Graph Neural Networks and Their Current Applications in Bioinformatics

Xiao-Meng Zhang[1], Li Liang[1], Lin Liu[1,2]* and Ming-Jing Tang[2,3]*

[1] School of Information, Yunnan Normal University, Kunming, China, [2] Key Laboratory of Educational Informatization for Nationalities Ministry of Education, Yunnan Normal University, Kunming, China, [3] School of Life Sciences, Yunnan Normal University, Kunming, China

Graph neural networks (GNNs), as a branch of deep learning in non-Euclidean space, perform particularly well in various tasks that process graph structure data. With the rapid accumulation of biological network data, GNNs have also become an important tool in bioinformatics. In this research, a systematic survey of GNNs and their advances in bioinformatics is presented from multiple perspectives. We first introduce some commonly used GNN models and their basic principles. Then, three representative tasks are proposed based on the three levels of structural information that can be learned by GNNs: node classification, link prediction, and graph generation. Meanwhile, according to the specific applications for various omics data, we categorize and discuss the related studies in three aspects: disease prediction, drug discovery, and biomedical imaging. Based on the analysis, we provide an outlook on the shortcomings of current studies and point out their developing prospect. Although GNNs have achieved excellent results in many biological tasks at present, they still face challenges in terms of low-quality data processing, methodology, and interpretability and have a long road ahead. We believe that GNNs are potentially an excellent method that solves various biological problems in bioinformatics research.

Keywords: bioinformatics, graph neural networks, deep learning, omics data, network biology

INTRODUCTION

In recent years, deep learning has met with great success in machine learning tasks such as speech recognition and image classification. Nevertheless, most of the theories of deep learning are focused on explaining regular Euclidean data (**Figure 1A**). With the rapid accumulation of non-Euclidean data represented by graph structure data (**Figure 1B**), more and more researchers begin to pay attention to the processing of graph structure data that can represent complex relationships between objects. For example, graph embedding algorithms are used to perform the mapping of graph structure data to simpler representations (Scarselli et al., 2008). However, this method may lose the topological information of the graph structure in the pre-treating stage, thereby affecting the final prediction result. Gori et al. (2005) proposed the concept of graph neural networks (GNNs) and designed a model that can directly process graph structure data based on research results in the field of neural networks. Scarselli et al. (2008) elaborated on this model, which showed that GNNs could deliver significantly better results than traditional methods due to using the topological information of graphs in an iterative process. Subsequently, new models and application research on

GNNs have been proposed. With the increasing interest in graph structure data mining, the research direction and application fields of GNNs have been greatly expanded.

In general, GNNs are actually a connectionist model that captures the dependence of graphs through message passing between nodes, which take into account the scale, heterogeneity, and deep topological information of input data simultaneously. At present, GNNs show reliable performance in mining deep-level topological information, extracting the key features of data, and realizing the rapid processing of massive data, such as predicting the properties of chemical molecules (Duvenaud et al., 2015), extracting text relationship (Peng et al., 2017), reasoning the structure of graphics and images (Wang et al., 2018), link prediction and node clustering of social networks (Zhang and Chen, 2018), network completion of missing information (Bojchevski and Günnemann, 2017), drug interaction prediction (Zitnik et al., 2018), etc.

In the era of biomedicine "big data," the aggregation and growth of large amounts of multiform data created enormous challenges to bioinformatics studies. In response to the characteristics and demands of these data, many algorithms in the field of machine learning, especially in deep learning, have been widely used in bioinformatics and propelling the development of bioinformatics. In many cases, biological data is constructed as a biological network in non-Euclidean domains, such as the molecular structure of proteins and RNAs, genetic disease association networks, and protein interaction networks. These biological networks have a great contribution to bioinformatics studies, especially for revealing the complex mechanisms of diseases. Network-based disease prediction methods had been proposed in 2011 (Barabasi et al., 2011), which was based on the assumption that "if a few disease components are identified, other disease-related components are likely to be found in their network-based vicinity." Goh et al. (2007) pointed out that the number of interactions between proteins in the same disease pathway was 10 times higher than in random experiments. Navlakha and Kingsford (2010) proved that the network topology method was effective in predicting the association of diseases and even the interaction of biomolecules. Compared with other models in deep learning, the natural advantage of GNNs in capturing hidden information in biological networks brings new opportunities to design computational models in the biology field. Besides this, GNNs are not only suitable for non-Euclidean data but also able to extract potential graph structures from data without apparent graph structures like images and make inferences and judgments based on this structure. Therefore, GNNs have been widely adopted in the field of medical imaging.

Through extensive literature investigation, we find that the application of GNNs in bioinformatics has been rapidly developed in recent years, and the number of research papers in this field has shown a rapid growth. **Figure 2** shows the statistics of published GNN articles in bioinformatics from 2015 to 2020.

Although previous studies have surveyed deep learning applications in bioinformatics (Wood and Hirst, 2004; Min et al., 2016; Sun et al., 2019) recently published a review paper of GCN in bioinformatics, these studies are confined to a narrow field such as drug discovery in reference. To the best of our knowledge, this is the first effort to review the application and development of GNNs for bioinformatics. The rest of this paper is carried out from the following aspects: (1) several standard GNN models are introduced for a better understanding on how GNNs extract potential information from biological data; (2) three levels of GNN applications (node level, edge level, and graph level) are illustrated in specific biological tasks. Meanwhile, existing applications of GNNs for bioinformatics are classified based on various biological problems and data forms, and the role of GNNs in these studies is discussed; (3) according to the discussions on existing studies, we summarize the limitations of this field, including imbalances of biological data, and the methodological and interpretability challenges of GNNs. Finally, future research directions on various applications are proposed.

MODEL PRINCIPLE AND DEVELOPMENT

There have been various GNN models for processing graph structure data. In this section, we present the original GNN and its variant models, including graph convolutional network (GCN), graph attention network (GAT), and graph autoencoders.

Graph Neural Network

Gori et al. (2005) proposed a novel neural network model capable of processing graph structure data–graph neural network in 2005. As a pioneering work of deep learning methods in non-Euclidean spaces, the goal of GNN is to learn how to generate an accurate state embedding vector h_i, that is, the state of the node is constantly updated with the information dissemination mechanism on the graph; each update depends on the state information of the neighboring nodes at the previous time.
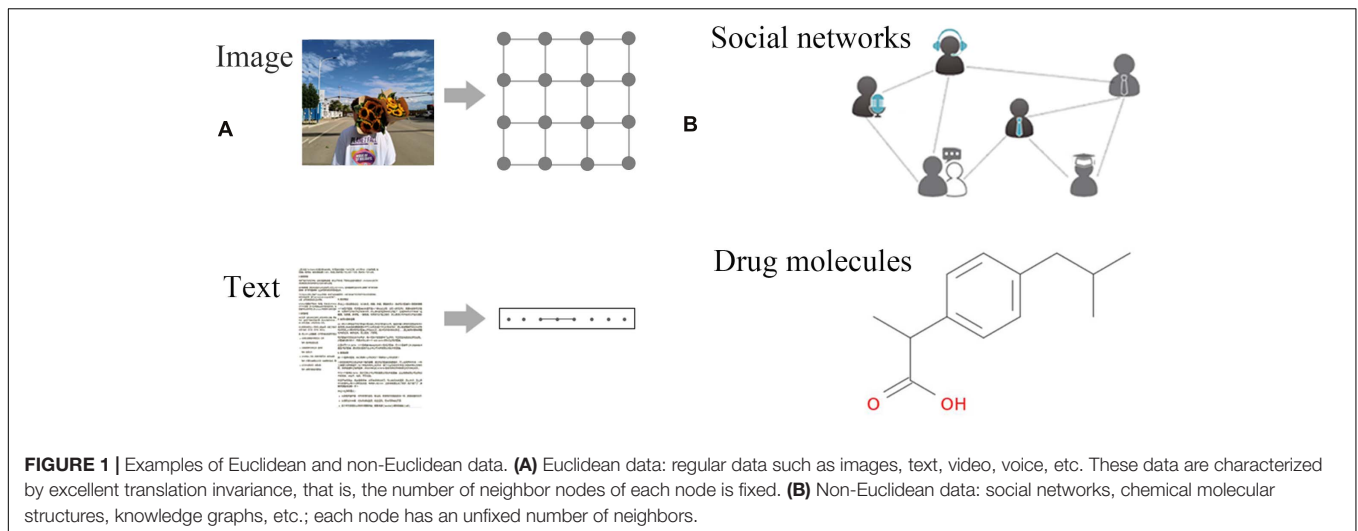
The related concepts are introduced as follows: let the input graph be $G = (V, E, X_V, X_E)$, $V = \{v_1, v_2, ..., v_n\}$ represents the set of nodes, and $E = \{(i, j) \,|\text{when } v_i \text{ is adjacent to } v_j\}$ is the set of edges. x_i denotes the feature vector of node v_i, and $X_V = \{x_1, x_2, ..., x_n\}$ is the set of feature vectors of all nodes. $x_{(i,j)}$ denotes the feature vector of edge (i, j), and $X_E = \{x_{(i,j)} | (i, j) \in E\}$ is the set of feature vectors of all edges.

The input graph G is converted into a dynamic graph $G^t = (V, E, X_V, X_E, H^t)$ in the graph neural network model, where $t = 1, 2, ..., T$ represents time and $H^t = (h_1^{(t)}, h_2^{(t)}, ..., h_n^{(t)})$, $h_i^{(t)}$ represents the state vector of node v_i at time t, which depends on the graph G^{t-1} at time $t - 1$. The equation of $h_i^{(t)}$ is as follows:

$$h_i^{(t)} = f_w(x_i, x_{co(i)}, h_{ne(i)}^{t-1}, x_{ne(i)}) \tag{1}$$

where $f_w(\cdot)$ denotes the local transformation function with parameter w, $x_{ne(i)}$ is the set of feature vectors of all nodes adjacent to node v_i, $x_{co(i)}$ is the set of feature vectors of all edges connected to node v_i, and $h_{ne(i)}^{(t)}$ is the set of state vectors of all nodes adjacent to node v_i at time t. GNN updates the node status in an iterative manner, and this process is shown in **Figure 3**.

The long-term dependency problem of the original GNN (it is difficult for the node features to affect the state after multiple

FIGURE 1 | Examples of Euclidean and non-Euclidean data. **(A)** Euclidean data: regular data such as images, text, video, voice, etc. These data are characterized by excellent translation invariance, that is, the number of neighbor nodes of each node is fixed. **(B)** Non-Euclidean data: social networks, chemical molecular structures, knowledge graphs, etc.; each node has an unfixed number of neighbors.

updates) makes it laborious to learn the deep structure. Based on GNN, some variant models appear successively.

Graph Convolutional Networks

The existing GCN models can be divided into two categories: spectral-based and spatial-based GCN. Spectral-based GCN is defined from the perspective of graph signal processing, which exploits the principle of Laplacian and Fourier transform to map the irregular structure of a graph to a regular Euclidean space for convolution operation. Spatial-based GCN directly utilizes the information dissemination mechanism on the graph to define the

FIGURE 2 | Statistics of published graph neural network (GNN) articles in bioinformatics from 2015 to 2020. The orange bar shows the total number of GNN papers in bioinformatics that year (note: the number of papers in 2020 is counted until October). The remaining colors, in turn, represent the number of papers related to GNNs in bioinformatics in terms of disease association prediction, drug research, and medical image processing, which are the components of the orange bar.

convolution operation, and its propagation method is similar to the original GNN. These two models will be discussed next.

Spectral-Based GCN

Spectral-based GCN uses the graph Laplace matrix as an important tool to extend the Fourier transform to the graph structure. Let A be the adjacency matrix with weighted undirected graph G, and the element $A(i, j)$ in the i-th row and j column of the matrix is the weight of the edge (i, j). Degree matrix D is defined as follows:

$$D(i,j) = \sum_{j=1}^{n} A(i,j) \tag{2}$$

The symmetric normalized Laplace matrix of graph G is defined as follows:

$$L = I - D^{-\frac{1}{2}} A D^{-\frac{1}{2}} \tag{3}$$

As a real symmetry positive semidefinite matrix, L can be decomposed into:

$$L = U \Lambda U^T \tag{4}$$

where $U = (u_0, u_1, \cdots, u_{n-1})$ is the eigenvector matrix, and
$\Lambda = \begin{bmatrix} \lambda_1 & \ldots & 0 \\ \ldots & \ldots & \ldots \\ 0 & \ldots & \lambda_n \end{bmatrix}$ is the diagonal matrix of eigenvalues. The
normalized Laplacian matrix L and its eigenvector u form an orthogonal space as the Fourier transform ecosystem on the graph. The graph signal represents the feature vector of all nodes in the graph, expressed by $x = (x_0, x_1, \cdots, x_{n-1}) \in R^n$. The Fourier transform of the graph signal x is given below.

$$\widehat{x} = U^T x \tag{5}$$

Calculate the convolution between the two signals as:

$$x * g = U((U^T x) \odot (U^T g)) \tag{6}$$

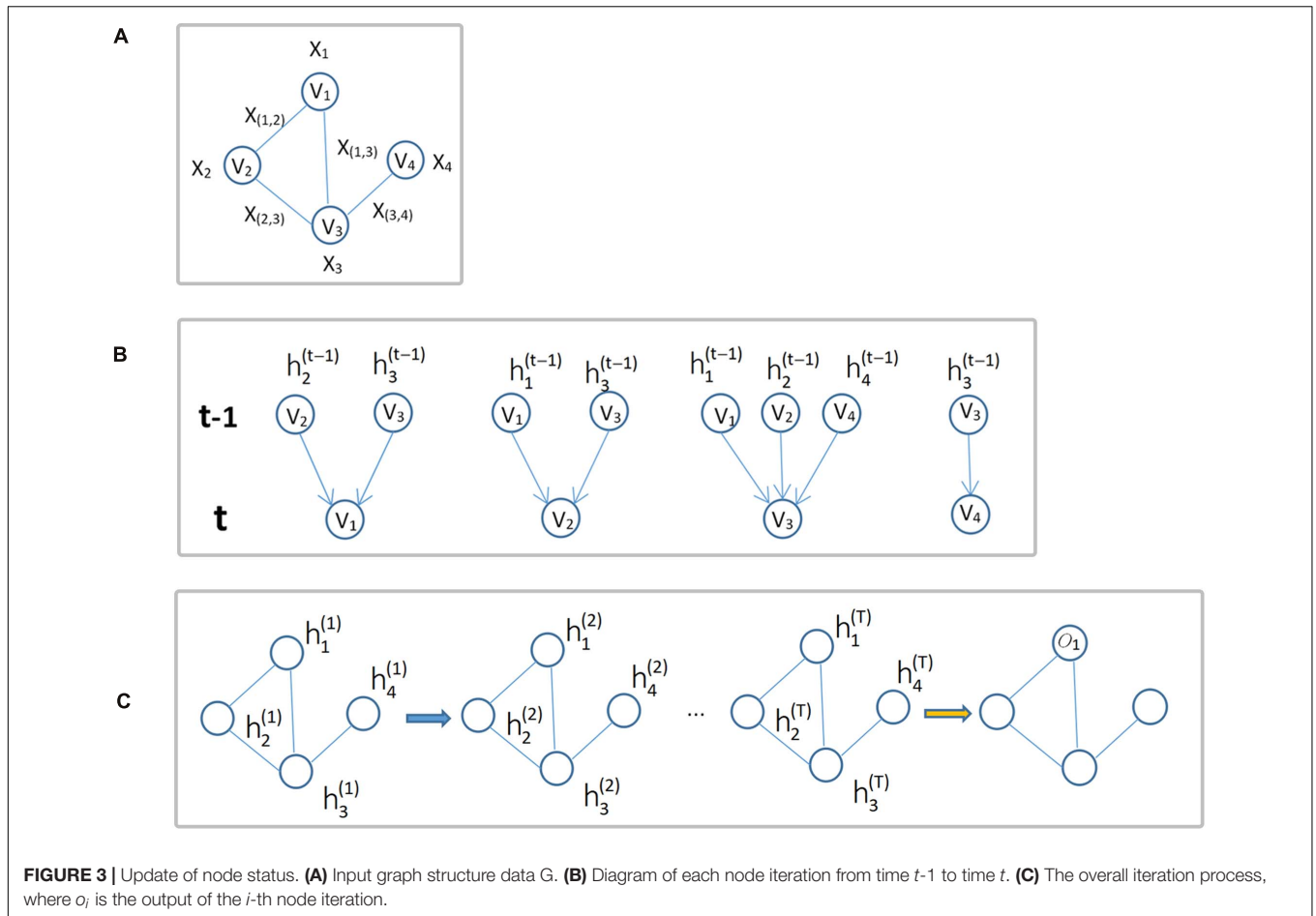**FIGURE 3** | Update of node status. **(A)** Input graph structure data G. **(B)** Diagram of each node iteration from time t-1 to time t. **(C)** The overall iteration process, where o_i is the output of the i-th node iteration.

If $\mathbf{g}_\theta = diag(\mathbf{U}^T\mathbf{g})$ is used as a filter for the graph signal x, we can define the graph convolution as follows:

$$\mathbf{x} * \mathbf{g}_\theta = \mathbf{U}\mathbf{g}_\theta(\Lambda)\mathbf{U}^T\mathbf{x} \qquad (7)$$

This is the first generation of a spectral-based GCN model proposed by Bruna et al. (2013), which contains multiple convolutional layers. Spectral-based GCN maps the graph structure to Euclidean space through the Laplacian matrix of the graph to realize the spectral convolution of the graph. Nevertheless, due to matrix–vector multiplication, the computational complexity of the model is relatively high, which is $O(n^2)$. To solve this problem, Defferrard et al. (2016) proposed a model ChebNets that uses a K-degree polynomial filter in the convolutional layer. The κ-th polynomial filter of the spectrum in the model is expressed as shown below.

$$\mathbf{g}_\theta = \sum_{k=0}^{K} \theta_k \lambda_l^k \qquad (8)$$

The K-th-order polynomial filter of the spectrum is expressed in the node domain as aggregating K-th-order neighborhoods to maintain spatial locality, and the number of filter parameters is also controlled to $O(K) = O(1)$. In order to further reduce the computational complexity, the model uses Chebyshev

polynomial $T_k(x) = 2T_{k-1}(x) - T_{k-2}(x)$ for recursive calculation, where $T_0(x) = 1$ and $T_1(x) = x$. Therefore, the convolution of the graph signal x and the filter is defined as shown below.

$$\mathbf{x} * \mathbf{g}_\theta = \mathbf{U}(\sum_{k=0}^{K} \theta_k T_k(\widetilde{\mathbf{L}}))\mathbf{U}^T\mathbf{x} \qquad (9)$$

As a simplification of the above-mentioned ChebNets, the graph convolutional network model proposed by Kipf and Welling truncates the Chebyshev polynomial to one time (Kipf and Welling, 2016a). For numerical stability, the adjacency matrix $\mathbf{A}$ is adjusted to obtain $\widetilde{\mathbf{A}}$, which results in a simplified combined convolutional layer.

$$\mathbf{H} = \mathbf{X} * \mathbf{g}_\theta = f(\widetilde{\mathbf{D}}^{-\frac{1}{2}}\widetilde{\mathbf{A}}\widetilde{\mathbf{D}}^{-\frac{1}{2}}\mathbf{X}\Theta) \qquad (10)$$

where $\widetilde{\mathbf{A}} = \mathbf{I} + \mathbf{A}$, and $\widetilde{\mathbf{D}}_{ij} = \sum j\widetilde{\mathbf{A}}_{ij}$, $f(\cdot)$ is the activation function; Θ is the filter parameter matrix.

Although the above-mentioned methods based on frequency domain perform well in feature extraction, their limitations are also obvious. First of all, due to the problem of data volume, the method based on the Laplacian matrix of graphs is hard to calculate on large graphs. Second, the trained GCN can only

be applied to a fixed graph structure rather than to arbitrary-structure graphs.

Spatial-Based GCN

The above-mentioned methods are based on the convolution theorem and define the graph convolution in the spectral domain, while the spatial method starts from the node domain and aggregates each central node and its neighboring nodes along the edge. Diffusion convolutional neural network (DCNN) (Atwood and Towsley, 2015) proposes that convolution is a process of diffusion between nodes and uses the k-hop transition probability obtained after random walking to define the weight between nodes. The structure of layer m is as follows:

$$H^{(m+1)} = f(WP^k H^m) \qquad (11)$$

where P^k denotes the k-hop reachability probability between two nodes in a random walk, and W is a learnable model parameter. DCNN describes the high-order information between nodes, but it is hard to extend to a large graph because the computational complexity of the model is $O(n^2 K)$.

GraphSage (Hamilton et al., 2017) randomly samples the neighboring nodes so that the neighboring nodes of each node are less than the given number of samples so as to adapt to the application on large-scale networks. The graph convolution operation is as follows:

$$h_v^{(k)} = \sigma(W^k \cdot f_k(h_v^{(k-1)}, \{h_u^{(k-1)}, \forall u \in S_{N_{(v)}}\})) \qquad (12)$$

where $f_k(\cdot)$ is the aggregate function, and $S_{N_{(v)}}$ is the random sampling result of neighbors of node v. GraphSage gives a variety of forms of aggregation functions, which are mean aggregator, LSTM aggregator, and pooling aggregator.

There are also some studies aimed at defining the general framework of GCNs. Among them, mixture model networks (MoNet) (Monti et al., 2017) focus on the lack of translation invariance on the graph and map the local structure of each node to a vector of the same size by defining a mapping function. Finally, learn the shared convolution kernel on the result of the mapping. The message passing neural network (MPNN) (Gilmer et al., 2017) is based on information dissemination and aggregation between nodes and proposes a framework by defining a general form of the aggregation function.

Mixture model networks defines a coordinate system on the graph and expresses the relationship between nodes as a low-dimensional vector in the new coordinate system. At the same time, a weight function is defined on all adjacent nodes centered on a node, and a vector representation of the same size is obtained for each node.

$$D_j(x)f = \sum_{y \in N(x)} w_j(u(x, y))f(y), \; j = 1, \cdots, J \qquad (13)$$

where $N(x)$ represents the set of adjacent nodes of x, $f(y)$ represents the value of node y on the signal f, $u(x, y)$ refers to the low-dimensional vector representation of the node relationship in the coordinate system u, w_j represents the j-th weight function, and J represents the weight function number. This operation

makes each node get a J-dimensional representation, and the shared convolution kernel is defined on this.

Differently from MoNet, MPNNs point out that the core of graph convolution is to define the aggregation function between nodes using the aggregation function to get the local structure expression of each node and its neighboring nodes and then applying the update function to itself and the local structure expression to get the new expression of the current node. The convolution operation is as follows:

$$h_v^{(k)} = U_k(h_v^{(k-1)}, \sum_{u \in N(v)} M_k(h_v^{(k-1)}, h_u^{(k-1)}, x_{vu}^e)) \qquad (14)$$

where U_k and M_k are update function and aggregate function, respectively. The aggregate function learned under the spatial framework can be adapted to the task and the specific graph structure and has greater flexibility.

Graph Attention Networks

In order to solve the shortcomings of GCN and its similar structure, GAT (Veličković et al., 2017) introduces the attention mechanism into the propagation step of the graph to learn the weight between two connected nodes. In the GAT model, input the set $x = \{x_1, x_2, ..., x_n\}$ of the node feature to an attention layer; a new learned set $h = \{h_1, h_2, ..., h_n\}$ of node feature will be output. The attention coefficient of edge (i, j) is represented by α_{ij}, and the equation is as follows:

$$\alpha_{ij} = \frac{exp(\text{LeakyRELU}(a^T [Wx_i || Wx_j]))}{\sum_{k \in N_i} exp(\text{LeakyRELU}(a^T [Wx_i || Wx_k]))} \qquad (15)$$

where N_i is a set composed of adjacent nodes of node V_i, a represents the learnable weight vector, and W is a shared linear transformation weight matrix. The output features of each node are calculated by the following equation:

$$h_i = \sigma(\sum_{j \in N_i} \alpha_{ij} Wx_j) \qquad (16)$$

Multi-head attention expands the attention layer into K independent attention mechanisms to make the learning process of self-attention more stable, and the final expression is given as shown below.

$$h_i = \sigma(\frac{1}{K} \sum_{k=1}^{K} \sum_{j \in N_i} \alpha_{ij}^k W^k x_j) \qquad (17)$$

Parallel computing operations give GAT a higher efficiency, and the applicability of GAT on completely unknown graphs makes up for the limitation of spectral GCN.

Graph Autoencoder Networks

The wide application of autoencoder (AE) and its variants in the field of unsupervised learning has led to an increasing number of AE-based graph generation models. Sparse autoencoder

(SAE) (Tian et al., 2014) is the source of AE-based graph neural network. It uses the following $\mathcal{L}_2$ reconstruction loss:

$$\min_\theta \mathcal{L}_2 = \sum_{i=1}^{N} ||\boldsymbol{P}(i,:) - \widehat{\boldsymbol{P}}(i,:)||_2, \qquad (18)$$

$$\widehat{\boldsymbol{P}}(i,:) = G(\boldsymbol{h}_i), \quad \boldsymbol{h}_i = F(\boldsymbol{P}(i,:)), \qquad (19)$$

where $\boldsymbol{P}$ is the transition matrix, and $\widehat{\boldsymbol{P}}$ is the reconstruction matrix; $\boldsymbol{h}_i \in \boldsymbol{R}^d$ represents the low-dimensional representation of node v_i, and F and G are encoder and decoder, respectively. d is the dimension of hidden variables, and $\mathcal{N}$ is the number of nodes and $d \ll N$. On the basis of SAE, Wang et al. (2016) proposed a structural deep network embedding (SDNE) model, which modified the reconstruction loss function as shown below.

$$\min_\theta \mathcal{L}_2 = \sum_{i=1}^{N} ||(\boldsymbol{A}(i,:) - G(\boldsymbol{h}_i)) \bigodot \boldsymbol{b}_i||_2 \qquad (20)$$

when $\boldsymbol{A}(i,j) = 0$, $b_{ij} = 1$; otherwise, $b_{ij} = \beta > 1$, and β is a hyperparameter. The supervised learning method is used to learn the first-order approximation. The loss function is as follows:

$$\mathcal{L}_1 = \sum_{i,j=1}^{N} (\boldsymbol{A}(i,j)|\,|\boldsymbol{h}_i - \boldsymbol{h}_j|\,|_2^2) \qquad (21)$$

Finally, the loss function of SDNE is obtained:

$$\mathcal{L} = \mathcal{L}_2 + \alpha \mathcal{L}_1 + \mathcal{L}_{\text{reg}} \qquad (22)$$

The variational autoencoder (VAE) (Kingma and Welling, 2013) is suitable for learning graph node representation without supervision information. Kipf and Welling (2016b) proposed a variational graph autoencoder (VGAE), which was the first time that VAE was extended to graphs. The generation model of VGAE is as follows:

$$p(\boldsymbol{A}|\boldsymbol{H}) = \prod_{i=1}^{N} \prod_{j=1}^{N} p(\boldsymbol{A}(i,j)|\boldsymbol{h}_i, \boldsymbol{h}_j) \qquad (23)$$

$$p(\boldsymbol{A}(i,j) = 1|\boldsymbol{h}_i, \boldsymbol{h}_j) = \text{sigmoid}(\boldsymbol{h}_i^T, \boldsymbol{h}_j) \qquad (24)$$

Variational graph autoencoder learns parameters by minimizing the lower bound of variation L.

$$L = E_{q(\boldsymbol{H}|F^V, \boldsymbol{A})}[\log p(\boldsymbol{A}|\boldsymbol{H})] - KL[q(\boldsymbol{H}|F^V, \boldsymbol{A})||p(\boldsymbol{H})] \qquad (25)$$

Among them, $KL(\cdot)$ represents the Kullback–Leibler divergence function, which is used to measure the distance between two distributions. The time complexity of the model is $O(N^2)$.

APPLICATION PRINCIPLE OF GNNS IN BIOINFORMATICS
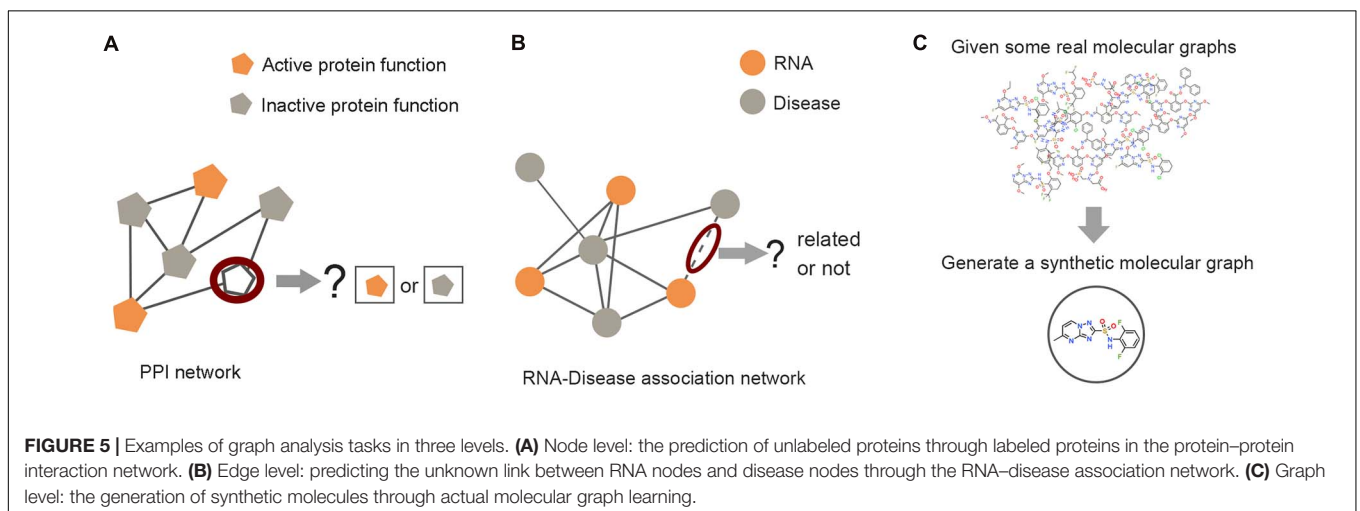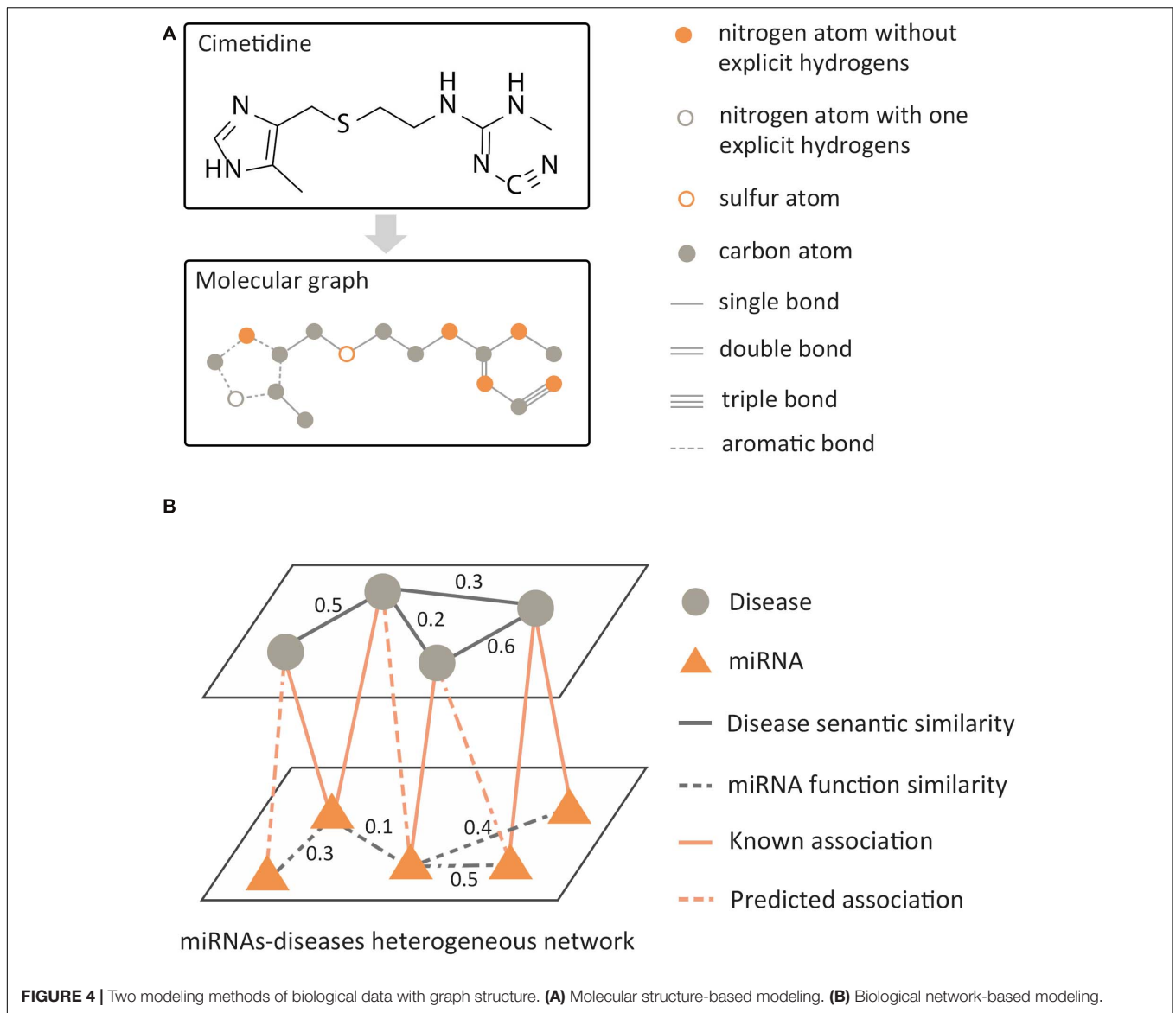
Modeling Methods

In the data analysis of bioinformatics, the biological data with graph structure can be modeled in two ways: molecular structure-based modeling and biological network-based modeling. For molecular structured-based modeling, atoms or valid chemical substructures (Jin et al., 2018) are used as nodes, bonds are used as edges, and then the molecular graph is constructed as shown in **Figure 4A**. Molecular graphs have a wide range of applications in predicting the properties of molecules and *de novo* molecular design. For biological network-based modeling, various entities are used as nodes, such as gene, disease, RNA, etc. The edges between nodes mean that there is a known association between pairs of entities, such as miRNA–disease interaction. Then, a relational network is generated, as shown in **Figure 4B**. GNNs are known to have a perfect performance in extracting potential information from graph structures, so they can process omics data in the biological field, including genomics, proteinomics, RNomics, and radiomics. Combined with the above-mentioned two modeling methods, applying GNNs in these omics data can be employed for a variety of tasks, such as molecular property prediction, *de novo* molecular design, link prediction, node classification in biological networks, etc.

The Tasks of GNNs in Bioinformatics

Based on the modeling methods cited above, the structural information learned by GNNs provided a basis for different levels of graph analysis tasks: node level, edge level, and graph level (shown in **Figure 5**).

Node Level

Node classification is the typical task at the node level (**Figure 5A**), which can be performed by way of supervised learning, unsupervised learning, and semi-supervised learning. As the most commonly used method of node classification, semi-supervised learning combines the characteristics of supervised learning and unsupervised learning. Compared with supervised learning and unsupervised learning, semi-supervised learning on the graph extracts high-level node representations through information dissemination, which does not need to label all nodes and make good use of some known associated information. This setting is powerful for the task of inferring the association between entities in the biological network. For example, Ioannidis et al. (2019) constructed multiple protein–protein interaction (PPI) networks based on protein connectivity for different types of cells and proposed a graph residual neural network (GRNN) architecture for semi-supervised learning over multi-relational graphs. The influence of different relations was measured by learnable parameters. For the protein function prediction in generic cell, brain cell, and circulation cell data sets, GRNN had a macro F1 score of 0.86, 0.77, and 0.80, which was far better than the baseline model. In allusion to population disease prediction, Parisot et al. (2017) modeled population information as a graph, medical imaging data as the feature of the subject

FIGURE 4 | Two modeling methods of biological data with graph structure. **(A)** Molecular structure-based modeling. **(B)** Biological network-based modeling.

FIGURE 5 | Examples of graph analysis tasks in three levels. **(A)** Node level: the prediction of unlabeled proteins through labeled proteins in the protein–protein interaction network. **(B)** Edge level: predicting the unknown link between RNA nodes and disease nodes through the RNA–disease association network. **(C)** Graph level: the generation of synthetic molecules through actual molecular graph learning.

node, and phenotype data as the weight of the edge. GCN was utilized to simultaneously model individual features and associations between subjects from potentially large populations. In the setting of semi-supervised learning, conditioning the GCN on the adjacency matrix provides the representation learning for all nodes. Compared with the standard linear classifier, their work improved the quality of prediction.

Edge Level

As the main task of the edge level, link prediction is defined as, given some graphs, an edge prediction model is trained based on the features of nodes or edges for predicting the connectivity probability between node pairs in these graphs or newly given graphs, as indicated in **Figure 5B**. The link prediction task has captured the attention of different research fields due to its broad applicability. Predicting the interaction between biological entities from complex biological networks also plays an important role in the research of bioinformatics and has become increasingly important and more challenging. The GNN models are also effective for solving the link prediction tasks. Zhang and Chen (2018) proposed the SEAL (learning from subgraphs, embeddings, and attributes for link prediction) model based on information dissemination, which used GNN to replace the fully connected neural network in the traditional Weisfeiler–Lehman neural machine method and learned general graph structure features from local subgraphs. Its performance on public biological network data sets such as yeast, *Caenorhabditis elegans*, and *Escherichia coli* was superior to the traditional graph embedding models. In addition, GCN has been utilized to predict various interactions in biological networks. For example, PPI networks with a small amount of label information that were encoded to predict the relationship between drugs and diseases (Bajaj et al., 2017), disease similarity networks, and microRNA (miRNA) similarity networks were built to indicate the association between miRNA and disease by VGAE (Ding et al., 2020).

Graph Level

The task of graph level is mainly related to graph generation (**Figure 5C**). Learning to generate graph structure data by training on a set of representative data is the core of graph generation tasks. For discovering new chemical structures, a graph generative model based on GNNs was first proposed with the motivation of molecular graphs generation. Simonovsky and Komodakis (2018) combined GNN and VAE to propose GraphVAE, which was used for small-scale molecular graph generation. Experiments on the QM9 database and ZINC database proved that GraphVAE has a higher accuracy than the previous methods. Jin et al. (2018) proposed a junction tree variational autoencoder (JT-VAE), which allowed the model to gradually expand the molecule while maintaining the chemical validity of each step. The experimental results on the ZINC database had shown that JT-VAE could generate better results than the traditional model and GraphVAE. MolGAN (De Cao and Kipf, 2018) is a generative model for small graphs, which is able to generate discrete graph structures and promote the generation of molecules with specific chemical properties through reinforcement learning methods (**Figure 6**).

The MolGAN model produced nearly 100% effective compounds in experiments on the QM9 chemical database. In practice, the abilities of graph generative models for protein structure prediction and chemical molecular map generation play an important role in related applications, such as drug design and protein structure design.

In general, as a novel type of graph embedding method, GNNs can perfectly integrate the features of nodes and the structural information between nodes in various specific applications. To better illustrate the application status and mechanisms of GNNs in bioinformatics, more comprehensive existing research on these three levels are summarized for specific biological problems, which are discussed in the next section.

TYPICAL APPLICATION OF GNNS IN BIOINFORMATICS

Based on various biological tasks, the existing application of GNNs in bioinformatics can be categorized into three typical topics: disease association prediction, drug development and discovery, and medical imaging. Note that these applications are also based on the three levels of graph analysis tasks in the previous section. In this section, the way and the development of GNNs handle representative problems are described in more detail.

Disease Association Prediction

Discovering the associated factors with various diseases is an important task in bioinformatics. At present, the existing methods of disease association prediction mainly include matrix decomposition (Koren et al., 2009; Wang et al., 2017) network propagation (Lee et al., 2011; Guan et al., 2012; Li and Li, 2012; Sun et al., 2014; Zhou et al., 2015), and machine learning (Luo et al., 2016; Zhou and Skolnick, 2016; Frasca, 2017; Xuan et al., 2019b; Jiang and Zhu, 2020). Essentially, some machine learning methods are also based on similarity measures and matrix decomposition. Nevertheless, matrix factorization methods map the features of entities to a latent space but ignore the representation of topological relationships between entities. In other methods, the shallow models ignore the rich structural information in disease-related networks, which ultimately affects the quality of entity feature representation. Recently, GNNs have been used to capture the nonlinear relationship between diseases and other entities in biological networks. More and more methods have introduced convolution operations into heterogeneous networks for extracting features of local sub-graphs. All of the studies discussed in this section have directly or indirectly contributed to the development of deep learning methods in the field of disease prediction. Different biological networks were constructed in these studies, which were based on RNA–disease associations, disease–gene associations, and other association information.

RNA–Disease Association

A large amount of evidence has shown that microRNAs (miRNAs), long non-coding RNAs (lncRNAs), circular RNAs (circRNAs), and Piwi-interacting RNA are widely involved in the

FIGURE 6 | Flow chart of MolGAN. The generator extracted a sample from the prior distribution and generated a molecular graph corresponding to a specific compound. The graph neural network-based discriminator and reward network directly operate on the representation of the graph structure. The discriminator learns to distinguish whether the molecular graph comes from the real data set or the generator. The reward network assigned rewards to each molecular graph based on samples and optimized the generation process through reinforcement learning methods.

FIGURE 7 | Flow chart of DimiG in RNA–disease association. The protein-coding gene (PCG)–miRNA association network was used as the input of the two-layer graph neural network. The PCGs were labeled, and the miRNAs were unlabeled. The yellow nodes in the network represented the weighted sum of neighbor embeddings. The final output could infer the probability between diseases and unlabeled miRNAs.

occurrence and development of diseases (Xuan et al., 2019a; Li J. et al., 2020; Wang L. et al., 2020; Zheng et al., 2020). Therefore, the identification of these RNA–disease associations plays a crucial role in exploring the pathogenesis of complex diseases. The RNA and disease data analysis methods based on computational models make up for the high cost and time-consuming defects of biological experimental verification methods.

Starting in 2019, GNNs have been introduced into this type of research. Pan and Shen (2019) proposed a semi-supervised multi-label graph convolution model (DimiG), which did not rely on known association information between miRNAs and diseases. DimiG integrated multiple networks related to protein-coding genes and used network knowledge transfer to indirectly predict the association between miRNAs and diseases; taking DimiG as an example, **Figure 7** shows how GCN uses the associated information in the network to generate features for unlabeled nodes. Li J. et al. (2020) used GCN to learn the feature representations of miRNAs and diseases from the miRNA

functional similarity network and disease semantic similarity network, respectively, and utilized the neural induction matrix to generate an association matrix, combining known miRNAs and disease association information to train the model. This model can predict all miRNAs related to breast cancer without any known related miRNAs. Based on the similarity method, the association prediction between RNAs and diseases can also integrate more useful information. Li C. et al. (2019) integrated miRNA–disease, miRNA–gene, disease–gene, and PPI networks. Furthermore, based on the information extracted by GCN, the top 10 unknown interactions between miRNAs and diseases were analyzed. Using the FastGCN algorithm and the Forest by Penalizing Attributes (Forest PA) classifier, Wang L. et al. (2020) can accurately predict potential circRNA disease associations. In order to better learn the hidden representation of node features, Zhang J. et al. (2019) used GCN combined with an attention mechanism to extract domain features and conducted experimental tests on two different RNA disease networks. There have also been some studies that used the autoencoder method on the graph to reconstruct node features. Wu et al. (2020) used GCN as an encoder to learn the feature representation of lncRNAs and diseases from the bipartite graph associated with lncRNA–disease, and the score of the lncRNA–disease interaction was calculated from the inner product of the two potential factor vectors. In the research of Ding et al. (2020), VGAE was used to reduce the noise effect caused by randomly selecting negative samples. In the prediction of disease-related RNAs with limited known data, the integration of multi-view information can help us understand complex biological networks more comprehensively. Therefore, to capture a deeper interaction mode between multiple related data, the integration method of different types of data by graph deep learning model needs to be further explored.

Disease–Gene Association

Single-cell RNA sequencing technology provides gene expression data for a single cell. GNNs can infer the interaction between cells (Jiahua et al., 2020; Zeng et al., 2020; Wang et al., 2021) and simulate cell differentiation (Bica et al., 2020) and disease state prediction (Ravindra et al., 2020). Precise gene–disease association prediction can help researchers reveal the function of disease-causing genes and provide evidence for disease prevention. Prioritizing candidate genes for various diseases is able to accelerate the development of early treatments and solve the DGP problem to a certain extent. Rao et al. (2018) proposed a rare disease gene sequencing method, which was different from the previous association network method. They integrated the combination pairwise ontological and curated associations into a heterogeneous network and used the frequency qualifier from Orphanet to calculate the edge weights. The qualifiers included terms such as "obligation," "very frequent," and "frequent." Since the learning algorithm outlined in the standard VGAE does not focus on learning the relationships between different node types, Singh and Lio (2019) proposed a constrained VGAE variant for predicting specific node associations on the disease–gene association network by improving the optimization objectives of an algorithm. Wang et al. (2019a) defined a new cluster loss

function and a dropout mechanism based on the GCN and graph embedding method to improve the generalization ability.

Although a large amount of medical data has been reserved in various database, accurate prediction of cancer remains a challenge. As a group of complex diseases, cancer is caused by multiple gene defects, and there is synthetic lethality between genes. Therefore, the interaction network of genes plays an important role in cancer prediction (Iglehart and Silver, 2009). In order to analyze the underlying mechanism of cancer, Schulte-Sasse et al. (2019) initially used GCN to classify and predict cancer genes. Layer-wise relevance propagation was used to identify the gene input signals and the network topology of the learned model, which is in the neighborhood of a gene. The synthetic lethality between genes is extremely sparse. In order to solve the over-fitting problem, Cai et al. (2020) proposed a new GCN model based on fine-grained edge dropout and coarse-grained node dropout to reduce the over-fitting in sparse graphs. Chereda et al. (2021) combined the PPI network and gene expression data for patients and utilized GCN to classify the nodes in the patient's sub-network for predicting breast cancer metastasis. In the classification of breast cancer subtypes, there are also related studies based on local GCN, which was used to combine with the PPI network and the gene expression matrix information of multiple patients (Rhee et al., 2018). The correlation generated by the GNNs for each data point not only improves the interpretability of the model but also makes it more advantageous in predicting tasks related to patient-specific disease networks.

Others

In addition to the studies listed above, GNNs are also introduced into some research of other related fields, for instance, the discovery of disease proteins (Eyuboglu and Freeman, 2004). The disease protein prediction problem can naturally be defined as a semi-supervised classification problem on the protein–protein interaction network. The realization of the neighborhood positioning for visualized disease pathways proved that most diseases do not have obvious neighborhood positioning. Some studies utilized graph structures to model RNA secondary molecular structures for RNA classification (Rossi et al., 2019) and RNA-binding proteins prediction (Uhl et al., 2019; Yan et al., 2020), where bases were considered as nodes in the graph and phosphodiester bonds and hydrogen bonds were two different types of edges. In other studies, miRNA, lncRNA, and other elements were used to construct heterogeneous networks for predicting the interactions between miRNA and lncRNA as well as lncRNA-targeted genes. For multi-group biomedical data classification, a weighted patient similarity network was constructed based on various omics data and cosine similarity method (Wang T. et al., 2020), and GCN performs a feature extraction on these networks so as to find the cross-omics correlation in label space for integrating multi-omics effectively.

Drug Development and Discovery

The drug development process mainly includes drug target determination, lead compound discovery and optimization, candidate drug determination, preclinical research, and clinical

research (Vohora and Singh, 2017). However, the lack of drug targets, the poor clinical transformation of animal models, disease heterogeneity, and the inherent complexity of biological systems have made drug development a long and arduous process. The purpose of modern drug development is to speed up the intermediate steps through machine learning methods so as to save development costs. Therefore, more and more researchers tend to utilize machine learning models for predicting early molecular properties, which can tremendously reduce the workload of later experiments. As the most concerning machine learning method, deep learning in the field of biomedicine showed the following limitations: first of all, most deep learning models cannot learn structural information directly from the original input data, which rely on high-quality, labeled data sets, and secondly, traditional CNN or other deep models have difficulties in directly processing unstructured data like molecular graphs, so the internal structure information of molecules is usually not fully taken into account. Therefore, GNNs, which extend deep learning methods to non-Euclidean domains, have become the latest method to deal with drug-related tasks.

Protein Structure and Function Prediction

Protein function research occupies an important position in life sciences, and most diseases are closely related to protein dysfunction. Anfinsen (1973) found that the denatured ribonuclease that only retained the primary structure could refold and restore biological activity, which indicated that the amino acid sequence representing the primary structure of the protein contains important information about the secondary and tertiary structure of proteins. At present, significant progress has been made in protein structure prediction. The most accurate structure prediction can fully clarify the biological mechanism of protein action on a molecular scale, and its application in drug development and other fields is of great significance to biochemical research.

High computational cost and interpretability are problems in common methods of molecular structure analysis, such as 3D CNN and 2D CNN. In recent years, some studies have shown the powerful capabilities of GNNs in learning the effective structure of proteins from simplified graphical representations. Zamora-Resendiz and Crivelli (2019) proposed a protein structure learning method that was more suitable for large data sets. Unlike the previous 3D and 2D representations, this model could apply to the natural spatial representation of molecular structures, which brought a high transferability to the application direction. Aiming at the inverse protein folding problem, Ingraham et al. (2019) proposed a protein design framework based on a similar graph attention method, which could construct a conditional generation model for a given target structure protein sequence directly, and greatly improved the design efficiency. For protein function prediction, there are two types of methods: based on protein structure (Ioannidis et al., 2019) and based on PPI networks (Gligorijevic et al., 2019). Similar to a previous work, Gligorijevic et al. (2019) modeled the protein structure as a graph to predict the protein function. Ioannidis et al. (2019) used a multi-relation diagram method based on PPI network modeling with semi-supervised learning. The structural characteristics of

a protein determine the breadth and complexity of its function. However, the large number of invalid fragments contained in the protein sequence may affect the judgment of its function. Using GNNs to integrate the feature of protein relationship networks is one of the ways to solve the problem of differences in protein sequences and functions.

Protein–Protein Interaction Prediction

Protein–protein interaction information is able to provide theoretical assistance for drug development indirectly. Fout et al. (2017) proposed a space-based convolution operator to predict the interface between protein pairs, which was suitable for graphs with any size and structure. In the identification of protein complexes, the high rate of "false positive/negative" in the PPI network makes the detection of protein complexes arduous. Therefore, Yao et al. (2020) proposed a denoising method based on a variational graph autoencoder. They embedded the PPI network into the vector space through multi-layer GCN and deleted some interactions with credibility below the threshold so as to obtain a reliable association network. The experimental results on multiple datasets showed that the recognition accuracy of protein complexes increases by 5–200%. Liu X. et al. (2020) used an unsupervised GNN to predict the changes in protein binding properties after mutations and recognized abnormal interactions between atoms without annotations. By improving on the existing sorting algorithm, Cao and Shen (2020) and Johansson-Åkhe et al. (2020), respectively, proposed a scoring mechanism for the evaluation of protein docking models and doctored peptides. The convolution operation on graph encodes the structure and features of protein into the graph embedding representation and aggregates information along the edges of the network nodes for association scores, which solves the spatial limitations of conventional convolution methods.

Ligand–Protein (Drug–Target) Interaction Prediction

Drug targets are relevant to the pathological state of diseases or biomolecules, so the identification of drugs and their targets is the core problem in the development of new drugs. Drug target interaction prediction is essentially the interaction prediction problem between ligands and proteins, and many related studies have been performed. Nonetheless, there exist some problems as follows: (1) Traditional machine learning algorithms express the prediction results in a binary classification, but the real association relationship is not limited to the binary level. For some target proteins that do not exist in the test set but appear in practical applications, the prediction accuracy cannot be grasped; (2) It is more difficult to deal with the chemical space where drug molecules can be synthesized; (3) The prediction results lack biological interpretation. Although the test results of the model seem good, it is still unconvincing; and (4) The information about the ligand bound to a specific protein is always easy to obtain, but there is a lack of data on the real negative ligand–protein relationship for learning.

In response to these existing problems, Feng et al. (2018) introduced GCN into drug target identification for the first time, which learned the molecular structure information of drugs, and combined protein information as input. This study realized

the prediction of the real-valued interaction strength between drugs and targets and solved the cold-target problem. There are also some studies similar to the general thinking of the above-mentioned method but differ in data processing (Gao et al., 2018; Nguyen et al., 2021). Miyazaki et al. (2020) provided a drug–target interaction prediction model that ligands were specifically targeting toward proteins without using true negative interaction information. Torng and Altman (2019) established an unsupervised graph autoencoder to learn the representation of protein pockets without relying on the target–ligand complexes, where features were extracted from the pocket graph and 2D ligand graph by GCN, respectively. Jiang et al. (2020) proposed an association prediction method that constructs a molecular graph and a protein contact graph. These graphs were based on the structural information of drug molecules and the sequence information of proteins, which were performed by a three-layer GCN for providing an accurate prediction.

Unlike the above-cited methods of representing drugs as graphs, Zhao et al. (2021) proposed a network-based prediction model that incorporated the drug–protein association network into existing methods. The feature information of each drug–protein pair learned by GCN was used as the input of a deep neural network for predicting the final label. In order to make up for the ignorance of cellular context-dependent effects in previous studies, Zhong et al. (2020) used the information in gene transcriptional profiles to indirectly predict drug-targeted binding.

Besides this, knowledge graphs are also critical in biomedicine, which can be processed by GNNs. In these studies, the result of all graph-based interaction predictions was inseparable from the quality of the knowledge graph, but the content of the knowledge graphs extracted in real life is complex and contains interference information. Therefore, Neil et al. (2018) proposed a model that could adapt to noise data and reduced the influence of noisy data on the overall prediction effect of the model by assigning low weights to unreliable edges.

Prediction of Molecular Properties
The prediction of molecular properties is a basic and important part of drug development. The first work to introduce graph convolution into the field of molecular properties learning (Duvenaud et al., 2015) was based on extended-connectivity circular fingerprints (ECFP), which created differentiable fingerprints to replace discrete operations in circular fingerprints and replaced hash functions with single-layer neural networks. The experimental results showed that, when the weights were large and random, this model exhibited a similar performance to ECFP, and as the weights were adjusted through training, its performance was better than ECFP. In view of the large space requirements of fingerprint-based methods and the large information noise of fingerprint encoding, Kearnes et al. (2016) proposed a molecular graph convolution method based on deep neural networks instead of molecular fingerprints. The molecular structure was represented by molecular graphs, and the distance between graphs forms the level of molecules. Although this method is not always superior in performance over molecular fingerprinting methods, it opens a new path for molecular

property prediction. In addition, the emergence of multi-task deep neural networks (MT-DNNs) makes neural networks more powerful in drug discovery. Liu et al. (2019) combined GCN with MT-DNNs to further improve the prediction accuracy and realized a completely data-driven deep learning method that did not rely on domain-specific feature descriptors or fingerprints for drug property prediction. It is well known that the electrostatic calculations are useful for the prediction of the chemical reactivity of molecules and their ability to form certain types of interactions. Rathi et al. (2019) proposed a method to generate electrostatic potential surfaces close to quantum mechanics quality for ligand molecules within the time frame of interactive drug design, which provided an effective tool for medicinal chemists and modelers. In the process of drug discovery, the false positives or false negatives of bioassay conclusions caused by unstable compounds in storage make it hard to complete the stability prediction of compounds. Differently from the traditional rule-based method, Li et al. (2019b) proposed an end-to-end, attention-based GCN model to predict the stability of compounds. The model dynamically learned structural information from molecular graphs instead of pre-defined structural features, thereby reducing the risk of false alarms. The graph convolution operation can capture the local features of molecular sub-structure effects, thereby generating accurate global descriptors from the composite structure data.

De novo Molecule Design
The ultimate goal of drug design is to discover molecules with ideal chemical properties. Nonetheless, the hugeness and complexity of chemical space and the discontinuity of the spatial structure of compounds make it demanding to explore the chemical space of new molecules. By reducing the consumption of labor costs, computer-aided drug design is dedicated to accelerating the process of *de novo* molecular design. Although the generative model in machine learning can effectively generate molecules based on SMILES strings (Weininger, 1988), it cannot effectively represent the topological information of molecular structure. Based on the above-mentioned analysis, GNNs can be directly used to generate molecular graphs. Therefore, GNN is a kind of high-precision and low-cost method to determine molecular properties by analyzing the topological information of graphs.

The arbitrary connectivity and discrete structures of the graph make it laborious to generate a graph from the vectors in continuous code space, but if the maximum number of nodes in the generated graph is constrained, it is still computationally controllable. **Figure 8** shows a continuous embedding method of VAE to generate small molecular graphs, which was proposed by Simonovsky and Komodakis (2018). This early generation mode avoids the difficulties that may be encountered in generating graphs; the upper bound of negative log-likelihood was minimized in the model training, but it only applies to the generation of small molecules. Another way is based on the probability description of GCN to generate the graph gradually (Li et al., 2018a) rather than directly generating the entire graph. Compared with the routine method, this method has achieved better results, but there are still challenges in

FIGURE 8 | Variational graph autoencoder (VGAE) in molecular graph generation. As the encoder of VGAE, the edge condition convolution embedded the original graph into a continuous vector space, and the decoder generated a probabilistic fully connected graph according to the predefined number of nodes and updated the parameters through approximate graph matching to improve the reconstruction ability of the autoencoder.

the generation of macromolecular graphs. Li et al. (2018b) proposed a conditional graph generation model and explored two types of graph generation architectures. One treated graph generation as a Markov process, and the other introduced molecular-level recursive units to increase the scalability of the model. You et al. (2018) combined the prior knowledge of the example molecule dataset to generate goal-directed molecules. This method integrated and expanded three ideas of graph representation, reinforcement learning, and adversarial training, where reinforcement learning and confrontation training were combined to form a unified framework so as to reach the desired goal by continuously guiding the generation process and limiting the output space according to basic chemical rules. The experimental results show that this method achieves the most superior performance in terms of optimizing chemical properties and constrained properties under conditions similar to known molecules. Differently from the node-by-node method of generating graphs, Jin et al. (2018) proposed a method of connecting tree self-encoding, which utilized effective sub-graphs as components to generate molecular graphs in two stages. The first stage generated a connection tree structure as a sub-graph component, and then these sub-graphs were combined into a complete molecular graph in the second stage. This model avoided the generation of invalid intermediate states of molecules and improved the work efficiency. Khemchandani et al. (2020) learned the interactive binding model from the binding data through GCN and proposed an attribute prediction module, which utilized a scoring mechanism to determine the more useful molecules with the specific attribute during the generation process.

At present, the graph-based method of molecular generation has more advantages than the grammar-based generation method. Although the new compound obtained by the molecular graph method has higher scores on various evaluation indicators, it has also been drawn into question. Besides this, the method of generating molecular graphs is still limited to 2D space, and the 3D information of molecules is completely ignored, which may become the focus in the future.

Drug Response Prediction

The combination of genomics data and drug information for drug response prediction has promoted the development of personalized medicine. Huang et al. (2020) combined GCN with autoencoder to predict the association between miRNA and drug resistance. In this study, the association prediction was regarded as a problem of semi-supervised learning, and a graph convolution model was built by combining the known miRNA expression profile and the drug structure fingerprint information. There are some other studies that focus on the effect of drug treatment on cell growth. Liu Q. et al. (2020) predicted the therapeutic effect of drugs on cancer cells by constructing a cancer cell information sub-network and a drug structure sub-network. Considering the complexity of cancer factors, Singha et al. (2020) integrated biological network, genomics, inhibitor analysis, and disease–gene association data into a large heterogeneous graph. Multiple graph convolution blocks and attention propagation were used to aggregate network topology information, and a graphic readout framework was constructed for predicting the final result. Hwang et al. (2020) adopted a similar structure with the study of Liu et al. (Huang et al., 2020), but introduced a set of information about the dosage and duration of drug administration for predicting drug-induced liver injury.

Drug–Drug Interaction Prediction

When a drug is taken with another drug, the expected efficacy of drugs may be significantly changed. Therefore, research on drug–drug interaction (DDI) is essential to reduce the occurrence of adverse drug events and maximize the synergistic benefits in the treatment of diseases. Obviously, the most practical way to explore the medicinal properties of drug combinations is computer-aided DDI detection. Zitnik et al. (2018) predicted the side effects between drugs from a multimodal heterogeneous network that was composed of PPI, drug–protein targets, and drug–drug interactions, where each side effect was represented by a different edge. Ma et al. (2018)

proposed a framework of a multi-view drug graph encoder based on the attention mechanism, which was used to measure the drug similarity. To make full use of the heterogeneous correlation between different views, each type of drug feature had been considered as a view, which was associated with a learnable attention weight in the similarity integration. This multi-view method could capture more similar information than the previous single view.

Medical Imaging

Medical images play an extremely important role in clinical disease diagnosis, classification, and treatment. In the field of medical imaging, deep learning methods combined with a computer-aided diagnosis are used for the early detection and evaluation of diseases. Similar to other image-related tasks, segmentation, classification, and recognition are the main tasks of concern in medical imaging. Meanwhile, image data can be represented as a graph structure appropriate for the use of GNNs. Therefore, GNNs have an extensive application space in the field of medical imaging, such as image segmentation (Gopinath et al., 2019; Wang et al., 2019b; Tian et al., 2020a,b), abnormal detection (Wu et al., 2019) of MRI images and pathological images, classification (Shi et al., 2019; Zhou et al., 2019; Adnan et al., 2020) and visualization (Levy et al., 2020; Sureka et al., 2020) of histological images, analysis of surgical images (Zhang et al., 2018), image enhancement (Hu et al., 2020), registration (Hansen et al., 2019), retrieval (Zhai et al., 2019), brain connection (Ktena et al., 2017, 2018; Li X. et al., 2019a; Mirakhorli and Mirakhorli, 2019; Grigis et al., 2020; Li et al., 2020; Zhang and Pierre, 2020; Zhang et al., 2021) and disease prediction (Parisot et al., 2017; Kazi et al., 2018,

2019a,b,c; Anirudh and Thiagarajan, 2019; Yang et al., 2019; Stankevičiūtė et al., 2020; Zhang and Pierre, 2020; Zhang et al., 2021), etc.

Image Segmentation

Before GNNs were introduced, CNN-based image segmentation technology that was widely used in various studies has been maturing. Ma et al. (2018) proposed a gated graph neural network for segmenting 3D images and utilized directed graph learning to predict the movement of points on the basis of the coarse segmentation, where a second segmentation was performed to obtain a smooth image. For processing surface data (such as MRI), Tian et al. (2020a) utilized spectral convolution for realizing cortical surface parcellation. The traditional spectral embedding can only be realized in orthogonal grid space, but this method was able to directly learn the surface features of the cortex. Gopinath et al. (2019) proposed an automatic and interactive prostate contour prediction method based on GCN blocks, which could consider the multiscale feature, and utilized a contour matching loss training method to preserve the details of the prostate boundary. In most cases, the segmentation model of deep learning adopts the pixel-wise segmentation method with high computational complexity. In contrast, the GNN method only uses object contours to segment, reducing the amount of calculation.

Brain Connectivity Research

Global similarity measurement between graphics represents the structural or functional connections within the brain by labeled maps, which is of great significance in the study of brain

FIGURE 9 | Similarity measure learning of graph neural network in functional brain graphs. The information contained in the fMRI image was integrated into the graph through the graph partition and connection matrix. Two graphs were, respectively, used as the input of the two graph convolutional networks with sharing parameters, and the output of the network was combined by the inner product layer. Finally, a fully connected layer outputs the estimated similarity between the graphs.

connectivity. At present, most models in deep learning use regular images as the default input data, but the processing of irregular brain connection images remains a problem. The pioneering work of depth graph networks in brain connection research was done by Ktena et al. (2018), and then this work was further extended about the way of cross-validation (Mirakhorli and Mirakhorli, 2019). **Figure 9** shows the process for the conversion and analysis of functional magnetic resonance imaging (fMRI) in the study of Ktena et al. which has been used in some later research of GNNs for brain connection. Zhai et al. (2019) constructed a generative model using graph convolution and VAE to predict the abnormal parts of input graphs. In order to lift the restrictions of fixed graphic structure for a model, Zhang and Pierre (2020) proposed a spatial GCN-based learning model, which could perform the classification of different brain connections and help predict the association between brain connection sub-networks and disease. Parisot et al. (2017) proposed an effective method to annotate human brain activity and then conducted a study on the transferability of brain decoding (Stankevičiūtė et al., 2020). Yang et al. (2019) predicted the state of consciousness of the brain by constructing a dynamic functional connectivity matrix that describes the state of the brain, which proved the effectiveness of graph convolution methods in predicting cortical signals.

Multimodal Fusion

All the studies discussed above are based on the analysis of a single image. Nevertheless, using only single-modal imaging data in disease prediction may lead to a lack of precision in the results. Yang et al. (2019) used the weighted edge-weighted graph attention network model to combine different modalities of medical imaging (brain structural magnetic resonance imaging and fMRI) for identifying bipolar disorder. In addition, more research tends to combine medical imaging data with non-imaging data in disease prediction. Based on multimodal fusion, the lack of some important information in monomodal data can be optimized and complemented to a certain extent. Parisot et al. (2017) introduced GCN into group-level medical applications for the first time. In this work, they proposed a population graph, which modeled people as nodes in the graph, and the feature vector of brain image was used as the feature representation of the node. Phenotypic data were combined for disease prediction in a semi-supervised way, such as gender, age, etc. In 2018, Parisot et al. extended this work with further in-depth analysis methods and modeling options (Stankevičiūtė et al., 2020). Since then, the use of population graph methods for disease prediction has become the choice of many researchers. Kazi et al. (2018) used a multi-level parallel GCN model to optimize the extraction of correlation information between nodes, which introduced an automatic learning layer for weight distribution and the attention mechanism for utilizing the features of each multimodal data (Kazi et al., 2019c). Aiming at the problem of insufficient feature extraction caused by fixed neighborhoods in GCN model, InceptionGCN (Kazi et al., 2019a) was proposed by Kazi et al., which considered the receptive field convolution kernels of different dimensions and utilized two aggregation methods to process all the features obtained by a convolution kernel.

Nevertheless, the performance of this model on various datasets is quite different, so the LSTM-based attention mechanism was introduced in later work to better integrate multi-modal data (Kazi et al., 2019b). For the research of autism spectrum disorder classification problem, Anirudh and Thiagarajan (2019) proposed a more robust method, which was based on the set of weakly trained G-CNN for reducing the model sensitivity to the choice of graph construction. Arya et al. (2020) believed that population phenotypic data is not suitable for defining the relationship between edges; then, the actual similarities between the brain's structures were used to directly extract the variables from brain MRI so as to establish the relationship between nodes. Stankevičiūtė et al. (2020) used a population graph to predict brain age but obtained unsatisfactory results. The topological characteristics between the various regions of the brain make GNNs more suitable for identifying related patterns of brain disease and effectively assisting in exploring the mechanism of the disease.

SUMMARY ON TYPICAL APPLICATIONS

In general, the applications of GNNs in node level, edge level, and graph level are not completely independent. According to the different levels of tasks and biological problems, we give a detailed summary on the above-mentioned typical applications in **Table 1**.

DISCUSSION AND FUTURE RESEARCH DIRECTIONS

The Problem and Trend of Methodology

Current GNNs have room for improvement in the methods of processing biological tasks. This section proposes the methodological problems and future development directions of GNNs for the three application fields of disease prediction, drug discovery, and biomedical imaging.

For disease prediction, improvements need to be made in three areas: the similarity evaluation of new nodes, the introduction of node attribute information, and heterogeneous information processing. First, most of the research in disease prediction adopted broad similarity methods. GNN models are used to extract the in-depth information of a heterogeneous network composed of disease semantic similarity, RNA functional similarity, and multiple association data. However, the construction of various similarity networks would have increased the complexity of the GNN model to a certain extent, and an efficient similarity evaluation paradigm needs to be improved for new diseases or RNA. In addition, more attention should be paid to the introduction of node attribute information into the modeling process, such as disease semantic features and RNA structural features, which can avoid the excessive dependence on associated information. Finally, for heterogeneous networks containing multi-source information, GNNs can deeply integrate their topological information. Nevertheless, current GNNs mainly focus on the processing of isomorphic graphs and cannot sufficiently capture the heterogeneity of nodes and edges in

TABLE 1 | Summary of typical applications.

Classification	Biological problems	Publications	Task of graph neural networks		
			Node level	Edge level	Graph level
Disease association prediction	RNA–disease association prediction	Wang L. et al., 2020	✓		
		Li C. et al., 2019; Zhang J. et al., 2019; Ding et al., 2020; Li J. et al., 2020; Wu et al., 2020; Zheng et al., 2020		✓	
	Inter-cell interaction prediction	Wang et al., 2021		✓	
	scRNA-Seq clustering	Zeng et al., 2020	✓		
	Impute the dropout events in scRNA-Seq data	Jiahua et al., 2020		✓	
	Simulate cell differentiation	Bica et al., 2020			✓
	Disease state prediction	Ravindra et al., 2020	✓		
	Disease genetic prioritization	Rao et al., 2018		✓	
		Li Y. et al., 2019		✓	
	Disease–gene association prediction	Singh and Lio, 2019		✓	✓
		Wang et al., 2019a		✓	
		Han et al., 2019			✓
	Cancer genetic judgment	Iglehart and Silver, 2009	✓		
	Synthetic lethality prediction	Cai et al., 2020		✓	
	Breast cancer metastasis prediction	Chereda et al., 2021			✓
	Breast cancer subtype classification	Rhee et al., 2018		✓	
	Disease protein judgment	Eyuboglu and Freeman, 2004	✓		
	Discussion on the relationship between drug and disease	Bajaj et al., 2017		✓	
	RNA classification	Rossi et al., 2019			✓
	Predict the binding site of RNA protein	Uhl et al., 2019			✓
		Yan et al., 2020			✓
	RNA interaction prediction	Huang et al., 2019			✓
	Targeted gene prediction of lncRNA	Zhao et al., 2020			✓
	Biomedical data classification	Wang T. et al., 2020			✓
Drug development and discovery	Protein structure prediction	Zamora-Resendiz and Crivelli, 2019	✓		
	Protein design	Ingraham et al., 2019; Strokach et al., 2020			✓
	Protein function prediction	Gligorijevic et al., 2019			✓
		Fout et al., 2017		✓	
	Protein functional sites prediction	Toomer, 2020	✓		
	Interface prediction between protein pairs	Yao et al., 2020			✓
	Protein–protein interaction network denoising	Liu X. et al., 2020		✓	
	Prediction of the influence of protein mutation on binding	Cao and Shen, 2020	✓		
	Protein docking model evaluation	Johansson-Åkhe et al., 2020			✓
	Assessment of docked peptide conformations	Feng et al., 2018			✓
	Drug–target interaction prediction	Gao et al., 2018; Miyazaki et al., 2020; Nguyen et al., 2021		✓	
		Neil et al., 2018; Torng and Altman, 2019; Jiang et al., 2020; Zhong et al., 2020; Zhao et al., 2021			✓
	Molecular structure coding	Duvenaud et al., 2015; Kearnes et al., 2016			✓
	Drug properties prediction	Liu et al., 2019			✓
	Esp surface detection of ligands	Rathi et al., 2019	✓		
	Compound–protein interaction prediction	Li S. et al., 2020		✓	
		Tsubaki et al., 2018			✓
	Compound stability prediction	Li et al., 2019b			✓
	De novo molecule design	Dai et al., 2018; Li et al., 2018a,b; You et al., 2018; Khemchandani et al., 2020			✓
	Prediction of the association between miRNA and drug resistance	Huang et al., 2020		✓	
	Prediction of the effect of drugs on cancer cell growth	Liu Q. et al., 2020; Singha et al., 2020			✓

(Continued)

TABLE 1 | Continued

Classification	Biological problems	Publications	Task of graph neural networks		
			Node level	Edge level	Graph level
	Prediction of drug-induced liver damage	Hwang et al., 2020			√
	Side effects prediction between drugs	Zitnik et al., 2018		√	
	Drug similarity prediction	Ma et al., 2018		√	
	Drug combination design	Long et al., 2020		√	
	Drug recommendation	Mao et al., 2019		√	
	Alkaloid classification	Eguchi et al., 2019			√
	Product prediction of organic reactions	Coley et al., 2019			√
	Chemical network prediction	Li Q. et al., 2018		√	
Medical image processing	Image segmentation	Gopinath et al., 2019; Wang et al., 2019b; Tian et al., 2020a,b	√		
	Image classification	Shi et al., 2019; Zhou et al., 2019; Adnan et al., 2020	√		
	abnormal detection	Wu et al., 2019	√		
	Image visualization	Levy et al., 2020; Sureka et al., 2020			√
	Image enhancement	Hu et al., 2020	√		
	Image registration	Hansen et al., 2019	√		
	Image retrieval	Zhai et al., 2019	√		
	Surgical image analysis	Zhang et al., 2018	√		
	Disease prediction	Parisot et al., 2017; Kazi et al., 2018, 2019a,b,c; Anirudh and Thiagarajan, 2019; Yang et al., 2019; Arya et al., 2020; Stankevičiūtė et al., 2020		√	
	Brain connection research	Ktena et al., 2017, 2018; Li X. et al., 2019a; Mirakhorli and Mirakhorli, 2019; Grigis et al., 2020; Zhang and Pierre, 2020; Zhang et al., 2021		√	
		Li et al., 2020		√	√

heterogeneous networks (Zhang C. et al., 2019). So, a new architecture needs to be studied, which can consider the feature of data in heterogeneous biological networks.

In drug discovery, the construction mode of chemical networks and the definition of molecular model structure need to be further explored. In the study of compound interactions, compounds and chemical networks are usually modeled as graphs. These graph-based methods have been successfully applied to the related tasks of chemical networks, but there are few studies that can simultaneously consider these two different types of graphs in an end-to-end manner. Harada et al. (2020) used molecular graphs as nodes in chemical networks and performed internal and outer convolution operations on them. The dual graph convolutional network can capture the feature of the individual molecular graph structure and the molecular relationship network simultaneously, making excellent results in dense networks. In addition, current molecular modeling is based heavily on the 2D graph structure, and the 3D structure that may affect the properties of molecules has rarely been considered. Therefore, the research of GNNs on the molecular 3D structure may be a future direction that has been neglected previously.

Multimodal Fusion

Graph neural networks also have limitations in multimodal data processing in medical imaging. The natural combination of multimodal deep learning (Ngiam et al., 2011) and multi-source omics data accelerates the development of bioinformatics. Some studies of GNNs in the multimodal fusion have been discussed in Section 4.3.3. It is not uncommon to find that the data is relatively balanced in these studies, but the involvement of unbalanced data cannot be avoided in actual tasks. Therefore, the method of processing unbalanced data in GNNs needs further research.

The Problem and Trend Caused by Biological Data

Existing research on biomolecular networks has proved that many biomolecular networks have the properties of sparseness and scale-free nature. Sparseness is expressed as if the network size is N, then the number of edges is $O(N)$ instead of N^2, which results from a particular optimization in the long-term evolution of organisms. The scale-free nature is reflected in the degree distribution of these networks that obeys the power-law distribution, where most nodes have a small number of connections and a few nodes have a large number of connections. This characteristic demonstrates that a few nodes represented as biomolecules play a key role in the dynamic changes of biomolecular networks. For the problems of sparseness and scale-free nature of biomolecular networks, two methods—dropout and regularization—can be adopted to alleviate the overfitting caused by them. With a fixed probability, the dropout method randomly sets each dimension of the weight to zero during the training process so that the model only updates part of the parameters each time. As an example, the GCN model proposed by Cai et al. (2020) is based on the methods of fine-grained edge dropout and coarse-grained node dropout to making GCN learn a more stable representation in the process of continuous adaptation. Dropout can alleviate the instability when there is

a fantastic amount of data in the training set; consequently, the dropout method is better suited to large data sets. The regularization method adds a regularization term to the loss function to limit the scale of parameters. In the study of disease–gene association prediction by Han et al. (2019), there were 3,209 diseases and more than 10,000 genes in the data set. Nevertheless, only 3,954 known disease–gene associations and a margin control loss function were defined to reduce sparse impact.

In addition, the collection of negative sample data is often ignored in the most current research, which leads to the fact that the biological data only contains positive samples. The lack of negative samples increases the difficulty of model training. Therefore, in the study of Eyuboglu and Freeman (2004), a simple method was proposed to solve the lack of negative labels, which randomly select k from the set of unlabeled nodes to mark them as negative. This seemingly unreasonable method causes most tags to be randomly selected, but in fact, randomly selected proteins may have a low correlation with disease. Hence, a certain number of random negative labels can be considered reasonable.

Finally, there is a large amount of noise information in the biomolecular network, so noise reduction processing is a very positive step for improving the model's performance. For example, GAT can assign low weight to noisy data or directly eliminate the associations with correlation degrees lower than the threshold in the network. More methods for reducing noise are worth further development.

The Lack of Interpretability

In bioinformatics, simply providing the computing results is just not far enough. The lack of interpretation is a persistent problem of the black box model like deep learning. As the entities and relationships in GNNs usually correspond to various types of objects that exist in the real world, then GNNs have abilities to support more interpretable analysis and visualization (Selsam et al., 2018). Take learning molecular fingerprint (Duvenaud et al., 2015) as an example; the fingerprint encoding method using neural graphs can take into account the similarity between molecular fragments to achieve a more meaningful feature representation, which is also ignored in traditional fingerprint encoding. In the prediction of metastasis for breast cancer patients (Chereda et al., 2021), the graph layer-wise relevance propagation was proposed to explain how GCN generates predictions based on patient-specific PPI sub-network data which could be potentially highly useful for the development of personalized medicine. In histological image analysis, Sureka et al. (2020) modeled histological tissue as a nuclear graph and established a graph convolutional network framework based on attention mechanism and node occlusion for disease diagnosis. This method visualized the relative contribution of each cell nucleus by a whole-slide image. In data analysis, GNNs generate relevant information for each data node, which makes the model more interpretative to some extent. Overall, the further exploration of interpretability in modeling biological networks by GNNs is still an essential direction of future research.

The Exploration of Deep Structure

The deep network structure is more common in deep learning. For instance, a residual network (ResNet) that excels in image classification has 152 layers (He et al., 2016), but the layers of most networks are below three in the field of GNNs (Zhou et al., 2020). Experiments have shown that, as the number of network layers increases, the characteristics of all nodes will approach the same value, which will reduce network performance (Li Q. et al., 2018). However, deeper networks can provide larger parameter space and stronger representation capabilities, so the feasibility of a deep graph neural network deserves to be explored.

CONCLUSION

Graph neural networks, as a branch of deep learning in non-Euclidean space, perform particularly well in various tasks that process graph structure data. In this paper, a systematic survey of GNNs and their advances in bioinformatics is presented from multiple perspectives. Three representative tasks are especially proposed based on the three levels of structural information that can be learned by GNNs: node classification, link prediction, and graph generation. Meanwhile, according to the specific applications for various omics data, we categorize and discuss the related studies in three aspects: disease prediction, drug discovery, and biomedical imaging. Finally, the limitations and future possibilities of applying GNNs to bioinformatics studies are illustrated.

Although GNN has achieved excellent results in many biological tasks at present, it still faces challenges in terms of low-quality data processing, methodology, and interpretability and has a long road ahead. We believe that GNNs are potentially a wonderful method that solves various biological problems in bioinformatics research. Furthermore, this paper can provide a valuable reference for new researchers joining the studies in this area.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/supplementary material, further inquiries can be directed to the corresponding author/s.

AUTHOR CONTRIBUTIONS

LLia and X-MZ contributed to conceptualization, methodology, and writing – original draft preparation. LLiu and M-JT contributed to formal analysis and writing – review and editing. X-MZ contributed to the investigation and data curation. All authors have read and agreed to the published version of the manuscript.

FUNDING

REFERENCES

Adnan, M., Kalra, S., and Tizhoosh, H. R. (2020). "Representation learning," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, Seattle. 988–989.

Anfinsen, C. (1973). Principles that govern the folding of protein chains. *Science* 181, 223–230.

Anirudh, R., and Thiagarajan, J. (2019). "Bootstrapping graph convolutional neural networks for autism spectrum disorder classification," in *Proceedings of the ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, (Brighton: IEEE).

Arya, D., Olij, R., Gupta, D. K., El Gazzar, A., Wingen, G., Worring, M., et al. (2020). "Fusing structural and functional MRIs using graph convolutional networks for autism classification," in *Medical Imaging With Deep Learning*, (Piscataway, NJ: PMLR), 44–61.

Atwood, J., and Towsley, D. (2015). Diffusion-convolutional neural networks. *NIPS* 2016, 1993–2001.

Bajaj, P., Heereguppe, S., and Sumanth, C. (2017). Graph convolutional networks to explore drug and disease relationships in biological networks. *Comput. Sci.*

Barabasi, A., Gulbahce, N., and Loscalzo, J. (2011). Network Medicine: a network-based approach to human disease. *Nat. Rev. Genet.* 12, 56–68. doi: 10.1038/nrg2918

Bica, I., Andres Terre, H., Cvejic, A., and Lio, P. (2020). Unsupervised generative and graph representation learning for modelling cell differentiation. *Sci. Rep.* 10:9790. doi: 10.1038/s41598-020-66166-8

Bojchevski, A., and Günnemann, S. (2017). Deep gaussian embedding of graphs: Unsupervised inductive learning via ranking. *arXiv* [Preprint]. arXiv:1707.03815.

Bruna, J., Zaremba, W., Szlam, A., and LeCun, Y. (2013). Spectral networks and locally connected networks on graphs. *arXiv* [Preprint]. arXiv:1312.6203.

Cai, R., Chen, X., Fang, Y., Wu, M., and Hao, Y. (2020). Dual-dropout graph convolutional network for predicting synthetic lethality in human cancers. *Bioinformatics* 2020:btaa211. doi: 10.1093/bioinformatics/btaa211

Cao, Y., and Shen, X. (2020). Energy-based graph convolutional networks for scoring protein docking models. *Prot. Struct. Funct. Bioinform.* 88, 1091–1099.

Chereda, H., Bleckmann, A., Menck, K., Perera-Bel, J., Stegmaier, P., Auer, F., et al. (2021). Explaining decisions of graph convolutional neural networks: patient-specific molecular subnetworks responsible for metastasis prediction in breast cancer. *Genome Med.* 13:845. doi: 10.1186/s13073-021-00845-7

Coley, C., Jin, W., Rogers, L., Jamison, T., Green, W., Jaakkola, T., et al. (2019). A graph-convolutional neural network model for the prediction of chemical reactivity. *Chem. Sci.* 10:4228D. doi: 10.1039/C8SC04228D

Dai, H., Tian, Y., Dai, B., Skiena, S., and Song, L. (2018). Syntax-directed variational autoencoder for structured data. *arXiv* [Preprint]. arXiv:1802.08786.

De Cao, N., and Kipf, T. (2018). MolGAN: An implicit generative model for small molecular graphs. *arXiv* [Preprint]. arXiv:1805.11973.

Defferrard, M., Bresson, X., and Vandergheynst, P. (2016). Convolutional neural networks on graphs with fast localized spectral filtering. *Adv. Neur. Inform. Proces. Syst. Barcelona NIPS* 2016, 3844–3852.

Ding, Y., Tian, L. P., Lei, X., Liao, B., and Wu, F. X. (2020). Variational graph auto-encoders for miRNA-disease association prediction. *Methods* 2:4. doi: 10.1016/j.ymeth.2020.08.004

Duvenaud, D., Maclaurin, D., Aguilera-Iparraguirre, J., Gómez-Bombarelli, R., Hirzel, T., Aspuru-Guzik, A., et al. (2015). "Convolutional Networks on graphs for learning molecular fingerprints," in *Advances in Neural Information Processing Systems (NIPS)*, Montreal, QC, 13.

Eguchi, R., Ono, N., Morita, A., Katsuragi, T., Nakamura, S., Huang, M., et al. (2019). Classification of alkaloids according to the starting substances of their biosynthetic pathways using graph convolutional neural networks. *BMC Bioinform.* 20:2963. doi: 10.1186/s12859-019-2963-6

Eyuboglu, E. S., and Freeman, P. B. (2004). Disease protein prediction with graph convolutional networks. *Genetics* 5, 101–113.

Feng, Q., Dueva, E., Cherkasov, A., and Ester, M. (2018). Padme: A deep learning-based framework for drug-target interaction prediction. *arXiv* [Preprint]. arXiv:1807.09741. doi: 10.1007/s12559-021-09840-x

Fout, A. M., Jonathon, B., Basir, S., and Asa, B. (2017). Protein Interface prediction using graph convolutional networks. *NIPS* 6533–6542.

Frasca, M. (2017). Gene2DisCo: gene to disease using disease commonalities. *Artif. Intel. Med.* 82:1. doi: 10.1016/j.artmed.2017.08.001

Gao, K. Y., Fokoue, A., Luo, H., Iyengar, A., Dey, S., and Zhang, P. (2018). Interpretable drug target prediction using deep neural representation. *IJCAI* 2018, 3371–3377.

Gilmer, J., Schoenholz, S. S., Riley, P. F., Vinyals, O., and Dahl, G. E. (2017). "Neural message passing for quantum chemistry," in *Proceedings of the International Conference on Machine Learning*, (Piscataway, NJ: PMLR), 1263–1272.

Gligorijevic, V., Renfrew, P., Kosciolek, T., Koehler, J., Cho, K., Vatanen, T., et al. (2019). Structure-based function prediction using graph convolutional networks. *Nat. Commun.* 12, 1–14.

Goh, K., Cusick, M., Valle, D., Childs, B., Vidal, M., and Barabasi, A. (2007). The human disease network. *Proc. Natl. Acad. Sci. U.S.A.* 104, 8685–8690. doi: 10.1073/pnas.0701361104

Gopinath, K., Desrosiers, C., and Lombaert, H. (2019). Graph convolutions on spectral embeddings for cortical surface parcellation. *Med. Image Anal.* 54, 297–305. doi: 10.1016/j.media.2019.03.012

Gori, M., Monfardini, G., and Scarselli, F. (2005). "A new model for learning in graph domains," in *Proceedings of the 2005 IEEE International Joint Conference on Neural Networks*, Vol. 2, (Piscataway, NJ: IEEE), 729–734. doi: 10.1109/IJCNN.2005.1555942

Grigis, A., Tasserie, J., Frouin, V., Jarraya, B., and Uhrig, L. (2020). Predicting cortical signatures of consciousness using dynamic functional connectivity graph-convolutional neural networks. *bioRxiv* [Preprint]. doi: 10.1101/2020.05.11.078535

Guan, Y., Gorenshteyn, D., Burmeister, M., Wong, A. K., Schimenti, J. C., Handel, M. A., et al. (2012). Tissue-specific functional networks for prioritizing phenotype and disease genes. *PLoS Comput. Biol.* 8:e1002694. doi: 10.1371/journal.pcbi.1002694

Hamilton, W., Ying, R., and Leskovec, J. (2017). Inductive representation learning on large graphs. *NIPS* 2017, 1024–1034.

Han, P., Yang, P., Zhao, P., Shang, S., Liu, Y., Zhou, J., et al. (2019). "GCN-MF: disease-gene association identification by graph convolutional networks and matrix factorization," in *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, Anchorage. 705–713.

Hansen, L., Dittmer, D., and Heinrich, M. P. (2019). "Learning deformable point set registration with regularized dynamic graph cnns for large lung motion in copd patients," in *Proceedings of the International Workshop on Graph Learning in Medical Imaging*, (Cham: Springer), 53–61.

Harada, S., Akita, H., Tsubaki, M., Baba, Y., Takigawa, I., Yamanishi, Y., et al. (2020). Dual graph convolutional neural network for predicting chemical networks. *BMC Bioinform.* 21:3378. doi: 10.1186/s12859-020-3378-0

He, K., Zhang, X., Ren, S., and Sun, J. (2016). "Deep residual learning for image recognition," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, New Orleans. 770–778.

Hu, X., Yan, Y., Ren, W., Li, H., Zhao, Y., Bayat, A., et al. (2020). Feedback graph attention convolutional network for medical image enhancement. *arXiv* [Preprint]. arXiv:2006.13863.

Huang, Y., Huang, Z., You, Z., Zhu, Z., Huang, W., Guo, J., et al. (2019). Predicting lncRNA-miRNA interaction via graph convolution auto-encoder. *Front. Genet.* 10:758. doi: 10.3389/fgene.2019.00758

Huang, Y. A., Hu, P., Chan, K. C., and You, Z. H. (2020). Graph convolution for predicting associations between miRNA and drug resistance. *Bioinformatics* 36, 851–858. doi: 10.1093/bioinformatics/btz621

Hwang, D., Jeon, M., and Kang, J. (2020). "A drug-induced liver injury prediction model using transcriptional response data with graph neural network," in *Proceedings of the 2020 IEEE International Conference on Big Data and Smart Computing (BigComp)*, (Piscataway, NJ: IEEE), 323–329.

Iglehart, J. D., and Silver, D. P. (2009). Synthetic lethality – A new direction in cancer drug development. *N. Engl. J. Med.* 361:189. doi: 10.1056/NEJMe0903044

Ingraham, J., Garg, V., Barzilay, R., and Jaakkola, T. (2019). *Generative Models for Graph-Based Protein Design. Advances in Neural Information Processing Systems 32 (NeurIPS 2019), December 2019, Vancouver, Canada, Neural Information Processing Systems Foundation, 2019*. Vancouver, BC: Neural Information Processing Systems Foundation.

Ioannidis, V. N., Marques, A. G., and Giannakis, G. B. (2019). "Graph neural networks for predicting protein functions," in *Proceedings of the 2019 IEEE 8th*

International Workshop on Computational Advances in Multi-Sensor Adaptive Processing (CAMSAP), (Piscataway, NJ: IEEE), 221–225.

Jiahua, R., Zhou, X., Lu, Y., Zhao, H., and Yang, Y. (2020). Imputing single-cell RNA-seq data by combining graph convolution and autoencoder neural networks. *iScience* 24:102393. doi: 10.1016/j.isci.2021.102393

Jiang, L., and Zhu, J. (2020). Review of MiRNA-disease association prediction. *Curr. Prot. Pept. Sci.* 21, 1044–1053.

Jiang, M., Li, Z., Zhang, S., Wang, S., Wang, X., Yuan, Q., et al. (2020). Drug-target affinity prediction using graph neural network and contact maps. *RSC Adv.* 10, 20701–20712. doi: 10.1039/D0RA02297G

Jin, W., Barzilay, R., and Jaakkola, T. (2018). "Junction tree variational autoencoder for molecular graph generation," in *Proceedings of the International Conference on Machine Learning*, (Piscataway, NJ: PMLR), 2323–2332. doi: 10.1186/s13321-019-0396-x

Johansson-Åkhe, I., Mirabello, C., and Wallner, B. (2020). InterPepRank: assessment of docked peptide conformations by a deep graph network. *bioRxiv* [Preprint]. doi: 10.1101/2020.09.07.285957

Kazi, A., Albarqouni, S., Kortüm, K., and Navab, N. (2018). Multi layered-parallel graph convolutional network (ML-PGCN) for disease prediction. *arXiv* [Preprint]. arXiv:1804.

Kazi, A., Shekarforoush, S., Sridhar, A., Burwinkel, H., Vivar, G., Kortüm, K., et al. (2019a). "InceptionGCN: receptive field aware graph convolutional network for disease prediction," in *Proceedings of the International Conference on Information Processing in Medical Imaging*, (Cham: Springer), 73–85. doi: 10.1007/978-3-030-20351-1_6

Kazi, A., Shekarforoush, S., Sridhar, A., Burwinkel, H., Wiestler, B., et al. (2019b). "Graph convolution based attention model for personalized disease prediction," in *Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention*, (Cham: Springer), 122–130. doi: 10.1007/978-3-030-32251-9_14

Kazi, A., Sridhar, A., Shekarforoush, S., Kortüm, K., Albarqouni, S., and Navab, N. (2019c). "Self-attention equipped graph convolutions for disease prediction," in *Proceedings of the 2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019)*, (Venice: IEEE), 1896–1899. doi: 10.1109/ISBI.2019.8759274

Kearnes, S., McCloskey, K., Berndl, M., Pande, V., and Riley, P. (2016). Molecular graph convolutions: moving beyond fingerprints. *J. Comput. Aided Mol. Des.* 30, 595–608. doi: 10.1007/s10822-016-9938-8

Khemchandani, Y., O'Hagan, S., Samanta, S., Swainston, N., Roberts, T. J., Bollegala, D., et al. (2020). DeepGraphMolGen, a multi-objective, computational strategy for generating molecules with desirable properties: a graph convolution and reinforcement learning approach. *J. Cheminform.* 12, 1–17. doi: 10.1186/s13321-020-00454-3

Kingma, D. P., and Welling, M. (2013). Auto-encoding variational bayes. *arXiv* [Preprint]. arXiv:1312.6114. doi: 10.1093/bioinformatics/btaa169

Kipf, T. N., and Welling, M. (2016a). Semi-supervised classification with graph convolutional networks. *arXiv* [Preprint]. arXiv:1609.02907.

Kipf, T. N., and Welling, M. (2016b). Variational graph auto-encoders. *arXiv* [Preprint]. arXiv:1611.07308.

Koren, Y., Bell, R., and Volinsky, C. (2009). Matrix factorization techniques for recommender systems. *Computer* 42, 30–37.

Ktena, S. I., Parisot, S., Ferrante, E., Rajchl, M., Lee, M., Glocker, B., et al. (2018). Metric learning with spectral graph convolutions on brain connectivity networks. *NeuroImage* 169, 431–442. doi: 10.1016/j.neuroimage.2017.12.052

Ktena, S. I., Parisot, S., Ferrante, E., Rajchl, M., Lee, M., Glocker, B., et al. (2017). "Distance metric learning using graph convolutional networks: Application to functional brain networks," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*, (Cham: Springer), 469–477.

Lee, I., Blom, U. M., Wang, P. I., Shim, J. E., and Marcotte, E. M. (2011). Prioritizing candidate disease genes by network-based boosting of genome-wide association data. *Genome Res.* 21, 1109–1121. doi: 10.1101/gr.118992.110

Levy, J., Haudenschild, C., Bar, C., Christensen, B., and Vaickus, L. (2020). Topological feature extraction and visualization of whole slide images using graph neural networks. *bioRxiv* [Preprint]. doi: 10.1101/2020.08.01.231639.

Li, C., Liu, H., Hu, Q., Que, J., and Yao, J. (2019). A novel computational model for predicting microRNA-disease associations based on heterogeneous graph convolutional networks. *Cells* 8:977. doi: 10.3390/cells8090977

Li, J., Zhang, S., Liu, T., Ning, C., Zhang, Z., and Zhou, W. (2020). Neural inductive matrix completion with graph convolutional networks for miRNA-disease association prediction. *Bioinformatics* 36, 2538–2546. doi: 10.1093/bioinformatics/btz965

Li, Q., Han, Z., and Wu, X. M. (2018). "Deeper insights into graph convolutional networks for semi-supervised learning," in *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 32.

Li, S., Wan, F., Shu, H., Jiang, T., Zhao, D., and Zeng, J. (2020). MONN: a multi-objective neural network for predicting compound-protein interactions and affinities. *Cell Syst.* 10, 308–322.e11. doi: 10.1016/j.cels.2020.03.002

Li, X., Dvornek, N. C., Zhou, Y., Zhuang, J., Ventola, P., and Duncan, J. S. (2019a). "Graph neural network for interpreting task-fmri biomarkers," in *Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention*, (Cham: New Orleans, Springer), 485–493. doi: 10.1007/978-3-030-32254-0_54

Li, X., Yan, X., Gu, Q., Zhou, H., Wu, D., and Xu, J. (2019b). DeepChemStable: chemical stability prediction with an attention-based graph convolution network. *J. Chem. Inform. Model.* 59, 1044–1049. doi: 10.1021/acs.jcim.8b00672

Li, X., Zhou, Y., Gao, S., Dvornek, N., Zhang, M., Zhuang, J., et al. (2020). Braingnn: Interpretable brain graph neural network for fmri analysis. *bioRxiv* [Preprint]. doi: 10.1101/2020.05.16.100057

Li, Y., and Li, J. (2012). Disease gene identification by random walk on multigraphs merging heterogeneous genomic and phenotype data. *BMC Genomics* 7(Suppl. 7):S27. doi: 10.1186/1471-2164-13-S7-S27

Li, Y., Kuwahara, H., Yang, P., Song, L., and Gao, X. (2019). *PGCN: Disease Gene Prioritization by Disease and Gene Embedding Through Graph Convolutional Neural Networks*. doi: 10.1101/532226

Li, Y., Vinyals, O., Dyer, C., Pascanu, R., and Battaglia, P. (2018a). Learning deep generative models of graphs. *arXiv* [Preprint]. arXiv:1803.03324.

Li, Y., Zhang, L., and Liu, Z. (2018b). Multi-objective *de novo* drug design with conditional graph generative model. *J. Cheminform.* 10, 1–24. doi: 10.1186/s13321-018-0287-6

Liu, K., Sun, X., Jia, L., Ma, J., Xing, H., Wu, J., et al. (2019). Chemi-Net: a molecular graph convolutional network for accurate drug property prediction. *Int. J. Mol. Sci.* 20:3389. doi: 10.3390/ijms20143389

Liu, Q., Hu, Z., Jiang, R., and Zhou, M. (2020). DeepCDR: a hybrid graph convolutional network for predicting cancer drug response. *Bioinformatics* 36(Suppl. 2), i911–i918. doi: 10.1093/bioinformatics/btaa822

Liu, X., Luo, Y., Song, S., and Peng, J. (2020). Pre-training of graph neural network for modeling effects of mutations on protein-protein binding affinity. *arXiv* [Preprint]. arXiv:2008.12473.

Long, Y., Wu, M., Kwoh, C., Luo, J., and li, X. (2020). Predicting human microbe-drug associations via graph convolutional network with conditional random field. *Bioinformatics* 36:598. doi: 10.1093/bioinformatics/btaa598

Luo, J., Ding, P., Liang, C., Cao, B., and Chen, X. (2016). Collective prediction of disease-associated miRNAs based on transduction learning. *IEEE/ACM Transact. Comput. Biol. Bioinform.* 14, 1468–1475. doi: 10.1109/TCBB.2016.2599866

Ma, T., Xiao, C., Zhou, J., and Wang, F. (2018). Drug similarity integration through attentive multi-view graph auto-encoders. *arXiv* [Preprint]. arXiv:1804.10850.

Mao, C., Yao, L., and Luo, Y. (2019). Medgcn: Graph convolutional networks for multiple medical tasks. *arXiv* [Preprint]. arXiv:1904.00326.

Min, S., Lee, B., and Yoon, S. (2016). Deep learning in bioinformatics. *Brief. Bioinform.* 18, 851–869. doi: 10.1093/bib/bbw068

Mirakhorli, J., and Mirakhorli, M. (2019). Graph-based method for anomaly detection in functional brain network using variational autoencoder. *bioRxiv* [Preprint]. 616367.

Miyazaki, Y., Ono, N., Huang, M., Altaf−Ul−Amin, M., and Kanaya, S. (2020). Comprehensive exploration of target−specific ligands using a graph convolution neural network. *Mol. Inform.* 39:1900095. doi: 10.1002/minf.201900095

Monti, F., Boscaini, D., Masci, J., Rodolà, E., Svoboda, J., and Bronstein, M. (2017). *Geometric Deep Learning on Graphs and Manifolds Using Mixture Model CNNs*. 5425–5434. doi: 10.1109/CVPR.2017.576

Navlakha, S., and Kingsford, C. (2010). The power of protein interaction networks for associating genes with diseases. *Bioinformatics (Oxf. Engl.)* 26, 1057–1063. doi: 10.1093/bioinformatics/btq076

Neil, D., Briody, J., Lacoste, A., Sim, A., Creed, P., and Saffari, A. (2018). Interpretable graph convolutional neural networks for inference on noisy knowledge graphs. *arXiv*. [Preprint]. arXiv:1812.00279.

Ngiam, J., Khosla, A., Kim, M., Nam, J., Lee, H., and Ng, A. Y. (2011). "Multimodal deep learning," in *ICML*.

Nguyen, T., Le, H., Quinn, T. P., Nguyen, T., Le, T. D., and Venkatesh, S. (2021). GraphDTA: Predicting drug–target binding affinity with graph neural networks. *Bioinformatics* 37, 1140–1147. doi: 10.1093/bioinformatics/btaa921

Pan, X., and Shen, H.-B. (2019). Inferring disease-associated MicroRNAs using semi-supervised multi-label graph convolutional networks. *iScience* 20:13. doi: 10.1016/j.isci.2019.09.013

Parisot, S., Ktena, S. I., Ferrante, E., Lee, M., Moreno, R. G., Glocker, B., et al. (2017). "Spectral graph convolutions for population-based disease prediction," in *Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention*, (Cham: Springer), 177–185.

Peng, N., Poon, H., Quirk, C., Toutanova, K., and Yih, W. (2017). Cross-sentence N-ary relation extraction with graph LSTMs. *Transact. Associat. Comput. Linguist.* 5:49. doi: 10.1162/tacl_a_00049

Rao, A., Vg, S., Joseph, T., Kotte, S., Sivadasan, N., and Srinivasan, R. (2018). Phenotype-driven gene prioritization for rare diseases using graph convolution on heterogeneous networks. *BMC Med. Genom.* 11:372. doi: 10.1186/s12920-018-0372-8

Rathi, P. C., Ludlow, R. F., and Verdonk, M. L. (2019). Practical high-quality electrostatic potential surfaces for drug discovery using a graph-convolutional deep neural network. *J. Med. Chem.* 63, 8778–8790. doi: 10.1021/acs.jmedchem.9b01129

Ravindra, N., Sehanobish, A., Pappalardo, J., Hafler, D., and Dijk, D. (2020). "Disease state prediction from single-cell data using graph attention networks," in *Proceedings of the ACM Conference on Health, Inference, and Learning (CHIL '20)*, (New York, NY: Association for Computing Machinery), 121–130. doi: 10.1145/3368555.3384449

Rhee, S., Seo, S., and Kim, S. (2018). Hybrid approach of relation network and localized graph convolutional filtering for breast cancer subtype classification. *arXiv* 3527–3534. [Preprint]. doi: 10.24963/ijcai.2018/490.

Rossi, E., Monti, F., Bronstein, M., and Lio, P. (2019). ncRNA Classification with Graph Convolutional Networks. *arXiv* [Preprint]. arXiv:1905.06515.

Scarselli, F., Gori, M., Tsoi, A. C., Hagenbuchner, M., and Monfardini, G. (2008). The graph neural network model. *IEEE Transact. Neur. Netw.* 20, 61–80. doi: 10.1109/TNN.2008.2005605

Schulte-Sasse, R., Budach, S., Hnisz, D., and Marsico, A. (2019). "Graph convolutional networks improve the prediction of cancer driver genes," in *Artificial Neural Networks and Machine Learning – ICANN 2019: Workshop and Special Sessions. ICANN 2019. Lecture Notes in Computer Science*, Vol. 11731, eds I. Tetko, V. Kùrková, P. Karpov, and F. Theis (Cham: Springer). doi: 10.1007/978-3-030-30493-5_60

Selsam, D., Lamm, M., Bünz, B., Liang, P., de Moura, L., and Dill, D. L. (2018). Learning a SAT solver from single-bit supervision. *arXiv* [Preprint]. arXiv:1802.03685.

Shi, J., Wang, R., Zheng, Y., Jiang, Z., and Yu, L. (2019). Graph convolutional networks for cervical cell classification. Cervical cell classification with graph convolutional network. *Comput. Methods Program. Biomed.* 198, 105807.

Simonovsky, M., and Komodakis, N. (2018). "Graphvae: towards generation of small graphs using variational autoencoders," in *Proceedings of the International Conference on Artificial Neural Networks*, (Cham: Springer), 412–422. doi: 10.1186/s12868-016-0283-6

Singh, V., and Lio, P. (2019). Towards probabilistic generative models harnessing graph neural networks for disease-gene prediction. *arXiv* [Preprint]. arXiv:1907.05628.

Singha, M., Pu, L., Busch, K., Wu, H. C., Ramanujam, J., and Brylinski, M. (2020). GraphGR: a graph neural network to predict the effect of pharmacotherapy on the cancer cell growth. *bioRxiv* [Preprint]. doi: 10.1101/2020.05.20.107458

Stankevičiūtė, K., Azevedo, T., Campbell, A., Bethlehem, R., and Lio, P. (2020). *Population GNNs for Brain Age Prediction*. doi: 10.1101/2020.06.26.172171

Strokach, A., Becerra, D., Corbi-Verge, C., Perez-Riba, A., and Kim, P. (2020). Fast and flexible protein design using deep graph neural networks. *Cell Syst.* 11:16. doi: 10.1016/j.cels.2020.08.016

Sun, J., Shi, H., Wang, Z., Zhang, C., Liu, L., Wang, L., et al. (2014). Inferring novel lncrna–disease associations based on a random walk model of a lncrna functional similarity network. *Mol. BioSyst.* 10:70608. doi: 10.1039/c3mb70608g

Sun, M., Zhao, S., Gilvary, C., Elemento, O., Zhou, J., and Wang, F. (2019). Graph convolutional networks for computational drug development and discovery. *Briefings Bioinform.* 21:bbz042. doi: 10.1093/bib/bbz042

Sureka, M., Patil, A., Anand, D., and Sethi, A. (2020). "Visualization for Histopathology Images using Graph Convolutional Neural Networks," in *Proceedings of the 2020 IEEE 20th International Conference on Bioinformatics and Bioengineering (BIBE)*, (Piscataway, NJ: IEEE), 331–335.

Tian, F., Gao, B., Cui, Q., Chen, E., and Liu, T.-Y. (2014). "Learning deep representations for graph clustering," in *Proceedings of the National Conference on Artificial Intelligence*, Piscataway, NJ, Vol. 2, 1293–1299.

Tian, Z., Li, X., Zheng, Y., Chen, Z., Shi, Z., Liu, L., et al. (2020a). Graph-convolutional-network-based interactive prostate segmentation in MR images. *Med. Phys.* 47, 4164–4176. doi: 10.1002/mp.14327

Tian, Z., Zheng, Y., Li, X., Du, S., and Xu, X. (2020b). Graph convolutional network based optic disc and cup segmentation on fundus images. *Biomed. Optics Exp.* 11, 3043–3057. doi: 10.1364/BOE.390056

Toomer, D. (2020). *Predicting Protein Functional Sites Through Deep Graph Convolutional Neural Networks on Atomic Point-Clouds*. Available online at: https://cs230.stanford.edu/ (accessed November 15, 2020).

Torng, W., and Altman, R. B. (2019). Graph convolutional neural networks for predicting drug-target interactions. *J. Chem. Inform. Model.* 59, 4131–4149.

Tsubaki, M., Tomii, K., and Sese, J. (2018). Compound-protein interaction prediction with end-to-end learning of neural networks for graphs and sequences. *Bioinformatics (Oxf. Engl.)* 35:535. doi: 10.1093/bioinformatics/bty535

Uhl, M., Tran Van, D., Heyl, F., and Backofen, R. (2019). GraphProt2: A novel deep learning-based method for predicting binding sites of RNA-binding proteins. *bioRxiv* [Preprint]. doi: 10.1101/850024

Veličković, P., Cucurull, G., Casanova, A., Romero, A., Lio, P., and Bengio, Y. (2017). Graph attention networks. *arXiv* [Preprint]. arXiv:1710.10903,

Vohora, D., and Singh, G. (2017). *Pharmaceutical Medicine and Translational Clinical Research*. Cambridge, MA: Academic Press.

Wang, D., Cui, P., and Zhu, W. (2018). *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. New York, NY: ACM, 1225–1234.

Wang, J., Ma, A., Chang, Y., Gong, J., Jiang, Y., Qi, R., et al. (2021). scGNN is a novel graph neural network framework for single-cell RNA-Seq analyses. *Nat. Commun.* 12, 1–11. doi: 10.1038/s41467-021-22197-x

Wang, L., You, Z.-H., Li, Y.-M., Zheng, K., and Huang, Y.-A. (2020). GCNCDA: a new method for predicting circRNA-disease associations based on graph convolutional network algorithm. *PLoS Comput. Biol.* 16:e1007568. doi: 10.1371/journal.pcbi.1007568

Wang, Q., Sun, M., Zhan, L., Thompson, P., Ji, S., and Zhou, J. (2017). "Multi-modality disease modeling via collective deep matrix factorization," in *Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, Halifax. 1155–1164.

Wang, T., Shao, W., Huang, Z., Tang, H., Ding, Z., and Huang, K. (2020). MORONET: multi-omics integration via graph convolutional networks for biomedical data classification. *bioRxiv* [Preprint]. doi: 10.1101/2020.07.02.184705

Wang, X., Gong, Y., Yi, J., and Zhang, W. (2019a). "Predicting gene-disease associations from the heterogeneous network using graph embedding," in *Proceedings of the 2019 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, (Piscataway, NJ: IEEE), 504–511.

Wang, X., Ye, Y., and Gupta, A. (2018). "Zero-shot recognition via semantic embeddings and knowledge graphs," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Salt Lake City. 6857–6866.

Wang, X., Zhang, L., Roth, H., Xu, D., and Xu, Z. (2019b). "Interactive 3D segmentation editing and refinement via gated graph neural networks," in *Graph Learning in Medical Imaging. GLMI 2019. Lecture Notes in Computer Science*, Vol. 11849, eds D. Zhang, L. Zhou, B. Jie, and M. Liu (Cham: Springer), doi: 10.1007/978-3-030-35817-4_2

Weininger, D. (1988). SMILES, a chemical language and information system. 1. Introduction to methodology and encoding rules. *J. Chem. Inform. Comput. Sci.* 28, 31–36.

Wood, M., and Hirst, J. (2004). "Recent applications of neural networks in bioinformatics. biological and artificial intelligence environments," in *Proceedings of the 15th Italian Workshop on Neural Nets, WIRN VIETRI 2004, Vietri sul Mare, Italy, 2004. DBLP*, Vietri sul Mare.

Wu, J., Zhong, J. X., Chen, E. Z., Zhang, J., Jay, J. Y., and Yu, L. (2019). "Weakly- and Semi-supervised Graph CNN for identifying basal cell carcinoma on pathological images," in *International Workshop on Graph Learning in Medical Imaging*, (Cham: Springer), 112–119.

Wu, X., Lan, W., Chen, Q., Dong, Y., Liu, J., and Peng, W. (2020). Inferring LncRNA-disease associations based on graph autoencoder matrix completion. *Comput. Biol. Chem.* 87:107282. doi: 10.1016/j.compbiolchem.2020.107282

Xuan, P., Pan, S., Zhang, T., Liu, Y., and Sun, H. (2019a). Graph Convolutional network and convolutional neural network based method for predicting lncRNA-disease associations. *Cells* 8:1012. doi: 10.3390/cells8091012

Xuan, P., Sun, H., Wang, X., Zhang, T., and Pan, S. (2019b). Inferring the disease-associated mirnas based on network representation learning and convolutional neural networks. *Int. J. Mol. Sci.* 20:3648. doi: 10.3390/ijms20153648

Yan, Z., Hamilton, W., and Blanchette, M. (2020). Graph neural representational learning of RNA secondary structures for predicting RNA-protein interactions. *Bioinformatics* 36, i276–i284. doi: 10.1093/bioinformatics/btaa456

Yang, H., Li, X., Wu, Y., Li, S., Lu, S., Duncan, J., et al. (2019). "Interpretable multimodality embedding of cerebral cortex using attention graph," in *Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention*, (Cham: Springer), 799–807.

Yao, H., Guan, J., and Liu, T. (2020). Denoising protein–protein interaction network via variational graph auto-encoder for protein complex detection. *J. Bioinform. Comput. Biol.* 18:2040010. doi: 10.1142/S0219720020400107

You, J., Liu, B., Ying, R., Pande, V., and Leskovec, J. (2018). Graph convolutional policy network for goal-directed molecular graph generation. *arXiv* [Preprint]. arXiv:1806.02473.

Zamora-Resendiz, R., and Crivelli, S. (2019). Structural learning of proteins using graph convolutional neural networks structural learning of proteins using graph convolutional neural networks. *bioRxiv* [Preprint]. doi: 10.1101/610444

Zeng, Y., Zhou, X., Jiahua, R., Lu, Y., and Yang, Y. (2020). "Accurately clustering single-cell rna-seq data by capturing structural relations between cells through graph convolutional network," in *Proceedings of the 2020 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, Vol. 2020, Piscataway, NJ, 519–522. doi: 10.1109/BIBM49941.2020.9313569

Zhai, Z., Staring, M., Zhou, X., Xie, Q., Xiao, X., Bakker, M. E., et al. (2019). "Linking convolutional neural networks with graph convolutional networks: application in pulmonary artery-vein separation," in *Proceedings of the International Workshop on Graph Learning in Medical Imaging*, (Cham: Springer), 36–43.

Zhang, C., Song, D., Huang, C., Swami, A., and Chawla, N. V. (2019). "Heterogeneous graph neural network," in *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, New Orleans. 793–803.

Zhang, J., Hu, X., Jiang, Z., Song, B., Quan, W., and Chen, Z. (2019). "Predicting disease-related RNA associations based on graph convolutional attention network," in *Proceedings of the 2019 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, San Diego, CA, 177–182. doi: 10.1109/BIBM47256.2019.8983191

Zhang, M., and Chen, Y. (2018). Link prediction based on graph neural networks. *arXiv* [Preprint]. arXiv:1802.09691.

Zhang, X., He, L., Chen, K., Luo, Y., Zhou, J., and Wang, F. (2018). "Multi-view graph convolutional network and its applications on neuroimage analysis

for parkinson's disease," in *Proceedings of the AMIA Annual Symposium Proceedings*, Vol. 2018, (Bethesda, MY: American Medical Informatics Association), 1147.

Zhang, Y., and Pierre, B. (2020). Transferability of brain decoding using graph convolutional networks. *bioRxiv* [Preprint]. doi: 10.1101/2020.06.21.163964 doi: 10.1016/j.compmedimag.2020.101748

Zhang, Y., Tetrel, L., Thirion, B., and Bellec, P. (2021). Functional annotation of human cognitive states using deep graph convolution. *NeuroImage* 231:117847. doi: 10.1016/j.neuroimage.2021.117847

Zhao, T., Hu, Y., Peng, J., and Cheng, L. (2020). GCN-CNN: a novel deep learning method for prioritizing lncRNA target genes. *Bioinformatics (Oxf. Engl.)* 36:428. doi: 10.1093/bioinformatics/btaa428

Zhao, T., Hu, Y., Valsdottir, L. R., Zang, T., and Peng, J. (2021). Identifying drug–target interactions based on graph convolutional network and deep neural network. *Brief. Bioinform.* 22, 2141–2150. doi: 10.1093/bib/bbaa044

Zheng, K., You, Z. H., Wang, L., Wong, L., and Chen, Z. H. (2020). "Inferring disease-associated Piwi-interacting RNAs via graph attention networks," in *Proceedings of the International Conference on Intelligent Computing*, (Cham: Springer), 239–250.

Zhong, F., Wu, X., Li, X., Wan, D., Zunyun, F., Liu, X., et al. (2020). Computational target fishing by mining transcriptional data using a novel Siamese spectral-based graph convolutional network. *bioRxiv* [Preprint]. doi: 10.1101/2020.04.01.019166

Zhou, H., and Skolnick, J. (2016). A knowledge-based approach for predicting gene–disease associations. *Bioinformatics* 32, 2831–2838. doi: 10.1093/bioinformatics/btw358

Zhou, J., Cui, G., Hu, S., Zhang, Z., Yang, C., Liu, Z., et al. (2020). Graph neural networks: a review of methods and applications. *AI Open* 1, 57–81. doi: 10.1016/j.aiopen.2021.01.001

Zhou, M., Wang, X., Li, J., Hao, D., Wang, Z., Shi, H., et al. (2015). Prioritizing candidate disease-related long non-coding RNAs by walking on the heterogeneous lncRNA and disease network. *Mol. BioSyst.* 11, 760–769. doi: 10.1039/c4mb00511b

Zhou, Y., Graham, S., Alemi Koohbanani, N., Shaban, M., Heng, P. A., and Rajpoot, N. (2019). "Cgc-net: Cell graph convolutional network for grading of colorectal cancer histology images," in *Proceedings of the IEEE/CVF International Conference on Computer Vision Workshops*, Seoul. doi: 10.1109/TMI.2020.2971006

Zitnik, M., Agrawal, M., and Leskovec, J. (2018). Modeling polypharmacy side effects with graph convolutional networks. *Bioinformatics* 34, i457–i466.

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.