaining the original advantage of GAT and expressing the polarity of the edge more accurately.
3. We define a gating mechanism to determine high-order neighbors that affect a node and innovatively define an attention mechanism to demonstrate that the low-order neighbors of a node have a relatively greater influence on a node. In comparison, high-order neighbors have less influence.
4. To obtain a good interpretative network embedding, we strengthen the constraint of the balance theory on the sign propagation process.

2. Materials and Methods

2.1. Materials

2.1.1. *Saccharomyces cerevisiae* Core Dataset

One of the PPI datasets used in this study was taken from the Saccharomyces cerevisiae core dataset (version: Scere20170205) in the database of interacting proteins (DIP) [21]. The 5470 protein pairs in the dataset were used as the positive sample set in the test. The *S. cerevisiae* core dataset is available online (https://dip.doe-mbi.ucla.edu/dip/Download.cgi?SM=7&TX=4932, accessed on 10 February 2021).

Since the non-interacting pairs were not readily available, we drew on the strategies in the literature [10,22,23] to construct the negative sample set. Here, we followed the assumption that there is no interaction between proteins at different subcellular localizations in the cell, i.e., we randomly paired proteins with different subcellular location information, and the resulting protein pair was considered non-interactive. The subcellular localization information of proteins can be obtained from the UniProt database [24]. Note that there is a small amount of subcellular localization information of proteins that cannot be queried, and this information was excluded during random matching. The UniProt database is available online (https://www.uniprot.org/, accessed on 10 February 2021). To ensure data balance, we constructed 5470 negative edges for the PPI network by random pairing. In total, the dataset contained 10,940 pairs of samples (50% positive samples and 50% negative samples).

2.1.2. Human Dataset

The second dataset we used is the Human dataset. The HIPPIE database collects human PPIs with experimental annotations [25]. By considering factors such as the number of publications or the type of experimental support, the reliability score of the interaction was scored. Later, Hampe et al. [26] chose 10% of the highest scoring interactions from the HIPPIE dataset to obtain a high-quality human protein subset. In order to obtain fair comparison results, Liu et al. [27] followed the same strategy as DPPIs [28] on the basis of high-quality human protein subset, eliminating the redundancy of the Human dataset, so that no two PPIs are similar at the sequence level. If at least two sequences have more than 40% sequence identity, two PPIs are considered to be similar.

In this paper, we used the Human dataset used by Liu et al. as the second test dataset, the Human dataset is available online (https://zenodo.org/record/3960077/files/Human.zip?download=1, accessed on 31 March 2021).

2.2. Related Definition

Here, we describe relevant definitions involved in the proposed SN-GGAT method. First, we describe some concepts of the basic PPI signed network. Then, we define node update rules for PPI signed networks.

2.2.1. PPI Signed Network

A PPI signed network is a graph structure that can be described by $G = (V, E)$, where V is the set of all protein nodes in graph G, and E is the set of edges between any two protein nodes in graph G. Here, the edges between protein nodes i and j are denoted as $e(i, j)$ and $e(i, j) \in E$, respectively; $e(i, j) = 1$ represents a positive interaction between protein nodes i and j; $e(i, j) = 0$ represents an unknown interaction between protein nodes i and j; and $e(i, j) = -1$ represents a negative interaction between protein nodes i and j.

For example, we can learn from the Drosophila melanogaster signed PPI database [29] constructed by Arunachalam V et al. that the interaction between Ribosomal protein LP0 (FlyBase ID: FBgn0000100) and Ataxin-2 protein (FlyBase ID: FBgn0041188) is positive; however, the interaction between Ribosomal protein LP0 and Dodeca-satellite-binding protein 1 (FlyBase ID: FBgn0027835) is negative. Thus, the PPI signed network comprising the Ribosomal protein LP0, Ataxin-2 protein, and Dodeca-satellite-binding protein 1 can be represented as shown in Figure 1. The signed PPI database is available online (https://www.flyrnai.org/SignedPPI/Download.jsp, accessed on 25 January 2021).

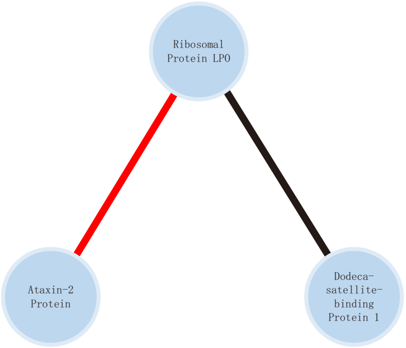


Figure 1. An example of positive and negative interactions between protein nodes. The red edge indicates positive interaction and the black edge indicates negative interaction.

$$A(i,j) = \begin{cases} -1, \; if \; e(i,j) = -1 \\ 0, \; if \; e(i,j) = 0 \\ 1, \; if \; e(i,j) = 1 \end{cases} \tag{1}$$

To define and explain the adjacency matrix A of the PPI signed network more intuitively, we selected seven protein nodes in a PPI network as an example and label these nodes with serial numbers. Simultaneously, we selected part of the edges between these protein nodes to form the PPI signed network (Figure 2). Adjacency matrix A is defined in Equation (1), and the adjacency matrix corresponding to Figure 2 is expressed by Equation (2). Here, the link sign prediction task must replace the unknown zeros in the adjacency matrix with the predicted signs.

$$A = \begin{pmatrix} 1 & 1 & 0 & -1 & 1 & 0 & 1 \\ 1 & 1 & 0 & 1 & -1 & 0 & -1 \\ 0 & 0 & 1 & -1 & 0 & 0 & 0 \\ -1 & 1 & -1 & 1 & 0 & 0 & -1 \\ 1 & -1 & 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 & 0 \\ 1 & -1 & 0 & -1 & 0 & 0 & 1 \end{pmatrix} \tag{2}$$

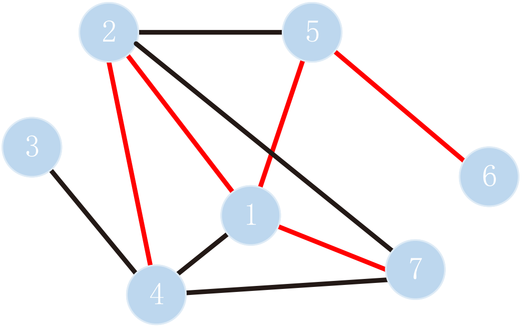


Figure 2. An example of protein–protein interaction (PPI) network. The PPI network consists of seven protein nodes and some connected edges.

2.2.2. Balance Theory

Heider proposed a structural balance model that can use positive and negative links to describe protein interactions or user relationships. The balance theory originally began with analyzing the balance of triangles in signed networks and has since been applied to large-scale link sign prediction. This theory considers all possible combinations of the triplet comprising three nodes, forming four intuitive understandings—a friend of a friend is a friend, an enemy of a friend is an enemy, a friend of an enemy is an enemy, and an enemy of an enemy is a friend.

According to the structural balance theory, if the product of the signs on the three edges of a triangle is positive, the triangle is structurally balanced; otherwise, it is unbalanced. Here, if we use "+" to represent positive edges, i.e., the positive relationship between two nodes, and "−" to represent the negative edges, i.e., the negative relationship between two nodes, then the structural network of the two triangles in Figure 3a,b is balanced. In contrast, the triangular structural networks in Figure 3c,d are unbalanced. Here, it is worth noting that there are times when the balance theory does not hold in real life, such as G-proteins, which play the role of molecular switches in the process of signal transduction. G-proteins do physically interact with several other proteins, but these proteins do not interact directly. Therefore, this paper states that balance theory has some limitations in specific biological networks. However, in a broad sense, in real life, compared with unbalanced triangles, there are more balanced triangles [30]. In addition, the balance theory has achieved remarkable results in our method. Therefore, we carefully considered balanced triangles in our test.

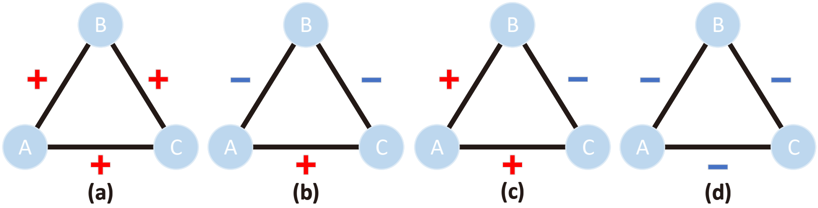


Figure 3. Examples of balanced and unbalanced triangles, where (**a**,**b**) are balanced triangles, and (**c**,**d**) are unbalanced triangles.

2.2.3. PPI Signed Accessibility Matrix

If there are s paths of length k between proteins i and j, the meaning of length is the number of edges that cross other proteins between two proteins. According to the balance theory, if p positive edges and q negative edges are obtained, and $p + q = s$, then the value of the i-th row and j-th column of the k-order PPI signed accessibility matrix M_k is $M_k(i, j)$. This is calculated using Equation (3), where the return value of the $sgn(x)$ function is 1, −1, or 0 when x is positive, negative, or zero, respectively.

$$M_k(i, j) = \text{sgn}(p - q) \tag{3}$$

The 0-order accessibility matrix M_0 and the first-order accessibility matrix M_1 are defined first. M_0 and M_1 are defined in Equations (4) and (5), respectively.

$$M_0 = I \tag{4}$$
$$M_1 = A \tag{5}$$

M_0 is equal to the identity matrix, i.e., every node can reach itself in zero steps, and every node is positively related to itself. M_1 is equal to the adjacency matrix A because A reflects nodes such that each node can reach itself in a single step, which exactly matches the definition of M_1.

In addition, we can derive the second-order accessibility matrix M_2 expression as follows.

$$M_2 = \text{sgn}(M_1 \cdot M_1) = \text{sgn}\left(\sum_{k=1}^{n} M_1(i,k) \cdot M_1(k,j)\right) \tag{6}$$

According to Equation (6), the second-order accessibility matrix M_2 corresponding to Figure 2 can be obtained as follows.

$$M_2 = \begin{pmatrix} 1 & -1 & 1 & -1 & 1 & 1 & 1 \\ -1 & 1 & -1 & 1 & -1 & -1 & -1 \\ 1 & -1 & 1 & -1 & 0 & 0 & 1 \\ -1 & 1 & -1 & 1 & -1 & 0 & -1 \\ 1 & -1 & 0 & -1 & 1 & 1 & 1 \\ 1 & -1 & 0 & 0 & 1 & 1 & 0 \\ 1 & -1 & 1 & -1 & 1 & 0 & 1 \end{pmatrix} \tag{7}$$

Similarly, the third-order accessibility matrix is given in Equation (8). We further derive the expression of the n-order accessibility matrix in Equation (9).

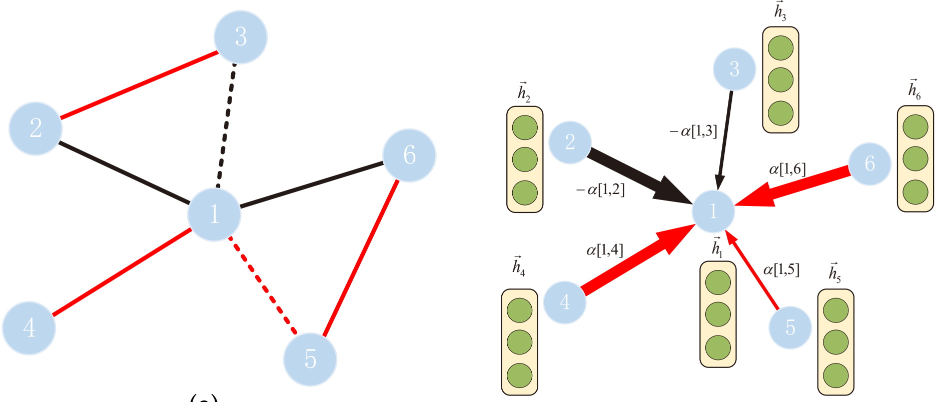
$$M_3 = \mathrm{sgn}(M_2 \cdot M_1) \tag{8}$$

$$M_n = \mathrm{sgn}(M_{n-1} \cdot M_1) \tag{9}$$

2.2.4. Node Update Rules for PPI Signed Network

The nodes in a PPI signed network are affected by their m-order neighbors when they update. Using the attention weight, the features of the m-step accessible neighbor nodes are aggregated to the central node to realize node updates. Of these m-order nodes, the attention weight of low-order neighbors is relatively high, and the attention weight symbols of positive and negative links are opposite.

In the PPI signed network shown in Figure 4a, the solid line represents the first-order neighbor, and the dotted line represents the second-order neighbor. If the value of m in the above definition is 2, the update rule of node 1 is shown in Figure 4b. Here, node 1 is affected by its first- and second-order neighbors. The thickness of the arrow represents the relative size of the attention weight. Note that attention α is used to aggregate the neighbor nodes to update node 1.



(**a**) (**b**)

Figure 4. (**a**) PPI signed network, where the solid line represents the first-order neighbor and the dotted line represents the second-order neighbor. (**b**) Update rules for node 1, where the thickness of the arrow represents the relative size of the attention weight.

The specific algorithm and implementation of the attention mechanism and node update are introduced in detail in Section 2.3.

2.3. Proposed SN-GGAT Method

Here, we describe the implementation of the proposed SN-GGAT model in detail, including the model structure, gating mechanism, attention mechanism, and algorithm implementation.

2.3.1. Model Framework

The model structure of SN-GGAT is shown in Figure 5, which includes the following four parts.

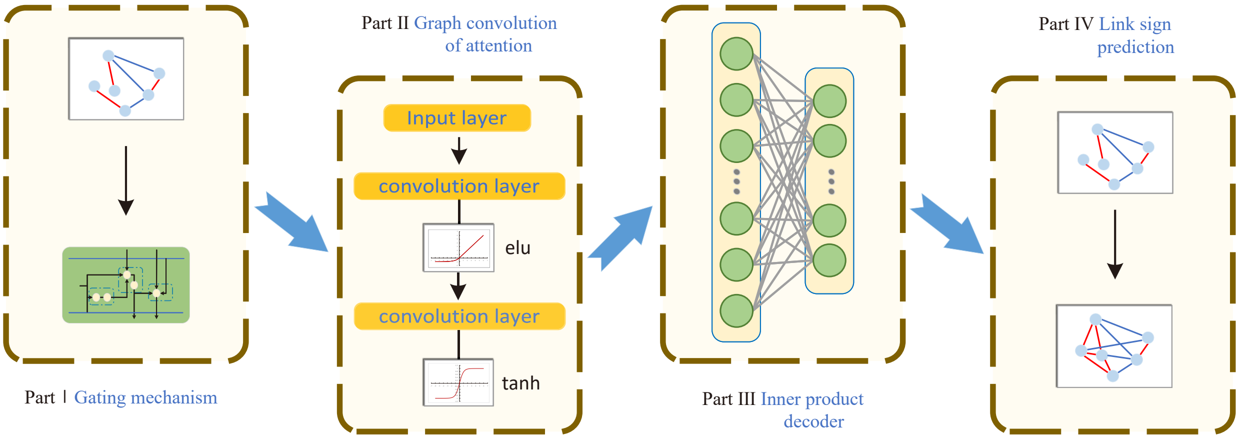

Part Ⅰ Gating mechanism
Part II Graph convolution of attention
Input layer
convolution layer
elu
convolution layer
tanh
Part III Inner product decoder
Part IV Link sign prediction

Figure 5. Framework of the proposed SN-GGAT model. Part I uses the gating mechanism to calculate the parameters required by the algorithm; part II executes the algorithm to obtain the feature representation of each node; part III uses the inner product decoder to obtain the reconstructed adjacency matrix; part IV uses the prediction result to verify the accuracy of the model.

In Part I, we calculated the adjacency matrix A, low-order memory accessibility matrix M'_m, and low-order attention accessibility matrix $\widetilde{M}_m$. These were fed into the gating mechanism, which outputs the corresponding high-order memory accessibility matrix M'_{m+1} and high-order attention accessibility matrix $\widetilde{M}_{m+1}$ as the input for Part II.

In part II, we considered M'_{m+1} as the adjacency matrix used by the model. Here, 0 in M'_{m+1} represent boundless and a non-zero value represents an edge. The model assigns an attention coefficient to all edges. The role of $\widetilde{M}_{m+1}$ is to modify and update these attention coefficients such that attention follows the following rules: positive edges are positive, negative edges are negative, low-order neighbor high attention, and high-order neighbor low attention. With these definitions, we employed two convolutional layers to train the node vector's feature representation. Finally, we added a nonlinear activation function layer, wherein the nonlinear activation function is a hyperbolic tangent function (tanh). We obtained the feature representation of each node through these layers, which forms the input for part III.

In part III, according to the feature representation of nodes, we obtained an adjacency matrix, which is reconstructed. This process is a relatively open problem that can be solved through deep learning. Here, we employed an inner product decoder to calculate node similarity to obtain the reconstructed adjacency matrix, i.e., the link prediction result.

In part IV, model accuracy was verified using the reconstructed adjacency matrix obtained in part III.

2.3.2. Model Interpretation

In the traditional GAT [20] algorithm, convolution is defined as using an attention mechanism to aggregate different neighborhoods differently. The function of the attention mechanism is to assign a weight coefficient to each neighbor node and subsequently updated the central node through convolution summation. Therein, different weights are assigned to different neighbor nodes through GAT, but only the first-order neighbor information of each node is considered. However, there may be other potential edge relationships that have not been mined out in the network. These edge relationships are not considered by GAT, which is inconsistent with the relationships existing in practice because it is not only the first-order neighbor node that affects the node, i.e., a higher-order neighbor node will also have a certain impact on the node or exhibit a certain relationship (positive or negative). For example, X and Y are friends, and their relationship is very good; however, Y and Z have a very poor relationship. According to the balance theory, the relationship between X and Z is very likely to be bad, which means that both Y and Z, i.e., the second-order neighbor of X, have an impact on X. However, Y and Z may have different effects on X, and the impact of first-order neighbors may be even greater, which we discuss in Section 2.3.4.

The algorithm to select high-order neighbor nodes based on the gating mechanism and improving the attention mechanism is employed in the proposed method.

2.3.3. Gating Mechanism

When updating the feature representation of a node, we should consider both the first-order neighbor information of the node and the fact that the high-order neighbor impacts the node. Here, inspired by Long Short-Term Memory (LSTM) [31] and Gate Recurrent Unit (GRU) [32], update, memory, and reset gates are proposed to obtain gating, memory, and attention accessibility matrices, respectively. The gating mechanism employed in the proposed SN-GGAT is shown in Figure 6.

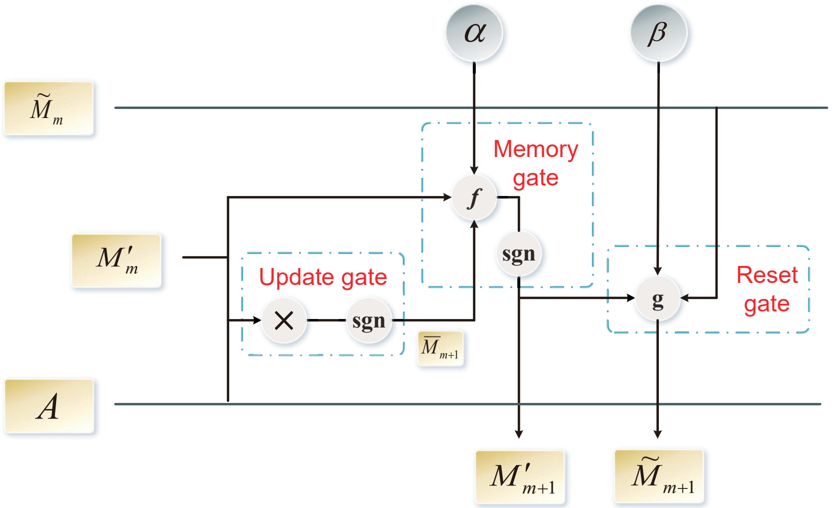


Figure 6. Gating mechanism in the proposed SN-GGAT. It includes update gate, memory gate, and reset gate.

The gating unit's input is the adjacency matrix A, memory accessibility matrix M'_m, and attention accessibility matrix $\widetilde{M}_m$ of the previous gating unit's output. The output is the memory accessibility matrix M'_{m+1} and attention accessibility matrix $\widetilde{M}_{m+1}$. The first-order memory accessibility matrix M'_1 and first-order attention accessibility matrix $\widetilde{M}_1$ are defined as adjacency matrices in Equations (10a) and (10b), respectively.

$$M'_1 = A \tag{10a}$$

$$\widetilde{M}_1 = A \tag{10b}$$

The update gate considers the m-order memory accessibility matrix M'_m and adjacency matrix A as input, and outputs the m + 1—order gating accessibility matrix $\overline{M}_{m+1}$. The specific calculation method is expressed as follows.

$$\overline{M}_{m+1} = \operatorname{sgn}\left(M'_m \cdot A\right) \tag{11}$$

In the update process, the high-order accessibility matrix may forget the low-order neighbor information. For example, in the first-order accessibility matrix (adjacency matrix A) and second-order accessibility matrix expressed by Equations (2) and (7) in Section 2.2, the edge from nodes 1 to 2 shown by the second-order accessibility matrix is -1, and the edge from nodes 1 to 2 in the first-order accessibility matrix is $+1$. Here, the higher-order accessibility matrix forgets the low-order edge information.

To solve this problem, the memory gate memorizes the low-order memory accessibility matrix and displays it in the high-order memory accessibility matrix. In this manner, the high-order memory accessibility matrix can represent the high-order neighbor information of the node and ensure that the low-order neighbor information is retained. The memory gate's inputs are the m-order memory accessibility matrix M'_m and $m+1$—order gating accessibility matrix $\overline{M}_{m+1}$. The $m+1$—order memory accessibility matrix M'_{m+1} can then be obtained from the memory gate as follows.

$$M'_{m+1} = \operatorname{sgn}\left((1-\alpha)\cdot\overline{M}_{m+1} + \alpha\cdot M'_m\right) \text{ where } 0.5 < \alpha < 1 \tag{12}$$

In the proposed method, low-order neighbor nodes are considered nodes with a higher influence. Here, the higher the order, the lower the influence of neighbor nodes, and the positive and negative links have opposite attention. The reset gate realizes this function, and its inputs are the m-order attention accessibility matrix $\widetilde{M}_m$ and $m+1$—order memory accessibility matrix M'_{m+1}. The $m+1$—order attention accessibility matrix $\widetilde{M}_{m+1}$ can then be obtained from the reset gate as follows.

$$\widetilde{M}_{m+1} = (1-\beta)\cdot\widetilde{M}_m + \beta\cdot M'_{m+1} \text{ where } 0 < \beta \le 1 \tag{13}$$

In the reset gate, the smaller the value of β, the lower the influence of high-order neighbors on nodes. For example, when $\beta = 1$, the influence of high-order neighbors is the same as that of low-order neighbors.

Therefore, the symbolic operation in Figure 6 represents matrix multiplication, function f represents $f(x,y) = \operatorname{sgn}((1-\alpha)\cdot x + \alpha\cdot y)$, and function g represents $g(x,y) = (1-\beta)\cdot x + \beta\cdot y$, where α and β are hyperparameters.

The high-order memory accessibility and high-order attention accessibility matrices are calculated recursively in the gating mechanism according to the value of m. When using GAT to update nodes, the attention mechanism simultaneously considers both low- and high-order neighbors, and reasonably allocates the attention coefficient according to the order, which makes node updates more appropriate and practical, and the prediction results are more accurate.

2.3.4. Attention Mechanism

We present the following explanations for the definition of the attention mechanism in signed networks.

1. The first-order neighbor of node i is the node directly associated with node i; therefore, first-order neighbor nodes have the greatest influence on node i.
2. With increasing order, the influence of high-order neighbors of node i on node i decreases gradually.
3. The positive and negative links of node i have the opposite influence on node i.
4. The above influences specifically refer to attention and are well implemented in the attention accessibility matrix output by the gating mechanism.

In the proposed method, the expression and updating rules of the attention weight are affected by the attention accessibility matrix $\widetilde{M}_m$, which makes the attention weight follow the rule of "low-order high attention, high-order low attention, different signs attention opposite" in a signed network. The attention weight is calculated as follows.

$$\operatorname{coef}[i,j] = \frac{\exp\left(\operatorname{LeakyReLU}\left(\vec{\mathbf{a}}^T\left[\mathbf{W}\vec{h}_i \| \mathbf{W}\vec{h}_j\right]\right)\right)}{\sum_{k\in N(i)}\exp\left(\operatorname{LeakyReLU}\left(\vec{\mathbf{a}}^T\left[\mathbf{W}\vec{h}_i \| \mathbf{W}\vec{h}_k\right]\right)\right)} \tag{14}$$

$$\alpha[i,j] = \frac{\mathrm{coef}[i,j] \cdot \widetilde{M}_m[i,j]}{\sum_{k\in N(i)} \mathrm{abs}\Big(\mathrm{coef}[i,k] \cdot \widetilde{M}_m[i,k]\Big)} \tag{15}$$

Here, $\alpha[i,j]$ is the attention weight between nodes i and j, $\cdot T$ represents transposition, $||$ represents the vector connection operation, parameter W is used to realize feature dimension transformation of each node, parameter $\vec{\alpha}$ is used to calculate the attention coefficient of nodes i to j, function $abs()$ is an operation that provides the absolute value, $N(i)$ is the set of neighbor nodes of node i, and the expression of $N(i)$ is given in Equation (16), where n_nodes is the total number of nodes.

$$N(i) = \{n \mid M'_m[i,n] \neq 0, 0 \leq n < n\_nodes\} \tag{16}$$

The updating rules of node features are shown in Equation (17). The low- and high-order neighbors of node i constitute set $N(i)$, which impacts node i.

$$\vec{h}_i = \tanh\left(\sum_{j\in N(i)} \alpha[i,j] \cdot \mathrm{W}\vec{h}_j\right) \tag{17}$$

2.3.5. Algorithm

The specific implementation algorithm of the proposed SN-GGAT is given in Algorithm 1.

Algorithm 1 Gated Graph Attention for Signed Network (SN-GGAT)

```
Input:
    PPIs adjacency matrix A;
    The number of nodes n;
    The order of accessibility matrix m;
    Epochs E.
Output:
    Node feature matrix Z;
    Reconstructed adjacency matrix $A_r$.
1:  $\overline{M}_1, M'_1, \widetilde{M}_1 \leftarrow A$
2:  if $m! = 1$ then
3:      for $i \in \{2, \ldots, m\}$ do
4:          $\overline{M}_i \leftarrow \mathrm{sgn}(M'_{i-1} \cdot A)$
5:          $M'_i \leftarrow \mathrm{sgn}((1-\alpha)\overline{M}_i + \alpha M'_{i-1})$
6:          $\widetilde{M}_i \leftarrow (1-\beta)\widetilde{M}_{i-1} + \beta M'_i$
7:      end for
8:  end if
9:  for epoch $\in \{1, \ldots, E\}$ do
10:     for $i \in \{0, \cdots, n-1\}$ do
11:         $coef[i,j] \leftarrow \mathrm{softmax}\Big(\mathrm{LeakyReLU}\Big(\vec{a}^T\Big[W\vec{h}_i \| W\vec{h}_j\Big]\Big)\Big)$
12:         $\alpha[i,j] \leftarrow \mathrm{coef}[i,j] \cdot \widetilde{M}_m[i,j] / \sum_{k\in N(i)} abs\Big(\mathrm{coef}[i,k] \cdot \widetilde{M}_m[i,k]\Big)$
13:         $\vec{h}_i \leftarrow \tanh\Big(\sum_{j\in N(i)} \alpha[i,j] \cdot W\vec{h}_j\Big)$
14:     end for
15:     $Z \leftarrow \mathrm{concat}\Big(\vec{h}_0, \vec{h}_1, \ldots, \vec{h}_{n-1}\Big)$
16:     $A_r \leftarrow \mathrm{sgn}(Z \cdot Z^T)$
17:     Update parameters with $A_r$
18: end for
```

Here, in lines 1 to 8, the m-order gating accessibility matrix, m-order memory accessibility matrix, and m-order attention accessibility matrix are obtained by iterating the gating mechanism, and lines 11 to 13 correspond to the node update process.

From lines 15 to 18, we obtained the node's feature matrix and used the idea of the inner product decoder to calculate node similarity to obtain the reconstructed adjacency matrix. Here, the loss function uses cross entropy with logits when updating parameters.

3. Results and Discussion

3.1. Link Sign Prediction of the Saccharomyces Cerevisiae PPI Signed Network

In this section, we used the Saccharomyces cerevisiae core dataset to test the validity of the proposed SN-GGAT, and introduce the computation settings, evaluation criteria, test results, and a discussion of the results.

3.1.1. Computation Settings

In the literature [33], the second-order accessibility matrix was used as the feature representation of nodes, and good link prediction results were obtained in signed networks. Therefore, in our algorithm, the second-order accessibility matrix was input as the feature set of the model.

In this test, we set up two attention convolutional layers. The output dimensions of each layer were 128 and 32, respectively; the multi-head attention mechanism was used in the first layer; and the number of heads was 6. The activation function of the first layer used the elu function, which is used in the original GAT algorithm, and the activation function of the second layer used the tanh function. During training, we used the Adam optimizer with a learning rate of 0.002 to optimize the parameters.

In the gating mechanism, we used the second-order memory accessibility matrix and the second-order attention accessibility matrix with the best test results in the node update process, where the hyperparameters were as follows: $\alpha = 0.8$ and $\beta = 0.2$. Finally, the node embedded representation of the model output participated in the subsequent link prediction.

3.1.2. Evaluation Criteria

In this test, we used three commonly used machine learning evaluation indexes, i.e., accuracy, precision, and recall, which are defined in Equations (18)–(20), respectively.

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN} \tag{18}$$

$$Precision = \frac{TP}{TP + FP} \tag{19}$$

$$Recall = \frac{TP}{TP + FN} \tag{20}$$

Here, TP denotes true positive (representing the number of samples with positive predictive values and positive labels), FP denotes false positive (representing the number of samples with positive predictive values but negative labels), TN denotes true negative (representing the number of samples with negative predictive values and negative labels), and FN denotes false negative (representing the number of samples with negative predictive values but positive labels).

3.1.3. Test Results

We compared the proposed method with the state-of-the-art PPI prediction methods, including Wong's method [34], Du's method [35], DeepFE-PPIs [23], and Song's method [36]. These methods are summarized as follows:

- Wong's method: this method is a combination of the Rotating Forest (RF) model and a new feature representation for PPIs detection. In this method, the response matrix (PR) method is used to transform the amino acid sequence into a matrix, and then the texture descriptor based on local phase quantization (LPQ) is used to extract the local phrase information in the matrix.

- Du's method: This method uses a deep neural network to learn protein representation from common protein descriptors effectively, and extracts useful features of protein pairs by a layer-wise abstraction.
- DEEPFE-PPIs: This method employs a new residue representation called Res2vec, which provides effective input for the downstream deep learning model. PPIs can be accurately inferred when protein structure knowledge is completely unknown.
- Song's method: In this method, a random projection ensemble classifier (RPEC) is used to identify new PPIs based on the evolutionary information contained in protein amino acid sequences.

We used 5-fold cross validation on the S.cerevisiae core dataset and compared the test results with other state-of-the-art methods. When performing 5-fold cross-validation, we divided the entire dataset into five parts in equal proportions, and took one part as the test set and the other four parts as the training set without repeating each time. In these five tests, each test was run repeatedly three times, and the index values obtained from the 3 times were averaged as the result of each test. The comparison results are shown in Table 1 and Figure 7, where the overall statistical significance level of all tests run is 5% (i.e., $p < 0.05$). Note that SN-GAT is the test result obtained without using the gating mechanism (or the parameter m of the gating mechanism is equal to 1), i.e., the memory accessibility and attention accessibility matrices used in the node update process are the adjacency matrix A of the PPI signed network.

The results demonstrate that SN-GGAT achieved remarkable PPI prediction results and outperformed the compared methods in all evaluation criteria. Introducing the gating mechanism allows the node to consider high-order neighbor information during the node update process and assigns different weights to both high- and low-order neighbors. The prediction results of SN-GGAT were more accurate than those of SN-GGAT; therefore, the gating mechanism plays an important role in SN-GGAT, which is the reason why our algorithm outperforms other existing prediction algorithms.

Table 1. Performance comparison of SN-GGAT with other state-of-the-art methods on the *S. cerevisiae* core dataset.

Method	Test Set	Accuracy (%)	Precision (%)	Recall (%)
SN-GGAT	1	96.43	98.01	94.96
	2	97.04	98.40	95.78
	3	96.25	97.82	94.78
	4	96.34	98.28	94.50
	5	96.81	98.57	95.14
	Average	96.57 ± 0.34	98.22 ± 0.30	95.03 ± 0.48
SN-GAT	1	95.72	97.61	93.95
	2	96.03	97.99	94.14
	3	95.19	96.69	93.85
	4	95.54	97.79	93.40
	5	95.91	98.26	93.67
	Average	95.68 ± 0.33	97.67 ± 0.60	93.80 ± 0.28
Song's method	Average	95.64 ± 0.52	96.75 ± 0.45	94.47 ± 0.47
DeepFE-PPIs	Average	94.78 ± 0.61	96.45 ± 0.87	92.99 ± 0.66
Du's method	Average	94.43 ± 0.30	96.65 ± 0.59	92.06 ± 0.36
Wong's method	Average	93.92 ± 0.36	96.45 ± 0.45	91.10 ± 0.31

In addition, we found a biologically interesting example. In our prediction, there is an interaction between Ctk1 and Snf1. The interaction between the two proteins is not shown in the database, at least not in the Saccharomyces cerevisiae core dataset of DIP database. However, Driessche et al. [37] found a physical interaction between Ctk1 and Snf1 in their two-hybrid system. Ctk1 is a kinase involved in transcriptional control, and Snf1 is a kinase that regulates glucose-dependent genes. Driessche et al. showed that Ctk1 and Snf1 co-regulate GSY2 in vivo by Northern blot analysis. This finding supports the view that Ctk1 interacts with Snf1 in the functional module of cell response to glucose restriction.

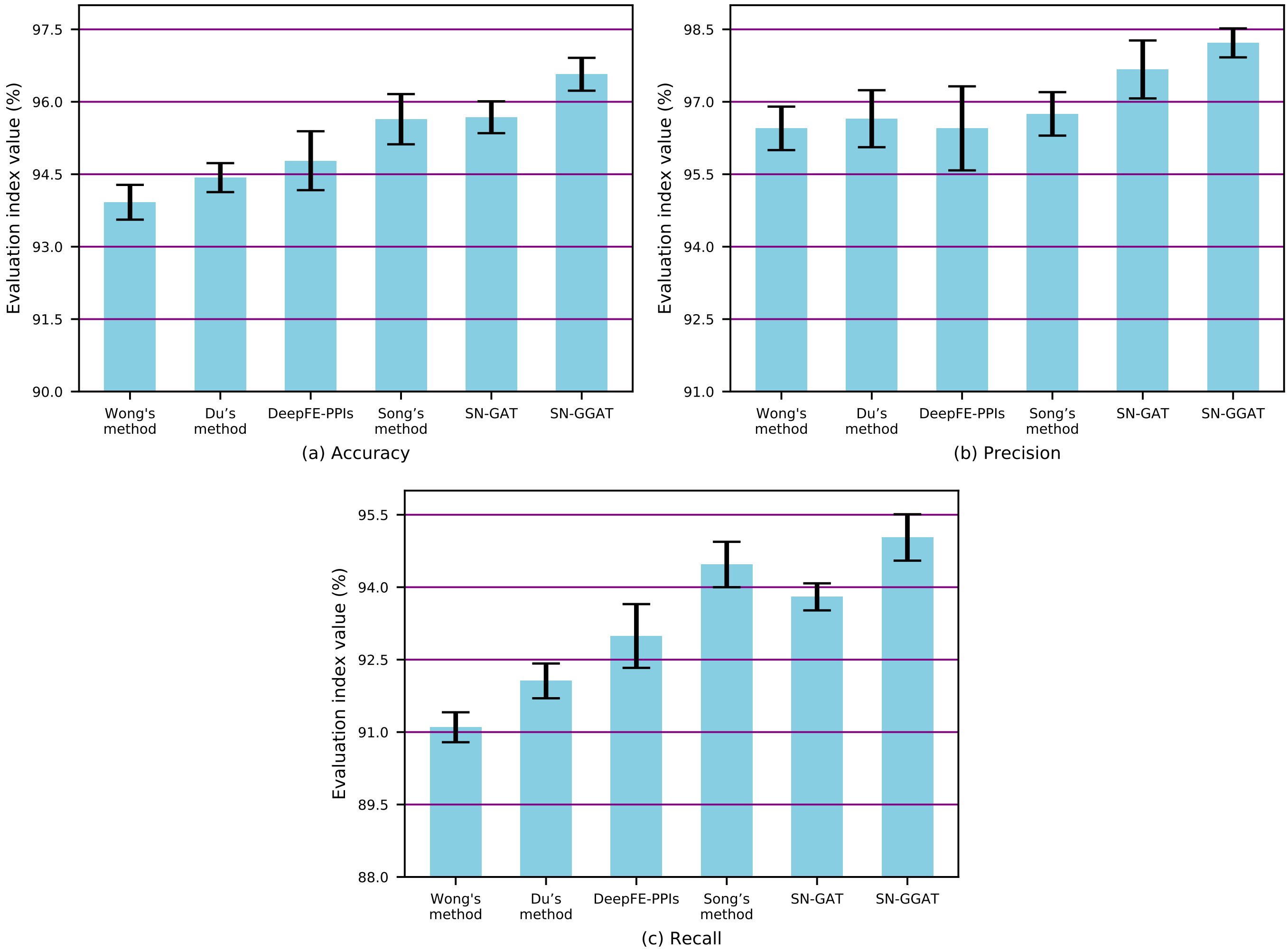


Figure 7. Accuracy, Precision, and Recall values (Mean ± SD) of SN-GGAT method and the compared methods.

3.1.4. Parameter Discussion

In our algorithm, the most important hyperparameter is parameter m of the gating mechanism (the order m of the accessibility matrix). After conducting nearly twenty tests, we found that when $m = 2$, the result obtained using the second-order memory accessibility matrix and second-order attention accessibility matrix was excellent. The test results obtained by considering different values of m are shown in Figure 8a–c.

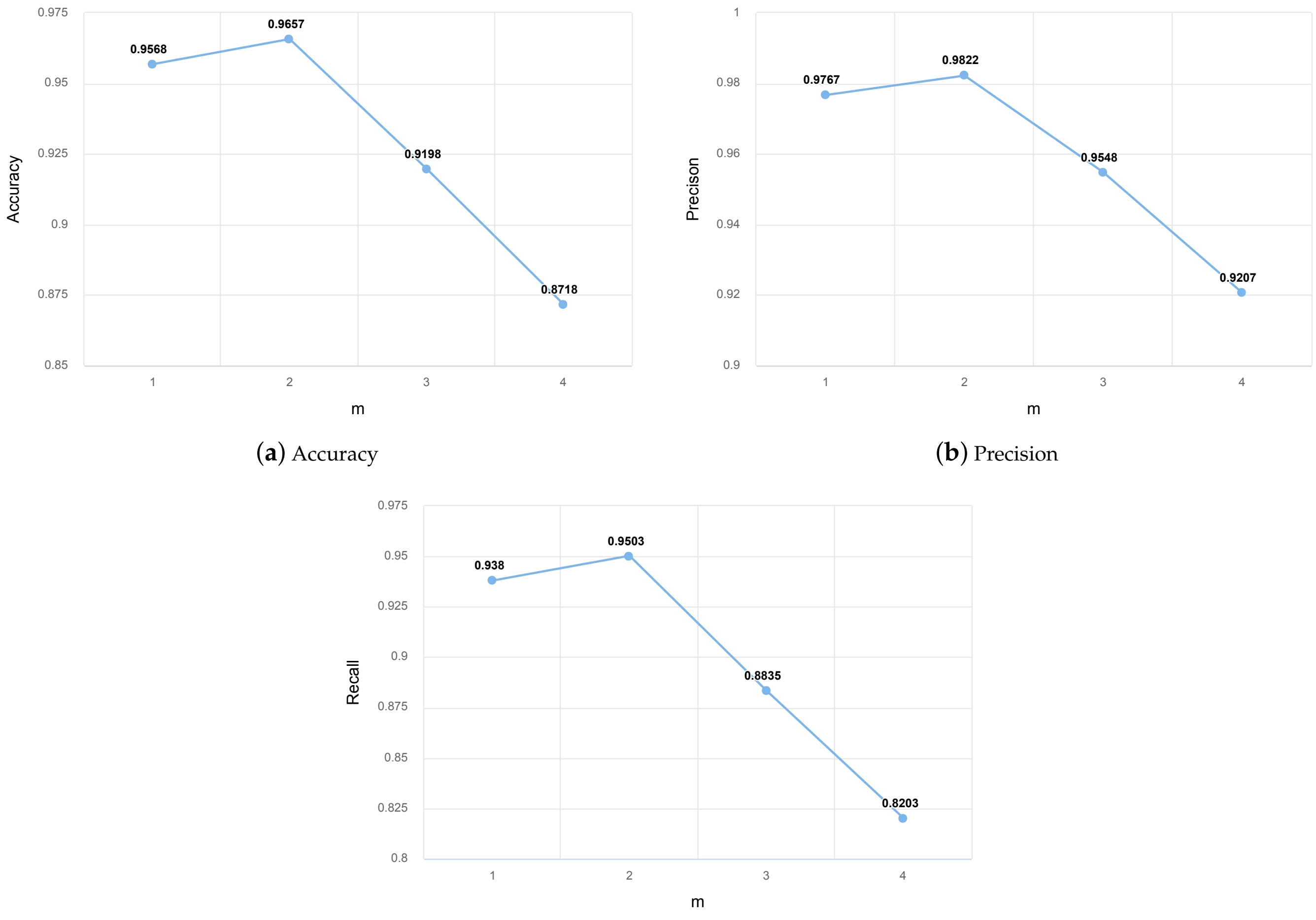
As can be observed in the results described above, the index values obtained using the second-order memory accessibility matrix are greater than those obtained using the adjacency matrix (first-order memory accessibility matrix), and model performance was optimal when using the second-order memory accessibility matrix. However, the model's performance was poor when the third- or fourth-order memory accessibility matrices were used. We summarize possible reasons for these results as follows.

1. The first-order neighbors of a node have the highest influence on the node; therefore, the result obtained using $m = 1$ were second only to the result obtained using $m = 2$.
2. As $m = 1$ only considers first-order neighbor nodes as neighbors and does not consider second-order neighbor nodes, $m = 2$ will consider high-order information

more comprehensively. The gating mechanism reasonably allocates the attention weight for the first- and second-order nodes, thus improving the test results when $m = 2$.

3. When $m > 2$, the model's result was very poor. There may be two reasons for this. First, according to the principle of the defined gating mechanism, the higher the accessibility matrix's order, the higher the number of high-order nodes that are assigned attention. When the attention weight is calculated using Equation (15), under the influence of high-order neighbors, the proportion of the influence of first-order neighbors on nodes decreases, thereby worsening the model's final effect. Second, we explain this phenomenon through an intuitive example: my friend (first-order neighbor) has a great influence on me, and my friend's friend (second-order neighbor) also has a certain influence on me; however, a friend of my friend's friend (third-order neighbors) may have a minimal influence on me, and I may not even meet them (third-order neighbors) in real life. Therefore, the test results obtained with $m = 3$ are not ideal.

Generally, when $m = 2$, our gating mechanism achieved highly satisfactory results. The subject of considering high-order information without reducing the attention of low-order neighbors to themselves will be researched in future work.



(**a**) Accuracy

(**b**) Precision

(**c**) Recall

Figure 8. Test results obtained with different m values on the S.cerevisiae core dataset. Where (**a**–**c**) are the Accuracy, Precision, and Recall with different m values, respectively.

3.2. *Link Sign Prediction of the Human PPI Signed Network*

In this section, we predict the human–protein interaction, verify the performance of our method by comparing it with three state-of-the-art methods, and discuss the biological meanings.

3.2.1. Test Results

We predicted the link of the signed network composed of human protein and evaluated the model by using the value of auPR, where auPR refers to the area under the PR (precision-recall) curve. The higher the value of auPR, the better the performance of the model. We performed 10-fold cross validation on the Human dataset, and the test results are shown in Figure 9.

We took the average value of ten results, drew the average curve of PR, and compared it with DPPIs [28], DeepFE-PPIs [23], and Liu's method [27]. The comparison results are shown in Figure 10. DPPIs uses a convolutional neural network combined with random projection and data expansion to predict PPIs, and its auPR value is 0.4127. DeepFE-PPIs employ a new residue representation called Res2vec, which provides effective input for the downstream deep learning model. The auPR value of this method is 0.4273. Liu's method uses GCNs to learn the location information of proteins in a PPI network and combines the sequence and location information of amino acids to generate strong protein characterization, with an auPR value of 0.4542. The SN-GGAT method has the best performance, and the auPR value is 0.5104, which is higher than the other three methods.

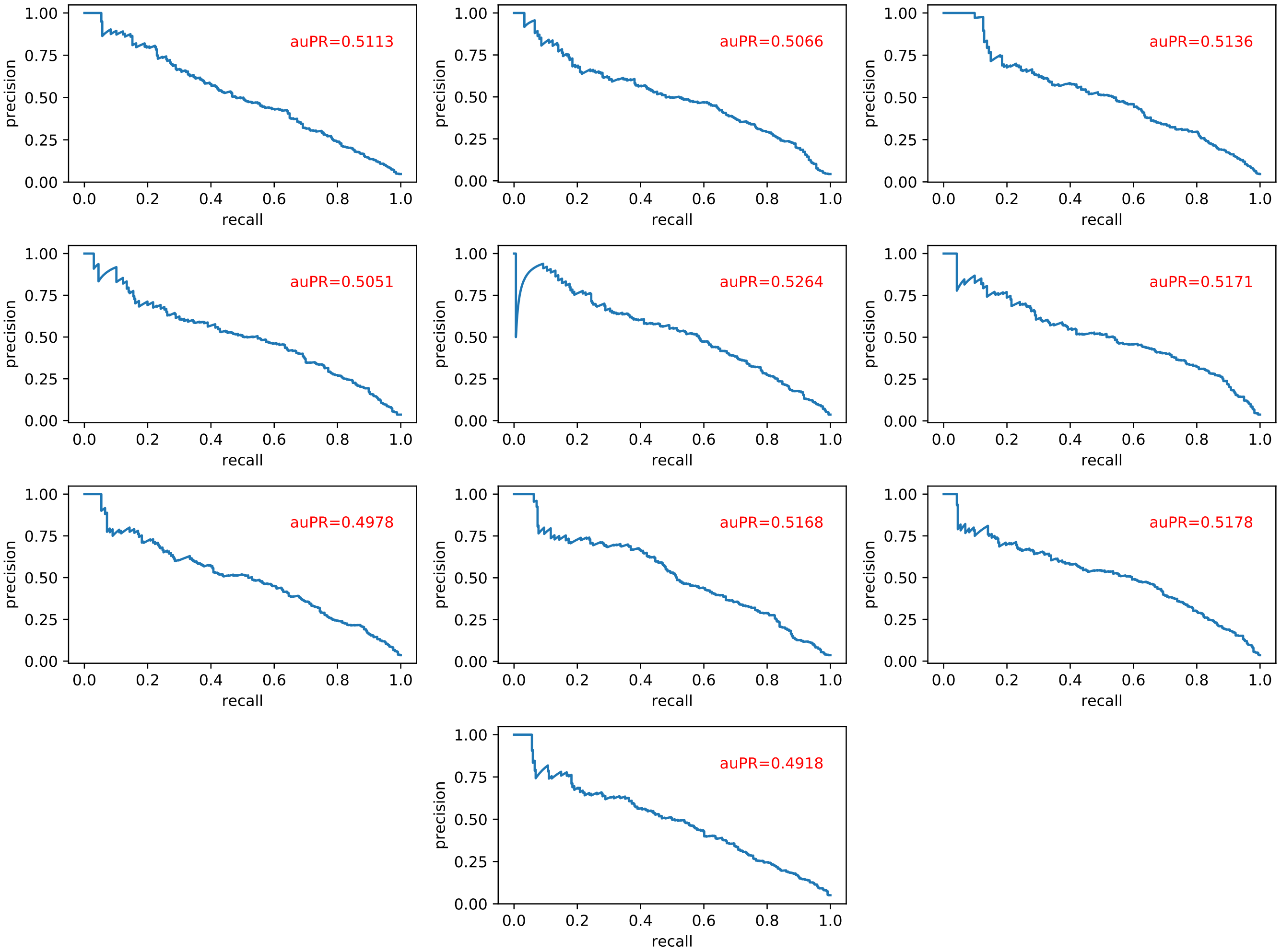


Figure 9. The test results obtained by using the SN-GGAT method to execute the link signed prediction task on the Human dataset, where each sub-graph represents one prediction result.

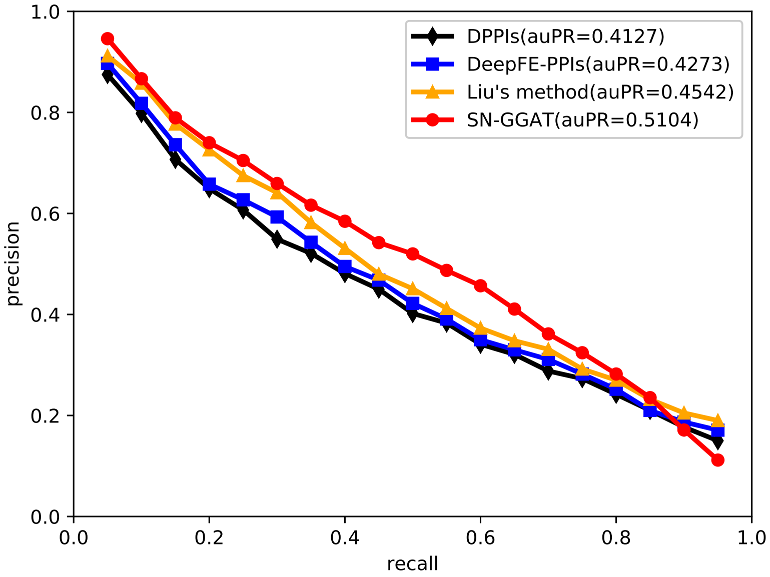


Figure 10. Performance comparison of SN-GGAT with other state-of-the-art methods on the Human dataset. The auPR is the mean of 10-fold cross validation.

By analyzing the prediction results, we found that most of the predicted interactions satisfy the balance theory in the entire signed network. The explanation of satisfying the balance theory is shown in Figure 11a, and the analysis result is shown in Figure 11b.

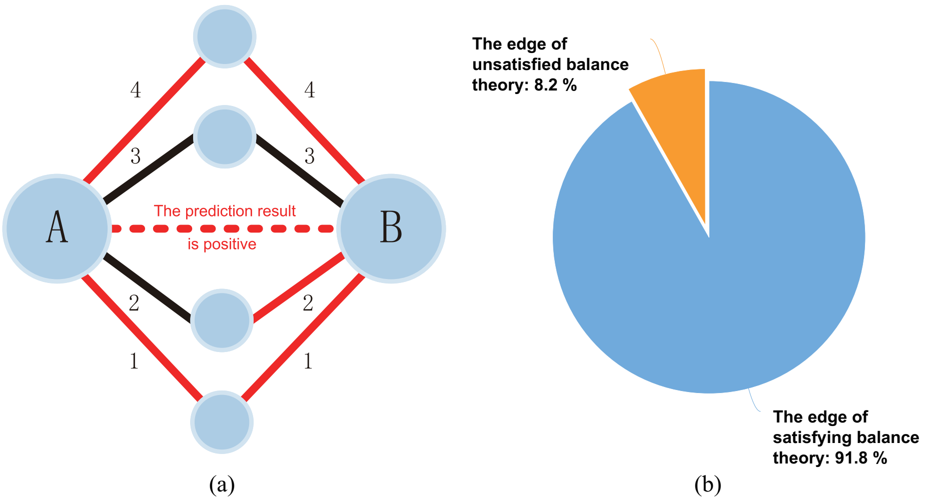


Figure 11. (**a**) An example of PPI satisfying balance theory in signed networks. According to the balance theory, it can be inferred that the connection between protein A and protein B is positive, negative, positive, and positive through routes 1, 2, 3, and 4, respectively. Since the theoretically positive quantity is more than the theoretically negative quantity, if the connection edge between protein A and protein B is predicted to be positive, it means that the predicted protein interaction satisfies the balance theory in the entire signed network. (**b**) The analysis result of the interaction network obtained after the prediction. It shows the proportion of edges satisfying balance theory and unsatisfied balance theory in the entire signed network.

This analysis result shows that in the entire signed network after prediction, 91.8% of the connected edges satisfy the balance theory. At the same time, it also shows that the balance theory plays a key role in the gating mechanism and sign propagation process. The introduction of balance theory into SN-GGAT has an important impact on the performance of the model.

3.2.2. Discussion in Biological Meanings

In our test, we constructed the protein–protein interactions as a signed network. The SN-GGAT method expands the PPI network by predicting unknown protein interactions (i.e., the connection edges in the PPI network were significantly increased) so that the entire PPI network is rich in more interactive information. From a biological point of view, SN-GGAT can provide important clues for the in-depth study of protein functions by accurately predicting protein interactions. In the current bioinformatics research, some new methods [38–40] using the PPI network to predict protein function have been proposed. SN-GGAT can expand the existing PPI network and increase the training samples of protein function prediction by increasing the number of edges in the protein–protein interaction network, so as to improve the accuracy of protein function prediction. In addition, the accurate prediction of protein interactions can also promote the in-depth study of gene expression regulatory networks in biological life activities and other complex life activities. In sum, the SN-GGAT method we proposed has made considerable contributions to biological research, especially in research based on PPI networks and other bioinformatics-related research.

4. Conclusions

Protein–protein interaction usually refers to the binding or chemical reaction between proteins through spatial conformation or chemical bond, while a protein–protein interaction network is composed of proteins through their interactions. These interactions are involved in cell cycle regulation, gene expression regulation, biological signal transduction, and energy and substance metabolism processes. The accurate prediction of protein–protein interactions in biological systems plays an important role in understanding the working principle of proteins, the functional connections between proteins, and the reaction mechanism of biological signals and energy metabolism.

In order to accurately predict PPIs, we characterized PPIs as a signed network and realized PPI prediction via link prediction of the signed network. We developed the SN-GGAT method, which combines the concepts of signed network, balance theory, and accessibility matrix. The proposed method selects high-order neighbor nodes based on a gating mechanism and improves the attention mechanism of the original GAT. In addition, node features are updated according to the rules of "low-order high attention, high-order low attention, different sign attention opposite"; thus, the concept of GAT is extended to signed networks and applied to PPI prediction. We compared the proposed SN-GGAT to four state-of-the-art methods on the Saccharomyces cerevisiae core dataset. The test results demonstrated that the proposed method obtained the highest accuracy and has strong competitiveness. Finally, we tested the proposed method again on the Human protein interaction dataset. We learned that most of the predicted PPIs satisfy the balance theory in the entire signed network. This conclusion shows that the idea of incorporating balance theory into the algorithm is correct and necessary.

In the future, we plan to further study the attention mechanism of signed networks to effectively mine and use hidden high-order information to improve the accuracy of PPI prediction. In addition, the PPI is predicted in this paper, and we will continue in-depth study in the future, such as experimental verification of the predicted results.

Author Contributions: Conceptualization, Z.X., W.G. and H.W.; methodology, Z.X. and H.W.; validation, Z.L. and H.W.; formal analysis, X.Y. and H.W.; data curation, W.G., J.W. and H.W.; writing—original draft preparation, Z.X. and H.W.; writing—review and editing, H.W.; supervision, H.W.; project administration, H.W.; funding acquisition, H.W. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the National Natural Science Foundation of China (no. 61672329, no. 62072290, no. 81871508, and no. 61773246), the Major Program of Shandong Province Natural Science Foundation (no. ZR2019ZD04 and no. ZR2018ZB0419), and the Shandong Provincial Project of Education Scientific Plan (no. SDYY18058).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The Saccharomyces cerevisiae core dataset was obtained from the database of interacting proteins (DIP) and are available at https://dip.doe-mbi.ucla.edu/dip/Download.cgi?SM=7&TX=4932 (accessed on 10 February 2021) with the permission of DIP. The UniProt database was obtained from UniProt and is available at https://www.uniprot.org/ (accessed on 10 February 2021) with the permission of UniProt. The Human PPI dataset was obtained from Liu's site and is available at https://zenodo.org/record/3960077/files/Human.zip?download=1 (accessed on 31 March 2021) with the permission of Liu.

Acknowledgments: The authors thank the laboratory of data science and technology, Shandong Normal University.

Conflicts of Interest: The authors declare no conflict of interest.

Abbreviations

The following abbreviations are used in this manuscript:

PPIs	Protein–protein interactions
SN-GGAT	Gated Graph Attention for Signed Networks
GAT	Graph Attention Network
GNNs	Graph Neural Networks
GCN	Graph Convolution Network

References

1. Bauer, A.; Kuster, B. Affinity purification-mass spectrometry: Powerful tools for the characterization of protein complexes. *Eur. J. Biochem.* **2003**, *270*, 570–578. [CrossRef]
2. Fields, S.; Song, O.K. A novel genetic system to detect protein–protein interactions. *Nature* **1989**, *340*, 245–246. [CrossRef]
3. Ong, S.E.; Blagoev, B.; Kratchmarova, I.; Kristensen, D.B.; Steen, H.; Pandey, A.; Mann, M. Stable isotope labeling by amino acids in cell culture, SILAC, as a simple and accurate approach to expression proteomics. *Mol. Cell. Proteom.* **2002**, *1*, 376–386. [CrossRef]
4. Lin, J.S.; Lai, E.M. Protein–protein interactions: Co-immunoprecipitation. In *Bacterial Protein Secretion Systems*; Springer: Berlin, Germany, 2017; pp. 211–219. [CrossRef]
5. Ledsgaard, L.; Kilstrup, M.; Karatt-Vellatt, A.; McCafferty, J.; Laustsen, A.H. Basics of antibody phage display technology. *Toxins* **2018**, *10*, 236. [CrossRef]
6. Syahir, A.; Usui, K.; Tomizaki, K.Y.; Kajikawa, K.; Mihara, H. Label and label-free detection techniques for protein microarrays. *Microarrays* **2015**, *4*, 228–244. [CrossRef] [PubMed]
7. Nguyen, H.H.; Park, J.; Kang, S.; Kim, M. Surface plasmon resonance: A versatile technique for biosensor applications. *Sensors* **2015**, *15*, 10481–10510. [CrossRef] [PubMed]
8. Pazos, F.; Valencia, A. Similarity of phylogenetic trees as indicator of protein–protein interaction. *Protein Eng.* **2001**, *14*, 609–614. [CrossRef]
9. Hsin Liu, C.; Li, K.C.; Yuan, S. Human protein–protein interaction prediction by a novel sequence-based co-evolution method: Co-evolutionary divergence. *Bioinformatics* **2013**, *29*, 92–98. [CrossRef] [PubMed]
10. Guo, Y.; Yu, L.; Wen, Z.; Li, M. Using support vector machine combined with auto covariance to predict protein–protein interactions from protein sequences. *Nucleic Acids Res.* **2008**, *36*, 3025–3030. [CrossRef]
11. Zaki, N.; Lazarova-Molnar, S.; El-Hajj, W.; Campbell, P. Protein-protein interaction based on pairwise similarity. *BMC Bioinform.* **2009**, *10*, 1–12. [CrossRef] [PubMed]
12. Roy, S.; Martinez, D.; Platero, H.; Lane, T.; Werner-Washburne, M. Exploiting amino acid composition for predicting protein-protein interactions. *PLoS ONE* **2009**, *4*, e7813. [CrossRef]
13. Valente, G.T.; Acencio, M.L.; Martins, C.; Lemke, N. The development of a universal in silico predictor of protein-protein interactions. *PLoS ONE* **2013**, *8*, e65587. [CrossRef]
14. Kern, C.; González, A.J.; Liao, L.; Vijay-Shanker, K. Predicting interacting residues using long-distance information and novel decoding in hidden markov models. *IEEE Trans. Nanobiosci.* **2013**, *12*, 158–164. [CrossRef]
15. Zhang, J.; Liu, X.; Ren, F.; Sun, X.; Yu, Q. The effects of group diversity and organizational support on group creativity. *Acta Psychol. Sin.* **2016**, *48*, 1551–1560. [CrossRef]
16. Kunegis, J.; Preusse, J.; Schwagereit, F. What is the added value of negative links in online social networks? In Proceedings of the 22nd International Conference on World Wide Web, Rio de Janeiro, Brazil, 13–17 May 2013; pp. 727–736. [CrossRef]

17. Wang, D.; Cui, P.; Zhu, W. Structural deep network embedding. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Francisco, CA, USA, 13–17 August 2016; pp. 1225–1234. [CrossRef]
18. Grover, A.; Leskovec, J. node2vec: Scalable feature learning for networks. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Francisco, CA, USA, 13–17 August 2016; pp. 855–864. [CrossRef]
19. Kipf, T.N.; Welling, M. Semi-supervised classification with graph convolutional networks. *arXiv* **2016**, arXiv:1609.02907.
20. Veličković, P.; Cucurull, G.; Casanova, A.; Romero, A.; Lio, P.; Bengio, Y. Graph attention networks. *arXiv* **2017**, arXiv:1710.10903.
21. Xenarios, I.; Salwinski, L.; Duan, X.J.; Higney, P.; Kim, S.M.; Eisenberg, D. DIP, the Database of Interacting Proteins: A research tool for studying cellular networks of protein interactions. *Nucleic Acids Res.* **2002**, *30*, 303–305. [CrossRef]
22. You, Z.H.; Zhu, L.; Zheng, C.H.; Yu, H.J.; Deng, S.P.; Ji, Z. Prediction of protein-protein interactions from amino acid sequences using a novel multi-scale continuous and discontinuous feature set. In *BMC Bioinformatics*; Springer: Berlin, Germany, 2014; Volume 15, pp. 1–9. [CrossRef]
23. Yao, Y.; Du, X.; Diao, Y.; Zhu, H. An integration of deep learning with feature embedding for protein–protein interaction prediction. *PeerJ* **2019**, *7*, e7126. [CrossRef]
24. Consortium, U. Update on activities at the Universal Protein Resource (UniProt) in 2013. *Nucleic Acids Res.* **2012**, *41*, D43–D47. [CrossRef]
25. Schaefer, M.H.; Fontaine, J.F.; Vinayagam, A.; Porras, P.; Wanker, E.E.; Andrade-Navarro, M.A. HIPPIE: Integrating protein interaction networks with experiment based quality scores. *PLoS ONE* **2012**, *7*, e31826. [CrossRef]
26. Hamp, T.; Rost, B. Evolutionary profiles improve protein–protein interaction prediction from sequence. *Bioinformatics* **2015**, *31*, 1945–1950. [CrossRef]
27. Liu, L.; Zhu, X.; Ma, Y.; Piao, H.; Yang, Y.; Hao, X.; Fu, Y.; Wang, L.; Peng, J. Combining sequence and network information to enhance protein–protein interaction prediction. *BMC Bioinform.* **2020**, *21*, 1–13. [CrossRef]
28. Hashemifar, S.; Neyshabur, B.; Khan, A.A.; Xu, J. Predicting protein–protein interactions through sequence-based deep learning. *Bioinformatics* **2018**, *34*, i802–i810. [CrossRef]
29. Vinayagam, A.; Zirin, J.; Roesel, C.; Hu, Y.; Yilmazel, B.; Samsonova, A.A.; Neumüller, R.A.; Mohr, S.E.; Perrimon, N. Integrating protein-protein interaction networks with phenotypes reveals signs of interactions. *Nat. Methods* **2014**, *11*, 94–99. [CrossRef] [PubMed]
30. Derr, T.; Ma, Y.; Tang, J. Signed graph convolutional networks. In Proceedings of the 2018 IEEE International Conference on Data Mining (ICDM), Singapore, 17–20 November 2018; pp. 929–934. [CrossRef]
31. Gers, F. Long Short-Term Memory in Recurrent Neural Networks. Ph.D. Thesis, Lausanne, Switzerland, 2001. [CrossRef]
32. Cho, K.; Van Merriënboer, B.; Gulcehre, C.; Bahdanau, D.; Bougares, F.; Schwenk, H.; Bengio, Y. Learning phrase representations using RNN encoder-decoder for statistical machine translation. *arXiv* **2014**, arXiv:1406.1078.
33. Cui, J.; Zhuang, H.; Liu, T.; Wang, H. Semi-Supervised Gated Spectral Convolution on a Directed Signed Network. *IEEE Access* **2020**, *8*, 49705–49716. [CrossRef]
34. Wong, L.; You, Z.H.; Li, S.; Huang, Y.A.; Liu, G. Detection of protein-protein interactions from amino acid sequences using a rotation forest model with a novel PR-LPQ descriptor. In *Proceedings of the International Conference on Intelligent Computing*; Springer: Berlin, Germany, 2015; pp. 713–720. [CrossRef]
35. Du, X.; Sun, S.; Hu, C.; Yao, Y.; Yan, Y.; Zhang, Y. DeepPPI: boosting prediction of protein–protein interactions with deep neural networks. *J. Chem. Inf. Model.* **2017**, *57*, 1499–1510. [CrossRef]
36. Song, X.Y.; Chen, Z.H.; Sun, X.Y.; You, Z.H.; Li, L.P.; Yang, Z. An Ensemble Classifier with Random Projection for Predicting Protein–Protein Interactions Using Sequence and Evolutionary Information. *Appl. Sci.* **2018**, *8*, 89. [CrossRef]
37. Van Driessche, B.; Coddens, S.; Van Mullem, V.; Vandenhaute, J. Glucose deprivation mediates interaction between CTDK-I and Snf1 in Saccharomyces cerevisiae. *FEBS Lett.* **2005**, *579*, 5318–5324. [CrossRef]
38. Zhao, B.; Wang, J.; Wu, F.X. Computational methods to predict protein functions from protein-protein interaction networks. *Curr. Protein Pept. Sci.* **2017**, *18*, 1120–1131. [CrossRef]
39. Prasad, A.; Saha, S.; Chatterjee, P.; Basu, S.; Nasipuri, M. Protein function prediction from protein interaction network using bottom-up L2L apriori algorithm. In *Proceedings of the International Conference on Computational Intelligence, Communications, and Business Analytics*; Springer: Berlin, Germany, 2017; pp. 3–16. [CrossRef]
40. Saha, S.; Prasad, A.; Chatterjee, P.; Basu, S.; Nasipuri, M. Protein function prediction from protein–protein interaction network using gene ontology based neighborhood analysis and physico-chemical features. *J. Bioinform. Comput. Biol.* **2018**, *16*, 1850025. [CrossRef]